Monday, 8:50-10:00

■ MA-01

Monday, 8:50-10:00 - Room: B100/1001

Plenary 1

Stream: Plenaries
Invited session
Chair: *Selin Ahipasaoglu*

1 - Constrained Optimization via Frank-Wolfe Algorithms
Sebastian Pokutta

This talk focuses on constrained optimization problems that can be efficiently solved using first-order methods, particularly Frank-Wolfe methods (also known as Conditional Gradients). These algorithms have emerged as a crucial class of methods for minimizing smooth convex functions over polytopes, and their applicability extends beyond this domain. Recently, they have garnered significant attention due to their ability to facilitate structured optimization, a key aspect in some machine learning applications. I will provide a broad overview of these methods, highlighting their applications and presenting recent advances in both traditional optimization and machine learning. Time permitting, I will also discuss an extension of these methods to the mixed-integer convex optimization setting.

Monday, 10:30-12:30

■ MB-01

Monday, 10:30-12:30 - Room: B100/1001

Advances in Large-Scale Derivative-Free Optimization

Stream: Zeroth and first-order optimization methods
Invited session
Chair: *Francesco Rinaldi*
Chair: *Andrea Cristofari*

1 - Dimensionality reduction techniques for derivative free optimization
Coralia Cartis

We will discuss linear, and also, time permitting, nonlinear techniques for reducing the dimension of problems and algorithms in the absence of derivatives. We will discuss various aspects of subspace techniques in model-based optimization problems, both randomised and deterministic, first- and second-order, low-rank and full rank. Numerical experiments and theoretical results will be presented. Time permitting, we will present machine learning based dimensionality reduction techniques that we use in the context of Bayesian optimization methods.

2 - Solving 10,000-Dimensional Optimization Problems Using Inaccurate Function Values: An Old Algorithm
Zaikun Zhang

We reintroduce a derivative-free subspace optimization framework originating from Chapter 5 of [Z. Zhang, On Derivative-Free Optimization Methods, PhD thesis, Chinese Academy of Sciences, Beijing, 2012 (supervisor Ya-xiang Yuan)]. At each iteration, the framework defines a low-dimensional subspace based on an approximate gradient, and then solves a subproblem in this subspace to generate a new iterate. We sketch the global convergence and worst-case complexity analysis of the framework, elaborate on its implementation, and present some numerical results on problems with dimension as high as 10,000.

The same framework was presented by Zhang during ICCOPT 2013 in Lisbon under the title "A Derivative-Free Optimization Algorithm with Low-Dimensional Subspace Techniques for Large-Scale Problems", although it remained nearly unknown to the community until very recently. An algorithm following this framework named NEWUOAs was implemented by Zhang in MATLAB in 2011 (https://github.com/newuoas/newuoas), ported to Modula-3 in 2016 by M. Nystroem, a Principle Engineer at the Intel Corporation, and made available in the open-source package CM3 (https://github.com/modula3/cm3/blob/master/caltech-other/newuoa/src/NewUOAs.m3). NEWUOAs has been used by Intel in the design of chips, including its flagship product Atom P5900.

3 - On the computation of the cosine measure in high dimensions.
Scholar Sun

In derivative free optimization, the cosine measure is a value that quantifies the uniform density of a set of vectors. This value often arises in the convergence analysis of direct search methods, whereby choosing a set of search directions with a greater cosine measure can often yield better performance. Given the increasing interest in tackling high-dimensional DFO problems, it is valuable to be able to compute the cosine measure in this setting. However the cosine measure is computed as the solution to a minimax problem and has recently been shown to be NP-hard, making it difficult to scale into higher dimensions. We propose a new formulation of the problem and heuristic to tackle this problem in higher dimensions and compare it with existing algorithms in the literature. In addition, new theorems are presented to facilitate the construction of sets with specific cosine measures, allowing for the creation of a test-set to benchmark the algorithms with.

4 - A Novel Stochastic Derivative-Free Trust-Region Algorithm with Adaptive Sampling and Moving Ridge Functions
Benjamin Rees, Christine Currie, Vuong Phan

In this talk, we present a novel stochastic, model-based, derivative-free trust-region algorithm, ASTROMoRF. Within simulation optimization (SO) problems, evaluations of the objective function are inherently noisy and often computationally expensive to obtain; therefore, when solving SO problems, it is good practice to limit the number of calls made to the simulation model. ASTROMoRF ensures a stable finite-run performance through the implementation of a trust-region framework. It is also designed with high-dimensional SO problems in mind, using dimensionality-reduction techniques during the model construction stage of the algorithm to remove the model construction's dependency on the dimension of the problem. ASTROMoRF applies an adaptive sampling scheme when obtaining responses from the model, in order to ensure stochastic sampling error from Monte Carlo runs is kept in lock-step with the model bias. ASTROMoRF employs a variable projection method to construct the surrogate model and active subspace matrix, with certification and geometry improvement on the interpolation set to ensure that the constructed model is fully linear. The talk will present an overview of the algorithm and numerical results showcasing the solvability of ASTROMoRF against other solvers for a range of problems and will demonstrate the superiority of using local active subspace construction over global sensitivity analysis.

■ MB-02

Monday, 10:30-12:30 - Room: B100/7011

Optimization and applications

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Morteza Rahimi*
Chair: *Arghya Sinha*

1 - A Strongly Convex Framework for Image Reconstruction Using Generic Smoothing Filters
Arghya Sinha, Kunal N. Chaudhury

We study a class of regularized optimization problems related to image reconstruction, which are implicitly parametrized by a denoiser. We show that, with a specific class of linear denoisers, the associated objective functions exhibit strong convexity. Generally, strong convexity arises when either the loss function or the regularizer is strongly convex. We show these denoisers are proximable in a suitable inner product space and correspond to a nonsmooth, convex regularizer. Furthermore, even if neither the loss function nor the regularizer is strongly convex, we develop a sufficient condition that ensures their sum is strongly convex. However, to make nonsymmetric denoisers proximable, we must either symmetrize or alter the inner product space, which can increase time complexity and reduce reconstruction quality. To address this issue, we use sufficient conditions similar to those before to consider an alternative optimization problem that achieves better reconstructions in a shorter time frame without requiring the denoiser to be proximable. Specifically, we demonstrate that the inherent properties of nonnegativity and stochasticity can be used to create a strongly convex optimization framework that can be efficiently solved using conjugate gradient methods. In practice, this approach is significantly faster—by an order of magnitude—than splitting algorithms that use the same denoisers, and it produces improved reconstructions when nonsymmetric denoisers are applied.

2 - Random embeddings for global optimization: convergence results beyond isotropy
Roy Makhlouf, Estelle Massart

Many real-world optimization problems are high-dimensional, requiring dimensionality reduction techniques to solve them efficiently. Recently, the use of random embeddings was shown to substantially outperform classical methods for Lipschitz continuous objectives with special structure, such as functions with low effective dimension. Tools from conic integral geometry have been used to explore the benefits of random embeddings for global optimization of Lipschitz continuous objectives with no additional structure. These tools allow to derive lower bounds on the probability that a random linear subspace intersects a ball of approximate minimizers, by using the circular cone tangent to the ball. We aim here to extend these results to functions that vary very slowly along a low-dimensional subspace, for which we replace the ball of approximate minimizers with an ellipsoid to account for the anisotropic structure of the objective. Our findings offer deeper insights into how the anisotropic structure in high-dimensional functions impacts optimization algorithms.

3 - Optimization Techniques for Sparse and Structured Solutions in Big Data
Maurine Wafula, Leah Mutanu

In recent years, the explosion of data has created both incredible opportunities and significant challenges for the field of data analysis. To effectively analyze massive datasets, we need optimization techniques that are not only efficient but also help us understand the underlying patterns. This paper explores a powerful approach: focusing on solutions that are both sparse where, they only use a small subset of the available information and structured, reflecting the inherent organization of the data. The research will look into the theoretical foundations of these techniques, showcasing how they can be used to identify the most important features in a dataset while minimizing noise and irrelevant information. We then introduce novel optimization algorithms specifically designed for large-scale problems, ensuring they can handle the sheer volume of data we face today. These algorithms are not just about speed; they also guarantee reliable results and can be adapted to work on constantly evolving datasets. The research will demonstrate the practical value of these techniques through real-world applications in fields like machine learning, image processing, and bioinformatics. For example, in medicine, these methods can help identify crucial genetic markers associated with diseases, paving the way for more personalized treatments. This research bridges the gap between theoretical advancements and practical needs. By carefully balancing accuracy, computational efficiency, and

4 - Mapping Financial Well-Being: A Bayesian Spatial Clustering Analysis of Regional Economic Disparities
Lok-Yung Wong

Conventional methods of measuring financial well-being, such as composite indexes and econometric models, lack the adaptability to reflect regional and behavioural variability. This paper presents a Bayesian-augmented DBSCAN framework to quantify clustering uncertainty by combining Monte Carlo Dropout with Bayesian Optimisation for adaptive parameter modification. Applied to English financial well-being data, the approach produces latent clusters representing geographic and behavioural aspects of financial resilience, savings, and borrowing stress. Unlike segmentation depending on stationary demographic characteristics, this method assesses confidence in every assignment by adjusting to local data structures. The main contribution is in creating a clustering approach that combines uncertainty estimation with density-based segmentation, thereby enabling more accurate and understandable detection of financial activity patterns. This paradigm facilitates transferable applications in policy targeting and behavioural risk assessment and is intended to increase the methodological rigidity of financial well-being analysis.

■ MB-03

Monday, 10:30-12:30 - Room: B100/4011

First-order methods in modern optimization (Part I)

Stream: Large scale optimization: methods and algorithms
Invited session
Chair: *Simone Rebegoldi*
Chair: *Andrea Sebastiani*

1 - Learning firmly nonexpansive operators
Jonathan Chirinos Rodriguez, Emanuele Naldi, Kristian Bredies

In this talk, we propose a data-driven approach for constructing (firmly) nonexpansive operators. We demonstrate its applicability in Plug-and-Play (PnP) methods, where classical algorithms such as Forward-Backward splitting, Chambolle–Pock primal-dual iteration, Douglas–Rachford iteration or alternating directions method of multipliers (ADMM), are modified by replacing one proximal map by a learned firmly nonexpansive operator. We provide sound mathematical background to the problem of learning such an operator via expected and empirical risk minimization. We prove that, as the number of training points increases, the empirical risk minimization problem converges (in the sense of Gamma-convergence) to the expected risk minimization problem. Further, we derive a solution strategy that ensures firmly nonexpansive and piecewise affine operators within the convex envelope of the training set. We show that this operator converges to the best empirical solution as the number of points in the envelope increases in an appropriate way. Finally, the experimental section details practical implementations of the method and presents an application in image denoising, where we consider a novel, interpretable PnP Chambolle–Pock primal-dual iteration.

2 - Line search based stochastic gradient methods for learning applications
Federica Porta

Empirical risk minimization problems arise in a variety of applications, including machine learning and deep learning. Stochastic gradient algorithms are a standard approach to solve these problems due to their low computational cost per iteration and a relatively simple implementation. The aim of this talk is to present stochastic gradient algorithms based on Armiko-like line search to automatically select the learning rate, reducing the need for manual hyperparameter tuning, which can be computationally intensive. The line search can be combined with both acceleration techniques such as inertial term and/or variable metric and dynamical variance reduction procedures. Convergence properties of the resulting algorithms can be proved under the assumptions of both non-convex and convex objective functions. Numerical experiments on classification tasks illustrate the effectiveness and practical benefits of the proposed strategies.

3 - Convergence rates of regularized quasi-Newton methods without strong convexity
Shida Wang, Jalal Fadili, Peter Ochs

In this paper, we study convergence rates of the cubic regularized proximal quasi-Newton method (Cubic SR1) for solving non-smooth additive composite problems that satisfy the so-called Kurdyka-Lojasiewicz (KL) property with respect to some desingularization function rather than strong convexity. After a number of iterations, Cubic SR1 exhibits non-asymptotic explicit super-linear convergence rates if the exponent of KL property is beyond 1/2. For the special case, i.e. functions which satisfy Lojasiewicz inequality (PL), the rate becomes global and non-asymptotic. This work presents, for the first time, non-asymptotic explicit convergence rates of regularized (proximal) SR1 quasi-Newton methods applied to non-convex non-smooth problems with KL property. Actually, the rates are novel even in the smooth non-convex case. Notably, we achieve this without employing line search or trust region strategies, without assuming the Dennis-Mor'e condition, and without assuming strong convexity. Furthermore, for convex problems, we focus on a more tractable gradient regularized quasi-Newton method (Gradient SR1) which can achieve results similar to those obtained with cubic regularization.

4 - LoCoDL: Communication-Efficient Distributed Optimization with Local Training and Compression
Laurent Condat, Arto Maranjyan, Peter Richtarik

In distributed optimization, and even more in federated learning, communication is the main bottleneck. We introduce LoCoDL, a communication-efficient algorithm that leverages the two techniques of Local training, which reduces the communication frequency, and Compression with a large class of unbiased compressors that includes sparsification and quantization strategies. LoCoDL provably benefits from the two mechanisms and enjoys a doubly-accelerated communication complexity, with respect to the condition number of the functions and the model dimension, in the general heterogenous regime with strongly convex functions. The paper "LoCoDL: Communication-Efficient Distributed Learning with Local Training and Compression" has been presented at the conference ICLR 2025.

■ MB-05

Monday, 10:30-12:30 - Room: B100/4013

Optimization and machine learning I

Stream: Optimization for machine learning
Invited session
Chair: *Laurent Condat*
Chair: *Avetik Karagulyan*

1 - Laplacian Regularization in Semi-Supervised Learning with Functional Data
Zhengang Zhong

We investigate a family of regression problems in a semi-supervised setting, where the goal is to assign real-valued labels to n sample points, given a small subset of N labeled points. A goal of semi-supervised learning is to leverage the (geometric) structure provided by the large number of unlabeled data when assigning labels. To capture this structure, we model the data set using random geometric graphs with a connection radius. In the context of optimization problems, the associated objective functionals reward the regularity of the labeling function. However, these problems degenerate in high dimensions when the labeling function lacks sufficient regularity. To address this issue, we study the point-wise asymptotic behavior of such objective functionals in the setting of functional sample data, as n goes to infinity and the connection radius tends to zero.

2 - When to Forget? Complexity Trade-offs in Machine Unlearning
Martin Van Waerebeke, Marco Lorenzi, Gioanni Neglia, Kevin Scaman

Machine Unlearning (MU) aims at removing the influence of specific data points from a trained model, striving to achieve this at a fraction of the cost of full model retraining. In this paper, we analyze the efficiency of unlearning methods and establish the first upper and lower bounds on minimax computation times for this problem, characterizing the performance of the most efficient algorithm against the most difficult objective function. Specifically, for strongly convex objective functions and under the assumption that the forget data is inaccessible to the unlearning method, we provide a phase diagram for the unlearning complexity ratio—a novel metric that compares the computational cost of the best unlearning method to full model retraining. The phase diagram reveals three distinct regimes: one where unlearning at a reduced cost is infeasible, another where unlearning is trivial because adding noise suffices, and a third where unlearning achieves significant computational advantages over retraining. These findings highlight the critical role of factors such as data dimensionality, the number of samples to forget, and privacy constraints in determining the practical feasibility of unlearning.

3 - Searching for optimal per-coordinate stepsizes with multidimensional backtracking
Frederik Kunstner, Victor Sanches Portella, Nick Harvey, Mark Schmidt

The backtracking line-search is an eective technique to automatically tune the step-size in smooth optimization. It guarantees similar performance to using the theoretically optimal step-size. Many approaches have been developed to instead tune per-coordinate step-sizes, also known as diagonal preconditioners, but none of the existing methods are provably competitive with the optimal per-coordinate step-sizes. We propose multidimensional backtracking, an extension of the backtracking line-search to find good diagonal preconditioners for smooth convex problems. Our key insight is that the gradient with respect to the step-sizes, also known as hypergradients, yields separating hyperplanes that let us search for good preconditioners using cutting-plane methods. As blackbox cutting-plane approaches like the ellipsoid method are computationally prohibitive, we develop an ecient algorithm tailored to our setting. Multidimensional backtracking is provably competitive with the best diagonal preconditioner and requires no manual tuning.

4 - Automatic recommendation of optimization methods through their worst-case complexity
Sofiane Tanji, *François Glineur*
Since the advent of modern computational mathematics, the literature on optimization algorithms has been ever-growing, and a wide range of methods are now available. Typically, these methods apply to specific templates of optimization problems and are accompanied by proofs of their convergence rates. Templates vary: an objective may be a single function or a sum of functions with different properties (e.g., smoothness, convexity, a computable proximal operator), and constraints may be linear, involve a feasible set with a projection operator, or be functional. Hence, for a given optimization problem, it is not always easy to identify which templates it can match, especially if one wants to consider equivalent reformulations of the problem. This makes the task of choosing an optimization method tedious. We propose a general framework to recommend optimization methods for any oracle-based problem. Given a user-provided optimization problem, the framework: (1) computes equivalent reformulations; (2) checks which formulations match known templates; (3) retrieves applicable optimization methods; and (4) order methods given their certified worst-case performance. Moreover, when algorithms or templates feature hyperparameters (such as stepsize or some amount of curvature shift), our approach will provide the best problem-dependent value for those hyperparameters according to their worst-case convergence rates.

■ MB-06

Monday, 10:30-12:30 - Room: B100/7013

Nonsmooth optimization: from continuous to discrete Part I

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Alberto De Marchi*

1 - Mixed-integer linearity in nonlinear optimization
Alberto De Marchi
Bringing together nonlinear optimization with polyhedral and integrality constraints enables versatile modelling, but poses significant computational challenges. To address these problems, the talk presents an algorithm that is inspired by proximal-gradient methods but replaces the proximal operator with calls to a generic mixed-integer linear solver. The technique computes feasible iterates based on sequential mixed-integer linearization with trust region safeguard. Convergence to critical, possibly suboptimal, feasible points is established for arbitrary starting points. The theoretical and algorithmic developments are corroborated by numerical applications.

2 - Lyapunov functions and control for differential inclusions by optimization
Sigurdur Hafstein
The stability of equilibria in dynamical systems can be characterized by the existence of Lyapunov functions, which are real-valued functions from the state-space that are decreasing along trajectories. Lyapunov functions can be constructed using linear programming or semidefinite optimization for various kinds of dynamical systems, including systems defined by ODEs. We will discuss a method using linear programming and how it can be extended for differential inclusions and systems with control, with leads to MIP problems.

3 - Finding and exploiting quadratic underestimators for optimal value functions of nonconvex all-quadratic problems via copositive optimization
Markus Gabl, *Immanuel Bomze*
Modeling parts of an optimization problem as an optimal value function that depends on a top-level decision variable is a regular occurrence in optimization and an essential ingredient for methods such as Benders Decomposition. If the problem is convex, duality theory can be used to build piecewise affine models of the optimal value function over which the top-level problem can be optimized efficiently. In this talk, we are interested in the optimal value function of an quadratically constrained quadratic problem problem (QCQP) which is not necessarily convex, so that duality theory can not be applied without introducing a generally unquantifiable relaxation error. This issue can be bypassed by employing copositive reformulations of the underlying QCQP. We investigate two ways to parametrize these by the top-level variable. The first one leads to a copositive characterization of an underestimator that is sandwiched between the convex envelope of the optimal value function and that envelope's lower-semicontinuous hull. The dual of that characterization allows us to derive affine underestimators. The second parametrization yields an alternative characterization of the optimal value function itself, which other than the original version has an exact dual counterpart. From the latter, we can derive any convex and nonconvex quadratic underestimators of the optimal value function. These underestimators can be exploited in a Benders decomposition fashion.

■ MB-07

Monday, 10:30-12:30 - Room: B100/5015

Hyperparameter Optimization for Classification

Stream: Bilevel and multilevel optimization
Invited session
Chair: *Yaru Qian*

1 - Bilevel hyperparameter optimization for RBF Kernel support vector machines
Qingna Li
The problem of tuning the hyperparameters of a Support Vector Machine (SVM) model via cross validation is intuitively a bilevel problem. Methods for solving this problem have been presented in the literature however these papers have addressed only the linear kernel SVM. The Radial Basis Function (RBF) or Gaussian kernel affords SVM models the ability to capture more complex relationships between the variables of our data. This however comes with the drawback of the primal form of the training problem containing the function $phi: mathbbR rightarrow mathbbRinfty$. To avoid this, we consider instead the dual formulation of problem. Therefore, we have to construct robust methods for deriving the primal parameters from the dual parameters. We perform the KKT single level reformulation and apply the Sholtes relaxation to solve the resulting mathematical program with equilibrium constraints (MPEC). The relaxed problem is then solved using fmincon in MATLAB. We also discussed the theoretical property of the resulting MPEC.

2 - A highly efficient single-loop smoothing damped Newton method for large-scale bilevel hyperparameter optimization of SVC
Yixin Wang

Bilevel hyperparameter optimization has received growing attention thanks to the fast development of machine learning. Due to the tremendous size of data sets, the scale of bilevel hyperparameter optimization problem could be extremely large, posing great challenges in designing efficient numerical algorithms. In this paper, we focus on solving the large-scale mathematical programs with equilibrium constraints (MPEC) derived from hyperparameter selection of L1-support vector classification (L1-SVC). We propose a highly efficient single-loop smoothing damped Newton method (SDN) for solving such MPEC. Compared with most existing algorithms where subproblems are involved and solved by on-shelf packages, our approach fully takes advantage of the structure of MPEC and therefore is single-loop. Moreover, the proposed SDN enjoys a quadratic convergence rate under proper assumptions. Extensive numerical results over LIBSVM dataset show the superior performance of SDN over other state-of-art algorithms including the Scholtes global relaxation method with subproblem solved by SNOPT and the Matlab built-in function fmincon, especially in CPU time. For example, for dataset w4a, SDN is 20 times faster than SGRM and 3 times faster than fmincon. Further numerical results also verifies the quadratic convergence of SDN as well as the fulfillment of the second order sufficient condition, while guarantees that SDN returns a strict local minimizer of the smoothing problem of MPEC.

3 - On Constraint Qualifications for MPECs with Applications to Bilevel Hyperparameter Optimization for Machine Learning
Jiani Li

Constraint qualifications for a Mathematical Program with Equilibrium Constraints (MPEC) are essential in analyzing different stationary points and establishing convergence results. In this paper, we explore various classical MPEC constraint qualifications and analyze connections between them. We subsequently study the behavior of these constraint qualifications in the context of a specific MPEC derived from the bilevel hyperparameter optimization (BHO) for L1-loss support vector classification. In particular, for such an MPEC, we provide a full characterization of the well-known MPEC linear constraint qualification, therefore, establishing conditions under which it holds or fails for our BHO for support vector machines.

4 - Bilevel hyperpaGlobal relaxation-based LP-Newton method for multiple hyperparameter selection in support vector classification
Yaru Qian

Support vector classification (SVC) is an effective tool for classification tasks in machine learning. Its performance relies on the selection of appropriate hyperparameters. This paper focuses on optimizing the regularization hyperparameter C and determining feature bounds for feature selection within SVC, leading to a potentially large hyperparameter space. It is very well-known in machine learning that this could lead to the so-called em curse of dimensionality. To address this challenge of multiple hyperparameter selection, the problem is formulated as a bilevel optimization problem, which is then transformed into a mathematical program with equilibrium constraints (MPEC). Our primary contributions are two-fold. First, we establish the satisfaction of the MPEC-MFCQ for our problem reformulation. Furthermore, we introduce a novel global relaxation based linear programming (LP)-Mewton method (GRLPN) for solving this problem and provide corresponding convergence results. Typically, in global relaxation methods for MPECs, the algorithm for the corresponding subproblem is treated as a blackbox. Possibly for the first time in the literature, the subproblem is specifically studied in detail. Numerical experiments demonstrate GRLPN's superiority in efficiency and accuracy over both grid search and traditional global relaxation methods solved using the well-known nonlinear programming solver, SNOPT.

■ MB-08

Monday, 10:30-12:30 - Room: B100/7007

Systematic and computer-aided analyses I: Analyses of proximal splittings methods & friends

Stream: Systematic and computer-aided analyses of optimization algorithms
Invited session
Chair: *Aymeric Dieuleveut*

1 - The Augmented Lagrangian Method for Infeasible Convex Optimization
Roland Andrews

The augmented Lagrangian approach is widely used for solving constrained optimization problems. This work presents a precise study of its behavior when the problem under consideration is infeasible. In particular, we show that the generated sequences of iterates converge to solutions of the closest feasible problem. In this context, we present a hierarchy of assumptions allowing to show a variety of precise results, from sequence convergence (without rates) to precise convergence rates. Our study leverages the classical relationship between augmented Lagrangian algorithms and proximal point methods on dual problems. In doing so, a key tool for us is a simple (and apparently new) result on the behavior of the proximal point algorithm applied to functions that are potentially unbounded below, specifically showing the convergence of convex conjugate values and subgradients, along with their respective convergence rates.

2 - Difference-of-convex algorithm with weakly convex functions: improved splitting technique and equivalence with proximal gradient descent
Teodor Rotaru, Panagiotis Patrinos, François Glineur

The difference-of-convex algorithm (DCA) is a classical, simple parameter-free algorithm designed to minimize a difference of convex functions. We present a comprehensive convergence analysis of DCA extended to handle the case where the second, subtracted function is weakly convex.

Assuming at least one function is smooth, we evaluate the algorithm's performance using the norm of the best (sub)gradient residual or the best (prox-)gradient mapping. Six distinct regimes are identified, two of them corresponding to the previously studied standard difference-of-convex setting. Our simplified proofs are based on deriving six distinct descent lemmas, inspired by solving the associated performance estimation problems (PEP). These establish sublinear convergence rates, which are provably tight for three of the regimes. We also conjecture the exact behavior across any number of iterations in all regimes, informed by extensive numerical simulations.

We also propose a new method to improve the best worst-case convergence rate based on a curvature-shifting technique, which can potentially transform the subtracted function into a weakly convex one. Additionally, we investigate the underexplored equivalence between proximal gradient descent (PGD) and DCA iterations, showing that PGD convergence rates can be derived from DCA rates, where in particular the weakly convex subtracted function enables to deduce rates for large PGD stepsizes.

3 - Forward-backward type splitting algorithms with minimal lifting I
Anton Åkerman, Emanuele Naldi, Enis Chenchene, Pontus Giselsson

We consider forward-backward type splitting methods for solving monotone inclusion problems involving maximally monotone and cocoercive operators, and present a characterization of all such methods that are frugal, averaged nonexpansive and with minimal lifting. In cases where simpler counterexamples seem difficult, we derive necessary and sufficient conditions for nonexpansiveness using Performance Estimation. Ultimately, we present a parameterization of this general algorithm class, based on which the second part of the talk will discuss practical performance and heuristics for (computer aided) algorithm design.

4 - Forward-backward type splitting algorithms with minimal lifting II
Emanuele Naldi, Anton Åkerman, Enis Chenchene, Pontus Giselsson

Using the general algorithm introduced in the first part that characterize all frugal, minimally lifted and averaged nonexpansive splitting algorithms, we show that many minimal lifting splitting schemes present in the literature are incorporated in our framework. We further demonstrate how our formulation can be used to systematically design new algorithms with provable convergence guarantees and that are efficient. In order to facilitate the use of the algorithms and the choice of the involved matrices and parameters, we propose several heuristics with clear practical indications. We finally explore the application of our methods to distributed optimization, where the inherent structural properties of our framework enable efficient decentralized implementations. We conclude with numerical experiments that highlight the practical performance of our proposed methods and heuristics, illustrating competitive convergence behavior on both centralized and distributed problem instances. These results confirm the potential of our approach in a variety of large-scale optimization scenarios.

■ MB-09

Monday, 10:30-12:30 - Room: B100/8013

Generalized convexity and monotonicity 1

Stream: Generalized convexity and monotonicity
Invited session
Chair: *Krzysztof Rutkowski*

1 - Proximal Algorithms for a class of abstract convex functions
The Hung Tran

In this paper, we analyze a class of nonconvex optimization problems from the viewpoint of abstract convexity. Using the respective generalizations of the subgradient, we propose an abstract notion of a proximal operator and derive several algorithms, namely abstract proximal point method, abstract forward-backward method, and abstract projected subgradient method. Global convergence results for all algorithms are discussed, and numerical examples are given.

2 - Lagrangian duality for generalized convex optimization problems
Monika Syga, Ewa Bednarczuk

We investigate Lagrangian duality in the context of nonconvex optimization problems. Utilizing the framework of generalized convexity, we provide criteria that guarantee the zero duality gap property. In our results, we do not need any linear structure in underlying spaces.

3 - On Mangasarian-Fromovitz condition in smooth inifinte programming
Krzysztof Rutkowski, Ewa Bednarczuk, Krzysztof Leśniewski

In this presentation we introduce a modified Mangasarian-Fromovitz condition for smooth infinite programming problem in Banach spaces under equality constraints and infinite number of inequality constraints, where the nonlinear operator defining the equality constraints has possible nonsurjective derivative at the local minimum. We will show the existence of Lagrange multipliers by using the proposed condition together with Nonlinear Farkas-Minkowski condition or weak*-closedness of Hurwicz set. At the end of the presentation we show some relationships to other conditions existing in the literature for the problem.

4 - Similarity-based fuzzy clustering scientific articles: potentials and challenges from mathematical and computational perspectives
Thi Huong Vu, Ida Litzel, Thorsten Koch

Fuzzy clustering, which allows an article to belong to multiple clusters with soft membership degrees, plays a vital role in analyzing large-scale publication data. This problem can be formulated as a constrained optimization model, where the goal is to minimize the discrepancy between the similarity observed from data and the similarity derived from a predicted distribution. While this approach benefits from leveraging state-of-the-art optimization algorithms, tailoring them to work with real, massive databases like OpenAlex or Web of Science - containing about 70 million articles and a billion citations - poses significant challenges. In this talk, we discuss potentials and challenges of the approach from both mathematical and computational perspectives. Among other things, second-order optimality conditions are established - providing new theoretical insights - and practical solution methods are proposed by exploiting the problem's structures. Specifically, we accelerate the gradient projection method with GPU-based parallel computing to handle large-scale data efficiently.

■ MB-10

Monday, 10:30-12:30 - Room: B100/8011

Optimization, Learning, and Games I

Stream: Optimization, Learning, and Games
Invited session
Chair: *Georgios Piliouras*

1 - Convex Markov Games
Ian Gemp

Behavioral diversity, expert imitation, fairness, safety goals and others give rise to preferences in sequential decision making domains that do not decompose additively across time. We introduce the class of convex Markov games that allow general convex preferences over occupancy measures. Despite infinite time horizon and strictly higher generality than Markov games, pure strategy Nash equilibria exist. Furthermore, equilibria can be approximated empirically by performing gradient descent on an upper bound of exploitability. Our experiments reveal novel solutions to classic repeated normal-form games, find fair solutions in a repeated asymmetric coordination game, and prioritize safe long-term behavior in a robot warehouse environment. In the prisoner's dilemma, our algorithm leverages transient imitation to find a policy profile that deviates from observed human play only slightly, yet achieves higher per-player utility while also being three orders of magnitude less exploitable.

2 - The Complexity of Two-Team Polymatrix Games with Independent Adversaries
Alexandros Hollender, Gilbert Maystre, Sai Ganesh Nagarajan

Adversarial multiplayer games are an important object of study in multiagent learning. In particular, polymatrix zero-sum games are a multiplayer setting where Nash equilibria are known to be efficiently computable. Towards understanding the limits of tractability in polymatrix games, we study the computation of Nash equilibria in such games where each pair of players plays either a zero-sum or a coordination game. We are particularly interested in the setting where players can be grouped into a small number of teams of identical interest. While the three-team version of the problem is known to be PPAD-complete, the complexity for two teams has remained open. Our main contribution is to prove that the two-team version remains hard, namely it is CLS-hard. Furthermore, we show that this lower bound is tight for the setting where one of the teams consists of multiple independent adversaries. On the way to obtaining our main result, we prove hardness of finding any stationary point in the simplest type of non-convex-concave min-max constrained optimization problem, namely for a class of bilinear polynomial objective functions

3 - From min-max to harmonic games: Regret, learning, and the role of optimism
Panayotis Mertikopoulos

Min-max games are often thought of as the strategic counterpart to potential / common interest games, but this view is overly simplistic - there are simple examples of two-player, zero-sum games that are potential, so this analogy is hardly apt. In this talk, we will discuss the class of harmonic games, originally due to Candogan et al. (2011), which can be shown to be the orthogonal complement of potential games, and thus provide a more principled counterpart thereof.

The main focus of the talk will be the long-run behavior of regularized learning algorithms - like mirror descent - in harmonic games. In a continuous-time setting, we show that the learning dynamics are Poincaré recurrent, i.e., they return arbitrarily close to their starting point infinitely often, so they fail to converge. In discrete time, the standard, "vanilla" implementation of the method may lead to even worse outcomes, eventually trapping the players in a perpetual cycle of best-responses. However, by adding a suitable extrapolation step - which includes as special cases the optimistic and mirror-prox variants of the algorithm - we show that learning converges to a Nash equilibrium from any initial condition, and all players are guaranteed at most O(1) regret. In this regard, harmonic games comprise the canonical complement of potential games not only from a strategic, but also from a dynamical viewpoint.

4 - Revenue Efficiency of First Price Auctions over Approximate Correlated Equilibrium
Efstratios Skoulakis

In this work we study the revenue of approximate correlated equilibrium in discrete first price auctions - the set of allowable bids is $mathcalB = 0, 1/k, dots, 1 - 1/k, 1$ for some $k in mathbbN$. We show that the revenue of any $epsilon$-textitapproximate correlated equilibrium is at least $v_2 - Theta(1/k)- Theta(epsilon k2)$, where $v_2 geq 0$ is the second-highest valuation. Our results establish the first polynomial convergence rates on the revenue generated by no-swap regret bidders in first-price auctions. For instance, if bidders admit the optimal swap regret of $mathcalO(sqrtk T)$, then the time-averaged revenue is at least $v_2 - Theta(1/k) - Theta(epsilon)$ after $mathcalO(k5/epsilon2)$ rounds.

■ MB-11

Monday, 10:30-12:30 - Room: B100/5017

Optimal and stochastic optimal control 1

Stream: Optimal and stochastic optimal control
Invited session
Chair: *Gerhard-Wilhelm Weber*

1 - Optimal model description of finance and human factor indices
Betül Kalaycı, Vilda Purutcuoglu, Gerhard-Wilhelm Weber

Economists have conducted research on several empirical phenomena regarding the behavior of individual investors, such as how their emotions and opinions influence their decisions. All those emotions and opinions are described by the word Sentiment. In finance, stochastic changes might occur according to investors sentiment levels. In this study, our main goal is to apply several operational research techniques and analyze these techniques' accuracy. Firstly, we represent the mutual effects between some financial process and investors sentiment with multivariate adaptive regression splines

(MARS) model. Furthermore, we consider to extend this model by using distinct data mining techniques and compare the gain in accuracy and computational time with its strong alternatives applied in the analyses of the financial data. Hence, the goal of this study is to compare the forecasting performance of sentiment index by using two-stage MARS-NN (neural network), MARS-RF (random forest), RF-MARS, RF-NN, NN-MARS, and NN-RF hybrid models. Furthermore, we aim to classify the peoples' feelings about economy according to their confidence levels. Moreover, to forecast the underlying state change of the consumer confidence index (CCI) and to observe the relationship with some macroeconomic data (CPI, GDP and currency rate) at a monthly interval, we apply hidden Markov model (HMM). The aim is to detect the switch between these states and to define a path of these states.

2 - Dynamic Programming using Average-Value-at-Risk Criteria
Kerem Ugurlu

We investigate dynamic programming equations on Averaga-Value-at-Risk (AVaR) using machine learning techniques and demonstrate several simulations. The dynamic programming equation on AVaR is specifically using a so called "state aggregation technique" that makes use of the sufficient statistic of the optimization problem. Via state aggregation, we present the "Markovian" framework and demonstrate several implementations using fully connected neural networks. The risk averseness level alpha on the model is also investigated via several simulations.

3 - Interpretable Machine Learning Approach for Nonlinear Control
Edmondo Minisci, Giulio Avanzini

The work is about an interpretable machine learning approach for control based on genetic programming with integrated search for continuous coefficients. The method can discover compact, human-readable control laws by combining symbolic expressions with embedded parameter tuning, thus bridging the gap between black-box learning and classical control. The main aim of the work is to demonstrate its potential on textbook cases, including standard systems (e.g., nonlinear oscillators, inverted pendulum) and provide initial results on aeronautical applications, such as stability augmentation of unstable, highly manoeuvrable aircraft. The obtained solutions maintain performance comparable to conventional controllers, while preserving transparency and ease of analysis, and the method enables the treatment of nonlinear systems without reliance on gain scheduling.

4 - Stochastic differential games for optimal investment problems in a Markov regime-switching jump-diffusion market
Gerhard-Wilhelm Weber, Emel Savku

In their paper from behavioral finance, we employ dynamic programming principle on two optimal investment problems in a continuous-time Markov regime-switching setting. They model different states of an economy and, hence, investors' floating levels of psychological reactions by a D-state Markov chain: a zero-sum game between an investor and the market, and a nonzero-sum stochastic differential portfolio game. We derive regime-switching Hamilton-Jacobi-Bellman-Isaacs equations, and explicit optimal portfolio strategies with Feynman-Kac representations of value functions. For a two-state special case the results are illustrated and the impact of regime switches observed by a comparison.

■ MB-12

Monday, 10:30-12:30 - Room: B100/8009

Portfolio optimization

Stream: Applications: Finance
Invited session
Chair: *Houduo Qi*

1 - An Optimization Study of Diversification Return Portfolios
Houduo Qi

The concept of Diversification Return (DR) was introduced by Booth and Fama in 1990s and it has been well studied in the finance literature mainly focusing on the various sources it may be generated. However, unlike the classical Mean-Variance (MV) model of Markowitz, DR portfolios lack optimization theory for justifying their often-outstanding empirical performance. In this paper, we first explain what the DR criterion tries to achieve in terms of portfolio centrality. A consequence of this explanation is that practically imposed norm constraints in fact implicitly enforce constraints on DR. We then derive the maximum DR portfolio under given risk and obtain the efficient DR frontier. We further develop a separation theorem for this frontier and establish a relationship between the DR frontier and Markowitz MV efficient frontier. Finally, we use DAX30 stock data to illustrate the obtained results and demonstrate an interesting link to the maximum diversification ratio portfolio studied by Choueifaty and Coignard.

2 - Coherent risk measurement and return analysis on non-reflexive Banach space
Hedvig Gal

The research papaer relies on the a priori estimation of Kountzakis (2011), where the focus of interest was on the measurement of financial risk on non-reflexive Banach space. In the current formulation the research question referes on what we can say about the axioms of coherence, risk measures and portfolio performance analysis under the conditon of non-reflexive Banach space? The financial portfolio compares the financial assets with high and low average number of daily transactions and their standard deviation, Conditional Value-at-Risk (CvaR), Value-at-Risk (VaR), performance of Sharpe ratio, Treynor indicator and the optimal portfolio execution strategy using Kissel-Glanz function.

3 - Adaptive minimum variance portfolio selection by linear shrinkage and expansion
Xiang Zhao, Selin Ahipasaoglu, Houduo Qi

We study the minimum variance portfolio optimization problem under covariance matrix uncertainty, arising from estimation errors and outliers. Shrinkage methods, which impose structure on the sample covariance matrix by moving it towards a target matrix, are effective when the target matrix is negatively correlated with estimation errors. We extend the linear matrix shrinkage framework to include "matrix expansion" in which the covariance matrix is allowed to move away from a given target. Using a target matrix constructed with Lagrange multipliers for the long-only portfolio problem, we identify the allowable range for shrinkage and expansion. The optimal modification, which could be shrinkage or expansion under the modified framework, is dynamically selected on the basis of in-sample performance. Our numerical experiments demonstrate that portfolios generated by this method can achieve higher expected returns, superior Sharpe ratios, and improved risk metrics, such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). These results suggest that combining shrinkage and expansion in a data-driven framework enhances portfolio performance and robustness.

■ MB-13

Monday, 10:30-12:30 - Room: B100/6009

True sparsity in Standard Quadratic Problems

Stream: Sparsity guarantee and cardinality-constrained (MI)NLPs
Invited session
Chair: *Immanuel Bomze*

1 - Novel and Tractable Convex Relaxations of Standard Quadratic Optimization Problems under Cardinality Constraints
E. Alper Yildirim, *Immanuel Bomze*, *Bo Peng*, *Yuzhou Qiu*

Standard quadratic optimization problems (StQPs) provide a versatile modelling tool in a multitude of applications such as mathematical finance, machine learning (clustering) and modelling in biosciences (e.g. selection and ecology). In this talk, we consider StQPs under an additional sparsity or cardinality constraint, which, even for convex objectives, renders NP-hard problems. One motivation to study StQPs under such sparsity restrictions is the high-dimensional portfolio selection problem with too many assets to handle, in particular in the presence of transaction costs. We present novel computational approaches to this relevant but difficult problem, involving modern conic optimization techniques and significant dimensional reduction, which is essential for the tractability of these methods when the problem size grows. In addition, we propose a particular generation procedure that systematically avoids instances that are too easy. We present extensive computational results demonstrating the versatility and strength of the proposed relaxations.

2 - From Keynes to Markowitz and back — optimize portfolios by strict cardinality control
Immanuel Bomze, *Paula Amaral*, *Yuan Chen*, *Bo Peng*

We will address portfolio selection problems with rigorous lower and/or upper bounds on the number of selected assets. In case of lower bounds, we will also impose a fixed positive lower bound on the amount of used (positive) assets, i.e., enter the domain of semi-continuous variables. This so-called minimum buy-in threshold may be incurred by transaction costs. The resulting models are QPs with additional binary or continuous variables, and additional constraints (of linear and/or complementarity type). Several formulations and algorithmic perspectives are discussed, as well as potential microfinance applications.

3 - Cardinality-Constrained Optimization for Large-Scale Portfolio
Yuan Chen, *Immanuel Bomze*, *Nikolaus Hautsch*, *Bo Peng*

We propose a portfolio optimization model that reconciles Keynes's advocacy for concentrated investments with Markowitz's emphasis on diversification. By incorporating cardinality constraints into the Markowitz mean-variance framework, we enable investors to focus on a small set of assets, fostering specialized expertise. Cardinality constraints allow investors to still use the sample covariance matrix in high-dimensional settings with limited data, balancing diversification needs while mitigating estimation errors inherent in such environments.

Monday, 14:00-16:00

■ MC-01

Monday, 14:00-16:00 - Room: B100/1001

Strategies to Improve Zeroth-Order Optimization Methods

Stream: Zeroth and first-order optimization methods
Invited session
Chair: *Francesco Rinaldi*
Chair: *Andrea Cristofari*

1 - Improving the robustness of zeroth-order optimization solvers
Stefan M. Wild

Zeroth-order optimization solvers are often deployed in settings where little information regarding a problem's conditioning or noise level is known. An ideal solver will perform well in a variety of challenging settings. We report on our experience developing adaptive algorithms, which leverage information learned online to adapt critical algorithmic features. We illustrate our approach in trust-region-based reduced-space methods and show how trained policies can even be deployed effectively in nonstationary cases, where the noise seen changes over the decision space. This is joint work with Pengcheng Xie.

2 - Derivative free optimization with structured random directions
Silvia Villa, Marco Rando, Cheik Traoré, Cesare Molinari, Lorenzo Rosasco

We consider the problem of minimizing an objective function in a black-box setting where only function evaluations are available. Finite-difference methods is a class of algorithms that mimic gradient-based methods by replacing gradients with approximations built using function evaluations along a set of directions. In this talk I will focus on the specific choice of the set of directions where the directions are randomly chosen but satisfy an orthogonality constraint. I will then analyze finite-difference methods which incorporate variance-reduction techniques to minimize a finite sum of functions. I will derive convergence rates for smooth non-convex functions and show that our algorithm achieves a convergence rate that matches the state-of-the-art. Finally,I will illustrate our method through numerical experiments.

3 - Enhancing finite-difference-based derivative-free optimization with machine learning
Geovani Grapiglia, Timothé Taminiau, Estelle Massart

Derivative-Free Optimization (DFO) involves methods that rely solely on evaluations of the objective function. One of the earliest strategies for designing DFO methods is to adapt first-order methods by replacing gradients with finite-difference approximations. The execution of such methods generates a rich dataset about the objective function, including iterate points, function values, approximate gradients, and successful step sizes. In this work, we propose a simple auxiliary procedure to leverage this dataset and enhance the performance of finite-difference-based DFO methods. Specifically, our procedure trains a surrogate model using the available data and applies the gradient method with Armijo line search to the surrogate until it fails to ensure sufficient decrease in the true objective function, in which case we revert to the original algorithm and improve our surrogate based on the new available information. As a proof of concept, we integrate this procedure with the derivative-free method proposed in (Optim. Lett. 18: 195–213, 2024). Numerical results demonstrate significant performance improvements, particularly when the approximate gradients are also used to train the surrogates.

4 - A derivative-free algorithm based on resilient positive spanning sets
Sébastien Kerleau, Clément Royer

Positive spanning sets (PSSs) are very useful in DFO algorithms. Optimization methods based on PSSs typically favor those with the best cosine measure, yet this metric does not fully account for their structure: in particular, it does not reflect the spanning capabilities of their subsets. This talk focuses on a particular class of PSSs, namely PkSSs, that remain positively spanning when losing elements. After showing how to construct PkSSs whose subsets all possess a large cosine measure, I present a derivative-free algorithm based on parallel computing and resilient to stragglers.

■ MC-02

Monday, 14:00-16:00 - Room: B100/7011

Matrix factorization

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Atharva Awari*

1 - Alternating Direction Method of Multipliers for Nonlinear Matrix Decompositions
Atharva Awari, Nicolas Gillis, Arnaud Vandaele

Low-rank matrix approximations are fundamental in data analysis, machine learning, and signal processing. Traditional approaches, such as the singular value decomposition and nonnegative matrix factorization (NMF), assume a linear relationship between the observed data matrix X and its low-rank factors W and H. However, many real-world datasets exhibit nonlinear structures that cannot be effectively captured using standard linear models. This has led to growing interest in nonlinear matrix decompositions (NMDs), where the goal is to find low-rank factors W and H such that Xf(WH), where f is an element-wise nonlinear function . Despite the increasing relevance of NMDs, existing algorithms lack the flexibility to handle a wide range of nonlinear functions commonly used in practice. Examples include the ReLU function with f(x)=max(0,x)) useful in the approximation of sparse datasets, the component-wise square with f(x)=x2 useful in representation of probabilistic circuits, and Min-Max models where the data lies in a certain interval (a,b) with f(x)=min(b,max(a,x)). To bridge this gap, we propose an Alternating Direction Method of Multipliers (ADMM) framework tailored for NMDs. Our method efficiently handles diverse nonlinear models while accommodating different loss functions, including least squares, 1-norm, and the Kullback-Leibler divergence.Additionally, our approach is easily adaptable to other nonlinear functions and loss functions, ensuring broad applicability

2 - Instance-wise Distributionally Robust Nonnegative Matrix Factorization
Amjad Seyedi, Nicolas Gillis

Nonnegative matrix factorization (NMF) is a widely used data representation model across diverse domains, including machine learning. At its core, NMF aims to minimize the distance between the original input and its lower-rank approximation. However, when data are noisy or contain outliers, NMF often struggles to provide accurate results. Existing robust methods depend on known distributional assumptions, which can limit their effectiveness in real-world scenarios where the noise distribution is unknown. To address this limitation, we introduce the instance-wise distributionally robust NMF (iDRNMF), a model designed to accommodate a wide range of noise distributions. By employing a weighted-sum multi-objective method, iDRNMF handles multiple noise distributions and their combinations. Furthermore, while entry-wise models assume noise contamination at individual matrix entries, our proposed instance-wise model assumes contamination at the level of entire instances. This perspective is often more suitable for data representation tasks, as it addresses noise affecting whole feature vectors rather than isolated features. To train the model, we develop a unified multi-objective optimization framework based on an iterative reweighted algorithm, maintaining computational efficiency as it is comparable to single-objective NMFs. This framework offers flexible updating rules, making it well-suited for optimizing a wide range of robust and distributionally robust objectives.

3 - Implicit Bias in Matrix Factorization and its Explicit Realization in a new Architecture
Yikun Hou, Suvrit Sra, Alp Yurtsever

Gradient descent for matrix factorization is known to exhibit an implicit bias toward approximately low-rank solutions. While existing theories often assume the boundedness of iterates, empirically the bias persists even with unbounded sequences. We thus hypothesize that implicit bias is driven by divergent dynamics markedly different from the convergent dynamics for data fitting. Using this perspective, we introduce a new factorization model: $Xapprox UDVtop$, where U and V are constrained within norm balls, while D is a diagonal factor allowing the model to span the entire search space. Our experiments reveal that this model exhibits a strong implicit bias regardless of initialization and step size, yielding truly (rather than approximately) low-rank solutions. Furthermore, drawing parallels between matrix factorization and neural networks, we propose a novel neural network model featuring constrained layers and diagonal components. This model achieves strong performance across various regression and classification tasks while finding low-rank solutions, resulting in efficient and lightweight networks.

■ MC-03

Monday, 14:00-16:00 - Room: B100/4011

First-order methods in modern optimization (Part II)

Stream: Large scale optimization: methods and algorithms
Invited session
Chair: *Simone Rebegoldi*
Chair: *Andrea Sebastiani*

1 - Optimization Techniques for Learning Multi-Index Models
Hippolyte Labarrière, Shuo Huang, Ernesto De Vito, Lorenzo Rosasco, Tomaso Poggio

Neural networks naturally identify low-dimensional structures that improve generalization, a property often lacking in classical kernel methods. In this talk, I will present how we can bridge this gap by integrating the multi-index model (MIM) into kernel regression. By introducing a hyper-kernel that incorporates a learned projection matrix, we enhance the adaptability of kernel methods. I will discuss the theoretical guarantees of this approach and introduce an optimization framework based on alternating minimization techniques and Nyström approximations

2 - Neural Blind Deconvolution for Poisson Data
Alessandro Benfenati, Ambra Catozzi, Valeria Ruggiero

The Blind Deconvolution problem arises in several scientific imaging fields, such as Microscopy, Medicine, and Astronomy. The Point Spread Function (PSF) may only be approximately known: blind deconvolution aims to reconstruct the image when only the recorded data is available. Among the standard variational approaches, Deep Learning techniques have gained attention due to their impressive performance. The Deep Image Prior framework has been employed to solve this task, giving rise to the so-called Neural Blind Deconvolution, where the unknown PSF and image are estimated through two separate neural networks. In this paper, we focus on microscopy images, where the predominant noise is of Poisson type: the objective function to minimize is hence the generalized Kullback-Leibler divergence, coupled with regularization terms for both the PSF and the image. Furthermore, we propose modifying the standard NBD formulation by incorporating an upper bound for the blur kernel, which depends on the optical instrument. A numerical solution is obtained via an alternating Proximal Gradient Descent-Ascent procedure, resulting in the Double Deep Image Prior for Poisson noise algorithm. We evaluate the proposed strategy on both synthetic and real-world images, achieving promising results and demonstrating that the correct choice of loss and regularization functions strongly depends on the specific application.

3 - Adaptively Inexact Bilevel Learning via Primal-Dual Differentiation
Mohammad Sadegh Salehi

Bilevel optimisation plays a crucial role in machine learning, particularly for tasks such as learning parameters in variational regularisation problems. In this talk, we introduce an Adaptively Inexact First-order Method for Bilevel Learning, which employs a primal-dual differentiation strategy and leverages the 'piggyback' algorithm to compute hypergradients. Our approach derives an a posteriori error bound for the primal-dual setting, enabling adaptive tolerance selection to balance computational efficiency and accuracy. Additionally, we introduce an adaptive step-size strategy to enhance upper-level optimisation.

From an application perspective, we showcase the efficient learning of total variation discretisation for image denoising and the first successful bilevel learning of input convex neural networks (ICNNs) as regularisers via reformulation for sparse-angle computed tomography. Moreover, our method outperforms existing approaches for training such data-adaptive regularisers.

4 - Alternate Through the Epochs Stochastic Gradient for Multi-Task Neural Networks
Stefania Bellavia, Francesco Della Santa, Alessandra Papini

We focus on the training phase of Neural Networks for Multi-Task Learning. We consider hard-parameter sharing Multi-Task Neural Networks (MTNNs) and discuss alternate stochastic gradient updates. Traditional MTNN training faces challenges in managing conflicting loss gradients, often yielding sub-optimal performance. The proposed alternate training method updates shared and task-specific weights alternately through the epochs, exploiting the multi-head architecture of the model. This approach reduces computational costs per epoch and memory requirements. Convergence properties similar to those of the classical stochastic gradient method are established. Empirical experiments demonstrate enhanced training regularization and reduced computational demands.

■ MC-05

Monday, 14:00-16:00 - Room: B100/4013

Optimization and machine learning II

Stream: Optimization for machine learning
Invited session
Chair: *Laurent Condat*

1 - Stabilized Proximal-Point Methods for Federated Optimization
Xiaowen Jiang, Anton Rodomanov, Sebastian Stich

In developing efficient optimization algorithms, it is crucial to account for communication constraints—a significant challenge in modern Federated Learning. The best-known communication complexity among non-accelerated algorithms is achieved by DANE, a distributed proximal-point algorithm that solves local subproblems at each iteration and that can exploit second-order similarity among individual functions. However, to achieve such communication efficiency, the algorithm requires solving local subproblems sufficiently accurately, resulting in slightly sub-optimal local complexity. Inspired by the hybrid-projection proximal-point method, in this work, we propose a novel distributed algorithm S-DANE. Compared to DANE, this method uses an auxiliary sequence of prox-centers while maintaining the same deterministic communication complexity. Moreover, the accuracy condition for solving the subproblem is milder, leading to enhanced local computation efficiency. Furthermore, S-DANE supports partial client participation and arbitrary stochastic local solvers, making it attractive in practice. We further accelerate S-DANE and show that the resulting algorithm achieves the best-known communication complexity among all existing methods for distributed convex optimization while still enjoying good local computation efficiency as S-DANE. Finally, we propose adaptive variants of both methods using line search, obtaining the first provably efficient adaptive algorithm.

2 - Adaptive strategies for a stochastic trust-region based algorithm
Simone Rebegoldi, Stefania Bellavia, Benedetta Morini

The TRust-region-ish (TRish) algorithm is a stochastic optimization method for finite-sum minimization problems proposed in [Curtis et al., INFORMS J. Optim., 2019]. At each iteration, TRish takes either a stochastic gradient step or a normalized step, the latter occuring when the norm of the stochastic gradient belongs to a prefixed positive interval. Although TRish shows a reduced dependence from the learning rate, it still requires the tuning of the size of the mini-batch, as well as two additional positive scalars involved in the definition of the normalized step. The tuning of these hyper-parameters can be computationally onerous, and represents the main limitation of the algorithm.

In this talk, we study the theoretical properties of TRish, under the assumption that the learning rate and the additional scalars defining the step are constant, and provide results that complement the convergence analysis already existing for TRish. Furthermore, we investigate the use of adaptive sampling techniques that vary dynamically the sample size for computing the stochastic gradient, and provide a practical implementation of TRish that complies with the theoretical assumptions of our analysis. We show experimentally that an adaptive strategy for building the stochastic gradient reduces the dependence of the performance of TRish from its hyper-parameters.

3 - A primal-dual algorithm for variational image reconstruction with learned convex regularizers
Hok Shing Wong

We address the optimization problem in a data-driven variational reconstruction framework, where the regularizer is parameterized by an input-convex neural network (ICNN). While gradient-based methods are commonly used to solve such problems, they struggle to effectively handle non-smooth problems which often leads to slow convergence. Moreover, the nested structure of the neural network complicates the application of standard non-smooth optimization techniques, such as proximal algorithms. To overcome these challenges, we reformulate the problem and eliminate the network's nested structure. By relating this reformulation to epigraphical projections of the activation functions, we transform the problem into a convex optimization problem that can be efficiently solved using a primal-dual algorithm. We also prove that this reformulation is equivalent to the original variational problem. Through experiments on several imaging tasks, we demonstrate that the proposed approach outperforms subgradient methods in terms of both speed and stability.

4 - Entropic Mirror Descent for Linear Systems: Polyak's Stepsize and Implicit Bias
Alexander Posch, Yura Malitsky

In this talk, we explore the use of entropic mirror descent for solving linear systems on the nonnegative orthant. We discuss the motivation behind this approach, such as its implicit bias when initialized close to the origin, and analyze its convergence properties. Although the method has a simple structure, its convergence behavior is more subtle, with a particular challenge being the unboundedness of the domain. We demonstrate how a Polyak-type stepsize can help overcome these challenges and lead to explicit rates. In particular, we establish sublinear convergence in the function values and give conditions under which the convergence is linear. We show how the convergence results generalize to the minimization of arbitrary convex L-smooth functions on the nonnegative orthant. Furthermore, we propose an alternative method that avoids the exponentiation in the mirror descent update, resembling gradient descent with Hadamard parametrization, but with provable convergence guarantees. Finally, we extend the convergence results to linear systems over the entire space.

■ MC-06

Monday, 14:00-16:00 - Room: B100/7013

Nonsmooth optimization: from continuous to discrete Part II

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Alberto De Marchi*

1 - Paley graphs, stable sets and the exact subgraph hierarchy: Bad and good news
Elisabeth Gaar, Dunja Pucher

To determine the stability number of a graph, defined as the cardinality of the largest set of pairwise non-adjacent vertices, is a fundamental NP-hard discrete optimization problem. The exact subgraph hierarchy (ESH) provides a sequence of increasingly tighter upper bounds on the stability number based on non-smooth optimization, starting with the Lovász theta function at the first level and including all exact subgraph constraints of subgraphs of order k into the semidefinite program to compute the Lovász theta function at level k.

In this talk, we investigate the ESH for Paley graphs, a class of strongly regular, vertex-transitive graphs. We show that for Paley graphs, the bounds obtained from the ESH remain the Lovász theta function up to a certain level, i.e., the bounds of the ESH do not improve up to a certain level.

To overcome this limitation, we introduce the vertex-transitive ESH for the stable set problem for vertex-transitive graphs such as Paley graphs. We prove that this new hierarchy provides upper bounds on the stability number of vertex-transitive graphs that are at least as tight as those obtained from the ESH. Additionally, our computational experiments reveal that the local ESH produces superior bounds compared to the ESH for Paley graphs.

2 - A modified proximal point algorithm involving generalized nonexpansive mappings
Izhar Uddin

In this paper, we propose a modified proximal point algorithm based on M-iteration to approximate a common element of the solution sets of convex minimization problems and the fixed points of nearly asymptotically quasi-nonexpansive mappings in CAT(0) spaces. Additionally, we establish some convergence results of the proposed algorithm for solving both the minimization and fixed point problems. We also present an application and provide numerical results demonstrating the effectiveness of our algorithm.

3 - Combining Gradient Information and Primitive Directions for High-Performance Mixed-Integer Optimization
Pierluigi Mansueto, Matteo Lapucci, Giampaolo Liuzzi, Stefano Lucidi

In this talk we consider bound-constrained mixed-integer optimization problems where the objective function is differentiable w.r.t. the continuous variables for every configuration of the integer variables. We mainly suggest to exploit derivative information when possible in these scenarios: concretely, we propose an algorithmic framework that carries out local optimization steps, alternating searches along gradient-based and primitive directions. The algorithm is shown to match the convergence properties of a derivative-free counterpart. Most importantly, the results of thorough computational experiments show that the proposed method clearly outperforms not only the derivative-free approach but also the main alternatives available from the literature to be used in the considered setting, both in terms of efficiency and effectiveness.

■ MC-07

Monday, 14:00-16:00 - Room: B100/5015

Bilevel Optimization in Data Science

Stream: Bilevel and multilevel optimization
Invited session
Chair: *Samuel Ward*

1 - Bi-level optimization in machine learning: instances, acceleration and implicit bias
Zhanxing Zhu

In this talk, I will review several important instances of bi-level optimization in machine learning. These include neural architecture search, hyperparameter optimization, adversarial training, data condensation, meta learning, etc.

I will also introduce several recent works my group did to advance this area. 1) Accelerating large-scale bi-level optimization (NeurIPS'24). We proposed a novel Forward Gradient Unrolling with Forward Gradient, abbreviated as (FG)2U, which achieves an unbiased stochastic approximation of the meta gradient for bi-level optimization. (FG)2U circumvents the memory and approximation issues associated with classical bi-level optimization approaches and delivers significantly more accurate gradient estimates than existing approaches. 2) Implicit bias of bi-level optimization (ICLR'22 and TPAMI'25). Adversarial training, as an instance of bi-level optimization, has been empirically demonstrated as an effective strategy to improve the robustness of deep neural networks (DNNs) against adversarial examples. However, the underlying reason of its effectiveness is still non-transparent. We conducted extensive theoretical analysis on the training dynamics of homogeneous DNNs and show that the adversarial training implicitly learns a generalized margin to improve the adversarial robustness, solving the long-standing conjecture in this area.

2 - Adversarial training under restricted data manipulation using pessimistic bilevel optimisation
David Benfield

Adversarial machine learning concerns situations in which learners face attacks from active adversaries. Such scenarios arise in applications such as spam email filtering, malware detection and fake-image generation, where security methods must be actively updated to keep up with the ever-improving generation of malicious data. Pessimistic Bilevel optimisation has been shown to be an effective method of training resilient classifiers against such adversaries. By modelling these scenarios as a game between the learner and the adversary, we anticipate how the adversary will modify their data and then train a resilient classifier accordingly. However, since existing pessimistic bilevel approaches feature an unrestricted adversary, the model is vulnerable to becoming overly pessimistic and unrealistic. When finding the optimal solution that defeats the classifier, it's possible that the adversary's data becomes nonsensical and loses its intended nature. Such an adversary will not properly reflect reality, and consequently, will lead to poor classifier performance when implemented on real-world data. By constructing a constrained pessimistic bilevel optimisation model, we restrict the adversary's movements and identify a solution that better reflects reality. We demonstrate through experiments that this model performs, on average, better than the existing approach.

3 - An Exploratory Study on Clustering Models under Perturbations Based on a Bilevel Optimization Framework

Yutong Zheng

Clustering, as one of the core tasks in unsupervised learning, plays a crucial and indispensable role in numerous disciplines and application domains. However, the robustness of clustering is increasingly threatened by noisy data. This presentation provides an overview of our preliminary investigation into the impact of perturbations on clustering results. Using a bilevel optimization framework, we have studied a perturbation bound that ensures cluster stability, thereby establishing a connection to adversarial data attacks. Furthermore, we examine a metric for evaluating the degree of change in clustering results and derive its explicit expression and monotonicity properties. Future research will focus on extending the framework to higher dimensions and larger scales, while further exploring perturbation robustness and adversarial data attacks, thus addressing real-world large-scale challenges.

4 - Mathematical programs with complementarity constraints and application to hyperparameter tuning for nonlinear support vector machines

Samuel Ward

We consider the Mathematical Program with Complementarity Constraints (MPCC). One of the main challenges in solving the problem is the systematic failure of standard Constraint Qualifications (CQs). The Mangasarian Fromovitz Constraint Qualification (MFCQ), for instance, does not hold for the MPCC when the problem is viewed from the lens of a standard optimisation problem with equality and inequality constraints. Carefully accounting for the combinatorial nature of the complementarity constraints, tractable versions of MPCC-tailored MFCQ have been designed and widely studied in the literature. In this presentation, we look closely at two such MPCC-tailored MFCQs and their influence in the convergence analysis of the sequential partial penalisation and Scholtes relaxation algorithms. Moreover, we go a step further and look at how these CQs would behave for the problem of tuning hyperparameters in a Support Vector Machine (SVM), a fundamental problem for classification algorithms in machine learning. Additionally, we present robust implementations and comprehensive numerical experimentation on real-world data sets, which show that the sequential partial penalisation method for the MPCC reformulation of the hyperparameter optimisation of a nonlinear SVM can outperform both the Scholtes relaxation technique and the state-of-the-art algorithms from the machine learning literature.

■ MC-08

Monday, 14:00-16:00 - Room: B100/7007

Systematic and computer-aided analyses II: Systematic algorithmic design approaches

Stream: Systematic and computer-aided analyses of optimization algorithms

Invited session

Chair: *François Glineur*

1 - Exact Verification of First-Order Methods via Mixed-Integer Linear Programming

Vinit Ranjan, Jisun Park, Stefano Gualandi, Andrea Lodi, Bartolomeo Stellato

We present exact mixed-integer linear programming formulations for verifying the performance of first-order methods for parametric quadratic optimization. We formulate the verification problem as a mixed-integer linear program where the objective is to maximize the infinity norm of the fixed-point residual after a given number of iterations. Our approach captures a wide range of gradient, projection, proximal iterations through affine or piecewise affine constraints. We derive tight polyhedral convex hull formulations of the constraints representing the algorithm iterations. To improve the scalability, we develop a custom bound tightening technique combining interval propagation, operator theory, and optimization-based bound tightening. Numerical examples, including linear and quadratic programs from network optimization, sparse coding using Lasso, and optimal control, show that our method provides several orders of magnitude reductions in the worst-case fixed-point residuals, closely matching the true worst-case performance.

2 - Tight Analysis of Second-Order Optimization Methods via Interpolation of generalized Hessian Lipschitz Univariate Functions

François Glineur, Nizar Bousselmi, Julien Hendrickx, Anne Rubbens

We show how to compute exact worst-case guarantees on the performance of second-order optimization methods on a wide range of univariate function classes. These classes include Hessian Lipschitz, self-concordant, and quasi-self-concordant functions, which we unify within a class of generalized Lipschitz functions.

Our results rely on the development of interpolation conditions for all classes of interest, for which we present a standard derivation procedure, combined with the Performance Estimation framework which we generalize to include second-order quantities. Use those, we improve or, in some cases, prove tightness of known rates of convergence of Newton's method and its regularized variants on the above-mentioned classes of univariate functions.

3 - Forward-backward algorithms with deviations

Sebastian Banert

We compare different variations of forward-backward-type algorithms for minimising a sum of two functions or finding a zero of the sum of two monotone operators. The focus will be on algorithms with, what we call, deviations or steering vectors. Such algorithms allow for a degree of freedom in the dimension of the optimisation variable instead of only scalar parameters, potentially leading to a greater adaptability to specific problem classes, for example by deep learning. We will compare different approaches to introduce deviations by their performance guarantees, expressivity, and numerical performance. This talk is based on joint work with Jevgenija Rudzusika, Ozan Öktem, Jonas Adler, Hamed Sadeghi, Pontus Giselsson, and Oskar Bircks.

■ MC-09

Monday, 14:00-16:00 - Room: B100/8013

Generalized convexity and monotonicity 2

Stream: Generalized convexity and monotonicity
Invited session
Chair: *Vuong Phan*

1 - Douglas–Rachford-based splitting for generalized DC programming with applications to signal recovery
Minh N. Dao

The difference-of-convex (DC) program is an important model in nonconvex optimization due to its structure, which encompasses a wide range of practical applications. In this work, we aim to tackle a generalized class of DC programs, where the objective function is formed by summing a possibly nonsmooth nonconvex function and a differentiable nonconvex function with Lipschitz continuous gradient, and then subtracting a nonsmooth continuous convex function. We develop a proximal splitting algorithm that utilizes proximal evaluation for the concave part and Douglas–Rachford splitting for the remaining components. The algorithm guarantees subsequential convergence to a critical point of the problem model. Under the widely used Kurdyka–Łojasiewicz property, we establish global convergence of the full sequence of iterates and derive convergence rates for both the iterates and the objective function values, without assuming the concave part is differentiable. The performance of the proposed algorithm is tested on signal recovery problems with a nonconvex regularization term and exhibits competitive results compared to notable algorithms in the literature on both synthetic data and real-world data.

2 - Robust Regression and Outlier Detection with DC Programming
Marah-Lisanne Thormann, Vuong Phan, Alain Zemkoho

Robust regression is a popular alternative to Ordinary Least Squares when outliers are present in the data. A commonly used robust regression technique is Least Trimmed Squares (LTS), utilizing only a subset of the observations to estimate the regression coefficients. Unfortunately, determining the exact solution corresponds to a combinatorial problem with an unmanageable computation time for larger data applications. Therefore, the most popular approach currently is a heuristic called Fast-LTS. In this talk, we alternatively propose the successive Boosted Difference of Convex Functions Algorithm (sBDCA) to solve the classical LTS problem. From a theoretical point of view, the approach can be seen as a combination of the Boosted Difference of Convex Functions Algorithm and Difference of Convex Functions Algorithm with successive DC decompositions. For the LTS problem, we prove that the algorithm converges to a local solution in the best case linearly, and in the worst case sublinearly. We additionally propose a problem-specific preconditioner that corrects the direction given by the gradient of the objective function, further improving the quality of the algorithmic output. In numerical experiments with synthetic and real-world data sets, we show that sBDCA with preconditioning is both significantly faster than Fast-LTS, and finds drastically lower objective function values especially in settings with many independent variables.

3 - Hilbert direct integrals of monotone operators
Nhut Minh Bui

Finite Cartesian products of operators play a central role in monotone operator theory and its applications. Extending such products to arbitrary families of operators acting on different Hilbert spaces is an open problem, which we address by introducing the Hilbert direct integral of a family of monotone operators. We present properties of this construct, as well as conditions under which the direct integral inherits the properties of the factor operators. Applications to monotone inclusion and variational problems are discussed. Joint work with Patrick Combettes.

4 - Curve Based Convex Functions and its Optimality
Qamrul Hasan Ansari

This paper introduces a new concept of convexity, namely α-convexity or curve-based convexity. It is found that an α-convex set can be disconnected—thereby making the proposed α-convexity completely different than geodesic convexity since a geodesically convex set must be path-connected. We discuss continuity, differentiability, a direction of descent, epigraph, and extremum properties of α-convex functions. It is found that a curve-based convex function may not have directional derivative. Necessary and sufficient optimality conditions of first and second orders for unconstrained optimization, and also a sufficient Karush-Kuhn-Tucker type condition for constrained optimization problems with α-convex functions are established. Further, we have proposed the concept of bifunction along a curve and discuss the nonsmooth variational inequality problems in terms of α-bifunctions. Finally, a relation providing a connection between the gap function and α-bifunctions for variational inequality problems is given. This work will allow the use of standard optimization methods and pave the road to various extensions of conventional optimization tools.

■ MC-10

Monday, 14:00-16:00 - Room: B100/8011

Optimization, Learning, and Games II

Stream: Optimization, Learning, and Games
Invited session
Chair: *Antonios Varvitsiotis*

1 - Robust Generalized Nash Equilibria
Mauro Passacantando, Sara Mattia

Robust optimization is a well-established technique for dealing with data uncertainty in optimization problems. However, most results relate to single decision-maker problems, while fewer results are available for settings with multiple decision-makers. These settings are prevalent in several applied contexts, where different stakeholders may be involved in the decision problem, with different roles and decision power. This is the case, for example, in health care, urban planning, or shift scheduling problems. The present study aims at generalizing the results known for robust optimization problems with a single decision-maker to non-cooperative games where multiple decision-makers (players) are present. In particular, we focus on the Generalized

Nash Equilibrium Problem (GNEP), where both the objective function and the feasible region of each player are affected by the actions of the other players. The robust version of a GNEP with uncertain parameters is defined, and its continuity, differentiability, convexity, and monotonicity properties are investigated. An existence result for robust equilibria is also given. In the case of linear or quadratic dependence of the objective functions and constraints on the uncertain parameters, equivalent reformulations of the robust GNEP are given.

2 - A Stable-Set Bound and Maximal Numbers of Nash Equilibria in Bimatrix Games
Constantin Ickstadt

Quint and Shubik (1997) conjectured that a non-degenerate n-by-n game has at most 2n-1 Nash equilibria in mixed strategies. The conjecture is true for n at most 4 but false for n=6 or larger. We answer it positively for the remaining case n=5, which had been open since 1999. The problem can be translated to a combinatorial question about the vertices of a pair of simple n-polytopes with 2n facets. We introduce a novel obstruction based on the index of an equilibrium, which states that equilibrium vertices belong to two equal-sized disjoint stable sets of the graph of the polytope. This bound is verified directly using the known classification of the 159,375 combinatorial types of dual neighborly polytopes in dimension 5 with 10 facets. Non-neighborly polytopes are analyzed with additional combinatorial techniques where the bound is used for their disjoint facets.

3 - Sum-of-squares for Optimal Transport: The Gromov Wasserstein Problem
Yong Sheng Soh

The Gromov-Wasserstein (GW) problem is an extension of the classical optimal transport problem to settings where the source and target distributions reside in incomparable spaces, and for which a cost function that attributes the price of moving resources is not available. The sum-of-squares (SOS) hierarchy is a principled method for deriving tractable semidefinite relaxations to generic polynomial optimization problems. The GW problem, as it is stated, is a non-convex quadratic program and is intractable to solve in general.

In this talk, we describe a moment-SOS hierarchy for solving the GW problem. We also describe extensions that remain valid for continuous measures. We establish convergence rates of the proposed SOS hierarchy, as well as sample complexity rates that arise from sampling the source and target distributions. We show that these hierarchies define pseudo-metrics over the space of (metric measure-) spaces. Finally, we discuss future directions that hint at applying SOS for general optimization problems over distributions that depend on the decision variable in a polynomial way.

4 - Inducing Cooperation in Social Dilemmas
Stefanos Leonardos

Social dilemmas illustrate situations where individual interests conflict with collective welfare, often leading to outcomes that harm the group whilst being rational for individuals. Despite this tension, real-life observations suggest that cooperation between individuals not only emerges but is key to the development of human societies. We first analyze social dilemmas through the lens of selfishness level, a game-theoretic metric that quantifies incentives for defection and prescribes the payoff modifications needed to induce prosocial behavior. We then address limitations of the canonical social dilemma model by introducing a novel multi-agent reinforcement learning mechanism that equips agents with partner choice, fostering human-aligned decision-making. Our approach promotes sustained cooperation across diverse social dilemmas and enhances learning robustness, even under unfavorable initial conditions.

■ MC-11

Monday, 14:00-16:00 - Room: B100/5017

Advances in conic optimization

Stream: Conic Optimization
Invited session
Chair: *Miguel Anjos*

1 - Efficient moment hierarchy for quadratic programming using moment cones
Shengding Sun, Fatma Kilinc-Karzan

Quadratic programming with quadratic constraints (QCQP) is a fundamental task in optimization that has many applications. The moment sum-of-squares hierarchy is a class of convex, semidefinite programming based relaxations that has been widely studied, and quantitative convergence results are known. However even the second level hierarchy is inefficient to compute in practice for moderate size inputs. We aim to mitigate the computational cost by proposing a more efficient moment hierarchy that only involves a subset of monomials or polynomials. In addition to reduction of the matrix size, it also uses the moment cone formulation instead of the usual localizing matrix, improving efficiency in important cases. We demonstrate its tightness and computational effectiveness in the case of ball constraints. Based on joint work with Fatma Kilinc-Karzan.

2 - Obtuse almost-equiangular sets
Bram Bekker

For t between -1 and 1, a set of points on the (n1)-dimensional unit sphere is called t-almost equiangular if among any three distinct points there is a pair with inner product t. We propose a semidefinite programming upper bound for the maximum cardinality α(n,t) of such a set based on an extension of the Lovász theta number to hypergraphs. This bound is at least as good as previously known bounds and for many values of n and t it is better.

We also refine existing spectral methods to show that α(n,t) is at most 2(n+1) for all n and nonpositive t, with equality only at t = 1/n. This allows us to show the uniqueness of the optimal construction at t = 1/n for n at most 5 and to enumerate all possible constructions for n equal to 2 and 3 and nonpositve t.

3 - A Positive Semidefinite Safe Approximation of Multivariate Distributionally Robust Constraints Determined by Simple Functions
Jan Rolfes, Jana Dienstbier, Frauke Liers

Single-level reformulations of (non-convex) distributionally robust optimization (DRO) problems are often intractable, as they contain semiinfinite dual constraints. Based on such a semiinfinite reformulation, we present a safe approximation, that allows for the computation of feasible solutions for DROs that depend on nonconvex multivariate simple functions. Moreover, the approximation allows to address ambiguity sets that can incorporate information on moments as well as confidence sets. The typical strong assumptions on the structure of the underlying constraints, such as convexity in the decisions or concavity in the uncertainty found in the literature can at least in part be overcome. In order to achieve algorithmic tractability, the presented safe approximation is then realized by a discretized counterpart for the semiinfinite dual constraints. The approximation leads to a computationally tractable mixed-integer positive semidefinite problem for which state-of-the-art software implementations are readily available. The tractable safe approximation provides sufficient conditions for distributional robustness of the original problem, i.e., obtained solutions are provably robust.

■ MC-12

Monday, 14:00-16:00 - Room: B100/8009

Robust optimisation and its applications

Stream: Applications: AI, uncertainty management and sustainability
Invited session
Chair: *Belen Martin Barragan*

1 - Solving a convex quadratic maximization problem appearing in some distributionally robust problem
Mathis Azema, Wim van Ackooij, Vincent Leclere

The distributionally robust optimization (DRO) framework has emerged as a powerful approach for dealing with uncertainty. In the context of Unit Commitment, where demand uncertainty affects the right-hand side of constraints, we investigate a DRO approach based on the Wasserstein distance with the $L2$-norm. This method can be addressed using Benders' Decomposition (BD) as in the risk-neutral approach. However, the key difference lies in the oracle problem that generates valid cuts for the master problem: while the risk-neutral approach requires solving a linear oracle problem at each iteration and for each scenario, the DRO formulation leads to a quadratic convex maximization oracle problem. This problem has been proven to be NP-hard, and directly solving it with a commercial solver like Gurobi may be inefficient, significantly slowing down the BD.

Since BD only requires a valid cut to progress toward convergence, solving the oracle problem to optimality at each iteration is unnecessary. However, computing an optimal solution is essential for determining an upper bound, which is needed to certify whether the BD has reached optimality.

Thus, we first propose to use the Frank-Wolfe algorithm to generate non-optimal but valid cuts by solving only linear programs. Second, we develop slower but exact methods to solve the problem to optimality, ensuring an accurate computation of the optimality gap at the end of the BD.

2 - Counterfactual explanations under uncertainty
Antonio Navas-Orozco, Emilio Carrizosa

In a Generalized Linear Model (GLM), finding a counterfactual decision of a record amounts to finding a vector with maximal outcome at a distance small enough of the record. In this paper, we extend this problem and address the challenge of building robust counterfactual decisions. Robustness is understood as guaranteeing that the predicted outcome for the counterfactual decision remains sufficiently high when the nominal probability distribution (the empirical distribution for the given training sample) is replaced by a probability distribution with the same support but different frequencies. Exploiting the structure of the Exponential Family, the problem of finding a robust counterfactual decision is expressed as a biobjective bilevel nonlinear optimization problem, whose structural properties are studied, and a numerical method is proposed.

3 - A robust variant to the smart predict-then-optimize approach.
Belen Martin Barragan, Aakil Caunhye, Xuefei Lu

In this research, we develop and explore a modified version of the smart predict-then-optimize (SPO) strategy, which considers uncertainties in data prediction and inputs when optimizing. Building on the fundamental principles of the SPO model, our method focuses on refining predictions to reduce regret when those predictions shape the parameters of an optimization problem. We shift from a fixed, deterministic approach to one where data inaccuracies introduce uncertainty, and we apply robust optimization methods to address these uncertainties. Specifically, we study three types of robustness (worst-case robustness, strict robustness, and intermediate robustness) that tolerate varying levels of suboptimality and thus replicate different robustness-enforcing strategies. We assess our robust optimization models considering both uncertainties in the predictions and in the covariates. Our numerical results show significant out-of-sample performance improvements under randomly generated covariate disturbances, compared to the classic SPO approach, even when a small sample size is used.

4 - Beyond One-Hot Labels: KL-Divergence Training with Empirical Distributions for Faster Optimization
Arman Bolatov

We introduce an alternative training approach for deep learning classification that enhances standard pipelines by minimizing KL divergence against approximated true underlying distributions rather than cross-entropy on one-hot labels. Our work demonstrates two effective strategies: first, training models using KL divergence against distributions that accurately reflect label relationships and potential ambiguities; second, leveraging locality-sensitive hashing to create empirical distributions from semantically similar examples. Experiments on image classification tasks show these approaches lead to more faster and stable optimization. We further establish the effectiveness of teacher-student knowledge transfer, where the teacher models trained with KL divergence on the approximated true distributions successfully guide new networks, outperforming traditional training methods particularly when labels contain noise or ambiguity.

■ MC-13

Monday, 14:00-16:00 - Room: B100/6009

Cardinality control in optimization problems for Data Science

Stream: Sparsity guarantee and cardinality-constrained (MI)NLPs
Invited session
Chair: *E. Alper Yildirim*

1 - An Efficient Network-aware Direct Search Method for Influence Maximization

Matteo Bergamaschi, Francesco Rinaldi, Francesco Tudisco, Sara Venturini

Networks play a crucial role in understanding real-world scenarios by revealing patterns, key nodes, and predicting information, disease, or innovation spread. The Influence Maximization problem, a significant challenge in network analysis, targets identifying influential individuals for various domains like marketing, public health, and social media. In this talk, we introduce a more comprehensive model for information propagation on networks, catering to a broader range of dynamics. Then we introduce Network-aware Direct Search (NaDS), a new direct search method tailored for the Influence Maximization problem. Unlike conventional methods, our approach guarantees good performances in diverse settings where traditional approaches falter, even in large-scale networks with thousands of nodes.

2 - Strict cardinality control in feature selection for linear Support Vectors: a scalable conic decomposition approach

Laura Palagi, Immanuel Bomze, Federico D'Onofrio, Bo Peng

In this talk, we present the embedded feature selection problem in linear Support Vector Machines (SVMs), in which an explicit cardinality constraint is employed. These models try to achieve fairness, transparency and explainability in AI applications, ranging from Math. Finance/Economics to social and life sciences. The problem is NP-hard due to the presence of the cardinality constraint, even though the original linear SVM amounts to a problem solvable in polynomial time. To handle the hard problem, we introduce two mixed-integer formulations for which novel semidefinite relaxations are proposed. Exploiting the sparsity pattern of the relaxations, we decompose the problems and obtain equivalent relaxations in a much smaller cone, making the conic approaches scalable. We propose heuristics using the information on the optimal solution of the relaxations. Moreover, an exact procedure is proposed by solving a sequence of mixed-integer decomposed semidefinite optimization problems. Numerical results on classical benchmarking datasets are reported, showing the efficiency and effectiveness of our approach.

3 - Exact sparsity control for multiclass linear Support Vector Machines

Pedro Duarte Silva, Immanuel Bomze, Laura Palagi, Bo Peng, Marta Monaci, Federico D'Onofrio

Recent advances in Machine Learning have lead to the pervasiveness of AI applications that rely on highly accurate classification algorithms. However, in many applications, because of transparency or legal requirements, algorithm interpretability is an essential condition that needs to be considered hand in hand with classification accuracy. In particular case of multiclass classification problems, while state of art deep learning neural networks and kernel based Support Vector Machines often excel in minimizing expected error rates, they also work as uninterpretable black boxes that cannot be use in many problems with strict transparency constraints. A viable alternative under these conditions is to rely on sparse linear multiclass Support Vector Machines, that impose a limit on the cardinality of the feature set effectively considered. Most existing approaches to tackle this problem add surrogate regularizers to the chosen SVM criteria, in order to reduce the number of features used by the classifier. However, this approach sometimes does not work as intended and, in this presentation, we will discuss a class of sparse multiclass linear SVMs that make a rigorous explicit control on the cardinality of feature set employed. Numerical results on classical benchmarking datasets will be reported, showing the efficiency and effectiveness of our approach.

Monday, 16:30-18:30

■ MD-01

Monday, 16:30-18:30 - Room: B100/1001

Derivative-Free Optimization Methods for challenging applications: Handling Nonsmoothness and Constraints

Stream: Zeroth and first-order optimization methods
Invited session
Chair: *Francesco Rinaldi*
Chair: *Andrea Cristofari*

1 - Derivative-free Penalty-IPM for nonsmooth constrained optimization
Andrea Brilli, Youssef Diouane, Sébastien Le Digabel, Giampaolo Liuzzi, Christophe Tribes

We propose an Penalty-Interior Point method to solve nonsmooth constrained optimization problems. The unconstrained subproblems are solved using a derivative-free Mesh Adaptive Direct Search approach. Assuming local Lipschitz continuity of the limit points, we prove various stationarity properties using regularity conditions for the constraints. The approach has been implemented into the NOMAD solver, and we have performed experiments to compare the new strategy against existing ones.

2 - Complexity of a Riemannian Direct-search algorithm
Bastien Cavarretta, Clément Royer, Florian Yger, Florentin Goyens

Direct-search algorithms are derivative-free optimization techniques that operate by polling the variable space along specific directions forming a positive spanning set (PSS). We investigate the construction of PSSs when the variables are constrained to a Riemannian manifold, and polling must be performed along tangent directions. We show that projecting a PSS from the ambient space to the tangent space may lead to worse complexity guarantees than generating directions directly in the tangent space. Our numerical experiments illustrate the practical benefit of the latter construction.

3 - TRFD: A derivative-free trust-region method based on finite differences for composite nonsmooth optimization
Dânâ Davar, Geovani Grapiglia

We present TRFD, a derivative-free trust-region method based on finite differences for minimizing composite nonsmooth functions. We establish a worst-case evaluation complexity bound that TRFD needs to find an approximate stationary point. For L1 and Minimax problems, the bound depends linearly on the number of variables for specific instances of TRFD, while this bound depends quadratically on the inverse of the desired accuracy. In addition, if the objective function is convex, the previous complexity can be improved for Minimax problems to a linear dependence on the inverse of the desired accuracy. Numerical results are also reported, illustrating the relative efficiency of TRFD.

4 - On the Pareto-efficient points of Simple Constrained Multiobjective Problems
Dimo Brockhoff

Numerical benchmarking is a crucial aspect of designing and recommending practically relevant multiobjective optimization algorithms. It helps in terms of the interpretation of benchmarking experiments if we fundamentally understand the used benchmark problems and their properties—in particular with respect to their Pareto-efficient points (also known as the Pareto set). There is no fundamental difference when we move towards more realistic *constrained* multiobjective problems.

In this talk, I will discuss, as a first step towards better understanding of multiobjective *constrained* problems, our recent results on the Pareto sets of some of the simplest multiobjective constrained problems. The focus of my talk will be on problems where each single objective function is convex quadratic and the constraints are linear. Building on classical results from convex analysis, our main result proves that the Pareto set of the constrained problem equals the projection of the unconstrained Pareto set into the feasible (convex) set. Besides discussing the theory, I will visually show the Pareto sets and fronts of some problems and compare them to the performance of the well-established NSGA-II algorithm.

This is joint work with Anne Auger (Inria and IP Paris, France), Jordan N. Cork and Tea Tušar (both Jožef Stefan Institute and Jožef Stefan International Postgraduate School, Slovenia).

■ MD-02

Monday, 16:30-18:30 - Room: B100/7011

Optimization in machine Learning

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Alireza Kabgani*
Chair: *Lorenzo Ciarpaglini*

1 - Provable Reduction in Communication Rounds for Non-Smooth Convex Federated Learning
Karlo Palenzuela, Ali Dadras, Alp Yurtsever

Multiple local steps are key to communication-efficient federated learning. However, theoretical guarantees for such algorithms, without data heterogeneity-bounding assumptions, have been lacking in general non-smooth convex problems. Leveraging projection-efficient optimization methods, we propose FedMLS, a federated learning algorithm with provable improvements from multiple local steps. FedMLS attains an $epsilon$-suboptimal solution in $mathcalO(1/epsilon)$ communication rounds, requiring a total of $mathcalO(1/epsilon2)$ stochastic subgradient oracle calls.

2 - Revisiting LocalSGD and SCAFFOLD: Improved Rates and Missing Analysis
Ruichen Luo, Sebastian Stich, Samuel Horvath, Martin Takac

LocalSGD and SCAFFOLD are widely used methods in distributed stochastic optimization, with numerous applications in machine learning, large-scale data processing, and federated learning. However, rigorously establishing their theoretical advantages over simpler methods, such as minibatch SGD (MbSGD), has proven challenging, as existing analyses often rely on strong assumptions, unrealistic premises, or overly restrictive scenarios. In this work, we revisit the convergence properties of LocalSGD and SCAFFOLD under a variety of existing or weaker conditions, including gradient similarity, Hessian similarity, weak convexity, and Lipschitz continuity of the Hessian. Our analysis shows that (i) LocalSGD achieves faster convergence compared to MbSGD for weakly convex functions without requiring stronger gradient similarity assumptions; (ii) LocalSGD benefits significantly from higher-order similarity and smoothness; and (iii) SCAFFOLD demonstrates faster convergence than MbSGD for a broader class of non-quadratic functions. These theoretical insights provide a clearer understanding of the conditions under which LocalSGD and SCAFFOLD outperform MbSGD.

3 - Convex Formulations For Training Two-Layer ReLU Neural Networks
Karthik Prakhya, Tolga Birdal, Alp Yurtsever

Solving non-convex, NP-hard optimization problems is crucial for training machine learning models, including neural networks. However, non-convexity often leads to black-box machine learning models with unclear inner workings. While convex formulations have been used for verifying neural network robustness, their application to training neural networks remains less explored. In response to this challenge, we reformulate the problem of training infinite-width two-layer ReLU networks as a convex completely positive program in a finite-dimensional (lifted) space. Despite the convexity, solving this problem remains NP-hard due to the complete positivity constraint. To overcome this challenge, we introduce a semidefinite relaxation that can be solved in polynomial time. We then experimentally evaluate the tightness of this relaxation, demonstrating its competitive performance in test accuracy across a range of classification tasks.

4 - Exploring Step Size Adaptation in Large-Scale Deep Learning Optimization
Lorenzo Ciarpaglini, Laura Palagi, Diego Scuppa, Marco Sciandrone

Scaling deep learning optimization remains a fundamental challenge, especially in relation to the choice of step size within first-order methods. In this work, we explore adaptive strategies for learning rate selection designed to improve both convergence behavior and robustness across different architectures, tasks, and datasets. Rather than relying on fixed schedules or heuristic tuning, our approach aims to incorporate meaningful information from the optimization landscape to guide parameter updates. The resulting methods are flexible and can be integrated into standard training procedures with minimal overhead. An empirical study is conducted to assess the performance of different adaptive step size strategies, with a focus on their stability, efficiency, and integration within standard training pipelines.

■ MD-03

Monday, 16:30-18:30 - Room: B100/4011

Variational Methods in Set and Vector Optimization

Stream: Multiobjective and Vector Optimization
Invited session
Chair: *Radu Strugariu*

1 - Subdifferential conditions for sharp and approximate solutions in nonsmooth optimization
Marius Durea

We present new results on optimality conditions for constrained scalar and set-valued nonsmooth optimization problems, with a focus on sharp, isolated, and approximate solutions. Utilizing various types of subgradients, coderivatives, and Shapiro properties for sets, we characterize different degrees of sharpness and extend existing results from the literature. Additionally, we explore approximate minimality for the same class of problems and establish necessary conditions under mild assumptions on the objective function. Our findings provide new insights into certain optimality conditions in nonsmooth optimization and further expand the existing theoretical framework.

2 - Optimality conditions for nondominated and approximate solutions in vector optimization with variable domination structure
Radu Strugariu

Within the framework of variable domination structures, we demonstrate that the approaches used in the literature cover each other. This observation enables us to design unified methods for deriving necessary optimality conditions in both cases. Moreover, we extend the method to the case of approximately nondominated points.

3 - On Derivative-Free Methods for Solving Set-based Robust Counterparts to Uncertain Multiobjective Optimization Problems
Christian Günther

This talk is devoted to derivative-free methods for solving set-based robust counterparts to uncertain multiobjective optimization problems. The motivation comes from uncertain multiobjective engineering problems for damage localization and quantification in flexible mechanical structures, which are in fact nonsmooth nonlinear multiobjective parameter estimation problems. We focus on robustness concepts (e.g., set-based min-max robust efficiency, set-based optimistic robust efficiency) given by solution concepts for the set-based robust counterpart problems based on preorder set relations (e.g., upper set less relation, lower set less relation). We illustrate an algorithmic pattern search procedure for approximating solutions to the set-based robust counterpart problems (with special emphasis on the case with a finite number of uncertainties), which is based on set-based (first- and higher-level) non-dominated sorting for finite families of sets and an infinite penalty approach, together with some implementation details of the procedure to improve numerical efficiency.

■ MD-05

Monday, 16:30-18:30 - Room: B100/4013

Relaxed Smoothness and Convexity Assumptions in Optimization for Machine Learning

Stream: Optimization for machine learning
Invited session
Chair: *Eduard Gorbunov*

1 - Loss Landscape Characterization of Neural Networks without Over-Parametrization
Rustem Islamov

Optimization methods play a crucial role in modern machine learning, powering the remarkable empirical achievements of deep learning models. These successes are even more remarkable given the complex non-convex nature of the loss landscape of these models. Yet, ensuring the convergence of optimization methods requires specific structural conditions on the objective function that are rarely satisfied in practice. One prominent example is the widely recognized Polyak-Łojasiewicz (PL) inequality, which has gained considerable attention in recent years. However, validating such assumptions for deep neural networks entails substantial and often impractical levels of over-parametrization. In order to address this limitation, we propose a novel class of functions that can characterize the loss landscape of modern deep models without requiring extensive over-parametrization and can also include saddle points. Crucially, we prove that gradient-based optimizers possess theoretical guarantees of convergence under this assumption. Finally, we validate the soundness of our new function class through both theoretical analysis and empirical experimentation across a diverse range of deep learning models.

2 - Methods for Convex (L0,L1)-Smooth Optimization: Clipping, Acceleration, and Adaptivity
Eduard Gorbunov

Due to the non-smoothness of optimization problems in Machine Learning, generalized smoothness assumptions have been gaining a lot of attention in recent years. One of the most popular assumptions of this type is (L0,L1)-smoothness (Zhang et al., 2020). In this talk, we focus on the class of (strongly) convex (L0,L1)-smooth functions and discuss new convergence guarantees for several existing methods. In particular, we discuss improved convergence rates for Gradient Descent with (Smoothed) Gradient Clipping and for Gradient Descent with Polyak Stepsizes. We also extend these results to the stochastic case under the over-parameterization assumption, propose a new accelerated method for convex (L0,L1)-smooth optimization, and derive new convergence rates for Adaptive Gradient Descent (Malitsky and Mishchenko, 2020).

3 - Optimizing (L_0, L_1)-Smooth Functions by Gradient Methods
Anton Rodomanov, Daniil Vankov, Angelia Nedich, Lalitha Sankar, Sebastian Stich

We study gradient methods for optimizing (L_0, L_1)-smooth functions, a class that generalizes Lipschitz-smooth functions and has gained attention for its relevance in machine learning. We provide new insights into the structure of this function class and develop a principled framework for analyzing optimization methods in this setting. While our convergence rate estimates recover existing results for minimizing the gradient norm in nonconvex problems, our approach significantly improves the best-known complexity bounds for convex objectives. Moreover, we show that the gradient method with Polyak stepsizes and the normalized gradient method achieve nearly the same complexity guarantees as methods that rely on explicit knowledge of (L_0, L_1). Finally, we demonstrate that a carefully designed accelerated gradient method can be applied to (L_0, L_1)-smooth functions, further improving all previous results.

4 - A Third-Order Perspective on Newton's Method and its Application in Federated Learning
Slavomír Hanzely

This paper investigates the global convergence of stepsized Newton methods for convex functions with Hölder continuous Hessians or third derivatives for the federated learning. With focus on single node setup, we propose several simple stepsize schedules with fast global convergence guarantees, up to $mathcalO (1/k3)$. For cases with multiple plausible smoothness parameterizations or an unknown smoothness constant, we introduce a stepsize linesearch and a backtracking procedure with provable convergence as if the optimal smoothness parameters were known in advance. Additionally, we present strong convergence guarantees for the practically popular Newton method with exact linesearch.

■ MD-06

Monday, 16:30-18:30 - Room: B100/7013

Smoothing techniques for nonsmooth optimization

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Olivier Fercoq*

1 - Second-order proximal-gradient methods for avoiding nonsmooth strict saddle points
Alexander Bodard, Masoud Ahookhosh, Panagiotis Patrinos

This work introduces two proximal-gradient-based algorithms – a trust-region method and a curvilinear linesearch method – for minimizing the sum of a smooth function f and a nonsmooth function g. We prove that these methods converge to second-order stationary points of the objective whenever the proximal mapping of g is locally of class C1 around around the forward points of the limit points. If g belongs to the broad class of C2-partly smooth functions, this last requirement is shown to hold under a mild strict complementarity condition. We highlight, in particular, that these results do not require local smoothness of the objective around the limit points. To the best of our knowledgde, the presented algorithms constitute the first second-order proximal-gradient algorithms which provably escape nonsmooth strict saddle points, irrespective of the initial iterate. Our approach is based on the so-called forward-backward envelope (FBE), which has previously enabled the design of globally convergent linesearch methods. As an intermediate step in our analysis, we describe conditions under which the FBE is locally of class C2 around critical points, and establish an equivalence between the second-order stationary points of the original objective and those of the FBE. Finally, numerical experiments validate the proposed methods, both on illustrative toy problems, and on sparse principal component analysis and phase retrieval problems.

2 - A Proximal Variable Smoothing for Nonsmooth Minimization of the Sum of Three Functions Including Weakly Convex Composite Function
Keita Kume, Isao Yamada

We propose a proximal variable smoothing algorithm for a nonsmooth optimization problem whose cost function is the sum of three functions including a weakly convex composite function. The proposed algorithm has a single-loop structure inspired by a proximal gradient-type method. More precisely, the proposed algorithm consists of two steps: (i) a gradient descent of a time-varying smoothed surrogate function designed partially with the Moreau envelope of the weakly convex function; (ii) an application of the proximity operator of the remaining function not covered by the smoothed surrogate function. We also present a convergence analysis of the proposed algorithm by exploiting a novel asymptotic approximation of a gradient-mapping-type stationarity measure.

3 - Analyzing the speed of convergence in nonsmooth optimization via the Goldstein epsilon-subdifferential
Bennet Gebken

In smooth optimization, Taylor expansion is a powerful tool when analyzing the speed of convergence of solution methods. In nonsmooth optimization, this tool cannot be applied anymore, as there is no suitable generalization of Taylor expansion for a nonsmooth function. As a result, while many different solution methods for nonsmooth problems have been proposed, the speeds of convergence of these methods are rarely analyzed. In this talk, I will present a novel approach based on the Goldstein epsilon-subdifferential for analyzing convergence behavior in nonsmooth optimization. More precisely, given a converging sequence and upper bounds for the distance of the epsilon-subdifferential to zero for vanishing epsilon, we will derive an estimate for the distance of said sequence to the minimum. The assumptions we require for this are polynomial growth around the minimum and, depending on the order of growth, a higher-order generalization of semismoothness. After giving an overview of the assumptions and the theoretical results, I will show how these results lead to a better understanding of the behavior of gradient sampling methods.

4 - Minimizing the smoothed gap to solve saddle point problems
Olivier Fercoq

In this work, we minimize the self-centered smoothed gap, a recently introduced optimality measure, in order to solve convex-concave saddle point problems. The self-centered smoothed gap can be computed as the sum of a convex, possibly nonsmooth function and a smooth weakly convex function. Although it is not convex, we propose an algorithm that minimizes this quantity, effectively reducing convex-concave saddle point problems to a minimization problem. Its worst case complexity is comparable to the state of the art, and the algorithm enjoys linear convergence in favorable cases.

■ MD-07

Monday, 16:30-18:30 - Room: B100/5015

Methods for simple and nonsmooth bilevel optimization

Stream: Bilevel and multilevel optimization
Invited session
Chair: *Sebastian Lämmel*

1 - A Simple Algorithm for Simple Bilevel Programming
Joydeep Dutta

In this talk we are going to discuss a simple yet effective algorithm for a simple bilevel programming problem. We seek to minimize a convex function over the solution set of a convex optimization problem with smooth objective. This is a problem where the Slater condition fails and we provide a simple yet novel way to develop an algorithm and prove it's convergence without the Lipschitz gradient assumption on the lower-level objective. We also provide some numerical experiments to show the effectiveness of our algorithm. This is joint work with S. Dempe, T. Pandit and K. M. Rao.

2 - Approximation and exact penalization in simple bilevel variational problems
Riccardo Tomassini, Giancarlo Bigi

In this talk we analyse a simple bilevel problem, where the lower-level is a monotone variational inequality and the upper-level is a convex optimization problem. Regularity is gained by adding a given degree of inexactness to the lower-problem and allows us to leverage techniques of exact penalization, bounding the growth of the penalization parameter. The resulting penalized problem relies on the reformulation of the lower-level through the Minty gap function, that we approximate with cutting planes techniques. We devise different algorithms to treat the affine and non-affine cases, we discuss their convergence and provide theoretical bounds on the final degree of inexactness. Numerical results are reported for the non-affine case to test the sensitivity of the problem to several parameters, showing that the true final inexactness is generally meaningfully lower than the theoretical one.

3 - Anomalies of the Scholtes regularization for mathematical programs with complementarity constraint
Sebastian Lämmel, Vladimir Shikhman

For mathematical programs with complementarity constraints (MPCC), we refine the convergence analysis of the Scholtes regularization. Our goal is to relate nondegenerate C-stationary points of MPCC with nondegenerate Karush-Kuhn-Tucker points of its Scholtes regularization. We detected the following anomalies: (i) in a neighborhood of a nondegenerate C-stationary point there could be degenerate Karush-Kuhn-Tucker points of the Scholtes regularization; (ii) even if nondegenerate, they might be locally non-unique; (iii) if nevertheless unique, their quadratic index potentially differs from the C-index of the C-stationary point under consideration. Thus, a change of the topological type for Karush-Kuhn-Tucker points of the Scholtes regularization is possible. In particular, a nondegenerate minimizer of MPCC might be approximated by saddle points. In order to bypass the mentioned anomalies, an additional generic condition for nondegenerate C-stationary points of MPCC is identified. Then, we uniquely trace nondegenerate Karush-Kuhn-Tucker points of the Scholtes regularization and successively maintain their topological type. As a byproduct, we refute a result on the well-posedness of Scholtes regularization wrongly proven in literature.

■ MD-08

Monday, 16:30-18:30 - Room: B100/7007

Systematic and computer-aided analyses III: noisy gradient methods and fixed-point algorithms

Stream: Systematic and computer-aided analyses of optimization algorithms
Invited session
Chair: *Baptiste Goujaud*
Chair: *François Glineur*

1 - Tight analyses of first-order methods with error feedback
Daniel Berg Thomsen

Error feedback (EF), including newer variants such as EF21, are used in distributed optimization to reduce communication bandwidth. Recently, methods using error feedback were proposed and analyzed in different settings. E.g. in single and multi-worker settings, with or without stochastic gradient oracles, etc. There are many different claims regarding both the theoretical and practical performance of these methods in different contexts. In this talk, we intend to provide a clearer picture of the pros and cons of some of these methods. We will first focus on the single-worker setting using deterministic gradients, as it presents a simpler problem statement with fewer possible variations. This allows us to get a better understanding of the performance of these methods and to meaningfully compare them with one another. To do so, we use an automated approach for identifying simple yet optimal Lyapunov functions based on the performance estimation problem (PEP) framework, along with functions achieving those bounds.

2 - On fundamental proof structures in first-order optimization
Aymeric Dieuleveut, Baptiste Goujaud, Adrien Taylor

First-order optimization methods have attracted a lot of attention due to their practical success in many applications, including in machine learning. Obtaining convergence guarantees and worst-case performance certificates for first-order methods have become crucial for understanding ingredients underlying efficient methods and for developing new ones. However, obtaining, verifying, and proving such guarantees is often a tedious task. Therefore, a few approaches were proposed for rendering this task more systematic, and even partially automated. In addition to helping researchers finding convergence proofs, these tools provide insights on the general structures of such proofs. We aim at presenting those structures, showing how to build convergence guarantees for first-order optimization methods.

3 - An invariance theory for complete characterization of exact optimal fixed-point algorithm family
TaeHo Yoon, Ernest Ryu, Benjamin Grimmer

For nonexpansive fixed-point problems, both Halpern's method with optimal parameters and its so-called H-dual algorithm exhibit the exact optimal worst-case convergence rates. In this work, we completely characterize the infinite family of distinct algorithms using predetermined step-sizes, represented as lower triangular H-matrices, all of which attain the same optimal convergence rates. The characterization is based on algebraic quantities that we call H-invariants, whose values stay constant over all optimal H-matrices. The theory of H-invariants offers a novel view of optimal acceleration in first-order optimization as an algebraic study of carefully selected invariants and structures induced by them.

4 - Performance Estimation framework for a range of stochastic first-order methods.
Anne Rubbens, Julien Hendrickx

We rely on a computer-assisted technique to automatically obtain worst-case performance guarantees for a variety of stochastic first-order methods, and illustrate how it enables improvements over existing guarantees on several numerical examples.

This technique allows dealing with a wide range of stochastic settings, including e.g. unified variance model and finite-sum optimization, and can be applied to the analysis of non variance reduced (e.g. SDG) and variance reduced (e.g. SAGA) methods.

■ MD-09

Monday, 16:30-18:30 - Room: B100/8013

Generalized convexity and monotonicity 3

Stream: Generalized convexity and monotonicity
Invited session
Chair: *Felipe Lara*

1 - On an alternative to convexity of the values in set-valued analysis: from theory to applications
Annagiulia Pezzola, Didier Aussel, Rossana Riccardi, Domenico Scopelliti

Set-valued maps play a crucial role in variational analysis, optimization and fixed point theory. In particular, the convexity of the values of a set-valued map is a key assumption in establishing fundamental results in these fields: indeed, moving toward the applications, it is a requirement that has to be satisfied although sometime it could be too demanding. This motivates the introduction of an alternative framework that can handle non-convex values while preserving the mathematical robustness. The aim of this talk is to describe this alternative approach with the related applications.

2 - Fixed-time convergent second-order time-varying dynamical systems
Lien Nguyen

Addressing optimization problems using the finite-time/fixed-time stability properties of dynamical systems has received increasing interest. In this work, we propose new fixed-time second-order time-varying dynamical systems. We derive a sufficient condition for the existence and uniqueness of solutions to a general ordinary differential equation (ODE) and present a rigorous proof establishing the well-posedness of our proposed algorithms, addressing a significant gap in the literature. We then employ the proposed dynamical systems to solve nonsmooth additive composite optimization problems.

3 - On finite termination of quasi-Newton methods on quadratic problems
Aban Ansari-Önnestam, Anders Forsgren

Quasi-Newton methods form an important class of methods for solving nonlinear optimization problems. In such methods, first order information is used to approximate the second derivative. The aim is to mimic the fast convergence that can be guaranteed by Newton-based methods. In the best case, quasi-Newton methods will far outperform steepest descent and other first order methods, without the computational cost of calculating the exact second derivative. These guarantees are a local result, that follow closely from the fact that such an objective function is strongly convex and can be approximated by a quadratic function close to the solution. Understanding the performance of quasi-Newton methods on quadratic problems with a symmetric positive definite Hessian is therefore of vital importance. This talk will discuss ways in which the reliance on line search and a full approximation of the Hessian can be relaxed, and how these changes affect the behavior of quasi-Newton methods. It is possible to achieve termination in at most one iteration more than termination of methods such as CG and BFGS based on only first order information, without any type of line search except for the last step.

■ MD-10

Monday, 16:30-18:30 - Room: B100/8011

Interactions between optimization and machine learning

Stream: Zeroth and first-order optimization methods
Invited session
Chair: *Cesare Molinari*
Chair: *Silvia Villa*

1 - From learning to optimize to learning optimization algorithms
Camille Castera, Peter Ochs

Towards designing learned optimization algorithms that are usable beyond their training setting, we identify key principles that classical algorithms obey, but have up to now, not been used for Learning to Optimize (L2O). Following these principles, we provide a general design pipeline, taking into account data, architecture and learning strategy, and thereby enabling a synergy between classical optimization and L2O, resulting in a philosophy of Learning Optimization Algorithms. As a consequence our learned algorithms perform well far beyond problems from the training distribution. We demonstrate the success of these novel principles by designing a new learning-enhanced BFGS algorithm and provide numerical experiments evidencing its adaptation to many settings at test time.

2 - The Surprising Agreement Between Convex Optimization Theory and Learning-Rate Scheduling for Large Model Training
Fabian Schaipp, Alexander Hägele, Adrien Taylor, Umut Simsekli, Francis Bach

We show that learning-rate schedules for large model training behave surprisingly similar to a performance bound from non-smooth convex optimization theory. We provide a bound for the constant schedule with linear cooldown; in particular, the practical benefit of cooldown is reflected in the bound due to the absence of logarithmic terms. Further, we show that this surprisingly close match between optimization theory and practice can be exploited for learning-rate tuning: we achieve noticeable improvements for training 124M and 210M Llama-type models by (i) extending the schedule for continued training with optimal learning-rate, and (ii) transferring the optimal learning-rate across schedules.

3 - Greedy learning to optimise with convergence guarantees
Patrick Fahy, Mohammad Golbabaee, Matthias J. Ehrhardt

Learning to optimise (L2O) leverages training data to accelerate solving optimisation problems. Many existing methods use unrolling to parameterise update steps, but this often leads to memory limitations and a lack of convergence guarantees. We introduce a novel greedy strategy that learns iteration-specific parameters by minimising the function value at the next step. This approach enables training over significantly more iterations while keeping GPU memory usage constant. We focus on preconditioned gradient descent with multiple parameterisations, including a novel convolutional preconditioner. Our method ensures that parameter learning is no harder than solving the initial optimisation problem and provides convergence guarantees. We validate our approach on inverse problems such as image deblurring and Computed Tomography, where our learned convolutional preconditioners outperform classical methods like Nesterov's Accelerated Gradient and L-BFGS.

4 - Learning from data via overparameterization
Cesare Molinari, Silvia Villa, Lorenzo Rosasco, Cristian Vega, Hippolyte Labarrière

Solving data driven problems requires defining complex models and fitting them on data, neural networks being a motivating example. The fitting procedure can be seen as an optimization problem which is often non convex, and hence optimization guarantees hard to derive. An opportunity is provided by viewing the model of interest as a redundant re-parameterization - an overparametrization - of some simpler model for which optimization results are easier to achieve. In this talk, after formalizing the above idea, we review some recent results and derive new ones. In particular, we consider the gradient flow of some classes of linear overparamtetrization and show they correspond to suitable mirror flow on the original parameters. Our main contribution relates to the study of the latter, for which we establish well posed-ness and convergence. The results yields insight on the role of overparametrization for implicit regularization and constrained optimization.

■ MD-11

Monday, 16:30-18:30 - Room: B100/5017

Applications of conic optimization

Stream: Conic Optimization
Invited session
Chair: *Miguel Anjos*

1 - The application of SDP-based relaxation for the nordic day-ahead energy market
Shudian Zhao, Jan Kronqvist

Social welfare maximization problems in the day-ahead energy market can be formulated as large-scale mixed-integer programming (MIP) problems. These problems are computationally challenging due to the integrality constraints and the high dimensionality introduced by temporal coupling across time periods. Standard linear relaxations often provide weak bounds, limiting the efficiency of branch-and-bound algorithms.

In this talk, we present a semidefinite programming (SDP)-based relaxation approach for these MIP problems. We analyze the modeling power of various classes of valid inequalities and demonstrate, through numerical results, how they contribute to tightening the relaxation and improving computational performance.

2 - Aggregated Formulations and Relaxations of the Stable Set Polytope
Haobang Zhou, Akshay Gupte, E. Alper Yildirim

The stable set polytope admits a binary quadratic representation, whose Shor relaxation is the well-known theta body. We consider aggregations of quadratic constraints before and after applying the Shor relaxation, thus yielding new families of semidefinite relaxations. Conditions are established for the well-known valid inequalities for the stable set polytope to remain valid for the aggregations. We also compare the closures of such aggregations and establish the surprising result that aggregations preserving exact formulations do not necessarily yield stronger relaxations than their inexact counterparts. In particular, one of our aggregation closures equals the theta body, whereas the other one equals a relaxation of the theta body obtained by relaxing equalities to inequalities.

This is joint work with Akshay Gupte and Alper Yildirim.

3 - Hybrid scalarization for multi-objective semi definite polynomial optimization
Sujeet Kumar Singh

Decision making is part of our lives, and almost all decision problems have conflicting criteria for evaluation. Such problems are referred to as multi-objective optimization problems (MOOPs). We intend to develop new scalarization techniques using a modified goal function for the MOOP. Although multiobjective optimization problems have a rich literature on scalarizing methods, the existing weighing scalarization methods have some deficiencies in assigning the weights and then finding the solution per the objectives' priority to tackle the incommensurability in heterogeneous objectives. Further, the weighing methods are unable to generate the Pareto points, which fall on the nonconvex part of the Pareto front. We consider these issues and propose the Gamma-connective scalarization technique to solve the multi-objective optimization problem. The underline functions are considered to be the higher-degree polynomials and the recently developed sum of squares(SOS) techniques is used for equivalent PSD conversion. Positive semidefinite (PSD) optimization has efficient tools like YALMIP and SOSTOOLS to work with efficiently. These tools make PSD a tractable convex optimization problem. The performance of the proposed method would be evaluated using some measure of closeness to the ideal solution for several test problems.

■ MD-12

Monday, 16:30-18:30 - Room: B100/8009

Applications of optimisation under uncertainty

Stream: Applications: AI, uncertainty management and sustainability
Invited session
Chair: *Nicole Cristina Cassimiro de Oliveira*

1 - Phase recovery from masked phaseless antenna measurements
Sakirudeen Abdulsalaam, Adrien Guth, Holger Rauhut, Dirk Heberling

The radiation characteristics of an antenna under test (AUT) is one of the most important antenna properties. Spherical Near Field (SNF) measurements are known to be the most accurate characterization method. Despite its accuracy, SNF measurements pose several challenges, including the need for a significant number of samples and complicated mathematical transformation to derive the AUT's far-field (FF) radiation pattern from the complex near-field (NF) measurements. Furthermore, the phase acquisition becomes more challenging at higher frequencies. Therefore, research into AUT's characterization with measurements and transformation techniques based on amplitude information only has gained traction. The key challenge in this case is to compute coefficients describing the AUT's radiation behaviour from amplitude NF measurements. PhaseLift, a convex programming technique, has been shown in the literature to be one of the successful methods for phase recovery from discrete Fourier Transform (DFT) measurements with random masks. In this work, we extend this technique to antenna measurements. We prove stable recovery by showing that the sampling operator is well-conditioned. Numerical experiments with noiseless data corroborate the theory.

2 - Controlled stochastic processes for simulated annealing
Vincent Molin, Axel Ringh

Simulated annealing solves optimization problems by means of a random walk in an energy landscape based on the objective function and a temperature parameter. By slowly decreasing the temperature, the algorithm converges to the global optimal solution, also for nonconvex functions. However, if the temperature is decreased too quickly, this procedure often gets stuck in local minima. To overcome this, we here present a new perspective on simulated annealing. More precisely, we consider the cooling landscape as a curve of probability measures and prove that there exists a minimal norm velocity field which solves the continuity equation. The latter is a differential equation which governs the evolution of the aforementioned curve. The solution is the weak gradient of an integrable function, which is in line with the interpretation of the velocity field as a derivative of optimal transport maps. We also show that controlling stochastic annealing processes by superimposing this velocity field would allow them to follow arbitrarily fast cooling schedules. Based on these findings, we design novel interacting particle based optimization methods, convergent optimal transport based approximations to this control, that accelerate simulated annealing processes. This acceleration behavior is also validated on a number of numerical experiments.

3 - Robust treatment planning in proton therapy
Nicole Cristina Cassimiro de Oliveira, Aurelio Oliveira, Juliana Campos de Freitas

Proton therapy is a precise cancer treatment technique, but uncertainties in the Bragg peak position can impact its effectiveness. This study focuses on mathematical optimization to enhance treatment robustness. A minimax-based model is proposed to minimize the maximum tumor dose while preserving critical organs. The model is validated with TROTS database cases, particularly in head and neck cancer, where structures like the oral cavity, larynx, and brainstem are highly sensitive to radiation. To address anatomical and positioning variations, a multi-objective optimization approach is introduced, incorporating goal programming. This strategy balances tumor coverage and organ-at-risk protection, resulting in more reliable treatment plans. The findings highlight the potential of mathematical models to improve proton therapy's safety and effectiveness.

■ MD-13

Monday, 16:30-18:30 - Room: B100/6009

Recent advances in optimization problems with cardinality constraints

Stream: Sparsity guarantee and cardinality-constrained (MI)NLPs
Invited session
Chair: *Laura Palagi*

1 - Critical Point Theory for Sparse Huber Recovery
Deniz Akkaya, Mustafa Pinar

We study the problem of sparse recovery in compressed sensing, aiming to minimize sensing error in linear measurements using sparse vectors with a fixed number of nonzero entries. To improve robustness against outliers, we employ the Huber loss, which effectively reduces the influence of Gaussian noise while preserving sensitivity to small deviations. Extending classical results from the sum of squared errors to the non-smooth setting of Huber loss presents nontrivial challenges, which we address in this work.

Our analysis is grounded in critical point theory, where we establish key properties such as non-degeneracy, genericity, and stability. Furthermore, we extend Morse theory to the Huber loss framework and conduct a detailed examination of saddle points. These results offer deeper insights into the theoretical landscape of sparse recovery, shedding light on its structural complexity and optimization challenges.

2 - Combinatorial, conic and linear bounds for expansions in a graph
Akshay Gupte

We consider expansion problems in graphs, which are a class of fractional quadratic combinatorial optimization with cardinality constraints, and have applications in network science, coding theory etc. Edge expansion is a NP-hard cut problem that asks for finding the Cheeger constant (isoperimetric number) of a graph which is the smallest ratio of cut size to the size of the smaller partition. We derive some lower bounds in terms of graph diameter that are tight upto a constant factor and which lead to establishing the famous Mihail-Vazirani conjecture for all polytopes of diameter 2, e.g. asymmetric TSP, and a polynomial version of this conjecture for all 0-1 polytopes. Next, we obtain lower bounds through SDP relaxations and analyse their strength in terms of the spectrum of the graph Laplacian. In contrast, SOCP and LP relaxations give trivial lower bounds. Since the problem involves a cut function, lifted LP relaxations can be derived by using polymatroid inequalities for a submodular function and some disjunctive arguments. On the algorithmic side, we compare various root-finding algorithms for a parametric approach to the problem. Similar analyses can also be carried out for vertex expansion problems.

This is an extension of previous work in https://arxiv.org/abs/2403.04657 (ISCO 2024 and journal preprint).

3 - Margin Optimal Regression Trees
Ilaria Ciocci, Marta Monaci, Laura Palagi

Interpretable machine learning models have gained increasing attention in recent years, as they provide explanatory and transparent insights into their decision-making process. Among these models, decision trees have been widely studied thanks to their intuitive structure and inherent interpretability. Along this research line, we extend to regression tasks the Margin Optimal Classification Trees (MARGOT) approach, introduced by D'Onofrio et al. (C&OR, 2024), which embeds Support Vector Machines along the binary tree structure. This leads to a quadratic mixed-integer formulation for optimal regression trees. We address the sparsity of the proposed model by introducing cardinality constraints on the features, selecting only the most relevant ones in order to enhance interpretability. To evaluate both the predictive and optimization performances of our approach, we conduct computational experiments on benchmark datasets, comparing it with state-of-the-art optimal tree methods.

Tuesday, 9:00-10:00

■ TA-01

Tuesday, 9:00-10:00 - Room: B100/1001

Plenary 2

Stream: Plenaries
Invited session
Chair: *Laura Palagi*

1 - Bayesian optimization for mixed feature spaces using tree kernels and graph kernels
Ruth Misener

Bayesian optimization is effectively a two-step iterative process that first trains a surrogate model using continuous optimization over hyperparameter space and then optimizes the acquisition function over the search space. We investigate Bayesian optimization for mixed-feature search spaces using both tree kernels and graph kernels for Gaussian processes. With respect to trees kernels, our Bayesian Additive Regression Trees Kernel (BARK) uses tree agreement to define a posterior over sum-of-tree functions. With respect to graph kernels, our acquisition function with shortest paths encoded allows us to optimize over graphs, for instance to find the best graph structure and/or node features. We formulate both acquisition functions using mixed-integer optimization and show applications to a variety of challenges in molecular design, engineering and machine learning.

This work is joint with Toby Boyne, Alexander Thebelt, Yilin Xie, Shiqiang Zhang, Jixiang Qing, Jose Folch, Robert Lee, Nathan Sudermann-Merx, David Walz, Behrang Shafei, and Calvin Tsay.

Tuesday, 10:30-12:30

■ TB-01

Tuesday, 10:30-12:30 - Room: B100/1001

Zeroth-Order Optimization Methods for Stochastic and Noisy Problems

Stream: Zeroth and first-order optimization methods
Invited session
Chair: *Francesco Rinaldi*
Chair: *Andrea Cristofari*

1 - Unifying Trust-region Algorithms with Adaptive Sampling for Nonconvex Stochastic Optimization
Sara Shashaani, Yunsoo Ha

Continuous simulation optimization is challenging due to its derivative-free and often nonconvex noisy setting. Trust-region methods have shown remarkable robustness for this class of problems. Each iteration of a trust-region method involves constructing a local model within a neighborhood of the incumbent that helps verify sufficient reduction in the function estimate at the trial step. When the local model approximates the function well, larger neighborhoods are preferred for faster progress. Conversely, unsuccessful approximations can be corrected by contracting the neighborhood. Traditional trust-region methods can be slowed down by incremental contractions that lead to numerous unnecessary iterations and significant simulation cost towards convergence to a stationary point. We propose a unified regime for adaptive sampling trust-region optimization (ASTRO) that can enjoy faster convergence in both iteration count and sampling effort by employing quadratic regularization and dynamically adjusting the trust-region size based on gradient estimates.

2 - Direct Search Methods for Stochastic Zeroth-Order Problems
Francesco Rinaldi, Andrea Cristofari

In this talk, we describe a direct search algorithm that handles Stochastic Zeroth-Order problems, i.e., problems whose objective is not computable in practice, with the only information available obtained by a stochastic zeroth-order oracle calculating an estimate of the function for any given point. Under standard assumptions on the accuracy and the variance of the random estimates used in the algorithm, we establish global convergence to stationary points. Finally, we report some numerical results to show the practical effectiveness of the method.

3 - Analysis of derivative-free algorithms on noisy problems
Alexandre Chotard, Anne Auger

In the context of derivative-free optimization, we discuss in a first part of the talk the robustness to noise of a few optimizers that include quasi-Newton algorithms (used with finite differences to estimate the gradient), the Nelder-Mead algorithm as well as stochastic or randomized algorithms like Evolution Strategies.

In a second part of the presentation we discuss proven sufficient conditions on the step-size adaptation scheme of a simple Evolution Strategy to ensure the linear convergence on the class of scaling-invariant functions perturbed with multiplicative noise.

4 - Derivative-Free Constrained Optimization in Hydraulics: Augmented Lagrangian Method with BOBYQA
Fabio Fortunato Filho, José Mario Martínez

This work proposes a Derivative-Free Safeguarded Augmented Lagrangian method for solving constrained optimization problems. The method was developed with a solid theoretical foundation and applied to the estimation of the hydraulic coefficient in river flow modeling, a problem widely studied in the literature. Additionally, an analysis of the neighborhood of each point generated in the minimization of subproblems was conducted to evaluate the behavior of each solver.

The channel modeling was performed using the Saint-Venant equations, applied to the East Fork River, with data collected over 31 days. These equations were solved using the finite difference method with artificial diffusion, chosen for its computational efficiency given the need to solve the equations repeatedly during the optimization process.

To minimize the error between observed and simulated data, derivative-free methods, such as PRIMA and BOBYQA, were tested in a box-constrained optimization problem.

The results demonstrate that the proposed approach is efficient and viable for all tested methods, reducing computational time and providing accurate estimates of the hydraulic coefficient. The study also highlights the advantages of optimization over traditional methods, pointing to potential future improvements and the integration of the method into a project compatible with HEC-RAS software, which is widely used for this type of problem.

■ TB-02

Tuesday, 10:30-12:30 - Room: B100/7011

Infinite-dimensional optimization - Part I

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Behzad Azmi*

1 - Asymptotic behavior of penalty dynamics for constrained variational inequalities
Siqi Qu, Mathias Staudigl, Juan Peypouquet

We propose a comprehensive framework for solving constrained variational inequalities via various classes of evolution equations displaying multi-scale aspects. In a Hilbertian framework, the class of dynamical systems we propose combine Tikhonov regularization and exterior penalization terms in order to yield simultaneously strong convergence of trajectories to least norm solutions in the constrained domain. Our construction thus unifies the literature on regularization methods and penalty-term based dynamical systems. We then move on showing how our base dynamics can be augmented by higher-order derivatives to introduce memory into the system.

2 - Inertial dynamics with vanishing Tikhonov regularization for multiobjective optimization
Konstantin Sonntag

This talk focuses on recent developments in continuous-time dynamical systems for multiobjective optimization. Research in this area is motivated by the strong connection between accelerated first-order optimization methods and corresponding gradient-like dynamical systems. In practice, it has been shown that the analysis of continuous-time systems is often simpler and can later be translated into discrete-time optimization schemes.

In this talk, we introduce, within a Hilbert space setting, a second-order dynamical system with asymptotically vanishing damping and vanishing Tikhonov regularization, which addresses a multiobjective optimization problem with convex and differentiable components of the objective function. We explore the asymptotic behavior of trajectory solutions to this system, particularly focusing on the fast convergence of function values, quantified in terms of a merit function. Depending on the system's parameters, we establish weak and, in some cases, strong convergence of trajectory solutions toward a weak Pareto optimal solution. To achieve this, we apply Tikhonov regularization individually to each component of the objective function.

■ TB-03

Tuesday, 10:30-12:30 - Room: B100/4011

Theoretical and algorithmic advances in large scale nonlinear optimization and applications Part 1

Stream: Large scale optimization: methods and algorithms
Invited session
Chair: *Stefania Bellavia*
Chair: *Benedetta Morini*

1 - Fully stochastic trust-region methods with Barzilai-Borwein steplengths
Benedetta Morini, Mahsa Yousefi, Stefania Bellavia

We discuss stochastic gradient methods using stochastic adaptations of Barzilai-Borwein steplengths for finite-sum minimization problems. Our approach builds on the Trust-Region-ish (TRish) framework, a first-order stochastic trust-region method based on careful step normalization. Our framework, TRishBB, is designed to enhance the performance of TRish while reducing the computational cost of its second-order variant.

In this talk, we introduce TRishBB in three variants, each leveraging Barzilai-Borwein steplengths in a stochastic setting. We will highlight the theoretical foundations of TRishBB, key insights, and properties from the convergence analysis, and discuss its practical impact on machine learning applications with numerical results.

2 - prunAdag: an adaptive pruning-aware gradient method
Giovanni Seraghiti, Margherita Porcelli, Philippe L. Toint

In this talk, we will discuss Objective Function Free Optimization (OFFO) in the context of pruning the parameter of a given model. OFFO algorithms are methods where the objective function is never computed; instead, they rely only on derivative information, thus on the gradient in the first-order case. One prominent example of OFFO methods are adaptive gradient algorithms. Pruning emerges as an alternative compression technique, mainly for neural networks, to matrix and tensor factorization or quantization. We will focus on pruning-aware methods that use specific rules to classify parameters as relevant or irrelevant at each iteration, promoting the former and penalizing the latter. The final purpose is to enhance convergence to a solution of the problem at hand, which is robust to pruning irrelevant parameters after training. Our contribution is a novel deterministic algorithm, termed prunAdag, which is both adaptive and pruning-aware. We develop a new adaptive strategy to separately update parameters that extends the distinction between relevant and irrelevant. Specifically, we introduce the concepts of optimisable and decreasable parameters. prunAdag uses the Adagrad step for updating the optimisable parameters and an Adagrad-like trust-region framework to gradually decrease the magnitude of the decreasable ones. We will also discuss the convergence properties of the algorithm, and we will illustrate preliminary results on different applications.

3 - An acceleration strategy for gradient methods in convex quadratic programming
Gerardo Toraldo, Serena Crisci, Anna De Magistris, Valentina De Simone

We propose a new acceleration strategy for gradient-based methods for solving strictly convex Quadratic optimization Problems (QP). The acceleration strategy is based on the idea of performing, at selected iterations, minimization steps along descent directions other than the negative gradient, or even in affine subspaces of low dimension. In particular, considering the contribution of the linear and quadratic part of the objective function could be useful in designing line searches in acceleration steps. We present numerical tests to understand how the acceleration steps incorporated into different gradient methods influence their behavior. We examined randomly generated QP and box constrained QP test problems, designed to assess the algorithms under various conditions, such as matrix dimensions, condition numbers, and initialization strategies. Our experiments show that the use of acceleration steps in some BB methods significantly improves computational results.

4 - Corrective Frank-Wolfe: Unifying and Extending Correction Steps
Jannis Halbey, Seta Rakotomandimby, Mathieu Besançon, Sebastian Pokutta

Fully-Corrective Frank-Wolfe (FCFW) is a powerful extension of the classical Frank-Wolfe algorithm, offering accelerated convergence by re-optimizing over the convex hull of previously selected atoms. Despite its strong theoretical guarantees, FCFW is rarely used in practice, as its correction step often requires solving a subproblem nearly as hard as the original one. Several conditional gradient-type methods with more light-weight corrective steps on the active set have since emerged. In this work, we propose Corrective Frank-Wolfe, a unifying framework for these methods that provides a general proof scheme establishing linear convergence for strongly convex objectives over polytopes. Furthermore, we introduce two corrective steps that leverage second-order information, yielding significant practical speedups at the cost of solving a linear system or a linear program, respectively. Computational experiments demonstrate substantial improvements, particularly on quadratic objectives. Notably, our approach mitigates a key practical bottleneck of the theoretically optimal Conditional Gradient Sliding (CGS) method, which heavily depends on solving a projection subproblem.

■ TB-04

Tuesday, 10:30-12:30 - Room: B100/5013

Stochastic and Deterministic Global Optimization

Stream: Global optimization
Invited session
Chair: *Dmitri Kvasov*
Chair: *Eligius M.T. Hendrix*

1 - Global Optimization Algorithm through High-Resolution Sampling

Daniel Cortild, Claire Delplancke, Nadia Oudjane, Juan Peypouquet

We present an optimization algorithm that can identify a global minimum of a potentially nonconvex smooth function with high probability, assuming the Gibbs measure of the potential satisfies a logarithmic Sobolev inequality. Our contribution is twofold: on the one hand we propose said global optimization method, which is built on an oracle sampling algorithm producing arbitrarily accurate samples from a given Gibbs measure. On the other hand, we propose a new sampling algorithm, drawing inspiration from both overdamped and underdamped Langevin dynamics, as well as from the high-resolution differential equation known for its acceleration in deterministic settings. While the focus of the work is primarily theoretical, we demonstrate the effectiveness of our algorithms on the Rastrigin function, where it outperforms recent approaches.

2 - ϵ-subdifferential methods for global DC optimization

Adil Bagirov

In this talk, we discuss methods for solving the difference of convex (DC) optimization problems subject to box constraints. These methods are based on the combination of local and global search methods where the local methods are used to find stationary points of the problem and the global methods are used to escape from these points. The escaping procedure is designed using ϵ-subdifferentials of DC components. Convergence of the proposed methods are discussed. We report results of numerical experiments using academic test problems and compare methods with the state of the-art global optimization solvers.

3 - On global optimization of some nonlinear problems involving disjunctive constraints

Sonia Cafieri, Marcel Mongeau, Sebastien Bourguignon, Gwenaël Samain

We propose Branch-and-Bound-based optimization methods for some nonlinear problems whose only combinatorial aspect comes from their disjunctive constraints. We build on the recently introduced continuous quadrant penalty formulation of disjunctive constraints, which constitutes a continuous-optimization alternative to the classical formulations (such as the bigM formulation) based on the introduction of binary variables. Such a formulation, based on the introduction of a smooth piecewice-quadratic penalty function, yields to a continuous nonconvex problem. The solution of this problem allows an efficient computation of upper bounds to be used within Branch-and-Bound. We apply the proposed approach to a problem from discrete geometry, where a rectangle has to be covered by minimal-radius disks, and to a problem from compressed sensing, where a sparsity-enhancing cardinality-constrained inverse problem in signal processing is addressed.

4 - Benchmarking tools for global optimization

Dmitri Kvasov, Yaroslav Sergeyev

We consider various tools for testing stochastic and deterministic constrained global optimization algorithms (such as, for example, operational zones and aggregated operational zones https://www.nature.com/articles/s41598-017-18940-4). We also present a recently proposed generator of constrained test problems https://ieeexplore.ieee.org/document/9664812 which is based on the popular GKLS generator for benchmarking box-constrained global optimization methods. We illustrate the considered techniques by testing some known optimization algorithms.

This work was partially supported by the Italian INdAM GNCS Project 2025, number CUP_ E53C24001950001.

■ TB-05

Tuesday, 10:30-12:30 - Room: B100/4013

Randomized Optimization algorithms I

Stream: Optimization for machine learning
Invited session
Chair: *Laurent Condat*

1 - Scalable Second-Order Optimization Algorithms for Minimizing Low-Rank Functions

Edward Tansley, Coralia Cartis, Zhen Shao

Second-order optimization algorithms, such as Adaptive Regularization framework using Cubics (ARC), often have faster convergence rates than first-order algorithms that only rely on gradient information. However, for high-dimensional problems, the computational cost of these methods can be a barrier to their use in practice. We present R-ARC-D, a random subspace variant of ARC that chooses the size of the subspace adaptively, based on the rank of the projected second derivative matrix. At each iteration, our variant only requires access to (low-dimensional) projections of first- and second-order problem derivatives and calculates a reduced step inexpensively. The ensuing method maintains the optimal global rate of convergence to an approximate first-order minimizer of (full-dimensional) cubic regularization, while showing improved scalability both theoretically and numerically, particularly when applied to low-rank functions (i.e., functions that vary in only a low-dimensional linear subspace of dimension r). When applied to these low-rank functions, our algorithm naturally adapts the subspace size to the true rank of the function, without knowing it a priori. We observe this behaviour through numerical experiments, which confirm that R-ARC-D effectively adapts to the function's true rank. Additionally, we demonstrate its strong performance and improved scalability compared to the full-dimensional ARC on high-dimensional problems.

2 - Distributed Optimization with Communication Compression

Yuan Gao, Sebastian Stich

Modern machine learning tasks often involve massive datasets and models, necessitating distributed optimization algorithms with reduced communication overhead. Communication compression, where clients transmit compressed updates to a central server, has emerged as a key technique to mitigate communication bottlenecks. In this work, we consider the theory of distributed optimization with contractive communication compression. The naive implementation of distributed optimization algorithm with contractive compression often leads to unstable convergence or even divergence. We discuss how recent variants of the popular and practical Error Feedback mechanism help to mitigate the aforementioned issues, and obtain provable convergence under standard assumptions. We further extend these theories and consider the decentralised setting or the incorporation of Nesterov's acceleration.

3 - Stochastic Gradient Descent without Variance Assumption: A Tight Lyapunov Analysis

Lucas Ketels, Daniel Cortild, Guillaume Garrigos, Juan Peypouquet

We derive new upper bounds on the performance of the stochastic gradient descent (SGD) algorithm in the convex and smooth setting under minimal variance assumptions and with a focus on a wide range of step- sizes. The analysis of SGD usually assumes the uniform boundedness of the variance of the gradient noise. Removing this assumption usually degrades the results in two ways: it restricts the range of step-sizes and increases the complexity by a polynomial factor. Our analysis, however, solely relies on the weak assumption that the variance of the gradient noise is bounded at a solution, and remains valid for a wide range of step-sizes while improving state-of-the-art results. The proof is based on a Lyapunov analysis obtained by solving a suitable semidefinite program obtained through the performance estimation problem framework, as done in Taylor and Bach (2021). We also present numerical and analytical evidence of the tightness of our analysis

4 - Communication-Efficient Algorithms for Federated Learning and Weakly Coupled Games

Sebastian Stich, Ali Zindari, Parham Yazdkhasti, Anton Rodomanov, Tatjana Chavdarova

Communication-efficient algorithms are crucial for distributed and federated learning, where reducing communication overhead can significantly enhance performance. Over the past decade, many such algorithms have been developed and refined. However, these algorithms often only reduce communication in certain regimes, and in the worst-case scenarios, they may not reduce the communication required to reach a predefined level of accuracy.

In this talk, we will explore this question and review the speedup properties of several algorithms. We will focus on those that can provably reduce the total communication cost of training. One such method is emphDecoupled SGDA, a novel adaptation we developed for distributed games. Our findings show that, in the context of emphWeakly Coupled Games, emphDecoupled SGDA can achieve communication speed-up, making it a promising approach for resource-constrained environments.

■ TB-06

Tuesday, 10:30-12:30 - Room: B100/7013

Advances in nonsmooth optimization

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Alireza Kabgani*
Chair: *Vladimir Shikhman*

1 - Monotone and nonmonotone linearized block coordinate descent methods for nonsmooth composite optimization problems

Yassine Nabou, Lahcen El Bourkhissi, Sebastian Stich, Tuomo Valkonen

In this paper, we introduce both monotone and nonmonotone variants of LiBCoD, a Linearized Block Coordinate Descent method for solving composite optimization problems. At each iteration, a random block is selected, and the smooth components of the objective are linearized along the chosen block in a Gauss-Newton approach. For the monotone variant, we establish a global sublinear convergence rate to a stationary point under the assumption of bounded iterates. For the nonmonotone variant, we derive a global sublinear convergence rate without requiring global Lipschitz continuity or bounded iterates. Preliminary numerical experiments on logistic regression highlight the promising performance of the proposed approach.

2 - Global convergence of a second-order augmented Lagrangian method under an error bound condition

Renan William Prado, Gabriel Haeser, Roberto Andreani, Maria Laura Schuverdt, Leonardo Delarmelina Secchin

In this talk, we deal with convergence to points satisfying the weak second-order necessary optimality conditions of a second-order safeguarded augmented Lagrangian method from the literature. To this end, we propose a new second-order sequential optimality condition that is, in a certain way, based on the iterates generated by the algorithm itself. This also allows us to establish the best possible global convergence result for the method studied, from which a companion constraint qualification is derived. The companion constraint qualification is independent of the Mangasarian-Fromovitz and constant-rank constraint qualifications and remains verifiable without them, as it can be certified by different known constraint qualifications. Furthermore, unlike similar results from previous works, the new constraint qualification cannot be weakened by another one with second-order global convergence guarantees for the method and assures second-order stationarity without needing constant rank hypotheses. To guarantee the latter result, we established the convergence of the method under a property slightly stronger than the error bound constraint qualification, which, until now, has not been known to be associated with nonlinear optimization methods.

3 - Non-euclidean proximal methods for convex-concave saddle-point problems

Eyal Cohen, Shoham Sabach, Marc Teboulle

Abstract. Motivated by the flexibility of the Proximal Alternating Predictor Corrector (PAPC) algorithm

which can tackle a broad class of structured constrained convex optimization problems via their convex- concave saddle-point reformulation, in this paper, we extend the scope of the PAPC algorithm to include

non-Euclidean proximal steps. This allows for adapting to the geometry of the problem at hand to produce simpler computational steps. We prove a sublinear convergence rate of the produced ergodic sequence, and under additional natural assumptions on the non-Euclidean distances, we also prove that the algorithm globally converges to a saddle-point. We demonstrate the performance and simplicity of the proposed algorithm through its application to the multinomial logistic regression problem.

4 - Stationarity in nonsmooth optimization between geometrical motivation and topological relevance
Vladimir Shikhman

The goal of this paper is to compare alternative stationarity notions in structured nonsmooth optimization (SNO). Here, nonsmoothness is caused by complementarity, vanishing, orthogonality type, switching, or disjunctive constraints. On one side, we consider geometrically motivated notions of stationarity in terms of Fréchet, Mordukhovich, and Clarke normal cones to the feasible set, respectively. On the other side, we advocate the notion of topologically relevant T-stationarity, which adequately captures the global structure of SNO. Our main findings say that (a) stationary points via Frechet normal cone include all local minimizers; (b) stationary points via Mordukhovich normal cone, which are not Frechet stationary, correspond to the singular saddle points of first order; (c) T-stationary points, which are not Mordukhovich stationary, correspond to the regular saddle points of first order; (d) stationary points via Clarke normal cone, which are not T-stationary, are irrelevant for optimization purposes. Overall, a hierarchy of stationarity notions for SNO is established.

■ TB-07

Tuesday, 10:30-12:30 - Room: B100/5015

Nonsmooth Bilevel Optimization

Stream: Bilevel and multilevel optimization
Invited session
Chair: *Shangzhi Zeng*

1 - A new problem qualification for Lipschitzian optimization problems
Andreas Fischer, Isabella Käming, Alain Zemkoho

In contrast to a constraint qualification (CQ), a problem qualification may not only rely on the constraints of an optimization problem but also on the objective function and, like a CQ, guarantees that a local minimizer is a Karush-Kuhn-Tucker (KKT) point. With the Subset Mangasarian-Fromovitz Condition (subMFC), a new problem qualification is introduced. A comparison of subMFC with several existing qualifications will be presented and reveals for example that subMFC is strictly weaker than quasinormality and can even hold if calmness in the sense of Clarke is violated. The power of the new problem qualification is also demonstrated for optimistic bilevel optimization by means of the lower-level value function reformulation.

2 - Solving Bilevel Optimization Problems using Finite Elements and Reduced Order Methods
Floriane Mefo Kue

We consider bilevel optimization problems which can be interpreted as inverse optimal control problems. The lower-level problem is an optimal control problem with a parametrized objective function. The upper-level problem is used to identify the parameters of the lower-level problem. We then consider the Karush-Kuhn-Tucker reformulation of the problem and the main focus is on the derivation of first-order necessary optimality conditions and solutions algorithms via finite elements methods and reduced order methods.

3 - Nonsmooth Implicit Differentiation: Deterministic and Stochastic Convergence Rates
Saverio Salzo

I will address the problem of efficiently computing a generalized derivative of the fixed-point of a parametric nondifferentiable contraction map. This problem has wide applications in machine learning, including hyperparameter optimization, meta-learning and data poisoning attacks. Two popular approaches are analyzed: iterative differentiation (ITD) and approximate implicit differentiation (AID). A key challenge behind the nonsmooth setting is that the chain rule does not hold anymore. Building upon the recent work by Bolte et al. (2022), who proved linear convergence of nondifferentiable ITD, I will show an improved linear rate for ITD and a slightly better rate for AID, both in the deterministic case. I will also introduce NSID, a new stochastic method to compute the implicit derivative when the fixed point is defined as the composition of an outer map and an inner map which is accessible only through a stochastic unbiased estimator. Rates for such stochastic method rates will be presented.

4 - Alternating Gradient-Type Algorithm for Bilevel Optimization with Inexact Lower-Level Solutions via Moreau Envelope-based Reformulation
Shangzhi Zeng

In this work, we study a class of bilevel optimization problems where the lower-level problem is a convex composite optimization model, which arises in various applications, including bilevel hyperparameter selection for regularized regression models. To solve these problems, we propose an Alternating Gradient-type algorithm with Inexact Lower-level Solutions (AGILS) based on a Moreau envelope-based reformulation of the bilevel optimization problem. The proposed algorithm does not require exact solutions of the lower-level problem at each iteration, improving computational efficiency. We prove the convergence of AGILS to stationary points and, under the Kurdyka-Lojasiewicz (KL) property, establish its sequential convergence. Numerical experiments, including a toy example and a bilevel hyperparameter selection problem for the sparse group Lasso model, demonstrate the effectiveness of the proposed AGILS.

■ TB-08

Tuesday, 10:30-12:30 - Room: B100/7007

Systematic and computer-aided analyses IV: Online & distributed gradient methods

Stream: Systematic and computer-aided analyses of optimization algorithms
Invited session
Chair: *Manu Upadhyaya*
Chair: *Baptiste Goujaud*

1 - Constructive approach to worst-case analysis and algorithm design in online convex optimization
Julien Weibel

In this talk, we will present methods based on semi-definite programming to perform worst-case analysis of online convex optimization algorithms. Those methods allow to construct simultaneously a worst-case scenario and a proof for a matching upper bound on the regret of a given online convex optimization algorithm. We will also present how to adapt those methods for algorithm design by formulating a semi-definite program that optimises simultaneously the choice of the algorithm with its regret upper bound.

2 - A Linear Parameter-Varying framework for the analysis of time-varying optimization algorithms.
Fabian Jakob, Andrea Iannelli

We present a framework to analyze general first-order optimization algorithms for time-varying convex optimization using robust control tools. We assume that the temporal variability of the objective is caused by time-varying parameters, and only past knowledge is available. We consider the case of strongly convex objective functions with Lipschitz continuous gradients. We model the first-order algorithms as discrete-time linear parameter varying (LPV) systems in feedback with a time-varying (sub)gradient. We leverage the approach of analyzing algorithms as uncertain control interconnections with integral quadratic constraints and generalize that framework to the time-varying case. We propose novel IQCs that capture the behavior of time-varying nonlinearities and leverage techniques from LPV control to establish upper bounds on the tracking error. These bounds can be computed offline by solving a semi-definite program and can be interpreted as input-to-state stability with respect to a disturbance signal related to the problem's temporal variability. An important benefit of our analysis framework is the ability to capture how convergence rates and robustness of general first-order algorithms depend on specific properties of the problem and the parameters rate of variation, yielding for example new insights on the role of acceleration in sequential decision-making problems.

3 - Computer-Aided Analysis of Decentralized Online Optimization Algorithms
Erwan Meunier, Julien Hendrickx

We present a novel application of the Performance Estimation Problem (PEP) framework to Decentralized Online Optimization (DOO), enabling the automatic computation of exact worst-case performance bounds for popular algorithms in this setting, would they be Projection-based, Projection-Free or relying on the Mirror-Descent scheme. Our findings reveal that existing theoretical guarantees can significantly overestimate true worst-case behavior, often by several orders of magnitude, potentially leading to suboptimal algorithmic choices in practice. Notably, we also uncover that certain algorithms do not benefit from inter-agent communication in the worst-case, challenging common intuitions about decentralized cooperation. By leveraging PEP-based analysis, we propose a principled way of refining step-size strategies. Thanks to this approach, we show that the worst-case regret of classical methods can be reduced by up to 20%. This work highlights the value of near tight performance estimation for the principled design and deployment of decentralized optimization algorithms.

■ TB-09

Tuesday, 10:30-12:30 - Room: B100/8013

Variational Analysis I

Stream: Variational analysis: theory and algorithms
Invited session
Chair: *Fernando García Castaño*

1 - The Douglas-Rachford algorithm with variable stepsizes as a relocated fixed-point iteration
Felipe Atenas, Heinz Bauschke, Minh N. Dao, Matthew Tam

The Douglas-Rachford algorithm is a popular proximal splitting method that breaks down complex problems with a sum structure into simpler pieces easier to handle. Traditional convergence guarantees assume constant stepsizes, while the theory with variable stepsizes is scarce. The fundamental challenge in varying this parameter stems from the stepsize-dependent nature of the fixed point set of the Douglas-Rachford operator, preventing the use of classical arguments to deduce convergence. To address this limitation, we propose a novel variant of the Douglas-Rachford algorithm that allows updating the stepsize between iterations, by composing the original Douglas-Rachford iteration with a "fixed point relocator" operator. For finding a zero of the sum of two maximally monotone operators in a Hilbert space, and under mild assumptions on the asymptotic behavior of the sequence of stepsizes, we show that the resulting relocated Douglas-Rachford method converges weakly to a fixed point of the limiting iteration operator. Furthermore, we establish that the corresponding shadow sequence converges weakly to a solution of the monotone inclusion problem.

2 - A Speed Restart Scheme for an Inertial System with Hessian-Driven Damping and Three Constant Coefficients
Huiyuan Guo, Juan Jose Maulen, Juan Peypouquet

The use of inertial methods is a popular first-order approach to smooth convex optimization problems. Numerous algorithms and methods have been proposed, gradient method, heavy ball with friction, Newton's method and so on. In recent years, researchers start to analyze the related differential equations or inclusions for investigating the dynamic behavior of these iterative algorithms. Despite the fast convergence rate guarantee, the trajectories as well as the sequences generated by inertial first-order methods exhibit chaotic behavior. Restart techniques represent an alternative way to accelerate gradient methods by reducing the oscillations. We analyze a speed restart scheme for an inertial system with Hessian-driven damping and three constant coefficients. We establish a linear convergence rate for the function values along the restarted trajectories without assuming the strong convexity of the objective function. We also report numerical experiments which show that dynamical system with speed restarting scheme together improve the performance of both continuous dynamics and inertial algorithms as a heuristic.

3 - Forward-backward algorithms devised by graphs
Francisco Javier Aragón Artacho, Rubén Campoy, César López Pastor

In this talk, we will present a methodology for devising forward-backward methods for finding zeros in the sum of a finite number of maximally monotone operators. We extend the techniques from [SIAM J. Optim., 34 (2024), pp. 1569-1594] to cover the case involving a finite number of cocoercive operators, which should be directly evaluated by the algorithm instead of computing their resolvent. The algorithms are induced by three graphs that determine how the algorithm variables interact with each other and how they are combined to compute each resolvent. The hypotheses on these graphs ensure that the algorithms obtained have minimal lifting and are frugal, meaning that the ambient space of the underlying fixed point operator has minimal dimension and that each resolvent and each cocoercive operator is evaluated only once per iteration. This framework not only allows to recover some known methods, but also to generate new ones, as the forward-backward algorithm induced by a complete graph. If time permits, we will present some numerical experiments showing how the choice of graphs influences the performance of the algorithms.

■ TB-10

Tuesday, 10:30-12:30 - Room: B100/8011

First order methods: new perspectives for machine learning

Stream: Large scale optimization: methods and algorithms
Invited session
Chair: *Cesare Molinari*
Chair: *Silvia Villa*
Chair: *Lorenzo Rosasco*

1 - Convergence Analysis of Nonlinear Parabolic PDE Models with Neural Network Terms Trained with Gradient Descent

Konstantin Riedl, Justin Sirignano, Konstantinos Spiliopoulos

Many engineering and scientific fields have recently become interested in modeling terms in partial differential equations (PDEs) with neural networks (NNs). The resulting PDE model, a function of the NN parameters, can be calibrated to available data by optimizing over the PDE using gradient descent, where the gradient is evaluated by solving an adjoint PDE. In this talk, we discuss the convergence of this adjoint optimization method for training NN-PDE models in the limit where both the number of hidden units and the number of training steps tend to infinity. Specifically, for a general class of nonlinear parabolic PDEs, we prove convergence of the NN-PDE solution to the target data (i.e., a global minimizer). The global convergence proof requires addressing several technical challenges, since the PDE system is both nonlinear and non-local. Although the adjoint PDE is linear, the NN training dynamics involve a non-local kernel operator in the infinite-width hidden layer limit, where the kernel lacks a spectral gap for its eigenvalues. This poses a unique mathematical challenge that is not encountered in finite-dimensional NN convergence analysis. We establish convergence by proving that an appropriate quadratic functional of the adjoint is globally Lipschitz and then applying a cycle of stopping times analysis to prove that the adjoint solution weakly converges to zero. Leveraging the definition of the adjoint PDE, this yields the global convergence of the original NN-PDE.

2 - Randomized trust-region method for non-convex mimimization

Radu-Alexandru Dragomir

We study second-order trust-region algorithms, with subproblem solvers based on the conjugate gradient method, for non-convex optimization. In this context, a desirable property is that the algorithm avoids saddle points and converges to a local minimizer. We introduce a trust-region method that (i) satisfies this property, (ii) has efficient complexity for large-scale problems, and (iii) is as close as possible to the version used in practice. Our scheme is based on the Steihaug-Toint truncated conjugate gradient (tCG) method, with two key modifications: a random perturbation of the initial point and a supplementary gradient step when the iterates reach the trust-region boundary. We provide theoretical guarantees for functions with isolated critical points. Our key assumption, the mu-Morse property, is a bound on the magnitude of the Hessian eigenvalues at critical points.

3 - Perspectives on the analysis and design of optimization algorithms: Lyapunov analyses and counter-examples

Adrien Taylor

Perspectives on the analysis and design of optimization algorithms: Lyapunov analyses and counter-examples

Lyapunov analysis is an essential tool in the study of simple first-order optimization methods, and for obtaining performance guarantees. This talk will cover a few recent constructive approaches to the discovery of such Lyapunov functions and of their structural properties in the context of first-order optimization. In addition, we cover a few constructive approaches to counter-examples for when no such Lyapunov exist.

The talk will be example-based, as many ingredients of those methodologies are already present when analyzing simple optimization algorithms such as gradient descent, the heavy-ball method, or the primal-dual hybrid gradient algorithm (Chambolle-Pock). We further illustrate how simple Lyapunov structure can be studied with existing software for computer-aided analyses (and in particular with PEPit) relying on different formulations of the analysis problem.

This talk is based on joint works with great colleagues, that include Francis Bach, Laurent Lessard, Bryan Van Scoy, Manu Upadhyaya, Sebastian Banert, Pontus Giselsson, Baptiste Goujaud, Daniel Berg Thomsen, Si Yi Meng, and Aymeric Dieuleveut.

4 - Accelerated Gradient Methods via Inertial Systems with Hessian-driven Damping

Juan Peypouquet

We analyze the convergence rate of a family of inertial algorithms, which can be obtained by discretization of an inertial system with Hessian-driven damping. We recover a convergence rate, up to a factor of 2 speedup upon Nesterov's scheme, for smooth strongly convex functions. As a byproduct of our analyses, we also derive linear convergence rates for convex functions satisfying a quadratic growth condition or Polyak-Łojasiewicz inequality. As a significant feature of our results, the dependence of the convergence rate on parameters of the inertial system/algorithm is revealed explicitly, which helps one get a better understanding of the acceleration mechanism underlying an inertial algorithm.

■ TB-11

Tuesday, 10:30-12:30 - Room: B100/5017

Advances in Manifold Optimization

Stream: Riemannian Manifold and Conic Optimization
Invited session
Chair: *Yukuan Hu*

1 - Infeasible Optimization on Manifolds: the Landing Approach

P.-a. Absil

Classic first-order optimization methods on a submanifold of a Euclidean space are based on two ingredients: (i) choosing a search direction at the current iterate and (ii) applying a retraction to produce points along a manifold-valued curve tangent to the search direction. It is now well established that ingredient (ii) can be computationally considerably more costly than ingredient (i) even for well-known manifolds such as the Stiefel manifold, notably in a stochastic gradient context where (i) can be particularly cheap. This has prompted the development of infeasible optimization methods that do not enforce the manifold constraint at the iterates but still exploit the manifold nature of the feasible set. The landing approach, which is the topic of this talk, belongs to this recent trend. It performs an update along a weighted sum of two terms. One term, tangent to the active "layered manifold", decreases the objective function while preserving the constraint function at the first order. The other term decreases infeasibility. Under mild assumptions and with sufficiently small step size, the method provably converges to critical points. Moreover, in the stochastic case where the update vector is affected by additive zero-mean bounded-variance noise, the landing algorithm with suitably diminishing step size is proved to converge in expectation. This talk is based on joint work with Pierre Ablin, Bin Gao and Simon Vary (arXiv:2303.16510, arXiv:2405.01702).

2 - The Distributionally Robust Optimization Model of Sparse Principal Component Analysis

Xin Liu

We consider sparse principal component analysis (PCA) under a stochastic setting where the underlying probability distribution of the random parameter is uncertain. This problem is formulated as a distributionally robust optimization (DRO) model based on a constructive approach to capturing uncertainty in the covariance matrix, which constitutes a nonsmooth constrained min-max optimization problem. We further prove that the inner maximization problem admits a closed-form solution, reformulating the original DRO model into an equivalent minimization problem on the Stiefel manifold. This transformation leads to a Riemannian optimization problem with intricate nonsmooth terms, a challenging formulation beyond the reach of existing algorithms. To address this issue, we devise an efficient smoothing manifold proximal gradient algorithm. We prove the Riemannian gradient consistency and global convergence of our algorithm to a stationary point of the nonsmooth minimization problem. Moreover, we establish the iteration complexity of our algorithm. Finally, numerical experiments are conducted to validate the effectiveness and scalability of our algorithm, as well as to highlight the necessity and rationality of adopting the DRO model for sparse PCA.

3 - A Unified Constrained Saddle Dynamics for Index-1 Saddle Point Search on Riemannian Submanifolds

Yukuan Hu

Index-1 saddle point search on Riemannian submanifolds is a fundamental task in various applications. However, most existing works concentrate on unconstrained settings, with only limited efforts devoted to the cases with special Riemannian submanifolds induced by global defining functions. In this talk, we introduce a unified constrained saddle dynamics applicable to general Riemannian submanifolds, where the position and direction variables are simultaneously evolved on the tangent bundle. In particular, the direction dynamics leverages the second fundamental form to maintain feasibility. The linear stability of the dynamics at index-1 saddle points is established. We further discretize the proposed dynamics using the Riemannian manifold tools. The local convergence properties of the resulting iterative methods are analyzed. Finally, the numerical experiments on electronic excited states calculations demonstrate the effectiveness of the proposed methods.

4 - Exploring Energy Landscapes: Stochastic Methods for Index-1 Saddle Points on Riemannian Submanifolds

Panos Parpas

We introduce a new stochastic algorithm to locate index 1 saddle points of a function defined on Riemannian submanifolds, which is particularly useful in high-dimensional settings. Our approach relies on two key ingredients: (i) the concentration properties of the first eigenmodes of the Witten Laplacian on 1 forms, specifically in the vicinity of index 1 saddle points, and (ii) a probabilistic representation of a partial differential equation involving this differential operator. Numerical examples on simple molecular systems underscore the potential of our algorithm for exploring complex potential energy landscapes in quantum chemistry applications.

■ TB-12

Tuesday, 10:30-12:30 - Room: B100/8009

Optimisation under uncertainty in the power sector

Stream: Applications: AI, uncertainty management and sustainability
Invited session
Chair: *Hongyu Zhang*

1 - Hybrid residential systems with battery storage: multi-objective optimization of self-sufficiency and economic viability

Paulo Rotella Junior, Arthur Leandro Guerra Pires, Rafael Miranda, Luiz Celio Souza Rocha, Karel Janda

This study aims to optimize the economic viability and energy autonomy of hybrid systems (photovoltaic and wind) with battery storage within the context of regulatory changes (Law 14,300/2022) and the available tariff models in Brazil. To achieve this, a multi-objective optimization model was proposed using two objective functions: Net Present Value and Self-Sufficiency. The objective functions were modeled using Response Surface Methodology, and Normal Boundary Intersection method was used for multi-objective optimization. Key results indicated that the changes in the compensation rule did not render investments unfeasible, mainly when the generation is based on photovoltaic energy. However, the non-compensable portion of the energy distribution costs for prosumers, related to the B Wire component, which could be avoided by not injecting energy into the grid, is insufficient to make battery use economically viable. The results emphasize the need for region-specific policies that can harness the distributed generation potential of each region, as seen in areas with wind generation potential, which can increase the Self-Sufficiency level and address voltage flow inversion issues. The depth of discharge was also evaluated, and from an economic perspective, optimal values were identified around 80%. The best residential tariff model was evaluated, with most cities favoring the conventional tariff from an economic point of view, indicating the need to review this tariff model.

2 - Renewable Energy Communities with Peer-to-Peer Exchange: a chance-constraint approach
Santo Saraceno, Elisabetta Allevi, Giorgia Oggioni, Rossana Riccardi, Abdel Lisser

This work presents a chance-constraint model for the management of Energy Communities, focusing on prosumers and peer-to-peer electricity exchanges.

The model aims to minimize the total operation costs of the community, while ensuring energy balance and satisfying technical constraints related to local production and the energy exchanges both inside the community and with the main grid.

Uncertainty in solar photovoltaic generation and electricity demand is addressed using individual and joint chance constraints that are modeled using normal distributions and approximated through piecewise-linear techniques when necessary.

The model is tested on a prototype example of community and is implemented in Python using Pyomo.

3 - Data-driven dynamic scenario generation using Benders decomposition with adaptive oracles
Hongyu Zhang, Ken McKinnon, Andrew Reeves

Scenario generation for a multi-stage stochastic program often aims to find a good finite discrete approximation of random variables, which is represented by a scenario tree. Many existing scenario generation algorithms perform scenario generation and reduction once prior to solving the problem and the tree is fixed. However, the information gained when solving a multi-stage stochastic program may detect other parameters that should have been treated uncertain and included in an expanded scenario tree. This paper proposes a data-driven dynamic scenario generation algorithm using Benders decomposition with adaptive oracles to dynamically adjust the scenario tree in the course of solving the underlying problem. The proposed algorithm detects the parameters that provide the most added value by being treated uncertain and returns a scenario tree and an -optimal solution. The proposed scenario generation algorithm concerns the class of multi-stage stochastic programs with block separable recourse. We apply the proposed scenario generation algorithm to a stochastic power system investment planning problem, and the results show that (1) the proposed rules can effectively identify which parameters to explore in an expanded tree, and (2) the proposed algorithm can estimate the impact of expanding a scenario tree to include additional parameters without much computational effort.

■ TB-13

Tuesday, 10:30-12:30 - Room: B100/6009

Numerical Methods and Applications I

Stream: Numerical Methods and Applications
Invited session
Chair: *Stefan Schwarze*

1 - The improvement function in branch-and-bound methods
Stefan Schwarze, Oliver Stein, Peter Kirst, Marc Rodestock

We present a new spatial branch-and-bound approach for treating optimization problems with nonconvex inequality constraints. It is able to approximate the set of all global minimal points in case of solvability, and else to detect unsolvability. The new technique covers the nonconvex constraints by means of an improvement function which, although nonsmooth, can be treated by standard bounding operations. The method is shown to be successful under a weak regularity condition, and we also give a transparent interpretation of the output in case that this condition is violated. Numerical tests illustrate the performance of the algorithm.

2 - LP relaxations for the elementary shortest path problem
Regina Schmidt, Mirjam Duer

The Elementary Shortest Path Problem (ESPP) is the problem of finding an elementary minimum-cost path between two nodes in a directed graph in such a way that each node on the path is visited exactly once. If negative arc costs are allowed, then the problem is NP-hard. We study an exact integer programming formulation for the ESPP and we discuss its LP relaxations. We present a solution approach based on these relaxations.

3 - Convergence Analysis of Modified Cubic-Regularized Newton Method
Vladislav Ryspayev

This paper presents a convergence analysis of a modified cubic-regularized Newton method for unconstrained optimization problems, considering the minimization of a twice-differentiable function with Lipschitz continuous Hessian and bounded eigenvalues. Our modification replaces the exact Hessian with a block diagonal matrix structure that combines a scaled identity matrix in one block and the original Hessian in another, effectively integrating first-order and second-order information. We derive explicit formulas for the step direction and size, establish a sufficient decrease property, and provide detailed convergence rates in different gradient regimes: for small gradient norms, we prove O(N(-1/2)) convergence, while for large gradient norms, we achieve the faster O(N(-2/3)) rate, with an overall global convergence guarantee of O(N(-1/2)) in the general mixed regime. The analysis reveals how the algorithm's behavior adapts to different problem characteristics, making the method particularly attractive for large-scale applications where full Hessian computations are expensive. Our experiments on matrix factorization problems demonstrate the method's practical efficiency in settings where partial second-order information can be advantageously combined with first-order approaches.

4 - On the Spectral Projected Gradient Method for the Molecular Distance Geometry Problem

Mariana da Rosa

The reconstruction of 3D protein structures from Nuclear Magnetic Resonance (NMR) data presents significant challenges due to uncertainties in the distance constraints between atoms within a protein molecule. We use a Spectral Projected Gradient algorithm to address these challenges, overcoming the limitations of previous methods that unrealistically assume all pairwise interatomic distances are known, are restricted to small problem instances, or only consider narrow distance ranges. Our approach incorporates fixed covalent bond lengths and angles while effectively handling larger interval uncertainties, offering a more accurate representation of the experimental conditions in NMR data. Computational experiments highlight the potential of this method to provide reliable solutions for molecular structure determination.

Tuesday, 14:00-16:00

■ TC-01

Tuesday, 14:00-16:00 - Room: B100/1001

First-Order Methods for Structured Optimization and Sampling

Stream: Zeroth and first-order optimization methods
Invited session
Chair: *Cesare Molinari*
Chair: *Silvia Villa*

1 - Constrained sampling with Primal-Dual Langevin Monte Carlo
Luiz Chamon

We consider the problem of sampling from a probability distribution known up to a normalization constant while satisfying a set of requirements. In contrast to traditional problem that considers only support constraints, however, our requirements are specified by the expected values of general nonlinear functions. Hence, methods based on mirror maps, barriers, and penalties, are not suited for this task. In this talk, we introduce a discrete-time primal-dual Langevin Monte Carlo algorithm (PD-LMC) that simultaneously constrains the target distribution and samples from it. By extending classical optimization arguments for saddle-point algorithms to the geometry of Wasserstein space, we derive convergence guarantees for target distributions satisfying (strong) convexity and log-Sobolev inequalities. We showcase the relevance and effectiveness of PD-LMC in fair and counterfactual Bayesian inference.

2 - A Fast Extra-Gradient Method with Flexible Anchoring
Enis Chenchene, Radu Ioan Bot

We introduce a novel Extra-Gradient method with anchoring governed by general parameters. Our method is derived from an explicit discretization of a dynamical system with Tikhonov regularization, which provides a theoretical foundation for analyzing its convergence properties. We establish strong convergence to specific points within the solution set, as well as convergence rates expressed in terms of the regularization parameters. Notably, our approach recovers the fast residual decay rate $O(k-1)$ for standard parameter choices. Numerical experiments highlight the competitiveness of the method and demonstrate how its flexible design enhances practical performance.

3 - Risk-averse guarantees for stochastic min-max problems
Yassine Laguel, Mert Gürbüzbalaban, Necdet Serhat Aybat, Yasa Syed

We investigate the stochastic accelerated primal-dual algorithm for strongly-convex-strongly-concave (SCSC) saddle point problems, common in distributionally robust learning, game theory, and fairness in machine learning. Our algorithm offers optimal complexity in several settings and we provide high probability guarantees for convergence to a neighbourhood of the saddle point. For quadratic problems under Gaussian perturbations, we derive analytical formulas for the limit covariance matrix together with lower bounds that show that our general analysis for SCSC problems is tight. Our risk-averse convergence analysis characterises the trade-offs between bias and risk in approximate solutions. We present numerical experiments on zero-sum games and robust learning problems. We finally extend our results on nonconvex-PL problems.

■ TC-02

Tuesday, 14:00-16:00 - Room: B100/7011

Infinite-dimensional optimization - Part II

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Behzad Azmi*

1 - A Fast Iterative Method for Variational Inequality and Classification Problems
Lateef Jolaoso

The study of variational inequality problems (VIPs) continues to drive progress in the design of iterative methods, particularly under more relaxed conditions on the cost operator. While several methods have been developed for solving VIPs under monotone and pseudomonotone conditions, recent research has shifted focus toward tackling the more general case of quasimonotone operators. In this paper, we propose a novel double inertial acceleration algorithm that integrates self-adaptive and relaxation strategies to solve quasimonotone VIPs. We establish both weak convergence and linear convergence rate results under mild assumptions. To validate the efficiency of the proposed method, we conduct extensive numerical experiments, which demonstrate its superior performance compared to several existing methods in the literature. Additionally, we explore an application in machine learning by reformulating the support vector machine (SVM) classification problem as a VIP, thereby highlighting the practical relevance and applicability of our approach.

2 - An Interior-Proximal Point Method for Nonsmooth, Nonconvex PDE-Constrained Problems with State Constraints
Behzad Azmi, Alberto De Marchi

We consider a class of nonconvex, nonsmooth problems governed by partial differential equations (PDEs) and subject to state constraints. To address these problems, we propose a flexible algorithm that combines interior point (IP) methods with proximal gradient techniques. While traditional IP methods face difficulties with nonsmooth objective functions and proximal algorithms are typically unable to handle state constraints, their combination effectively overcomes these individual limitations. We discuss a theoretical analysis of the algorithm, including convergence and complexity results. Numerical experiments are also discussed to demonstrate the algorithm's performance.

■ TC-03

Tuesday, 14:00-16:00 - Room: B100/4011

Theoretical and algorithmic advances in large scale nonlinear optimization and applications Part 2

Stream: Large scale optimization: methods and algorithms
Invited session
Chair: *Stefania Bellavia*
Chair: *Benedetta Morini*

1 - Advanced Techniques for Portfolio Optimization Under Uncertainty
Valentina De Simone

Portfolio optimization under uncertainty remains a critical challenge in financial decision-making, requiring advanced numerical optimization techniques to handle the complexities of real-world markets. This talk examines two contrasting approaches, the Worst-Case Approach and Cumulative Prospect Theory (CPT), within a mean-variance framework. The worst-case approach prioritizes robustness, minimizing potential losses under the most adverse market conditions. In contrast, CPT integrates behavioural factors such as probability weighting and loss aversion, providing a more psychologically realistic framework for decision-making. Both approaches pose significant computational challenges due to nonsmooth or nonconvex optimization problems. We discuss how leveraging problem structure and specialized numerical techniques enable efficient portfolio construction, even under severe uncertainty. By combining rigorous mathematical methods with practical insights, we explore how portfolio managers can navigate risk,uncertainty, and behavioural biases to develop more resilient investment strategies.

2 - Inexact derivative-free methods for constrained bilevel optimization with applications to machine learning
Marco Viola, Matteo Pernini, Gabriele Sanguin, Francesco Rinaldi

Bilevel Optimization is a powerful framework for addressing complex machine learning challenges such as hyperparameter tuning, meta-learning, data distillation, and adversarial training to name a few. Traditional gradient-based strategies to solve such BO problems involve the computation of a hypergradient for the function at the upper level. This can be a computationally demanding task since a correct evaluation of such gradients not only requires accurate solutions at the lower level, but also the storage of large Jacobian matrices, making such strategies impractical in case of large scale problems.

This work investigates the development and application of inexact derivative-free methods for solving constrained BO problems arising from machine learning, with the specific interest in problems subject to polyhedral constraints at the upper level. In particular, starting from the inexact direct-search method introduced in [Diouane et al., COAP 2024] for unconstrained setting, we propose a new method for the case of constrained problems in which the construction of feasible directions is inspired by zeroth-order Frank-Wolfe methods [Dzahini et al, 2024]. We discuss the theoretical properties of the proposed algorithm and test its effectiveness on both synthetic and real-life instances.

3 - Three Alternating Projection Methods for Matrix Completion
Mattia Silei, Stefania Bellavia, Simone Rebegoldi

A Feasibility Problem (FP) consists of finding a point in the intersection of two sets. Since the 1950s [1], the Alternating Projection Method (APM) has been used to solve FP, and more recently a regularized version has been proposed. In the first part of our work, we design inexactness criteria to extend the regularized version to use inexact projection onto one of the sets, and describe a practical procedure to implement the criteria. In the second part, we introduce the Matrix Completion (MC) problem, which consists of reconstructing the missing entries of a partially observed matrix M by leveraging the assumption that M has low rank [2]. Then, we show how MC can be formulated as a FP, and describe how implement inexactness criteria check in the case of the rank level set, which is a nonconvex set. In particular, it consists of a new stopping criterion for the Lanczos method used for the truncated SVD computation. Finally, we present numerical experiments to compare the three variants of AP on the task of MC with both random matrices and matrices from applications. [1] Von Neumann J. Functional Operators (AM-22), Volume 2: The Geometry of Orthogonal Spaces, 1951 [2] Cand'es E. J. and Recht B., Exact matrix completion via convex optimization, Found. Comput. Math., 9 (2009), pp. 717-772.

4 - Bi- and Multi-Level Optimization Strategies for Sparse and Interpretable Learning in NMF
Laura Selicato

Learning algorithms require careful hyperparameter tuning, especially in penalized optimization where constraints aid in extracting interpretable, low-dimensional data representations. This work explores bi-level optimization problems for Nonnegative Matrix Factorization (NMF), a powerful tool for capturing latent structures while preserving nonnegativity. We introduce AltBi, a novel algorithm integrating penalty hyperparameter tuning into the Kullback-Leibler NMF update rules as a bi-level optimization problem. Additionally, we present AltBi-J, a constrained Frobenius norm approach using a diversity-based penalty which shows superior sparsity capabilities compared to traditional penalties. Extending this to the spectral domain, we propose the SHINBO algorithm, a bi-level optimization method applied to Itakura-Saito NMF, which adaptively tunes row-wise penalties to enhance the extraction of periodic signals from noisy data, with effective application to fault detection in mechanical systems. Motivated by learning applications involving multiple constraints, bi-level methodology can be generalized to a multi-level optimization framework capable of handling nested nonconvex problems, advancing the theoretical and algorithmic foundations for learning in structured optimization paradigms. This is joint work with Flavia Esposito and Nicoletta Del Buono from University of Bari Aldo Moro, Andersen Ang from University of Southampton, and Rafał Zdunek from Politechnika Wrocławska.

■ TC-04

Tuesday, 14:00-16:00 - Room: B100/5013

Global Optimization advances

Stream: Global optimization
Invited session
Chair: *Eligius M.T. Hendrix*
Chair: *Sonia Cafieri*

1 - On polytopal branch and bound
Eligius M.T. Hendrix, Boglárka G.-Tóth, Leocadio G. Casado, Frédéric Messine

Using Interval Arithmetic or simplicial Branch and bound exploiting monotonicity, our investigation turned to the question of using polytopal subsets for global optimization over a polytope feasible set. We derived underlying theory. However, to come to practical implementations of a Branch and bound code requires well thinking on how to store information of polytopes and the feasibele set. In this contribution, we report on our investigation during the last year and the challenges that remain.

2 - A branch-and-bound algorithm for nonconvex Nash equilibrium problems
Peter Kirst, Stefan Schwarze, Oliver Stein

In this talk, a spatial branch-and-bound method for the computation of the set of all epsilon-Nash equilibria of continuous box-constrained nonconvex Nash equilibrium problems is presented. Thereby, the existence of epsilon-Nash equilibria is not assumed, but the algorithm is also able to detect their absence. We explain appropriate discarding and fathoming techniques. The proposed algorithm is proven to terminate with a prescribed approximation tolerance. Finally, we report our computational experience.

3 - Global and Robust Optimisation for Non-Convex Quadratically Constrained Quadratic Programs
Asimina Marousi, Vassilis Charitopoulos

Solving non-convex problems under uncertainty introduces additional complexity compared to their deterministic counterparts. State-of-the-art robust optimisation algorithms typically follow two approaches: utilising local solvers to obtain robust feasible solutions or leveraging global solvers to ensure robust optimality, often at a significantly higher computational cost. In this work, we propose a novel algorithm integrating global and robust optimisation methods to solve continuous non-convex quadratically constrained quadratic programming (QCQP) problems under convex uncertainty sets. The proposed Robust spatial branch-and-bound (RsBB) algorithm combines the principles of spatial branch-and-bound (sBB) with robust cutting planes. Our research hypothesis is that integrated exploration of global and robust optimality can yield computational benefits. The RsBB algorithm utilises McCormick envelopes to obtain convex lower bounds. At each node, a local solver tackles the non-convex problem. If the solution matches the best found so far, an infeasibility test assesses its robustness; otherwise, cutting planes are added to both the non-convex and convex problems. The performance of the RsBB algorithm is compared with state-of-the-art methods that rely on global solvers. The findings of our work highlight the efficiency of the RsBB algorithm and provide insights to the advantages of combining robustness and optimality search.

■ TC-05

Tuesday, 14:00-16:00 - Room: B100/4013

Randomized Optimization algorithms II

Stream: Optimization for machine learning
Invited session
Chair: *Laurent Condat*

1 - Derivative-free stochastic bilevel optimization for inverse problems
Mathias Staudigl, Simon Weissmann, Tristan van Leeuwen

Inverse problems are key issues in several scientific areas, including signal processing and medical imaging. Data-driven approaches for inverse problems aim for learning model and regularization parameters from observed data samples, and investigate their generalization properties when confronted with unseen data. This approach dictates a statistical approach to inverse problems, calling for stochastic optimization methods. In order to learn model and regularisation parameters simultaneously, we develop in this paper a stochastic bilevel optimization approach in which the lower level problem represents a variational reconstruction method formulated as a convex non-smooth optimization problem, depending on the observed sample. The upper level problem represents the learning task of the regularisation parameters. Combining the lower level and the upper level problem leads to a stochastic non-smooth and non-convex optimization problem, for which standard gradient-based methods are not straightforward to implement. Instead, we develop a unified and flexible methodology, building on a derivative-free approach, which allows us to solve the bilevel optimization problem only with samples of the objective function values. We perform a complete complexity analysis of this scheme. Numerical results on signal denoising and experimental design demonstrate the computational efficiency and the generalization properties of our method.

2 - Proximal splitting algorithms in nonlinear spaces
Russell Luke

In the setting of CAT(????) spaces, common fixed point iterations built from prox mappings (e.g. prox-prox, Krasnoselsky-Mann relaxations, nonlinear projected-gradients) converge locally linearly under the assumption of linear metric subregularity. Linear metric subregularity is in any case necessary for linearly convergent fixed point sequences, so the result is tight. To show this, we develop a theory of fixed point mappings that violate the usual assumptions of nonexpansiveness and firm nonexpansiveness in p-uniformly convex spaces. This is specialized to the computation of Frechet means in nonlinear spaces using deterministic and radomized methods.

3 - SPAM: Stochastic Proximal Point Method with Momentum Variance Reduction for Non-convex Cross-Device Federated Learning
Avetik Karagulyan

Cross-device training is a crucial subfield of federated learning, where the number of clients can reach the billions. Standard approaches and local methods are prone to client drift and insensitivity to data similarities. We propose a novel algorithm (SPAM) for cross-device federated learning with non-convex and non-smooth losses. We provide a sharp analysis under second-order (Hessian) similarity, a condition satisfied by various machine learning problems in practice. Additionally, we extend our results to the partial participation setting, where a cohort of selected clients communicate with the server at each communication round. We then conduct a complexity analysis of our convergence results, showing the improvement of our methods upon prior work.

4 - A Stochastic Newton-type Method for Non-smooth Optimization

Titus Pinta

We introduce a new framework for analyzing (Quasi-Newton type methods applied to non-smooth optimization problems. The source of randomness comes from the evaluation of the (approximation) of the Hessian. We derive, using a variant of Chernoff bounds for stopping times, expectation and probability bounds for the random variable representing the number of iterations of the algorithm until approximate first order optimality conditions are validated. As an important distinction to previous results in the literature, we do not require that the estimator is unbiased or that it has finite variance. We then showcase our theoretical results in a stochastic Quasi-Newton method for X-ray free electron laser orbital tomography and in a sketched Newton method for image denoising.

■ TC-06

Tuesday, 14:00-16:00 - Room: B100/7013

Structured nonsmooth optimization – Part I

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Karl Welzel*

1 - Equivalence between Linear Convergence and Error Bounds for Optimization Algorithms

Kira van Treek, Javier Peña, Juan Vera, Luis Zuluaga

Most common iterative optimization algorithms are fundamentally fixed-point iterations of an averaged operator, which typically lead to sublinear convergence rates in worst-case scenarios. We establish that a specific error bound condition for these algorithms is equivalent to their linear convergence. This result has significant implications, since proving linear convergence is reduced to verifying a regularity condition on the corresponding operator, which enables the use of tools such as the Hoffman bounds to compute convergence rates. As an illustration, we enhance existing results on the linear convergence of the Alternating Direction Method of Multipliers (ADMM) and Douglas-Rachford algorithm for structured problems, including piecewise linear, quadratic and linear optimization. For linear optimization we show a convergence rate that is independent of the problem data. Furthermore, we use our results to prove finite termination of the Douglas-Rachford method when applied to polyhedral feasibility problems.

2 - Blockwise DC Programming

Hoomaan Maskan, Pouria Fatemi, Alp Yurtsever, Suvrit Sra

Difference of convex (DC) programs are a class of structured non-convex problems. Due to their general form, many applications find their abstractions through the lens of DC programming. In this work, we target the class of blockwise DC (BDC) functions. Careful formulation reveals the broad applicability of these programs in, for example, training deep neural networks. We propose blockwise DC algorithms (BDCA) to treat these problems under various conditions, such as generalized smoothness and stochastic gradients.

3 - Presolve techniques for quasi-convex chance constraints with finite-support low-dimensional uncertainty

Guillaume Van Dessel, François Glineur

Chance-constrained programs (CCP) represent a trade-off between conservatism and robustness in optimization. In many CCPs, one optimizes an objective under a probabilistic constraint continuously parameterized by a (random) vector. In this work, we study the case where the constraint is a quasi-convex function of this vector. Moreover, vector's support is a collection of N scenarios in dimension p = 2 or p = 3. In general, even when both the constraint and the objective are convex in the decision variable, the feasible region of a CCP is nonconvex, turn- ing it into a difficult problem. However, under mild assumptions, many CCPs can be recast as big-M mixed-integer convex programs (MICP). Unfortunately, the difficulty of these MICPs explodes with the number of scenarios, restricting the instances practically solvable in decent time. To cut down the effective number of scenarios considered in MICP re- formulations and accelerate their solving, we propose and test presolve techniques based on computational geometry. Our techniques produce certificates to discard or select a priori some scenarios before solving a regular MICP. Moreover, the information aggregated during presolve leverages the possibility to strengthen big-M constants. Our numerical experiments suggest that spending some time in presolve is more efficient than a direct solve for a class of probabilistic projection problems, including an interesting type of facility location problem

4 - Local convergence of adaptively regularized tensor methods

Karl Welzel, Coralia Cartis, Raphael Hauser, Yang Liu

Tensor methods are methods for unconstrained continuous optimization that can incorporate derivative information of up to order p > 2 by computing a step based on the pth-order Taylor expansion at each iteration. The most important among them are regularization-based tensor methods which have been shown to have optimal worst-case iteration complexity of finding an approximate minimizer. We are interested in how local convergence rates depend on the order of the Taylor expansion p. In the case of functions that are uniformly convex of order q (ones that around the minimizer x* grow like the distance to x* to the qth power) and a fixed regularization parameter, the answer is known: we get (p/(q-1))th-order local convergence of function values and gradient norms if p > q-1. In particular, when the Hessian is positive definite at the minimizer (q=2) we get pth-order convergence, but also when the Hessian is singular at x* (q>2) superlinear convergence (compared to Newton's linear convergence) is possible as long as enough derivatives are available. We extend these convergence results to locally uniformly convex functions and fully adaptive methods. We discuss how for p > 2 it becomes crucial to select the "right" minimizer of the regularized local model in each iteration to ensure all iterations are eventually successful. If the right minimizer is used, the (p/(q-1))th-order local convergence is preserved, otherwise the rate degrades but stays superlinear.

■ TC-07

Tuesday, 14:00-16:00 - Room: B100/5015

Mixed-Integer Bilevel Optimization

Stream: Bilevel and multilevel optimization
Invited session
Chair: *Abdellah Bulaich Mehamdi*

1 - A Bilevel Critical Node Detection Problem
Ashwin Arulselvan

In this work, we formulate a bilevel critical node detection problem for a given threat level and a budget. A leader has a budget to immunize a subset of nodes. An attacker, with the knowledge of the leader's choice, will remove any set of non-immunized nodes within their budget, which is the threat level. The leader seeks to maximise the pairwise connectivity of the nodes for the worst case removal strategy of the attacker. We solve this problem using a high point relaxation within a branch-and-bound framework. We introduce some valid inequalities to strengthen the formulation and introduce a branching strategy to deal with the bilevel infeasibility. We test this procedure on two graph families with varying number of nodes, edge densities and budgets and share our computational experience. We conclude with the note that the framework could be applied to any network interdiction problem, provided some properties are preserved.

2 - On a class of interdiction problems with partition matroids: complexity and polynomial-time algorithms
Sergei Ketkov

In this study, we consider a class of linear matroid interdiction problems, where the feasible sets for the upper-level decision-maker (referred to as a leader) and the lower-level decision-maker (referred to as a follower) are induced by two distinct partition matroids with a common weighted ground set. Unlike classical network interdiction models where the leader is subject to a single budget constraint, in our setting, both the leader and the follower are subject to several independent capacity constraints and engage in a zero-sum game. While the problem of finding a maximum weight independent set in a partition matroid is known to be polynomially solvable, we prove that the considered bilevel problem is NP-hard even when the weights of ground elements are all binary. On a positive note, it is revealed that, if the number of capacity constraints is fixed for either the leader or the follower, then the considered class of bilevel problems admits several polynomial time solution schemes. Specifically, these schemes are based on a single-level dual reformulation, a dynamic programming-based approach, and a greedy algorithm for the leader.

3 - Bi-Level Optimization for Electricity Pricing: Links with the Rochet-Choné Model
Abdellah Bulaich Mehamdi, Wim van Ackooij, Luce Brotcorne, Stephane Gaubert, Quentin Jacquet

Pricing strategies in electricity markets are crucial for optimizing supply and demand dynamics, ensuring that both producers and clients make effective decisions. These strategies help to manage grid stability, to encourage efficient energy use, and to promote competition among suppliers, ultimately benefiting clients through more competitive prices and better options.

In this talk, we model the pricing problem using a Principal-Agent framework under adverse selection: the Rochet-Choné model. This model addresses pricing in a monopolistic market under reservation constraints and captures the interaction between the provider and consumers. The goal is to determine tariffs that balance profitability with social welfare. To solve this problem, we discretize the client space and solve the corresponding bi-level problem using the value function approach.

Additionally, we introduce quantization methods to improve the representation of the provider's offer. Using a menu of pricing contracts derived from the Rochet-Choné model, we show how to identify a subset that best represents the full menu. First we model the problem as a mixed-integer bi-level optimization problem. Next, we describe a greedy heuristic to compute the reduced menu, leveraging tools such as the Newton polygon and Fenchel conjugacy. Numerical experiments on real-life data from EDF highlight the method's effectiveness.

■ TC-08

Tuesday, 14:00-16:00 - Room: B100/7007

Systematic and computer-aided analyses V: Tools for systematic studies of first-order algorithms

Stream: Systematic and computer-aided analyses of optimization algorithms
Invited session
Chair: *Manu Upadhyaya*
Chair: *Adrien Taylor*

1 - Data-driven Analysis of First-order Methods via Distributionally Robust Optimization
Jisun Park, Vinit Ranjan, Bartolomeo Stellato

In this talk, we present a tractable formulation using convex optimization for the probabilistic analysis of first-order methods. We introduce a tractable conic linear program which combines the performance estimation problem (PEP) with the distributionally robust optimization (DRO) approach. Solving this program provides probabilistic performance guarantees in terms of the expectation or conditional value-at-risk (CVaR) of the chosen performance measure, assuming the problem instances follow distribution accessible only through samples. Our framework bridges the worst-case and average-case analyses of first-order methods, by incorporating the data-driven information of the optimization problems. We provide numerical experiments including the probabilistic analysis of convex quadratic minimization or smooth convex minimization, obtaining significant reductions in pessimism compared to classical worst-case analysis.

2 - Computer-aided analysis of relatively inexact gradient descent in smooth convex optimization
Pierre Vernimmen, François Glineur

In this work, we analyze relatively inexact gradient descent, which is a version of gradient descent using at each step an approximation of the gradient. We call it relatively inexact as we require that the norm of the error is bounded by a fraction of the norm of the gradient itself. We prove that the method converges in gradient norm, and derive its exact worst-case convergence rate for any stepsize when performing one iteration. We also show that these tight rates belong to three distinct regimes. Two regimes (short and long steps) are simple adaptations of the exact case, corresponding to simple univariate worst-case functions (Huber or quadratic). The third intermediate regime does not occur in the exact case and corresponds to a bivariate worst-case function.

We also demonstrate how relatively inexact gradients naturally occur in the context of compressed gradient descent, in which the gradient is stored using a small number of bits of information. This compression technique is frequently used when computing on GPUs.

Finally, we conduct a numerical experiment using several first-order methods (classical gradient descent, long-step methods and accelerated methods) to check their robustness to relative inexactness empirically in a more practical setting. We also propose and test a simple and universal way to adapt methods to deal with relative inexactness, namely, to shorten all step sizes by a well-chosen factor.

3 - AutoLyap: A Python package for computer-assisted Lyapunov analyses for first-order methods
Manu Upadhyaya, Adrien Taylor, Sebastian Banert, Pontus Giselsson

We introduce AutoLyap, a Python package designed to automate Lyapunov analysis for a wide class of first-order methods for solving structured optimization and inclusion problems. Lyapunov analyses are structured proof patterns commonly used to establish convergence results for first-order methods. Building on previous work, the core idea behind AutoLyap is to recast the verification of the existence of a Lyapunov analysis as a semidefinite programming (SDP) problem, which can then be solved numerically using standard SDP solvers. Users of the package specify (i) the optimization or inclusion problem, (ii) the first-order method in question, and (iii) the type of Lyapunov analysis they wish to verify. Once these inputs are provided, AutoLyap handles the SDP modeling and proceeds with the numerical solution of the SDP. We numerically verify (and sometimes extend) numerous established convergence results, demonstrating practical relevance.

■ TC-09

Tuesday, 14:00-16:00 - Room: B100/8013

Variational Analysis II

Stream: Variational analysis: theory and algorithms
Invited session
Chair: *Miguel Angel Melguizo Padial*

1 - Bilevel variational inequalities and its applications
Aviv Gibali

In this talk, we give a brief introduction to fixed point problem (FPP), variational inequality problem (VIP) and a class of hierarchical VIP known as the bilevel variational inequality problem (BVIP).

We present a novel fixed point iterative method to approximate the solution of BVIP in real Hilbert spaces. The convergence analysis of this method with application to optimal control problems are discussed.

2 - On the variational inequality approach to Network Games
Fabio Raciti

Network Games are a class of games where players are identified with the nodes of a graph and the (social or economic) relationships between any two players are represented by arcs of the graph. The interaction between a generic player and their neighbors is thus described by means of the adjacency matrix of the graph and, under suitable assumptions, the Nash equilibrium of the game can be expressed as a power expansion of the adjacency matrix. Since the seminal paper [1], many game theorists have written numerous papers on this topic, but the powerful tool of variational inequalities has not been fully exploited yet, with some exceptions. In the last years, the author, (F.R.) , has carried out a systematic analysis of the variational inequality approach to Network Games from both the theoretical and numerical point of view. In this talk we will describe the methodology above mentioned and some recent results.

[1] C. Ballester, Calvò-Armengil, Y. Zenou: WHO'S WHO IN NETWORKS. WANTED: THE KEY PLAYER, Econometrica, Vol. 74, No. 5 (September, 2006), 1403-1417.

3 - Inexactness in Variational Problems
Lorenzo Lampariello, Giancarlo Bigi, Simone Sagratella, Valerio Giuseppe Sasso

We investigate the relationships between solution sets of some variational problems, including, e.g., equilibrium problems, fixed point problems and variational inequalities. Given that inexactness naturally arises both in modeling real-world phenomena and in designing numerical solution methods, we consider inexact formulations of these problems. Our goal is to establish quantitative relations that describe how inexactness in one problem propagates to another.

4 - Henig Proper Efficient Points: Existence and Density Beyond Convexity
Fernando García Castaño, Miguel Angel Melguizo Padial

In this talk, we present new results on the existence and density of Henig global proper efficient points in vector optimization problems within arbitrary normed spaces. Our approach does not require convexity and, in some cases, applies to unbounded sets. However, a weak compactness condition –either on the set or a section of it– remains essential, along with a key separation property between the order cone and its conical neighborhoods. These conditions ensure the necessary convergence properties and allow the interpolation of a family of Bishop-Phelps cones between the order cone and its neighborhoods. This interpolation, together with a careful interaction between two types of conic neighborhoods, plays a fundamental role in our proofs. Our results generalize some existing results in the literature that were obtained under more restrictive assumptions.

■ TC-10

Tuesday, 14:00-16:00 - Room: B100/8011

Continuous Multi-Objective Optimization: Algorithms and Complexity Analyses

Stream: Multiobjective and Vector Optimization
Invited session
Chair: *Pierluigi Mansueto*

1 - On the convergence of outer approximation algorithms for convex vector optimization problems
Firdevs Ulus, Cagin Ararat, Muhammad Umer

Recently in [Ç. Ararat, F. Ulus, M. Umer, SIAM Journal on Optimization, 34:3 (2024), pp 2169-3166] we studied the convergence rate of an outer approximation algorithm for solving bounded convex vector optimization problems. The algorithm solves a norm minimizing scalarization in each iteration and terminates after finitely many iterations ensuring that the Hausdorff distance between the outer approximation and the upper image is less than a given approximation error. The norm is arbitrary but fixed through the iterations. The strong convergence rate of O(k(2/(1-q))) was proved for the Euclidean norm.

We now prove the same convergence rate for any arbitrary fixed norm. Moreover, if the norm used in the scalarization changes through the iterations, then it is possible to obtain the same convergence rate under some assumptions on the set of norms used. We show that Pascoletti-Serafini scalarization can be seen as a norm-minimizing scalarization, where the norm depends on the parameters of the scalarization.

2 - Effective Front-Descent Algorithms with Convergence Guarantees
Matteo Lapucci, Pierluigi Mansueto, Davide Pucci

We address continuous unconstrained multi-objective optimization problems and we discuss descent type methods for the reconstruction of the Pareto set. Specifically, we describe the class of Front Descent methods, which generalizes the Front Steepest Descent algorithm allowing the employment of suitable, effective search directions (e.g., Newton, Quasi-Newton, Barzilai-Borwein). We give a deep characterization of the behavior and the mechanisms of the algorithmic framework, and we underline how, under reasonable assumptions, standard convergence results and some complexity bounds hold for the generalized approach. Moreover, we remark that popular search directions can indeed be soundly used within the framework. We furthermore present a completely novel type of convergence results, concerning the sequence of sets produced by the procedure. These results concern convergence to stationarity for any sequence of induvidual points in these sets; additionally, some interesting properties are discovered that are useful in finite precision settings. Finally, the results from a large experimental benchmark are shown.

3 - Worst-case complexity analysis of a-posteriori methods for multi-objective optimization
Giampaolo Liuzzi, Andrea Cristofari, Marianna De Santis, Stefano Lucidi

In this work, we consider unconstrained a-posteriori multi-objective optimization problems. The main aim of the work is to give worst-case complexity bounds for multi-objective optimization methods which adopt a linesearch technique along steepest descent directions. We show that the considered methods enjoy the same worst-case complexity bounds recently proved in the literature for derivative-free methods.

4 - Worst-case Complexity in Continuous Multi-Objective Optimization
Rohollah Garmanjani

In this talk, we delve into the worst-case complexity of continuous optimization, which quantifies the computational effort required for an algorithm to reduce a stationarity measure below a given positive threshold in the worst-case scenario. We begin by providing an overview of worst-case complexity in single-objective optimization, outlining foundational results to serve as a benchmark.

We then shift our focus to the more complex realm of multiobjective optimization, highlighting its distinct challenges and recent advancements. Lastly, we examine the worst-case complexity of a trust-region algorithm, analyzing its performance under both convexity and strong convexity assumptions.

■ TC-11

Tuesday, 14:00-16:00 - Room: B100/5017

Advances in Manifold and Conic Optimization

Stream: Riemannian Manifold and Conic Optimization
Invited session
Chair: *Panos Parpas*

1 - Conjectures in Real Algebra and Polynomial Optimization through High Precision Semidefinite Programming
Michal Kocvara, Lorenzo Baldi

We study degree bounds for the denominator-free Positivstellens"atze in real algebra, based on sums of squares (SOS), or equivalently the convergence rate for the moment-sums of squares hierarchy in polynomial optimization, from a numerical point of view. As standard semidefinite programming (SDP) solvers do not provide reliable answers in many important instances, we use a new high-precision SDP solver, Loraine.jl, to support our investigation.

We study small instances (low-degree, small number of variables) of one-parameter families of examples, and propose several conjectures for the asymptotic behavior of the degree bounds. Our objective is twofold: first, to raise awareness on the bad performance of standard SDP solvers in such examples, and then to guide future research on the Effective Positivstellens"atze.

2 - Riemannian Optimization on Nonnegative Matrix Factorizations: chordal, oblique and more
Man Shun Andersen Ang, Flavia Esposito

Nonnegative Matrix Factorization (NMF) approximates a nonnegative matrix as a conic combination of two nonnegative, low-rank matrices. We explore Riemannian optimization in NMF through two examples.

1. Chordal-NMF: We replace the "point-to-point" distance in NMF with a ray-to-ray chordal distance, casting Chordal-NMF onto a manifold. Unlike prior methods requiring smooth manifolds, nonnegativity here yields a non-differentiable one. We propose a Riemannian Multiplicative Update (RMU) that ensures convergence without smoothness, showing its efficacy on multispectral images.

2. Sparse Orthogonal NMF: We add a nonconvex ell_0.5-quasi-norm for sparsity and orthogonal constraints to NMF factors. Replacing nonnegativity with a Hadamard product, we solve it via RMU on the Oblique manifold.

If time permit, we'll discuss NMF on the Stiefel manifold, SO(n), and entropic regularization in Lie groups.

Joint works with Flavia Esposito (Università degli Studi di Bari Aldo Moro).

3 - Second-order differential operators, stochastic differential equations and Brownian motions on embedded manifolds
Du Nguyen, Stefan Sommer

The article provides a framework to simulate Riemannian Brownian and Riemannian Langevin equations on embedded manifolds in global coordinates using retraction, important in MCMC sampling on manifolds, and is relevant to simulated annealing on manifolds. We specify the conditions when a manifold M embedded in an inner product space E is an invariant manifold of a stochastic differential equation (SDE) on E, linking it with the notion of second-order differential operators on M. When M is given a Riemannian metric, we derive a simple formula for the Laplace-Beltrami operator in terms of the gradient and Hessian on E and construct the Riemannian Brownian motions on M as solutions of conservative Stratonovich and Ito SDEs on E. We derive explicitly the SDE for Brownian motions on several important manifolds in applications, including left-invariant matrix Lie groups using embedded coordinates, Stiefel, Grassmann and symmetric positive definite (SPD) manifolds. Numerically, we propose three simulation schemes to solve SDEs on manifolds. In addition to the stochastic projection method, to simulate Riemannian Brownian motions, we construct a second-order tangent retraction of the Levi-Civita connection using a given E-tubular retraction. We also propose the retractive Euler-Maruyama method to solve a SDE, taking into account the second-order term of a tangent retraction. We provide software to implement the methods in the paper, including Brownian motions of the manifolds discussed

4 - Faster Randomized Methods for Orthogonality Constrained Problems
Boris Shustin, Haim Avron

Recent literature has advocated the use of randomized methods for accelerating the solution of various matrix problems arising in machine learning and data science. One popular strategy for leveraging randomization in numerical linear algebra is to use it as a way to reduce problem size. However, methods based on this strategy lack sufficient accuracy for some applications. Randomized preconditioning is another approach for leveraging randomization in numerical linear algebra, which provides higher accuracy. The main challenge in using randomized preconditioning is the need for an underlying iterative method, thus, randomized preconditioning so far has been applied almost exclusively to solving regression problems and linear systems. In this talk, we show how to expand the application of randomized preconditioning to another important set of problems prevalent in machine learning: optimization problems with (generalized) orthogonality constraints. We demonstrate our approach, which is based on the framework of Riemannian optimization and Riemannian preconditioning, on the problem of computing the dominant canonical correlations.

■ TC-12

Tuesday, 14:00-16:00 - Room: B100/8009

Optimization for sustainable energy systems

Stream: Applications: AI, uncertainty management and sustainability
Invited session
Chair: *Miguel Anjos*

1 - Optimal management of electric vehicle fleets: An optimistic bilevel optimization model
Eleni Michaelidou, Miguel Anjos

We consider an electric vehicle (EV) charging service provider that operates a set of charging stations and uses time-based pricing to minimize the impact of a large EV fleet — as operated by transportation and delivery companies, among others — on the power grid by encouraging the fleet operator to shift its flexible load to off-peak hours and benefit from lower electricity rates. We propose an optimistic bilevel optimization problem that captures the hierarchical interaction between the provider (leader) and the EV fleet owner (follower). The latter can participate in vehicle-to-grid (V2G) schemes, where EV batteries can be used as energy storage devices to supply energy back to the grid during peak hours. At the upper level, the provider defines the charging and discharging prices to minimize peak demand. In response, at the lower level, the EV fleet owner decides when to charge and whether and when to discharge while considering several operational constraints of the fleet, mainly related to the delivery schedule of the fleet, to minimize its total charging cost. The bilevel optimization problem is reformulated into a single-level formulation using a Karush-Kuhn-Tucker (KKT)-based heuristic approach along with the Special Ordered Sets 1 (SOS1) technique, which can be solved using off-the-shelf solvers. We report preliminary computational results for the proposed model.

2 - A Stochastic Optimization Model for the Water Pump Scheduling Problem with Demand Response in Large and High Altitude Water Supply Systems
Miguel Anjos

The water pump scheduling problem is to determine which water pumps will be turned on or off at each time period over a given time horizon for a given water supply system. Water networks, as energy-intensive infrastructures, are promising candidates to offer the power system a reduction in their energy consumption during certain hours of the day, a service known as demand response. The reduction is typically at times with a positive difference between the electricity spot price and the contracted energy price. However, both the spot price and the water demand are uncertain. We propose a new two-step stochastic optimization model for demand response in large and high-altitude water supply systems. We use a binary expansion approach to efficiently model the nonlinearities by reducing the computational difficulties while maintaining a good representation of the physical phenomena involved. The first stage uses a robust water profile optimization model, and the second stage achieves the power profile with a demand response to obtain the optimum water pump and demand response bidding schedule. We tested this approach using a case study from a mining company's water supply system. Our findings concluded that different seasons and energy policies, such as the minimum power requirement and availability bonus, can significantly impact the water supply system's total costs and the amount of demand response available to offer on the capacity market.

3 - A stochastic programming model for planning CO2 transport infrastructure
Lihan Zhang

Carbon capture and storage (CCS) can be regarded as a significant method to address climate change challenges and reduce greenhouse gases. Since some CCS technologies are novel and in the early stages, it is vital to consider the influence of uncertainty of CO2 capture rate on potential stakeholders in making informed transport investment decisions. Firstly, one cluster model is constructed, this research aims to develop a suitable scenario tree for the CO2 capture project and apply it in a multistage stochastic programming. When constructing the tree of CO2 capture scenarios, it is important to consider possible shut down and re-establishment situations to account for the risks associated with real-world projects. By considering various cases of scenario trees with their associated probabilities, the model offers investors the flexibility to assess different potential outcomes of CO2 capture development. In addition, the flexibility of ships is considered to test whether the shutdown possibilities influence the chosen of transport mode. To make the model more realistic, it would be meaningful to include multiple clusters and observe how they cooperate. This will foster collaboration between clusters and the development of CCS projects in the UK. We are now working on the study of how clusters in the UK can cooperate to create a transport network, taking into account the existence of CO2 capture uncertainty in each capture cluster.

Tuesday, 16:20-17:30

■ TD-01

Tuesday, 16:20-17:30 - Room: B100/1001

Plenary 3 (EUROPT Lecture)

Stream: Plenaries
Invited session
Chair: *Giancarlo Bigi*
Chair: *Immanuel Bomze*

1 - Contextual Stochastic Bilevel Optimization
Daniel Kuhn

We introduce contextual stochastic bilevel optimization (CSBO) - a stochastic bilevel optimization framework with the lower-level problem minimizing an expectation conditioned on contextual information and on the upper-level decision variable. We also assume that there may be multiple (or even infinitely many) followers at the lower level. CSBO encapsulates important applications such as meta-learning, personalized federated learning, end-to-end learning, and Wasserstein distributionally robust optimization with side information as special cases. Due to the contextual information at the lower level, existing single-loop methods for classical stochastic bilevel optimization are not applicable. We thus propose an efficient double-loop gradient method based on the Multilevel Monte-Carlo (MLMC) technique. When specialized to stochastic nonconvex optimization, the sample complexity of our method matches existing lower bounds. Our results also have important ramifications for three-stage stochastic optimization and challenge the long-standing belief that three-stage stochastic optimization is harder than classical two-stage stochastic optimization.

Wednesday, 9:00-10:00

■ WA-01

Wednesday, 9:00-10:00 - Room: B100/1001

Plenary 4

Stream: Plenaries
Invited session
Chair: *Alain Zemkoho*

1 - Non-Smooth Optimization on Stiefel Manifold and Beyond
Anthony So

Optimization problems on smooth manifolds often exhibit properties similar to those of unconstrained optimization problems. This has motivated the development of many elegant theoretical tools and efficient numerical methods in manifold optimization over the past decades. Most existing works focus on the setting of optimizing smooth functions on smooth manifolds. In recent years, due to various applications in data science and signal processing, there has been growing interest in developing numerical methods that are provably and practically efficient for optimizing non-smooth functions on smooth manifolds. In this talk, we focus on non-smooth optimization problems on the Stiefel manifold, in which the objective function either has certain composite form or is weakly convex. We present proximal and subgradient-type algorithms for tackling these problems and discuss their convergence properties. We also discuss possible extensions of these developments for non-smooth optimization on other smooth manifolds.

Wednesday, 10:30-12:30

■ WB-01

Wednesday, 10:30-12:30 - Room: B100/1001

Advances in stochastic and non-euclidean first order methods

Stream: Zeroth and first-order optimization methods
Invited session
Chair: *Cesare Molinari*
Chair: *Silvia Villa*

1 - Non-monotone stochastic line search without overhead for training neural networks
Andrea Cristofari, *Leonardo Galli*, *Stefano Lucidi*

In this work, we investigate the use of a new non-monotone strategy for stochastic gradient. In order to reduce the number of stepsize computations in the non-monotone line search, appropriate tests are used during the iterations to check if the Polyak stepsize can be accepted independently on the objective decrease. A preliminary theoretical analysis is illustrated and numerical results are given for the training of artificial neural networks on different datasets.

2 - On the convergence of stochastic Bregman proximal gradient algorithms with biased gradient estimators
Thomas Guilmeau, *Emilie Chouzenoux*, *Víctor Elvira*

Bregman proximal-gradient algorithms generalize proximal-gradient algorithms by measuring proximity between iterates with a Bregman divergence instead of the Euclidean distance. The choice of Bregman divergence can match the geometry of the objective function, making Bregman proximal-gradient algorithms well-suited for optimization problems with ill-behaved curvature. The convergence of deterministic Bregman proximal-gradient algorithms can be established as in the Euclidean case by leveraging the properties of relative smoothness and relative strong convexity, which generalize the Euclidean properties of smoothness and strong convexity. The stochastic setting is however more difficult, due to the difficulty of controlling the variance of the gradient estimators with Bregman divergences. As a consequence, most convergence results for stochastic Bregman-based methods assume the (Euclidean) strong convexity of the considered Bregman function. Moreover, the convergence of Bregman-based methods with biased gradient estimators is seldom studied. Motivated by applications in computational statistics (variational inference, adaptive importance sampling) where these assumptions fail, we present asymptotic and non-asymptotic convergence results for a stochastic Bregman proximal-gradient algorithm with biased gradient estimators, without assuming strong convexity of the Bregman function. This presentation is based on results from https://arxiv.org/abs/2211.04776.

3 - Riemannian gradient descent improves parameter-efficient fine-tuning
Bingcong Li

Low rank adapters (LoRA) leverage the classical Burer-Monteiro (BM) factorization to enable parameter-efficient fine-tuning of large language models (LLMs). However, recent studies have theoretically demonstrated that gradient descent (GD) can suffer from exponential slowdowns in LoRA training when an improper rank is chosen. To address this, we propose semi-orthogonal low-rank Adapters (SoLA), inspired by SVD-based factorization. We prove that Riemannian Gradient Descent (RGD) with semi-orthogonal constraints (i.e., Stiefel manifolds) overcomes slow convergence in such unfavorable cases. Guided by our theoretical insights, we apply SoLA to fine-tune LLMs and demonstrate its efficiency on large-scale tasks.

4 - General Tail Bounds for Non-Smooth Stochastic Mirror Descent
Andrea Paudice, *Khaled Eldowa*

We consider the problem of minimizing a convex non-smooth Lipschitz function over a convex domain when the optimizer is given access to noisy stochastic estimates of its subgradients. We analyze the classical Stochastic Mirror Descent method and provide novel tail bounds on the optimization error for both the average and the last iterate. Our results extend the existing tail bounds from the classical light-tailed Sub-Gaussian noise case to heavier-tailed noise regimes. We instantiate our results in two important cases: a class of noise with exponential tails and one with polynomial tails. A remarkable feature of our results is that they do not require an upper bound on the diameter of the domain. Finally, we support our theory with illustrative experiments that compare the behavior of the average of the iterates with that of the last iterate in heavy-tailed noise regimes.

■ WB-02

Wednesday, 10:30-12:30 - Room: B100/7011

High-order and tensor methods

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Wenqi Zhu*

1 - Efficient Adaptive Regularized Tensor Methods
Yang Liu, *Karl Welzel*, *Coralia Cartis*, *Raphael Hauser*, *Wenqi Zhu*

High-order tensor methods employing local Taylor approximations have attracted considerable attention for convex and nonconvex optimization. The pth-order adaptive regularization (ARp) approach builds a local model comprising a pth-order Taylor expansion and a (p+1)th-order regularization term, delivering optimal worst-case global and local convergence rates. However, for p2, subproblem minimization can yield multiple local minima, and while a global minimizer is recommended for p=2, effectively identifying a suitable local minimum for p3 remains elusive. This work extends interpolation-based updating strategies, originally proposed for p=2, to cases where p3, allowing the regularization parameter to adapt in response to interpolation models. Additionally, it introduces a new prerejection mechanism to discard unfavorable subproblem minimizers before function evaluations, thus reducing computational costs for p3. Numerical experiments, particularly on Chebyshev-Rosenbrock problems with p=3, indicate that the proper use of different minimizers can significantly improve practical performance, offering a promising direction for designing more efficient high-order methods.

2 - A low-rank augmented Lagrangian method for doubly nonnegative relaxations of mixed-binary quadratic programs
Tianyun Tang

Doubly nonnegative (DNN) programming problems are known to be challenging to solve because of their huge number of (n2) constraints and (n2) variables. In this work, we introduce RNNAL, a method for solving DNN relaxations of large-scale mixed-binary quadratic programs by leveraging their solutions' possible low-rank property. RNNAL is a globally convergent Riemannian augmented Lagrangian method (ALM) that penalizes the nonnegativity and complementarity constraints while preserving all other constraints as an algebraic variety. After applying the low-rank decomposition to the ALM subproblem, its feasible region becomes an algebraic variety with favorable geometric properties. Our low-rank decomposition model is different from the standard Burer-Monteiro (BM) decomposition model in that we make the key improvement to equivalently reformulate most of the quadratic constraints after the BM decomposition into fewer and more manageable affine constraints. This modification is also important in helping us to alleviate the violation of Slater's condition for the primal DNN problem. Moreover, we make the crucial step to show that the metric projection onto the algebraic variety, although non-convex, can be transformed into a solvable convex optimization problem under certain regularity conditions, which can be ensured by a constraint-relaxation strategy. Numerous numerical experiments are conducted to validate the efficiency of the proposed RNNAL method.

3 - On Global Rates for Regularization Methods Based on Secant Derivative Approximations
Sadok Jerad, Coralia Cartis

An approximation framework for adaptive regularization methods is presented, in which approximations are allowed only for the pth-order tensor. Between each recalculation of the pth-order derivative approximation, a high-order secant equation can be used to update the pth-order tensor as proposed in (Welzel2024), or the approximation can be kept constant in a lazy manner. When refreshing the pth-order tensor approximation after m steps, an exact evaluation of the tensor or a finite differences can be used with an explicit discretization stepsize. For all these new introduced variants, we establish the standard complexity bound on the number of iterations of standard adaptive regularization methods. Results are also specified for quasi-Newton methods for $p=2$.

4 - Beyond Second Order Methods for Nonconvex Optimization
Wenqi Zhu

Traditionally, first-order gradient-based techniques, such as stochastic gradient descent (SGD), and second-order methods, such as the Newton method, have dominated the field of optimization. In recent years, high-order tensor methods with regularization for nonconvex optimization have garnered significant research interest. These methods offer superior local convergence rates, improved worst-case evaluation complexity, enhanced insights into data geometry through higher-order information, and better parallelization compared to SGD. The most critical challenge in implementing the pth-order method ($p geq 3$) lies in efficiently minimizing the pth-order subproblem, which typically consists of a pth-degree multivariate Taylor polynomial combined with a $(p+1)$th-order regularization term. In this talk, we address the research gaps by characterizing the local and global optimality of the subproblem and investigating its potential NP-hardness. In this talk, we will introduce and discuss a series of provably convergent and efficient algorithms for minimizing the regularized subproblem both locally and globally, including the Quadratic Quartic Regularization Method (QQR), the Cubic Quartic Regularization Method (CQR), and the Sums-of-Squares Convex Taylor Method (SoS-C). More interestingly, our research adopts an AI-integrated approach, using the mathematical reasoning capabilities of large language models (LLMs) to verify the nonnegativity of multivariate polynomials.

■ WB-03

Wednesday, 10:30-12:30 - Room: B100/4011

Recent Advances in Line-Search Based Optimization

Stream: Large scale optimization: methods and algorithms
Invited session
Chair: *Matteo Lapucci*

1 - Line Search Methods are Sharpness-Aware and Operate at the Edge of Stability
Leonardo Galli

The traditional convergence analysis of Gradient Descent (GD) assumes the step size to be bounded from above by twice the reciprocal of the Lipschitz smoothness constant. However, recent numerical observations on neural networks have shown that GD also converges with larger step sizes. In this case, GD may enter into the so-called edge of stability phase in which the objective function decreases faster than with smaller steps, but nonmonotonically. Interestingly enough, this same behavior was already observed roughly 40 years ago when the first nonmonotone line search was proposed. In this paper, we show that nonmonotone line searches are able to operate at the edge of stability regime right from the start of the training. The numerical observations are supported by a theoretical analysis showing that line search methods yield step sizes that are indeed dependent on the Lipschitz smoothness constant and, consequently, on the sharpness (i.e., the largest eigenvalue of the hessian of the loss). We thus compare nonmonotone line search methods with SAM (Sharpness-Aware Minimization) showing that the first outperform the second in terms of speed of convergence, flatness of the solutions, and per-iteration costs.

2 - A Gradient Method with Momentum for Riemannian Manifolds
Diego Scuppa, Filippo Leggio, Marco Sciandrone

In this work we consider smooth optimization problems on Rieman- nian manifolds and we propose gradient methods with momentum. Convergence results are established and computational experiments are performed, on both academic tasks and real-world problems. The obtained computational results show the validity of the proposed approach that extends that recently presented.

3 - A Variable Dimension Sketching Strategy for Nonlinear Least-Squares
Greta Malaspina, Stefania Bellavia, Benedetta Morini

We present a stochastic inexact Gauss-Newton method for the solution of nonlinear least squares. To reduce the computational cost with respect to the classical method, at each iteration the proposed algorithm approximately minimizes the local model on a random subspace. The dimension of the subspace is adaptive, and two strategies are considered for its update, one that is based solely on Armijo condition, and one that takes into account the true Gauss-Newton model. Under suitable assumptions on the objective function and the random subspace, we prove a probabilistic bound on the number of iterations needed by the method to reduce the norm of the gradient below any given threshold. The numerical experiments demonstrate the effectiveness of the proposed method, compared to classical Gauss-Newton method and to the method that employs random subspaces with non-adaptive dimension.

4 - Stochastic line-search-based optimization for training overparameterized models: convergence conditions and effective approaches to leverage momentum
Davide Pucci, Matteo Lapucci

In recent years, the adoption of line-search techniques within incremental gradient-based methods for finite-sums problems has gathered considerable interest among researchers. Most recent works in the literature focus on incorporating line-searches into Stochastic Gradient Descent (SGD), as the use of a descent direction for the mini-batch objective is essential to ensure the line-search terminates in a finite number of steps. In this talk, we analyze how different search directions can be soundly used alongside stochastic line-searches. We define conditions on the sequence of search directions that guarantee finite termination and provide bounds for the backtracking procedure. Moreover, we shed light on the additional property of directions that is required to prove fast (linear) convergence of this general class of algorithms when applied to PL functions in the interpolation regime. We then focus on the special case of SGD with Polyak's momentum, analyzing the challenges arising when using line-searches with this search direction, and proposing a solution to overcome them. We present an algorithmic framework that effectively leverages the momentum direction alongside stochastic line-search, using a conjugate-gradient-type rule for the definition of the momentum parameter. Finally, we present a computational comparison, carried out on convex and nonconvex problems, showing the strong empirical performance of our method, which outperforms state-of-the-art approaches.

■ WB-04

Wednesday, 10:30-12:30 - Room: B100/5013

Optimization and learning for estimation problems

Stream: Optimization for machine learning
Invited session
Chair: *Laurent Condat*

1 - Multilevel Plug-and-Play Image Restoration
Nils Laurent, Julian Tachella, Elisa Riccietti, Nelly Pustelnik

Plug-and-play (PnP) image reconstruction methods leverage pretrained deep neural network denoisers as image priors to solve general inverse problems, and can obtain a competitive performance without having to train a network on a specific problem. Despite their flexibility, PnP methods often require several iterations to converge and their performance can be highly sensitive to the choice of the initialization and of the hyperparameters. In this talk, we propose a new multilevel PnP framework that accelerates the convergence by combining iterations at different scales, i.e. involving different resolutions of the image, and improves the robustness to initialization and hyperparameters setting using a coarse-to-fine strategy. In a series of experiments, including image inpainting and demosaicing, we show that the proposed multilevel PnP method outperforms other PnP methods in both speed and reconstruction performance.

2 - Anomaly Detection Using the Cloud of Spheres Classification Method
Paula Amaral, Tiago Dias

Machine learning models have been extensively applied across various domains, often without a deep understanding of their underlying mechanisms. Black-box models, such as Deep Neural Networks, present significant challenges in counterfactual analysis, interpretability, and explainability.

In this presentation, we introduce a novel binary classification model called the Connected Cloud of Spheres. This model is formulated as a Mixed-Integer Nonlinear Programming (MINLP) problem, aiming to minimize the number of spheres required to accurately classify data points. The method is particularly effective for datasets with highly non-linear and non-convex structures while remaining adaptable to linearly separable cases. Unlike neural networks, our approach operates directly in the original feature space, eliminating the need for kernel functions or extensive hyperparameter tuning.

Although primarily designed for binary classification, this method can be extended for anomaly detection, particularly in scenarios where negative examples are unavailable at the outset. Additionally, we discuss heuristic strategies for outlier identification and explainability, offering insights into how this approach enhances model transparency and interpretability.

3 - Learning Proximal Neural Networks at Equilibrium Without Jacobian
Leo Davy, Nelly Pustelnik, Luis Briceño-Arias

Proximal Neural Networks (PNNs) have been introduced as a principled framework for designing neural networks with desirable properties such as interpretability, stability, and convergence guarantees. In this talk, we focus on using PNNs to solve parametric variational formulations and address the bilevel optimization problem of learning hyperparameters. We consider the Deep Equilibrium formalism and demonstrate how the convergence properties of convex optimization algorithms can simplify the learning process by justifying the use of a technique known in the literature as Jacobian-Free Backpropagation. Finally, we present numerical results in image restoration to illustrate the effectiveness of our approach.

4 - Resource-Constrained Plug-and-Play Imaging: a block proximal heavy ball approach
Andrea Sebastiani, Federica Porta, Simone Rebegoldi

In this talk, we propose a novel memory-efficient optimization framework for Plug-and-Play (PnP) approaches in computational imaging. We introduce a block proximal variant of an heavy ball method that enables effective image reconstruction in resource-constrained settings. In particular, the memory required to perform the denoising step on the single block is significantly reduced, compared to the full image counterpart. Theoretical convergence guarantees are established under mild assumptions, and extensive experimental results demonstrate the efficacy of our method across multiple imaging problems including deblurring and super-resolution. The proposed framework extends the applicability of PnP reconstruction techniques to limited-memory scenarios without sacrificing performance or convergence properties.

■ WB-05

Wednesday, 10:30-12:30 - Room: B100/4013

Recent advances in min-max optimization

Stream: Optimization for machine learning
Invited session
Chair: *Ali Kavis*
Chair: *Aryan Mokhtari*

1 - Steering Towards Success: Efficient Methods for Nonconvex-Nonconcave Minimax Problems

Pontus Giselsson, Anton Åkerman, Max Nilsson, Manu Upadhyaya, Sebastian Banert

Nonconvex-nonconcave minimax problems frequently arise in applications, yet finding even a first-order stationary point is generally intractable without additional structure. The recently introduced Weak Minty variational inequality framework imposes such structure, enabling specialized extragradient-type methods with global convergence guarantees. Building on one of them, AdaptiveEG+, we propose three new algorithms that retain the same global convergence guarantees. One integrates momentum, while the other two incorporate Anderson acceleration directions through a novel one-shot line search strategy to combine the global convergence of AdaptiveEG+ with the fast local convergence of Anderson acceleration. All three methods derive from our new D-FLEX framework for solving firmly quasinonexpansive fixed-point problems, which leverages steering vectors to enhance practical performance. Indeed, numerical experiments showcase superior performance for the proposed methods on some challenging nonconvex-nonconcave minimax problems.

2 - A Universally Optimal Primal-Dual Method for Minimizing Heterogeneous Compositions

Benjamin Grimmer

This talk will propose a universal, optimal algorithm for convex minimization problems of the composite form g0(x)+h(g1(x),. . . ,gm(x))+u(x), which take a minimax form when h is dualized. We allow each gj to independently range from being nonsmooth Lipschitz to smooth, from convex to strongly convex, described by notions of Hölder continuous gradients and uniform convexity. Note that, although the objective is built from a heterogeneous combination of such structured components, it does not necessarily possess smoothness, Lipschitzness, or any favorable structure overall other than convexity. Regardless, we provide a universal optimal method in terms of oracle access to (sub)gradients of each gj. The key insight enabling our optimal universal analysis is the construction of two new constants, the Approximate Dualized Aggregate smoothness and strong convexity, which combine the benefits of each heterogeneous structure into single quantities amenable to analysis. As a key application, fixing h as the nonpositive indicator function, this model readily captures functionally constrained minimization g0(x)+u(x) subject to gj(x)0.

3 - Parameter-free second-order methods for min-max optimization

Ali Kavis, Ruichen Jiang, Qiujiang Jin, Sujay Sanghavi, Aryan Mokhtari

I will talk about an adaptive, line-search-free second-order methods with optimal rate of convergence for solving convex-concave min-max problems. By means of an adaptive step size, our algorithms feature a simple update rule that requires solving only one linear system per iteration, eliminating the need for line-search or backtracking mechanisms. Specifically, we base our algorithms on the optimistic method and appropriately combine it with second-order information. Moreover, distinct from common adaptive schemes, we define the step size recursively as a function of the gradient norm and the prediction error in the optimistic update. We first analyze a variant where the step size requires knowledge of the Lipschitz constant of the Hessian. Under the additional assumption of Lipschitz continuous gradients, we further design a parameter-free version by tracking the Hessian Lipschitz constant locally and ensuring the iterates remain bounded. We also evaluate the practical performance of our algorithm by comparing it to existing second-order algorithms for minimax optimization. Inspired by the adaptive design of the step size, we propose a heuristic initialization rule that performs competitively across different problems and scenarios and eliminates the need to fine tune the step size.

■ WB-06

Wednesday, 10:30-12:30 - Room: B100/7013

Structured nonsmooth optimization – Part II

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Behzad Azmi*

1 - Douglas-Rachford splitting algorithm for projected solution of quasi variational inequality with non-self constraint map

Maede Ramazannejad

In this paper, we present a Douglas-Rachford splitting algorithm within a Hilbert space framework that yields a projected solution for a quasi variational inequality. This is achieved under the conditions that the operator associated with the problem is Lipschitz continuous and strongly monotone. The proposed algorithm is based on the interaction between the resolvent operator and the reflected resolvent operator.

2 - On the relation between approximate and exact stationarity in disjunctive optimization

Isabella Käming, Patrick Mehlitz

We consider approximate stationarity conditions, constraint qualifications, and qualification conditions for optimization problems with disjunctive constraints, i.e., geometrically constrained problems with a constraint set that can be written as a union of convex polyhedral sets. This class of optimization problems covers, among others, optimization problems with complementarity, vanishing, switching, or cardinality constraints. In this talk, we first investigate suitable concepts of approximate stationarity and their connection to Mordukhovich- and, more importantly, even strong stationarity. The respective connections are established using strict constraint qualifications, particularly, so-called approximate regularity conditions. As an alternative approach to using strict constraint qualifications, we then introduce a new qualification condition which, based on an approximately stationary point, can again be used to infer its Mordukhovich- or strong stationarity. In contrast to the approximate regularity conditions, this condition depends on the involved sequences justifying approximate stationarity and, thus, is not a constraint qualification in the narrower sense. However, we will see that the new condition offers the benefits of being strictly weaker and much easier to verify than the strict constraint qualifications.

3 - Constrained Optimization in the Presence of Noise
Figen Oztoprak

In this work, we consider solving nonlinear optimization problems when only noisy evaluations of the objective and constraint functions are available. In particular, we consider two variants of the line-search sequential quadratic programming (SQP) method. The first variant is designed to work with noisy equality constraints, whereas the second one is designed to work robustly when noisy inequality constraints exist. We give convergence analysis for both methods under the assumption of bounded noise. We also present numerical experiments to give insights on the practical behavior of those methods, and to highlight the issues to be solved in a practical implementation (such as the termination criteria and the computation of quasi-Newton approximations). We motivate the problem setup studied in this work with an application example from robust design optimization. Another important application of our work is in the field of derivative-free optimization, when finite differences are employed to estimate gradients.

4 - A sharp augmented Lagrangian-based method for solving constrained DC optimization
Sona Taheri

In this talk, we will focus on constrained difference of convex (DC) optimization problems. Most existing methods for solving such problems do not guarantee finding feasible stationary points if the starting point is not feasible. We propose combining a method based on the sharp augmented Lagrangian with local techniques to solve constrained DC optimization problems. It is well-known that a subgradient of the dual function, formulated using sharp Lagrangians, increases the dual function, making these methods effective at quickly finding a feasible point. From there, a local method can be applied to compute a feasible stationary point. We develop a two-step method incorporating these approaches and study its convergence. Additionally, we compare its performance with a number of constrained nonsmooth optimization solvers using standard test problems.

■ WB-07

Wednesday, 10:30-12:30 - Room: B100/5015

Theory and methods for bilevel optimization

Stream: Bilevel and multilevel optimization
Invited session
Chair: *Kuang Bai*

1 - Stability analysis for two-level value functions and application to numerically solve a pessimistic bilevel program
Alain Zemkoho

We present some stability results for a two-level value function, which is the optimal value function of a parametric optimization problem constrained by the optimal solution set of another parameteric optimization problem. We then show how to use these stability results to write down (and subsequently compute) the necessary optimality conditions of a pessimistic bilevel optimization problem. We then demonstrate how the corresponding relaxation-based numerical process can be used to calculate local and global-type optimal solutions for the pessimistic bilevel program if one is equipped with a solver for minmax programs involving coupled inner constraints.

2 - Bilevel programming via set-valued optimization
Kuntal Som

Bilevel programming is one of the very active areas of research with many real-life applications in economics and engineering. Bilevel problems are hierarchical problems consisting of lower-level and upper-level problems, respectively. The leader or the decision-maker for the upper-level problem decides first, and then the follower or the lower-level decision-maker chooses his/her strategy. In the case of multiple lower-level solutions, the bilevel problems are not well defined, and there are many ways to handle such a situation. One standard way is to put restrictions on the lower level problems (like strict convexity) so that nonuniqueness does not arise. However, those restrictions are not viable in many situations. Therefore, there are two standard formulations, called pessimistic formulations and optimistic formulations of the upper-level problem (see Dempe (Foundations of Bilevel Programming. Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Dordrecht, 2002)). A set-valued formulation has been proposed in the the literature. However, the study is limited to the continuous set-up with the assumption of value attainment, and the general case has not been considered. In this talk, we focus on the general case and study the connection among various notions of solution. Our main findings suggest that the set-valued formulation may not hold any bigger advantage than the existing optimistic and pessimistic formulation.

3 - Lifting and Penalisation approach for Mathematical Program with Vanishing Constraints
Anand Kumar, K. S. Mallikarjuna Rao

In this work, we study infinite dimensional mathematical program with vanishing constraints (MPVC) as well as mathematical programs with complementary constraints. One of the major difficulty of MPVC and MPCC is that standard constraint qualifications are violated. In the literature, there are many relaxation studied. In this work, we propose a new relaxation scheme which considers a penalized problem in higher dimensions. Using this approach, we provide convergence results along with characterization of stationary points under appropriate assumptions.

4 - Directional combined approach for bilevel programs
Kuang Bai

In this paper, by employing the combined approach, we reformulate the bilevel programs into an equivalent set-constrained single-level problem, which however is nonsmooth. We introduce the directional partial calmness condition to penalize nonsmooth constraints into the objective function, then derive the directional necessary optimality condition under the directional metric subregularity/quasi-normality conditions. The directional partial calmness condition is generally weaker than the nondirectional one and thus more likely to hold. Furthermore, we establish sufficient conditions for directional partial calmness. An example is provided to illustrate that the directional partial calmness condition is satisfied for the bilevel programs, whereas the nondirectional condition does not hold, thereby validating the sufficiency of the conditions we have presented.

■ WB-08

Wednesday, 10:30-12:30 - Room: B100/7007

Theoretical advances in nonconvex optimization

Stream: Large scale optimization: methods and algorithms
Invited session
Chair: *Annette Dumas*
Chair: *Clément Royer*

1 - Continuized Nesterov Acceleration to improve convergence speed in non convex optimization

Julien Hermant, Jean-François Aujol, Charles Dossal, Aude Rondepierre

In the realm of smooth and convex functions, it is well known that in many scenarios, the Nesterov Accelerated Gradient (NAG) algorithm converges to the minimum significantly faster than Gradient Descent. Dropping the convexity assumption, this statement of convergence acceleration is challenged. We show that a variant, the continuized version of (NAG), introduced in [1], offers the opportunity to achieve new convergence results in settings where non convexity hinders the traditional NAG algorithm.

[1] E.Even,R.Berthier el al., "A Continuized View on Nesterov Acceleration for Stochastic Gradient Descent and Randomized Gossip"

2 - Complexity of Newton-type methods with quadratic regularization for nonlinear least squares

Iskander Legheraba, Clément Royer

Equipping globalized versions of Newton's method with complexity guarantees in a nonconvex setting has been a popular topic of research in recent years, motivated in part by the development of cubic regularization methods. In particular, quadratic regularization schemes can be adapted to achieve optimal iteration complexity bounds on general nonconvex problems.

In this talk, we describe a novel quadratic regularization approach where the quadratic subproblems are solved using a capped conjugate gradient technique. Such inexact steps lead to best-known operation and iteration complexity bounds, in both a first- and a second-order sense. By specializing our results to nonlinear least squares problems, we provide new complexity bounds that improve over existing results, through a reasoning that can be extended to other algorithms such as cubic regularization. We finally discuss the practical relevance of our approach on matrix least squares problems.

3 - Cubic regularized Newton methods with stochastic Hessian evaluations and momentum-based variance reduction

Yiming Yang

Second-order methods have received considerable attention in recent years, especially as many of them have been developed for efficiently achieving second-order stationary points in nonconvex optimization problems. However, second-order information, i.e., the Hessian matrix, could be computationally cost to evaluate for many real-world applications, which limits its practical utility. In this work, we introduce a new variant of the cubic regularized Newton method, which does not require exact evaluation of the Hessian but relaxes it to stochastic Hessian evaluations. This relaxation reduces the per-iteration costs of second-order methods and allows for greater flexibility in incorporating second-order information into optimization algorithm design. We establish the worst-case iteration complexity for our proposed cubic regularized Newton method with stochastic Hessian evaluations for finding approximate second-order stationary points. In addition, we show that, similar to the vanilla cubic regularized Newton method, our proposed variant can still achieve better complexity than first-order methods.

4 - Algorithms for nonconvex optimization on measure spaces.

Annette Dumas, Clément Royer

Super-resolution is a common task in medical imaging or microscopy, that consists in reconstructing a signal from a vector of measurements. This typically ill-posed problem can be formulated as the reconstruction of a sparse measure, yielding an optimization problem over a measure space. In this talk, we focus on super-resolution problems where the measurements correspond to Fourier coefficients up to a given frequency. We compare several algorithms for solving this problem, based on conditional gradient and projected gradient techniques, from both a numerical and theoretical perspective.

■ WB-09

Wednesday, 10:30-12:30 - Room: B100/8013

Variational Analysis III

Stream: Variational analysis: theory and algorithms
Invited session
Chair: *Francisco Javier Aragón Artacho*

1 - Geometry and the complexity of first-order methods for Lipschitz optimization

Adrian Lewis

We study first-order optimization algorithms from three geometric viewpoints. We first consider the simple idea of "slope" - a purely metric notion - and its role in popular convergence analyses based on the Kurdyka-Lojasiewicz inequality. We next introduce a measure of nonconvexity, and investigate its influence on the dimension-independent complexity of recent algorithms (Zhang et al, 2020) for optimizing objectives - common in machine learning - that may not be even weakly convex. Finally, we introduce a new local growth condition, inspired by a classical idealized iteration (Goldstein, 1977) for Lipschitz optimization, and explore its impact on near-linear local convergence.

Joint work with Siyu Kong and Tonghua Tian

2 - Inner approximations of convex sets and intersections of projectionally exposed cones
Vera Roshchina

we develop a technique for constructing arbitrarily tight inner convex approximations of compact convex sets with desired facial structure. These inner approximations have the property that all proper faces are extreme points, with the exception of a specific exposed face of the original set. We use this construction to prove that there exists a pair of projectionally exposed cones in dimension 5 whose intersection is not projectionally exposed (providing a negative solution to a couple of conjectures in conic geometry).

The talk is based on joint work with Bruno Lourenço and James Saunderson.

3 - Nonlinear Separation of Coradiant Sets: Optimality Conditions for Approximate Solutions
Miguel Angel Melguizo Padial, Fernando García Castaño

This talk investigates ϵ-efficiency and proper ϵ-efficiency in vector optimization via coradiant sets. We present a novel nonlinear separation theorem for coradiant sets in normed spaces, without convexity assumptions. Using this, we derive scalarization results that yield necessary and sufficient conditions for both ϵ-efficiency and proper ϵ-efficiency.

4 - Nonsmooth optimization techniques for computing projected quasi-equilibria
Giancarlo Bigi

Projected solutions to a quasiequilibrium problem allow overcoming the possible lack of (standard) solutions when the constraining set-valued map is not a self-map. This paper aims at providing a descent algorithm for computing projected solutions by relying on a reformulation of the problem asa nonsmooth optimization problem. The nonsmoothness of the gap function can be dealt with successfully through the nonexpansiveness of the projectionand tools such as Clarke subdifferentials. Nonetheless, some additional difficulties arise since the projection brings in nonsmoothness also in constraints that are provided by differentiable bifunctions. Monotonicity assumptions on the constraints have to cope with this further issue both to devise the algorithm and prove its convergence. Preliminary numerical tests show a promising behaviour of the algorithm.

The talk is based on joint papers with Marco Castellani and Sara Latini

---

■ WB-10

Wednesday, 10:30-12:30 - Room: B100/8011

Global Multi-Objective Optimization

Stream: Multiobjective and Vector Optimization
Invited session
Chair: *Pierluigi Mansueto*

1 - An unconstrained method for nonconvex Pareto front approximation
Ina Lammel, Karl-Heinz Küfer

In multi-objective optimization, approximating a Pareto front is a common task to help a decision-maker understand his or her options. Most scalarization-based approximation schemes for nonconvex Pareto fronts use constraints on the feasible set to guide the next solution to be computed to a particular region of the Pareto front. However, many powerful specialized solvers cannot handle constraints. We propose a new algorithm for the approximation of nonconvex Pareto fronts that only solves unconstrained optimization problems if the underlying multi-objective optimization problem is unconstrained. The algorithm uses a utility function-based scalarization method that incorporates penalty functions to approximate constraints. We show the completeness of the approach, i.e. every Pareto point can be the solution of a scalarization problem. The approximation algorithm is based on approximated boxes. We provide conditions under which a solution computed with approximated constraints actually fulfills the constraints and discuss the implications for the approximation algorithm and the use of domination cones to bound trade-offs. We illustrate the performance of the algorithm with numerical and practical examples and compare it to a box-based algorithm that uses a Pascoletti-Serafini scalarization approach.

2 - Local Upper Bounds for General Ordering Cones
Gabriele Eichfelder, Firdevs Ulus

The concept of local upper bounds plays an important role for numerical algorithms in nonconvex, integer, and mixed-integer multiobjective optimization with respect to the componentwise partial ordering, that is, where the ordering cone is the nonnegative orthant. In this talk, we answer the question on whether and how this concept can be extended to arbitrary ordering cones. This question has frequently arisen after talks, and this presentation aims to provide a comprehensive answer. We define local upper bounds with respect to a closed pointed solid convex cone and study their properties. We show that for special polyhedral ordering cones the concept of local upper bounds can be as practical as it is for the nonnegative orthant.

3 - Multi-Objective branch-and-bound algorithms for binary (non-convex) quadratic problems
Yue Zhang, Marianna De Santis, Pierre Fouilhoux, Lucas Létocart

We propose the first exact generic branch-and-bound framework for Multi-Objective Binary Quadratic programs (MObBQ). The MObBQ framework explores the decision space dividing the original problem into subproblems by variable fixing. The efficiency of the MObBQ approach is to repeatedly solve the easier relaxed problems and enumerate the complete non-dominated set by branching. Such a fully implicit enumeration stops until all nodes/subproblems are evaluated, where constructing valid lower and upper bound sets remains an essential ingredient. In this work, we propose different strategies for computing lower bound sets using the Quadratic Convex Reformulation techniques. The use of a preprocessing phase allows a significant time saving. We prove both theoretically and numerically the validity of our algorithm. Tests on bi- and tri-objectives instances from different multi-objective combinatorial quadratic problems show the effectiveness of the approach, able to outperform the epsilon-constraint method on bi-objective quadratic maxcut instances strongly.

■ WB-11

Wednesday, 10:30-12:30 - Room: B100/5017

Interior point methods and applications - Part I

Stream: Interior point methods and applications
Invited session
Chair: *Jordi Castro*

1 - **Proximal Augmented Lagrangian Regularization: A Primal-Dual Modular Interior-Point QP Solver.**
Jeremy Bertoncini, Alberto De Marchi, Matthias Gerdts
Quadratic Programming (QP) problems are fundamental in many fields such as optimization, machine learning or control theory. Our research introduces an interior point (IP) primal-dual solver that leverages both proximal and augmented Lagrangian regularization techniques, to efficiently address QP optimization problems. By combining the proximal regularization technique for higher convexity and the augmented Lagrangian method for enforcing inequality and equality constraints, the approach effectively balances computational efficiency with solution accuracy. Introduced in a suitable merit function, the convexified proximal sub-problem is iteratively solved using an inexact linesearch having conditions on both primal and dual variables. To further improve robustness, IP-slack variables are regularized ensuring a stable handling of the log-barrier term resulting into improved convergence properties. The solver integrates such methods to handle redundant linear constraints and achieve superlinear convergence properties. Additionally, the solver is object-oriented and modular, including multiple linesearch conditions, KKT-schemes, and termination strategies, allowing for greater flexibility and adaptability in different optimization scenarios. The numerical experiments demonstrate the solver's effectiveness in tackling the Maros-Mészáros Testset, showcasing its ability to achieve high precision even with ill-conditioned, redundant constraints and badly scaled problems.

2 - **Polynomial worst-case complexity for quasi-Newton Interior Point Methods**
Francisco Sobral, Jacek Gondzio
Quasi-Newton methods are well known techniques for large-scale numerical optimization. They use an approximation of the Hessian in optimization problems or the Jacobian in system of nonlinear equations. In the Interior Point context, quasi-Newton algorithms compute low-rank updates of the matrix associated with the Newton systems, instead of computing it from scratch at every iteration. In this work, we show that a simplified quasi-Newton primal-dual interior point algorithm for linear programming enjoys polynomial worst-case iteration complexity. Feasible and infeasible cases of the algorithm are considered and the most common neighborhoods of the central path are analyzed. Quasi-Newton updates are very attractive for large-scale optimization problems where the cost of factorizing the matrices is much higher than the cost of solving linear systems.

3 - **Proximal Stabilized Interior Point Methods and Applications**
Stefano Cipolla, Fabio Durastante, Jacek Gondzio, Beatrice Meini, Filippo Zanetti
In this talk, we present recent advances in the use of Proximal-Stabilization techniques within Interior Point Methods, highlighting both theoretical developments and practical applications across a range of optimization problems. The talk is based on (chronological order): [1] S. Cipolla and J. Gondzio. "Proximal Stabilized Interior Point Methods and Low-Frequency-Update Preconditioning Techniques". In: J. Optim. Theory Appl. (2023) [2] S. Cipolla, J. Gondzio, and F. Zanetti. "A regularized Interior Point Method for sparse Optimal Transport on Graphs". In: European J. Oper. Res. (2024) [3] S. Cipolla and J. Gondzio. "Proximal-Stabilized Semidefinite Programming". In: Comput Optim Appl (2024) [4] S. Cipolla, F. Durastante, and B. Meini. Enforcing Katz and PageRank Centrality Measures in Complex Networks. 2025. arXiv: 2409.02524

4 - **Comparison of Interior-point and active set quadratic programming for regularised column generation**
Wim Vanroose, Sebastian Van Thienen
Regularized column generation is equivalent to a bundle method that maximizes the dual objective. In each iteration, a cut/column is added, and the solution is updated by solving a quadratic programming problem. In this talk, we compare the performance of a warm-started interior point method, where multiple columns are added simultaneously, with a hot-started active-set QP approach, where columns are added one by one while maintaining a factorization of the KKT matrix using rank-one updates. We discuss the challenges in high-performance numerical linear algebra and present numerical results for large-scale planning problems in the airline industry.

■ WB-12

Wednesday, 10:30-12:30 - Room: B100/8009

Chemical and Energy Systems Optimization

Stream: Applications: AI, uncertainty management and sustainability
Invited session
Chair: *Gabriel Patron*
Chair: *Lavinia Ghilardi*

1 - **Complementarity-based modelling for sustainable dynamic gas network operation**
Lavinia Ghilardi
In the framework of the energy transition, optimization tools play a key tole in supporting decision-making for sustainable process operations. However, modelling process systems often involves nonlinearities along with nonsmooth or switching conditions, due to the underlying system physics and the presence of control components. At the same time, these applications require solutions to be obtained in fast computational time to enable effective real-time decisions.

In this work, we propose a modelling approach for optimizing gas networks operations to minimize CO2-eq emissions related to compressor energy consumption. The model includes nonlinear dynamic conservation equations in pipelines, compressor maps, and valves operation. We formulate the problem as a Mathematical Program with Complementarity Constraints (MPCC), where complementarity constraints and reformulations enable the modelling of nonsmooth or disjuctive operations while retaining a continuous nonlinear framework. In gas pipelines operations, these methods can be used to model flow reversals, hydrogen mixing, or control valves operations.

The proposed methodology is effectively applied to large-scale problems with branched-looped networks. The approach allows to retain local solution in fast computational times, enabling its application in real-time controllers.

2 - Risk-constrained two-stage demand response scheduling for green hydrogen production
Gabriel Patron

Hydrogen is increasingly being incorporated into global net zero plans with a particular emphasis on electrolysis-based "green" hydrogen. When integrated with the power grid, the demand response scheduling of a hydrogen production plant can be posed as a cost minimization problem that relies on having predicted electricity price signals as inputs. In reality, these price signals are often subject to considerable uncertainty and, especially when participating in several power markets (e.g., day-ahead, intraday), the performance of a deterministic scheduling solution can be significantly suboptimal. In this presentation, we propose a stochastic risk-aware formulation to determine the optimal hydrogen production schedule in a multi-market context with uncertain price signals. A multi-stage formulation allows for recourse actions upon price perturbations and prediction errors, while risk-aware objective functions such as conditional value-at-risk (CVaR) can minimize shortfall. Further, the proposed approach determines the proportion of participation in each power market.

3 - Distillation column optimization: A formal method using stage-to stage computations and distributed streams
Tobias Seidel

Distillation in distillation columns is a widely used process to separate liquid mixtures and plays a key role in the energy-intensive sector of chemical processes. To minimize the required energy for separation, an optimal design and optimal operation are needed. Using the MESH (mass, equilibrium, summation, heat) equations in an equilibrium stage model results in a highly nonlinear description with hundreds to thousands of optimization variables and equality constraints. Including the design in the optimization introduces discrete degrees of freedom related to the number of stages, further complicating the problem as a nonlinear mixed-integer optimization problem. To simplify the problem, we employ distributed streams, eliminating discrete degrees and replacing them with continuous ones. To avoid sophisticated initialization procedures, we combine this smooth formulation with stage-to-stage calculations that presolve most of the nonlinear equations, reducing the problem to maintaining only the MESH equations for a single stage. For this nonlinear reduction, we can prove the existence and uniqueness of solutions, ensuring a well-defined formulation and equivalence of the optimization problems. Our numerical experiments demonstrate the efficiency and stability of solving the proposed optimization problem in various scenarios, including single and multiple distillation columns.

Wednesday, 14:00-16:00

■ WC-01

Wednesday, 14:00-16:00 - Room: B100/1001

Advances in Multiobjective and Bilevel Optimization without Derivatives

Stream: Zeroth and first-order optimization methods
Invited session
Chair: *Francesco Rinaldi*
Chair: *Andrea Cristofari*

1 - A Direct Multisearch Approach (DMS) for Many-Objective Derivative-Free Optimization
Everton Silva, Ana Luisa Custodio

DMS was originally developed for continuous constrained multiobjective derivative-free optimization, with a general number of objective function components. Still, its performance was never tested for problems with more than three objectives. We propose DMS-Reduction, a variant that uses reduction methods, such as sketching techniques and correlation approaches, to reduce the number of objective function components and related variables to be addressed at each iteration. We detail the algorithmic structure and report promising numerical results in tackling many-objective optimization problems.

2 - Exploring Polynomial Models in the Search Step of Direct Multisearch
Marta Pozzi, Everton Silva, Ana Luisa Custodio

Direct Multisearch (DMS) is a direct search method suited for multiobjective derivative-free optimization. Its structure is organized in an optional search and a poll step, the latter responsible for convergence.

Quadratic models were recently introduced in the definition of the search step, improving the numerical performance of the algorithm. In this work, new strategies that make use of these models are explored and evaluated.

The MinMax scalarization approach is compared with the Epsilon and NBI methods, and with a strategy that uses the recently released Improved Steepest Descent algorithm.

3 - Derivative-Free Bilevel Optimization with Inexact Lower-Level Solutions
Edoardo Cesaroni, Giampaolo Liuzzi, Stefano Lucidi

In this work, we introduce derivative-free frameworks for bilevel optimization. We consider both the upper and lower-level problems with bound constraints on the variables, as well as general nonlinear constraints, assuming that first-order information is not available or it is impractical to obtain. The lower-level problem is solved with an accuracy that is progressively refined throughout the optimization process. Initially, we analyze the case where the upper-level problem is subject only to bound constraints, establishing convergence to Clarke-Jahn stationary points when the refinement process is allowed to reach its maximum precision. When a stricter limit is imposed on accuracy, we prove convergence to approximate stationary points using an extended notion of Goldstein stationarity. Finally, we extend the proposed approach to handle more complex constraints via an exact penalty function approach, proving convergence to stationary points under suitable assumptions.

■ WC-02

Wednesday, 14:00-16:00 - Room: B100/7011

Systematic and computer-aided analyses VI: Systematic approaches to the analyses of proximal and higher-order methods

Stream: Systematic and computer-aided analyses of optimization algorithms
Invited session
Chair: *Manu Upadhyaya*
Chair: *Baptiste Goujaud*

1 - SDP performance estimation for the boosted DCA
Etienne De Klerk, Hadi Abbaszadehpeivasti

The difference-of-convex algorithm (DCA) is a well-established nonlinear programming technique that solves successive convex optimization problems. These sub-problems are obtained from the difference-of-convex (DC) decompositions of the objective and constraint functions. We investigate the worst-case performance of the unconstrained DCA, with and without boosting, where boosting simply performs an additional step in the direction generated by the usual DCA method. We show that, for certain classes of DC decompositions, the boosted DCA is provably better in the worst-case than the usual DCA. The proof technique relies on semidefinite programming (SDP) performance estimation, as introduced by Drori and Teboulle. This talk is based on joint work with Hadi Abbaszadehpeivasti.

2 - A unified Lyapunov analysis for first and second order methods attached to bilevel optimization
David Alexander Hulett, Radu Ioan Bot, Enis Chenchene, Robert Csetnek

In a real Hilbert space, we address a convex bilevel optimization problem, where both the inner and outer levels have a composite smooth+nonsmooth structure. We study two algorithms, which at each step execute a proximal gradient step on a dynamically regularized objective function: one without momentum, in the spirit of the Bi-SG algorithm introduced by Merchav and Sabach in 2023, and one with momentum, in the spirit of the FBi-PG algorithm introduced by Merchav, Sabach and Teboulle in 2024. Attached to these algorithms we formulate a family of Lyapunov energy functionals which satisfy a certain descent property, and which allow us, in a streamlined way, to show convergence rates statements which match or surpass those already present in the literature. When the inner function satisfies a Hölderian error bound, we achieve small o rates for the inner and outer functional values along the last iterate, which is a first for this class of problems; additionally, the iterates converge weakly to a solution of the bilevel problem. These algorithms can be seen as forward discretizations of their continuous-time counterparts, which are respectively a first- and a second-order differential equation. The simpler, continuous-time Lyapunov analysis provides a guide and suggests the inequalities which should be reached in the discrete-time case.

3 - Designing Monotone Operator Splitting Algorithms with Steering Vectors
Max Nilsson, Sebastian Banert, Pontus Giselsson

In this work, we study iterative algorithms for solving the monotone inclusion problem involving the sum of m maximally monotone operators over a Hilbert space. We take a novel approach by deriving algorithms directly from Lyapunov analysis, reversing the conventional workflow of algorithm research. Our methodology is based on iteratively simplifying a Fejér monotonic quadratic inequality, which we reformulate using an LDLT factorization. This transforms the problem of algorithm design into an inertia eigenvalue problem, allowing us to characterize and construct provably convergent schemes. Our framework provides insights into existing resolvent-splitting methods and the underlying structure of their convergence.

■ WC-03
Wednesday, 14:00-16:00 - Room: B100/4011

Acceleration Methods in Optimization

Stream: Large scale optimization: methods and algorithms
Invited session
Chair: *Vuong Phan*
Chair: *Yingxin Zhou*

1 - Anderson acceleration with adaptive relaxation for convergent fixed-point iterations
Nicolas Lepage-Saucier

Two adaptive relaxation strategies are proposed for Anderson acceleration. They are specifically designed for applications in which mappings converge to a fixed point. Their superiority over alternative Anderson acceleration is demonstrated for linear contraction mappings. Both strategies perform well in three nonlinear fixed-point applications that include partial differential equations and the EM algorithm. One strategy surpasses all other Anderson acceleration implementations tested in terms of computation time across various specifications, including composite Anderson acceleration.

2 - Anderson Acceleration for Primal-Dual Hybrid Gradient
Yingxin Zhou, Stefano Cipolla, Vuong Phan

The use of first-order optimization methods has gained increasing attention for solving large-scale linear programming (LP) problems due to their low iteration computational cost. However, a drawback of such methods is their relatively slow convergence when high-accuracy solutions are required. To overcome this disadvantage, in this talk, we consider solving large-scale LP problems using the first-order method known as Primal-Dual Hybrid Gradient (PDHG), combined with an acceleration technique. Specifically, we formulate the PDHG method as a fixed-point iteration and apply Anderson Acceleration (AA) method to enhance its convergence speed. Unlike standard fixed-point schemes, we leverage the projection operator to ensure that the iterates generated by AA remain within the feasible domain of the operator, thereby avoiding the issue where the next iterate may violate the constraints of the LP problem. Under the assumption that the fixed-point operator is non-expansive, we establish a global convergence guarantee for the proposed algorithm. Numerical results will showcase how using of Anderson Acceleration in this framework allows to consistently improve the robustness and efficiency of the baseline PDHG. This talk is based on Y. Zhou, S. Cipolla, V. Phan. "Anderson Acceleration for Primal-Dual Hybrid Gradient."

3 - Accelerating Convergence of MPGP Algorithm
Jakub Kruzik, David Horak

Modified proportioning with gradient projections (MPGP) is an active set method that employs the conjugate gradient method to minimize a quadratic objective function on the free set, gradient projection to expand the active set, and proportioning to reduce the active set. Preconditioners are widely used to accelerate the solution of systems of linear equations. However, applying preconditioners in conjugate gradient-based algorithms for constrained quadratic programming is not straightforward. This is because preconditioning transforms variables, thereby converting constraints into more general forms, such as turning bound constraints into linear inequality constraints. Preconditioning applied to the submatrix of the Hessian corresponding to the free set (preconditioning in face) is a successful strategy that does not transform the constrained variables. However, a major drawback is that the preconditioner must be recomputed every time the free set changes. To overcome this limitation, we propose a new strategy for integrating preconditioning into MPGP. Our approach avoids both the transformation of constrained variables and the need to recompute the preconditioner based on the free set. This new preconditioning strategy is particularly effective when combined with alternative techniques for fast expansion of the active set. The speedup of the resulting algorithm will be demonstrated on a series of benchmarks using PERMON, an open-source library for quadratic programming.

■ WC-04
Wednesday, 14:00-16:00 - Room: B100/5013

Large Scale Optimization for Statistical Learning

Stream: Optimization for machine learning
Invited session
Chair: *Selin Ahipasaoglu*

1 - Spatial branch-and-bound methods for solving the k-Hyperplane clustering

Stefano Coniglio, Montree Jaidee

We address the k-Hyperplane Clustering problem, which consists in finding k hyperplanes that minimize the sum, over all data points, of squared Euclidean (2-norm) distances to their closest hyperplanes. We propose two algorithms that solve this problem to global optimality, both grounded in spatial-branch-and-bound (SBB) techniques. In the first algorithm, we strengthen a basic mixed integer quadratically-constrained quadratic-programming (MIQCQP) formulation with additional constraints derived from p-norm formulations of the problem defined for values of p different from 2. The second algorithm introduces a problem-specific branching scheme, where the infeasible regions of the problem (complements of Euclidean balls) are approximated by a polyhedron whose vertex set becomes increasingly rich as branching progresses. Experimental results demonstrate that both approaches substantially improve the problem's solvability to global optimality.

2 - Solving the Optimal Experiment Design with Mixed-Integer Convex Methods

Deborah Hendrych, Mathieu Besançon, Sebastian Pokutta

We tackle the Optimal Experiment Design Problem, which consists of choosing experiments to run or observations to select from a finite set to estimate the parameters of a system. The objective is to maximize some measure of information gained about the system from the observations, leading to a convex integer optimization problem. We leverage Boscia.jl, a recent algorithmic framework, which is based on a nonlinear branch-and-bound algorithm with node relaxations solved to approximate optimality using Frank-Wolfe algorithms. One particular advantage of the method is its efficient utilization of the polytope formed by the original constraints which is preserved by the method, unlike alternative methods relying on epigraph-based formulations. We assess our method against both generic and specialized convex mixed-integer approaches. Computational results highlight the performance of our proposed method, especially on large and challenging instances.

3 - A column generation approach to exact experimental design

Selin Ahipasaoglu, Stefano Cipolla, Jacek Gondzio

We propose an algorithm for the exact optimal experimental design problem under Kiefer's criteria. Our method first employs column generation to solve the continuous relaxation of the problem quickly. The support of this solution is used to construct a feasible solution that is provably close to the optimal design. We demonstrate that for large-scale problems where the number of regression points is significantly larger than the number of experiments, this approach is preferable over existing algorithms.

■ WC-05

Wednesday, 14:00-16:00 - Room: B100/4013

Recent Advances in Stochastic Optimization

Stream: Optimization for machine learning
Invited session
Chair: *Chuan He*

1 - Almost sure convergence rates for stochastic gradient methods

Simon Weissmann

In this talk, we present recent advances in establishing almost sure convergence rates for stochastic gradient methods. Stochastic gradient methods are among the most important algorithms in training machine learning problems. While classical assumptions such as strong convexity allow a simple analysis, they are rarely satisfied in applications. In recent years, global and local gradient domination properties have shown to be a more realistic replacement of strong convexity. They were proved to hold in diverse settings such as (simple) policy gradient methods in reinforcement learning and training of deep neural networks with analytic activation functions. We prove almost sure convergence rates $f(X_n)-f*in obig(n-frac14beta-1+epsilonbig)$ of the last iterate for stochastic gradient descent (with and without momentum) under global and local $beta$-gradient domination assumptions. The almost sure rates get arbitrarily close to recent rates in expectation. Finally, we demonstrate how to apply our results to the training task in both supervised and reinforcement learning.

2 - A Hessian-Aware Stochastic Differential Equation for Modelling SGD

Xiang Li

Understanding how Stochastic Gradient Descent (SGD) escapes from stationary points is essential for advancing optimization in machine learning, particularly in non-convex settings. While continuous-time approximations using stochastic differential equations (SDEs) have been instrumental in analyzing such behavior, existing models fall short in accurately capturing SGD dynamics—even for simple loss landscapes. In this work, we introduce a novel SDE model derived through a stochastic backward error analysis framework. This new formulation incorporates second-order information from the objective function into both its drift and diffusion components, leading to a more faithful representation of SGD's behavior.

Our new SDE improves the theoretical weak approximation error among existing models, reducing the dependence on the smoothness parameter. Importantly, we demonstrate that, for quadratic objectives, our model is the first to exactly replicate the distributions of SGD iterates.

Empirical evaluations on neural network loss surfaces further validate the practical advantages of our SDE. Additionally, the improved approximation allows for a better analysis of the escape time of SGD near stationary points.

3 - Complexity guarantees for risk-neutral generalized Nash equilibrium problems

Meggie Marschner, Mathias Staudigl

In this paper we address stochastic generalized Nash equilibrium problem (SGNEP) seeking with risk-neutral agents. In this work, the stochastic variance-reduced gradient (SVRG) technique is modified to contend with general sample space a stochastic forward-backward-forward splitting scheme with variance reduction (DVRSFBF) is proposed for resolving structured monotone inclusion problems. In DVRSFBF, the mini-batch gradient estimator is computed periodically in the outer loop, while only cheap sampling is required in the frequently activated inner loop, thus achieving significant speedups when sampling costs cannot be overlooked. The algorithm is fully distributed and it guarantees almost sure convergence under appropriate batch size and strong monotonicity assumptions. Moreover, it exhibits a linear rate with possible biased estimators, which is rather mild and imposed in many simulation-based optimization schemes. A numerical study on a class of networked Cournot games reflects the performance of our proposed algorithm.

4 - Stochastic first-order methods can leverage arbitrarily higher-order smoothness for acceleration
Chuan He

Stochastic first-order optimization methods play a crucial role in modern artificial intelligence. From a theoretical perspective, worst-case sample complexity is an important measure of the computational cost of stochastic algorithms. In this talk, I will introduce a new stochastic first-order method with multi-extrapolated momentum to leverage the Lipschitz continuity of arbitrarily high-order derivatives for acceleration. Surprisingly, this work highlights that higher-order smoothness can play an important role in the analysis of stochastic first-order methods.

■ WC-06

Wednesday, 14:00-16:00 - Room: B100/7013

Structured nonsmooth optimization – Part III

Stream: Nonsmooth and nonconvex optimization
Invited session
Chair: *Bastien Massion*
Chair: *Alireza Kabgani*

1 - Linear Convergence Rate in Convex Setup is Possible! Gradient Descent Method Variants under (L0,L1)-Smoothness
Aleksandr Lobanov

The gradient descent (GD) method – is a fundamental and likely the most popular optimization algorithm in machine learning (ML), with a history traced back to a paper in 1847 (Cauchy, 1847). In this paper, we provide an improved convergence analysis of gradient descent and its variants, assuming generalized smoothness (L0,L1). In particular, we show that GD has the following behavior of convergence in the convex setup: At first, the algorithm has linear convergence, and approaching the solution, has standard sublinear rate. Moreover, we show that this behavior of convergence is also common for its variants using different types of oracle: Normalized Gradient Descent as well as Clipped Gradient Descent (the case when the oracle has access to the full gradient); Random Coordinate Descent (when the oracle has access only to the gradient component); Random Coordinate Descent with Order Oracle (when the oracle has access only to the comparison value of the objective function). In addition, we also analyze the behavior of convergence rate of GD algorithm in a strongly convex setup.

2 - Primal-Dual Coordinate Descent for Nonconvex-Nonconcave Saddle Point Problems Under the Weak MVI Assumption
Iyad Walwil, *Olivier Fercoq*

We introduce two novel primal-dual algorithms for tackling non-convex, non-concave, and non-smooth saddle point problems characterized by the weak Minty variational inequality (MVI). The first algorithm generalizes the well-known Primal-Dual Hybrid Gradient (PDHG) method to address this challenging problem class. The second algorithm introduces a randomly extrapolated primal-dual coordinate descent approach, extending the Stochastic Primal-Dual Hybrid Gradient (SPDHG) algorithm. Designing a coordinated algorithm to solve non-convex, non-concave saddle point problems is unprecedented, and proving its convergence posed significant difficulties. This challenge motivated us to utilize PEPit, a Python-based tool for computer-assisted worst-case analysis of first-order optimization methods. By integrating PEPit with automated Lyapunov function techniques, we successfully derived our second novel algorithm. Both methods are effective under a mild condition on the weak-MVI parameter, achieving convergence with constant step sizes that adapt to the problem's structure. Numerical experiments on sigmoid and perceptron-regression problems validate our theoretical findings.

3 - A full splitting algorithm for structured difference-of-convex programs
Rossen Nenov

We address a class of nonconvex and nonsmooth problems in which the objective function is formed as the sum of a smooth function and the difference of two convex function composed with different linear operators. This structure gives rise to challenges from both nonsmoothness and nonconvexity. To tackle this, we introduce an adaptive double-proximal, full-splitting algorithm that separates the linear mappings from the nonsmooth terms through a moving center technique in the final subproblem. Our analysis establishes that the iterates converge subsequentially to an approximate stationary point and, under the Kurdyka-Łojasiewicz property, achieve global convergence. We also provide a counterexample demonstrating that the search for exact solutions can lead to divergence, underscoring the necessity of an approximate approach.

This is joint work with Radu Bot and Min Tao.

4 - An accelerated projected gradient descent for Grassmannian frame computation
Bastien Massion, *Estelle Massart*

This paper addresses the approximation of real and complex Grassmannian frames, namely sets of unit-norm vectors with minimum mutual coherence. We recast this problem as a collection of feasibility problems aiming to design frames with given target coherence, that evolves during the execution of the algorithm. The feasibility problems are solved by an accelerated alternating projection algorithm inspired by accelerated proximal methods, leveraging a Gram matrix representation of the frames. Numerical experiments indicate that our proposed Targeted coherence with Accelerated Alternating Projection (TAAP) algorithm outperforms state-of-the-art methods regarding the mutual coherence vs computational cost criterion, exhibiting the largest performance gap with existing methods when the frame dimension is comparable to the dimension of the ambient space.

■ WC-07

Wednesday, 14:00-16:00 - Room: B100/5015

Numerical Methods and Applications II

Stream: Numerical Methods and Applications
Invited session
Chair: *Brendan Ames*

1 - Extensions of Consensus-Based Methods
Sara Veneruso, Michael Herty, Lorenzo Pareschi

Solving non-convex minimization problems using multi-particle metaheuristic, derivative-free optimization methods has recently gained a lot of interest. Popular approaches, such as Consensus-Based Optimization and Particle Swarm Optimization, iteratively update a population of particles based on dynamics inspired by social interactions. Due to their tractable analytic structure, these methods can be adapted to different scenarios, including constrained optimization problems. Furthermore, it is possible to exploit the hierarchical structure of these methods to introduce a micro-macro decomposition of them. A first approach is to write the probability density of particles as a convex combination of microscopic and macroscopic contributions, and then to evolve both parts separately. An alternative approach is to study the marginals of the particle distribution. Several simulations are performed to show the validity of the extensions of the presented metaheuristic methods.

2 - A Hybrid Conjugate Gradient-Like Algorithm for Solving Nonlinear Equations and Image Restoration
Supak Phiangsungnoen

In this work, we present a hybrid spectral-conjugate gradient (SCG) method designed to efficiently solve nonlinear monotone operator equations. Our approach integrates a hybrid parameter that unifies key conjugate gradient methods, including Polak-Ribière-Polyak (PRP), Liu-Storey (LS), Fletcher-Reeves (FR), and conjugate descent (CD), as special cases. To ensure that the search direction satisfies the sufficient descent condition, we derive an appropriate spectral parameter. Additionally, the search direction is constructed to remain bounded, and under specific conditions, we establish the convergence of the sequence generated by our hybrid SCG algorithm. To demonstrate its effectiveness, we conduct extensive numerical experiments comparing the proposed method with existing algorithms. These experiments, based on benchmark nonlinear monotone operator equations, highlight the superior efficiency and practical potential of our approach.

3 - Finding a planted clique hidden among many dense subgraphs
Brendan Ames

Much recent research has focused on sufficient conditions for recovery of planted cliques or, more generally, dense submatrices, within random graphs. Although the maximum clique problem is NP-Hard, these results suggest that we can efficiently identify the maximum clique under certain assumptions on the input graph. Unfortunately, the overwhelming majority of this research focuses on the case where the input graph consists of a single large clique hidden by diversionary nodes and edges. On the other hand, many recent analyses have established recovery guarantees for stochastic block models via various graph clustering heuristics. We build upon these earlier results and establish sufficient conditions for recovery of a single planted clique from the optimal solution of a certain convex optimization problem within a random graph containing many planted cliques obscured by diversionary edges and nodes. We also discuss strategies for solution of this optimization problem and extensions of these recovery guarantees to non-convex quadratic programming relaxations of the maximum clique and densest submatrix problems.

■ WC-08

Wednesday, 14:00-16:00 - Room: B100/7007

Advances in non-convex optimization

Stream: Optimization for machine learning
Invited session
Chair: *Radu-Alexandru Dragomir*

1 - Benign landscapes for synchronization on spheres via normalized Laplacian matrices
Andrew McRae

I will present new results on the nonconvex landscapes of synchronization problems on spheres. The key tool is a deterministic landscape condition extending a recent result of Rakoto Endor and Waldspurger. If a certain problem-dependent Laplacian matrix (potentially with diagonal preconditioner, e.g., in the normalized graph Laplacian) has small enough condition number, the nonconvex landscape is benign. Our extension, which is to use a preconditioner, gives tighter results for many problems. This has consequences for several applications. For a spherical relaxation of the problem of synchronization over the two-element group Z_2, we show that, for problem instances coming from several popular statistical models, relaxing to any higher-dimensional sphere (even the unit circle, which was proposed by Burer et al. for max-cut and by Bandeira et al. for synchronization) is tight, has a benign landscape, and gives exact statistical recovery throughout the asymptotic statistical regimes where exact recovery is possible. Second, we consider the global synchronization of networks of coupled oscillators under the (homogeneous) Kuramoto model. We prove new and optimal asymptotic results for certain random signed networks, and we give new and simple proofs for several existing state-of-the-art results. For certain network types, the Laplacian condition number result is optimal. This work appears in the preprint https://arxiv.org/abs/2503.18801.

2 - Diagonal linear networks and the regularization path of the LASSO
Raphael Berthier

Diagonal linear networks is a crude model of neural networks in which we can analyze rigorously the effect of the composition of layers on the optimization trajectory. In the overparametrized setting, diagonal linear networks enjoy an implicit regularization: they converge to the interpolator of the data with minimum l1 norm, even without any explicit regularization on the network weights. In this talk, I will discuss the implicit regularization of diagonal linear networks when early stopped. I will show that the full optimization trajectory of diagonal linear networks is closely related to the regularization path of the LASSO, where the iteration number plays the role of the inverse regularization parameter.

3 - Long-time convergence of a consensus-based optimization method
Victor Priser, Pascal Bianchi, Radu-Alexandru Dragomir

This paper studies a global optimization, derivative-free algorithm for a non-convex function f. In particular, the proposed algorithm is a variant of the Consensus-Based Optimization (CBO) algorithm. CBO is a particle-based algorithm that consists of two terms: a drift term attracting particles toward a consensus, modeling the local best results of the particles, and a noise term allowing particles to explore the domain. Unlike other existing approaches that focus on a finite time window, we are interested in the long-time convergence of the algorithm. The study of this algorithm is first conducted in the mean-field limit framework with an infinite number of particles, where we show the long-time convergence of the law of the particles to the Dirac measure centered at the minimizer. In the second step, we demonstrate the long-time convergence, for a finite number of particles, of the law of the particles to a set of measures which are concentrated around the minimizer of f . The strength of our result is that we control the distance of the law of the particles to the Dirac measure centered at the minimizer at any time during the algorithm, whereas other studies lose control of this distance at some time, as their analyses are conducted within a finite time window.

4 - Simplicity Bias of Two-Layer Networks beyond Linearly Separable Data
Nikola Konstantinov

Simplicity bias, the propensity of deep models to over-rely on simple features, has been identified as a potential reason for limited out-of-distribution generalization of neural networks(Shah et al., 2020). Despite the important implications, this phenomenon has been theoretically confirmed and characterized only under strong dataset assumptions, such as linear separability (Lyu et al., 2021). In this work, we characterize simplicity bias for general datasets in the context of two-layer neural networks initialized with small weights and trained with gradient flow. Specifically, we prove that in the early training phases, network features cluster around a few directions that do not depend on the size of the hidden layer. Furthermore, for datasets with an XOR-like pattern, we precisely identify the learned features and demonstrate that simplicity bias intensifies during later training stages. These results indicate that features learned in the middle stages of training may be more useful for OOD transfer. We support this hypothesis with experiments on image data.

■ WC-09

Wednesday, 14:00-16:00 - Room: B100/8013

Generalized convexity and monotonicity 4

Stream: Generalized convexity and monotonicity
Invited session
Chair: *Minh N. Dao*

1 - Capra-Cutting Plane Method Applied to Sparse Optimization
Seta Rakotomandimby, Adrien Le Franc, Jean-Philippe Chancelier, Michel De Lara

In 1960, Kelley proposed a cutting plane method to minimize a continuous convex objective function over a compact set. This iterative method consists at each step in minimizing a polyhedral approximation of the objective function before improving the approximation by a valid affine cut. In generalized convexity, the affine functions giving the cuts are replaced by some family of base functions. This base functions are chosen so that the objective function is their supremum, making it abstract convex. In this framework, the Kelley's algorithm has been generalized to continuous abstract convex functions. We continue the generalization of the cutting plane method by providing a convergence result that can be applied to lower semicontinuous objective functions. This convergence result is motivated by the Capra-convexity results on the 0 pseudonorm, which is lower semicontinuous. As explicit formulas for the Capra-subdifferential of the 0 pseudonorm have been calculated, we can now implement Capra-cutting plane methods for the sparse problem of minimizing 0 over a compact set.

2 - Evaluation of Multi-agent algorithms in Multi-objective Environments with Weighted Sum Method
Carlos Ignacio Hernández Castellanos, José Olivas Díaz

In many complex problems, collaboration among multiple agents is essential to achieve a common goal. This structure is commonly known as collaborative multi-agent systems. The rise of deep reinforcement learning (DRL) has allowed agents to go further in the environment where they can learn. However, while much of the current MARL research focuses on single-objective tasks, many real-world collaborative multi-agent systems (MAS) problems involve conflicting objectives. This gap drives our study of stochastic multi-objective optimization on MARL (MOMARL) configurations.

Recent literature on Deep Multi-Task Learning (MTL) has defended using weighted sum over more complex multi-task optimizers in supervised and reinforcement learning domains. On the other hand, other works advocate using specialized multi-objective optimization algorithms.

Our study extends this discussion to the multi-agent domain by empirically evaluating nine state-of-the-art MARL algorithms across three environments. Note that these algorithms can be seen as a special case of gradient-based stochastic optimization. This work aims to provide insights into the success of weighted sum on MTL and DRL. For instance, our empirical results show that the Pareto fronts we obtained are highly convex. Notably, extreme weights failed to produce efficient solutions in most instances. Future work will focus on providing a theoretical analysis of these observations and developing effective methods for MOMARL problems.

3 - Linear Convergence of the Proximal Point Algorithm Beyond Convexity
Felipe Lara

We study the fundamental properties of proximal operators for nonsmooth (strongly) quasar-convex functions [1], and we apply them to the proximal point algorithm (PPA). Our analysis focuses on minimizing strongly quasar-convex functions, where the solution is unique. We establish that the sequence generated by PPA is a minimizing sequence and converges linearly under strong quasar-convexity. Additionally, we provide a complexity analysis, proving that PPA achieves an (mathcalO(ln(1/varepsilon))) convergence rate for both the iterates and the function values. Finally, and if the time allow us, numerical examples will be also presented.

References:

1-. O. Hinder, A. Sidford, N. Sohoni, Near-optimal methods for minimizing star-convex functions and beyond. Proc. 33th Conf. on Lear. Theory, 125, 1894–1938, (2020).

2-. Y.E. Nesterov, B.T. Polyak, Cubic regularization of Newton method and its global performance, Math. Programm., 108, 177–205, (2006).

■ WC-10

Wednesday, 14:00-16:00 - Room: B100/8011

Computational Aspects in Multiobjective Optimization

Stream: Multiobjective and Vector Optimization
Invited session
Chair: *Felix Neussel*

1 - Stochastic approximation in convex multiobjective optimization
Elena Molho, Carlo Alberto De Bernardi, Enrico Miglierina, Jacopo Somaglia

Given a multiobjective optimization problem with N strictly convex objective functions, we focus on a minimizer obtained by a linear scalarization technique with a specific choice of the coefficients in the simplex. Our main result is to find an estimation of the averaged error that we make if we approximate that minimizer with the minimizers of the convex combinations of n functions randomly chosen among the original N objective functions with probabilities that coincide with the original scalarization coefficients and uniformly weighted by the reciprocal of n. We prove that the averaged error considered above converges to 0 uniformly with respect to the coefficients. The key tool in the proof of our stochastic approximation theorem is a geometrical property, called small diameter property, ensuring that the minimizer of a convex combination of the objective functions continuously depends on the coefficients of the convex combination.

2 - Benchmarking Nonlinear Multi-Objective Optimizers in Julia
Manuel Berkemeier

We present tools and techniques to benchmark solver software for multi-objective optimization. When trying to compare solvers, one is often faced with various difficulties: For example, the solvers might target different problem classes, or their interfaces differ significantly. Additionally, there is (luckily!) no single agreed-upon programming language for research code. We will restrict ourselves to optimizers for nonlinear multi-objective optimization, possibly with constraints, and leverage features of the Julia language to overcome some of the other difficulties. Thanks to shared library loading, we can dynamically use Fortran software that previously required static compilation. It is also possible to dynamically interact with Python or Matlab scripts. Lastly, we will show the results of benchmarking a list of select solvers on a curated set of test problems, which we provide as a Julia package.

3 - Parametrized convex MINLP: Warm-starting with Outer Approximation for Sequence of MINLPs
Erik Tamm, Gabriele Eichfelder, Jan Kronqvist

This work addresses the challenge of efficiently solving parametrized sequences of convex Mixed-Integer Nonlinear Programming (MINLP) problems through warm-starting techniques. While solving such sequences is known to be difficult, the general case of convex MINLPs has received limited attention in the literature. Our research introduces warm-starting for the outer approximation algorithm and explores its potential to enhance computational efficiency. This approach is particularly relevant for applications involving multiobjective MINLPs using scalarization techniques.

The main theoretical contribution of this paper is the identification of conditions under which warm-starting significantly improves the performance of solving these problem sequences. We support our theoretical findings with the implementation of three proposed warm-starting rules, demonstrating noticeable performance improvements. Our methods are especially effective for problems where the optimal integer part remains optimal for several problems in the sequence.

These results highlight the potential for further research to enhance the computational efficiency of solving parametrized convex MINLPs by developing warm-starting techniques.

4 - On image space transformations in multiobjective optimization
Felix Neussel, Oliver Stein

It is well known that one can apply a strictly increasing function to the objective function of any standard single-objective optimization problem without altering the set of optimal points. This can be helpful in generating properties like smoothness or convexity of the objective.

As a generalization to multiobjective optimization, we consider monotone transformations of the objective space that leave the set of efficient points invariant. Under mild assumptions, for the standard ordering cone, we show that such transformations must be component-wise transformations, which means that a univariate strictly increasing function is applied to each of the objectives.

The same class of transformations also leaves the sets of weakly and of Geoffrion properly efficient points invariant. In addition, our approach allows us to specify trade-off bounds of properly efficient points after the transformation. We apply our results to prove some previously unknown properties of the compromise approach.

■ WC-11

Wednesday, 14:00-16:00 - Room: B100/5017

Interior point methods and applications - Part II

Stream: Interior point methods and applications
Invited session
Chair: *Stefano Cipolla*

1 - A tailored, matrix free interior point method for fast optimization on gas networks
Rowan Turner, Lars Schewe, John Pearson

We consider a PDE-constrained optimization problem arising from the prospective use of hydrogen as an energy carrier to support fully renewable electric grids. One important question is whether existing natural gas infrastructure can be reused for hydrogen to this end, and the challenges this brings for the control of these networks. We expect that a hydrogen network which uses gas generated from excess renewable electricity would be more difficult to control as the patterns of injection and withdrawal would be much less regular than today. Additional challenges arise from new operating parameters required for hydrogen – such as controlling for pressure fluctuations to prevent pipe-ageing. Motivated by a need for instationary optimization methods on networks at scale, we present a specialized, matrix free interior point method for gas problems. Our test problem is a line-pack optimization problem using a discretization of the 1d isothermal Euler equations, as a step towards understanding the important questions above. By incorporating a bespoke preconditioned iterative solver to tackle the linearized systems at each iteration of the interior point method, which form the key computational bottleneck in such a method, we utilize the highly structured nature of the problem to gain efficiency. The expectation is that the method will scale well with both network size and time windows, and be generalizable to broader PDE-constrained network optimization problems.

2 - An interior point approach for multi horizon risk averse stochastic optimization.
Jordi Castro, Laureano F. Escudero, Juan Francisco Monge

In a previous paper, the authors presented a novel approach based on a specialized interior-point method (IPM) for solving large-scale multistage continuous stochastic optimization problems. This approach considered both strategic (long-term) and operational (short-term) uncertainties and decisions.

This work extends the previous approach by incorporating risk-averse constraints: either expected conditional value-at-risk or expected conditional stochastic dominance. As in the earlier risk-neutral approach, the new risk-averse model is reformulated using splitting variables. The reformulated model remains compatible with the specialized IPM, which computes the Newton direction by combining Cholesky factorizations with preconditioned conjugate gradients (PCG).

It is shown that the new reformulated risk-averse constraints simply extend the preconditioner of the PCG with an additional diagonal matrix, preserving the efficient solution of systems using the preconditioner.

Preliminary results are reported for the solution of real-world problems involving several million variables and constraints.

3 - A higher-order extrapolation interior-point method for nonlinear optimization
Pim Heeman, Anders Forsgren

Interior-point methods for smooth optimization problems transform the problem into a problem without explicit inequalities by a logarithmic barrier transformation of the constraints. This gives an approximate problem where the accuracy of this approximation is controlled by a positive scalar, the so-called barrier parameter. The problem is solved for a sequence of decreasing barrier parameters, warm-starting the problem for the current parameter with the solution for the previous. A trade-off is here to be made between decreasing the barrier parameter fast enough to not solve too many optimization problems that have a fixed minimum cost associated with them on one hand and ensuring that the perturbed problem is not changed too drastically that warm-starting helps significantly on the other hand. In this talk, an accelerator for primal-dual interior-point methods following this scheme is proposed that uses higher-order derivatives, for asymptotic convergence at a rate proportional to the order of derivatives used. For problems of reduced complexity like convex quadratic programming problems, computational test results will be shown based on a proof-of-concept method using this accelerator.

4 - Low-Rank Semidefinite Programming for Symmetric Tensor Decomposition: An Experimental Study
Yassine Koubaa

Semidefinite programming (SDP) has long been a cornerstone in global optimization, particularly in polynomial optimization. Yet its application to tensor decomposition—via the matrix completion of moment (Hankel) matrices—remains relatively underexplored. In this work, we propose an SDP-based formulation aimed at recovering full moment matrices from partial observations, leveraging the intrinsic low-rank structure of Hankel matrices. We analyze and compare different low-rank oriented interior-point methods, based on effective preconditioner for low rank Schur complement matrix such as Loraine, and perturbation-based methods such as IPLR- BB, that naturally drives the solution toward a low-rank structure. Preliminary experiments will benchmark these methods against classical SDP solvers to assess computational efficiency and the quality of the low-rank solutions. We will also present experiments, assessing the strengths and weakness of the approach for tackling the challenges of tensor decomposition.

■ WC-12

Wednesday, 14:00-16:00 - Room: B100/8009

Optimisation under uncertainty for sustainability

Stream: Applications: AI, uncertainty management and sustainability
Invited session
Chair: *Parisa Ahani*

1 - Optimization of Dynamic Service Systems with Adaptive Resource Allocation and Pricing Strategies Based on Random Environments and Customer Feedback
Chesoong Kim

This paper models and optimizes a dynamic service system that adaptively adjusts resource allocation and service pricing in real-time, considering both random environmental changes and customer feedback. The system integrates features of multi-server and retrial queuing systems, modeling scenarios where the number of available servers fluctuates due to random environmental changes, such as in communication networks, and arrival rates dynamically change based on customer ratings. Service prices are also adjusted in real-time based on queue length and customer feedback to maximize system efficiency. A multi-dimensional Markov chain model is used to analyze the system's stability and performance. Optimal resource allocation and pricing policies are derived to maximize service provider revenue and enhance customer satisfaction. The efficiency of the proposed model is validated through numerical experiments under various scenarios, demonstrating its applicability in diverse fields such as communication networks, online platforms, and transportation systems.

2 - Efficient Design of Water Distribution Networks: An Approximation and a Heuristic
Nitish Dumoliya, Ashutosh Mahajan

A novel approximation of the Hazen-Williams equation and an heuristic are proposed to design cost-effective water distribution networks (WDNs). WDN design is challenging in cyclic networks due to nonlinear, nonconvex headloss constraints. Pipe size selection is limited to commercially available diameters. WDNs design model optimizes pipe sizes, lengths, flows, and node heads to minimize cost while meeting demand and hydraulic constraints. WDN models have two main challenges. First, some solvers abort, flagging an undefined first derivative due to headloss equation formulation, even though it is mathematically well defined. Second, the headloss equation's second derivative is undefined at zero flow. We develop an approximation of the Hazen-Williams equation, ensuring differentiability. We compare the behaviour of benchmark instances on solvers and modeling languages to highlight these challenges. Comparative results are given for the approximate model. We also propose a heuristic based on the acyclic flows in the water network. The idea is to solve a sequence of nonlinear optimization models to find locally optimal solutions. In each iteration, flow direction is fixed in some carefully selected arcs based on the last local solution of the nonconvex model, and it allows us to explore promising locally optimal solutions that may be close to the global solution. We compare the performance of our heuristic on benchmark water network instances to global and local solvers.

3 - Optimizing decision-making framework under uncertainty for fleet replacement problem
Parisa Ahani, Maria Isabel Gomes

This study explores the fleet composition challenge faced by transportation operators over a designated planning horizon, accounting for uncertainties in energy prices, vehicle acquisition costs and operational expenditures. Within the constraints of a limited budget, the decision-maker must select an optimal mix of vehicle types, each offering unique trade-offs in terms of cost and performance. For example, diesel-powered vehicles often have lower upfront costs but incur higher fuel and maintenance expenses, whereas electric vehicles tend to be more expensive initially but offer reduced operating costs over time. Given the fluctuating nature of energy markets particularly oil factoring uncertainty into the fleet planning process becomes essential for avoiding high long-term ownership costs. A strategy that prioritizes vehicles with lower purchase prices may prove more expensive when considering total cost of ownership. To address these complexities, this research introduces a novel multi-objective mixed-integer quadratic programming model. The model aims to simultaneously minimize the overall cost encompassing purchase price, energy use, maintenance, depreciation and emissions and the financial risks tied to uncertain factors such as fuel prices and vehicle cost variability. By generating a Pareto front, the model equips decision-makers with a robust framework to assess trade-offs between cost and risk. This enables the development of cost-efficient and resilient strategies.

Wednesday, 16:05-16:15

■ WD-01

Wednesday, 16:05-16:15 - Room: B100/1001

Closing

Stream: Plenaries
Invited session
Chair: *Alain Zemkoho*
Chair: *Giancarlo Bigi*
Chair: *Selin Ahipasaoglu*

Applications: AI, uncertainty management and sustainability

Selin Ahipasaoglu
University of Southampton
sda1u20@soton.ac.uk

Hongyu Zhang
University of Southampton
hongyu.zhang@soton.ac.uk

Track(s): 12

Applications: Finance

Panos Parpas
Imperial College London
panos.parpas@imperial.ac.uk

Houduo Qi
The Hong Kong Polytechnic University
houduo.qi@polyu.edu.hk

Track(s): 12

Bilevel and multilevel optimization

Qingna Li
Beijing Institute of Technology
qnl@bit.edu.cn

Alain Zemkoho
University of Southampton
a.b.zemkoho@soton.ac.uk

Track(s): 7

Conic Optimization

Miguel Anjos
University of Edinburgh
anjos@stanfordalumni.org

Track(s): 11

Generalized convexity and monotonicity

Sorin-Mihai Grad
ENSTA Paris
sorin-mihai.grad@ensta-paris.fr

Felipe Lara
Universidad de Tarapacá
felipelara@udec.cl

Vuong Phan
University of Southampton
t.v.phan@soton.ac.uk

Track(s): 9

Global optimization

Sonia Cafieri
ENAC - Ecole Nationale d'Aviation Civile
sonia.cafieri@enac.fr

Eligius M.T. Hendrix
Universidad de Málaga
eligius@uma.es

Dmitri Kvasov
University of Calabria
kvadim@dimes.unical.it

Track(s): 4

Interior point methods and applications

Stefano Cipolla
University of Southampton
s.cipolla@soton.ac.uk

Track(s): 11

Large scale optimization: methods and algorithms

Stefania Bellavia
Universita di Firenze
stefania.bellavia@unifi.it

Benedetta Morini
Universita di Firenze
benedetta.morini@unifi.it

Track(s): 3 8 10

Multiobjective and Vector Optimization

Gabriele Eichfelder
Technische Universität Ilmenau
gabriele.eichfelder@tu-ilmenau.de

Firdevs Ulus
Bilkent University
firdevs@bilkent.edu.tr

Track(s): 3 10

Nonsmooth and nonconvex optimization

Masoud Ahookhosh
University of Antwerp
masoud.ahookhosh@uantwerp.be

Behzad Azmi
University of Konstantz
behzad.azmi@uni-konstanz.de

Susan Ghaderi
KU Leuven
susan.ghaderi@kuleuven.be

Track(s): 2 6

Numerical Methods and Applications

Track(s): 7 13

Optimal and stochastic optimal control

Gerhard-Wilhelm Weber
Poznan University of Technology
gerhard-wilhelm.weber@put.poznan.pl

Track(s): 11

Optimization for machine learning

Ahmet Alacaoglu
University of British Columbia
alacaoglu@math.ubc.ca

Alp Yurtsever
Umeå University
alp.yurtsever@umu.se

Track(s): 4 5 8

Optimization, Learning, and Games

Georgios Piliouras
Singapore University of Technology and Design
georgios.piliouras@gmail.com

Antonios Varvitsiotis
Singapore University of Technology and Design
antonios@sutd.edu.sg

Track(s): 10

Plenaries

Selin Ahipasaoglu
University of Southampton
sda1u20@soton.ac.uk

Giancarlo Bigi
Universita' di Pisa
giancarlo.bigi@unipi.it

Alain Zemkoho
University of Southampton
a.b.zemkoho@soton.ac.uk

Track(s): 1

Riemannian Manifold and Conic Optimization

Panos Parpas
Imperial College London
panos.parpas@imperial.ac.uk

Track(s): 11

Sparsity guarantee and cardinality-constrained (MI)NLPs

Immanuel Bomze
University of Vienna
immanuel.bomze@univie.ac.at

Michel De Lara
École des Ponts ParisTech
michel.delara@enpc.fr

E. Alper Yildirim
The University of Edinburgh
e.a.yildirim@ed.ac.uk

Track(s): 13

Systematic and computer-aided analyses of optimization algorithms

Pontus Giselsson
Lund University
pontusg@control.lth.se

Adrien Taylor
Inria/ENS
adrien.taylor@inria.fr

Track(s): 2 8

Variational analysis: theory and algorithms

Francisco Javier Aragón Artacho
University of Alicante
francisco.aragon@ua.es

Fernando García Castaño
University of Alicante
fernando.gc@ua.es

Track(s): 9

Zeroth and first-order optimization methods

Andrea Cristofari
University of Rome "Tor Vergata"
andrea.cristofari@uniroma2.it

Cesare Molinari
Università di Genova
cesare.molinari@edu.unige.it

Francesco Rinaldi
Università di Padova
rinaldi@math.unipd.it

Silvia Villa
università di Genova
silvia.villa@unige.it

Track(s): 1 10

Session Chair Index

Gorbunov, Eduard MD-05
eduard.gorbunov@mbzuai.ac.ae
Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dabi, United Arab Emirates

Goujaud, Baptiste MD-08, TB-08
baptiste.goujaud@gmail.com
CMAP, Ecole Polytechnique, Palaiseau, IDF, France

Goujaud, Baptiste WC-02
baptiste.goujaud@polytechnique.edu
CMAP, Ecole Polytechnique, Palaiseau, France

He, Chuan WC-05
chuan.he@liu.se
Linköping University, Sweden

Hendrix, Eligius M.T. TB-04, TC-04
eligius@uma.es
Computer Architecture, Universidad de Málaga, Malaga, Spain

Hu, Yukuan TB-11
ykhu@lsec.cc.ac.cn
CERMICS, École nationale des ponts et chaussées, Institut Polytechnique de Paris, Champs-sur-Marne, Marne-la-Vallée Cedex 2, France

Kabgani, Alireza MD-02, TB-06, WC-06
alireza.kabgani@uantwerpen.be
Department of Mathematics, University of Antwerp, Antwerp, BELGIUM, Belgium

Karagulyan, Avetik MB-05
avetik.karagulyan@cnrs.fr
CNRS, Orsay, Ile de France, France

Kavis, Ali WB-05
kavis@austin.utexas.edu
ECE Department, University of Texas at Austin, Austin, Texas, United States

Kvasov, Dmitri TB-04
kvadim@dimes.unical.it
DIMES, University of Calabria, Rende (CS), Italy

Lapucci, Matteo WB-03
matteo.lapucci@unifi.it
Department of Information Engineering, University of Florence, Firenze, Italy

Lara, Felipe MD-09
felipelaraobreque@gmail.com
Universidad de Tarapacá, Arica, Arica y Parinacota, Chile

Lämmel, Sebastian MD-07
sebastian.laemmel@mathematik.tu-chemnitz.de
Chemnitz University of Technology, Chemnitz, Germany

Mansueto, Pierluigi TC-10, WB-10
pierluigi.mansueto@unifi.it
Department of Information Engineering, University of Florence, Florence, Italy

Martin Barragan, Belen MC-12
belen.martin@ed.ac.uk
University of Edinburgh Business School, The University of Edinburgh, Edinburgh, United Kingdom, United Kingdom

Massion, Bastien WC-06
bastien.massion@uclouvain.be
Icteam (inma), UCLouvain, Belgium

Melguizo Padial, Miguel Angel TC-09
ma.mp@ua.es
Mathematics, University of Alicante, Alicante, Comunidad Valenciana, Spain

Mokhtari, Aryan WB-05
mokhtari@austin.utexas.edu
ECE Department, University of Texas at Austin, Austin, Texas, United States

Molinari, Cesare TC-01, WB-01, MD-10, TB-10
cecio.molinari@gmail.com
Università di Genova, Genova, Italy

Morini, Benedetta TB-03, TC-03
benedetta.morini@unifi.it
Dipartimento di Ingegneria Industriale, Universita di Firenze, Firenze, Italy

Neussel, Felix WC-10
felix.neussel@kit.edu
Institute for Operations Research, KIT, Germany

Palagi, Laura TA-01, MD-13
palagi@diag.uniroma1.it
Department of Computer, Control, and Management Engineering A. Ruberti, Sapienza University of Rome, Rome, Italy

Parpas, Panos TC-11
panos.parpas@imperial.ac.uk
Computing, Imperial College London, London, United Kingdom

Patron, Gabriel WB-12
gabriel.patron17@imperial.ac.uk
Imperial College London, United Kingdom

Phan, Vuong WC-03, MC-09
t.v.phan@soton.ac.uk
University of Southampton, Southampton, United Kingdom

Piliouras, Georgios MB-10
georgios.piliouras@gmail.com
Singapore University of Technology and Design, Singapore

Qi, Houduo MB-12
houduo.qi@polyu.edu.hk
Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong

Qian, Yaru MB-07
yq4n22@soton.ac.uk
Electronics and computer science, University of Southampton, Southampton, United Kingdom

Author Index

Rinputprogram-europt2025.ind