SIAM 2023

Searchable Abstracts Document

Conference on Optimization

(OP23)

May 31–June 3, 2023

This document was current as of May 22, 2023. Abstracts appear as submitted.

SIAM
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 U.S.
Telephone: 800-447-7426 (U.S. & Canada) +1-215-382-9800 (Worldwide)
meetings@siam.org

IP1

Acceleration of First-order Optimization Algorithms via Damped Inertial Dynamics

We report recent progress regarding the acceleration of first-order algorithms for convex optimization. We consider algorithms obtained by temporal discretization of damped inertial dynamics. The optimization properties come from the design of the damping term. Nonsmooth optimization is addressed via splitting algorithms taking into account smooth terms by gradient methods, and non-smooth terms by proximal methods. We report recent improvements concerning the Nesterov accelerated gradient method and then introduce the Ravine method, which shares convergence properties similar to Nesterov's method. We show that the high-resolution ODE of the Nesterov and Ravine methods makes appear the geometric damping driven by the Hessian of the function to be minimized. The Hessian driven damping significantly dampen the oscillations that occur with inertial systems. We present fast algorithms for general convex optimization which are obtained by introducing a Tikhonov regularization with vanishing coefficient in Polyak's heavy ball. We present a new approach based on temporal rescaling and averaging of the continuous steepest descent. This provides fast convergence rates with the advantage that we do not need to do a Lyapunov analysis. Finally, we present some recent results concerning the extension of the above results to the stochastic framework, and the stochastic differential equation approach to solve convex optimization problems with a noisy gradient input.

Hedy Attouch
Université Montpellier II, France
hedy.attouch@umontpellier.fr

IP2

Convex-Composite Optimization

In convex-composite optimization one minimizes an objective function formed by composing a convex function with a smooth function. Such problems have roots going back to the work of Gauss on nonlinear least-squares. The convex-composite framework nicely unifies much of the theory and algorithms for a wide range of optimization problems including nonlinear programming, feasibility, minimax optimization, l_1 optimization, Kalman smoothing, parameter selection, and nonlinear maximum likelihood. The systematic study of convex-composite problems began in the 1970's concurrent with the emergence of modern nonsmooth variational analysis. The synergy between these ideas was natural since convex-composite functions are neither convex nor smooth. Recently there has been a resurgence of interest in this problem class due to emerging methods for approximation, regularization and smoothing as well as the relevance to a number of problems in global health, environmental modeling, image segmentation, dynamical systems, signal processing, machine learning, and AI. In this talk we review the convex-composite problem structure and variational properties. We then discuss algorithm design and review a few recent advances in first- and second-order methods.

James V. Burke
University of Washington
Department of Mathematics
jvburke@uw.edu

IP3

Towards a Taxonomy of All Pivot Rules for the Simplex Method

Dantzigs Simplex method is one of the most famous and popular algorithms in optimization. But despite its popularity, there are truly basic mathematical questions we cannot yet answer. At the core of any version of the simplex method is a pivot rule that selects the outgoing arc for a current vertex. Pivot rules come in many forms and types, but after 80 years we are still ignorant whether there is one that can make the simplex method run in polynomial time. This talk surveys what we know and our attempts to geometrically classify pivot rules. In the first part of my talk I will review the state of the art of pivot rules, including two recent positive results: For 0/1 polytopes there are explicit pivot rules for which the number of non-degenerate pivots is polynomial and even linear (joint work with A. Black, S. Kafer, L. Sanita). In the second half of my talk, I present a parametric analysis for all pvot rules using a geometric by-product we called its footprint. Memoryless pivot rules emerge as the most important of all rules and for every linear program we can construct a polytope, the pivot rule polytope, that parametrizes all memoryless pvot rules of a given LP. Thus we uncover hidden structure on the space of all pivot rules of an LP (joint work with A. Black, N. Lu tjeharms, and R. Sanyal).

Jesús A. De Loera
University of California, Davis, U.S.
deloera@math.ucdavis.edu

IP4

Exactness in Semidefinite Program Relaxations and Its Implications

Semidefinite program (SDP) relaxation is a popular technique for building high quality relaxations of nonconvex optimization problems in a variety of fields. In this talk, we review recent advancements on SDP exactness and its computational implications. We will examine three notions of SDP exactness for general quadratically constrained quadratic programs (QCQPs): (i) objective value exactnessthe condition that the optimal value of the QCQP and the optimal value of its SDP relaxation coincide, (ii) convex hull exactnessthe condition that the convex hull of the QCQP epigraph coincides with the (projected) SDP epigraph, and (iii) the rank-one generated (ROG) propertythe condition that a particular conic subset of the positive semidefinite cone related to a given QCQP is generated by its rank-one matrices. We will see how a geometric treatment of the projected SDP relaxation while crucially considering the interaction of the objective function with the constraints leads to sufficient conditions for (i) and (ii). The ROG property complements these results by offering a sufficient condition for both (i) and (ii) that is oblivious to the objective function. We will give a variety of sufficient conditions for these exactness conditions and discuss settings where these conditions are also necessary. We will conclude by highlighting how SDP exactness can be exploited in designing accelerated storage-optimal first-order methods for solving low-rank exact SDPs.

Fatma Kilinc-Karzan
Tepper School of Business
Carnegie Mellon University

fkilinc@andrew.cmu.edu

IP5

Optimization Over Probability Distributions

In modern applications of machine learning, we frequently encounter optimization problems over probability distributions. In such applications, the underlying probability distributions are represented by random samples (particles) drawn from the distributions. A first order optimization procedure moves these particles along the steepest descent direction, which induces a change of the underlying proability distribution they represent. In this talk, I will discuss some of such optimization problems in machine learning, including optimization algorithms, convergence results, and challenges.

Tong Zhang
The Hong Kong University of Science and Technology
tongzhang@tongzhang-ml.org

IP6

Large-Scale and Data-Driven Markov Decision Processes

Markov decision processes (MDPs) constitute one of the predominant modeling and solution paradigms for dynamic decision problems affected by uncertainty. MDPs model the dynamics of a system through a random state evolution that generates rewards over time. The decision maker aims to select actions that influence this state evolution so as to maximize rewards. In this talk, we review recent advances in MDPs along two directions: (i) the construction of data-driven policies that combine the (traditionally separated) tasks of estimating the systems behavior and selecting actions that maximize rewards in the estimated system, and (ii) the exploitation of structure to solve large-scale problems. In view of (i), we will show how the consideration of data-driven policies naturally leads to the study of robust MDPs, where the decision maker combats overfitting by hedging against the worst system dynamics that are plausible under some given training data. We will also discuss how alternative models of robustness offer different tradeoffs between the competing goals of out-of-sample performance and complexity of the involved policies and computations. As for (ii), we will review two types of structure that allow us to alleviate the well-known curse of dimensionality: weakly coupled MDPs that combine a potentially large number of MDPs via a small number of linking constraints, and factored MDPs whose states are represented by assignments of values to state variables that evolve and contribute to the system's rewards largely independently.

Wolfram Wiesemann
Department of Computing
Imperial College London
wwiesema@imperial.ac.uk

IP7

Optimization Applications and Challenges across Amazon Operations

Amazon operations faces many decisions while designing and executing the processes to fulfill orders for Amazons customers. These range from long term investment choices about facilities to online fulfillment decisions regarding delivery speed and how to route and sequence package deliveries. Many of these decisions have been automated by casting them as optimization problems. These result in large-scale optimization problems, both continuous and discrete, many of which need to be solved quickly and frequently. Most of these decisions are made in the face of uncertainty, and even decisions that may be cast as deterministic optimization incorporate inputs and parameters that have been estimated. This talk will examine optimization problems of different types that arise across Amazons fulfillment operations during the automation of the decision processes. Some of the most common and recurring structures and subproblems that Amazon encounters will be illustrated along with the applications from which they stem, highlighting what characteristics and application requirements make the optimization problems challenging. These problems arise throughout the design, planning, and execution of the fulfillment network.

Russell Allgor
Amazon, U.S.
rallgor@amazon.com

SP1

SIAM Prize Session: 2023 SIAG/OPT Best Paper Prize Lecture: Stochastic Model-Based Minimization of Weakly Convex Functions

We consider a family of algorithms that successively sample and minimize simple stochastic models of the objective function. We show that under reasonable conditions on approximation quality and regularity of the models, any such algorithm drives a natural stationarity measure to zero at the rate $O(k^{-1/4})$. As a consequence, we obtain the first complexity guarantees for the stochastic proximal point, proximal subgradient, and regularized Gauss–Newton methods for minimizing compositions of convex functions with smooth maps. The guiding principle, underlying the complexity guarantees, is that all algorithms under consideration can be interpreted as approximate descent methods on an implicit smoothing of the problem, given by the Moreau envelope. Specializing to classical circumstances, we obtain the long-sought convergence rate of the stochastic projected gradient method, without batching, for minimizing a smooth function on a closed convex set.

Damek Davis
Cornell University
ORIE
dsd95@cornell.edu

Dmitriy Drusvyatskiy
University of Washington
ddrusv@uw.edu

SP2

SIAM Prize Session: 2023 SIAG/OPT Test of Time Award Lecture: Two Decades of Low-Rank Optimization

In this talk, we offer our personal perspective on low-rank algorithms for semidefinite programs and related optimizaton problems. We touch on their history and describe connections with other trends in optimization, including first-order methods and benign nonconvexity. Finally, we highlight recent theoretical insights into low-rank algorithms—as well as various applications for which the low-rank approach has proven successful. Thank you to session sponsor

IBM.

Samuel Burer
University of Iowa
Dept of Management Sciences
samuel-burer@uiowa.edu

Renato Monteiro
Georgia Institute of Technology
rm88@gatech.edu

JP1

Welcome Remarks and Joint Plenary with ACDA23 - On the Optimization of Nonsmooth Problem without Generalized Derivatives

Numerous applications may require the solution of nonsmooth optimization problems, e.g., due to the action of values, a bilevel structure or complementarity conditions. For a large class of target functions, the concept of abs-linearization allows the generation of a piecewise linear local model that is provable of second order. Similar to the quadratic model generated by a truncated Taylor series in the smooth situation, this piecewise linear model can be used as a building block for optimization algorithms targeting nonsmooth problems of different kinds. In this talk, first we define abs-smooth functions covering a wide range of applications like clustering, image restoration, and robust gas network optimization. Also various mathematical models like complementarity problems or bilevel optimization tasks can be formulated as abs-smooth functions. Subsequently, the abs-linearization approach and the properties of the resulting local model will be illustrated. Then, we discuss the solution of piecewise linear optimization problems. Based on this capability we can derive optimization approaches that allow also the handling of constraints and/or nonlinearities. For each class of the nonsmooth optimization problems and the corresponding solution approach, convergence results and numerical examples will be given.

Andrea Walther
Humboldt-Universität zu Berlin
andrea.walther@math.hu-berlin.de

CP1

Newton-Pmr: Newton Subspace Methods with Complexity Guarantees for Non-Convex Optimization

Recently, Newton-MR methods have been introduced for solving nonconvex unconstrained smooth optimization problems. Leveraging the inherent ability of the minimum residual (MINRES) inner solver to detect directions of nonpositive curvature (NPC), Newton-MR variants enjoy a variety of optimal complexity guarantees. However, the application of these methods to modern high-dimensional problems remains challenging. To address this, we present novel variants that incorporate certain dimensionality reduction techniques. In particular, our proposed methods are based on recent results that have shown that preconditioning MINRES with a positive semidefinite but singular preconditioner is in fact equivalent to solving a low-dimensional problem whose dimension corresponds to the nullity of the preconditioning matrix. Utilizing these dimensionality reduction properties of preconditioned MINRES, we present novel variants of Newton-MR, called Newton-PMR, which can be readily applied to high-dimensional problems, while achieving desirable complexity guarantees.

Yang Liu
University of Oxford
yang.liu@maths.ox.ac.uk

Andre Milzarek
The Chinese University of Hong Kong, Shenzhen
School of Data Science
andremilzarek@cuhk.edu.cn

Fred Roosta
University of Queensland, Australia
fred.roosta@uq.edu.au

CP1

Decomposed Newton Steps

We consider Newton's method for computing roots $x \in \mathbb{R}^n$ of differentiable residuals $F(x)$ with dense Jacobians $F'(x)$ iteratively as x^k, $k = 0, \ldots$. The residual is assumed to be of the form $v_q = F(v_0) = (F_q \circ \ldots \circ F_1)(v_0)$ where $v_0 = x^k$. From $F'(v_0) = F'_q(v_{q-1}) \cdot \ldots \cdot F'_1(v_0)$ it follows that $-F'(v_0)^{-1} \cdot F(v_0) = -F'_1(v_0)^{-1} \cdot \ldots \cdot F'_q(v_{q-1})^{-1} \cdot F(v_0)$. Tridiagonal F'_i require $q \geq n$ in order for $F'(v_0)$ to become dense. Factorization yields $F_i \equiv F_i(v_{i-1}) = L_i(v_{i-1}) \cdot U_i(v_{i-1}) \equiv L_i \cdot U_i$ at a cost of $\mathcal{O}(q \cdot n)$. The Newton steps can hence be computed as $-U_1^{-1} \cdot \left(L_1^{-1} \cdot \ldots \cdot \left(U_q^{-1} \cdot \left(L_q^{-1} \cdot F_i\right)\right) \ldots\right)$ with a cost of $\mathcal{O}(n)$ yielding a total cost of $\mathcal{O}(q \cdot n)$. Accumulation of the dense $F'(x^k)$ at the cost of $\mathcal{O}(q \cdot n^2)$ and solution of the dense Newton systems can be avoided. The talk presents an outline of the method including generalization to banded local Jacobians and run time results. Nonregular sparsity patterns yield the usual challenges imposed by sparse linear algebra. Implications for the efficient implementation of this method as part of tape-based software tools for algorithmic differentiation are discussed.

Uwe Naumann
RWTH Aachen University
Software and Tools for Computational Engineering
naumann@stce.rwth-aachen.de

CP1

Accelerating the Spherical Triangle Algorithm for the Convex-Hull Membership Problem

In this study, show how to accelerate the spherical triangle algorithm (S-TA), a geometrical algorithm, recently introduced to solve the convex hull membership problem (CHMP). S-TA is a modification of Triangle Algorithm, which has its foundation on a theorem of alternatives, known as distance duality and its iteration mechanism is quite similar to a iteration of the well-now conditional gradient (Frank-Wolfe) method. Although S-TA convergence rate, for general problems, is $\mathcal{O}\left(1/\varepsilon^2\right)$, an improved iteration complexity of $\mathcal{O}(M/\varepsilon)$, with $M \geq 1$ was recently presented by [Kalantari and Zhang, ACM Trans. Math. Softw., 2022], under a mild assumption, called ε-property, that an iterate must comply. We propose here a heuristic that provides an iterate which satisfy the ε-property and show numerical results revealing that this heuristic has advantages, when compared with both TA and S-TA.

Luiz-Rafael Santos, Douglas Gonçalves, Rafaela Filippozzi
UFSC
l.r.santos@ufsc.br, douglas.goncalves@ufsc.br,

rafaela.filippozzi@gmail.com

CP2

The Lasserre Hierarchy for Equiangular Lines with a Fixed Angle

In sphere packing and spherical coding problems, calculable bounds coming from semidefinite programming play an important role. These bounds are mathematically interesting and the implementation is challenging. One example of such a bound is the Lasserre hierarchy, with each level giving a better, but more difficult to compute, bound. In our paper we have computed the second and third levels of the Lasserre hierarchy for the spherical finite distance problem. To do so, a connection is used between invariants in representations of the orthogonal group and representations of the general linear group. This allows computations in high dimensions and high precision interpolation is used to obtain solutions for all large dimensions. This gives new upper bounds for a range of spherical finite distance configurations. Our methods give a new bound on $N_\alpha(n)$: the maximum number of equiangular lines in dimension n with common angle $\arccos(\alpha)$. This is the first time an asymptotic bound is obtained from semidefinite programming for spherical codes. When setting up the semidefinite program, the main problem is computing invariants of representations and computing the Fourier inverse. In the talk, we will discuss the mathematical problem and the implementation of the semidefinite program.

Willem De Muinck Keizer
TU Delft
willemdmk@gmail.com

David de Laat
Delft University of Technology
d.delaat@tudelft.nl

Fabricio Machado
Universidade de São Paulo
fabcm1@gmail.com

CP2

Experiments on Dynamic Null Step Updates in the ConicBundle Subproblem

Within the ConicBundle approach for large scale semidefinite programming solving the quadratic semidefinite subproblem typically dominates the cost per iteration. In the case of a poor semidefinite model a potential null step may be detected early by preevaluating the eigenvalue oracle already at intermediate interior point solutions of the subproblem, somewhat akin to the proximal analytic center cutting plane method. This information may then be used to improve the semidefinite cutting model dynamically while solving the subproblem. We discuss a first tentative implementation of these ideas within the callable library ConicBundle and present some preliminary computational results.

Christoph Helmberg
TU Chemnitz
Germany
helmberg@mathematik.tu-chemnitz.de

CP2

Solving Sdp Relaxations of Max-Cut Problem with Large Number of Hypermetric Inequalities by L-Bfgs-B

We present a computational study of SDP-based bounds for Max-Cut that use a subset of hypermetric inequalities as cutting planes to strengthen the basic relaxation. Solving these relaxations is computationally challenging due to the potentially large number of violated constraints. To overcome these difficulties, we describe a heuristic separation algorithm for hypermetric inequalities and propose to use the augmented Lagrangian method as a bounding routine. Computational experiments show that the resulting relaxations provide very tight bounds for the Max-Cut.

Timotej Hrga
University of Ljubljana
timotej.hrga@fmf.uni-lj.si

Janez Povh
University of Ljubljana
Faculty of Mechanical Engineering
janez.povh@lecad.fs.uni-lj.si

CP2

Approximation Hierarchies for Copositive Cone over Symmetric Cone and Their Comparison

We first provide an inner-approximation hierarchy described by a sum-of-squares (SOS) constraint for the copositive (COP) cone over a general symmetric cone. The hierarchy is a generalization of that proposed by Parrilo (2000) for the usual COP cone (over a nonnegative orthant). Second, we characterize the COP cone over a symmetric cone using the usual COP cone. By replacing the usual COP cone appearing in this characterization with the inner- or outer-approximation hierarchy provided by de Klerk and Pasechnik (2002) or Yildirim (2012), we obtain an inner- or outer-approximation hierarchy described by semidefinite but not by SOS constraints for the COP cone over the direct product of a nonnegative orthant and a second-order cone. We then compare them with the existing hierarchies provided by Zuluaga et al. (2006) and Lasserre (2014). Theoretical and numerical examinations imply that we can numerically increase a depth parameter, which determines an approximation accuracy, in the approximation hierarchies derived from de Klerk and Pasechnik (2002) and Yildirim (2012), particularly when the nonnegative orthant is small. In such a case, the approximation hierarchy derived from Yildirim (2012) can yield nearly optimal values numerically. Combining the proposed approximation hierarchies with existing ones, we can evaluate the optimal value of COP programming problems more accurately and efficiently.

Mitsuhiro Nishijima
Tokyo Institute of Technology
nishijima.m.ae@m.titech.ac.jp

Kazuhide Nakata
Tokyo Institute of Technology
Department of Industrial Engineering and Management
nakata.k.ac@m.titech.ac.jp

CP2

On Semidefinite Representations of Second-Order Conic Optimization Problems

Second-order conic optimization (SOCO) can be considered as a special case of semidefinite optimization (SDO). Current literature states that a SOCO problem can be embedded in a semidefinite optimization problem using the arrow-head matrix transformation. However, a primal-dual solution pair cannot be mapped simultaneously using the arrow-head transformation as we might lose complementarity and duality in some cases. To address this issue, we investigate the relationship between SOCO problems, and their SDO counterpart. Through derivation of standard semidefinite representations of SOCO problems, we introduce admissible mappings. We show that the proposed mappings preserves both feasibility and optimality. Further, we discuss how the optimal partition of a SOCO problem maps to the optimal partition of its SDO counterpart.

Pouya Sampourmahani
Lehigh University
pos220@lehigh.edu

Mohammadhossein Mohammadisiahroudi
Industrial & System Engineering Dep, Lehigh University
625 Montclair Ave, Bethlehem, PA, USA, 18015
mom219@lehigh.edu

Tamas Terlaky
Lehigh University
Department of industrial and Systems Engineering
terlaky@lehigh.edu

CP3

Efficient Quantum Algorithm for Regularized Optimization

Regularized optimization methods, such as proximal methods and regularized Newton methods, are among the most theoretically sound and applicable approaches to convex and nonconvex optimization. Within the emerging trend of quantum computation, while there has been extensive work on quantum numerical linear algebra, previous quantum algorithms for general nonlinear optimization have complexity exponential in the number of iterations. In this work, we develop the first efficient quantum algorithm for high-dimensional optimization problems based on regularized methods with provable convergence guarantee and computational complexity, which scales polynomially in the number of iterations and polynomial-logarithmically in dimension. This is an exponential improvement over previous quantum algorithms, as well as a potential exponential speedup in terms of dimension over existing classical approaches under the discussion of end-to-end settings. To achieve this, we offer a novel analysis on the regularized gradient flows as continuous-time limits of first-order and second-order methods, for which we employ and improve upon the quantum Carleman linearization methodology in Liu et al. [PNAS, 118 (35), 2021] with new convergence criteria. We discuss potential applications of our quantum algorithm to various scenarios arising in computer science, statistics, signal processing, and machine learning.

Jin-Peng Liu
University of Maryland
jliu1219@terpmail.umd.edu

Weijie Su
University of Pennsylvania
suw@wharton.upenn.edu

Nikita Doikov
École Polytechnique Fédérale de Lausanne
nikita.doikov@epfl.ch

Aram Harrow
MIT
aram@mit.edu

CP3

On the Implementation of a Parallel Decomposition Algorithm for Large-Scale Nonlinear Two-Stage Problems

We present a scalable decomposition algorithm for large-scale nonlinear two-stage optimization problems in which the problem can be decomposed into a master problem and a number of subproblems. The distinguishing feature of the algorithm is that it is based on a smoothing technique using barrier formulations of the second-stage subproblems. The advantage of this framework is that speedup can be obtained by solving subproblems in parallel. The performance of an efficient C++ implementation is studied on a set of problems on power models and the scalability of algorithm is shown.

Xinyi Luo
Northwestern University
xinyiluo2023@u.northwestern.edu

Andreas Wächter
Northwestern University
IEMS Department
waechter@iems.northwestern.edu

CP3

On Global and Local Convergence of Iterative Linear Quadratic Optimization Algorithms for Discrete Time Nonlinear Control

A classical approach for solving discrete time nonlinear control on a finite horizon consists in repeatedly minimizing linear quadratic approximations of the original problem around current candidate solutions. While widely popular in many domains, such an approach has mainly been analyzed locally. We observe that global convergence guarantees can be ensured provided that the linearized discrete time dynamics are surjective, costs on the state variables are strongly convex and no costs on the control are present. We present how the surjectivity of the linearized dynamics can be ensured by appropriate discretization schemes given the existence of a feedback linearization scheme. We present complexity bounds of algorithms based on linear quadratic approximations through the lens of generalized Gauss-Newton methods. Our analysis uncovers several convergence phases for regularized generalized Gauss-Newton algorithms.

Vincent Roulet
Google
vincent.roulet1509@gmail.com

Siddhartha Srinivasa, Maryam Fazel, Zaid Harchaoui
University of Washington
siddh@cs.washington.edu, mfazel@uw.edu, zaid@uw.edu

CP3

Optimization under Rare Chance Constraints

Chance constraints provide a principled framework to mitigate the risk of high-impact extreme events by modifying the controllable properties of a system. The low probability and rare occurrence of such events, however, impose

severe sampling and computational requirements on classical solution methods that render them impractical. This work proposes a novel sampling-free method for solving rare chance constrained optimization problems affected by uncertainties that follow general Gaussian mixture distributions. By integrating modern developments in large deviation theory with tools from convex analysis and bilevel optimization, we propose tractable formulations that can be solved by off-the-shelf solvers. Our formulations enjoy several advantages compared to classical methods: their size and complexity are independent of event rarity, they do not require linearity or convexity assumptions on system constraints, and under easily verifiable conditions, serve as safe conservative approximations or asymptotically exact reformulations of the true problem. Computational experiments on linear, nonlinear, and PDE-constrained problems from applications in portfolio management, structural engineering, and fluid dynamics illustrate the broad applicability of our method and its advantages over classical sampling-based approaches in terms of both accuracy and efficiency.

Shanyin Tong
Columbia University
st3503@columbia.edu

Anirudh Subramanyam
Pennsylvania State University
subramanyam@psu.edu

Vishwas Rao
Argonne National Laboratory
vhebbur@anl.gov

CP4

Optimization of Renewable Energy Storage Using Iron As Energy Carrier

We present strategies for solving optimization problems arising within an interdisciplinary research project in which iron is investigated as energy carrier in a circular process allowing the large-scale transport and storage of renewable energies. The circular process consists of the energy storage via thermochemical reduction of iron oxide to iron and the energy release via thermochemical oxidation of iron back to its oxides. Our work focuses on the modeling and the optimization of the technical components occurring in the overall system and especially the optimal design and operation of the energy storage process. The latter comprises renewable energy generation plants, an electrolyzer for the production of green hydrogen as well as an innovative iron oxide reduction plant. This results in challenging optimization problems including ODE constraints for modeling the reduction reaction kinetics, nonlinear constraints describing the underlying physical processes of the technical components as well as multi-criteria economic objectives. In addition, uncertainties in economic data and the availability of renewable energies are considered. Aiming for global optimal solutions, we use SCIP as solver relying on a spatial branch-and-bound approach.

Elisa Corbean
TU Darmstadt
corbean@mathematik.tu-darmstadt.de

Stefan Ulbrich
Technische Universitaet Darmstadt
Fachbereich Mathematik
ulbrich@mathematik.tu-darmstadt.de

CP4

Mixed-Integer Nonlinear Optimization of Heating Networks with Time-Coupled Components

To reduce carbon emissions the heating sector plays a major role. In particular, district heating in combination with renewable energy generation and waste heat is of great importance. To lower the operating temperatures and transform to decentralized structures the development of new operating strategies is needed. Therefore we consider a global optimization approach of heating networks that aims at finding the cost-optimal operating strategy. The system state in heating networks is described by the underlying physical equations including operating parameters such as the mass flow rate, temperature and pressure. We focus on a stationary case coupled over time via heating storages. This leads to nonlinear equations with binary variables for determining the flow direction in the network pipes. The obtained Mixed-Integer Nonlinear Optimization Problem is solved with the solver SCIP. For real-sized district heating networks the solving process leads to high computational costs, especially for networks with high flexibilities. To lower the costs arising by the coupling of multiple time steps we consider a heuristic approach to investigate in which time steps the storage is charged or discharged based on the costs for heat generation. We evaluate the developed methods with numerical results based on real district heating networks.

Lea Rehlich
TU Darmstadt
rehlich@mathematik.tu-darmstadt.de

Stefan Ulbrich
Technische Universitaet Darmstadt
Fachbereich Mathematik
ulbrich@mathematik.tu-darmstadt.de

Marc Pfetsch
Research Group Optimization, Department of Mathematics
TU Darmstadt
pfetsch@mathematik.tu-darmstadt.de

CP4

Dynamic Modeling and Optimization of Mixed Hydrogen-Natural Gas Flow in Pipeline Networks

The decarbonization of the energy grid is a necessary step to reduce global greenhouse emission levels and limit the effects of climate change. One aspect of this decarbonization process includes the use of low or zero emission hydrogen fuel. At present, pure hydrogen cannot be transported using the current gas pipeline network and needs to be mixed with natural gas for transmission. Moreover, the optimal operation of gas pipeline network is a nontrivial problem which requires the modeling of pressure drop and flow equations inside the pipes and nodes of the network. In this talk, we present a dynamic model for the mixing and transport of hydrogen-natural gas (H2-NG) along a pipeline network. The dynamic model which is derived by lumping the full-scale partial differential algebraic equation (PDAE) model into a differential algebraic equation (DAE) model. The mixed gas network model captures the mixing, pressure drop and flow equations inside each node and pipe of the network. We present multiple reformulations for the nonlinear and non-smooth equations of mixing and compare the performance when solved using standard optimization solvers. We present both dynamic simulation and optimization model for the H2-NG mixed gas flow problem and solve literature test cases to show the effectiveness and need for optimization.

Saif R. Kazi, Anatoly Zlotnik, Kaarthik Sundar
Los Alamos National Laboratory
saifrahaman26@gmail.com, azlotnik@lanl.gov, kaarthik@lanl.gov

Shriram Srinivasan
Computational Earth Sciences Group (EES-16)
Los Alamos National Laboratory
shrirams@lanl.gov

CP5

How to Choose When and How to Rely on Low Fidelity Data in Bi-Fidelity Expensive Black-Box Problems

In the field of Multi-fidelity Expensive Black-Box (Mf-EBB), a problem instance consists of an expensive yet accurate source of information, and one or more cheap yet less accurate sources of information. The field aims to provide techniques either to accurately explain how decisions affect design outcome, or to find the best decisions to optimise design outcomes. Despite the existence of many techniques which provide solutions to both aims, only in recent years have researchers begun to explore the conditions under which the information sources of lesser accuracy can be exploited reliably. Focusing on model creation for problems with a single low-fidelity function, known as Bi-fidelity Expensive Black-Box (Bf-EBB) problems, this talk will present recent findings on what characteristics are required for the low-fidelity function to be beneficial (rather than harmful) when constructing surrogate models of the objective function. Based on these findings, new adaptive surrogate models which selectively choose when to rely on the low-fidelity function will be presented, and their performance compared with classical surrogate models.

Nicolau Andres Thio, Kate Smith-Miles, Andres Muñoz Acosta
The University of Melbourne
ARC Training Centre OPTIMA
nandres@student.unimelb.edu.au, smith-miles@unimelb.edu.au, munoz.m@unimelb.edu.au

CP5

Multi-Fidelity Bayesian Optimization for Composition Design under Uncertainty

Design of many material systems relies on solving complex optimization problems. This complexity stems from multiple sources such as lack of high-fidelity experimental data, inaccuracy and bias of simulations (due to, e.g., missing physics), high cost of data acquisition, and existence of categorical variables that give rise to combinatorially expanding design spaces. To holistically address these challenges, we propose a multi-fidelity Bayesian optimization (MFBO) framework for designing materials composition under uncertainty. Unlike existing MFBO methods, our framework (1) can handle high-dimensional input spaces that have categorical variables, (2) can leverage an arbitrary number of low-fidelity data sources, and (3) can predict which low-fidelity sources are trustworthy. We will demonstrate the advantages of our framework over existing technologies using a host of analytic and engineering examples.

Ramin Bostanabad
Northwestern University
raminb@uci.edu

CP5

Bayesian Optimization Formulations for Nonlinear Model Calibration

Nonlinear model calibration is difficult with expensive computational models such as molecular simulations as each evaluation can take weeks and must be evaluated hundreds of times. This makes nonlinear regression via gradient-based optimization cost prohibitive. Thus, we propose Gaussian process Bayesian optimization (GPBO) to calibrate parameters in computationally expensive simulation-based models, especially molecular modeling force fields. These force fields represent the intramolecular and intermolecular energies of a system and are useful in process and material design when calibrated effectively. We demonstrate the use of three formulations of GPBO and compare them to a traditional nonlinear regression approach. The first method uses a GP to model the (logarithmic) error between simulation predictions and experimental results. The second method uses a GP to emulate the input-output mapping of the expensive simulation-based model using an independence approximation to compute expected improvement. The third method replaces this approximation with a sparse grid to estimate the integral used to compute expected improvement. We find the second and third formulations capture the overall model behavior best for parameter calibration while the first formulation produces the highest accuracy and computational speed on a set of test problems. Future work aims to benchmark performance for all formulations in higher dimensions and apply them to real refrigerant force fields.

Montana N. Carlozo, Bridgette J. Befort, Nilay Kumar, Edward Maginn, Alexander Dowling
University of Notre Dame
mcarlozo@nd.edu, bbefort@nd.edu, nkumar@nd.edu, ed@nd.edu, adowling@nd.edu

CP5

Constrained Bayesian Optimization with an Exact Augmented Lagrangian

Bayesian optimization is a popular method for efficiently finding minima of functions that are expensive to evaluate. Two major challenges for Bayesian optimizers are the curse of dimensionality and satisfying nonlinear constraints. The curse of dimensionality can be mitigated for a Bayesian optimizer by using a gradient-enhanced Gaussian process, which is the approach used in this work. Various methods have been developed to apply a Bayesian optimizer to problems with nonlinear inequality constraints. Unfortunately, most of these methods do not allow for nonlinear equality constraints. A novel approach has been developed that utilizes an exact augmented Lagrangian to perform Bayesian optimization on problems with both nonlinear inequality and equality constraints. A traditional augmented Lagrangian uses Lagrange multipliers that are constant at each optimization iteration. Meanwhile, for an exact augmented Lagrangian, the Lagrange multipliers are free to vary in the parameter space. This is advantageous since it can accelerate the convergence to a local minimum. Furthermore, by having the Lagrange multipliers vary in the parameter space, this allows the Bayesian optimizer to consider a parameter space with several local minima. Test cases with nonlinear inequality and equality constraints will be presented.

Andre Marchildon, David Zingg
University of Toronto Institute for Aerospace Studies
andre.marchildon@mail.utoronto.ca,
dwz@utias.utoronto.ca

CP5

Deterministic Langevin Optimization with Normalizing Flows

We introduce a global, gradient-free surrogate optimization strategy for expensive black-box functions inspired by the Fokker-Planck and Langevin equations. These can be written as an optimization objective where one subtracts from the target function one is maximizing the logarithm of the current density of evaluated samples. This objective balances exploitation of the target objective with exploration of low-density regions. The method, Deterministic Langevin Optimization (DLO), relies on a Normalizing Flow density estimate to perform active learning and select proposal points for evaluation. This strategy differs qualitatively from the widely-used acquisition functions employed by Bayesian Optimization methods. We demonstrate competitive and superior progress toward objective optima on standard synthetic test functions, as well as on non-convex and multi-modal posteriors of moderate dimension. On real-world objectives, such as scientific and neural network hyperparameter optimization, DLO is competitive with state-of-the-art baselines.

James M. Sullivan
University of California, Berkeley
jmsullivan@berkeley.edu

Uros Seljak
University of California, Berkeley
U.S.
useljak@berkeley.edu

CP6

Gradient-Based Design Optimization of Concentric Cylindrical Offshore Structures

The optimal hydrodynamic behavior of floating offshore structures is important for engineering systems such as offshore wind turbines, wave energy converters, oil and gas platforms, and other marine devices. Such optimizations require repeatedly solving a boundary value problem for Laplace's equation to determine the hydrodynamic response to ocean waves. This is often done with numerical PDE solvers like the boundary element method, which is computationally costly and not analytically differentiable. In this work, we instead implement a previously-developed matched eigenfunction method to solve the diffraction and radiation problems analytically. We then integrate this solution into a novel gradient-based optimization using exact analytical gradients for the hydrodynamic response. Results are presented for the optimization of a compound truncated cylinder oscillating in heave. Finally, we compare the result and computational complexity of this novel matched-eigenfunction-based optimization with that of the existing state-of-the-art boundary element method-based optimization. The framework developed here can accelerate the development of offshore systems by enabling efficient integration of hydrodynamics into complex gradient-based optimization workflows.

Kapil Khanal, Rebecca McCabe, Maha N. Haji
Cornell University
kk733@cornell.edu, rgm222@cornell.edu,
maha@cornell.edu

CP6

Optimal Design of a Markovian (s, Q) Inventory System with Backorder

A Markovian (s, Q) inventory problem with backorder is investigated, where demand is a random variable over the infinite periods. When the inventory position falls below the reorder point s at any period, a fixed quantity of Q is ordered and the replenishment arrives after a constant lead time. Shortage is backordered and filled upon restocking. The reorder point and fixed quantity are to be determined to minimize the expected total relevant cost per unit time, which consists of ordering cost, inventory carrying cost and shortage cost. For the Markovian (s, Q) model where an order placed at the end of one period will arrive at the beginning of next period, an exact analysis is performed. Two properties have been established, and are used to develop an efficient heuristic for finding the optimal (s, Q). An approximate analysis is then performed for the extended model with lead time greater than one. A 2-D search procedure is developed to find the optimal or near optimal policy. Computational results have shown that the probability of finding the optimum policy is approximately 81.7%, and the average deviation of expected cost from those in the local enumeration is less than 0.7%. It also indicates that the heuristic is robust and efficient.

Shine-Der Lee, Zih-Ling Liu
National Cheng Kung University
Dept. of Industrial & Information Management
sdlee@mail.ncku.edu.tw, z8108151@email.ncku.edu.tw

CP6

Mitigating Landside Congestion at Airports Through Predictive Control of Diversionary Messages

We present a data-driven control framework for adaptively managing landside congestion at airports. Ground traffic significantly impacts airport operations and critical efficiency, environmental, and safety metrics. Our framework models a real-world traffic intervention currently deployed at Seattle-Tacoma International Airport (SEA), where a digital signboard recommends drivers to divert to Departures or Arrivals depending on current congestion. We use measured vehicle flow/speed and passenger volume data, as well as time-stamped records of diversionary messages, to build a macroscopic system dynamics model. We then design a model predictive controller that uses our estimated dynamics to recommend diversions that optimize for overall congestion. Finally, we evaluate our approach on 50 real-world historical scenarios at SEA where no diversions were deployed despite significant congestion. Our results suggest that our framework would have improved speed in the congested roadway by up to three times and saved between 20 and 80 vehicle-hours of cumulative travel time in every hour of deployment. Overall, our work emphasizes the opportunity of algorithmic decision- making to augment operator judgment and intuition and yield better real-world outcomes

Nawaf Nazir

Pacific Northwest National Laboratory
nawaf.nazir@pnnl.gov

Shushman Choudhury
Lacuna Technologies
shushman.choudhury@lacuna.ai

Soumya Vasisht
Pacific Northwest National Laboratory
soumya.vasisht

Stephen Zoepf
Lacuna Technologies
stephen.zoepf@lacuna.ai

Chase Dowling
Pacific Northwest National Laboratory
chase.dowling@pnnl.gov

CP6

A MINLP Formulation for Stringer Centerline Design

Fuselage stringers are load-bearing components that run lengthwise along an aircraft and transfer aerodynamic loads acting on the skin into the stringers and frames. The design of stringer centerlines in commercial airplanes has traditionally been performed manually by structural engineers as stringer configurations are subject to a wide variety of design and integration requirements. These requirements include minimum and maximum spacing between pairs of centerlines, maximum area constraints on frame bays, and a host of integration requirements with additional features in the fuselage. This presentation will describe the problem formulation in a tool called ASCEND (Automated Stringer Centerline Design) which automates the generation of stringer centerline designs in the aft body of a vehicle. We formulate the centerline design problem as a mixed-integer nonlinear program (MINLP) where discrete variables assign stringers to drop out of the design at specific frame stations and continuous variables control the path each stringer takes on the fuselage surface. We will highlight how this design space may be efficiently enumerated by checking the feasibility of a number of linear programs representing relaxations of the stringer spacing constraints. This enumeration is incorporated into a branch and bound algorithm that solves the full centerline design problem to global optimality.

Jeffrey D. Poskin, Troy Winfree
The Boeing Company
jdposkin@gmail.com, troy.winfree@boeing.com

CP6

Retrospective Analysis of Equity-Based Optimization for COVID-19 Vaccine Allocation

There are significant disparities in COVID-19 infection, hospitalization, and death rates between different racial or ethnic groups. To study this inequity, we construct a mathematical model of SARS-CoV-2 transmission that is stratified by both age and race. We consider two racial groups: non-Hispanic white persons and persons belonging to minority groups (including non-Hispanic black or African American persons, non-Hispanic Asian persons, non-Hispanic American Indian or Alaska Native persons, and Hispanic or Latino persons). After calibrating our model according to data from Oregon at the beginning of 2021, we allocate a very limited amount of vaccine among the age and racial groups to minimize overall disease burden (deaths or years of life lost), inequity in disease outcomes between racial groups, or both. Older age groups, who are at a greater risk of severe disease and death, are prioritized when minimizing measures of disease burden, and younger marginalized groups, who face the most inequities, are prioritized when minimizing measures of inequity. We find that, when allocating small amounts of vaccine, there is a trade-off between minimizing overall disease burden and minimizing inequity. The allocation strategies that minimize combinations of measures can produce middle-ground solutions that similarly improve both overall mortality and inequity, but the trade-off can only be mitigated by increasing the vaccine supply.

Erin Stafford
University of Washington
Department of Applied Mathematics
estaff@uw.edu

Dobromir Dimitrov
Fred Hutchinson Cancer Center
Vaccine and Infectious Disease Division
ddimitro@scharp.org

Rachel Cellabos
Fred Hutchinson Cancer Center
Public Health Sciences Division
rceballo@fredhutch.org

Laura Matrajt
Fred Hutchinson Cancer Center
Vaccine and Infectious Disease Division
laurama@fredhutch.org

CP7

A Unified Framework for Optimal Design of Experiment for Science and Engineering Grey-Box Models

Automation and optimization of engineering systems require careful design of experiments (DoE) to minimize the time-, material-, and labor costs of gathering data. In model-based DoE (MBDoE), design space samples are sequences of experimental conditions selected to maximize information gain, improve parameter precision, or discriminate between rival candidate models using Fisher information-based metrics such as A-, D-, and E- optimality. Yet, MBDoE methods assume the model structure is correct, and are thus inappropriate for DoE under epistemic (i.e., model-form) uncertainty. In contrast, adaptive design of experiments (ADoE) considers black-box models trained from data. In recent years, Bayesian optimization (BO) for ADoE has emerged, which uses probabilistic black-box surrogate models for efficient global optimization. In BO, an acquisition function determines the location of the next experiment by balancing the tradeoff between exploration (sampling in areas with high uncertainty) and exploitation (sampling based on distance from the previous point). At the intersection of white and black-box modeling paradigms are grey-box models which augment white-box models with black-box discrepancy terms. In this work, we present a unifying DoE framework for such models by synergizing MBDoE and BO for ADoE. Moreover, we benchmark the performance of popular acquisition functions in BO literature and demonstrate how these models enable limited-data optimal DoE under uncertainty.

Kyla D. Jones
University of Notre Dame
kjones29@nd.edu

CP7

A novel algorithm for a broad class of nonconvex optimization problems

In this paper, we propose a new global optimization approach for solving nonconvex optimization problems in which the nonconvex components are sums of products of convex functions. A broad class of nonconvex problems can be written in this way, such as concave minimization problems, difference of convex problems, and fractional optimization problems. Our approach exploits two techniques: first, we introduce a new technique, called the Reformulation-Perspectification Technique (RPT), to obtain a convex approximation of the considered nonconvex continuous optimization problem. Next, we develop a spatial Branch and Bound scheme, leveraging RPT, to obtain a global optimal solution. Numerical experiments on four different convex maximization problems, a quadratic constrained quadratic programming problem, and a dike height optimization problem demonstrate the effectiveness of the proposed approach. In particular, our approach solves more instances to global optimality for the considered problems than BARON and SCIP. Moreover, for problem instances that can be solved by our approach, BARON, and SCIP, for instances of small dimension BARON overall performs better on computation time, while for problem instances of larger dimension our approach outperforms both BARON and SCIP on computation time for most problem instances.

Theodoros Koukouvinos
Massachussets Institute of Technology
tkoukouv@mit.edu

Dimitris Bertsimas
Massachusetts Insitute of Technology
dbertsim@mit.edu

Dick den Hertog, Danique de Moor
University of Amsterdam
d.denhertog@uva.nl, d.demoor@uva.nl

Jianzhe Zhen
N.A.
trevorzhen@gmail.com

CP7

Niching Sparrow Search Algorithm for Solving Benchmark Optimization Problems, Speed Reducer Design, and Himmelblaus Nonlinear Optimization Problem

Metaheuristic algorithms are invented or modified in order to solve complex optimization problems at the global level. With the development of technology, almost every domain such as engineering, industrial, medical etcetera is facing the problem of optimization. In order to solve these problems, a number of algorithms have been discovered. One of the most recent optimization algorithms is the Sparrow Search Algorithm (SSA) which is famous for its good optimal ability along with fast convergence, Although, the SSA has a lot of merits, it is still facing numerous drawbacks namely falling into the local optima, steady convergence, etc. Therefore, we have proposed Niching SSA (NSSA) by introducing the niching technique in SSA for updating the position of scroungers and sparrows at the edge of the group. This NSSA has been tested on 18 benchmark functions, speed reducer design, and also on Himmelblaus nonlinear optimization problem. In this work, we have examined NSSA from various aspects like optimal value, average mean for convergence accuracy, the standard deviation for stability, and also have drawn the convergence curves through Matlab to check the convergence rate. Moreover, we have applied the Wilcoxon Signed rank test on NSSA. In all these aspects, computational results reveal that the performance of NSSA is superior with respect to SSA, GWO, PSO, and GSA.

Gagandeep Kaur Sidhu, Jatinder Kaur
Chandigarh University, Gharuan, Punjab
GAGAN.KW777@GMAIL.COM, jatindere1486@cumail.in

CP7

Some New Algorithms for Best Subset Selection in Linear Regression.

Based on the linear regression model $y = X\beta + \epsilon$ where $X \in \mathbb{R}^{n\times p}$ is the design matrix, $\beta \in \mathbb{R}^{p\times 1}$ is the coefficient vector and $y, \epsilon \in \mathbb{R}^{n\times 1}$ are the response and noise vectors respectively, we consider the best subset selection (BSS) problem of selecting a subset of k predictors from the available p predictors which minimizes the residual sum of squares $\|y - X\beta\|^2$. The discrete nature of the problem makes it of non-polynomial complexity. BSS problem can be formulated as a cardinality constrained quadratic minimization problem, which is a concave problem of the mixed integer type. In this work a new approach based on the well-known framework of interval branch and bound (IBB) method has been investigated. IBB has been successfully used for global continuous optimization over a fixed box domain. The proposed method finds an optimal solution by going over all the subsets systematically but not exhaustively, and employ certain deletion conditions to discard the subsets which cannot contain an optimal solution without explicitly exploring them. For comparison, a few existing algorithms are adopted in the study and additional sub-optimal techniques are also proposed for solving the BSS problem at a lower computational cost. The performance of these methods has been compared using a variety of test data sets.

Vikram Singh, Min Sun
The University of Alabama
vsingh12@crimson.ua.edu, msun@ua.edu

CP8

Cost Optimization and Analysis of Multi-Server Queue with Working Vacation and Imperfect Service

In the proposed model, the cost analysis and optimization of a Markovian queueing system with two-stage service, imperfect service during working vacations, and a second optional service with multiple servers are investigated in this study. This research dealt with a general cost optimization queueing model for retail stores that operate in a hybrid mode with numerous semi-attended checkout counters. Matrix geometric method is used to analyze the model and various performance indices have also been derived. Sensitivity analysis has been derived to validate the proposed model. The particle swarm optimization technique is employed to investigate cost minimization. This study helps the decision makers to improve the service quality with the help of optimal design of system descriptors.

Anshul Kumar, Madhu Jain
Indian Institute of Technology Roorkee
akumar@ma.iitr.ac.in, madhu.jain@ma.iitr.ac.in

CP8

Optimal Control and Application to Reduction of Uhi Intensity in Porous City

The architecture and urbanization of cities have a very important role in preventing overheating, and the generation of urban heat islands. The contribution of nature in the city is certainly one of the most interesting solutions to reduce the effects of urban heat island, but also for better management of the water cycle, while making the interurban more pleasant. The work is intended as an attempt to explore the use of the optimal control techniques to design green spaces and how to deal with the environmental problems related to the urban heat island appearing in cities. A porous media-based model (Navier-Stokes-Forschheimer-Fourier system) is developed in order to analyze the mitigation of the urban heat island effect. The focus was placed on presenting the applicability and relevance of a better combination between optimization techniques, and optimal control theory of partial differential equations to ensure the comfort in urban areas. To validate the method, we implement a numerical scheme, based on a finite element method, Spectral Projected Gradient algorithm and employing the free software FreeFem++. The quality of the results with a realistic model assesses the efficiency and the robustness of the approach.

Mohammed Louaked
Caen University
mohammed.louaked@unicaen.fr

Houari Mechkour
Ecole Centrale dElectronique (ECE-Paris)
houari.mechkour@ece.fr

Nacer sellila
Caen University
nacer.sellila@unicaen.fr

CP8

Methodology to Generate Bounded Solutions for Hard Knapsack Problems

In this talk, we will present a simple methodology that iteratively solves Combinatorial Optimization problems using commercial software to generate solutions that are within a tight tolerance of the optimums, using a reasonable amount of computer time. This robust procedure allows the user to specify two sequences of tolerance and maximum execution time. We will apply this methodology to solve a series of NP hard knapsack problems including MKP, MMKP and MDMKP.

Yun Lu
Kutztown University
lu@kutztown.edu

CP8

Iterative Local Methods Based on Second-Order Cone Programming for Ship Routing Problem with Two Drones

In this talk, we consider a routing problem for a cooperative delivery system that consists of one ship and two drones. When take-off and landing positions are continuous variables, such a problem can be formulated as a mixed-integer second-order cone programming problems. In particular, the computational difficulty is from the combination in the flights of the two drones. We propose local iterative methods to obtain a favorable solution in a short time using solutions of second-order cone programming problems.

Ningwen Yang
Tokyo Institute of Technology
yang.n.aa@m.titech.ac.jp

Makoto Yamashita
Tokyo Institute of Technology
Dept. of Mathematical and Computing Science
makoto.yamashita@c.titech.ac.jp

CP9

On Solving Retail Portfolio Maximization Problems Using Constrained Active Signature Methods

The retail industry is governed by crucial decisions on inventory management, discount offers (promotions), and stock clearing (markdowns). These present two sets of optimization problems. The former is an estimation problem, where the underlying objective is to predict the coefficients of demand elasticity with respect to product prices. The latter is the dynamic revenue maximization problem, that takes in these coefficients as inputs. While both present nonsmooth optimization problems, the latter is a challenging nonlinear problem in massive dimensions. This is further subject to constraints on inventory, inter-product relationships, and price bounds. Traditional approaches to solve such problems relied on using reformulations by introducing additional binary variables. In this work, we retain the nonsmooth structure generated by the absolute value function and sightly reformulate the problem(s) to a abs-quadratic form, i.e., in a quadratic matrix-vector-product based representation. Subsequently, we present an adaptation of the Constrained Active Signature Method (CASM) that explicitly exploits the abs-quadratic structure. In process, we also guarantee convexity of the objectives under some mild realistic assumptions on the market demand and structure. Two real world retail examples (UK and US market data from 2017-2019) are taken as case studies. Numerical results show promising improvement in computational time, in comparison to binary programming solvers.

Timo Kreimeier
Humboldt-Universität zu Berlin
Timo.Kreimeier@hu-berlin.de

Aswin Kannan
Humboldt University
aswin.kannan@hu-berlin.de

Andrea Walther
Humboldt-Universität zu Berlin
andrea.walther@math.hu-berlin.de

CP9

Numerical Method for Approximately Optimal Solutions of Two-Stage Distributionally Robust Op-

timization with Marginal Constraints

We consider a general class of two-stage distributionally robust optimization (DRO) problems which includes prominent instances such as task scheduling, the assemble-to-order system, and supply chain network design. The ambiguity set is constrained by fixed marginal distributions that are not necessarily discrete. We derive a relaxation of the problem via replacing the marginal constraints with finitely many linear constraints where we can control the relaxation error to be arbitrarily small. We then develop duality results and transform the inf-sup problem into an inf-inf problem. This leads to a numerical algorithm which solves a linear semi-infinite optimization problem. Besides an approximately optimal solution, the algorithm computes both an upper bound and a lower bound for the optimal value of the problem. The difference between the computed bounds provides a direct sub-optimality estimate of the computed solution. Most importantly, one can choose the inputs of the algorithm such that the sub-optimality is controlled to be arbitrarily small. We demonstrate the proposed algorithm by applying it to a supply chain network design problem where the ambiguity set involves a large number of discrete and continuous marginals. The numerical results showcase that the proposed algorithm computes high-quality robust decisions along with their corresponding sub-optimality estimates with practically reasonable magnitudes that are not over-conservative.

Qikun Xiang
School of Physical and Mathematical Sciences
Nanyang Technological University
qikun001@e.ntu.edu.sg

Ariel Neufeld
Nanyang Technological University
Singapore
ariel.neufeld@ntu.edu.sg

CP9

Robust Optimization with Continuous Decision-Dependent Uncertainty

We consider a robust optimization problem with continuous decision-dependent uncertainty (RO-CDDU), which has two new features: an uncertainty set linearly dependent on continuous decision variables and a convex piecewise-linear objective function. We prove that RO-CDDU is NP-hard in general and reformulate it into an equivalent mixed-integer nonlinear program (MINLP) with a decomposable structure to address the computational challenges. Such an MINLP model can be further transformed into a mixed-integer linear program (MILP) given the uncertainty set's extreme points. We propose an alternating direction algorithm and a column generation algorithm for RO-CDDU. We model a robust demand response (DR) management problem in electricity markets as RO-CDDU, where electricity demand reduction from users is uncertain and depends on the DR planning decision. Extensive computational results demonstrate the promising performance of the proposed algorithms in both speed and solution quality. The results also shed light on how different magnitudes of decision-dependent uncertainty affect the demand response decision.

Haoxiang Yang
Los Alamos National Laboratory
yanghaoxiang@cuhk.edu.cn

Hongfan Chen
The Chinese University of Hong Kong
kevinchen@cuhk.edu.hk

Andy Sun
Operations Research Center
Massachusetts Institute of Technology
sunx@mit.edu

CP9

Economic Capital Based Portfolio Management with Robust Optimization

Economic capital based decision-making is currently experiencing a revival in banking and business investment. Risk Adjusted return on capital (RAROC) is bound to play an important role in financial decision-making, not only for regulatory-capital requirements and performance tracking of business units but also as a risk metric in pricing and transaction decision-making. In this paper, we study portfolio management by maximizing the RAROC of the portfolio. Robust techniques are used to handle the uncertainty of the return, and the optimal portfolio can be selected by solving a second-order cone optimization problem or a linear optimization when the economic capital is measured by VaR or CVaR. Multi-period investment problem is also considered. Numerical tests on the market data show that the investment with RAROC strategies has better performance than that based on the Sharpe ratio and Hangseng index.

Hongxia Yin
Minnesota State University Mankato
hongxia.yin@mnsu.edu

Daoyu Wu
Beijing Shixinronghe Investment Ltd., Beijing, China
wudaoyu05@mails.ucas.edu.cn

CP10

The Emergence of League and Sub-League Structure in the Population Lotto Game

In order to understand if and how strategic resource allocation can constrain the structure of pair-wise competition outcomes in competitive human competitions we introduce a new multiplayer resource allocation game, the Population Lotto Game. This new game allows agents to allocate their resources across a continuum of possible specializations. While this game allows non-transitive cycles between players, we show that the Nash equilibrium of the game also forms a hierarchical structure between discrete 'leagues' based on their different resource budgets, with potential sub-league structure and/or non-transitive cycles inside individual leagues. We provide an algorithm that can find a particular Nash equilibrium for any finite set of discrete sub-population sizes and budgets. Further, our algorithm finds the unique Nash equilibrium that remains stable for the subset of players with budgets below any threshold.

Joel D. Nishimura
Arizona State University
joel.nishimura@asu.edu

Giovanni Artiglio
University of Wisconsin-Madison

gartigli@asu.edu

CP10

Combatting Gerrymandering with Social Choice: the Design of Multi-Member Districts

Every representative democracy must specify a mechanism under which voters choose their representatives. The most common mechanism in the United States – Winner takes all single-member districts – both enables substantial partisan gerrymandering and constrains 'fair' redistricting, preventing proportional representation in legislatures. We study the design of multi-member districts (MMDs), in which each district elects multiple representatives, potentially through a non-Winner takes all voting rule. We carry out large-scale empirical analyses for the U.S. House of Representatives under MMDs with different social choice functions, under algorithmically generated maps optimized for either partisan benefit or proportionality. Doing so requires efficiently incorporating predicted partisan outcomes – under various multi-winner social choice functions – into an algorithm that optimizes over an ensemble of maps. We find that with three-member districts using Single Transferable Vote, fairness-minded independent commissions would be able to achieve proportional outcomes in every state up to rounding, and advantage-seeking partisans would have their power to gerrymander significantly curtailed. Simultaneously, such districts would preserve geographic cohesion, an arguably important aspect of representative democracies. In the process, we advance a rich research agenda at the intersection of social choice and computational gerrymandering.

David B. Shmoys
Cornell University
david.shmoys@cornell.edu

Nikhil Garg
School of Operations Research & Information Engineering
Cornell Tech and the Technion
ng343@cornell.edu

Wes Gurnee
Operations Research Center
MIT
wesg@mit.edu

David Rothschild
Microsoft Research
david@researchdmr.com

CP11

Minimizing a Quadratic over Stiefel Manifolds and Placement Applications

In this paper, we present one trust region method for the initialization scheme, which has applications in VLSI placement. The proposed algorithm is basically one sequential subspace method(SSM) for minimizing a quadratic over r-dimensional Stiefel manifolds, where r is 2 or 3. For each subproblem, we apply sequential quadratic programming to compute the local minimizer. In general, the local minimizer is not necessarily a global minimizer. When the system matrix has r identical smallest eigenvalues, the iterates of SSM actually converge to a global minimizer, whenever each SSM subspace contains the following $3r$ vectors: (i)r orthogonal unit vectors associated with the current iterate, (ii) r vectors corresponding to the gradient of the objective function at the current iterate, and (iii) r orthogonal unit eigenvectors associated with the smallest eigenvalue of the system matrix.

Pengwen Chen
National Chung Hsing University
pengwen@nchu.edu.tw

CP11

On a Sequence of Quasi-Nonexpansive Mappings in a Geodesic Space with Curvature Bounded Above

The purpose of this talk is to present $\triangle$-convergence and strong convergence theorems for quasi-nonexpansive sequences in the setting of a geodesic space with curvature bounded above by one. The results can be applied to the image recovery problem for a countable family of closed convex subsets of such spaces and also applied to the optimization problem for convex functions.

Pongsakorn Yotkaew, Nawarat Ekkarntrong
Khon Kaen University
pongyo@kku.ac.th, nawaek@kku.ac.th

Nuttapol Pakkaranang
King Mongkut's University of Technology Thonburi
nuttapol.pakka@gmail.com

Bancha Panyanak
Chiangmai University
bpanyanak@yahoo.com

CP12

Online Contextual Decision Making with a Smart Predict-Then-Optimize Method

We study an online contextual decision-making problem with resource constraints. At each time period, the decision-maker first predicts a reward vector and resource consumption matrix based on a given context vector and then makes a decision. The final goal is to maximize the summation of the reward and the utility from resource consumption, while satisfying the resource constraints. We propose an algorithm that mixes a prediction step based on the "Smart Predict-then-Optimize (SPO)" method with a dual update. The regret bound is sublinear and depends on the risk bounds of the surrogate loss used to learn the prediction model. Our algorithm and regret bounds apply to a general convex feasible region for the resource constraints, including both hard and soft constraints, and they apply to a wide class of prediction models in contrast to linear context or finite policy spaces.

Paul Grigas
University of California- Berkeley
pgrigas@berkeley.edu

Heyuan Liu
University of California, Berkeley
heyuan_liu@berkeley.edu

CP12

A Recursive Multilevel Algorithm for Deep Learning

Increasing the complexity of modern neural networks offers superior classification accuracy but makes the training computationally more expensive. Multilevel methods, traditionally applied in solving differential equations using a hierarchy of discretizations, offer the potential to reduce computational effort. In this talk, we present a recursive multilevel stochastic gradient descent algorithm that accelerates learning by a multilevel strategy. A gradient correction term is needed to establish first-order consistency. We discuss further conditions to ensure convergence of the method. To demonstrate the usefulness of our approach, we apply it to residual neural networks in image recognition using image resolution to establish a hierarchy. We construct neural networks of decreasing number of variables and corresponding prolongation and restriction operators. Numerical results are presented.

Isabel Jacob
Technical University of Darmstadt
jacob@mathematik.tu-darmstadt.de

Stefan Ulbrich
Technische Universitaet Darmstadt
Fachbereich Mathematik
ulbrich@mathematik.tu-darmstadt.de

CP12

Maximal Independent Set for Boolean Pattern Recognition

Logical Analysis of Data (henceforth, LAD) is a rule-based supervised learning methodology that utilizes mixed integer and linear programs (henceforth, MILP) to find (Pareto-)optimal logical patterns that distinguishes one type of data from the other(s). An independent set in a graph is a collection of vertices that are pairwise disjoint, and finding maximum independent set is a NP-complete problem in combinatorics. In this paper, we establish a relationship between a Boolean pattern for a specific decision and the maximum independent set in a graph under the homogeneity condition. However, we exploit a maximal independent set instead for efficiency, and generate new type of pattern containing the set. We demonstrate the efficiency and efficacy of proposed combinatorial algorithm through computational experiments on benchmark machine learning datasets. Comparing results obtained from the MILP approach for LAD pattern generation, our new method yields competitive testing accuracy, requires at least 2 orders of magnitude less time for pattern discovery, and generates fewer logical patterns. Our new approach not only provides a combinatorial and more efficient way of analyzing data but also sheds light on new ways to tackle a hard problem in combinatorics and optimization.

Dongwoo Kang, Hong Ryoo
Korea University
towarmer@korea.ac.kr, hsryoo@korea.ac.kr

CP12

Mixed-Integer Quadratic Optimization and Clustering for Semi-Supervised Support Vector Machines

Among the most famous algorithms for solving classification problems are support vector machines (SVMs), which find a separating hyperplane for a set of labeled points. Sometimes, however, labels are only available for a subset of points. Furthermore, this subset can be non-representative, e.g., due to self-selection in a survey. Semi-supervised SVMs tackle the setting of labeled and unlabeled points and can often improve the reliability of the results. Moreover, additional information on the cardinality of the classes can be available from undisclosed sources. We propose a mixed-integer quadratic programming (MIQP) formulation to consider the setting of labeled and unlabeled points as well as the total number of points in each class. Since the MIQP's computation overhead rapidly grows as the number of variables increases, we introduce a clustering approach to reduce the model's size. Our preliminary numerical tests show that this approach leads to a similar accuracy than the MIQP formulation but at much lower computation cost, thus allowing to solve larger problems. With respect to the original SVM formulation, we observe that our approach has even better accuracy for biased samples.

Maria Eduarda Pinheiro
Trier University
pinheiro@uni-trier.de

Martin Schmidt
Universität Trier
martin.schmidt@uni-trier.de

Jan Pablo Burgard
Economic and Social Statistics Department
University of Trier, Germany
burgardj@uni-trier.de

CP12

Model-Based Feature Selection for Neural Networks: A Mixed-Integer Programming Approach

n this work, we develop a novel input feature selection framework for ReLU-based deep neural networks (DNNs), which builds upon a mixed-integer optimization approach. While the method is generally applicable to various classification tasks, we focus on finding input features for image classification for clarity of presentation. The idea is to use a trained DNN, or an ensemble of trained DNNs, to identify the salient input features. The input feature selection is formulated as a sequence of mixed-integer linear programming (MILP) problems that find sets of sparse inputs that maximize the classification confidence of each category. These "inverse' problems are regularized by the number of inputs selected for each category and by distribution constraints. Numerical results on the well-known MNIST and FashionMNIST datasets show that the proposed input feature selection allows us to drastically reduce the size of the input to $\sim$15% while maintaining a good classification accuracy, dropping less than 5 percentage points. This allows us to design DNNs with significantly fewer connections, reducing computational effort and producing DNNs that are more robust toward adversarial attacks.

Shudian Zhao, Jan Kronqvist
KTH Royal Institute of Technology
shudian@kth.se, jankr@kth.se

Calvin Tsay
Imperial College, London
c.tsay@imperial.ac.uk

CP13

Networked Policy Gradient Play in Markov Potential Games

We propose a networked policy gradient play algorithm for solving Markov potential games. In a Markov game, each agent has a reward function that depends on the actions of all the agents and a common dynamic state. A differentiable Markov potential game admits a potential value function that has local gradients equal to the gradients of agents' local value functions. In the proposed algorithm, agents use parameterized policies that depend on the state and other agents' policies. Agents use stochastic gradients and local parameter values received from their neighbors to update their policies. We show that the joint policy parameters converge to a first-order stationary point of a Markov potential game in expectation for general action and state spaces. Numerical results on the lake game exemplify the convergence of the proposed method.

Sarper Aydin, Ceyhun Eksin
Texas A&M University
Industrial and Systems Engineering Department
sarper.aydin@tamu.edu, eksinc@tamu.edu

CP13

Computing Perfect Bayesian Equilibria: From a Mathematical Characterization to a Differentiable Path-Following Method

As a slight relaxation to the requirements of sequential equilibrium, the notion of perfect Bayesian equilibrium was established by demanding Kreps and Wilson's global sequential rationality, belief consistency and subgame perfection, which has broad applications in economic analysis. Nonetheless, the establishment lacks a practicable and effective formulation for computationally finding such an equilibrium. To address this concern, this paper develops a mathematical characterization of perfect Bayesian equilibrium. The characterization comes from a new consistency notion called self-strategy-independent consistency, which introduces an explicit system for determining the players' beliefs. A salient point of self-strategy-independent consistency is to characterize a player's belief at an information set as a probability distribution irrelevant to his own behavioral strategies but induced by his previous beliefs. As a result of the characterization, we attain a polynomial system as a necessary and sufficient condition for a perfect Bayesian equilibrium. An application of the characterization leads to a differentiable path-following method for computing perfect Bayesian equilibria. Numerical results further confirm the effectiveness and efficiency of the method.

Yiyin Cao, Chuangyin Dang
City University of Hong Kong
yiyincao2-c@my.cityu.edu.hk, mecdang@cityu.edu.hk

CP13

A Polynomial-System Characterization of Nash Equilibrium in Behavioral Strategies and Its Computational Applications

As one of the most important and elegant ideas in game theory, Nash equilibrium was formulated by Nash (1951). Nonetheless, due to global rationality in the formulation, the existing approaches have to invoke a normal-form or sequence-form representation for finding a Nash equilibrium in behavioral strategies. To directly compute such an equilibrium, we develop in this paper an equivalent definition of Nash equilibrium in behavioral strategies of an extensive-form game with perfect recall. The development exploits an auxiliary behavioral strategy profile and a self-independent belief system to acquire global rationality from local rationality through the conditional expected payoff on every information set. The equivalent definition naturally leads to a polynomial system as a necessary and sufficient condition for a Nash equilibrium. A restriction of the equivalent definition on every subgame yields an equivalent definition of subgame perfect equilibrium in behavioral strategies and a polynomial system as a sufficient and necessary condition for a subgame perfect equilibrium. Moreover, we establish with the equivalent definitions the existence of both a smooth path to a Nash equilibrium and a smooth path to a subgame perfect equilibrium and secure two differentiable path-following methods. Comprehensive numerical experiments are presented to further verify the effectiveness and efficiency of the methods.

Chuangyin Dang, Yiyin Cao
City University of Hong Kong
mecdang@cityu.edu.hk, yiyincao2-c@my.cityu.edu.hk

CP14

A Variational Calculus for Optimal Control of Networks of Scalar Conservation Laws

Networks of scalar conservation laws are useful models for applications such as vehicular traffic flow, supply chains or transmission of data. Such networks consist of initial boundary value problems of scalar conservation laws on every edge which are coupled by node conditions. For optimal control of these networks a variational calculus is desirable that implies the differentiability of objective functionals with respect to controls. In the last decade research on initial boundary value problems of scalar conservation laws successfully introduced a variational calculus based on the concept of shift differentiability which implies differentiability of objective functionals of tracking type and yields an adjoint based gradient representation. This talk will present recent progress in the development of a variational calculus for networks of scalar conservation laws based on the concept of shift-differentiability. Regarding the node conditions we will introduce a framework for their representation and analysis which has proven compatible with the known approach of shift-differentiability on a single edge. In this framework we are able to extend results of the initial boundary value problem such as continuous Frchet differentiability and an adjoint based gradient representation of particular functionals of the control-to-solution operator on the network.

Marcel M. Steinhardt
Technical University Darmstadt
steinhardt@mathematik.tu-darmstadt.de

Stefan Ulbrich
Technische Universitaet Darmstadt
Fachbereich Mathematik
ulbrich@mathematik.tu-darmstadt.de

CP14

Design Optimization via the Continuous Stochastic Gradient Method

We consider large scale optimization problems with an objective function depending on random or distributed parameters. Here, standard deterministic optimization approaches rely on a discretization of the appearing integrals. Thus, the underlying quadrature rule can introduce artificial local minima into the objective function, resulting in an overall poor performance of the optimizer. Since circumventing this effect severely increases the computational cost, stochastic gradient type methods have gained popularity over the years. However, standard schemes from literature are typically limited to expected loss functions

and overall still require many iterations, i.e., state equation solutions. Thus, we present the continuous stochastic gradient method (CSG), which provides an efficient hybrid approach, without these limitations. In CSG, samples calculated in previous iterations are collected in an optimal linear combination to obtain an approximation to the full gradient and objective function value. It can be shown that the approximation error for both of these quantities vanishes during the optimization process. Therefore, CSG inherits many convergence properties known from full gradient methods, like convergence for constant step sizes or even step sizes obtained by an Armijo-type line search, based on the CSG approximations. After detailing the main theoretical properties of CSG, its efficiency is demonstrated for several applications from nanoparticle optics.

Andrian Uihlein
Friedrich-Alexander-Universität Erlangen-Nürnberg
andrian.uihlein@fau.de

Lukas Pflug
FAU Erlangen-Nuremberg
lukas.pflug@fau.de

Michael Stingl
Friedrich-Alexander-Universität Erlangen-Nürnberg
michael.stingl@fau.de

CP14

Computing P-Harmonic Descent Directions and Their Limits for Shape Optimization

Shape optimization constrained to partial differential equations is a vivid field of research with high relevance for industrial grade applications. Recent development suggests that using a p-harmonic approach to determine descent directions is superior to classical Hilbert space methods, but features the solution of a vector-valued p-Laplace problem with a boundary force in each iteration. We present the extension of an algorithm for scalar Dirichlet problems to solve these problems with the advantage of not requiring an iteration over the order p and thus compute higher-order solutions efficiently. Results are verified by numerical experiments in a fluid dynamic setting. A general requirement on the transformations of the computational domain is to keep it of Lipschitz type. While solutions for finite p yield approximations in $W^{1,p}$, analytically only descent directions in $W^{1,\infty}$ are admissible. However, this is challenging since the limit of the p-Laplace problem features in general non-unique solutions, in particular arising from a change of sign in the force term required to fulfill geometric constraints in shape optimization. Therefore, we make progress towards an algorithm for computing admissible descent directions, that are the limit of the corresponding p-harmonic descent directions. The resulting deformations then still preserve the quality of the underlying mesh, which is crucial for applications.

Henrik Wyschka
Universität Hamburg
henrik.wyschka@uni-hamburg.de

Martin Siebenborn
University Hamburg
Mathematics Department
martin.siebenborn@uni-hamburg.de

Winnifried Wollner
University of Hamburg
Department of Mathematics
winnifried.wollner@uni-hamburg.de

CP14

Binary Level-Set Method for Variational Implicit Solvation Model

The Variational Implicit Solvent Model (VISM) is a theoretical and computational tool to study biomolecular systems with complex topology. Central in VISM is an effective free-energy of all possible interfaces separating solutes (e.g., proteins) from solvent (e.g., water). Such a functional can be minimized numerically by a level-set method to determine the stable equilibrium conformations and solvation free energies. However, the PDE-based approach is time-consuming.We developed a discrete formulation of the problem using the binary level set method. The interface is approximated by a binary level set function that takes value ± 1 on the solute/solvent region. The VISM energy can be minimized by iteratively flipping the binary level set function. The new method is fast enough to be coupled with the Monte Carlo method for biomolecular simulations. We also show that with the discrete formulation, the VISM energy functional can be globally minimized by min-cut max-flow algorithms.

Zirui Zhang, Li-Tien Cheng
University of California, San Diego
ziruz16@uci.edu, l3cheng@ucsd.edu

CP15

Lattices Enumeration via Linear Programming

Given a positive integer d and $\mathbf{a}_1, \ldots, \mathbf{a}_r$ vectors in $\mathbb{R}^d$, $\{k_1\mathbf{a}_1 + \cdots + k_r\mathbf{a}_r : k_1, \ldots, k_r \in \mathbb{Z}\} \subset \mathbb{R}^d$ is the so-called lattice generated by the family of vectors, or by the matrix $\mathbf{A} = (\mathbf{a}_1|\ldots|\mathbf{a}_r) \in \mathbb{R}^{d\times r}$. In high dimensional integration, prescribed lattices are used for constructing reliable quadrature schemes. The quadrature points are the lattice points lying on the integration domain, typically the unit hypercube $[0,1)^d$ or a shifted hypercube. It is crucial to be able to enumerate the lattice points in such domains inexpensively. Undeniably, the lack of fast enumeration procedures hinders the applicability of lattice rules. Existing enumeration procedures exploit intrinsic properties of the lattice at hand, such as $\mathbb{Z}$-periodicity, orthogonality, recurrences, etc. We present a general-purpose lattice enumeration strategies based on linear programming. We demonstrate how to combine duality and parametric linear programming in order to accelerate these strategies, producing performances comparable to the enumeration strategies that are fine-tuned to special lattices. In addition, we discuss a variety of relaxation and reduction techniques that allow further acceleration of the introduced algorithms. Numerical experiments in high dimension are also presented.

Abdellah Chkifa
Mohamed VI Polytechnic University
abdellah.chkifa@um6p.ma

CP15

Quadratic Program Networks

A Quadratic Program Network (QPNet) is a collection of interdependent QPs to be solved jointly. Well-known problem classes which can be described as QPNets include bilevel QPs and linear-quadratic Nash equilibrium problems. However, these instances are among the simplest QPNets conceivable, typically comprised of few QPs and in basic network patterns. In this work, a theory is presented describing general QPNets, such as those comprised of many QPs organized in complex, multi-level networks. It is argued that this generalization is not merely for theoretical interest but to provide a natural framework for describing important real-world problems. To this end, an open-source solver is developed to compute local solutions to QPNets, the capabilities of which are demonstrated on interesting example instances.

Forrest Laine
Vanderbilt University
forrest.laine@vanderbilt.edu

CP15

Random Block Coordinate Descent Methods for Computing Optimal Transportation Problems

The optimal transport problems are reduced to a linear programming (LP) problem through discretization. In this paper, we introduced the random block coordinate descent (RBCD) methods to directly solve this LP. In each iteration, we restrict the potential large-scale optimization problem to small LP subproblems constructed via randomly chosen working sets. Using a random Gauss-Southwell-q rule to select these working sets, we equip the RBCD method with almost sure convergence and linear convergence rate. By further exploring the special structure of constraints in the optimal transport problems, we proposed several approaches to refine the random working set selection and accelerate the RBCD method. Preliminary numerical experiments demonstrate several merits of the accelerated random block coordinate descent (ARBCD) over the Sinkhorn method, in terms of solution accuracy and numerical stability.

Yue Xie
Pennsylvania State University
yxie21@hku.hk

CP15

On Facet-Pivot Method for Linear Programming

The simplex method for linear programming (LP) has been extensively studied for many decades. Currently, almost all pivot rules in LP algorithms are vertex pivot, which searches from one vertex to the next vertex along an edge of the convex polytope. In this paper, we propose a facet pivot method that updates a facet in every iteration by crossing a ridge of the convex polytope. The main advantage of using the facet pivot rules is that the method is more suitable for the general LP problem with equality, inequality, and boundary constraints without introducing relaxation variables. Our preliminary numerical experience shows that the facet pivot method is very promising compared to the traditional vertex pivot method.

Yaguang Yang
NRC
yaguang.yang@nasa.gov

CP15

A Hybrid Interior-Point-Column-Generation Method for Discrete Optimal Transport Problems

Discrete Optimal Transport problems give rise to very large linear programs (LP) with a particular structure of the constraint matrix. In this talk we present a hybrid algorithm that mixes an interior point method (IPM) and column generation, specialized for the LP originating from the Kantorovich Optimal Transport problem. Knowing that optimal solutions of such problems display a high degree of sparsity, we propose a column-generation-like technique to force all intermediate iterates to be as sparse as possible. The algorithm is implemented nearly matrix-free. Indeed, most of the computations avoid forming the huge matrices involved and solve the Newton system using only a much smaller Schur complement of the normal equations. We prove theoretical results about the sparsity pattern of the optimal solution, exploiting the graph structure of the underlying problem. We use these results to mix iterative and direct linear solvers efficiently, in a way that avoids producing preconditioners or factorizations with excessive fill-in and at the same time guaranteeing a low number of conjugate gradient iterations. We compare the proposed method with two state-of-the-art solvers and show that it can compete with the best network optimization tools in terms of computational time and memory usage. We perform experiments with problems reaching more than a billion variables and demonstrate the robustness of the proposed method.

Filippo Zanetti
University of Edinburgh
f.zanetti@sms.ed.ac.uk

Jacek Gondzio
The University of Edinburgh, United Kingdom
J.Gondzio@ed.ac.uk

CP16

Solution of Constrained Multi-Objective Optimization Problems Using Trust Region Technique

We propose a trust region based algorithm for constrained multi-objective optimization problems. At every iteration of the algorithm, a sub-problem is formulated using the quadratic approximation of all the objective functions and linear approximation of all the constraints at that iterating point. The solution to this sub-problem provides the descent direction. The step is evaluated using the notion of actual reduction and predicted reduction. A non-differentiable penalty function is used to handle the constraint violations. An adaptive BFGS update rule is introduced to update the matrix. The proposed method can adjust the trust region radius automatically at every iteration, which helps to reduce the number of iterations. In addition, a spreading technique is introduced to derive the well-spread Pareto front. Under some reasonable assumptions, the global convergence of the proposed algorithm is proved. Furthermore, the superlinear rate convergence property of the algorithm is established. Numerical results and comparisons with existing methods are provided using a set of test problems to show the efficiency of the proposed method, both in the quality of the approximated Pareto front and in the computational perspective.

Nantu K. Bisui
Research Scholar
nantukrbisuimath@gmail.com

Geetanjali Panda
Professor, Indian Institute of Technology Kharagpur

geetanjali@maths.iitkgp.ac.in

CP16

Multivariate Optimization for Sampling Instrumentation

In fields involving sampling, optimization of instrument parameters is a common practice. However, there is a gap between the current optimization methods used and modern, accurate optimization that should be applied. There is a need to employ a generalized strategy, suitable for almost any analytical instrumentation/sample prep/analysis scheme. The novelty of the proposed work is performing multivariate, multi-objective optimization with Karush-Kuhn-Tucker conditions to limit the optimization space to real-world or realistic solutions that are bounded by physical limitations of the instrument or process. In addition, we propose a novel concept for broadening the options for the quantities to be optimized, for example, the shape of the curve, height and/or width, rather than area. This will allow for more easily processed samples in computational analysis, statistical analysis, machine learning algorithms, etc. These methods applied to various analytical systems/samplers/sample preparation methods are relevant to government agencies, nuclear nonproliferation offices, as well as food industries. The result is a procedure to reduce time invested in optimizing analytical methods using traditional single variate trial-and-error optimization and potentially improve the sensitivity of current techniques. This methodology could serve to reduce the cost of developing and optimizing future systems and instruments.

Stephanie Gamble
Savannah River National Laboratory
stephanie.gamble@srnl.doe.gov

CP16

On the Set Order Topology and Its Application to Well-Posedness in Set-Valued Optimization

In this work, we introduce a topology on the power set $\mathcal{P}(Z)$ of a partially ordered normed space Z from which we derive a topological convergence on $\mathcal{P}(Z)$ along with new concepts of continuity and semicontinuity for set-valued mappings. Our goal is to propose a suitable framework in order to address set optimization problems involving set relations based on a cone ordering. Taking advantage of this new framework, we establish several results regarding the well-posedness of set-valued optimization problems that are consistent with the state of the art. We also exemplify the practicability of our theoretical results by establishing the well-posedness of the portfolio investment optimization problem without requiring a convexity property for the composite risk measure and with a not necessarily bounded set of minimal solutions.

James Larrouy, Michel H. Geoffroy
Université des Antilles
Department of Mathematics
james.larrouy@univ-antilles.fr, michel.geoffroy@univ-antilles.fr

CP16

C-Pearl: Multi-Objective Reinforcement Learning-Based Approach for Large-Scale Constrained Nuclear Reactor Loading Pattern Optimization.

The Constrained Pareto Envelope Augmented with Reinforcement Learning (C-PEARL) aims at solving constrained multi-objective problems, especially in the field of engineering for which candidate solutions are evaluated with an often-slow physic-based code. The Nuclear Reactor loading pattern optimization falls into this category. It belongs to the class of combinatorial optimization problem and is characterized by a remarkably large design space with many conflicting constraints and objectives. There, designers have stringent time constraints imposed by result-driven assignments which contrast to academia from which many of the popular algorithms originate. Hand-optimization in which the trade-off between solutions is evaluated thanks to the intuition of the designers is still the preferred method in the industry. Therefore, they would greatly benefit from methods offering a wealth of high quality solutions within a reasonable amount of time. C-PEARL was developed to tackle this issue and combined single-policy multi-objective-based Reinforcement Learning (RL) methods to draw the pareto envelope and curriculum learning to handle the constraints. The novelty lies in having RL agents learn not to generate unfeasible solutions while receiving rewards depending on the quality of the envelope directly to improve it. C-PEARL offers many promises to solve the problem at hand, and its application could seamlessly extend to any type of engineering design.

Paul Seurin
Massachusetts Institute of Technology (MIT)
paseurin@mit.edu

Koroush Shirvan
Massachusetts Institute of technology (MIT)
kshirvan@mit.edu

CP16

Nonsmooth Multiobjective Optimization Problems with Switching Constraints

we consider nonsmooth multiobjective optimization problems with switching constraints (NMOPSC). Further, we introduce the notion of Weak stationarity, Mordukhovich stationarity, and Strong stationarity, i.e., W-stationarity, M-stationarity, and S-stationarity, respectively, for the MOPSC and establish that the M-stationary conditions are sufficient optimality conditions for the MOPSC using generalized convexity. Further, we propose a Wolfe-type dual model for the NMOPSC and establish weak duality and strong duality results under generalized convexity assumptions.

Vinay Singh
National Institute of Technology, Mizoram
vinaybhu1981@gmail.com

CP17

Local Convergence Rates for Weakly Convex Douglas-Rachford Splitting and Admm via Descent of Proximal Merit Functions

We present a framework for analyzing local rates of convergence for nondescent methods in structured weakly convex optimization, such as the Douglas-Rachford splitting scheme and ADMM. The analysis is based on descent properties of a merit function that serves as a proxy of the true objective function, and a subdifferential-based error bound on the distance to critical points. This work corresponds to an extension of [F. Atenas, C. Sagastizbal, P. J. Silva, and M. Solodov. A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods. Accepted in SIOPT, 2022], in the sense that local rate of convergence properties for descent methods can be directly transferred to nondescent methods through an appropriate merit function defined using primal-dual information, for which sufficient descent along the generated iterates is guaranteed.

Felipe Atenas
University of Campinas
f262900@dac.unicamp.br

CP17

An Asynchronous Proximal-Subgradient Method for Solving Additive Convex Optimization Problems

In this work we consider additive convex optimization problems in which the objective function is the sum of a large number of convex nondifferentiable cost functions. We assume that each cost function is specifically written as the sum of two convex nondifferentiable function in which one function is appropriate for the subgradient method and the another one is suitably treated with the proximal iteration. To this end, we propose a distributed optimization algorithm based on the subgradient method and the proximal method. The proposed method is governed by asynchronous feature in which it allows time-varying delays when computing the subgradients. We prove the convergence of function values of iterates to the optimal value. To demonstrate of the efficiency of the presented theoretical result, we investigate the binary classification problem via the support vector machine learning.

Tipsuda Arunrat
Khon Kaen University
tipsudalake@gmail.com

Sakrapee Namsak
Vidyasirimedhi Institute of Science and Technology (VISTEC)
sakrapee_namsak@kkumail.com

Nimit Nimana
Khon Kaen University
nimitni@kku.ac.th

CP17

A Normal Map-Based Perspective on Second Order Theory for Composite Problems: Second Order Conditions, Metric Regularity, and Nonsingularity

Strong metric subregularity and strong metric regularity of the natural residual and the normal map are of particular importance in the convergence analysis of first-order and second-order algorithms. In this paper, we characterize the strong metric subregularity of the natural residual and the normal map for general nonsmooth nonconvex composite functions, and show the equivalence between these conditions, the strong metric subregularity of the subdifferential, and the quadratic growth condition of the objective function. For strong metric regularity of the subdifferential, we show that when the nonsmooth part is C^2-strictly decomposable with additional mild assumptions, then the strong metric regularity of the subdifferential is equivalent to those of natural residual and the normal map, and a counterpart of strong second-order sufficient condition in the composite optimization setting. Counterexamples illustrate the necessity of our assumptions.

Wenqing Ouyang
University of Science and Technology of China
220049034@link.cuhk.edu.cn

Andre Milzarek
The Chinese University of Hong Kong, Shenzhen
School of Data Science
andremilzarek@cuhk.edu.cn

CP17

A Dynamic Distributed Conjugate Gradient Method for Variational Inequality Problem over the Common Fixed-Point Constraints

In this paper, we propose a dynamic distributed conjugate gradient method for solving the strongly monotone variational inequality problem over the intersection of fixed-point sets of firmly nonexpansive operators. The proposed method allows the independent computation of a firmly nonexpansive operator along with the dynamic weight which is updated at each iteration. This strategy aims to speed up the convergence behavior of the algorithm by updating control factors to drive each iterative step. Under some suitable control conditions on corresponding parameters, we show a strong convergence of the iterate to the unique solution of the considered variational inequality problem. We consider the numerical experiments and discuss some observation points by applying the model to solve the image classification problem via the support vector machine learning.

Mootta Prangprakhon
Khon Kaen University
mootta_prangprakhon@hotmail.com

Narin Petrot, Porntip Promsinchai
Naresuan University
narinp@nu.ac.th, p.promsinchai@hotmail.com

Nimit Nimana
Khon Kaen University
nimitni@kku.ac.th

CP17

Minimizing the Difference of Convex and Weakly Convex Functions Via Bundle Method

We consider optimization problems with objective and constraints being the difference of convex and weakly convex functions. This framework allows covering a vast family of nonsmooth and nonconvex optimization problems, in particular, those involving Difference-of-Convex (DC) functions with known or unknown DC decomposition, functions whose gradient is Lipschitz continuous, as well as problems that comprise certain classes of composite and nonconvex value functions. We investigate several stationary conditions and present a proximal bundle method to compute critical points for problems of this class. Our algorithm, which employs an improvement function and an original rule to update the proximal parameter to ensure convergence, relies on cutting-plane approximations of the convex functions and linearizations of the weakly convex ones to construct a sequence of convex quadratic subproblems yielding new iterates. The practical performance of the method is illustrated by numerical experiments on some nonconvex stochastic problems.

Ksenia Syrtseva
Mines Paris, Universite PSL, CMA
EDF R&D
ksenia.syrtseva@minesparis.psl.eu

Sophie Demassey, Welington de Oliveira
Mines Paris, Universite PSL, CMA
sophie.demassey@minesparis.psl.eu,
welington.oliveira@minesparis.psl.eu

Wim van Ackooij
EDF R&D
wim.van-ackooij@edf.fr

CP18

gH-Dini Hadamard Subdifferential for Nonsmooth Interval Optimization Problem and Its Applications in Nonsmooth Control Systems

In this article, we investigate the notions of gH-Dini Hadamard subdifferential and gH-Dini Hadamard superdifferential for nonsmooth interval-valued functions (IVFs). For convex IVFs, the gH-Dini Hadamard subdifferential and the gH-subdifferential are found to coincide. The gH-Dini Hadamard subdifferential is proved to be convex and closed. Furthermore, we show that the gH-Dini Hadamard subdifferential obeys the one-sided sum rule and the partial chain rule. A criterion for an IVF to be gH-Dini Hadamard subdifferentiable is given. We also propose a smooth variational description of gH-Dini Hadamard subgradients, demonstrating that each gH-Dini Hadamard subgradient can be described equivalently by the gH-Hadamard derivative of a smooth IVF. We derive a Fritz-John-type necessary optimality condition and a KKT-type necessary optimality condition using the smooth variational description of gH-Dini Hadamard subgradients. To observe an application, we use the gH-Dini Hadamard subdifferential to investigate the asymptotic controllability of nonsmooth dynamical control systems. Finally, an example of an unstable nonsmooth control system is presented.

Suprova Ghosh
IIT-BHU, Varanasi
suprovaghosh.rs.mat19@itbhu.ac.in

CP18

Epsilon-Subdifferentiability for Interval-Valued Functions and Its Applications

A problem is called an interval optimization problem if it is associated with interval-valued function. Due to the uncertain nature of many real-world occurrences, the study of interval optimization problems has become a noteworthy research topic. However, there are many interval optimization problems whose exact solutions cannot be found. In this study, to find approximate solutions to such problems, we propose the notion of ϵ-subdifferentiability for convex interval-valued functions with the help of generalized Hukuhara difference. Several important characteristics of ϵ-subdifferential set, such as nonemptiness, closedness, convexity, boundedness, etc. are given in this study. Furthermore, a new solution concept, namely approximate solution or ϵ-solution, is introduced for an interval optimization problem. Using the proposed ϵ-subdifferentiability, we develop two necessary and sufficient optimality conditions to find an ϵ-solution for an unconstrained interval optimization problem. Lastly, a theorem has been given to solve interval minimax optimization problems using ϵ-subdifferentiability. Some numerical examples are also presented to support the whole study.

Krishan Kumar
IIT (BHU), Varanasi
krishankumar.rs.mat19@itbhu.ac.in

Debdas Ghosh
IIT (BHU) Varanasi
debdas.mat19@iitbhu.ac.in

CP18

Totally Relaxed Inertial Self-Adaptive Projection and Contraction Method for Solving Non-Lipschitz Variational Inequalities

Due to the difficulty posed in calculating projection onto the feasible set when solving variational inequality problem (VIP), several attempts are being made by researchers in recent times, to eliminate the need for an actual projection onto the feasible set, so as to ease computation. In this paper, we study the classical variational inequality problem defined on a finite intersection of sub-level sets of convex functions, for which the cost operator is non-Lipschitz (or uniformly) continuous. We propose a new iterative method, called the Totally relaxed inertial self-adaptive projection and contraction method (TRISPCM), in which the projection onto the feasible set is replaced with a projection onto some half-space. We adopt a self-adaptive step size, void of the line search procedure. We also utilize the inertial technique for achieving a faster convergence rate. Furthermore, under some mild assumptions we prove that the sequence generated by our proposed algorithm converges strongly to a minimum-norm solution of the problem. Finally, we conduct some numerical experiments to vividly exhibit the performance of our method in comparison with some of the existing methods in the literature. Our result in this paper extends and improves several of the existing results in the literature in this direction.

Victor A. Uzor
University of KwaZulu NAtal, South Africa
University of KwaZulu NAtal, South Africa
uzorvictor50@gmail.com

Timilehin Alakoya, oluwatosin mewomo
University of KwaZulu NAtal, South Africa
timimaths@gmail.com, mewomoo@ukzn.ac.za

Aviv Gibali
Braude College
P.O. Box 78, Karmiel 2161002, Israel
avivg@braude.ac.il

CP18

Moreau Envelope of Supremum Functions with Applications to Infinite and Stochastic Programming

In this talk, we discuss the Moreau envelope of the supremum of a family of convex, proper, and lower semicontinuous functions. Under mild assumptions, we prove that the Moreau envelope of a supremum is the supremum of Moreau envelopes, which allows us to approximate possibly nonsmooth supremum functions by smooth functions that are also the suprema of functions. Consequently, we propose and study approximated optimization problems from infinite and stochastic programming for which we obtain zero-duality results and optimality conditions without the verification of constraint qualification conditions.

Emilio Vilches
O'Higgins University
emilio.vilches@uoh.cl

Pedro Pérez-Aros
Universidad de O'Higgins
Universidad de O'Higgins
pedro.perez@uoh.cl

CP19

Optimized Implicit Runge-Kutta Method for Direct Integration of Third Order Boundary Value Problems

An Implicit Runge-Kutta method for the numerical approximation of the solution of general third order boundary value problems is derived via a "direct" method as those invented by Nystr6m. This method considers three intra-step points which are adequately selected so as to optimize the local truncation errors of the main formulas. The new method is zero-stable, consistent and convergent. Numerical examples from literature shows the efficiency of this method in terms of the global errors obtained.

Zamurat A. Adegboye
NIGERIAN DEFENCE ACADEMY
,KADUNA,NIGERIA
zamurat.adegboye@imsp-uac.org

CP19

Variational-Like Inclusion Problems Involving A-Maximal m-Relaxed η-Accretive Mappings

We introduce a new class of accretive mappings in the framework of q-uniformly smooth Banach spaces called A-maximal m-relaxed η-accretive mappins in the sense of semi-inner product of type (p) (SIPp)). The resolvent operator associated with such a mapping is defined; its Lipschitz continuity is proved and an estimate of its Lipschitz constant is computed under some suitable conditions. Generalized nearly asymptotically nonexpansive mappings is studied and the problem of finding a common element of the set of fixed points of a given generalized nearly asymptotically nonexpansive mapping and the set of solutions of a variational-like inclusion problem involving A-maximal m-relaxed η-accretive mappings in the sense of SIPp is investigated. To achieve this goal, a new iterative algorithm is constructed. Using graph convergence and resolvent operator associated with an A-maximal m-relaxed η-accretive mapping in the sense of SIPp, a new equivalence relationship between the graph convergence and the resolvent operator convergence of a sequence of A-maximal m-relaxed η-accretive mappings in the sense of SIPp is established. As an application of this equivalence, the strong convergence of the sequence generated by our proposed iterative algorithm to a common point of the above two sets is demonstrated under some suitable assumptions imposed on the parameters. The results presented in this paper unify, improve and generalize some recent works in this field.

Suliman Al-Homidan
King Fahd University of Petroleum & Minerals
P.O. Box 119 Dhahran 31261, Saudi Arabia
homidan@kfupm.edu.sa

CP19

Kinodynamic Control Systems and Discontinuities in Clearance

We investigate the structure of discontinuities in clearance (or minimum time) functions for nonlinear control systems with general, closed obstacles (or targets). We establish dynamic results regarding interactions between admissible trajectories and clearance discontinuities: e.g. instantaneous increases in clearance when passing through a discontinuity, and propagation of discontinuity along optimal trajectories. Then, investigating sufficient conditions for discontinuities, we explore a common directionality condition for velocities at a point, characterized by strict positivity of the minimal Hamiltonian. Elementary consequences of this common directionality assumption are explored before demonstrating how, in concert with corresponding obstacle configurations, it gives rise to clearance discontinuities both on the surface of the obstacle and propagating out into free space. Minimal assumptions are made on the structure of obstacle sets.

Jeremy LeCrone
University of Richmond
jlecrone@richmond.edu

Niles Armstrong
Milwaukee School of Engineering
armstrongn@msoe.edu

Jory Denny
Independent Researcher
joryldenny@gmail.com

CP19

On Optimistic, Pessimistic and Mixed Fuzzy-Programming Based Approaches to Solve Multi-Objective Fully Intuitionistic Fuzzy Linear Fractional Programming Problems

A decision-maker often encounters a state of uncertainty and hesitancy due to various unpredictable factors while solving real-world optimization problems. Due to ambiguity in the decision-making process, the parameters and decision variables of an optimization model generally fail to have the exact values. To deal with such realistic uncertain conditions, the concept of intuitionistic fuzzy numbers has been employed widely. This work presents a model of the linear fractional programming problem with multiple objectives having all the parameters and decision variables expressed as intuitionistic fuzzy numbers. To solve the problem, we first linearize the fractional model using appropriate transformations, and after that utilizing the accuracy function and intuitionistic fuzzy programming, the equivalent optimization model is obtained. The conflicting nature of the multiple objectives in the linearized model has been handled using the linear and exponential membership/non-membership functions applying normal, optimistic, pessimistic, and mixed approaches. Moreover, a practical application related to the production process in the textile industry has been constructed and solved using the proposed algorithm, and further, a comparison is drawn amidst various approaches to establish the efficacy of the proposed modelling and developed algorithm.

Manisha Malik, S. K. Gupta
Indian Institute of Technology Roorkee

mmalik@ma.iitr.ac.in, skgiitr@gmail.com

CP19

Continuous and Discrete Control of Potential Flow Networks

We consider a class of networks in which the potential across the edges drives the flow through them, and the potential loss increases non-linearly with the flow rate. In these systems, as the nodal demands are usually not proportional to the equilibrium flow rates, flow control elements are required to match supply to demand. The control elements can broadly be classified into two types discrete and continuous. Discrete control elements can have only two operational states: fully open or fully closed. On the other hand, continuous control elements may be operated in any intermediate position in addition to the fully open and fully closed states. Naturally, with their increased flexibility, continuous control elements can provide better performance. The less intuitive question is: *By what extent?* In the present work, we quantify the performance of networks based on the time required to transport a given quantum of material. We define R to be the ratio of minimal operational times with either type of control. In a class of networks with a single source and multiple sinks, and the potential loss across edges increases with flow rate raised to the power of n, we analytically show that $1 \leq R \leq m^{(1-1/n)}$, where m is the maximum depth of the network. The result highlights the role of network topology in the variations in operational time. An extension also throws light on the optimal and selfish operation of non-linear flow networks.

Varghese Kurian, Sajay Velmurugan
Department of Chemical Engineering, IIT Madras
varghese@alumni.iitm.ac.in, sajayv@alumni.iitm.ac.in

Sridharakumar Narasimhan
Department of Chemical Engineering,
Indian Institute of Technology Madras
sridharkrn@iitm.ac.in

CP20

Fair Maximal Covering Location Problems

We provide a general mathematical optimization based framework to incorporate fairness measures from the facilities' perspective to Discrete and Continuous Maximal Covering Location Problems. The main ingredients to construct a function measuring fairness in this problem are the use of: (1) ordered weighted averaging operators, a family of aggregation criteria very popular to solve multiobjective combinatorial optimization problems; and (2) α-fairness operators which allow to generalize most of the equity measures. A general mathematical optimization model is derived which captures the notion of fairness in maximal covering location problems. The models are firstly formulated as mixed integer non-linear optimization problems for both the discrete and the continuous location spaces. Suitable mixed integer second order cone optimization reformulations are derived using geometric properties of the problem. Finally, the paper concludes with the results obtained on an extensive battery of computational experiments on real datasets. The obtained results support the convenience of the proposed approach.

Victor Blanco
Universidad de Granada
vblanco@ugr.es

Ricardo Gázquez
Universidad de Málaga
rgazquez@ugr.es

CP20

Minimum Spanning Trees in Infinite Graphs: Theory and Algorithms

Consider the Minimum Spanning Tree (MST) problem on an undirected graph with countably many nodes of finite degree and absolutely summable costs on edges. Is an MST guaranteed to exist? Can one of the greedy algorithms used to find MSTs in finite graphs serve as a building block for an algorithm in the countably infinite setting? What assumptions are sufficient for this algorithm to converge to optimality in value? Moreover, when do the finite-sized iterate subgraphs of the algorithm converge to an MST of the original infinite graph, and what is the appropriate notion of convergence to use in the analysis? Finally, can we detect whether an edge included in an iterate subgraph is in fact an edge of an MST? We propose a layered greedy algorithm which proceeds by finding MSTs on nested finite subgraphs (called layers) using Prim's algorithm. If the infinite graph has an MST, this algorithm converges in objective value to optimality as the sizes of the layers grow to infinity. Moreover, in the setting where the underlying graph has the finite cycle property (i.e., every edge is contained in at most finitely many cycles) and distinct edge costs, we show that a unique MST T^* exists and the layered greedy algorithm produces iterates that converge to T^* by eventually "locking in" edges after finitely many iterations.

Marina A. Epelman
University of Michigan
Industrial and Operations Engineering
mepelman@umich.edu

Christopher Ryan
UBC Sauder School of Busines
chris.ryan@sauder.ubc.ca

Robert Smith
University of Michigan
rlsmith@umich.edu

CP20

Sensitivity Analysis for Integer Programs Using the K-Neighborhood

We investigate the behavior of an integer program when the objective function changes. In particular, we are interested in conditions for which a given integer solution stays optimal. The set of all such objective vectors forms the normal cone at the solution, i.e., for all objectives in this cone, the solution stays optimal. This cone is described by all active constraints of the integer hull. Since the integer hull is not known in general and expensive to calculate, we are looking for sufficient conditions for an objective to lie in this cone. As one example for such a condition, we investigate the k-neighborhood. Under the assumption that feasible neighboring binary solutions have at least k different components, we derive an inner approximation of the normal cone. Another approach for approximating the normal cone is by repeatedly adding feasible neighboring solutions to an existing sub-cone at the given point. The latter approach requires to solve an integer program in each iteration and thus is computationally expensive. We will present computational results for this approach in the context of analyzing the behavior of optimal energy networks.

Erik Jansen
TU Darmstadt
jansen@mathematik.tu-darmstadt.de

Marc Pfetsch
Research Group Optimization, Department of Mathematics
TU Darmstadt
pfetsch@mathematik.tu-darmstadt.de

CP20

Influence Maximization in Boolean Networks

The optimization problem aiming at the identification of minimal sets of nodes able to drive the dynamics of Boolean networks toward desired long-term behaviors is central for some applications, as for example the detection of key therapeutic targets to control pathways in models of biological signaling and regulatory networks. Unfortunately, the complexity of the optimization problem is exponential, making it exactly solvable on very small systems only. Some scalable approaches exist but they rely on linear approximations; other approaches estimate nonlinear effects but they are generally not scalable. In this talk, we introduce an alternative method inspired by those used in the solution of the well-studied problem of influence maximization for spreading processes in social networks. The computational time of the proposed method scales cubically with the network size. This is achieved thanks to some strong approximations, as for example neglecting dynamical correlations among Boolean variables. However, the method has the desirable feature of fully accounting for the nonlinear nature of Boolean dynamics. We validate the method on small gene regulatory networks whose dynamical landscapes are known by means of brute-force analysis. We then systematically apply it to a large collection of gene regulatory networks revealing that for about 65% of the analyzed networks, the minimal driver sets contain less than 20% of their nodes.

Filippo Radicchi
Center for Complex Networks and Systems Research
Indiana University
f.radicchi@gmail.com

Thomas Parmer
Indiana University
tjparmer@indiana.edu

Luis Rocha
State University of New York at Binghamton
rocha@binghamton.edu

CP20

A Modified Splitting Lagrangian Multiplier Method for Semi-Integer Problem

In various real-world applications, there are some cases that the decision variable should be confined in a disconnected set, that is, the feasible solution is either the lower bound or an integer in an inclusive range; and it is called a semi-integer variable. Such conditions practically exist in one of the Sustainable Development Goals (SDGs) in forest management. The important problem is arising from tree breeding where tree seeds is genetically selected to maximize the breeding quality so that it keeps under the diversity. The existence of the semi-integer variables, particularly in tree breeding, are usually more challenging to solve than continuous variables. Therefore, this research discusses several methods such as cone decomposition method (CDM), outer approximation (OA) method and the modified splitting augmented Lagrangian method (SALMM) to handle the difficulty of the constraints. The result shows that, as the data is increased, SALMM performs its efficiency than other method. However, for the certain data size, the SALMM failed to generate the optimal solution. In this case, the generated solution does not satisfy the quadratic (diversity) constraint. To remove such problem, we proposed another method which is a combination between CDM and SALM. Using the proposed method, we successfully and efficiently obtained the optimal solution of the semi-integer arising from tree breeding problem. The proposed method can be also used for similar structure.

Sena Safarina
Tokyo Institute of Technology
safarinasena@gmail.com

Makoto Yamashita
Tokyo Institute of Technology
Dept. of Mathematical and Computing Science
makoto.yamashita@c.titech.ac.jp

CP21

Extended Karush Kuhn Tucker type conditions in Interval Optimization Problems with the help of generalized Hukuhara subdifferentiability and analysis of Nonconvex Nonsmooth Composite Model in Interval Optimization

The optimization problem with interval-valued functions (IVFs) has become a significant research topic due to the inherent presence of imprecise and uncertain events in different real-world situations. These problems are known as interval optimization problems (IOPs). In this analysis, firstly, we study calculus for gH-subdifferential of convex interval-valued functions (IVFs). Next, with the help of developed results, we derive a Fritz-John-type and a Karush-Kuhn-Tucker-type efficiency condition for weak efficient solutions of IOPs. Thereafter, we apply our developed theories in a nonconvex nonsmooth composite model of an interval optimization problem (IOP). To support our results, we have performed a comparison of our theory with the existing Karush-Kuhn-Tucker and Fritz-John conditions for IOPs. After that, we report a characterization of the weak efficient solutions of the nonconvex nonsmooth composite model by applying the proposed concepts. Finally, we presented applications of developed results on the nonconvex nonsmooth model in 'Convex constrained nonconvex programming problems' for interval optimization problems (IOPs) and in 'Lasso optimality conditions' for interval optimization problems (IOPs).

Anshika Anshika
IIT-BHU
anshika.rs.mat19@itbhu.ac.in

Debdas Ghosh
IIT BHU, Varanasi, INDIA
Assistant Professor

debdas.mat@iitbhu.ac.in

CP21

Lagrangian Duality In Nonconvex Optimization

The Lagrangian function associated with an optimization problem has been of the utmost importance to develop algorithmic solution approaches for optimization problems. This is particularly the case when the problem of interest is convex, as in that case, optimizing the Lagrangian function provides a dual problem (Lagrangian dual) to the primal (original) one satisfying strong duality. Here, using techniques borrowed from both polynomial optimization and real analysis, we show that the classical Lagrangian function associated with an optimization problem can be used to obtain a Lagrangian dual of the problem that satisfies strong duality, even when the problem is nonconvex, as long as some qualification assumptions on the constraints and/or feasible set of the problem are satisfied. Furthermore, we present some consequences of our results when they are applied to polynomial optimization problems, in particular, to several quadratic optimization problems ubiquitous in the literature.

Rodolfo A. Quintero Ospina
Lehigh University
Deparment of Industrial and Systems Engineering
roq219@lehigh.edu

CP21

Global Well-Posedness of Set-Valued Optimization with Application to Uncertain Problems

Well-posedness for optimization problems is a well-known notion and has been studied extensively for scalar, vector and set-valued optimizationproblems. There is a broad classification in terms of pointwise and global wellposedness notions in vector and set-valued optimization problems. We focus on global well-posedness for set-valued optimization problems in this paper. We will discuss the equivalence between some existing notions of global well-posedness for set-valued optimization problems and also found scope of improving and extending the research in that field. On the other hand, robust approach towards uncertain optimization problems is another growing area of research. The well-posedness for the robust counterparts have been explored in very few papers, and that too only in the scalar and vector cases Therefore, we will focus on some global well-posedness properties of the robust formulation of uncertain set-valued optimization problems that generalize the concept of the well-posedness of robust formulation of uncertain vector optimization problems.

Vetrivel Vellaichamy
Indian Institute of Technology, Madras, Chennai India
vetri@iitm.ac.in

CP22

Vehicle Routing Problem with Simultaneous Pick-Up and Delivery in Omni-Channel Retailing

With mobile technology today, different retail formats have evolved increasingly. Traditional offline retailers have included online channels to serve online customers via their websites and social media. However, more and more retailers have ventured into omnichannel retailing to offer seamless services within different channels. Omni-channel retailing reduces friction in customers' transactions including purchases and customer returns - a reverse logistics issue gaining urgency in online retail. This thesis presents the Vehicle Routing Problem with Simultaneous Pickup and Delivery in Omni-channel Retailing (OCVRP-SPD). The classical SPD problem in VRP is a cost-minimizing optimization model with customer nodes having both pickup and delivery demands of multiple products. All deliveries to the customer nodes are fulfilled from the warehouse while all pickups along the route are delivered back to the warehouse. Here, the SPD problem is extended and explored under an Omni-channel format with those features: (1) offline stores are used as fulfillment centers to fulfill online customer orders, (2) the warehouse may also be used as fulfillment centers for some online customer orders, and (3) retail store and customers can return the product to the warehouse which should be picked up along the route.

Ting-Yang Chou, Ching-Hsuan Lin
National Taiwan University of Science and Technology
tychou100@gmail.com, m10901001@mail.ntust.edu.tw

CP22

Tighter LP-Relxations for the Railway Rolling Stock Rotation Problem with Vehicle Maintenance

The Rolling Stock Rotation Problem (RSRP) is the task to group railway rolling stock to vehicles compositions and assign them to trips of a given railway timetable in an operational feasible and most cost efficient way. In real world applications several different rules have to be considered to claim solutions to the RSRP as operational feasible. An example for these type of constraints are distance based vehicle maintenance requirements which define a maximum mileage for individual vehicles between two stops at maintenance facilities. In Giacco et al. 2014 and Borndoerfer et al. 2016 MILP formulations were presented to solve the RSRP while considering maintenance requirements. In both applications a continuous flow to track the vehicle mileage is linked to a binary vehicle (hyper) flow to solve the problem. The LP-relaxations of both models suffer from a systematical error resulting from the Big-M like linking of the variables. In this talk we present a reformulation of the latter model using path variables to overcome this error. Due to the worst case exponential number of possible paths we present a dynamic way to generate promising candidates to solve the LP-Relaxation. The generation process is a shortest path computation in a two layered graph to decrease the computation time of the approach. We evaluate our approach on real world long distance train instances and compare the results to the ones found with the algorithm of Borndoerfer et al. 2016.

Boris Grimm
Zuse Institute Berlin (ZIB)
grimm@zib.de

Ralf Borndoerfer, Julian Bushe
Zuse Institute Berlin
borndoerfer@zib.de, bushe@zib.de

CP22

Vehicle Assignment Variables in Branch-and-Price Algorithms for Vehicle Routing Problems

The heterogeneity in a Vehicle Routing Problem (VRP) introduces a new level of decision one has to deal with: Which customer gets assigned to which vehicle? These decisions can be mathematically represented by the vehicle assignment variables. Thus far, these variables did not receive much attention in branch-and-price algorithms, although their valuable information can be utilized to speed up the optimization process. In particular, this is achieved by excluding fractional solutions and simplifying the underlying instances of the pricing problems, known to be the bottleneck for VRPs. We adapt two state-of-the-art techniques commonly used for homogeneous VRPs: the fixing of and the branching on arc flow variables. While the adaption of the latter already enjoys some attention in the literature, vehicle assignment variable fixing seems yet to be exploited for a branch-and-price algorithm. We discuss the theoretical and practical challenges implied by the implementation of these techniques and present suitable solutions. Computational experiments performed on a variety of heterogeneous fleet VRPs showed that the site-dependent VRP and the VRP with heterogeneous time windows are most receptive, as the success correlates strongly with the symmetry of the variables. Overall, our results show that the vehicle assignment variable fixing and branching speed up the pricing algorithm and therefore can improve the general performance of state-of-the-art solvers.

Lukas Schürmann
University of Bonn
Institute for Computer Science
lschuermann@cs.uni-bonn.de

Petra Mutzel
University of Bonn
petra.mutzel@cs.uni-bonn.de

CP22

Numerical Optimization of Price and Return Strategies for Retail Consumers Products

The product return is widely utilized in retail markets as an option when customers have uncertain valuation of a product. To analyze the return strategies of a retailer who sells a product in a monopoly market, we develop an analytical framework capable of capturing the impacts of major related factors. These factors, in addition to other commonly studied factors in the literature, include return leniency, customers heterogeneity, and endowment effect. Using our analytical framework, we characterize the probability of a customer purchasing the product and the probabilities of keeping and returning it. These probabilities in turn characterize the retailers demand and return volume. The complexity of these models prevents us from finding closed form relations for the optimal price and refund level. We, therefore, use a set of numerical experiments over a wide range of parameters values that could reciprocate most practical circumstances. On this basis, we demonstrate optimal pricing and return strategies under various circumstances. Our analysis shows, among other results, that the endowment effect reduces the benefit of refund optimization and the retailers expected profit is more sensitive to the customers endowment effect when the product is not expensive for the retailer to acquire, and the return process is not very easy for the customers.

Ali Shirzadeh Chaleshtari
Operations and Information Management Department, Isenberg S
University of Massachusetts Amherst
ashirzadeh@umass.edu

Ehsan Elahi
Department of Management Science and Information Systems
University of Massachusetts Boston
ehsan.elahi@umb.edu

CP23

Stochastic Alternating Optimization Methods for Solving Large-Scale Machine Learning Problems

In this paper, we propose the Stochastic Alternating Linearization (SALIN) method for solving structured regularization problems, which have been widely used in machine learning and data mining. The algorithm is a stochastic extension of Alternating Linearization (ALIN) in solving convex optimization problems with complex non-smooth regularization terms. SALIN linearizes the loss function and penalty function alternatively at each iteration, based on the stochastic approximation of sub-gradients. By applying a special update test at each iteration, and a carefully designed sub-gradients update scheme, the algorithm achieves fast and stable convergence. The update test just relies on a fixed pre-defined set, and we show that the choice of the test set has little influence on the overall performance of the algorithm. Therefore SALIN is a robust method. We present the results of extensive numerical experiments for structured regularization problems such as Fused LASSO and graph-guided SVM, with both synthetic and real-world datasets. The numerical results demonstrate the efficacy and accuracy of our method.

Yu Du
University of Colorado Denver
University of Colorado Denver
yu.du@ucdenver.edu

Kaicheng Wu
Amazon
kaichengacademy@gmail.com

CP23

First-Order Algorithms to Optimize Functions with Biased Stochastic Oracles

This work considers the optimization of functions available through biased first-order stochastic oracles where the bias can be controlled/reduced but it requires higher computational cost. This problem setup is motivated by applications in infinite-time Markov decision processes and stochastic composition optimization. We propose first-order adaptive stochastic optimization algorithms and establish their finite-time overall computational and sample complexities under different convexity settings. Some numerical studies on the performance of the proposed algorithms and their comparison with other methods will be presented.

Yin Liu, Sam Davanloo Tajbakhsh
The Ohio State University
Department of Integrated Systems Engineering
liu.6630@osu.edu, davanloo.1@osu.edu

CP23

Stochastic Optimization for Spectral Risk Measures

Spectral risk objectives – also called L-risks – have gained attention in areas such as quantitative finance, distributionally robust machine learning, and operations research due to their ability to measure worst-case performance across an error distribution. We develop stochastic algorithms to optimize these quantities by characterizing their subdifferential and addressing challenges such as biasedness of subgradient estimates and non-smoothness of the objective. We explore these algorithms in machine learning problems such as regression, classification, and clustering. We present numerical results comparing the proposed approach to simple baselines such as the stochastic subgradient algorithm and regularized dual averaging.

Ronak Mehta
University of Washington
ronakdm@uw.edu

CP23

An Adaptive Subsampled Hessian-Free Optimization Method for Statistical Learning

We consider nonconvex statistical learning problems and propose a variable sample-path method, where the sample size is dynamically updated to ensure a decrease in the true objective function with high probability. We integrate this strategy in a subsampled Hessian-free trust-region method with truncated conjugate gradient, relying on outer product approximations. The approach is compared to various adaptive sample approximation algorithms and stochastic approximation methods proposed in stochastic optimization and machine learning. The efficiency of the approach is illustrated on various large size datasets and different regression models.

Jérémy Rieussec, Jean Laprés-Chartrand, Fabian Bastin
University of Montreal
jeremy.rieussec@gmail.com, jean.lapres-chartrand@umontreal.ca, bastin@iro.umontreal.ca

CP24

Fast Misalignment Correction in Cone-Beam Tomography via Variable Projection

In cone-beam computed tomography with circular source trajectory, only an approximation of the acquisition geometry is available, such parameters need to be corrected before any reconstruction strategy is used. First, we propose a fast and automatic method for determining the center of rotation position of the implicit two-dimensional fan-beam sinogram lying in the center of the cone-beam projections. This is done based on symmetry properties of fan-beam sinograms via an error-reduction strategy, by iteratively reducing the shift of symmetrical projections. This will be proved to have a faster convergence than any second-order algorithm and can already be used for any fan-beam scanning device. Then, both the in-plane detector angle (about the focal axis) and the detector horizontal shift (on the tomographic rotation plane) are estimated as a joint least-squares problem based on symmetrical properties of now a *tilted* fan-beam sinogram within the cone using a variable projection approach. The inner subproblem is the fan-beam center of rotation problem described above; while the remaining outer problem, on the in-plane detector angle, is proved to has a locally strongly convex and differentiable loss function under some realistic assumptions on the scanned object; therefore it is solved here via the BFGS algorithm. Finally, the expansion of the method to the remaining geometrical variables will be discussed. Our method will be validated with real tomographic data.

Patricio Guerrero, Wim Dewulf
Department of Mechanical Engineering, KU Leuven
patricio.guerrero@kuleuven.be, wim.dewulf@kuleuven.be

CP24

Data Assimilation for Quantum Nv Diamond Spectroscopy

Nitrogen-vacancy (NV) defect centers in diamond have generated much interest for their uses in quantum information and sensing. Negatively charged centers are used for high spatial-resolution sensing and for quantum information . Despite the rapid NV applications development, our grasp of basic NV properties is incomplete, which is important to understand to fully exploit potential uses. In this work we construct a statistical model for NV spectroscopy and use it in synthetic experiments to solve inverse problems. Our principal application is to develop a primary sensor based on the NV diamond quantum optical properties. This is a significant challenge because the NV diamond structure is sensitive to temperature and pressure as well as magnetic and electric fields. First, using the Hamiltonian for the effects of local strain and the environmental variables, we identify the observable components based on the invertibility of various observation systems. Next, we observe the influence of temperature and pressure on the NV center by solving the Schrdinger Equation and computing the theoretical spectroscopy curve. We assume that the observed photon counts are Poisson random variables with rates proportional to the theoretical spectroscopy. Then, using the Maximum Likelihood Estimation we find the parameter values that maximize the likelihood. Last but not the least we determine the robustness of the model using sensitivity analysis

Shraddha Verma
George Mason University
srajpal4@gmu.edu

CP25

Intuitionistic Fuzzy Twin Proximal Support Vector Machine with Fuzzy Hyperplane

Twin support vector machine (TWSVM) is a contemporary machine learning technique for classification and regression problems. However, TWSVM is sensitive to noises as it ignores the positioning of the input data samples and hence fails to distinguish between support vectors and noises. To overcome this issue, we propose a novel intuitionistic fuzzy twin proximal SVM with fuzzy hyperplane (IFTPSVM-FH). Instead of addressing two quadratic programming problems like in TWSVM, two non-parallel classifiers are obtained by solving two systems of linear equations. The two significant features of the proposed approach are that it gives an intuitionistic fuzzy number based on the relevance to each data vector, and the parameters for the hyperplane are fuzzified variables. With the use of fuzzy variables, the proposed method effectively captures the ambiguous character of real-world categorization tasks by representing vagueness in the training data. The proposed approach uses both membership and nonmembership weights to reduce the effect of noise. By incorporating nonlinear kernel functions into the feature space, the method can be used to detect complex patterns or nonlinearity in the datasets. We have applied our method to real-world classification tasks and conclude that it performs incredibly well in comparison to other approaches. In order to demonstrate the practical application of the proposed model, we have applied the technique for predicting the trends of the stock market.

Yash Arora, Shiv Gupta
Indian Institute of Technology Roorkee
yash_a@ma.iitr.ac.in, s.gupta@ma.iitr.ac.in

CP25

Fair Machine Learning on Cluster Samples

In many social surveys the data collection process is a cluster sampling, e.g. due to cost restrictions. In cluster samples the assumption of independence between the observation is not fulfilled. Hence, if the machine learning models do not account for the cluster correlations, the results may be biased. Especially high is the bias in cases where the cluster assignment is correlated to the variable of interest. When using machine learning for automated prediction, it is important to account for fairness in the prediction. Fairness in machine learning aims to ensure that biases in the data and model inaccuracies do not lead to discriminatory decisions. E.g., predictions from fair machine learning models should not discriminate against sensitive variables such as sexual orientation and ethnicity. We show the effect of cluster sampling on the quality of fair machine learning predictions in a reproducible simulation study. Further, we propose a correction for the clustering effect. The newly developed algorithm will be evaluated within a simulation and applied exemplary on the Adult survey.

João Vitor Pamplona
Trier University
pamplona@uni-trier.de

Jan Pablo Burgard
Economic and Social Statistics Department
University of Trier, Germany
burgardj@uni-trier.de

CP26

Choosing Penalty Parameters for Solving Constrained Binary Models with Qubos

Binary optimization models with constraints are common in the field of mathematical optimization. Applications abound in logistics, finance, engineering and industry. Forming a quadratic unconstrained binary optimization (QUBO) model from a constrained model can be advantageous for many reasons, including the expanding space of available solvers. Quantum technologies suggest avenues for solving more complex problems than previously possible. A promising approach in quantum computing operates on QUBOs, but there exists a challenge in translating a constrained problem to a QUBO in choosing appropriate penalty parameters. This choice will affect the ultimate solution by preventing the discovery of feasible solutions, hindering the search for optimal solutions or allowing the solver to reveal optimal bit vectors. Our work builds on the foundation laid by Glover, Kochenberger and Du. By forming QUBOs for a large number of constrained binary optimization problems using a range of penalty parameters, we can compare the minimum QUBO value found with a solver to the value obtained with a known optimal solution. This data helps to drive a general methodology for choosing penalty parameters before the first QUBO is created for a model and after the first solution is evaluated for fitness to tune it. By defining an optimality condition for the penalty parameter, we can utilize optimization methods to automate the search for penalty parameters.

Wesley Dyk, David Haycraft
Quantum Computing, Inc
wdyk@quantumcomputinginc.com,
dhaycraft@quantumcomputinginc.com

Raouf Dridi, Catherine Potts
Quantum Computing, Inc.
rdridi@quantumcomputinginc.com,
cpotts@quantumcomputinginc.com

CP26

Exploiting Structure in MILP: a Modelers Perspective

Many real-life problems can be tackled as Structured Mixed Integer Linear Programs(MILP). The workflow when dealing with MILPs usually involves two tools: a modelling tool and a solver. This presentation focuses on modelling tools. Most modelling tools either enable an encoding close to the mathematical one but lack support for structures or offer predefined components to assemble but adding components or modifying them is difficult. In this lecture, we explain the advantages of having a mixed approach by illustrating with the Graph-Based Optimization Modelling Language (GBOML). GBOML is a modelling tool that combines the strengths of both worlds. It supports structure and offers an encoding close to the mathematical one and library-like functionalities such as reuse and component assembling. We explain how exploiting structure can lead to a natural encoding of problems, faster time to build the intermediate representation and sometimes faster solving time. We explain how GBOML takes advantage of its structure to perform vectorization and parallelization. We propose a benchmark of GBOML, JuMP, Plasmo and Pyomo on a published energy problem tacked as a structured MILP. We show that GBOML and JuMP take a similar time to build the intermediate representation and outperform Plasmo and Pyomo. With parallelism, GBOML outperforms the three tools. In terms of memory footprint, GBOML uses the least memory. The presentation ends with some open directions of research.

Bardhyl Miftari, Mathias Berger
University of Liege
bmiftari@uliege.be, mathias.berger@alumni.duke.edu

Derval Guillaume
University of Liège
gderval@uliege.be

Damien Ernst
University of Liège
Telecom Paris
dernst@uliege.be

CP26

The Flow Control Game Model

We analyze strategies for controlling the flow on a network of nodes differentiated by the utility they enjoy from flows. For this, we propose a two-stage game played on a network, wherein the network defender decides on a flow control threshold in the first stage and then decides on interventions on the network, given the realization of flow changes. Our model is directly applicable to the control of substances among consumers and suppliers, e.g., opioid networks.

Abhishek Ray
George Mason University
George Mason University
aray8@gmu.edu

Monika Tomar, John Biechele-Speziale
Purdue University
tomarm@purdue.edu, jbiechel@purdue.edu

CP27

Globally Convergent Mpcc Based Algorithm to Solve Hybrid Dynamical Optimization Problems

Hybrid Dynamical Systems are an important way to model multiple physical processes like contact based mechanical systems and vapor liquid equilibrium systems. Optimal control of hybrid dynamical systems requires modeling and solving the non-smooth dynamics with kinks using complementarity constraints. Unfortunately, higher order numerical integration methods fail to achieve desired accuracy with uniform discretization grid. Adaptive discretization methods which can detect these non-smooth points accurately are required to solve these optimal control problems (OCP) accurately. Recently, Nurkanovic et al. (arXiv:2205.05337) developed a finite elements with switch detection (FESD) method for solving the non-smooth OCP. The method builds on previous idea (Baumrucker et al. (DOI:10.1016/j.jprocont.2009.02.006)) of using variable step sizes and additional constraints with cross-complementarities to enforce the non-smooth points to be at the boundary of the finite elements. Since the complementarity problem is solved using the penalty formulation, the approach is not suitable for highly nonlinear dynamical systems. Moreover, the algorithm can converge to spurious solutions which are not local optimum of the original OCP. We propose a hybrid strategy based on constraint relaxation and sequential linearization for solving the dynamic complementarity system which overcomes the drawbacks of the penalty approach. We present examples which show the efficacy of the proposed method.

Saif R. Kazi
Los Alamos National Laboratory
saifrahaman26@gmail.com

Mandar Thombre
Equinor Research Center, Norway
thombre.mandar@gmail.com

Lorenz Biegler
Carnegie Mellon University
Department of Chemical Engineering
lb01@andrew.cmu.edu

CP27

Optimizing with Attractor

This presentation studies local search algorithms on combinatorial problems from the perspective of dynamical systems. A heuristic local search algorithm on a combinatorial problem is a discrete dynamical system. The attractor theory in dynamical systems provides the necessary and sufficient theoretical foundation to study the search behavior of a local search system. In a local search system, the search trajectories converge into an attractor in the solution space, which contains the most promising solutions. The attractor can reduce the problem size exponentially, and make the exhaustive search feasible. Thus this new search paradigm is called optimizing with attractor. We illustrate this paradigm using TSP and 2-opt search. We use the edge configurations of local search trajectories to show how these trajectories converge to an attractor, and also show a local search system actually is a global and deterministic system. Convergence of a local search system means that edge configurations of search trajectories converge to a small set of edges. The search trajectories can prune TSP from a full tree into a tree with extremely sparse branches, and thus most unnecessary tours are eliminated. Under this paradigm, we can design a search system that combines local search and exhaustive search, in which local search process exponentially reduces the problem size in polynomial time and exhaustive search process finds the optimal solution in the attractor in polynomial time.

Weiqi Li
University of Michigan-Flint
weli@umflint.edu

Lavinia Stanica
Romanian-American University
stanica.lavinia.justina@profesor.rau.ro

CP27

Optimization of Control Free Time Sweeping Processes and Applications

This project addresses a free time optimal control problem of Mayer type for sweeping processes in which the perturbation is non-smooth. We develop a constructive finite-difference approximation procedure that allows us to establish necessary optimality conditions for the original continuous-time controlled sweeping process, and then show how these optimality conditions are applied to solve several applications in the real life.

Trang Nguyen
Wayne State University
ha2288@wayne.edu

Boris Mordukhovich
Wayne State University
Department of Mathematics
boris@math.wayne.edu

Giovanni Colombo
University Padova
Dipartimento di Matematica
colombo@math.unipd.it

Nguyen-Truc-Dao Nguyen
University of Michigan
nntdao@umich.edu

CP28

Monge-Kantorovich Framework in Smoothing Irregularities

There have been many studies working on statistical distances in probability theory, statistics and machine learning, however, for high dimensional data, we need to overcome the problems related to local irregularities in the measured distributions. One recent proposed way is smoothing these irregularities through convolution with a Wasserstein kernel. Monge-Kantorovich distance (Wasserstein distance) has attracted many researchers in statistics and

probability theory. We can find several applications of Monge-Kantorovich distances in kinetic theory to discuss the uniqueness of the kinetic equations and their stability. Methods of kernel illustrates one of the most effective nonparametric smoothing techniques. These methods are simple to understand and they possess very good statistical properties. There are efforts to construct kernels on discrete structures like strings, trees and graphs in which family of kernels have been used to generate the radial basis kernels, constructing the reproducing kernel Hilbert space to define transform theory associated with a positive definite kernel (gaussian processes), and in machine learning techniques which employ positive definite kernels to construct learning and approximating the reproducing kernel Hilbert space (RKHS) problems. We review the Monge-Kantorovich or Wasserstein distance theory and its application in optimization theory and we use it to derive several results mentioned in this area.

Tahmineh Azizi
Kansas State University
tazizi@ksu.edu

CP28

Tangent similarity gap in feedforward neural networks

We consider the problem of learning feedforward neural networks, $f : \mathbb{R}^{d_{in}} \times \Theta \to \mathbb{R}^{d_{out}}$ with an adversarial generator in its input space. We assume that this adversarial generator is smooth and that for any input x and some norm p the norm of the difference $\|x - \phi(x)\|_p$ is bounded. In this context we propose a novel similarity measure based on the expected change in the tangent mapping $(\mathbf{E}_{x-\phi(x)\sim\mathcal{N}(0,c\mathbf{I})}[\|\nabla_\theta f(x;\theta) - \nabla_\theta f(\phi(x);\theta)\|_2^2])$ defined over the manifold of the network instead of the loss used for learning. We show how the similarity gap, the difference between the expected similarity between pairs with the same label and pairs with different labels [Balcan and Blum, On a theory of learning with similarity functions, ICML 2006], based on the sensitivity of the tangent mapping under a proper metric changes during learning and indicate the relationship between the measure and the Neural Tangent Kernel [Jacot, Gabriel and Hongler, Neural tangent kernel: Convergence and generalization in neural networks, NeurIPS, 2018] and its connection to the generalization gap and to the complexity of the network. Our experiments suggest that the change in the proposed, label independent similarity measure correlates with the empirical generalization gap.

Balint Daroczy
Institute for Computer Science and Control (ELKH SZTAKI)
daroczyb@ilab.sztaki.hu

Daniel Racz
Institute for Computer Science and Control (ELKH SZTAKI)
Budapest, Hungary
racz.daniel@sztaki.hu

CP28

Finite-Time Analysis of Single-Timescale Actor-Critic

Decentralized Actor-Critic (AC) algorithms have been widely utilized for multi-agent reinforcement learning (MARL) and have achieved remarkable success. Apart from its empirical success, the theoretical convergence property of decentralized AC algorithms is largely unexplored. Most of the existing finite-time convergence results are derived based on either double-loop update or two-timescale step sizes rule. In practice, the *single-timescale* update is widely utilized, where actor and critic are updated in an alternating manner with step sizes being of the same order. In this work, we study a decentralized *single-timescale* AC algorithm.Theoretically, using linear approximation for value and reward estimation, we show that the algorithm has sample complexity of $\tilde{\mathcal{O}}(\varepsilon^{-2})$ under Markovian sampling, which matches the optimal complexity with double-loop implementation (here, $\tilde{\mathcal{O}}$ hides a logarithmic term). The central to establishing our complexity results is *the hidden smoothness of the optimal critic variable* we revealed. We also provide a local action privacy-preserving version of our algorithm and its analysis. Finally, we conduct experiments to show the superiority of our algorithm over the existing decentralized AC algorithms.

Qijun Luo
The Chinese University of Hong Kong Shenzhen
qijunluo@link.cuhk.edu.cn

Xiao Li
The Chinese University of Hong Kong, Shenzhen
lixiao@cuhk.edu.cn

CP28

Universal Approximation Property of Random Neural Networks

In this talk, we study single-hidden-layer feedforward neural networks with randomly initialized weights and biases, which are inspired by the works on extreme learning machines, random feature models, and reservoir computing. After the random initialization, only the linear readout needs to be trained, which can be performed, e.g., by a linear regression. Despite the popularity of this approach in empirical tasks, only little is known about the approximation properties. By considering these "random neural networks" as Banach space-valued random variables, we prove some universal approximation theorems within Bochner spaces. Hereby, the Banach space can be a more general function spaces such as L^p-spaces or Sobolev spaces, where the latter includes the approximation of the derivatives. In addition, we provide some approximation rates for neural networks, which shows in particular that random neural network overcome the curse of dimensionality on a dimension-scaled neighborhood of the training data. This talk is based on joint work with Ariel Neufeld.

Philipp Schmocker
Nanyang Technological University
philippt001@e.ntu.edu.sg

Ariel Neufeld
Nanyang Technological University
Singapore
ariel.neufeld@ntu.edu.sg

CP28

An introduction to the perceptron-augmented extended sigmoid function (SIGTRON) for imbalanced linear classification

The sigmoid function is a vital ingredient of machine learning and artificial intelligence. In the logistic regression on the linear classification problem, the role of the sigmoid function is that it provides probabilistic interpretability on the gradient of the logistic loss function and a natural probability interpretation during the classification testing phase. However, due to the symmetric and unbounded structure of the sigmoid function, the performance of the logistic regression is limited when training data is not balanced or includes outliers. To overcome the imbalance issue, e.g., in the one-vs-all strategy of the multiclass classification, in this talk, we introduce a perceptron-augmented extended sigmoid function (SIGTRON) as a probabilistic interpretable gradient for a loss function, defined through the integration of SIGTRON. An attractive property of the proposed method induced from the SIGTRON is that, due to the asymmetric and inherent robustness structure of the SIGTRON, it is well fit on the imbalanced training data set. Note that the loss function of the proposed model can be roughly estimated. Numerical experiments with more than one hundred data sets justify that the SIGTRON-based classification model outperforms all the classification models in LIBLINEAR. We use a simple trapezoidal integration for SIGTRON-induced loss function and L-BFGS algorithm with the Armijo-Wolfe line search.

Hyenkyun Woo
Logitron X
hyenkyun@gmail.com

Duk-Kyun Woo
Electronics and Telecommunications Research Institute (ETRI)
dkwu@etri.re.kr

MS1

Subspace Methods for Nonconvex Optimization

We discuss subspace methods - in both their random and deterministic variants, for nonconvex optimization problems. We are interested in functions with low effective dimensionality, that vary only along certain important directions or components. We show that the effective subspace of variation can be efficiently learned in advance of the optimization process; we contrast this with random techniques that focus directly on optimization rather than learning. Time permitting, various problem classes will be addressed such as local and global optimization.

Coralia Cartis
University of Oxford
Mathematical Institute
cartis@maths.ox.ac.uk

Estelle Massart
University of Louvain, Belgium
massart@uclouvain.be

Adilet Otemissov
The Alan Turing Institute for Data Science
London, UK
aotemissov@turing.ac.uk

Zhen Shao
University of Oxford
shaoz@maths.ox.ac.uk

Lindon Roberts, Lindon Roberts
University of Sydney
lindon.roberts@sydney.edu.au, lindon.roberts@sydney.edu.au

MS1

A Newton-MR Algorithm With Complexity Guarantees for Nonconvex Smooth Unconstrained Optimization

We consider variants of Newton-MR algorithm for solving unconstrained, smooth, but non-convex optimization problems. Unlike the overwhelming majority of Newton-type methods, which rely on conjugate gradient algorithm as the primary workhorse for their respective subproblems, Newton-MR employs minimum residual (MINRES) method. Recently, it has been established that MINRES has inherent ability to detect non-positive curvature directions as soon as they arise and certain useful monotonicity properties will be satisfied before such detection. We leverage these recent results and show that our algorithms come with desirable properties including competitive first and second-order worst-case complexities. Numerical examples demonstrate the performance of our proposed algorithms.

Yang Liu
University of Oxford
yang.liu@maths.ox.ac.uk

Fred Roosta
University of Queensland, Australia
fred.roosta@uq.edu.au

MS1

Inexact Sequential Quadratic Programming for Noisy Equality Constrained Optimization with Rank-Deficient Jacobians

We develop a class of practical stochastic algorithms for solving constrained optimization problems where only noisy evaluations of the objective and constraint functions are available. It is assumed that the noise level of those functions evaluations is known. The algorithm we propose computes the gradient approximations by derivative-free methods (e.g., finite differences) and computes a step by leveraging the sequential quadratic programming (SQP) framework. Our algorithm is inexact and dynamic and selects a step size either via an adaptive rule or a relaxed line search, and is endowed with strong theoretical guarantees even in the setting in which the constraint Jacobians are rank deficient. Finally, numerical experiments demonstrate the advantages of the method.

Jiahao Shi
University of Michigan
jiahaos@umich.edu

Albert S. Berahas
Department of Industrial and Operations Engineering
University of Michigan
albertberahas@gmail.com

Baoyu Zhou
University of Chicago
zbaoyu@uchicago.edu

MS2

A Single Cut Proximal Bundle Method for Stochas-

tic Convex Composite Optimization

This talk considers optimization problems where the objective is the sum of a function given by an expectation and a closed convex composite function. For such problems, a stochastic composite proximal bundle (SCPB) method with optimal complexity is proposed. The method does not require estimation of parameters involved in the assumptions on the objective functions. Moreover, to the best of our knowledge, this is the first proximal bundle method for stochastic programming able to deal with continuous distributions. Finally, computational results are show that SCPB substantially outperforms the robust stochastic approximation method on all instances considered.

Renato C. Monteiro
Georgia Institute of Technology
School of ISyE
monteiro@isye.gatech.edu

Jiaming Liang
Yale University
jiaming.liang@yale.edu

MS2

Progressive Decoupling as a Primal-Dual Method in Convex Optimization

The progressive decoupling algorithm minimizes a convex function over a given subspace through iterations that subtract a linear function, add a proximal term, and minimize over the entire space. The subspace stands for linear relationships among variables that in this process are decoupled so that separability and other features supportive of problem decomposition can be enabled. Among the many specializations are various basic schemes of splitting. What is not apparent on the surface, is that progressive decoupling is actually a primal-dual method based on representing the space as the product of the given subspace and its orthogonal complement, and then taking a partial conjugate to get an associated Lagrangian function. The proximal point algorithm is applied to that Lagrangian in saddle point mode and in a new variable-metric implementation. That' is behind the scenes and hidden from view, but the outcome is a method that is sure to converge globally to an optimal primal-dual pair and generically exihibit asymptotic linear convergence.

R. Tyrrell Rockafellar
University of Washington, U.S.
rtr@uw.edu

MS2

Adaptive Methods for Nonconvex Min-Max Optimization

Adaptive gradient methods such as AdaGrad and Adam have shown their ability to adjust the stepsizes on the fly in a parameter-agnostic manner and are successful in nonconvex minimization. When it comes to nonconvex minimax optimization, direct extensions of such adaptive optimizers without proper time-scale separation may fail to work in practice. In fact, even for a quadratic example, the naive combination of Gradient Descent Ascent with any existing adaptive stepsizes is proven to diverge if the initial primal-dual stepsize ratio is not carefully chosen. We introduce two simple fixes for these adaptive methods, allowing automatic adaptation to the time-scale separation necessary for fast convergence. The resulting algorithms are fully parameter-agnostic and achieve near-optimal complexities in deterministic and stochastic settings of nonconvex-strongly-concave minimax problems, without a priori knowledge about problem-specific parameters.

Junchi Yang, Xiang Li, Niao He
ETH Zurich
junchi.yang@inf.ethz.ch, xiang.li@inf.ethz.ch, niao.he@inf.ethz.ch

MS3

New Developments of First Order Methods for Linear Programming

Linear programming (LP) is a fundamental tool in operations research with wide applications in practice. The state-of-the-art LP solvers are essentially based on either simplex method or barrier method, which are quite mature and reliable at delivering highly accurate solutions. However, it is highly challenging to further scale up these two methods. The computational bottleneck of both methods is the matrix factorization when solving linear equations, which usually requires significantly more memory usage and cannot be directly applied on the modern computing resources, i.e., distributed computing and/or GPUs. In contrast, first-order methods (FOMs) only require matrix-vector multiplications, which work very well on these modern computing infrastructures and have massively accelerated the machine learning training process during the last 15 years. In this talk, Ill present new FOMs for LP. On the computational side, we build up a new LP solver based on the proposed FOMs and Ill present a comprehensive numerical study on the proposed FOMs. The solver has been open-sourced through Google OR-Tools.

Haihao Lu
University of Chicago
haihao.lu@chicagobooth.edu

David Applegate
Google
dapplegate@google.com

Miles Lubin
Hudson River Trading
miles.lubin@gmail.com

Oliver H. Hinder
University of Pittsburgh
ohinder@pitt.edu

Mateo Diaz
California Institute of Technology,U.S.
mateodd@caltech.edu

Jinwen Yang
University of Chicago
Department of Statistics
jinweny@statistics.uchicago.edu

MS3

Algorithms for Large Saddle-Point Problems with Bilinear Coupling

We consider a convex-concave primal-dual optimization framework in which the coupling between primal and dual variables is bilinear. This framework admits linearly constrained optimization together with a variety of interesting problems in machine learning, including (linear) empirical risk minimization with various regularization terms. It also includes a formulation that we term "generalized linear programming' (GLP) in which regularization terms and constraints are added to the traditional linear programming formulation, provided they admit efficient prox operations. Problems from differentially robust optimization (DRO), using either f-divergence metrics or Wasserstein metrics, can be formulated as GLPs. We describe algorithms for our framework that take prox-gradient steps alternately in the primal and dual variables, but incorporate such additional features as coordinate descent, variance reduction, dual averaging, importance sampling, and iterate averaging. Our methods can also exploit sparsity in the matrix that couples primal and dual variables. Our methods match or improve on the best known worst-case complexity bounds in various settings. Computational experiments indicate that our methods also have good practical performance. The talk represents joint work with Ahmet Alacaoglu, Jelena Diakonikolas, Chaobing Song, Eric Lin, and Volkan Cevher

Stephen Wright
University of Wisconsin-Madison
Dept. of Computer Sciences
swright@cs.wisc.edu

MS3

Geometric Condition Numbers and the Primal-Dual Hybrid Gradient Method for Linear Programming

There has recently been significant progress in solving huge-scale linear optimization problems (LP) using the primal-dual hybrid gradient method (PDHG) with restarts in conjunction with heuristic enhancements. However, there is still a lack of theoretical understanding regarding the behavior of PDHG for LP. With the aim of understanding what "natural characteristics of LP are relevant for PDHG, we study further properties and complexity implications for several geometric condition numbers for LP related to sharpness and error ratios. We show that these measures enter into the computational guarantees of PDHG, and that these measures are also interpretable as stability measures under data perturbation. Our geometric condition numbers offer tighter (and more computable) theoretical guarantees, compared with the existing theory for PDHG using the Hoffman constant of optimality conditions. Additionally, our condition numbers, we provide theoretical evidence for the practical success of several significant heuristics employed in PDHG implementations for LP. Our experiments on LP relaxations of the MIPLIB 2017 dataset lead to provable improvements in these heuristics, and demonstrate the impact of these condition numbers on the actual performance of PDHG.

Zikai Xiong
MIT
Operations Research Center
zikai@mit.edu

Robert M. Freund
MIT
Sloan School of Management
rfreund@mit.edu

MS4

Warehouse Problem with Bounds, Fixed Costs and Complementarity Constraints

We discuss an open question in the warehouse problem where a merchant trading a commodity tries to find an optimal inventory-trading policy that decides on quantities of purchases and sales during a fixed time horizon in order to maximize their total pay-off, making use of fluctuations in sale and cost prices. We provide the first known polynomial-time algorithms for the case when there are fixed costs for purchases and sales, optional complementarity constraints that prohibit purchasing and selling during the same time period, and bounds on purchase and sales quantities. We do so by providing an exact characterization of the extreme points of the feasible region and using this to construct a suitable network where a min-cost flow computation provides an optimal solution. We are also able to provide polynomial extended linear formulations for the original feasible regions. Our methods build on the work by Wolsey and Yaman. We also consider the problem without fixed costs and provide a fully polynomial time approximation scheme in a setting with time-dependent bounds.

Ishan Bansal
Cornell University
ib332@cornell.edu

MS4

On Aggregations of Quadratic Inequalities and Hidden Hyperplane Convexity

We study properties of the convex hull of a set S described by quadratic inequalities. A simple way of generating inequalities valid on S is to to take a nonnegative linear combinations of the defining inequalities of S. We call such inequalities aggregations. Special aggregations naturally contain the convex hull of S, and we give sufficient conditions for such aggregations to define the convex hull. We introduce the notion of hidden hyperplane convexity (HHC), which is related to the classical notion of hidden convexity of quadratic maps. We show that if the quadratic map associated with S satisfies HHC, then the convex hull of S is defined by special aggregations. To the best of our knowledge, this result generalizes all known results regarding aggregations defining convex hulls. Using this sufficient condition, we are able to recognize previously unknown classes of sets where aggregations lead to convex hull. We show that the condition known as positive definite linear combination together with hidden hyerplane convexity is a sufficient condition for finitely many aggregations to define the convex hull. All the above results are for sets defined using open quadratic inequalities. For closed quadratic inequalities, we prove a new result regarding aggregations giving the convex hull, without topological assumptions on S. This is joint work with Grigoriy Blekherman and Shengding Sun.

Santanu Dey
Georgia Institute of Technology
santanu.dey@isye.gatech.edu

MS4

Disjoint Bilinear Programming: Convex Hulls, Facial Disjunctions, and Perspective Relaxations

We study convex hull characterisations of a generalised bilinear set given by $(x, y, Z) : Z = xy, x \in P, y \in Y$ where Z is a matrix variable whose each entry is the bilinear term $x_i * y_j$, and P is a polytope and Y is a closed convex set.

Akshay Gupte

School of Mathematics
University of Edinburgh
akshay.gupte@ed.ac.uk

Laura Galli
University of Pisa
laura.galli@unipi.it

Adam N. Letchford
Lancaster University Management School
a.n.letchford@lancaster.ac.uk

MS5

Value of Stochastic Solution with Right-Hand Side Uncertainty

We revisit the value of stochastic solution (VSS) in the context of distributional ambiguity. When the uncertainty arises from the right-hand side of a two-stage stochastic program, we consider upper and lower bounds of VSS using distributionally robust and optimistic optimization. We discuss the computation of these bounds and demonstrate them through numerical examples.

Haoming Shen
Industrial and Operations Engineering
University of Michigan
hmshen@umich.edu

Ruiwei Jiang
Industrial and Operations Engineering
University of Michigan
ruiwei@umich.edu

MS5

New Algorithms for Robust Optimization with Decision-Dependent Uncertainty

Robust optimization problems where the realizations of one or more unknown parameters depend on the choice of the optimal decisions can be modeled using so-called decision-dependent or endogeneous uncertainty sets. Existing solution methods for these models rely on classical duality-based reformulations, which are plagued with several limitations including lack of scalability and require assumptions on the structure of the decision-dependency. To overcome these limitations, we propose a primal branch-and-bound algorithm that generalizes the classical cutting set method to the case of decision-dependent uncertainty sets. We present convergence results as well as findings from computational experiments that showcase the advantages of the proposed algorithm over that of existing methods.

Anirudh Subramanyam
Pennsylvania State University
subramanyam@psu.edu

MS5

Adjustability in Robust Linear Optimization

We investigate the concept of adjustability – the difference in objective values between two types of robust optimization formulations: one where (static) decisions are made before uncertainty realization, and one where uncertainty is resolved before (adjustable) decisions. This difference reflects the value of information and decision timing in optimization under uncertainty, and is related to several other concepts such as interchangeability in games and optimality of decision rules in robust optimization. We develop a theoretical framework to quantify adjustability of problems with linear objective, linear constraints, and fixed recourse. We make very few additional assumptions. In particular, we do not assume constraint-wise separability or parameter nonnegativity that are commonly imposed in the literature. This allows us to study important but previously under-investigated problems, such as formulations with equality or both upper and lower bound constraints. Based on the discovery of a connection between the reformulations of the two problems, we provide a necessary and sufficient condition – in the form of a theorem-of-the-alternatives – for adjustability to be zero when the uncertainty set is polyhedral. Then, we develop a constructive approach to quantify adjustability when the uncertainty set is general, which results in an efficient and tight algorithm to bound adjustability. We demonstrate the efficiency and tightness via theoretical and numerical analyses.

Ningji Wei
Texas Tech University
ningji.wei@ttu.edu

Peter Zhang
Carnegie Mellon University
pyzhang@cmu.edu

MS6

Fantastic Lifts and Where to Find Them: What to Look for in a Parameterization

Given one optimization problem, consider pairing it with another by smoothly parametrizing the domain. This is done either for practical purposes (e.g., to use smooth optimization algorithms with good guarantees) or for theoretical purposes (e.g., to reveal that the landscape satisfies a strict saddle property). The central question is: how do the landscapes of the two problems relate? The relation is often determined by the parametrization itself; it is almost entirely independent of the cost function. In this talk, I will present a geometric framework for studying parametrizations according to their effect on landscapes. Applications include: optimization over low-rank matrices and tensors by optimizing over a factorization; the Burer-Monteiro approach to semidefinite programs; training neural networks by optimizing over their weights and biases; and quotienting out symmetries. Joint work with Eitan Levin (Caltech) and Nicolas Boumal (EPFL).

Joe Kileel
University of Texas at Austin, U.S.
jkileel@math.utexas.edu

MS6

Relaxation, Reparameterization, and Global Optimality in Geometric Machine Perception

Many fundamental geometric estimation tasks in robotics and computer vision naturally take the form of high-dimensional nonconvex optimization problems over Riemannian manifolds; this class includes (for example) the fundamental problems of simultaneous localization and mapping (in robotics), rotation averaging (in computer vision), and sensor network localization (in distributed sensing). These problems are known to be computationally hard in general, with many suboptimal local minima that can entrap the smooth local optimization methods typically applied to solve them. The result is that standard machine perception algorithms can be surprisingly brittle, often returning egregiously wrong estimates in practice. In this talk, we describe a novel class of *certifiably correct* estimation algorithms that can efficiently recover globally optimal solutions of these challenging geometric estimation problems in many practical settings. Our approach is based upon a sequence of relaxations and reparameterizations of the original estimation problem that we prove enables the recovery of an *exact*, *globally optimal* solution under moderate measurement noise, and can be easily implemented using off-the-shelf optimization libraries.

David M. Rosen
Northeastern University
d.rosen@northeastern.edu

MS6

Improved Global Guarantees for the Nonconvex BurerMonteiro Factorization Via Rank Overparameterization

We consider minimizing a twice-differentiable, L-smooth, and μ-strongly convex objective ϕ over an $n \times n$ positive semidefinite matrix $M \succeq 0$, under the assumption that the minimizer $M^\star$ has low rank $r^\star \ll n$. Following the Burer–Monteiro approach, we instead minimize the nonconvex objective $f(X) = \phi(XX^T)$ over a factor matrix X of size $n \times r$. This substantially reduces the number of variables from $O(n^2)$ to as few as $O(n)$ and also enforces positive semidefiniteness for free, but at the cost of giving up the convexity of the original problem. In this paper, we prove that if the search rank $r \geq r^\star$ is overparameterized by a *constant factor* with respect to the true rank $r^\star$, namely as in $r > \frac{1}{4}(L/\mu - 1)^2 r^\star$, then despite nonconvexity, local optimization is guaranteed to globally converge from any initial point to the global optimum. This significantly improves upon a previous rank overparameterization threshold of $r \geq n$, which is known to be sharp if ϕ is allowed to be nonsmooth and/or non-strongly convex, but would increase the number of variables back up to $O(n^2)$. Conversely, without overparameterization, we prove that such a guarantee is possible if and only if ϕ is almost perfectly conditioned, with $L/\mu < 3$. Therefore, we conclude that a small amount of overparameterization can lead to large improvements in theoretical guarantees.

Richard Zhang
University of Illinois at Urbana-Champaign
ryz@illinois.edu

MS7

Frank-Wolfe Type Methods for Nonconvex Inequality-Constrained Problems

The Frank-Wolfe method (also known as the conditional gradient method) implements efficient linear optimization oracles for minimizing smooth functions over compact convex sets. This method and its variants form a prominent class of projection-free first-order methods for a large variety of application problems such as matrix completion. In this talk, we extend this method to minimize smooth functions over a possibly nonconvex compact set, which is defined as the level set of a difference-of-convex function that satisfies mild regularity conditions. The key to our extension is the introduction of a new linear optimization oracle for the nonconvex compact constraint set. We discuss convergence and present numerical experiments to illustrate the empirical performance of the proposed algorithm.

Liaoyuan Zeng
College of Science
Zhejiang University of Technology
zengly@zjut.edu.cn

Yongle Zhang
School of Mathematical Sciences
Sichuan Normal University, China
yongle-zhang@163.com

Guoyin Li
The University of New South Wales
Sydney
g.li@unsw.edu.au

Ting Kei Pong
Department of Applied Mathematics
The Hong Kong Polytechnic University, Hong Kong
tk.pong@polyu.edu.hk

MS7

A Correlative Sparse Lagrange Multiplier Expression Relaxation for Polynomial Optimization

We consider polynomial optimization with correlative sparsity. We construct correlative sparse Lagrange multiplier expressions (CS-LMEs) and propose the CS-LME reformulation for polynomial optimization using the Karush-Kuhn-Tucker optimality conditions. Correlative sparse sum-of-squares (CS-SOS) relaxations are applied to solve the CS-LME reformulation. We show that the CS-LME reformulation inherits the original correlative sparsity pattern, and the CS-SOS relaxation provides sharper lower bounds when applied to the CS-LME reformulation, compared with when it is applied to the original problem. Moreover, the convergence of our approach is guaranteed under mild conditions. In numerical experiments, our new approach usually finds the global optimal value (up to a negligible error) with a low relaxation order, for cases where directly solving the problem fails to get an accurate approximation. Also, by proper exploitation of the correlative sparsity, our CS-LME approach requires less computational time than the original LME approach for reaching the same accuracy level.

Zheng Qu
The University of Hong Kong
zhengqu@hku.hk

Xindong Tang
The Hong Kong Polytechnic University
xindong.tang@polyu.edu.hk

MS7

The Effect of Implicit Discretizations of Gradient Flow on Badly Conditioned Nonconvex Problems

The gradient flow (GF) is an ODE for which its explicit Euler's discretization is the gradient descent method. In this work, we investigate a family of methods derived from approximate implicit discretizations of (GF), drawing the connection between larger stability regions and less sensitive hyperparameter tuning. We focus on the implicit τ-step backwards differentiation formulas (BDFs), approximated in an inner loop with a few iterations of vanilla gradient descent, and give their convergence rate when the objective function is convex, strongly convex, or nonconvex. Numerical experiments show the wide range of effects of these different methods on extremely poorly conditioned problems, especially those brought about in training deep neural networks.

Yifan Sun
Stony Brook University
yifan.sun@stonybrook.edu

MS8

Risk-Averse Perspectives on Min-Max Problems

Risk averse optimization plays a major role in the design of safe and robust models for decision-making. In this talk, we revisit the bias-variance trade-off of min-max stochastic first-order methods from a robust perspective. Precisely, we study the convergence properties of accelerated methods on saddle-point problems for diverse convex risk measures. We deduce new ways to set the associated hyperparameters in order to stabilise the algorithm at the equilibrium. We illustrate the interest of such approach on a series of numerical experiments inspired from game theory and distributionally robust optimization.

Yassine Laguel
Rutgers Business School
yassine.laguel@rutgers.edu

Mert Gurbuzbalaban
Rutgers University
mert.gurbuzbalaban@gmail.com

Necdet S. Aybat
Penn State University
nsa10@psu.edu

MS8

First-Order Algorithms for Min-Max Optimization in Geodesic Metric Spaces

From optimal transport to robust dimensionality reduction, a plethora of machine learning applications can be cast into the min-max optimization problems over Riemannian manifolds. Though many min-max algorithms have been analyzed in the Euclidean setting, it has proved elusive to translate these results to the Riemannian case. Zhang et al. [2022] have recently shown that geodesic convex concave Riemannian problems always admit saddle-point solutions. Inspired by this result, we study whether a performance gap between Riemannian and optimal Euclidean space convex-concave algorithms is necessary. We answer this question in the negative-we prove that the Riemannian corrected extragradient (RCEG) method achieves last-iterate convergence at a linear rate in the geodesically strongly-convex-concave case, matching the Euclidean result. Our results also extend to the stochastic or non-smooth case where RCEG and Riemanian gradient ascent descent (RGDA) achieve near-optimal convergence rates up to factors depending on curvature of the manifold.

Tianyi Lin
UC Berkeley
darren_lin@berkeley.edu

MS8

Robust Accelerated Primal-Dual Methods for Computing Saddle Points

We consider strongly convex/strongly concave saddle point problems assuming we have access to unbiased stochastic estimates of the gradients. We propose a stochastic accelerated primal-dual (SAPD) algorithm and show that SAPD sequence, generated using constant primal-dual step sizes, linearly converges to a neighborhood of the unique saddle point, where the size of the neighborhood is determined by the asymptotic variance of the iterates. Interpreting the asymptotic variance as a measure of robustness to gradient noise, we obtain explicit characterizations of robustness in terms of SAPD parameters and problem constants. Based on these characterizations, we develop computationally tractable techniques for optimizing the SAPD parameters, i.e., the primal and dual step sizes, and the momentum parameter, to achieve a desired trade-off between the convergence rate and robustness on the Pareto curve. This allows SAPD to enjoy fast convergence properties while being robust to noise as an accelerated method. We also show that SAPD admits convergence guarantees for the distance metric with a variance term optimal up to a logarithmic factor which can be removed by employing a restarting strategy. Finally, we illustrate the efficiency of our approach on distributionally robust logistic regression.

Xuan Zhang, Necdet S. Aybat
Penn State University
xxz358@psu.edu, nsa10@psu.edu

Mert Gurbuzbalaban
Rutgers University
mert.gurbuzbalaban@gmail.com

MS9

Generalized Optimistic Methods for Convex-Concave Saddle Point Problems

The optimistic gradient method has seen increasing popularity as an efficient first-order method for solving convex-concave saddle point problems. To analyze its iteration complexity, a recent work [Mokhtari et al., SIOPT 2020] proposed an interesting perspective that interprets the optimistic gradient method as an approximation to the proximal point method. In this talk, we follow this approach and distill the underlying idea of optimism to propose a generalized optimistic method, which encompasses the optimistic gradient method as a special case. Our general framework can handle constrained saddle point problems with composite objective functions and can work with arbitrary norms with compatible Bregman distances. Moreover, we also develop an adaptive line search scheme to select the stepsizes without knowledge of the smoothness coefficients. We instantiate our method with first-order, second-order and higher-order oracles and give sharp global iteration complexity bounds. Moreover, our line search scheme provably only requires an almost constant number of calls to a subproblem solver per iteration on average, making our first-order and second-order methods particularly amenable to implementation.

Ruichen Jiang
University of Texas at Austin

rjiang@utexas.edu

MS9

Differential Privacy Meets Adaptive Optimization

Adaptive optimization methods have become the default solvers for many machine learning tasks. Unfortunately, the benefits of adaptivity may degrade when training with differential privacy, as the noise added to ensure privacy reduces the effectiveness of the adaptive preconditioner. In this talk, I will focus on two directions we explore to improve privacy/utility/computation tradeoffs in the context of adaptive optimization. We begin by introducing a general framework that uses non-sensitive side information to precondition the gradients, allowing the effective use of adaptive methods in private settings. Then, we will discuss techniques to estimate and efficiently adapt to gradient geometry without auxiliary data or side information. In particular, motivated by the observation that adaptive methods can tolerate stale preconditioners, we construct delayed but less noisy preconditioners from historical iterates. In both works, we formally analyze the convergence properties, showing that the proposed methods can reduce the amount of noise needed to achieve the same privacy guarantees. We demonstrate significantly improved empirical performance across image, text, and recommendation benchmarks. I will conclude the talk with applications of the proposed approaches to federated learning (i.e., training statistical models across heterogeneous networks while keeping user data local), as well as future directions.

Tian Li
Carnegie Mellon University
tianli@cmu.edu

MS9

Will Bilevel Optimizers Benefit from Loops

Bilevel optimization has arisen as a powerful tool for solving a variety of machine learning problems. Two current popular bilevel optimizers AID-BiO and ITD-BiO naturally involve solving one or two sub-problems, and consequently, whether we solve these problems with loops (that take many iterations) or without loops (that take only a few iterations) can significantly affect the overall computational efficiency. Existing studies in the literature cover only some of those implementation choices, and the complexity bounds available are not refined enough to enable rigorous comparison among different implementations. In this work, we first establish unified convergence analysis for both AID-BiO and ITD-BiO that are applicable to all implementation choices of loops. We then specialize our results to characterize the computational complexity for all implementations, which enable an explicit comparison among them. Our result indicates that for AID-BiO, the loop for estimating the optimal point of the inner function is beneficial for overall efficiency, although it causes higher complexity for each update step, and the loop for approximating the outer-level Hessian-inverse-vector product reduces the gradient complexity. For ITD-BiO, the two loops always coexist, and our convergence upper and lower bounds show that such loops are necessary to guarantee a vanishing convergence error, whereas the no-loop scheme suffers from an unavoidable non-vanishing convergence error.

Kaiyi Ji
University at Buffalo
kaiyiji@buffalo.edu

Mingrui Liu
George Mason University
mingruil@gmu.edu

Yingbin Liang
Ohio State University
yingbinliang@gmail.com

Lei Ying
University of Michigan
leiying@umich.edu

MS10

Pajarito: An Extensible Mixed-Integer Conic Solver In Julia

Pajarito.jl is an outer approximation solver for mixed integer conic problems. In early 2022, we redesigned Pajarito to support MathOptInterface. Pajarito now has a generic cone interface that allows adding support for new convex cones through a small list of oracles. In the supplemental package PajaritoExtras.jl,we have extended Pajarito by adding support for several cones recognized by Hypatia.jl (a continuous conic solver with its own generic cone interface). In the examples folder of PajaritoExtras, we formulate a wide variety of applied mixed-integer conic problems. Our experiments compare Pajarito'sperformance under different but equivalent formulations. This talk gives an overview of Pajarito's new architecture and presents some of these computational results.

Chris Coey
MIT Operations Research Center
coey.chris@gmail.com

Juan Pablo Vielma
Google
juan.pablo.vielma@gmail.com

MS10

Uno: A Unified Nonlinear Optimization Framework

We present a new nonlinear optimization framework, called UNO, that can be used to implement a broad range of optimization methods. Our framework allows us to implement different subproblem solves (e.g. LP or QP steps), provides a flexible choice of globalization strategy (including penalty function or filter methods), a choice of globalization mechanism (e.g. trust-region or line-search), and different constraint relaxation strategies (l1, augmented Lagrangian, or restoration). We illustrate the flexibility of our framework by implementing a range of existing optimization solvers, and show how we can easily adapt the framework to experiment with new optimization approaches for solving constrained nonlinear optimization problems.

Sven Leyffer
Argonne National Laboratory
leyffer@anl.gov

Charlie Vanaret
Technische Universität Berlin
vanaret@zib.de

MS10

Recent Advances in the OSQP Solver: Differen-

tiable Optimization, Accelerated Linear Algebra, and More

OSQP is a fast, light, and widely popular open-source solver for quadratic optimization based on the alternating direction method of multipliers. We present the latest changes, improvements, and additions in OSQP 1.0. We introduce a new modular linear algebra framework to seamlessly run OSQP on three different platforms: default single-threaded for embedded applications, the Intel Math Kernel Library for multicore CPUs, and CUDA for GPUs. We discuss new OSQP code generation routines to create a light embeddable version of the solver. In contrast to prior OSQP code generation, our new implementation runs directly in C and can be called from any high level OSQP interface, including Python, Matlab, and Julia. In addition, OSQP now supports differentiable optimization and it can efficiently compute, directly in C, the derivatives of its solution with respect to the problem parameters. With convenient wrappers in JAX and Pytorch, OSQP is, to our knowledge, the first solver with internal automatic differentiation capabilities for the optimization problem itself. We illustrate the benefits of the new functionalities in OSQP on various numerical examples.

Ian McInerney
The University of Manchester
i.mcinerney17@imperial.ac.uk

Vineet Bansal
Princeton University
vineetb@princeton.edu

Paul Goulart
University of Oxford
paul.goulart@eng.ox.ac.uk

Bartolomeo Stellato
Princeton University
bstellato@princeton.edu

MS11

Robust Explainable Prescriptive Analytics

We propose a new robust explainable prescriptive analytics framework that minimizes a risk-based objective function under distributional ambiguity by leveraging the data collected on the past realizations of the uncertain parameters affecting the decision model and the side information that have some predictive power on those uncertainties. The framework solves for an explainable response policy that transforms the side information directly to implementable here-and-now decisions. Such a policy should endow with the properties of facilitating explanation of the decisions, ensuring that the solutions are implementable, and maintaining the computational tractability of the optimization problem. We show that affine and tree-based policies could achieve these salient properties. We adopt the data-driven robust satisficing framework to address the issue of overfitting when the empirical distribution is used for evaluating the risk-based objective function. To address linear optimization models with recourse, we provide a new safe tractable approximation that ensures the feasibility of the robust satisficing model, which can deal with constraints that are biaffine in the outcome variables and the side information. We also introduce a new biaffine recourse adaptation to improve the quality of the approximation. We provide a simulation case study on how the framework can be applied to in a risk minimizing portfolio optimization problem using past returns as side information.

Li Chen
National University of Singapore
chen_l@u.nus.edu

Melvyn Sim
NUS Business School
melvynsim@nus.edu.sg

Xun Zhang
Shanghai Jiao Tong University (SJTU) - Antai College of Econ
xunzhang@sjtu.edu.cn

Minglong Zhou
Fudan University - School of Management
zhouminglong111@gamil.com

MS11

A Practical Scenario-Based Approach for Uncertain Convex Optimization Problems

Optimization under uncertainty has undergone rapid development in the past decades. However, despite the growing recognition of the importance of uncertainty, we observe that the existing methods are still underutilized in practice. We propose a simple, yet effective, data-driven approach that offers a number of practical advantages over existing methods. Our approach is: (i) accessible, (ii) able to deal with a wide variety of problems and (iii) provides dimension-free statistical feasibility guarantees. We propose an iterative generation and evaluation algorithm in which the robustness of solutions is analyzed a posteriori. We demonstrate the effectiveness of our approach over existing methods using numerical experiments.

Justin Starreveld, Dick den Hertog
University of Amsterdam
j.s.starreveld@uva.nl, d.denhertog@uva.nl

Roger Laeven
Department of Quantitative Economics
University of Amsterdam
r.j.a.laeven@uva.nl

Guanyu Jin
University of Amsterdam
g.jin@uva.nl

MS11

Stochastic Optimization with Optimal Importance Sampling

In solving stochastic optimization problems that involve rare events running crude Monte Carlo can be prohibitively inefficient. Importance sampling can be considered instead to lower the sampling error to more desirable levels. However, selecting a best in class importance sampler typically requires knowledge of a minimizer of the stochastic optimization problem. We show that by simultaneously solving an auxiliary stochastic optimization problem we can untie this circular challenge.

Bart Van Parys
Massachusetts Institute of Technology
vanparys@mit.edu

Liviu Aolaritei
ETH Zurich
aliviu@ethz.ch

Henry Lam
Columbia University
henry.lam@columbia.edu

MS12

Randomized Linear Algebra for Interior Point Methods

Linear programming is a central problem in computer science and applied mathematics with numerous applications across a wide range of domains, including machine learning and data science. Interior point methods (IPMs) are a common approach to solving linear programs with strong theoretical guarantees and solid empirical performance. The time complexity of these methods is dominated by the cost of solving a linear system of equations at each iteration. In common applications of linear programming, particularly in data science and scientific computing, the size of this linear system can become prohibitively large, requiring the use of iterative solvers which provide an approximate solution to the linear system. Approximately solving the linear system at each iteration of an IPM invalidates common analyses of IPMs and the theoretical guarantees they provide. In this talk we will discuss how randomized linear algebra can be used to design and analyze theoretically and practically efficient IPMs when using approximate linear solvers.

Petros Drineas
Purdue University
Computer Science
pdrineas@purdue.edu

Haim Avron
Tel Aviv University
haimav@tauex.tau.ac.il

Agniva Chowdhuri
Postdoctoral Research Associate
Oak Ridge National Laboratory
chowdhurya@ornl.gov

Gregory Dexter
Purdue University
gdexter@purdue.edu

Palma London
Cornell University
londonpalma@gmail.com

MS12

Combining Proximal Point and Interior Point Methods: the Linear Algebra Perspective

We will present recent results concerning the interpretation of Interior Point Methods using the tools of Proximal Point Algorithms. In particular, we will show that it is possible to design a class of second-order methods characterized by a reduced computational footprint of the related linear algebra routines, comparable to that of the first-order methods.

Stefano Cipolla
University of Edinburgh
scipolla@ed.ac.uk

Jacek Gondzio
The University of Edinburgh, United Kingdom
J.Gondzio@ed.ac.uk

MS12

New Developments of ADMM-based Interior Point Methods for Linear Programming and Conic Programming

: The ADMM-based interior point method (ABIP, Lin et al. 2021) is a hybrid algorithm which effectively combines the interior point method and the first-order method to achieve performance boost in large-scale linear programming. Different from the standard interior point method which relies on a costly Newton step, ABIP applies the alternating direction method of multipliers (ADMM) to approximately solve the barrier penalized problem. In this paper, we provide a new version of ABIP with multiple improvements. First, we develop several new implementation strategies to accelerate ABIPs performance for linear programming. Next, we extend ABIP to solving the more general linear conic programming and establish the associated iteration complexity of the algorithm. Finally, we conduct extensive numerical experiments in both synthetic and real-world datasets to demonstrate the empirical advantage of our developments. In particular, the enhanced ABIP achieves a 5.8x reduction in the geometric mean of run time on 105 LP instances from Netlib and it compares favorably against state-of-the-art open-source solvers in a wide range of large-scale problems. Moreover, it is even comparable to the commercial solvers in some particular datasets.

Bo Jiang, Qi Deng
Shanghai University of Finance and Economics
jiang.bo@mail.shufe.edu.cn, qideng@sufe.edu.cn

Qing Feng
Cornell University, U.S.
qf48@cornell.edu

Dongdong Ge, Wenzhi Gao
Shanghai University of Finance and Economics, China
dongdong@mail.shufe.edu.cn, gwz@163.shufe.edu.cn

Yinyu Ye
Stanford University
yinyu-ye@stanford.edu

Chunwen Zhang
Shanghai University of Finance and Economics, China
chuwzhang@gmail.com

MS13

Convergent Discretization of Optimal Boundary Control Problems for Discontinuous Solutions of Hyperbolic Conservation Laws

We study the convergence of discretization schemes for the adjoint equation arising in the adjoint-based derivative computation for optimal boundary control of conservation laws with source term. As boundary control we consider piecewise continuously differentiable controls with possible discontinuities at switching times, where the smooth parts as well as the switching times serve as controls. Since entropy solutions of hyperbolic conservation laws can develop discontinuities, the derivation of differentiability results for objective functions and of optimality conditions is involved. Moreover, a numerical approximation that yields consistent derivatives and optimality conditions is delicate. It can be shown that the derivative of tracking-type objective functionals with respect to the smooth parts and the switching times of the boundary controls can be represented by an adjoint-based formula. The adjoint state is the so-called reversible solution of a transport equation with discontinuous coefficient and boundary conditions. We study discrete adjoint schemes of monotone difference schemes in conservation form such as Engquist-Osher or Lax-Friedrichs scheme. We also allow that the state is computed by another numerical scheme satisfying certain convergence properties. We proof convergence results of the discrete adjoint to the reversible solution. This implies also the convergence of the discrete sensitivities. The findings are illustrated by numerical examples.

Stefan Ulbrich
Technische Universitaet Darmstadt
Fachbereich Mathematik
ulbrich@mathematik.tu-darmstadt.de

Paloma Schäfer Aguilar
Technical University Darmstadt (Germany)
aguilar@mathematik.tu-darmstadt.de

MS13

A Directional Newton Derivative Scheme for Shape Optimization Problems Constrained by Variational Inequalities

Shape optimization problems constrained by variational inequalities (VI) are non-smooth and non-convex optimization problems. The non-smoothness arises due to the VI constraint, which makes it challenging to derive optimality conditions. Besides the non-smoothness there are complementary aspects due to the VIs as well as distributed, nonlinear, non-convex and infinite-dimensional aspects due to the shapes which complicate to set up an optimality system and, thus, to develop fast and higher order solution algorithms. In this talk, we consider directional Newton-derivatives in order to formulate optimality conditions. In this context, we set up a directional Newton-shape derivative scheme and formulate an algorithm based on the novel derivatives. Examples show the application of the proposed scheme.

Kathrin Welker
TU Freiberg
Germany
kathrin.welker@math.tu-freiberg.de

Volker H. Schulz
University of Trier
Department of Mathematics
Volker.Schulz@uni-trier.de

Nico Goldammer
University of the Federal Armed Forces Hamburg
goldammer@hsu-hh.de

MS14

Recent Advances in Sdp Performance Estimation for Iterative Methods

Semidefinite programming (SDP) performance analysis was introduced in 2014 in a seminal paper by Drori and Teboulle. Since then, it has become a standard tool in establishing the (often exact) rates of convergence of many iterative methods, by formulating the worst-case convergence rate as the optimal value of an SDP problem. This has yielded new insight on many methods, including gradient descent, Newtons method for self-concordant functions, ADMM, DCA, etc. In this talk we will review some recent results on SDP performance analysis for iterative methods. The talk will be based on joint work with Hadi Abbaszadehpeivasti and Moslem Zamani.

Etienne De Klerk
Tilburg University
E.deKlerk@tilburguniversity.edu

Hadi Abbaszadehpeivasti
Tilburg University, The Netherlands
h.abbaszadehpeivasti@tilburguniversity.edu

Moslem Zamani
Tilburg University
m.zamani_1@tilburguniversity.edu

MS14

Exact Optimal Accelerated Complexity for Fixed-Point Iterations

Despite the broad use of fixed-point iterations throughout applied mathematics, the optimal convergence rate of general fixed-point problems with nonexpansive nonlinear operators has not been established. This work presents an acceleration mechanism for fixed-point iterations with nonexpansive operators, contractive operators, and nonexpansive operators satisfying a Hölder-type growth condition. We then provide matching complexity lower bounds to establish the exact optimality of the acceleration mechanisms in the nonexpansive and contractive setups. This result will also be extended to the nonexpansive setup without fixed points. Finally, we provide experiments with CT imaging, optimal transport, and decentralized optimization to demonstrate the practical effectiveness of the acceleration mechanism.

Jisun Park, Ernest K. Ryu
Seoul National University, South Korea
colleenp0515@snu.ac.kr, ernestryu@snu.ac.kr

MS14

The speed-robustness trade-off for first-order methods with additive gradient noise

We study the trade-off between convergence rate and sensitivity to stochastic additive gradient noise for first-order optimization methods. Ordinary Gradient Descent (GD) can be made fast-and-sensitive or slow-and-robust by increasing or decreasing the stepsize, respectively. However, it is not clear how such a trade-off can be navigated when working with accelerated methods such as Polyak's Heavy Ball (HB) or Nesterov's Fast Gradient (FG) methods, or whether any of these methods can achieve an optimal trade-off. We consider three classes of functions: (1) strongly convex quadratics, (2) smooth strongly convex functions, and (3) functions whose gradients are bounded in a sector. For each function class, we present a tractable way to compute convergence rate and sensitivity to additive gradient noise for a broad family of first-order methods, and we present algorithm designs that are near-Pareto-optimal with respect to our analysis. Each design consists of a simple analytic update rule with two states of memory, similar to HB and FG. Moreover, each design has a scalar tuning parameter that explicitly trades off convergence rate and sensitivity to additive gradient noise. When tuned as aggressively as possible, our proposed algorithms recover the algorithms with fastest-known convergence rates for each function class. When tuned to be more robust, our algorithms are novel and provide a practical way to control noise sensitivity while maintaining fast convergence.

Bryan Van Scoy
Miami University
bvanscoy@miamioh.edu

Laurent Lessard
Northeastern University
l.lessard@northeastern.edu

MS15

Wasserstein Generative Adversarial Uncertainty Quantification and Transfer Learning in Physics-Informed Neural Networks

In this paper, we study a physics-informed algorithm for Wasserstein Generative Adversarial Networks (WGANs) for uncertainty quantification in solutions of partial differential equations. By using groupsort activation functions in adversarial network discriminators, network generators are utilized to learn the uncertainty in solutions of partial differential equations observed from the initial/boundary data. Under mild assumptions, we show that the generalization error of the computed generator converges to the approximation error of the network with high probability when the number of samples is sufficiently taken. According to our established error bound, we also find that our physics-informed WGANs have a higher requirement for the capacity of discriminators than that generators. Numerical results on synthetic examples of partial differential equations are reported to validate our theoretical results and demonstrate how uncertainty quantification can be obtained for solutions of partial differential equations and the distributions of initial/boundary data. However, the quality or the accuracy of the uncertainty quantification theory in all the points in the interior is still a theoretical vacancy and required for further research.

Yihang Gao
Department of Mathematics
University of Hong Kong
gaoyh@connect.hku.hk

Michael Ng
Department of Mathematics
The University of Hong Kong
mng@maths.hku.hk

MS15

Stochastic Optimization with Constraints: Algorithm, Convergence, and Statistical Inference

Constrained stochastic optimization problems appear widely in numerous applications in statistics, machine learning, and engineering. These include constrained maximum likelihood estimation, constrained deep neural networks, physics-informed machine learning, and optimal control. In this talk, I will describe how to design adaptive algorithms for solving constrained stochastic nonlinear optimization problems based on sequential quadratic programming (SQP). This talk is divided into two parts. In the first part, I will describe the development of a stochastic line search procedure to adaptively select the stepsize in a setting where batches of samples can be accessed. The algorithm adaptively selects the batch size in each iteration to estimate the objective model while satisfying suitable probabilistic conditions on the model estimation accuracy. The SQP equipped with this line search procedure provably converges to a KKT point almost surely. In the second part, I will switch to a fully online setup, where the algorithm only has access to a single sample to estimate the objective model in each iteration. I provide the global and local almost sure convergence results and perform statistical inference by showing the asymptotic normality with a Berry-Esseen bound for the SQP iterates.

Sen Na
ICSI and Department of Statistics, UC Berkeley
senna@berkeley.edu

MS15

Differentiable Optimization-Based Layers

Constraints, such as physical constraints or resource-based constraints, are ubiquitous in real-world models, and have been a long time subject of study for the optimization field. Recently, using the implicit function theorem and its software implementations, we can make constrained optimization solvers differentiable with regards to some parameters. This allows us to incorporate constraints-based modeling into deep-learning models. In this talk, we provide a small introduction to this subfield, and showcase some recent applications of this way of thinking: designing differentiable PDE-constrained functions, and designing models on very large graphs by selecting subgraphs in a differentiable way.

Geoffrey Negiar
UC Berkeley
BAIR
geoffrey_negiar@berkeley.edu

MS16

Variational Inequality and Saddle Point Problems with Functional Constraints

In the first part of this talk, we discuss the monotone Variational Inequality (mVI) problem which involves function constraints. In the existing literature, algorithms for the mVI problem assume accurate projections onto the feasible set. This is violated when the mVI problem involves complicated function constraints. Here, we show a principled approach to handling such function constraints. We show unified complexity results for the problem in smooth/nonsmooth and stochastic operator and/or stochastic function constraint settings. For the deterministic and smooth setting, we show an adaptive version of the algorithm. In the second part, we show that all of our algorithms can be easily extended for convex-concave saddle point (SP) problems where primal and dual variables are coupled by a constraint. Note that the usual SP problems studied in the literature assume that the feasible domain is a product form of the primal domain and the dual domain. This assumption is violated when the variables are coupled. Hence, our algorithms solve a new problem class of even deterministic SP problem setting (which is widely explored in recent years). Borrowing developments from the mVI problem, we further develop (I) adaptive versions for smooth deterministic SP problems, and (II) unified complexity for smooth/nonsmooth stochastic and/or deterministic SP problems.

Digvijay Boob
Southern Methodist University
dboob@smu.edu

MS16

Functional Constrained Optimization for Risk Aversion and Sparsity Control

In this paper, we focus on projection-free methods that generate a sparse trajectory for solving challenging functional constrained optimization problems. Specifically, for the convex setting, we propose a Level Conditional Gradient (LCG) method, which leverages a level-set framework to update the estimation of the optimal value and an inner conditional gradient oracle for solving mini-max subproblems. We show that the method achieves $\mathcal{O}\big(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\big)$ iteration complexity for solving both smooth and nonsmooth cases without dependency on a possibly large magnitude of optimal dual Lagrange multiplier. For the nonconvex setting, we introduce the Level Inexact Proximal Point (IPP-LCG) method and the Direct Nonconvex Conditional Gradient (DNCG) method. The first approach taps into the advantage of LCG by transforming the problem into a series of convex subproblems and exhibits an $\mathcal{O}\big(\frac{1}{\epsilon^3}\log\frac{1}{\epsilon}\big)$ iteration complexity for finding an approximate KKT point. DNCG is the first single-loop projection-free method, with iteration complexity bounded by $\mathcal{O}\big(\frac{1}{\epsilon^4}\big)$ for computing a so-called ϵ-Wolfe point. We demonstrate the effectiveness of LCG, IPP-LCG and DNCG by devising formulations and conducting numerical experiments on two risk averse sparse optimization applications: a portfolio selection problem and a radiation therapy planning problem.

Yi Cheng
ISyE, Georgia Tech
Atlanta, GA, USA
chengyi95@outlook.com

Guanghui Lan
Georgia Institute of Technology
george.lan@isye.gatech.edu

Edwin Romeijn
ISyE, Georgia Tech
edwin.romeijn@isye.gatech.edu

MS16

On Well-Structured Convex-Concave Saddle Point Problems and Variational Inequalities with Monotone Operators

For those acquainted with CVX (aka disciplined convex programming) of M. Grant and S. Boyd, the motivation of this work is the desire to extend the scope of CVX beyond convex minimization—to convex-concave saddle point problems and variational inequalities with monotone operators. To attain this goal, given a family $\mathcal{K}$ of cones (e.g., Lorentz, semidefinite, geometric, etc), we introduce the notions of $\mathcal{K}$-conic representation of a convex-concave saddle point problem and of variational inequality with monotone operator. We demonstrate that given such a representation of the problem of interest, the latter can be reduced straightforwardly to a conic problem on a cone from $\mathcal{K}$ and thus can be solved by (any) solver capable to handle conic problems on cones from $\mathcal{K}$ (e.g., Mosek or SDPT3 in the case of semidefinite cones). We also show that $\mathcal{K}$-representations of convex-concave functions and monotone vector fields admit a fully algorithmic calculus which helps to recognize the cases when a saddle point problem or variational inequality can be converted into a conic problem on a cone from $\mathcal{K}$ and to carry out such conversion.

Anatoli Juditsky
LJK, Université Grenoble Alpes
France
anatoli.juditsky@univ-grenoble-alpes.fr

Arkadi Nemirovski
School of Industrial and Systems Engineering
Georgia Tech
nemirovs@isye.gatech.edu

MS17

Dynamic Pricing for Electric Vehicles Charging Station Management

Due to numerous political and environmental incentives, the number of electric cars has increased dramatically in recent years. This rapid growth raises a number of challenges in terms of charging management. More precisely, it is important to manage the location of charging stations, the size of the stations to avoid queues and the impact of charging on the electrical grid. In this talk, we propose a dynamic pricing model that aims at distributing users in time and space on charging stations in order to reduce the impact on the electrical grid and thus avoid black outs. In order to represent the hierarchical decision process, this pricing problem is modeled by a bi-level optimization problem where the institution managing the charging stations (the leader) determines dynamic tariffs by explicitly taking into account the preferences of the users (the followers) with respect to their charging choices (space and time preferences). Relying on the particular structure of the problem, the bilevel problem is reformulated as a MIP one. Experimental results put into highlight the efficiency of the solution approach. Sensitivity results are provided and discussed

Clemence Alasseur
EDF
clemence.alasseur@edf.fr

Miguel F. Anjos
University of Edinburgh
anjos@stanfordalumni.org

Luce Brotcorne
INRIA Lille Nord Europe
luce.brotcorne@inria.fr

Gaël Guillot
INRIA , Lille Nord Europe
gael.guillot@inria.fr

Riadh Zorgati
EDF R&D
Department OSIRIS
riadh.zorgati@edf.fr

MS17

Deep Learning with Nontrivial Constraints

Despite the sweeping success and impact of deep learning in numerous domains, imposing explicit constraints is relatively new but increasingly pressing in deep learning (DL), stimulated by, e.g., trustworthy AI that performs robust optimization over complicated perturbation sets and scientific and engineering applications that need to respect physical laws and constraints. In this talk, we will (1) survey DL problems with nontrivial constraints across science, engineering, and medicine, (2) highlight the NCVX computing framework we have recently built, which provides deterministic solvers to solve constrained DL problems, and (3) invite the optimization community to solve the stochastic constrained DL problems.

Ju Sun
University of Minnesota
jusun@unm.edu

MS17

Aggregations of Quadratic Inequalities and Hidden Hyperplane Convexity

Abstract: we study sets defined by quadratic inequalities, and whether the convex hull can be described using nonnegative weighted linear combinations of the constraints, which we call aggregations. We introduce the notion of hidden hyperplane convexity (HHC), which gives a sufficient condition for convex hull to be given by aggregations. HHC encompasses known results for two or three quadratics as special cases, and is closely related to the classical notion of hidden convexity in convex geometry. We apply our results to sets defined by linear and sphere constraints. Joint work with Greg Blekherman and Santanu S. Dey.

Shengding Sun
Georgia Institute of Technology
Department of Mathematics
ssun313@gatech.edu

MS18

Optimistic Posterior Sampling Methods for Sample-efficient Reinforcement Learning

Posterior sampling for Reinforcement Learning (RL) refers to a class of methods motivated by a Bayesian framework, which is quite attractive due to its flexibility in incorporating a variety of models. In this talk, we will discuss a recent line of work which has endowed these methods with worst-case sample-efficiency guarantees through an optimistic enhancement of the basic posterior sampling method. We will present a novel structural measure called a decoupling coefficient, which captures the difficulty of the underlying RL problem, and bound the sub-optimality of the learning agent in terms of this measure. Along the way, we will also see how standard no-regret learning ideas for posterior sampling can be leveraged and adapted to obtain our results.

Alekh Agarwal
Google
alekhagarwal@google.com

Tong Zhang
The Hong Kong University of Science and Technology
tongzhang@tongzhang-ml.org

MS18

Efficiently Computing Nash Equilibria in Adversarial Team Markov Games

Computing Nash equilibrium policies is a central problem in multi-agent reinforcement learning that has received extensive attention both in theory and in practice. However, provable guarantees have been thus far either limited to fully competitive or cooperative scenarios or impose strong assumptions that are difficult to meet in most practical applications. In this work, we depart from those prior results by investigating infinite-horizon *adversarial team Markov games*, a natural and well-motivated class of games in which a team of identically-interested players – in the absence of any explicit coordination or communication – is competing against an adversarial player. This setting allows for a unifying treatment of zero-sum Markov games and Markov potential games, and serves as a step to model more realistic strategic interactions that feature both competing and cooperative interests. Our main contribution is the first algorithm for computing stationaryϵ-approximate Nash equilibria in adversarial team Markov games with computational complexity that is polynomial in all the natural parameters of the game, as well as$1/\epsilon$. The proposed algorithm is particularly natural and practical, and it is based on performing independent policy gradient steps for each player in the team, in tandem with best responses from the side of the adversary; in turn, the policy for the adversary is then obtained by solving a carefully constructed linear program.

Ioannis Panageas
University of California, Irvine
ipanagea@ics.uci.edu

MS19

Optimizing Large Scale Temporal Multi-Commodity Capacitated Flow Problem

We study the temporal fixed-charge multi-commodity flow problem. This problem is used in modeling large scale fulfillment networks where timing impacts both consumer experience as well as operations in the network. To this end, we wish to maximize the volume of the flow that arrive within the promised times while obeying capacity and cost constraints. Due to the sheer size of this problem, it cannot be solved directly even on moderate size transportation networks. Lara et al. proposed a combination heuristics that approximately solve the MILP formulation of this problem. In this work, we show a simple greedy heuristic approach which involves iteratively finding the best possible path in the time expanded network to flow packages between the origin of a demand and its destination. This subroutine resembles the dynamic programming approach to the shortest path problem. A post-processed module is then used to find the timed-path that delivers most on-time flow per unit of cost incurred. We show through experiments that on average this heuristic can close a 96% gap with the approach in Lara et al. while being 7X faster.

Arash Haddadan
Amazon
ahaddada@amazon.com

MS19

Faster Solutions of Very Large Timing Optimization Problems Using Dynamic Discretization Discovery.

In this talk, we present novel algorithms for large instances of temporal, fixed-charge multicommodity flow problems. At the heart of our algorithms lie strong, but extremely large mixed integer programming formulations that in turn are based on time-expanded networks arising from a given base graph. We discuss novel solution algorithms that extend the so called dynamic discretization discovery methodology (DDD) due to Boland et al. (Operations Research, 2017). In particular, we show that DDD can be applied when we face hard node- and arc capacity constraints, as well as when the underlying formulation is cyclic in nature.

Jochen Koenemann
University of Waterloo
kjochen@amazon.com

MS19

Dynamic discretization discovery for region-based networks

Practical network design applications are often complicated by temporal considerations such as transit times and time-varying network parameters. While these problems can be modeled with time-indexed formulations, this often results in huge time-expanded graphs. The recently-developed dynamic discretization discovery (DDD) method allows many time-dependent problems to become more tractable by iteratively solving instances of the problem on smaller networks where each node has its own discrete set of departure times. However, in the current implementation of DDD, all arcs departing a common node share the same set of departure times. This causes DDD to be ine?ective for solving problems where all near-optimal solutions require many distinct departure times at the majority of the nodes in the network. To address this shortcoming, we develop a DDD framework where the set of departure times is determined on the arc level rather than the node level. This approach is particularly effective for region-based networks.

Madison Van Dyk
University of Waterloo
madison.vandyk@uwaterloo.ca

MS20

On the Steplength Selection in Gradient-Based Methods for Special Constrained Optimization Problems

Gradient-based methods are widely used for solving nonlinear optimization problems; their popularity is essentially related to the computational simplicity combined with low cost per iteration and low memory requirements for solving such problems with medium accuracy. Steplength selection is a fundamental task for improving the practical efficiency of gradient-based algorithms; it requires to balance the need of accelerating the convergence rate without increasing the overall computational cost. In the last years, very efficient gradient-based approaches have been designed, exploiting special strategies for the steplength selection. This talk aims to show how the presence of the feasible region in quadratic and non-quadratic minimization problems influences the spectral behaviour of well-known steplength strategies, like the Barzilai-Borwein rules and the limited memory steplength selections developed by R. Fletcher in 2012, employed in standard gradient projection-like methods. Starting from a spectral analysis aimed at clarifying the relation between subsequent gradients along the iterations, we propose different ideas to generalize steplength selection rules to gradient projection algorithms taking into account the constraints. Some hints on how these concepts can be extended to proximal gradient methods for a special non-smooth minimization problem will be provided, together with numerical evaluations of the proposed approaches.

Serena Crisci
University of Campania Luigi Vanvitelli
serena.crisci@unicampania.it

Federica Porta
Università di Modena e Reggio Emilia
Dipartimento di Scienze Fisiche, Informatiche e Matematiche
federica.porta@unimore.it

Valeria Ruggiero
Dipartimento di Matematica e Informatica
valeria.ruggiero@unife.it

Luca Zanni
Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università di Modena e Reggio Emilia
luca.zanni@unimore.it

MS20

The Harmonic Rayleigh Quotient with Target for Stepsize Selection in Gradient Methods

We study the use of the harmonic Rayleigh quotient with target for the stepsize selection in gradient methods for nonlinear optimization problems. This stepsize provides a framework to reinterpret existing stepsizes and introduce new ones. In particular, we analyze the adaptive Barzilai-Borwein stepsize (Zhou, Gao and Dai, Gradient methods with adaptive step-sizes, 2006) and propose a target-based version of it. While existing stepsizes usually correspond to negative values for the target, positive targets are also considered. We extend the classical convergence analysis for quadratic problems (Dai and Liao, R-linear convergence of the Barzilai and Borwein gradient method, 2002) to these new stepsizes, and show they also satisfy a certain secant condition. Numerical experiments on generic unconstrained problems show that the introduction of a target may decrease the computational cost of the spectral gradient method combined with a nonmonotone line search.

Giulia Ferrandi, Michiel E. Hochstenbach
Eindhoven University of Technology
Department of Mathematics and Computer Science
g.ferrandi@tue.nl, m.e.hochstenbach@tue.nl

Nataša Krejic
University of Novi Sad
Faculty of Sciences
natasak@uns.ac.rs

MS20

Two Spectral Gradient Methods with Application to Eigenvalue Problems

In this work we introduce two Barzilai and Borwein-like steps-sizes for the classical gradient method. Both stepsizes are derived from two unconstrained optimization models that involve approximate information of the Hessian of the objective function. A convergence analysis of the proposed algorithm is provided. Some numerical experiments are performed in order to compare the efficiency and effectiveness of the proposed procedures with similar methods in the literature. Finally, we apply the new gradient methods in the solution of trace maximization problems over the Stiefel manifold, in order to compute eigenvalues of symmetric matrices.

Harry Oviedo
Universidad Adolfo Ibáñez, Facultad de Inegiería y Ciencias
harry.oviedo@uai.cl

MS21

Randomized Branch-and-Cut Algorithm for Stochastic Mixed-Integer Programming

Stochastic mixed-integer programs are among the most challenging optimization problems that find many practical applications. This talk presents a novel randomization-based branch and cut algorithm for two-stage stochastic mixed-integer programs. A stage-decomposition algorithm is embedded within a branch-and-cut procedure to handle the computational difficulty of solving the stochastic program. Unlike the deterministic branch-and-cut, where node relaxations are deterministic optimization problems, the node relaxations in our algorithm are stochastic. We solve these stochastic node relaxation problems using the sequential sampling-based stochastic decomposition algorithm. Instead of using a fixed sample set, the stochastic decomposition algorithm generates observations on the fly while concurrently performing the optimization step. This approach results in a partitioning process guided by stochastic lower and upper bounds. We discuss the convergence analysis and illustrate the computational performance of the algorithm on well-known instances in the literature.

Harsha Gangammanavar
Southern Methodist University, Department of Engineering
Management, Information, and Systems
harsha@smu.edu

MS21

Ptas for Stochastic Knapsack with Chance Constraints and Random Profits

We consider a stochastic knapsack problem where each item has a random profit and a random size of known distributions independent of other items. The goal is to find a set of "probably approximately optimal" items that satisfies the following two conditions. (1) The probability of the total size exceeding the knapsack bound is at most a given threshold. (2) The probability of the total profit underneath the deterministic average optimal value within an δ-tolerance is at most another given threshold. We present a greedy algorithm for the problem, which ensures a $1-O(\epsilon)$ approximation ratio under a mild additional assumption, and a Polynomial Time Approximation Scheme (PTAS) for the problem, which achieves a near-optimal profit.

Guanyi Wang
National University of Singapore
guanyi.w@nus.edu.sg

MS23

On Adaptive stochastic Trust-Region approaches for finite sum minimization

In this talk we discuss recent advances in stochastic first and second-order trust-region methods for solving finite-sum minimization problems. At each iteration, the proposed methods approximate the function and the derivatives by sampling, employing adaptive sample sizes that may vary from one iteration to the other. The choice of the function sample size is ruled by the inexact restoration approach, whereas the derivatives approximations can be computed averaging in smaller sample sets. The trust-region step is accepted or rejected according to whether a sufficient decrease condition on a suitable merit function holds or not. The proposed methods can also handle the case of bound constraints. We provide convergence results in probability and worst-case complexity results in expectation. We also report numerical results showing the advantages of adaptive approaches in reducing the tuning efforts in training neural network.

Stefania Bellavia
Universita di Firenze
stefania.bellavia@unifi.it

Nataša Krejic
University of Novi Sad
Faculty of Sciences
natasak@uns.ac.rs

Benedetta Morini
Dipartimento di Ingegneria Industriale
Universita' di Firenze
benedetta.morini@unifi.it

Simone Rebegoldi
Università degli studi di Firenze
simone.rebegoldi@unifi.it

MS23

A Stochastic-Gradient-based Interior-Point Algorithm for Bound-Constrained Stochastic Optimization

An interior-point algorithm for minimizing a stochastic objective function subject to bound constraints is presented, analyzed, and tested. The algorithm is unique from typical interior-point methodologies due to its definition of an inner neighborhood of the feasible region—dependent on a positive neighborhood parameter—in which the iterates are forced to remain. This avoids the need to employ a tool such as a fraction-to-the-boundary rule (which is common in the deterministic regime) since the use of such a tool makes it difficult to ensure convergence guarantees in the stochastic regime. It is shown that with a careful balance between the barrier, step size, and neighborhood parameter sequences, the algorithm enjoys convergence guarantees in the deterministic setting as well as in the context of stochastic and probabilistic analyses when stochastic gradients are employed. The results of numerical experiments show that in stochastic settings the algorithm can outperform state-of-the-art projected stochastic gradient techniques.

Frank E. Curtis
Industrial and Systems Engineering
Lehigh University
frank.e.curtis@gmail.com

Vyacheslav Kungurtsev
Czech Technical University in Prague
Department of Computer Science
vyacheslav.kungurtsev@fel.cvut.cz

Daniel Robinson, Qi Wang
Lehigh University
daniel.p.robinson@gmail.edu, qiw420@lehigh.edu

MS23

A smoothing-based decomposition for nonlinear nonconvex two-stage optimization

We present a decomposition framework for nonlinear nonconvex continuous optimization. As distinguishing feature, we use a smoothing technique based on barrier methods to render the response of the second-stage problem differentiable. This general setup makes it possible to utilize efficient implementations of existing nonlinear optimization solvers for both the master and the subproblems. We discuss local convergence properties of the method and present some numerical experiments.

Andreas Wächter
Northwestern University
IEMS Department
waechter@iems.northwestern.edu

Xinyi Luo
Northwestern University
xinyiluo2023@u.northwestern.edu

Shenyinying Tu
LinkedIn
shenyinyingtu2021@u.northwestern.edu

Ermin Wei
Northwestern University
ermin.wei@northwestern.edu

MS24

Scalable Projection-Free Optimization Methods Via Multicenter Radial Duality

We will discuss an approach to constrained optimization replacing orthogonal projections with simple one-dimensional linesearches with respect to known feasible points satisfying each constraint. Through a radial reformulation, any original constrained, potentially non-Lipschitz optimization problem can be cast as a root-finding problem over a series of unconstrained Lipschitz subproblems (extending the prior works of Renegar and Grimmer which required one point feasible to every constraint). This bares notable similarities to classic level-set methods, but with improved conditioning/Lipschitz continuity structure. Based on this, we develop new projection-free first-order methods that provably converge towards optimality while maintaining feasibility. Applications to solving second-order cone programs will be presented, for which we show acceleration based on the smoothness and strong convexity of the underlying constraints.

Benjamin D. Grimmer
Cornell University
grimmer@jhu.edu

Thabo Samakhoana
Johns Hopkins University
tsamakh1@jhu.edu

MS24

Robust Nonconvex Optimization in the Presence of Stochastic Constraints

We consider the problem of robust (nonconvex) optimization in the presence of constraints that are stochastic. Such optimization problems arise in modern applications such as training fair machine learning models. For such high-dimensional applications, first-order methods are desirable due to their scalability. In this talk, we first discuss how certain existing algorithms such as the Alternating Direction Method of Multipliers or the penalty-based methods can be extended to this regime. We provide extensive numerical experiments on the scalability and sensitivity of these methods. We also discuss some convergence analyses on these algorithms.

Meisam Razaviyayn
University of Southern California
razaviya@usc.edu

MS24

Policy Gradient and Duality: Estimation, Convergence and Beyond

Consider the estimation of policy gradient (PG) using data that maybe collected in off-policy manners. The talk presents a Policy Gradient Bellman Equation and shows how to leverage Q function approximator for policy gradient estimation via a double fitted iteration. We show that the double fitted iteration is equivalent to a plug-in estimator. Further, the PG estimation error bound is determined by a restricted chi-square divergence that quantifies the interplay between distribution shift and function approximation. A matching Cramer Rao lower bound is also provided. Next, we show how to extend policy gradient method to reinforcement learning beyond cumulative rewards. In particular, consider the objective that is a general concave utility function of the state-action occupancy measure, containing as special cases maximal exploration, imitation and safety constrained RL. Such generality invalidates the Bellman equation and Suttons Policy Gradient Theorem. We derive a new Variational Policy Gradient Theorem for RL with general utilities, which establishes that the parametrized policy gradient may be obtained as the solution of a stochastic saddle point problem. We also exploit the hidden convexity of the Markov decision process and prove that the variational policy gradient scheme converges globally to the optimal policy for the general objective, though the optimization problem is nonconvex.

Mengdi Wang
Princeton University
mengdiw@princeton.edu

MS25

Improving the Geometry of Linear Optimization Problems for the Primal-Dual Hybrid Gradient Method

The primal-dual hybrid gradient method (PDHG) (with restarts and heuristic enhancements) has shown significant success in solving huge-scale LP problems, and our recent research has shown that this success relies on certain geometric condition measures of the LP related to error ratios and sharpness. However, these condition measures may take on extreme values and lead to poor performance both in theory and in practice, which begs the question of whether these geometric condition measures can be improved by applying suitable transformations of the given LP instance? In this study we show in particular how column rescaling can improve these condition measures in theory (and lead to improved performance of PDHG in practice). Our theoretical development leads to guidelines for practical implementation of re-scaling based on ideas from analytic centers. Also, our experiments on LP relaxations of the MIPLIB 2017 dataset demonstrate the impact of rescaling on the actual performance of PDHG.

Robert M. Freund
MIT
Sloan School of Management
rfreund@mit.edu

Zikai Xiong
MIT
Operations Research Center
zikai@mit.edu

MS25

On the Geometry and Refined Rate of Primal-Dual Hybrid Gradient for Linear Programming

Linear programming (LP) is a fundamental class of optimization problems with various applications. Recently, there is a new trend of researches on using first-order methods (FOMs) for LP with the goal to scale up LP by taking advantage of distributed computing. A notable example is the implementation of PDLP, an FOM LP solver that is based on primal-dual hybrid gradient (PDHG). Despite its numerical success, the theoretical understanding of PDHG for LP is far from complete; its existing complexity result depends on the global Hoffman constant of the KKT system, which is known to be loose and uninformative. In this work, we aim to develop a fundamental understanding of the geometric behaviors of PDHG for LP and a refined complexity rate that is not relied on the global Hoffman constant. We show that there are two major stages of PDHG for LP: in Stage I, PDHG identifies active variables and the length of the first phase that is driven by a certain quantity which measures the closeness to degeneracy; in Stage II, PDHG effectively solves a homogeneous linear inequality system, and the complexity of the second stage is driven by a well-behaved local sharpness constant of the system. This finding is closely related to the concept of partial smoothness in non-smooth optimization, and it is the first complexity result of partial smoothness without the non-degeneracy assumption. Our results suggest that degeneracy itself does not slow down the convergence, but near-degeneracy does.

Jinwen Yang
University of Chicago
Department of Statistics
jinweny@statistics.uchicago.edu

Haihao Lu
University of Chicago
haihao.lu@chicagobooth.edu

MS25

Fast Potential-Reduction Algorithms for Linear Programming

We develop a few first-order, PDHG, and/or dimension-reduced second-order potential reduction techniques for linear programming. The introduction of potential reduction removes the barrier parameter tuning such as in ABIP. We establish the convergence theory and demonstrate their practical performances.

Yinyu Ye
Stanford University
yinyu-ye@stanford.edu

MS26

Ideal Polyhedral Relaxations of Non-Polyhedral Sets.

Modern statistical and decision-making problems call for mixed-integer nonlinear optimization (MINLO) formulations, which inherently lead to non-polyhedral relaxations. There has been a substantial progress in extending and adapting techniques for both the mixed-integer linear optimization and continuous nonlinear literatures, but there may a fundamental limit on the effectiveness of such approaches as they fail to exploit the specific characteristics of MINLO problems. In this talk, we discuss recent progress in studying the fundamental structure of MINLO problems. In particular, we show that such problems have a hidden polyhedral substructure that captures the non-convexities associated with discrete variables. Thus, by exploiting this substructure, convexification theory and methods based on polyhedral theory can naturally be used study non-polyhedral sets. We also provide insights into how to design algorithms that tackle the ensuing relaxations.

Andres Gomez
University of Southern California
University of Southern California
gomezand@usc.edu

MS26

Explicit Convex Hull Description of Bivariate Quadratic Sets with Indicator Variables

We present an explicit description for the convex hull of bivariate quadratic sets with indicator variables in the space of original variables. Computational results will be presented.

Aida Khajavirad
Lehigh University
aida@lehigh.edu

Antonio De Rosa
University of Maryland
anderosa@umd.edu

MS26

On Constrained Mixed-Integer Dr-Submodular Minimization

DR-submodular functions encompass a broad class of functions which are generally non-convex and non-concave. We study the problem of minimizing any DR-submodular function, with continuous and general integer variables, under box constraints and possibly additional monotonicity constraints. We propose valid linear inequalities for the epigraph of any DR-submodular function under the constraints. We further provide the complete convex hull of such an epigraph, which, surprisingly, turns out to be polyhedral. We propose a polynomial-time exact separation algorithm for our proposed valid inequalities, with which we first establish the polynomial-time solvability of this class of mixed-integer nonlinear optimization problems.

Kim Yu, Simge Kucukyavuz

Northwestern University
kim.yu@u.northwestern.edu, simge@northwestern.edu

MS27

Bounds for Multistage Mixed-Integer Distributionally Robust Optimization

Multistage mixed-integer distributionally robust optimization (DRO) forms a class of extremely challenging problems. In this talk, new bounding criteria for this class of difficult decision problems are provided through scenario grouping using the ambiguity sets associated with various commonly used ϕ-divergences and the Wasserstein distance. Our approach does not require any special problem structure such as linearity, convexity, or stagewise independence. Therefore, while we focus on multistage mixed-integer DRO, the proposed bounds can be applied to a wide range of DRO problems including two-stage and multistage, with or without integer variables, convex or non-convex, and nested or non-nested formulations. We provide conditions on how to choose the ambiguity sets' radii of the subgroup problems as well as the radius of the ambiguity set to combine the optimal values of subgroup problems. We also propose a way to dissect the scenario tree and combine the subgroups in a nested fashion in the multistage setting to guarantee lower bounds (LBs). Upper bounds are calculated by fixing some decisions obtained through the LBs, finding a feasible policy, and using the cost of that policy. Numerical results on a multistage mixed-integer production problem show the efficiency of the proposed approach through different choices of partition strategies, ambiguity sets, and levels of robustness.

Guzin Bayraksan
The Ohio State University
bayraksan.1@osu.edu

Francesca Maggioni
Dep. of Mathematics, Statistics Computer Science and Applications, University of Bergamo
francesca.maggioni@unibg.it

Daniel Faccini
University of Bergamo
daniel.faccini@unibg.it

Arthur Yang
Ohio State University

MS27

On the Sparsity of Optimal Linear Decision Rules in Dynamic Robust Optimization

In this talk, we consider the widely-studied class of production-inventory problems from the seminal work of Ben-Tal et al. (2004) on linear decision rules in dynamic robust optimization. We prove that there always exists an optimal linear decision rule for this class of problems in which the number of nonzero parameters in the linear decision rule is equal to a small constant times the number of parameters in a static decision rule. This result demonstrates that the celebrated performance of linear decision rules in such dynamic robust optimization problems can be obtained without sacrificing the simplicity of static decision rules. From a practical standpoint, our result lays a theoretical foundation for the growing stream of literature on harnessing sparsity to develop practicable algorithms for computing optimal linear decision rules in operational planning problems with many time periods.

Bradley Sturt
Department of Information and Decision Sciences
University of Illinois at Chicago
bsturt@uic.edu

Haihao Lu
University of Chicago
haihao.lu@chicagobooth.edu

MS27

New Advances in Algorithms and Complexity Analysis for Multistage Stochastic and Distributionally Robust Optimization

In this talk, we will present some recent advances in the algorithm design and iteration complexity analysis of dual dynamic programming (DDP) type algorithms for solving multistage stochastic and, more broadly, distributionally robust optimization (MS-DRO) problems. In particular, we will present a phi-divergence based algorithmic framework for obtaining exact cuts for multistage stochastic mixed-integer nonlinear programs, and present the first iteration complexity bounds that essentially characterize the relation of the number of stages, the dimension of the nodal problems, and the iteration complexity of DDP type algorithms. We will also present two new variants of the DDP algorithm for solving MS-DRO and the associated complexity bounds. Finally, we will present comprehensive computational experiments that compare the performance of MS-DRO, robust optimization, and risk-averse stochastic optimization.

Andy Sun
Operations Research Center
Massachusetts Institute of Technology
sunx@mit.edu

Shixuan Zhang
Brown University
shixuan_zhang@brown.edu

MS28

Sparse Optimization on Measures with over-Parameterized Gradient Descent

Minimizing a convex function of a measure with a sparsity-inducing penalty is a typical problem arising, e.g., in sparse spikes deconvolution or two-layer neural networks training. We show that this problem can be solved by discretizing the measure and running non-convex gradient descent on the positions and weights of the particles. For measures on a d-dimensional manifold and under some non-degeneracy assumptions, this leads to a global optimization algorithm with a complexity scaling as $\log(1/\epsilon)$ in the desired accuracy ϵ, instead of ϵ^{-d} for convex methods. The key theoretical tools are a local convergence analysis in Wasserstein space and an analysis of a perturbed mirror descent in the space of measures. Our bounds involve quantities that are exponential in d which is unavoidable under our assumptions.

Lénaic Chizat
EPFL

lenaic.chizat@epfl.ch

MS28
Geometry of Linear Convolutional Networks

We study the family of functions that are represented by a linear convolutional neural network (LCN). These functions form a semi-algebraic subset of the set of linear maps from input space to output space. In contrast, the families of functions represented by fully-connected linear networks form algebraic sets. We observe that the functions represented by LCNs can be identified with polynomials that admit certain factorizations, and we use this perspective to describe the impact of the network's architecture on the geometry of the resulting function space. We further study the optimization of an objective function over an LCN, analyzing critical points in function space and in parameter space, and describing dynamical invariants for gradient descent. Overall, our theory predicts that the optimized parameters of an LCN will often correspond to repeated filters across layers, or filters that can be decomposed as repeated filters. We also conduct numerical and symbolic experiments that illustrate our results and present an in-depth analysis of the landscape for small architectures.

Guido F. Montufar
UCLA
guidomontufar@gmail.com

Kathlen Kohn
KTH
kathlen@kth.se

Thomas Merkh
UCLA
tbd@tbd.org

Matthew Trager
New York University
matthew.trager@cims.nyu.edu

MS28
On the Implicit Bias of Initialization Shape: Beyond Infinitesimal Mirror Descent

We propose a general framework for understanding the dynamics and implicit bias of gradient descent on a chosen parametrization in the space of functions, relating it to Riemannian Gradient Flow with a metric tensor specified by the Tangent Kernel, which in turn might or might not relate to Mirror Descent with respect to a particular potential function.

Nati Srebro
Toyota Technological Institute at Chicago
nati@ttic.edu

MS29
An Accelerated Method for Stochastic Optimization with State-Dependent Noise

In this work, we present an adaptive stochastic gradient extrapolation (ASGE) method for solving stochastic optimization problems with state-dependent noise. We show that ASGE can accelerate the vanilla stochastic approximation (SA) methods and achieve optimal convergence regarding the deterministic error (iteration complexity) for both convex and strongly convex objective functions. We establish convergence guarantees both in expectation and with high probability. Moreover, we develop a multi-epoch ASGE method for recovery of sparse solutions to stochastic optimization problems with state-dependent noise, e.g., spare generalized linear regression.

TianJiao Li
Georgia Institute of Technology
tli432@gatech.edu

MS29
Maximizing Sum of Coupled Traces with Applications

This paper concerns maximizing the sum of coupled traces of quadratic and linear matrix forms. The coupling comes from requiring the matrix variables in the quadratic and linear matrix forms packed together to have orthonormal columns. At a maximum, the KKT condition becomes a nonlinear polar decomposition (NPD) of a matrix-valued function with dependency on the orthogonal polar factor. An self-consistent-field (SCF) iteration, along with a local optimal CG (LOCG) acceleration, are proposed to compute the NPD. It is proved both methods are convergent and the LOCG acceleration is extremely effective. As applications, we numerically demonstrated our methods on the MAXBET subproblem and the MvPC subproblem, both of which sit at the computational kernels of two multi-view subspace learning models, respectively.

Li Wang
Department of Mathematics
University of Texas at Arlington
li.wang@uta.edu

Leihong Zhang
Soochow University
longzlh@suda.edu.cn

Rencang Li
University of Texas at Arlington
rcli@uta.edu

MS29
Beyond Wasserstein DRO: Gradient Structures for Distributionally Robust Optimization

The principles of robust machine learning (ML) under the so-called distribution shift are far from clear. This shift can be a consequence of causal confounding, unfairness due to data biases, and adversarial attacks. In such cases, recent optimizers adopt robustification strategies derived from distributionally robust optimization (DRO). For example, one of the most interesting directions of DRO is the adoption of the optimal transport distance, the Wasserstein distance. While the Wasserstein DRO literature has exploded, the actual robustness gained from DRO for ML is often no better than existing regularization techniques. To make matters worse, many state-of-the-art DRO methods based on optimal transport place severe limitations on the learning models and scalability, making them inapplicable to modern ML tasks. In this talk, I will introduce mathematical tools beyond the Wasserstein DRO using unbalanced optimal transport and kernel geometry. I will also discuss ML applications such as robust learning under distribution shift.

Jia-Jie Zhu
Weierstrass Institute for Applied Analysis and Stochastics

zhu@wias-berlin.de

MS30

Stochastic Variance Reduction for Variational Inequality Methods

We propose stochastic variance reduced algorithms for solving convex-concave saddle point problems, monotone variational inequalities, and monotone inclusions. Our framework applies to extragradient, forward-backward-forward, and forward-reflected-backward methods both in Euclidean and Bregman setups. All proposed methods converge in the same setting as their deterministic counterparts and they either match or improve the best-known complexities for solving structured min-max problems. Our results reinforce the correspondence between variance reduction in variational inequalities and minimization. We also illustrate the improvements of our approach with numerical evaluations on matrix games.

Ahmet Alacaoglu
University of Wisconsin-Madison
alacaoglu@wisc.edu

Yura Malitsky
Linkoping University
y.malitsky@gmail.com

MS30

Stochastic Differentially Private and Fair Learning Via Min-Max Optimization

Machine learning models are increasingly used in high-stakes decision-making systems. In such applications, a major concern is that these models sometimes discriminate against certain demographic groups such as individuals with certain race, gender, or age. Another major concern in these applications is the violation of the privacy of users. While fair learning algorithms have been developed to mitigate discrimination issues, these algorithms can still leak sensitive information, such as individuals' health or financial records. Utilizing the notion of differential privacy (DP), prior works aimed at developing algorithms that are both private and fair. However, existing DP fair learning algorithms are either not guaranteed to converge or require full batch of data in each iteration. In this paper, we provide the first stochastic DP algorithm for fair learning that is guaranteed to converge. Our framework permits different fairness notions, including demographic parity and equalized odds. In addition, our algorithm can be applied to non-binary classification tasks with multiple (non-binary) sensitive attributes. As a byproduct of our convergence analysis, we provide the first utility guarantee for a DP algorithm for solving nonconvex-strongly concave min-max problems. Our numerical experiments show that the proposed algorithm consistently offers significant performance gains over the state-of-the-art baselines, and can be applied to larger scale problems.

Andrew Lowy
University of Southern California
lowya@usc.edu

Devansh Gupta
Indraprastha Institute of Information Technology, Delhi.
devansh19160@iiitd.ac.in

Meisam Razaviyayn
University of Southern California
razaviya@usc.edu

MS30

A Primal-dual Method for Stochastic Saddle Point Problems

In this talk, we consider a stochastic saddle point problem where the objective function has a finite-sum and nonbilinear structure that commonly arises in various machine learning problems, e.g., robust classification, kernel matrix learning, and generative adversarial network. To contend with the challenge of computing gradient, we develop a stochastic variance-reduced accelerated primal-dual algorithm. Furthermore, we were able to incorporate Bregman-distance functions in the proximal step of the proposed method which provides significant flexibility in the algorithm and facilitates the proximal computation of important instances of functions. Finally, we investigate the convergence of the proposed method and provide the corresponding complexity results.

Erfan Yazdandoost Hamedani
University of Arizona
erfany@arizona.edu

MS31

FeDXL: Provable Federated Learning for Deep X-Risk Optimization

We tackle a novel federated learning (FL) problem for optimizing a family of X-risks, to which no existing FL algorithms are applicable. In particular, the objective has the form of $E_{z\sim S1} f(E_{z'\sim S2} L(w; z, z'))$, where two sets of data S1, S2 are distributed over multiple machines, $L(\cdot)$ is a pairwise loss that only depends on the prediction outputs of the input data pairs (z, z'), and $f(\cdot)$ is possibly a non-linear non-convex function. This problem has important applications in machine learning, e.g., AUROC maximization with a pairwise loss, and partial AUROC maximization with a compositional loss, etc. The challenges for designing a FL algorithm lie at the non-decomposability of the objective over multiple machines and the interdependency between different machines. We propose two provable FL algorithms (FeDXL) for handling linear and nonlinear f, respectively. To tackle the challenges, we decouple the gradient's components with two types namely active parts and passive parts, where the active parts depend on local data that can be computed with the local model and the passive parts depend on other machines that are communicated/computed based on historical models. We develop a novel theoretical analysis to address the issue of latency of passive parts and interdependency between the local gradient estimators and the involved data. We establish both iteration and communication complexities.

Zhishuai Guo
Texas A&M University
zhishguo@tamu.edu

MS31

Efficient First-Order Methods in Bilevel Optimization

Many machine learning problems such as game theory, meta-learning and hyperparameter optimization can be formulated as a bilevel optimization problem. In this talk, we first introduce a double-looped bilevel optimization method based on Bregman distance. We show that this method achieves a lower computation complexity than the best-known results in the deterministic case. After that, we present a fully single-looped bilevel optimization method that is in addition hessian-free. This method has a lower iterative cost for a subclass of bilevel optimization problems. We further analyze the theoretical convergence rate of the proposed method and the complexity of its stochastic variant.

Heng Huang
University of Pittsburgh
heng.huang@pitt.edu

MS31

Bring Your Own Algorithm for Optimal Differentially Private Stochastic Minimax Optimization

We study differentially private (DP) algorithms for smooth stochastic minimax optimization, with stochastic minimization as a byproduct. The holy grail of these settings is to guarantee the optimal trade-off between the privacy and the excess population loss, using an algorithm with a linear time-complexity in the number of training samples. We provide a general framework for solving differentially private stochastic minimax optimization (DP-SMO) problems, which enables the practitioners to bring their own base optimization algorithm and use it as a black-box to obtain the near-optimal privacy-loss trade-off. Our framework is inspired from the recently proposed Phased-ERM method for nonsmooth differentially private stochastic convex optimization (DP-SCO), which exploits the stability of the empirical risk minimization (ERM) for the privacy guarantee. The flexibility of our approach enables us to sidestep the requirement that the base algorithm needs to have bounded sensitivity, and allows the use of sophisticated variance-reduced accelerated methods to achieve near-linear time-complexity. To the best of our knowledge, these are the first near-linear time algorithms with near-optimal guarantees on the population duality gap for smooth DP-SMO, when the objective is (strongly-)convex(strongly-)concave. Additionally, based on our flexible framework, we enrich the family of near-linear time algorithms for smooth DP-SCO with the near-optimal privacy-loss trade-off.

Liang Zhang
ETH Zürich
liang.zhang@inf.ethz.ch

Kiran K. Thekumparampil
Amazon
kirankoshy@gmail.com

Sewoong Oh
University of Washington
sewoong@cs.washington.edu

Niao He
ETH Zurich
niao.he@inf.ethz.ch

MS32

Applying Modern Software Development Approaches to Accelerate Building An Optimization Package in Industry

Design Explorer is the Boeing enterprise standard tool for design space exploration and optimization. Developed over the last 30 years, first as a FORTRAN/C/Perl tool, then re-envisioned as a FORTRAN/C/C++ core tool exposed as a Jython package or as a Java package. After moving to make the main interface a Python package leveraging the same core tool 2022 we decided to move to a Python first approach, and replace compiled custom code with Python packages, native code, or Python packages that wrap FORTRAN code. We introduced modern software development practices to the team, and have greatly accelerated how quickly we can develop new capabilities and deliver them to our internal customers at Boeing. We will provide an overview of Design Explorer, review the transitions we made, and explain our use of modern software development techniques: Usage of continuous integration and continuous delivery tools (GitLab, PyLint, SonarCube, Pytest, etc.) Strict adherence to Python coding standards, static typing, unit tests, and documentation standards. Code review of all new code submitted, with rapid merging and short cycles. Agile project management and single source of data for reports. Leveraging these technologies and approaches has significantly improved how fast we can develop capabilities, and also improved the maintainability of the code. Enforcing good documentation practices throughout the code allows the automatic generation of end user documentation.

Joerg M. Gablonsky
Mathematics & Engineering Analysis
Phantom Works - The Boeing Company
Joerg.M.Gablonsky@boeing.com

Jeffrey D. Poskin, Tyler Smith, Mikel Woo, Nathan Pitts, Micah Goldade, Anjali Prasad
The Boeing Company
jdposkin@gmail.com, tyler.b.smith@boeing.com, mikel.j.woo@boeing.com, nathan.a.pitts@boeing.com, micah.l.goldade@boeing.com, anjali.prasad@boeing.com

MS32

Recent Trends in Software for Optimization and whats new in HiGHS

The field of software for optimization, including both solvers and modelling systems, is currently very active. New commercial optimization companies have been formed in recent years, major new open-source software projects have been created, and the development of new solvers for specialised mathematical optimization problems continues. This talk will present an overview of the state-of-the-art, and then discuss recent developments in the HiGHS linear optimization software project. Having established itself in recent via world-leading open-source performance on LP and MIP problems, HiGHS is currently developing a new interior point (IPM) solver for large scale sparse LP and convex QP problems. Although general in applicability, this solver is aimed at bridging the gap in performance between the current HiGHS IPM solver and commercial solvers in the context of energy systems modelling. The need for such a solver, and reasons for inferior performance of the current HiGHS IPM solver, will be discussed. The challenges of developing the new solver, as well as the future developments in HiGHS that it will facilitate, will also be outlined.

Julian Hall, Ivet Galabova
University of Edinburgh
jajhall@ed.ac.uk, I.L.Galabova@sms.ed.ac.uk

MS32

OR-Tools MathOpt: Solver Independent Modeling

for Mathematical Optimization at Google

OR-Tools MathOpt is a software library for algebraic modeling of mathematical optimization problems. MathOpt supports solver-independent modeling of a wide range of optimization problems including continuous and mixed-integer problems with linear or quadratic objectives and various classes of constraints (including linear, quadratic, second order cone and some specialized constraints). In this talk, we describe various aspects of MathOpts design and implementation including multi-language support (including C++ and Python), support for solver-independent callbacks and incremental solves, and precise contracts for termination statuses and dual solutions. We illustrate these topics through concrete use cases that consider both performance and code simplicity.

Juan Pablo Vielma
Sloan School of Management
Massachusetts Institute of Technology
jvielma@google.com

MS33

Wasserstein Distributionally Robust Regression in High Dimensions: Performance Analysis and Optimal Tuning of the Parameters

We propose a Wasserstein distributionally robust regression framework to estimate an unknown parameter from noisy linear measurements, and we focus on the task of analyzing the squared error performance of such estimators. Our study is carried out in the modern high-dimensional proportional regime, where both the ambient dimension and the number of samples go to infinity, at a proportional rate which encodes the under/over-parametrization of the problem. Under an isotropic Gaussian features assumption, we first show that the squared error can be precisely recovered as the solution of a convex-concave minimax problem which, surprinsingly, involves at most four scalar variables. We then explain how the resulting minimax problem can be employed to efficiently and optimally tune the ambiguity radius.

Liviu Aolaritei
ETH Zurich
aliviu@ethz.ch

Soroosh Shafieezadeh-Abadeh
Carnegie Mellon University
sshafiee@andrew.cmu.edu

Florian Dorfler
Mechanical Engineering
University of California at Santa Barbara
dorfler@ethz.ch

MS33

Holistic Robust Data-Driven Decisions via Distributionally Robust Optimization

We study precisely what are the sources of overfitting in data-driven decision-making problems. The goal is to define formally what are the desired robustness properties in algorithms and ultimately to develop robust algorithms protecting against all these sources. We identify three overfitting sources naturally present in any dataset: (i) statistical error, as a result of working with finite sample data, (ii) noise, which occurs when the data points are measured only with finite precision, and finally (iii) data misspecification, in which a small fraction of all data may be wholly corrupted. The last two sources can also be seen as natural causes of distribution shift. We show that existing data-driven formulations, such as LASSO and Wasserstein DRO, are typically robust against one of these sources in isolation but do not provide holistic protection against all overfitting sources simultaneously. We design a novel distributionally robust formulation, based on a combination of Kullback-Leibler and Levy-Prokhorov ambiguity sets, which does guarantee such holistic protection. We further show that this novel formulation provides optimal robustness. Our formulation can be applied to any loss function to provide robustness with minor computational cost. We finally apply our formulation to neural networks and show in experiments that the resulting novel robust neural networks considerably outperform state-of-art robust deep learning approaches.

Amine Bennouna, Bart Van Parys
Massachusetts Institute of Technology
amineben@mit.edu, vanparys@mit.edu

MS33

Differential Privacy via Distributionally Robust Optimization

In recent years, differential privacy has emerged as the de facto standard for sharing statistics of datasets while limiting the disclosure of private information about the involved individuals. This is achieved by randomly perturbing the statistics to be published, which in turn leads to a privacy-accuracy trade-off: larger perturbations provide stronger privacy guarantees, but they result in less accurate statistics that offer lower utility to the recipients. Of particular interest are therefore optimal mechanisms that provide the highest accuracy for a pre-selected level of privacy. To date, work in this area has focused on specifying families of perturbations a priori and proving their asymptotic and/or best-in-class optimality. We develop the first class of mechanisms that enjoy non-asymptotic and unconditional optimality guarantees. To this end, we formulate the mechanism design problem as an infinite-dimensional distributionally robust optimization problem. We show that this problem affords a strong dual, and we develop converging hierarchies of finite-dimensional upper and lower bounding problems. Our upper bounds correspond to implementable perturbations whose suboptimality can be bounded by our lower bounds. Both bounding problems can be solved within seconds via cutting plane techniques that exploit the inherent problem structure. Our numerical experiments demonstrate that our perturbations can outperform the previously best results from the literature.

Aras Selvi
Imperial College London
a.selvi19@imperial.ac.uk

Huikang Liu
Shanghai University of Finance and Economics
hkliu2014@gmail.com

Wolfram Wiesemann
Imperial College London
ww@imperial.ac.uk

MS34

Infeasibility Detection with Primal-Dual Hybrid

Gradient for Large-Scale Linear Programming

In this talk, we consider the problem of detecting infeasibility of large-scale linear programming problems using the primal-dual hybrid gradient method (PDHG) of Chambolle and Pock (2011). The literature on PDHG has mostly focused on settings where the problem is assumed to be feasible. When the problem has no solution, the iterates of the algorithm do not converge. In this scenario, we show that the iterates diverge at a controlled rate towards a well-defined ray. The direction of this ray is known as the infimal displacement vector. The first contribution of our work is to prove that this vector recovers certificates of primal and dual infeasibility whenever they exist. Based on this fact, we propose a simple way to extract approximate infeasibility certificates from sequences generated by the iterates of PDHG. Our second contribution is to establish tight convergence rates for these sequences. We prove a convergence rate of $O(k^{-1})$, improving over the known rate of $O(k^{-1/2})$. This rate is general and applies to any fixed-point iteration of a nonexpansive operator. Thus, it is a result of independent interest since it covers a broad family of algorithms, including, for example, ADMM. Further, we show that, under non-degeneracy assumptions, one of these sequences exhibits eventual linear convergence. Numerical experiments support our theoretical findings.

David Applegate
Google
dapplegate@google.com

Mateo Diaz
California Institute of Technology,U.S.
mateodd@caltech.edu

Haihao Lu
University of Chicago
haihao.lu@chicagobooth.edu

Miles Lubin
Hudson River Trading
miles.lubin@gmail.com

MS34

An Operator Perspective on PDHG, with Application to Quadratic Programming

Recently, Applegate et al. have investigated how the primal-dual hybrid gradient (PDHG) method of Chambolle and Pock behaves on infeasible or unbounded instances of linear programming. In particular, they prove that under some assumptions, the method yields a certificate of primal or dual infeasibility that is the minimum-norm point in the fixed-point set of an operator in a certain weighted norm. We provide an alternative point of view on their result, which leads us to a generalization of their results to quadratic programming. If time permits, we will also discuss extensions to other conic problems.

Tao Jiang
Cornell University
tj293@cornell.edu

Walaa M. Moursi
University of Waterloo
walaa.moursi@uwaterloo.ca

Stephen A. Vavasis
University of Waterloo
Dept of Combinatorics and Optimization
vavasis@uwaterloo.ca

MS34

Filling Nonzero Duality Gaps in SDP by Way of Perturbation: Hidden Continuity Behind Nonzero Duality Gaps in SDP

Let us consider a primal-dual pair (P) and (D) of SDP, and assume that both of them are either weakly feasible or weakly infeasible. Then there might be nonzero duality gap between (P) and (D). On the other hand, there exists arbitrary small perturbation on both (P) and (D) which would make perturbed primal and dual strongly feasible so that they have a common optimal value, zeroing the duality gap. Thus, perturbation analysis of the optimal value function under the existence of a nonzero duality gap in SDP is a challenging subject. Recently, we tackled this problem without making any restrictive assumption. Consider perturbation adding matrices proportional to identity matrices in (P) and (D) to make them strongly feasible, and denote by v the optimal value of the perturbed problem. By using the Tarski-Seidenberg Theorem of real algebraic geometry, we proved that the directional limit vlim of v exists when perturbation is reduced to zero along a line keeping the ratio of primal and dual perturbation constant. Furthermore, it was shown that vlim is a monotone decreasing function of the direction of approach (parametrized by an angle between zero and pi/2) whose value ranges from the optimal value of (P) (at angle=0) down to that of (D) (at angle=pi/2). We also proved that vlim is continuous except for the both end, angle=0 and pi/2. In this talk, we review the main results developed so far, and discuss conditions where continuity holds at angle=0 and pi/2.

Takashi Tsuchiya
National Graduate Institute for Policy Studies
tsuchiya@grips.ac.jp

Bruno Lourenço
The Institute of Statistical Mathematics
bruno@ism.ac.jp

Masakazu Muramatsu
The University of Electro-Communications
masakazumuramatsu@uec.ac.jp

MS35

A Descent Algorithm for the Optimal Control of ReLU Neural Network Informed PDEs Based on Approximate Directional Derivatives

We propose and analyze a numerical algorithm for solving a class of optimal control problems for learning-informed semilinear partial differential equations. The latter is a class of PDEs with constituents that are in principle unknown and are approximated by nonsmooth ReLU neural networks. We first show that a direct smoothing of the ReLU network with the aim to make use of classical numerical solvers can have certain disadvantages. This motivates us to devise a numerical algorithm that treats directly the nonsmooth optimal control problem, by employing a descent algorithm inspired by a bundle-free method. Several numerical examples are provided and the efficiency of the algorithm is shown.

Michael Hintermüller
Weierstrass Institute for Applied Analysis and Stochastics

Humboldt University Berlin
michael.hintermueller@wias-berlin.de

MS35

New Newton Differentiability Results for Operators Arising in PDE-Constrained Optimization Involving Nonsmoothness or Uncertainties

The efficient treatment of PDE-constrained problems involving inequalities or nonsmoothness often relies on generalized differentiability concepts such as Bouligand differentiability or Newton differentiability. We analyze these properties for convex functions in normed spaces and relate them to the norm upper semicontinuity of the subdifferential. We also discuss the Newton differentiability of operators of the form $F(x)(p) := f(x,p)$ under suitable conditions on f. Operators of this type have applications in the context of PDE constrained optimization under uncertainty and also for other problem classes connected to PDE constraints. Part of this is joint work with Martin Brokate.

Michael Ulbrich
Technical University of Munich
Chair of Mathematical Optimization
mulbrich@ma.tum.de

MS36

Convergence Rate Analysis of the Gradient Descent-Ascent Method for Convex-Concave Saddle-Point Problems

In this talk, I present our study on the gradient descent-ascent method for convex-concave saddle-point problems. We derive a new non-asymptotic global convergence rate in terms of distance to the solution set by using the semidefinite programming performance estimation method. The given convergence rate incorporates most parameters of the problem, and it is exact for a large class of strongly convex-strongly concave saddle-point problems for one iteration. We also investigate the algorithm without strong convexity, and we provide some necessary and sufficient conditions under which the gradient descent-ascent enjoys linear convergence.

Moslem Zamani
Tilburg University
m.zamani_1@tilburguniversity.edu

Hadi Abbaszadehpeivasti
Tilburg University, The Netherlands
h.abbaszadehpeivasti@tilburguniversity.edu

Etienne Klerk, de
Faculty of Economic Sciences
Tilburg University
e.deklerk@uvt.nl

MS36

Worst-Case Performance of Block Coordinate Descent Algorithms on Coordinate-Wise Smooth Convex Functions

We propose a unifying framework for the automated computer-assisted worst-case analysis of block coordinate descent algorithms for smooth unconstrained convex optimization. We first derive necessary interpolation conditions for the class of functions with coordinate-wise Lipschitz gradients. Those allow us to write Performance Estimation Problems (PEP) to compute worst-case bounds for several algorithms including cyclic coordinate descent (for various step sizes) and alternating minimization. We obtain sublinear rates for both algorithms that improve existing worst-case bounds, in some cases by a large constant factor. We demonstrate the flexibility of our approach by providing bounds that hold for simpler and more natural assumptions than those usually encountered in the literature. Finally, we provide numerical evidence for the fact that a standard scheme that provably accelerates random coordinate descent to a $O(1/k^2)$ complexity is actually inefficient when used in a (deterministic) cyclic algorithm.

Yassine Kamri
UCLouvain, Belgium
yassine.kamri@uclouvain.be

Julien Hendrickx
Université catholique de Louvain
julien.hendrickx@uclouvain.be

Francois Glineur
UCLouvain - CORE - INMA
francois.glineur@uclouvain.be

MS36

Non-Nesterov Acceleration Methods in First-Order Optimization

Since the pioneering work of Nesterov on accelerated gradient methods, finding efficient and optimal first-order methods has been the focus of the study of large-scale optimization. Recently, renewed vitality was injected into this classical line of research through the emergence of computer-assisted proof methodologies. In this talk, we first present recent advances in accelerated first-order optimization algorithms, focusing on the new non-Nesterov acceleration mechanisms. We then present the performance estimation problem (PEP) methodology, the essential computer-assisted tool for discovering the new acceleration mechanisms. Finally, we conclude by presenting the Branch-and-Bound Performance Estimation Programming (BnB-PEP), which significantly advances the computer-assisted proof methodology.

Ernest K. Ryu
Seoul National University, South Korea
ernestryu@snu.ac.kr

MS37

Physics-Based Machine Learning with Differentiable Solvers

Modeling the differential equations governing physical systems is a central aspect of many science and engineering tasks. These governing equations are derived from first-principles, and correspond to the laws of physics. Recently, neural network methods have shown some promise in solving differential equations through adding the underlying governing equations as a soft constraint to the loss function. However, such approaches only approximately enforce the constraints on a physical system. I will first discuss the challenges associated with such an approach. I will then discuss how we overcome these challenges by developing a neural network architecture that incorporates differential equation constrained optimization, which outputs solutions that verify the desired physical constraints exactly over a given spatial and/or temporal domain. I will show that this architecture allows us to accurately and efficiently fit solutions to new problems, and demonstrate this on fluid flow and transport phenomena problems.

Aditi Krishnapriyan
Chemical Engineering and EECS, UC Berkeley
aditik1@berkeley.edu

MS38

Accelerated Primal-Dual Scheme for a Class of Stochastic Nonconvex-Concave Saddle Point Problems

Stochastic nonconvex-concave min-max saddle point problems appear in many machine learning and control problems including distributionally robust optimization, generative adversarial networks, and adversarial learning. In this talk, we consider a class of nonconvex-concave saddle point problems and we propose a novel single-loop accelerated primal-dual algorithm with a convergence guarantee. We demonstrate that the proposed method provides an improvement on the oracle complexity under certain assumptions. To validate our theoretical findings, we provide some preliminary numerical experiments.

Afrooz Jalilzadeh
University of Arizona
afrooz@arizona.edu

MS38

Data-Driven Minimax Optimization with Expectation Constraints

Attention to data-driven optimization approaches, including the well-known stochastic gradient descent method, has grown significantly over recent decades, but data-driven constraints have rarely been studied, because of the computational challenges of projections onto the feasible set defined by these hard constraints. In this paper, we focus on the non-smooth convex-concave stochastic minimax regime and formulate the data-driven constraints as expectation constraints. The minimax expectation constrained problem subsumes a broad class of real-world applications, including two-player zero-sum game and data-driven robust optimization. We propose a class of efficient primal-dual algorithms to tackle the minimax expectation-constrained problem, and show that our algorithms converge at the optimal rate of $O(1/\sqrt{N})$. We demonstrate the practical efficiency of our algorithms by conducting numerical experiments on large-scale real-world applications.

Shuogang Yang
Hong Kong University of Science and Technology
yangsg@ust.hk

Xudong Li
Fudan University
lixudong@fudan.edu.cn

Guanghui Lan
Georgia Institute of Technology
george.lan@isye.gatech.edu

MS38

Escape from the Limit Cycle for Nonconvex-Nonconcave Minimax Optimization

Nonconvex-nonconcave minimax optimization has been the focus of intense research over the last decade due to its broad applications in machine learning and operation research. Unfortunately, most existing algorithms cannot be guaranteed to converge and always suffer from *limit cycles*. In this paper, we develop a provably convergent single-loop algorithm called *doubly smoothed gradient descent ascent method*, which gets rid of the limit cycle for constrained problems under the one-sided Kurdyka-Lojasiewicz condition. We further show that with KL exponent $\theta \in (0, 1)$, the algorithm has an iteration complexity of $\mathcal{O}(\epsilon^{-2\max\{2\theta,1\}})$ for finding a game stationary point. For general nonconvex-concave and convex-nonconcave problems, our algorithm converges to a game stationary point with an iteration complexity of $\mathcal{O}(\epsilon^{-4})$, which matches the best iteration complexity of single-loop algorithms under these settings. To the best of our knowledge, this is the first time that a simple and unified single-loop algorithm is proposed for solving nonconvex-nonconcave, nonconvex-concave, and convex-nonconcave problems. Our numerical results show that the proposed algorithm is efficient and can escape limit cycles when applied to some classical and challenging nonconvex-nonconcave examples.

Taoli Zheng, Linglingzhi Zhu
The Chinese University of Hong Kong
zhengtaoli360@gmail.com, llzzhu@se.cuhk.edu.hk

Anthony Man-Cho So
The Chinese University of Hong Kong, Hong Kong
manchoso@se.cuhk.edu.hk

Jose H. Blanchet, Jiajin Li
Stanford University
jose.blanchet@stanford.edu, jiajinli@stanford.edu

MS39

Infinite-Dimensional Moment-SOS Hierarchy for Nonlinear PDEs

We formulate a class of nonlinear partial differential equations (PDEs) as linear optimization problems on moments of positive measures supported on infinite-dimensional vector spaces. Using sums of squares (SOS) representations of polynomials in the space of distributions, we can prove convergence of a hierarchy of semidefinite finite-dimensional truncations solving approximately these infinite-dimensional optimization problems. As an illustration, we report on numerical experiments for solving the heat equation subject to a nonlinear perturbation.

Didier Henrion
LAAS-CNRS, University of Toulouse, France
and Czech Technical University in Prague, Czechia
henrion@laas.fr

Maria Infusino
University of Cagliari, Italy
maria.infusino@unica.it

Salma Kuhlmann
University of Konstanz , Germany
salma.kuhlmann@uni-konstanz.de

Victor Vinnikov
Ben Gurion University of the Negev
Beer Sheva, ISRAEL

vinnikov@math.bgu.ac.il

MS39

Generalized Semi-Infinite Programming Problems

We talk about Generalized Semi-Infinite Programming (GSIP) problems, which are given by polynomial functions. We propose Moment-SOS relaxations and polynomial extensions to solve GSIPs. The classical Semi-Infinite Programming (SIP) problem can also be solved as a special case of GSIPs. Numerical experiments are presented to demonstrate the efficiency of the proposed method.

Xiaomeng Hu, Jiawang Nie
University of California, San Diego
x8hu@ucsd.edu, njw@math.ucsd.edu

MS39

Combined Approach with Second-Order Optimality Conditions for Bilevel Programming Problems

In this work, we propose a combined approach with second-order optimality conditions of the lower level problem to study constraint qualifications and optimality conditions for bilevel programming problems. The new method is inspired by the combined approach developed by Ye and Zhu in 2010, where the authors combined the classical first-order and the value function approaches to derive new necessary optimality conditions. In our approach, we add a second-order optimality condition to the combined program as a new constraint. We show that when all known approaches fail, adding the second-order optimality condition as a constraint makes the corresponding partial calmness condition and the resulting necessary optimality condition easier to hold. We also give some discussions on advantages and disadvantages of the combined approaches with the first-order and the second-order information.

Xiaoxiao Ma
University of Victoria
xiaoxiaoma@uvic.ca

Wei Yao
Southern University of Science and Technology
yaow@sustech.edu.cn

Jane Ye
University of Victoria
Dept of Math & Statistics
janeye@uvic.ca

Jin Zhang
Southern University of Science and Technology, Shenzhen
zhangj9@sustech.edu.cn

MS40

Partially Observable RL with B-Stability: Unified Structural Condition and Sharp Sample-Efficient Algorithms

Partial Observability, where agents can only observe partial information about the true underlying state of the system, is ubiquitous in real-world applications of Reinforcement Learning (RL). Theoretically, learning a near-optimal policy under partial observability is known to be hard in the worst case due to an exponential sample complexity lower bound. While recent work has identified several tractable subclasses that are learnable with polynomial samples, a more unified understanding of when such sample-efficient learning is possible is still lacking, and the existing rates for the known tractable subclasses are far from sharp. In this talk, we will present our recent advances in Partially Observable RL in both aspects above, in the general setting of Predictive State Representations (PSRs). First, we propose a natural and unified structural condition for PSRs called B-stability. B-stable PSRs encompasses the vast majority of known tractable subclasses such as weakly revealing POMDPs, low-rank future-sufficient POMDPs, decodable POMDPs, and regular PSRs. Next, we show that any B-stable PSR can be learned with polynomial samples in relevant problem parameters. When instantiated in the aforementioned subclasses, our sample complexities improve substantially over the current best ones. Finally, we will also present new information-theoretic lower bounds for learning in revealing POMDPs—a widely-studied tractable subclass of POMDPs, which admit a mild gap from the upper bounds obtained by our algorithms, providing a solid starting point for future studies.

Yu Bai
Salesforce Research, U.S.
yu.bai@salesforce.com

Fan Chen
Peking University, China
chern@pku.edu.cn

Song Mei
University of California, Berkeley, U.S.
songmei@berkeley.edu

MS40

Faster Last-Iterate Convergence of Policy Optimization in Zero-Sum Markov Games

Multi-Agent Reinforcement Learning (MARL) – where multiple agents learn to interact in a shared dynamic environment – permeates across a wide range of critical applications. While there has been substantial progress on understanding the global convergence of policy optimization methods in single-agent RL, designing and analysis of efficient policy optimization algorithms in the MARL setting present significant challenges, which unfortunately, remain highly inadequately addressed by existing theory. In this paper, we focus on the most basic setting of competitive multi-agent RL, namely two-player zero-sum Markov games, and study equilibrium finding algorithms in both the infinite-horizon discounted setting and the finite-horizon episodic setting. We propose a single-loop policy optimization method with symmetric updates from both agents, where the policy is updated via the entropy-regularized optimistic multiplicative weights update (OMWU) method and the value is updated on a slower timescale. We show that, in the full-information tabular setting, the proposed method achieves a finite-time last-iterate linear convergence to the quantal response equilibrium of the regularized problem, which translates to a sublinear last-iterate convergence to the Nash equilibrium by controlling the amount of regularization. Our convergence results improve upon the best known iteration complexities, and lead to a better understanding of policy optimization in competitive Markov games.

Shicong Cen
Carnegie Mellon University

shicongc@andrew.cmu.edu

MS40

Policy Optimization in Adversarial MDPs: Improved Exploration Via Dilated Bonuses

Policy optimization is a widely-used method in reinforcement learning. Due to its local-search nature, however, theoretical guarantees on global optimality often rely on extra assumptions on the Markov Decision Processes (MDPs) that bypass the challenge of global exploration. To eliminate the need of such assumptions, in this work, we develop a general solution that adds dilated bonuses to the policy update to facilitate global exploration. To showcase the power and generality of this technique, we apply it to several episodic MDP settings with adversarial losses and bandit feedback, improving and generalizing the state-of-the-art. Specifically, in the tabular case, we obtain $\widetilde{O}(\sqrt{T})$ regret where T is the number of episodes, improving the $\widetilde{O}(T^{2/3})$ regret bound by Shani et al. (2020). When the number of states is infinite, under the assumption that the state-action values are linear in some low-dimensional features, we obtain $\widetilde{O}(\sqrt{T})$ regret with the help of a simulator and a magnitude-reduced loss estimator, improving the $\widetilde{O}(T^{2/3})$ regret of Neu and Olkhovskaya (2020) as well as removing the need of an exploratory policy that their algorithm requires. When a simulator is unavailable, we further consider a linear MDP setting and obtain $\widetilde{O}(T^{8/9})$ regret.

Haipeng Luo
University of Southern California
haipengl@usc.edu

MS41

Demand Balanced Hub Location

We formulate a new clustering problem of allocating demand and supply nodes to facilities so that the amount of supply and demand assigned to a facility are balanced. The resulting clustering problem is an interesting generalization of uncapacitated facility location and minimum cost bipartite matching, and permits interesting algorithmic approaches which we describe.

R Ravi
Carnegie Mellon University
ravi@andrew.cmu.edu

MS41

Designing Scalable Ecommerce Transportation Networks

Omnichannel retail today requires fulfillment of customer orders through a transportation network that is both low-cost and fast. In addition, to provide access to the retailer's entire selection (which may include third parties selling through the retailer's marketplace), the network needs to connect thousands of nodes to end customers, on a continental scale, using large fleets of vehicles. Designing and optimizing such a transportation network presents an algorithmic challenge, because simple solutions (such as a straightforward mixed-integer program) are not scalable. We solve this overall problem through a combination of techniques: decomposing into different pieces which are loosely coupled, reducing the search space by removing decision variable choices that do not make business sense, and accepting different optimality tolerances for different business decisions.

Amitabh Sinha
Amazon
amitabsi@amazon.com

MS41

Generating Data-Driven Uncertainty Sets for Robust Optimization for Large Scale Network Design

In practice, large-scale network design problems typically solve deterministic mixed integer programming models. However, although the resulting scenario is optimal for the deterministic scenario, it may result in poor performance under other scenarios. In this talk, we address two important questions. First, how can we utilize historical data to uncover correlation patterns in origin-destination flows? Addressing this question allow us to define data-driven uncertainty sets for possible scenarios. Second, how can we use the uncertainty set to model and solve a robust network design optimization model? The objective of our work is to design network structures (e.g., path consolidation) that are robust to realistic origin-destination flow variability.

Joline Uichanco
Amazon
jolineu@amazon.com

MS42

A Spectral Conjugate Subgradient Method

The spectral subgradient method is used to solve nonsmooth unconstrained optimization problems. It combines a nonmonotone linesearch and the classical subgradient approach using the spectral step length, which does not require any previous knowledge of the optimal value. We focus on the interesting case in which the objective function is convex and continuously differentiable almost everywhere, and it is often non-differentiable at minimizers. We present a modification of the original spectral subgradient method by adapting a conjugate direction with different choices for the beta parameter. We use performance profiles to compare the performance of the different betas, including beta zero, to represent the original spectral subgradient method.

Milagros Loreto
Division of Engineering and Mathematics, School of STEM, Uni
mloreto@uw.edu

MS42

Bregman Proximal DC Algorithms and Their Application to Blind Deconvolution with Nonsmooth Regularization

Nonconvex optimization problems are known to be difficult to obtain their global or local optimal solution. Besides, it is even challenging to quickly obtain stationary points. In this presentation, we propose efficient algorithms to obtain them for nonconvex optimization problems. For dealing with nonconvex optimization problems, a difference of convex functions (DC) structure has been attractive. Exploiting DC structure for nonconvex optimization problems, we propose the Bregman proximal DC algorithm (BPDCA) and the BPDCA with extrapolation (BPDCAe). These algorithms converge to a limiting stationary point under the L-smooth adaptable (L-smad) property, which is a generalization of L-smoothness. For applications, we consider blind deconvolution. Blind deconvolution is a technique to recover an original signal without knowing a convolving filter. It is naturally formulated as a minimization of a quartic objective function under some assumption. Because its differentiable part is not L-smooth, existing first-order methods are not theoretically supported. In this presentation, we reformulate the objective function as a difference of convex (DC) functions and add nonsmooth regularization. Then, we apply our proposed algorithms, BPDCA and BPDCAe. BPDCAe outperformed other existing algorithms in image deburring applications.

Shota Takahashi
The University of Tokyo
shota@mist.i.u-tokyo.ac.jp

MS43

Accelerating Benders decomposition for solving a sequence of sample average approximation problems

Sample average approximation (SAA) is a technique to get good approximate solutions to stochastic programs (SP). When applying SAA, it is often useful to solve multiple SAA problems to obtain a confidence interval on the true optimal value of the SP and also get a better solution. We study techniques to accelerate the solution of this sequence of SAA problems, when solving them via Benders decomposition. We exploit similarities in the problem structure, as the problems just differ in the realizations of the random samples. Our extensive computational experiments of large scale problems provide empirical evidence of the improved efficiency of our algorithm. In addition, we also present theoretical results that provide insight into the algorithm's performance.

Harshit Kothari, James Luedtke
University of Wisconsin-Madison
hkothari2@wisc.edu, jim.luedtke@wisc.edu

MS43

Piecewise Linear Decision Rules Via Adaptive Partition for Two Stage Stochastic Mixed Integer Linear Programs

Two-stage stochastic mixed-integer linear programming (TS-SMILP) is a very challenging class of problems, especially for problems with mixed-integer second-stage variables. In this presentation, we propose an algorithm based on piecewise linear decision rules for TS-SMILP with continuous distributions. The domain of the uncertainty parameter is partitioned into several subsets. In each subset, the second stage continuous variables adopt a linear decision rule with respect to the uncertain parameters and the second-stage binary variables are constant. Several theoretical properties are proved. First, we prove that there exists a piecewise decision rule that is optimal for TS-SMILP. Second, under uniform partition, the convergence rate of the piecewise decision rule is proved and compared with the convergence rate of sample average approximation. In the proposed algorithm, the partition of the uncertainty set is adaptively updated. Several adaptive partition schemes including uniform partition, strong partition, and reliability partition, are proposed. The proposed algorithms are applied to an energy storage problem under demand and renewable output uncertainty. Numerical experiments suggest that the proposed algorithms have superior computational performance compared with sample average approximation in problems with low-dimension uncertain parameters.

Can Li
Purdue University
canli@purdue.edu

Kibaek Kim
Argonne National Laboratory
kimk@anl.gov

MS43

Lagrangian Dual Decision Rules for Stochastic Programs with Decision-Dependent Uncertainty

We study the stochastic mixed integer programming with endogenous uncertainty. We build a new class of facility location problem with an extra stage at the beginning with binary probing decision that can determine the realization time of uncertainty in the model, whether in the second stage or in the third stage. We implement dual decomposition to relax the non-antipacivity constraints and enable parallel computation. To solve larger instances, we propose a new method called Lagrangian dual decision rule to find a linear combination of basis function as the Lagrangian multiplier, which can significantly reduce the search space.

Zhichao Ma
University of Wisconsin-Madison
zma59@wisc.edu

MS44

Hybrid Classical-Quantum Algorithms for Mixed-Integer Optimization

In a recent paper [arXiv:2207.13630], the authors proposed a hybrid copositive cutting-plane algorithm that leverages heuristic Ising solvers to solve mixed-binary quadratic programs (MBQPs) to global optimality. The approach relies on an exact reformulation of such MBQPs as copositive programs. Despite being convex, these programs are generally intractable since checking copositivity of a matrix is co-NP-complete. We exploit the synergy between checking copositivity, which can be posed as a continuous quadratic minimization problem solved classically, and Ising problems to design a hybrid approach. Ising solvers are responsible for checking the feasibility (i.e., the "hard part") of the copositive program, while the classical computer directs the search towards efficiently reducing the search space. The proposed algorithm does not rely on finding the true ground state of each Ising problem. Instead, it can tolerate inexact solutions, even in the absence of guarantees on the optimality gap. Furthermore, we provide analysis showing that the complexity of the algorithm's portion handled classically has polynomial scaling, suggesting that when applied to NP-hard problems, the solution's complexity is shifted onto the subroutine handled by the Ising solver. In summary, the hybrid copositive cutting-plane algorithm promises to offer resilience to the heuristic nature of novel and potentially quantum Ising solvers while taking advantage of any speed-up they may offer.

David E. Bernal Neira
NASA Ames
dbernalneira@usra.edu

Robin Brown
Stanford University
rabrown1@stanford.edu

Davide Venturelli
Research Institute of Advanced Computer Science, USRA
Quantum Artificial AI Lab NASA Ames
dventurelli@usra.edu

Marco Pavone
Autonomous Systems Laboratory
Stanford University
pavone@stanford.edu

MS45

An Objective-Function-Free Optimization regularization algorithm with optimal complexity rates

An adaptive regularization algorithm for unconstrained nonconvex optimization is presented in which the objective function is never evaluated, but only derivatives are used. This algorithm belongs to the class of adaptive regularization methods, for which optimal worst-case complexity results are known for the standard framework where the objective function is evaluated. It is shown in this paper that these excellent complexity bounds are also valid for the new algorithm, despite the fact that significantly less information is used. In particular, it is shown that, if derivatives of degree one to p are used, the algorithm will find an ϵ_1-approximate first-order minimizer in at most $\mathcal{O}(\epsilon_1^{-(p+1)/p})$ iterations, and an (ϵ_1, ϵ_2)-approximate second-order minimizer in at most $\mathcal{O}(\max(\epsilon_1^{-(p+1)/p}, \epsilon_2^{-(p+1)/(p-1)}))$ iterations. As a special case, the new algorithm using first and second derivatives, when applied to functions with Lipschitz continuous Hessian, will find an iterate x_k at which the gradient's norm is less than ϵ_1 in at most $\mathcal{O}(\epsilon_1^{-3/2})$ iterations.

Serge Gratton
University of Toulouse, IRIT, ANITI, France
serge.gratton@enseeiht.fr

Sadok Jerad
University of Toulouse, ANITI, France
sadok.jerad@toulouse-inp.fr

Phillippe L. Toint
University of Namur, (naXys)
philippe.toint@unamur.be

MS45

Principles of Noisy Nonlinear Optimization

Highly successful methods have been developed in the last few decades for the solution of deterministic optimization problems. It is natural to ask whether these methods need to be drastically redesigned to be robust to noise, or whether relatively small changes suffice. We argue that one can preserve the main underlying properties of nonlinear optimization methods and create robust variants by following three design principles that we discuss in detail. We test the resulting algorithms on constrained and unconstrained optimization problems.

Jorge Nocedal
Department of Industrial Engineering and Management Sciences
Northwestern University
j-nocedal@northwestern.edu

Shigeng Sun, Yuchen Luo
Northwestern University
shigengsun2024@u.northwestern.edu, yuchenlou2026@u.northwestern.edu

MS45

An Augmented Lagrangian Method for Constrained Stochastic Optimization

We develop an Augmented Lagrangian based method for constrained stochastic optimization and derive a worst case complexity analysis. Computational experiments with other recently proposed algorithms are presented.

Michael O'Neill
Lehigh University
mikeoneill@unc.edu

Michael O'Neill
University of North Carolina at Chapel Hill
mikeoneill@unc.edu

MS46

A First Order Primal-Dual Method for Solving Constrained Variational Inequalities

Solving for equilibria is a large problem class that has considerable potential for impactful applications ranging from robustifications of standard minimization, GANs, and multi-agent reinforcement learning, to economics, social or market dynamics, and finance. Furthermore, often we have constraints imposed on the decision variables, such as fixed or limited resources or ensuring that the solution belongs to the probability simplex, for example. Variational Inequalities (VIs) are a tool to formalize such general problems. Motivated by the lack of a first-order method to solve constrained VI (cVI) problems with general constraints, we propose a framework that combines (i) interior-point methods (IPMs)—to handle general constraints—with (ii) a primal-dual approach known as the alternating direction method of multipliers (ADMM) method, so as to avoid the Newton step of classical interior-point methods. Using this approach, we derive "ACVI"—a first-order ADMM-based interior point method for cVIs. We show that ACVI converges when only assuming that the associated operator is monotone—and not necessarily L-Lipschitz—with a rate that matches the known lower bound for this problem class. Furthermore, when the sub-problems of ACVI are solved approximately, we show that by using a standard warm-start technique the convergence rate stays the same, provided that the errors decrease at appropriate rates.

Tatjana Chavdarova
University of California, Berkeley
tatjana.chavdarova@berkeley.edu

MS46

Survey Descent: A Multipoint Generalization of Gradient Descent for Nonsmooth Optimization

For strongly convex objectives that are smooth, the classical theory of gradient descent ensures linear convergence relative to the number of gradient evaluations. An analogous nonsmooth theory is challenging. Even when the objective is smooth at every iterate, the corresponding local models are unstable and the number of cutting planes invoked by traditional remedies is difficult to bound, leading to convergences guarantees that are sublinear relative to the cumulative number of gradient evaluations. We instead propose a multipoint generalization of the gradient descent iteration for local optimization. While designed with general objectives in mind, we are motivated by a "max-of-smooth' model that captures the subdifferential dimension at optimality. We prove linear convergence when the objective is itself max-of-smooth, and experiments suggest a more general phenomenon.

XY Han
Cornell University
xh332@cornell.edu

Adrian Lewis
Cornell University
ORIE
adrian.lewis@cornell.edu

MS46

An Online Convex Optimization Framework for Convex Bilevel Optimization

We propose a new framework for solving the convex bilevel optimization problem, where one optimizes a convex objective over the optimal solutions of another convex optimization problem. As a key step of our framework, show that the convex bilevel optimization problem can be solved via an online convex optimization (OCO) problem in which the objective function remains fixed while the domain changes over time. We develop two OCO algorithms that work under different assumptions and provide their theoretical convergence rates. Our first algorithm works under minimal convexity assumptions, while our second algorithm is equipped to exploit structural information on the objective function, including smoothness, lack of first-order smoothness, and strong convexity. In the context of convex bilevel optimization, our results lead to rates of convergence in terms of both inner and outer objective functions simultaneously, and in particular without assuming strong convexity in the outer objective function. Specifically, after T iterations, our first algorithm achieves $O(T^{-1/3})$ error bound in both levels, and this is further improved to $O(T^{-1/2})$ by our second algorithm. We illustrate the numerical efficacy of our algorithms on standard linear inverse problems and a large-scale text classification problem.

Lingqing Shen
Carnegie Mellon University
lingqins@andrew.cmu.edu

Nam Ho-Nguyen
Discipline of Business Analytics
The University of Sydney
nam.ho-nguyen@sydney.edu.au

Fatma Kilinc-Karzan
Tepper School of Business
Carnegie Mellon University
fkilinc@andrew.cmu.edu

MS47

Variance Reduction and Low Sample Complexity in Stochastic Optimization via Proximal Point Methods

This paper studies novel variance reduction techniques in stochastic convex composite optimization via proximal point methods. It further applies those techniques to obtain improved sample complexity in high probability results without assuming the standard light-tail conditions. The paper develops both accelerated and unaccelerated proximal points methods. Another interesting result of the paper is that it establishes the iteration complexity of Nesterovs acceleration method with a restart in the deterministic setting.

Jiaming Liang
Yale University
jiaming.liang@yale.edu

MS47

Gauges and Accelerated Optimization over Smooth And/or Strongly Convex Sets

In this talk, we will present feasibility and constrained optimization problems defined over smooth and/or strongly convex sets. These notions mirror their popular function counterparts but are much less explored. In these settings, we propose new scalable, projection-free, accelerated first-order methods. Unlike splitting methods or classic projected methods, we do not require orthogonal projections, only using cheap one-dimensional linesearches and normal vector computations. Despite this, we derive the accelerated convergence guarantees of $O(1/T)$ for strongly convex sets, $O(1/T^{2})$ for smooth sets, and accelerated linear convergence given both. Our algorithms and analysis are based on novel characterizations of the Minkowski gauge of smooth and/or strongly convex sets, which may be of independent interest.

Ning Liu
Johns Hopkins University
nliu15@jhu.edu

Benjamin D. Grimmer
Cornell University
grimmer@jhu.edu

MS48

When Global Optimization Meet Clustering: Challenges and Opportunities

Clustering is a fundamental unsupervised machine learning task that plays a vital role in various fields of applications. In this talk, we concentrate on three fundamental centroid-based clustering problems: k-means, k-medoids, and k-center clustering problems. Although many heuristic algorithms are developed for these tasks, none of these algorithms can guarantee a global optimal solution. This work provides a practical global optimization algorithm for these tasks based on a reduced-space spatial branch and bound scheme. This algorithm can guarantee convergence to the global optimum by only branching on the centers of clusters, which is independent of the datasets cardinality. Notably, our algorithm can address datasets with millions of samples (i.e., 1000 times larger than the state-of-the-art methods in the literature). Compared with heuristic methods, the global optima obtained by our algorithm can reduce the objective function on average by 35% for k-center clustering, but only 1% for k-means and k-medoids clustering.

Yankai Cao
Chemical and Biological Engineering, UBC, Vancouver, BC, Can

yankai.cao@ubc.ca

MS48

A Simple Algorithm for Online Decision Making

Motivated by recent progress on online linear programming (OLP), we study the online decision making problem (ODMP) as a natural generalization of OLP. In ODMP, there exists a single decision maker who makes a series of decisions spread out over a total of T time stages. At each time stage, the decision maker makes a decision based on information obtained up to that point without seeing into the future. The task of the decision maker is to maximize the accumulated reward while overall meeting some predetermined m-dimensional long-term goal (linking) constraints. ODMP significantly broadens the modeling framework of OLP by allowing more general feasible regions (for local and goal constraints) potentially involving both discreteness and nonlinearity in each local decision making problem. We propose a Fenchel dual-based online algorithm for ODMP. At each time stage, the proposed algorithm requires solving a potentially nonconvex optimization problem over the local feasible set and a convex optimization problem over the goal set. Under the uniform random permutation model, we show that our algorithm achieves $O(\sqrt{mT})$ constraint violation deterministically in meeting the long-term goals, and $O(\sqrt{m \log m}\sqrt{T})$ competitive difference in expected reward with respect to the optimal offline decisions. We also extend our results to the grouped random permutation model.

Rui Chen
Cornell Tech
rui.chen@cornell.edu

Oktay Gunluk
Cornell University
ong5@cornell.edu

Andrea Lodi
Cornell Tech
andrea.lodi@cornell.edu

Guanyi Wang
National University of Singapore
guanyi.w@nus.edu.sg

MS48

Two-Sided Assortment Optimization with Simultaneous and Sequential Matches

We study the two-sided assortment problem recently introduced by Rios et al. (2021). A platform must choose an assortment of profiles to show each user on each side of the market in each period. Users can either like/dislike as many profiles as they want, and a match occurs if two users see and like each other, potentially in different periods. The platform's goal is to maximize the expected number of matches generated. Given the complexity of the problem, we focus on the one look-ahead version of the problem (i.e., two periods), and we provide algorithms and performance guarantees for different variants of the problem. Finally, we test the performance of our methods using data from our industry partner (a dating app in the US). We numerically show that allowing simultaneous matches leads to a higher number of matches, but the improvement is small relative to the case with sequential matches.

Ignacio Rios
UT Dallas
ignacio.riosuribe@utdallas.edu

Alfredo Torrico
Cornell University
alfredo.torrico@cornell.edu

Victor Verdugo
Universidad de O'higgins
victor.verdugo@uoh.cl

MS49

Asymptotically Tight Conic Approximations of Two-Sided Chance Constraints under Gaussian Mixture Distribution

In this talk, we present an efficient second-order cone programming (SOCP) approximation of the two-sided chance constraints (TCCs) under Gaussian Mixture (GM) distribution via a piecewise linear (PWL) approximation. Compared to the conventional normality assumption for forecast errors, the GM distribution adds an extra level of accuracy representing the uncertainties. Moreover, we show that our SOCP formulation has adjustable rates of accuracy and its optimal value enjoys asymptotic convergence properties. Furthermore, an algorithm is proposed to speed up the solution procedure by optimally selecting the PWL segments. Finally, we demonstrate the effectiveness of our proposed approaches through a numerical study of AC optimal power flow problems.

Jianqiang Cheng, Abolhassan Mohammadi Fathabad
University of Arizona
jqcheng@arizona.edu, amohammadi@email.arizona.edu

Kai Pan
Hong Kong Polytechnic University
kai.pan@polyu.edu.hk

Boshi Yang
Clemon University
boshiy@clemson.edu

MS49

Probability Quantification and Optimization with Chance-Constraints for Network Routing

We develop techniques to estimate probability of events where the optimal value of a convex program exceeds a given threshold. We leverage recent advances to approximately compute the probability that binary random variables from Bernoulli distributions belong to a specially-structured union of sets. We apply our techniques to the network reliability problem, which quantifies the probability of failure scenarios that cause network utilization to exceed one. We provide preliminary computational evaluation that shows that these techniques significantly improve the quality of bound.

Mohit Tawarmalani
Krannert School of Management
mtawarma@purdue.edu

Ashish Chandra
Purdue University

aashishc@purdue.edu

MS49

Also-X#: Convex Approximations for Distributionally Robust Chance Constrained Programs

This paper proposes a new algorithm, "ALSO-X#," for solving distributionally robust chance constrained programs (DRCCPs). We show that when ALSO-X# admits a unique optimal solution, ALSO-X# always outperforms ALSO-X, where the latter is recently proven to outperform the conditional value-at-risk (CVaR) approximation of a regular chance constrained program (Jiang and Xie, 2022). We also show that ALSO-X# can deliver an optimal solution to a DRCCP and numerically demonstrate its efficiency.

Weijun Xie
Virginia Tech
wxie@gatech.edu

Nan Jiang
Georgia Tech
nanjiang@gatech.edu

MS50

The Power of Preconditioning in Overparameterized Low-Rank Matrix Sensing

We propose ScaledGD(?), a preconditioned gradient descent method to tackle the low-rank matrix sensing problem when the true rank is unknown, and when the matrix is possibly ill-conditioned. Using overparametrized factor representations, ScaledGD(?) starts from a small random initialization, and proceeds by gradient descent with a specific form of damped preconditioning to combat bad curvatures induced by overparameterization and ill-conditioning. At the expense of light computational overhead incurred by preconditioners, ScaledGD(?) is remarkably robust to ill-conditioning compared to vanilla gradient descent (GD) even with overprameterization. Specifically, we show that, under the Gaussian design, ScaledGD(?) converges to the true low-rank matrix at a constant linear rate after a small number of iterations that scales only logarithmically with respect to the condition number and the problem dimension. This significantly improves over the convergence rate of vanilla GD which suffers from a polynomial dependency on the condition number. Our work provides evidence on the power of preconditioning in accelerating the convergence without hurting generalization in overparameterized learning.

Xingyu Xu
Carnegie Mellon University
xingyuxu@andrew.cmu.edu

Yandi Shen
University of Chicago
ydshen@uchicago.edu

Yuejie Chi
Carnegie Mellon University
yuejiechi@cmu.edu

Cong Ma
University of Chicago
congm@uchicago.edu

MS50

Nonconvex Matrix Factorization Is Geodesically Convex: Global Landscape Analysis for Matrix Factorization From a Riemannian Perspective.

We study a general matrix optimization problem with a fixed-rank positive semidefinite (PSD) constraint. We perform the Burer-Monteiro factorization and consider a particular Riemannian quotient geometry in a search space that has a total space equipped with the Euclidean metric. When the original objective f satisfies standard restricted strong convexity and smoothness properties, we characterize the global landscape of the factorized objective under the Riemannian quotient geometry. We show the entire search space can be divided into three regions: (R1) the region near the target parameter of interest, where the factorized objective is geodesically strongly convex and smooth; (R2) the region containing neighborhoods of all strict saddle points; (R3) the remaining regions, where the factorized objective has a large gradient. Our results provide a fully geometric explanation for the superior performance of vanilla gradient descent under the Burer-Monteiro factorization. When f satisfies a weaker restricted strict convexity property, we show there exists a neighborhood near local minimizers such that the factorized objective is geodesically convex. Our conclusions are also based on a result of independent interest stating that the geodesic ball centered at Y with a radius 1/3 of the least singular value of Y is a geodesically convex set under the Riemannian quotient geometry. The convexity radius obtained is sharp up to constants.

Yuetian Luo
University of Chicago
yuetian@uchicago.edu

Nicolas Garcia Trillos
Department of Statistics
University of Wisconsin-Madison
garciatrillo@wisc.edu

MS50

Linear Convergence of a Proximal Alternating Minimization Method with Extrapolation for L1-Norm Principal Component Analysis

A popular robust alternative of the classic principal component analysis (PCA) is the ℓ_1-norm PCA (L1-PCA), which aims to find a subspace that captures the most variation in a dataset as measured by the ℓ_1-norm. L1-PCA has shown great promise in alleviating the effect of outliers in data analytic applications. However, it gives rise to a challenging non-smooth non-convex optimization problem, for which existing algorithms are either not scalable or lack strong theoretical guarantees on their convergence behavior. In this paper, we propose a proximal alternating minimization method with extrapolation (PAMe) for solving a two-block reformulation of the L1-PCA problem. We then show that for both the L1-PCA problem and its two-block reformulation, the Kurdyka-Lojasiewicz exponent at any of the limiting critical points is 1/2. This allows us to establish the linear convergence of the sequence of iterates generated by PAMe and to determine the criticality of the limit of the sequence with respect to both the L1-PCA problem and its two-block reformulation. To complement our theoretical development, we show via numerical experiments on both synthetic and real-world datasets that PAMe is competitive with a host of existing methods. Our results not only significantly advance the convergence theory of iterative methods for L1-PCA but also demonstrate the potential of our proposed method in applications.

Peng Wang
University of Michigan
pengwa@umich.edu

MS51

Projection onto a Capped Rotated Second Order Cone with Applications to Sparse Regression

A closed-form expression for projecting onto a capped rotated-second order cone is proposed. This object arises as a part of the feasible region of the perspective relaxation of mixed integer nonlinear programs (MINLP) with binary indicator variables. As an application for the closed-form for solving this projection, we cosider the perspective relaxation of the sparse regression problem with L0 and L2 penalties. We compare a few different first-order methods using the closed-form expression of the projection and show performance advantages compared with a state-of-the art interior point solver even for moderately large datasets. As an extension, we also consider a generalized variant of the sparse regression problem with a group sparsity penalty.

Ishy Zagdoun, Noam Goldberg
Bar Ilan University
yosefza@walla.com, noam.goldberg@biu.ac.il

MS52

Bi/trilevel Optimization Approach for Hyperparameter Selection

Various machine learning (ML) models are equipped with parameters that need to be prefixed, and such parameters are often called hyperparameters. Needless to say, the prediction performance of ML models significantly depends on the choice of hyperparameters. In this talk, we will show several application examples where bi/trilevel optimization approach is used for hyperparameter selection. One of them is a trilevel hyperparameter learning model in the adversarial setting and we will present its solution method and experimental results.

Akiko Takeda
The University of Tokyo
RIKEN
takeda@mist.i.u-tokyo.ac.jp

Ryo Sato
The University of Tokyo
sato-ryo566@g.ecc.u-tokyo.ac.jp

Mirai Tanaka
The Institute of Statistical Mathematics
mirai@ism.ac.jp

Takayuki Okuno
Seikei University
s03034@cc.seikei.ac.jp

MS52

Capacity Expansion of Renewables Within An Electricity Market

Integration of renewable energy generations is an essential problem for modern power industry. In this paper, we develop a bilevel mixed integer optimization model to study the capacity expansion problem on renewable generations in an electricity market environment. Different from existing formulations, we consider the challenge of uncertain generation associated with the renewable energy sources, which makes this bilevel optimization formulation even more complex. To solve such a sophisticated problem, structural properties are analyzed and several reduction methods are proposed to render the problem computable. Numerical results on typical IEEE test beds show the advantage of bilevel optimization in capacity expansion, and the practical feasibility of the solution methods.

Bo Zeng, Liang Xu
University of Pittsburgh
bzeng@pitt.edu, lix21@pitt.edu

MS52

Maximizing Flexibility in Power Grids: A Trilevel Formulation

Uncertainty is a central concern in the robust operation of power grids since supply and demand are rarely known in advance. To operate power grids robustly, it is crucial to quantify and improve the flexibility of the grid, i.e., the maximal amount of uncertainty under which safe operation can be guaranteed. Flexibility of power systems, in a broad sense of the word, is actively explored in research. However, contrary to the prevalence in process systems literature, flexibility as an explicit measure of the range of permissible uncertain values is typically not considered in the power system literature. In this contribution, we maximize the flexibility of the grid as a concrete indicator for the amount of uncertainty the grid can tolerate while still ensuring nominal operation in the worst case. The resulting flexibility maximization problem is a trilevel optimization problem, including binary variables in the lowest level. We use a specialized solution algorithm for existence-constrained semi-infinite programming problems to solve this challenging problem numerically. We apply the proposed flexibility maximization in the context of DC flow to find both the maximal allowable range of uncertain injections and a range for the manageable exchange between two parts of a network. Results for a small and a medium-sized grid instance are presented.

Aron Zingler
RWTH Aachen University
aron.zingler@avt.rwth-aachen.de

Stéphane Fliscounakis
Reseau de transport delectricite
stephane.fliscounakis@rte-france.com

Patrick Panciatici
RTE France
patrick.panciatici@rte-france.com

Alexander Mitsos
RWTH Aachen University, Process Systems Engineering
Forschungszenturm Jülich, IEK-10
amitsos@alum.mit.edu

MS53

On the Generalized Nash Game Proposed by Rosen

We deal with the generalized Nash game proposed by Rosen, which is a game with strategy sets that are coupled across players through a shared constraint. We show a reformulation of this generalized game as a classical game. We also show that Rosen's result can be deduced from the one given by Arrow and Debreu. Finally, we establish necessary and sufficient conditions for a point to be a generalized Nash equilibrium by means of the Minty variational inequality. [D. Aussel and J. Dutta, Generalized Nash Equilibrium Problem, Variational Inequality and Quasiconvexity, Oper. Res. Lett., 36(4):461–464, 2008] [O. Bueno, C. Caldern, and J. Cotrina, A note on coupled constraint Nash games, Preprint on ArXiv: 2201.04262, pp. 11, 2022] [E. Cavazzuti, M. Pappalardo, and M. Passacantando, Nash equilibria, variational inequalities, and dynamical systems, J. Optim. Theory Appl., 114(3):491–506, 2002] [J.B. Rosen, Existence and uniqueness of equilibrium points for concave n-person games, Econometrica, 33(3):520–534, 1965]

John Cotrina
Universidad del Pacifico
cotrina_je@up.edu.pe

MS53

An Equilibrium Problem for the Analysis of the Integrated Water System

This work studies a regional water market where several water utilities operate. Water utilities can expand their water capacity, reduce their water losses, increase the percentage of wastewater treated. These new investments are rewarded by the water utilities as an increase in the marginal water tariff. All the water utilities maximize their individual profits simultaneously and, as a result, the market may reach an equilibrium; at the same time, a water authority aims to reduce water scarcity, water losses, and untreated water quantities. This behavior can be modeled as a bilevel stochastic programming problem. In the upper-level, the authority wants to minimize the social cost, that is, the increase in the total cost due to the new investments and the social cost for water scarcity. The upper-level problem is constrained by the set of the equilibrium solutions of the water utilities. Each water utility maximizes its own profit, subject to technological constraints. The upper and lower-level problems are interrelated since the upper-level problem determines the optimal water marginal tariff curve to submit to the lower-level problem, whereas the lower-level problem sets the water quantities distributed by the different water utilities and the investments decision variables. The existence of an equilibrium for the bilevel problem has been proved. In addition, taking into account that the problem is NP-Hard, an ad hoc algorithmic procedure has been studied.

Elisabetta Allevi
University of Brescia
elisabetta.allevi@unibs.it

Riccardo Cambini
University of Pisa
riccardo.cambini@unipi.it

Rossana Riccardi, Domenico Scopelliti
University of Brescia
rossana.riccardi@unibs.it, domenico.scopelliti@unibs.it

MS53

Economic Equilibrium in a Dynamical Framework

In the classical framework of Walras, agents have quantities of various goods that can be adjusted through buying and selling in a market subject to the budget constraint that the cost of purchases must match the revenue from sales. A vector of prices is said to furnish an equilibrium with respect to the initial holdings of goods if the total demands of the agents in these circumstances are balanced by the total supplies. But there are serious shortcomings to this portrayal of an equilibrium. It is is static without past or future, incompatible with buying and selling as ordinarily understood, and offers no economically plausible mechanism for how the prices in question might be generated. In a different approach, an equilibrium at a given moment is envisioned as a configuration of goods and prices in which no agent is interested in making adjustments. External influences can add or subtract from the current holdings, though, and thereby create a new incentive for trading. This leads to a dynamic model in continuous time in which the equilibrium evolves according to a differential equation that stems from the agents' utility functions alone. The prices emerge naturally through an idealization of micro-interactions in which agents get together in pairs to trade a single good for money in order to both gain higher utility in their holdings.

R. Tyrrell Rockafellar
University of Washington, U.S.
rtr@uw.edu

MS54

Recent Advances in Mosek

Mosek is a widely used optimization software package for large-scale, possibly mixed-integer linear and conic optimization problems. In this presentation, we will discuss improvements in the recently released Mosek version 10. Moreover, if time permits some computational results will be presented documenting the potential performance improvement obtained when upgrading from Mosek version 9 to version 10.

Erling D. Andersen
MOSEK ApS
Copenhagen, Denmark
e.d.andersen@mosek.com

MS54

Recent Computational Advances for Global Optimization of NonConvex Quadratic Programs

Branching algorithms for solving Mixed Integer Linear Programs (MILPs) have benefitted from over 50 years of computational progress that has dramatically thinned the herd of challenging industrial MILP models. More recently, the last 15 years have seen major progress in using similar branching algorithms to solve nonconvex quadratic programs to global optimality. This talk will discuss some key features of the Gurobi Optimizer for nonconvex quadratic programs, focusing on some of the cuts specific to nonconvex quadratic expressions that have contributed to significant improvements in the computational state of the art.

Ed Klotz
Gurobi
klotz@gurobi.com

MS54

Recent Advances in the Artelys Knitro Optimiza-

tion Solver

Artelys Knitro is a solver focused on large-scale, nonlinear, (potentially non-convex) optimization problems. Knitro offers both interior-point and active-set algorithms for continuous models, as well as tools for handling problems with integer variables and other discrete structures. This talk will highlight some of the latest developments in Knitro focusing on some of the recent advances in solving mixed-integer nonlinear problems, heuristics for finding global (or improved local) solutions for non-convex problems, and techniques for solving large-scale, security-constrained optimal power flow (SCOPF) models with millions of variables from the latest ARPA-E Grid Optimization (GO) Challenge 3 Competition.

Richard A. Waltz
Artelys Corp
richard.waltz@artelys.com

MS55

Robust Optimization for Rank-Dependent Models with Uncertain Probabilities

This paper studies distributionally robust optimization for a broad class of risk measures with uncertainty sets defined by ϕ-divergence measures. The risk measures are allowed to be non-linear in probabilities, are characterized by a Choquet integral induced by a probability weighting function, and include many well-known examples (CVaR, Mean Median Deviation, Gini-type). Optimization for this class of preference functionals is challenging due to their rank-dependent nature. We show that for concave probability weighting functions, the DRO problem can be reformulated into a tractable counterpart. We prove that the conic representability of the reformulated robust counterpart is determined jointly by that of the probability weighting and ϕ-divergence functions. While the number of constraints in general scales exponentially with the dimension of the state space, we provide four different approximation methods to circumvent this curse of dimensionality. They yield tight upper and lower bounds on the optimal value of the exact problem, and are formally shown to converge to the exact solution asymptotically. This is illustrated numerically in two examples given by a robust newsvendor problem and a robust portfolio choice problem. We also analyze non-concave probability weighting functions and show that the corresponding robust optimization problems are typically not equivalent to their concave envelope approximations, as opposed to other settings in the literature.

Guanyu Jin, Roger J.A.Laeven, Dick den Hertog
University of Amsterdam
g.jin@uva.nl, r.j.a.laeven@uva.nl, d.denhertog@uva.nl

MS55

Smoothed f-Divergence Distributionally Robust Optimization: Information-Theoretic Optimality and Complexity Independence

In measuring model discrepancy against empirical data in distributionally robust optimization (DRO), f-divergence is known to face the restriction of absolute continuity due to its involvement of likelihood ratios. On the other hand, distances such as Wasserstein and Levy-Prokhorov operate via support point perturbation and are free of such restriction, but they do not satisfy the information-theoretic optimality enjoyed by KL-divergence, a special f-divergence. We study a new computationally tractable distance by smoothing the f-divergence with Wasserstein or Levy-Prokhorov distance via a variational-representation-type optimization problem. We show that DRO based on such a distance achieves the large deviations asymptotic optimality enjoyed by KL-divergence for general, possibly continuous, ground-truth distributions, and the optimality is at the level of the ambiguity set and thus applies to a wide class of possibly discontinuous objective functions. Intriguingly, the large deviations decay rate in our result is independent of hypothesis class complexity, thus revealing the potential statistical generalization benefit of DRO over conventional empirical optimization that is heavily affected by the hypothesis class complexity.

Zhenyuan Liu
Columbia University
zl2817@columbia.edu

Bart Van Parys
Massachusetts Institute of Technology
vanparys@mit.edu

Henry Lam
Columbia University
henry.lam@columbia.edu

MS55

Maximum Mean Discrepancy Distributionally Robust Nonlinear Chance-Constrained Optimization with Finite-Sample Guarantee

This paper is motivated by addressing open questions in distributionally robust chance-constrained programs (DRCCP) where Wasserstein ambiguity sets have gotten significant attention recently. Specifically, the computational techniques for those programs typically place restrictive assumptions on the constraint functions. The size of the Wasserstein ambiguity sets is often set using costly cross-validation (CV) procedures or conservative measure concentration bounds. In order to address some of these limitations, we propose a practical DRCCP algorithm using kernel maximum mean discrepancy (MMD) ambiguity sets, which we term MMD-DRCCP. Our proposed framework can treat general nonlinear non-convex constraints without using ad-hoc reformulation techniques. Moreover, we present various exact reformulations of the conservative Conditional Value-at-Risk(CVaR) approximation of the DRCCP enabled by the MMD ambiguity set. We provide a finite-sample constraint satisfaction guarantee of a dimension-independent $\mathcal{O}(1/\sqrt{N})$ rate for our method, achievable by a practical algorithm. Further, we propose an efficient bootstrap scheme for constructing sharp MMD ambiguity sets in practice without resorting to CV and validate our new approach empirically in multiple experimental settings.

Yassine Nemmour, Heiner Kremer, Bernhard Schoelkopf
Max Planck Institute for Intelligent Systems
ynemmour@tuebingen.mpg.de,
heiner.kremer@tuebingen.mpg.de, bs@tuebingen.mpg.de

Jia-Jie Zhu
Weierstrass Institute for Applied Analysis and Stochastics

zhu@wias-berlin.de

MS56
LMIs in Matrix Variables

Free real algebraic geometry (RAG) concerns a noncommutative polynomial p, and inequalities involving its evaluation $p(X)$ on matrix variables $X = (X_1, \ldots, X_g)$. The matrices X can have any dimension; so are dimension free. A familiar example is a Riccati inequality based on polynomial whose form does not change but ingredients can be matrices of any compatible size. Free RAG bears on random matrix theory, quantum information theory and systems engineering. We focus on $L(X) \succeq 0$, a Free LMI; call the solution set $\mathcal{D}$ a free spectrahedron. These sets are closed under matrix linear combinations, i.e $\mathcal{D}$ is matrix convex. The talk will describe basic facts, recent results and conjectures about these. There are three natural types of extreme points for free spectrahedra:

$Ordinary\ Euclidean\ extreme \supset matrix\ extreme \supset free\ extreme.$

Matrix convex combinations of (even the smallest class) free extreme points span a free spectrahedron. Extensive experiments show that: Fixing a dimension of a free spectrahedron and randomly generating a linear functional yields a maximizer which is very likely free extreme. One never sees serious evidence that a matrix extreme point exists,hence there was the possibility that they do not exist for free spectrahedra.Alas they do exist. This is work with Aidan Epperly, Eric Evert, and Igor Klep.

J.William Helton
University of California, San Diego
Math Dept.
helton@math.ucsd.edu

Eric Evert
Northwestern University
eric.m.evert@gmail.com

MS56
Any-dimensional convex sets

We develop a systematic framework to think about "free" descriptions of convex sets, which are descriptions of sets which can be instantiated in any dimension. We view such freely described sets as families of sets, one in every relevant dimension, and explain how such descriptions arise from certificates of group symmetry and consistency conditions that the sets in the family must satisfy. We find connections with the recently-identified phenomenon in representation theory called "representation stability", which helps explain the ubiquity of free descriptions. We also use representation stability to explain why the size of many symmetry-reduced conic programs stabilize, unifying some of the arguments in the literature. The goal of this talk is to draw attention to the study of free descriptions and the structures involved in them, and highlight their connections to other areas.

Eitan Levin
CalTech
eitanl@caltech.edu

MS56
Exploiting Symmetry in Optimization Problems

The arithmetic mean/geometric mean inequality (AM/GM inequality) facilitates classes of nonnegativity certificates and of relaxation techniques for polynomials and, more generally, for exponential sums. Here, we present a first systematic study of the AM/GM-based techniques in the presence of symmetries under the linear action of a finite group. We prove a symmetry-adapted representation theorem and develop techniques to reduce the size of the resulting relative entropy programs. We study in more detail the complexity gain in the case of the symmetric group. In this setup, we can show in particular certain stabilization results. We exhibit several sequences of examples in growing dimensions where the size of the reduced problem stabilizes. Finally, we provide some numerical results, emphasizing the computational speedup.

Cordian Riener
UiT The Arctic University of Norway
cordian.riener@uit.no

MS57
Sample Size Estimates for Risk-Neutral Semilinear PDE-Constrained Optimization

We apply the sample average approximation (SAA) approach to risk-neutral optimization problems governed by semilinear elliptic partial differential equations with random inputs. After constructing a compact set that contains the SAA critical points, we derive nonasymptotic sample size estimates for SAA critical points using the covering number approach. Thereby, we derive upper bounds on the number of samples needed to obtain accurate critical points of the risk-neutral PDE-constrained optimization problem through SAA critical points. We quantify accuracy using expectation and exponential tail bounds. Numerical illustrations are presented.

Johannes Milz
Georgia Institute of Technology
johannes.milz@isye.gatech.edu

Michael Ulbrich
Technical University of Munich
Chair of Mathematical Optimization
mulbrich@ma.tum.de

MS57
Optimal design of large-scale nonlinear Bayesian inverse problems under model uncertainty

In this talk we propose a scalable computational framework for optimal experimental design (OED) for Bayesian nonlinear inverse problems governed by partial differential equations (PDEs) under model uncertainty. In particular, we consider inverse problems in which, in addition to the inversion parameters, the governing PDEs include secondary uncertain parameters. We build on the Bayesian approximation error (BAE) framework, to incorporate modeling uncer-tainties in the Bayesian inverse problem, and methods for A-optimal design of infinite-dimensional Bayesian nonlinear inverse problems. The OED problem is formulated as a binary bilevel PDE constrained optimization problem and a greedy algorithm is used to find optimal designs. We demonstrate the effectiveness of the proposed approach for a model inverse problem governed by an elliptic PDE on a three-dimensional domain. Our computational results also highlight the pitfalls of ignoring modeling uncertainties in the OED and/or inference stages.

Noemi Petra
University of California, Merced
npetra@ucmerced.edu

Alen Alexanderian
North Carolina State University, U.S.
alexanderian@ncsu.edu

Ruanui Nicholson
Department of Engineering Science
University of Auckland
ruanui.nicholson@auckland.ac.nz

MS57

Almost sure state constraints: relaxation, online stochastic approximation, and solution statistics

Treating almost sure, pointwise state constraints in PDE-constrained optimization under uncertainty presents a significant challenge both theoretically and practically. We present an approach that starts by reformulating the stochastic state constraint as a (degenerate) scalar expectation constraint. Upon relaxing the bound, we obtain a computationally feasible relaxation of the original problem. The theoretical properties of the relaxed problem are investigated as needed for a sample-based optimization algorithm. This includes uniform bounds on stochastic gradients of the objective and constraint functional. We then present a function-space-based online stochastic approximation (OSA) algorithm with a full convergence theory including rates of convergence for feasibility and the optimal value. These results are then illustrated by a numerical study. In particular, we provide a post-optimization analysis via a Kolmogorov-Smirnov test of the distributions of the random objective and constraint functional.

Thomas M. Surowiec
Department of Mathematics and Computer Science
Philipps University of Marburg
thomasms@simula.no

Drew P. Kouri
Optimization and Uncertainty Quantification
Sandia National Laboratories
dpkouri@sandia.gov

Mathias Staudigl
Maastricht University
Mathias.staudigl@gmail.com

MS58

Finding Counter-Examples in First Order Optimization. Application to the Heavy-Ball Method

While many approaches were developed for obtaining worst-case complexity bounds for first-order optimization methods in the last years, there remain theoretical gaps in cases where no such bound can be found. In such cases, it is often unclear whether no such bound exists (e.g., because the algorithm might fail to systematically converge) or simply if the current techniques do not allow finding them. In this work, we propose an approach to automate the search for cyclic trajectories generated by first-order methods. This provides a constructive approach to show that no appropriate complexity bound exists, thereby complementing the approaches providing sufficient conditions for convergence. Using this tool, we provide ranges of parameters for which some of the famous heavy-ball, Nesterov accelerated gradient, inexact gradient descent, and three-operator splitting algorithms fail to systematically converge, and show that it nicely complements existing tools searching for Lyapunov functions. We will present this framework focusing on the example of the Heavy-ball method and finally answer the question: does Heavy-ball accelerate on smooth strongly convex function?

Baptiste Goujaud, Aymeric Dieuleveut
Ecole Polytechnique, France
baptiste.goujaud@polytechnique.edu,
aymeric.dieuleveut@polytechnique.edu

Adrien Taylor
Inria (France)
Adrien.taylor@inria.fr

MS58

Fast Extra Gradient Methods for Smooth Structured Nonconvex-Nonconcave Minimax Problems

Modern minimax problems, such as generative adversarial network and adversarial training, are usually under a nonconvex-nonconcave setting, so developing an efficient method for such setting is of interest. Recently, two variants of the extragradient (EG) method are studied in that direction. First, a double stepsize EG, named EG+, was proposed under a smooth structured nonconvex-nonconcave setting, with a slow $O(1/k)$ rate on the squared gradient norm, where k denotes the number of iterations. Second, another variant of EG with an anchoring technique, named extra anchored gradient (EAG), was studied under a smooth convex-concave setting, yielding a fast $O(1/k^2)$ rate on the squared gradient norm. Built upon both EG+ and EAG, this work proposes a double stepsize EG with anchoring, named fast extragradient (FEG), that has a fast $O(1/k^2)$ rate on the squared gradient norm for smooth structured nonconvex-nonconcave problems; the corresponding saddle-gradient operator satisfies the negative comonotonicity condition.

Sucheol Lee, Donghwan Kim
Korea Advanced Institute of Science & Technology (KAIST)
csfh1379@kaist.ac.kr, donghwankim@kaist.ac.kr

MS58

Performance Analysis of Gradient Method on Smooth Hypoconvex Functions

We perform the first tight convergence analysis of the gradient method with varying step sizes when applied to smooth hypoconvex (weakly convex) functions. Hypoconvex functions are smooth nonconvex functions whose curvature is bounded and assumed to belong to the interval $[\mu, L]$, with $\mu < 0$. Our convergence rates improve and extend the existing analysis for smooth nonconvex functions with L-Lipschitz gradient (which corresponds to the case $\mu = -L$), and smoothly interpolate between that class and the class of smooth convex functions. We obtain our results using the performance estimation framework adapted to hypoconvex functions, using adapted interpolation conditions. We derive explicit upper bounds on the minimum gradient norm of the iterates for the entire range of step sizes $(0, 2/L)$, prove that these rates are tight when step sizes are shorter or equal to some threshold larger than $1.5/L$ and conjecture the tightness above the threshold.

Finally, we identify the optimal constant step size that minimizes the worst-case of the gradient method applied to hypoconvex functions.

Teodor Rotaru
KULeuven and UCLouvain, Belgium
teodor.rotaru@kuleuven.be

Francois Glineur
UCLouvain - CORE - INMA
francois.glineur@uclouvain.be

Panagiotis Patrinos
KU Leuven
panos.patrinos@esat.kuleuven.be

MS59

Efficient Natural Gradient Method for Large-Scale PDE Optimization Problems

Large-scale optimization is at the forefront of modern applied mathematics with areas of interest, including high-dimensional PDE, inverse problems, machine learning, etc. First-order methods are workhorses for large-scale optimization due to modest computational cost and simplicity of implementation. Nevertheless, these methods are often agnostic to the structural properties of the problem under consideration and suffer from slow convergence, being trapped in bad local minima, etc. Natural gradient descent is an acceleration technique in optimization that takes advantage of the problems geometric structure and preconditions the objective functions gradient by a suitable natural metric. Despite its success in statistical inference and machine learning, the natural gradient descent method is far from a mainstream computational technique due to the computational complexity of calculating and inverting the preconditioning matrix. This work aims at a unified computational framework and streamlining the computation of a general natural gradient flow in the context of PDE optimization problems via the systematic application of efficient tools from numerical linear algebra. We obtain efficient and robust numerical methods for natural gradient flows without directly calculating, storing, or inverting the dense preconditioning matrix. We treat Euclidean, Wasserstein, Sobolev, and FisherRao natural gradients in a single framework for a general loss function.

Levon Nurbekyan
UCLA
lnurbek@math.ucla.edu

MS59

Taming Hyper-Parameter Tuning in Continuous Normalizing Flows Using the JKO Scheme

A normalizing flow (NF) is a mapping that transforms a chosen probability distribution to a normal distribution. Such flows are one of the most popular techniques used for data generation and density estimation in machine learning and data science. The density estimate obtained with a NF requires a change of variables formula, which involves the computation of the Jacobian determinant of the NF transformation. In order to tractably compute this determinant, continuous normalizing flows (CNF) estimate the mapping and its Jacobian determinant using neural ODE. Optimal transport (OT) theory has been successfully used to assist in finding CNFs by formulating it as an OT problem with a soft penalty for enforcing the standard normal distribution as a target measure. This method results in the addition of a hyperparameter, α, that represents the strength of the soft penalty and requires significant tuning. We present JKO-Flow, an algorithm to solve OT-based CNF without the need of tuning α. This is achieved by integrating the OT CNF framework into a Wasserstein gradient flow framework, also known as the JKO scheme. Instead of tuning α, we repeatedly solve the optimization problem for a fixed α effectively performing a JKO update with a time-step α. Hence we obtain a "divide and conquer" algorithm by repeatedly solving simpler problems instead of solving a potentially harder problem with large α.

Alexander Vidal
Colorado School of Mines
vidal@mines.edu

MS60

Some applications of implicit function theorems from variational analysis

We present some applications of implicit function theorems from variational analysis to sensitivity analysis of prominent optimization problems, including the LASSO, square root LASSO or the proximal operator as a function of the proximal parameter. These are leveraged in numerical computations. This is based on joint work with Aaron Berk (McGill), Simone Brugiapaglia (Concordia), Michael Friedlander (UBC), and Ariel Goodwin (Cornell).

Tim Hoheisel
McGill University
Montreal, QC, Canada
tim.hoheisel@mcgill.ca

MS60

An Efficient HPR Algorithm for the Wasserstein Barycenter Problem with $O(Dim(P)/\varepsilon)$ Computational Complexity

We propose and analyze an efficient Halpern-Peaceman-Rachford (HPR) algorithm for solving the Wasserstein barycenter problem (WBP) with fixed supports. While the Peaceman-Rachford (PR) splitting method itself may not be convergent for solving the WBP, the HPR algorithm can achieve an $O(1/\varepsilon)$ non-ergodic iteration complexity with respect to the KarushKuhnTucker (KKT) residual. More interestingly, we propose an efficient procedure with linear time computational complexity to solve the linear systems involved in the subproblems of the HPR algorithm. As a consequence, the HPR algorithm enjoys an $O(\mathrm{Dim(P)}/\varepsilon)$ non-ergodic computational complexity in terms of flops for obtaining an ε-optimal solution measured by the KKT residual for the WBP, where Dim(P) is the dimension of the variable of the WBP. This is better than the best-known complexity bound for the WBP. Moreover, the extensive numerical results on both the synthetic and real data sets demonstrate the superior performance of the HPR algorithm for solving the large-scale WBP.

Defeng Sun
The Hong Kong Polytechnic University, Hong Kong
defeng.sun@polyu.edu.hk

MS60

An optimization algorithm for nonsmooth stochastic optimization with weakly concave objective

In this talk, we present optimization algorithms for a group of constrained nonsmooth optimization problems that arise in solutions to parametric optimization problems and applications such as security-constrained AC optimal power flow problems. These problems typically have objectives that are difference-of-convex (DC) functions but not weakly convex. Consequently, we focus on the weak concavity of the objective to design practical algorithms for large-scale problems. To ensure progress in the outer loop, the constraints are linearized.We present global convergence analysis and numerical examples in the form of two-stage optimization problems. Further, the algorithms are extended to the stochastic setting where the exact function values and subgradients are unavailable. The capabilities of the stochastic algorithm are demonstrated by solving a realistic optimal power flow problem as used in current power grid industry practice.

Jingyi Wang
Lawrence Livermore National Laboratory
7000 East Avenue, Livermore, CA 94550
wang125@llnl.gov

MS61

Perfect Games and Noncommutative Nullstellenstze

The foundation of classical algebraic geometry is the Nullstellensatz. Over the last two decades the basic analogous theorems for matrix (noncommutative) variables have emerged. This talk concerns commuting operator strategies for nonlocal games. The main results are two characterizations, based on a Nullstellensatz, which apply to games with perfect commuting operator strategies. The first applies to all games and reduces the question of whether or not a game has a perfect commuting operator strategy to a question involving left ideals and sums of squares. The second characterization is based on a new Nullstellensatz. It applies to a class of games we call torically determined games, special cases of which are XOR and linear system games. For these games we show the question of whether or not a game has a perfect commuting operator strategy reduces to instances of the subgroup membership problem and, for linear systems games, we further show this subgroup membership characterization is equivalent to the standard characterization of perfect commuting operator strategies in terms of solution groups. Both the general and torically determined games characterizations are amenable to computer algebra techniques, e.g., using semidefinite programming.

Igor Klep
University of Ljubljana, Department of Mathematics
Slovenia
igor.klep@fmf.uni-lj.si

Adam Bene Watts
University of Waterloo
Waterloo, Ontario, CA
adam.benewatts1@uwaterloo.ca

J.William Helton
University of California, San Diego
Math Dept.
helton@math.ucsd.edu

MS61

Rational Dual Certificates for Weighted Sums-of-Squares Polynomials: Algorithms and Complexity Bounds

Sum-of-squares (SOS) polynomials with rational coefficients in the interior of the SOS cones are sums of rational squares, meaning that they can be written as a sum of squares of rational polynomials, allowing for easy certification of their nonnegativity. The complexity of these certificates is measured by their bit sizes but are challenging to bound even in the univariate case. We study this problem in the context of dual certificates, which are rational vectors from the dual-SOS cone that can be interpreted as efficiently verifiable nonnegativity certificates. The talk will outline a very simple algorithm to compute a sequence of SOS lower bounds converging linearly to the optimal SOS lower bound and corresponding "small" rational dual certificates, whose bit sizes can be bounded as a function of parameters such as the degree and the number of variables of the polynomials, or their distance from the boundary of the cone. After providing a general complexity bound, we explore a few frequently studied special cases.

David Papp, Maria Davis
North Carolina State University
dpapp@ncsu.edu, mlmacaul@ncsu.edu

MS61

Rational Generalized Nash Equilibrium Problems

We study generalized Nash equilibrium problems that are given by rational functions. Rational expressions for Lagrange multipliers and feasible extensions of KKT points are introduced to compute generalized Nash equilibria (GNEs). We give a hierarchy of rational optimization problems to solve rational generalized Nash equilibrium problems. The existence and computation of feasible extensions are studied. The Moment-SOS relaxations are applied to solve the rational optimization problems. Under some general assumptions, we show that the proposed hierarchy can compute a GNE if it exists or detect its nonexistence. Numerical experiments are given to show the efficiency of the proposed method.

Jiawang Nie
University of California, San Diego
njw@math.ucsd.edu

Xindong Tang
The Hong Kong Polytechnic University
xindong.tang@polyu.edu.hk

Suhan Zhong
University of California, San Diego
suzhong@tamu.edu

MS62

On Policy Optimization Methods for Control

Policy Optimization methods enjoy wide practical use in reinforcement learning (RL) for applications ranging from robotic manipulation to game-playing, partly because they are easy to implement and allow for richly parameterized policies. Yet their theoretical properties, from optimality to statistical complexity, are still not fully understood. To help develop a theoretical basis for these methods, and to bridge the gap between RL and control theoretic approaches, recent work has studied whether gradient-based policy optimization can succeed in designing feedback control policies. In this talk, we start by showing the convergence and optimality of these methods for linear dynamical systems with quadratic costs, where despite nonconvexity, convergence to the optimal policy occurs under mild assumptions. Next, we make a connection between convex parameterizations in control theory on one hand, and the Polyak-Lojasiewicz property of the nonconvex cost function, on the other. Such a connection between the nonconvex and convex landscapes provides a unified view towards extending the results to more complex control problems.

Maryam Fazel
University of Washington
mfazel@uw.edu

MS62

Direct Policy Search for Robust Control: A Nonsmooth Optimization Perspective

Direct policy search has been widely applied in modern reinforcement learning and continuous control. However, the theoretical properties of direct policy search on nonsmooth robust control synthesis have not been fully understood. The optimal H-infinity control framework aims at designing a policy to minimize the closed-loop H-infinity norm, and is arguably the most fundamental robust control paradigm. In this talk, we discuss the recently developed global convergence theory of direct policy search on the H-infinity state-feedback control design problem. Notice that policy search for optimal H-infinity control leads to a constrained nonconvex nonsmooth optimization problem, where the nonconvex feasible set consists of all the policies stabilizing the closed-loop dynamics. We show that for this nonsmooth optimization problem, all Clarke stationary points are global minimum. Next, we identify the coerciveness of the closed-loop H-infinity objective function, and prove that all the sublevel sets of the resultant policy search problem are compact. Based on these properties, we show that Goldsteins subgradient method and its implementable variants can be guaranteed to stay in the nonconvex feasible set and eventually find the global optimal solution of the H-infinity state-feedback synthesis problem. We conclude our discussion with several possible extensions.

Bin Hu
UIUC
binhu7@illinois.edu

MS62

Policy Optimization for Optimal Control with Robustness and Risk-Sensitive Concerns: Implicit Regularization and Global Convergence

Policy optimization (PO) is a key ingredient for modern reinforcement learning (RL). For control design, certain *constraints* are usually enforced on the policies, accounting for either the stability, robustness, or safety concerns on the system. Policy gradient methods have been shown to converge to the global optimum of linear quadratic regulator (LQR), though it is a nonconvex PO problem. This striking result is built upon the property that the cost function is *coercive*, ensuring that the iterates remain feasible as the cost decreases. We study the convergence theory of PO

for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$-norm *robustness guarantee*. This general framework includes *risk-sensitive* linear control as a special case. One significant new feature of this problem is the *lack of coercivity*, i.e., the cost may have *finite* value around the boundary of the robustness constraint set, breaking the existing analysis for LQR. Interestingly, we show that two PO methods enjoy the *implicit regularization* property, i.e., the iterates preserve the $\mathcal{H}_\infty$ robustness constraint automatically, as if they are regularized by the algorithms. Furthermore, despite the nonconvexity, we show that these algorithms converge to the *globally optimal* policies with *globally sublinear* rates, avoiding all suboptimal stationary points/local minima, and with *locally (super-)linear* rates.

Kaiqing Zhang
University of Maryland
kaiqing@umd.edu

MS63

Capacity Portfolio Optimization in a 2-Sided Freight Marketplace

In the Amazon linehaul transportation marketplace, Amazon acquires capacity to ship truckloads in its middle mile network, and provides capacity to external shippers. Because of Amazons commitment to rapid delivery of goods to its customers, capacity must frequently be acquired long before specific demand for that capacity is identified. Amazons large internal demand for linehaul shipping means that Amazon is well-positioned to do exactly this, acquiring a portion of capacity through medium and long-term contracts before specific demand for that capacity is identified, while delaying the acquisition of capacity for highly uncertain demand. Amazon has developed a transportation marketplace, and a series of models and tools that allow Amazon to optimize its transportation capacity portfolio. This approach enables Amazon to effectively align capacity risk, supply risk and price, and to provide carriers and shippers with tools to effectively manage their operations. We give an overview of this marketplace and survey models that we have built to optimize our capacity portfolio.

Philip Kaminsky
Amazon
UC Berkeley
philipka@amazon.com

Idil Arsik, Timothy Jacobs, Roger Lederman
Amazon
arsik@amazon.com, timojaco@amazon.com, rlled-erm@amazon.com

MS63

Supply Chain Network Planning Under Stochastic Demands

We consider a supply chain activity where demands are modeled as a volume required at a given delivery terminal. Release and due dates, a terminal of origin, and a class are associated with these input demands. The network is described by its node set, terminals acting as warehouses, cross-docks, or delivery centers. The arcs correspond to the lanes operated. We model demands by their weekday quantities over the seven days. Terminals have their activities modeled to consider the processing rate of different classes of demands. The aim of the study is twofold: first, we provide a space-time network representation of the weekly operation of the supply chain where the lanes have a fixed schedule with associated capacities. We assume demands have an independent stochastic behavior, each represented by a probability distribution, to propose a stochastic programming model to determine the expected loss of demands produced by the given schedule. Second, we propose a procedure to search for limited schedule changes implying an increase in performance and a decrease in the expected loss.

Marcus Poggi de Aragao, Daniel Ulch
Amazon
mvpoggi@amazon.com, danulch@amazon.com

MS63

From Inbound to Outbound: Sustainable Supply Chain Network Design via Process Flexibility

We propose a novel computational framework of evaluating the interaction between numerous operational costs and carbon emission affects in the inbound network and outbound network of a supply chain system. In our framework, we adopt an epsilon-constraint method to investigate the cost of sustainability, via restricting the cost-related objectives within certain values while minimizing the total carbon emissions produced by the transportation decisions in the entire network. Our formulation can be regarded as a sampling-based two-stage stochastic program with a simple recourse, where in the first stage, decisions about inventory allocation and truck assignment are made, and in the second stage all customer demands realize and get fulfilled.

Hanzhang Qin
National University of Singapore
Amazon
hzqin@amazon.com

Salal Humair
Amazon
salal@amazon.com

Huseyin Topaloglu
Cornell University
Amazon
htopalog@amazon.com

MS64

TTRISK: Tensor Train Decomposition Algorithm for Risk Averse Optimization

This talk introduces TTRISK, an algorithm to solve high-dimensional risk-averse optimization problems governed by differential equations (ODEs and/or PDEs) under uncertainty. As an example, we focus on the so-called Conditional Value at Risk (CVaR), but the approach is equally applicable to other coherent risk measures. Both the full and reduced space formulations are considered. The algorithm is based on low rank tensor approximations of random fields discretized using stochastic collocation. To avoid non-smoothness of the objective function underpinning the CVaR, we propose an adaptive strategy to select the width parameter of the smoothed CVaR to balance the smoothing and tensor approximation errors. Moreover, unbiased Monte Carlo CVaR estimate can be computed by using the smoothed CVaR as a control variate. To accelerate the computations, we introduce an efficient preconditioner for the KKT system in the full space formulation. Further extensions to state constrainted problems are considered.

Harbir Antil
George Mason University
hantil@gmu.edu

Sergey Dolgov
University of Bath
Department of Mathematical Sciences
s.dolgov@bath.ac.uk

Akwum Onwunta
Lehigh University
ako221@lehigh.edu

MS64

Efficient Bi-Level Optimization for EPI-MRI Distortion Correction and Diffusion Tensor Imaging

Echo-Planar Imaging (EPI), a fast method of Magnetic Resonance Imaging (MRI) acquisition, is widely used for Diffusion Tensor Imaging (DTI) to study the movement of water throughout tissues and other anatomical structures. However, EPI leaves susceptibility artifact distortions in the resulting images that limit the accuracy of DTI analysis. In this talk, we describe epic4D, a tool for distortion correction designed to efficiently handle large four-dimensional DTI inputs. epic4D leverages techniques from optimal transport and parallelization for GPU acceleration to efficiently and accurately solve an optimization problem based on the known physical model for EPI distortions. We show that epic4D is faster on three-dimensional non-diffusion weighted inputs than existing EPI distortion correction tools, and we further show how it can efficiently and accurately perform distortion correction on four-dimensional DTI volumes. Finally, we demonstrate how the efficiency of epic4D on four-dimensional inputs allows for bilevel optimization of hyperparameters.

Abigail Julian
Emory University
abigail.julian@emory.edu

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

MS64

FedCBO: Reaching Group Consensus in Clustered Federated Learning through Consensus-Based Optimization

Federated learning is an important framework in modern machine learning that seeks to combine the training of learning models from multiple users, each with their own local data sets, in a way that is sensitive to data privacy and communication cost constraints. In clustered federated learning one assumes an additional and unknown clustering structure among users, and the goal is to train models that are useful for each cluster of users, rather than training a single global model. In this talk, I will present a novel solution to the problem of clustered federated learning that is inspired by ideas in consensus-based optimization (CBO). Our new CBO-type method is based on a system of interacting particles that is oblivious to cluster memberships. Our model is supported by theoretical justification, including a mean field analysis describing the large particle limit of the model, as well as convergence guarantees for the simultaneous global optimization of general non-convex objective functions (corresponding to the loss functions of each cluster of users) in the mean-field regime. I will also present experimental results that demonstrate the efficacy of the FedCBO algorithm compared to other state-of-the-art methods, thus complementing our theoretical findings.

Sixu Li
University of Wisconsin
sli739@wisc.edu

Nicolas Garcia Trillos
Department of Statistics
University of Wisconsin-Madison
garciatrillo@wisc.edu

Yuhua Zhu
Stanford University
yuz244@ucsd.edu

José A. Carrillo
University of Oxford
carrillo@maths.ox.ac.uk

MS65

A Hybrid Decomposition Method for Capacity Expansion Planning under Uncertainty

This work investigates possible ways of combining column generation, Lagrangian relaxation solved using bundle-type approaches, and Benders decomposition in a hybrid decomposition method. The aim is to improve convergence by exploiting the strengths of the different decomposition methods. The proposed hybrid decomposition is implemented using distributed computing. In this way, each decomposition approach is started in parallel, with the possibility of exchanging bounds and solutions between the methods. The solutions obtained at each iteration of the Benders decomposition and bundle method are used to generate new columns for the column generation approach. Likewise, the columns obtained by column generation are used to derive new cuts for Benders decomposition and the bundle method. The performance of the hybrid decomposition is analyzed on instances of the multi-stage stochastic capacity planning problem of hydrothermal power systems.

Angela Flores
Department of Electrical Engineering, University of Chile
Instituto Sistemas Complejos de Ingeniería (ISCI)
angflore@ing.uchile.cl

MS65

Optimal Power Flow for Resilient Electric Grids under Uncertain Hurricane Forecasts: A Multi-Stage Stochastic Programming Approach

Due to extreme weather disasters in the USA, an estimated 1,542 power outages occurred between 2000 and 2021, each affecting at least 50,000 customers. The process of upgrading the grid infrastructure for resilience against such natural disasters is a slow and expensive undertaking, but it is possible to use weather forecast data to make informed grid operating decisions to mitigate power outages. These weather forecasts are associated with high degrees of uncertainty because climate models are significantly affected by atmospheric measurements, which are typically sparse. In case of hurricanes, forecasts are periodically updated and include parameters that define tracks predicted by the climate simulation ensemble. For each potential hurricane track, the known fragility curves of grid components yield failure probabilities that can be used to generate power disruption scenarios. This is used within a rolling horizon framework of multi-stage stochastic programming to minimize load shed while accounting for uncertainty and the operational constraints of grid operation. We provide an overview of this multi-stage stochastic programming approach and illustrate the value of the stochastic solution using a medium-scale IEEE benchmark instance affected by a Category 5 Atlantic hurricane. We also showcase the importance of co-optimization with restorative actions so that it is possible to take operational decisions that truly minimize power supply interruption to the customer.

Mithun Goutham
The Ohio State University
goutham.1@buckeyemail.osu.edu

Rohit Kannan, Donatella Pasqualini
Los Alamos National Laboratory
rohitk@alum.mit.edu, dmp@lanl.gov

Deepjyoti Deka
Theory Division and the Center for Nonlinear Studies
Los Alamos National Laboratory
deepjyoti@lanl.gov

Harsha Nagarajan, Russell Bent
Los Alamos National Laboratory
harsha@lanl.gov, rbent@lanl.gov

MS65

Chance-Constrained Motion Planning under Wasserstein Ambiguity

Obstacle avoidance constraints (OA) appear frequently in motion planning problems and are non-convex in general. In this talk, we first propose a mixed-integer reformulation for OA and generalize it to chance-constrained OA (CC-OA) and Wasserstein chance-constrained OA (WCC-OA). Then, we derive the convex hull of (OA), (CC-OA), and (WCC-OA), and show that they can be separated efficiently in practice. In the end, we demonstrate the effectiveness and efficiency of the proposed separation approach using Dubin's car example.

Haoming Shen
Industrial and Operations Engineering
University of Michigan
hmshen@umich.edu

Ruiwei Jiang
Industrial and Operations Engineering
University of Michigan
ruiwei@umich.edu

MS66

Solving the Semidefinite Relaxation of QUBOs in Matrix Multiplication Time, and Faster with a Quantum Computer

Recent works on quantum algorithms for solving semidefinite optimization (SDO) problems have leveraged a quantum-mechanical interpretation of positive semidefinite matrices to develop methods that obtain quantum speedups with respect to the dimension n and number of constraints m. While their dependence on other parameters suggests no overall speedup over classical methodologies, some quantum SDO solvers provide speedups in the low-precision regime. We exploit this fact to our advantage, and present an iterative refinement scheme for the Hamiltonian Updates algorithm of Brandão et al. (*Quantum* 6, 625 (2022)) to exponentially improve the dependence of their algorithm on precision. As a result, we obtain a classical algorithm to solve the semidefinite relaxation of Quadratically Unconstrained Binary Optimization problems (QUBOs) in matrix multiplication time, and in time $\mathcal{O}\left(\left(n^{1.5}s^{0.5+o(1)}\right)\cdot \text{polylog}\left(n, \|C\|_F, \frac{1}{\epsilon}\right)\right)$ with a quantum computer using the sparse-access input model, where s is the sparsity parameter of the coefficient matrix C. Using the quantum operator input model, the running time of the algorithm is shown to be $\mathcal{O}\left(ns + n^{1.5}\cdot \text{polylog}\left(n, \|C\|_F, \frac{1}{\epsilon}\right)\right)$.

Brandon Augustino
Lehigh University
bra216@lehigh.edu

Giacomo Nannicini
University of Southern California
g.nannicini@usc.edu

Tamas Terlaky
Lehigh University
Department of industrial and Systems Engineering
terlaky@lehigh.edu

Luis Zuluaga
Lehigh University
luis.zuluaga@lehigh.edu

MS66

Approximation and Hardness Results for Quantum Max Cut

Nature is continually solving optimization problems, in valiantly trying to drive physical systems towards low-energy states. For quantum physical systems, this perspective presents a natural synergy with discrete optimization. We will discuss how classical discrete optimization techniques can be generalized to address fundamental problems in quantum physics. In particular we will present a recent body of work on approximation algorithms and hardness results for Quantum Max Cut, a natural (QMA-)hard quantum generalization of Max Cut that is motivated by quantum physical applications. This talk is aimed at a broad audience and assumes no background in quantum physics.

Ojas Parekh, Kevin Thompson
Sandia National Laboratories
odparek@sandia.gov, kevthom@sandia.gov

MS66

Improving QAOA Via Classical Warm-Starts

In recent years, there has been increased interest in using quantum computing for the purposes of solving problems in combinatorial optimization. We consider the Max-Cut problem and develop a new quantum algorithm that utilizes classically-obtained warm-starts in order to improve upon the standard Quantum Approximate Optimization Algorithm (QAOA) proposed by Farhi et al. in 2014; this talk will discuss both theoretical and experimental results associated with our approach with our main results being that we obtain a 0.658-approximation for Max-Cut, our approach provably converges to the Max-Cut as the circuit depth increases, and (on small graphs) our approach is able to empirically beat the (classical) Goemans-Williamson algorithm at a relatively low quantum circuit-depth.

Reuben Tate
Georgia Tech
rtate30@gatech.edu

Jai Moondra
Georgia Institute of Technology
jmoondra3@gatech.edu

Bryan Gard, Greg Mohler
Georgia Tech Research Institute
bryan.gard@gtri.gatech.edu, greg.mohler@gtri.gatech.edu

Swati Gupta
Georgia Tech
swatig@gatech.edu

MS67

An Interior Point Method for Nonlinear Optimization Problems with Noise

We propose an interior-point algorithm for solving nonlinear constrained optimization problems in which the objective and constraints values and derivatives are contaminated by non-diminishing and bounded noise. We identify the key components of the classical algorithm that should be modified to achieve noise tolerance. Given a fixed barrier parameter, we establish that the iterates converge into a neighborhood of stationary points that takes into account the proximity to the boundary of the feasible region. We also present an efficient barrier parameter update strategy and showcase the numerical performance of the algorithm.

Shima Dezfulian
Northwestern University
shimadezfulian2023@u.northwestern.edu

Andreas Wächter
Northwestern University
IEMS Department
waechter@iems.northwestern.edu

Jorge Nocedal
Department of Industrial Engineering and Management Sciences
Northwestern University
j-nocedal@northwestern.edu

MS67

Stochastic Algorithms in the Large: Batch Size Saturation, Stepsize Criticality, Generalization Performance, and Exact Dynamics

In this talk, I will present a framework for analyzing dynamics of stochastic optimization algorithms (e.g., stochastic gradient descent (SGD) and momentum (SGD+M)) when both the number of samples and dimensions are large. For the analysis, I will introduce a stochastic differential equation, called homogenized SGD. We show that homogenized SGD is the high-dimensional equivalent of SGD – for any quadratic statistic (e.g., population risk with quadratic loss), the statistic under the iterates of SGD converges to the statistic under homogenized SGD when the number of samples n and number of features d are polynomially related. By analyzing homogenized SGD, we provide exact non-asymptotic high-dimensional expressions for the training dynamics and generalization performance of SGD. The analysis is formulated for data matrices and target vectors that satisfy a family of resolvent conditions, which can roughly be viewed as a weak form of delocalization of sample-side singular vectors of the data. By analyzing these limiting dynamics, we can provide insights into learning rate, momentum parameter, and batch size selection. Finally we show this model matches performances on real data sets.

Courtney Paquette
McGill University and Google Brain Team
cypaquette@google.com

Elliot Paquette, Kiwon Lee, Andrew Cheng
McGill University
elliot.paquette@mcgill.ca, kiwon.lee@mail.mcgill.ca, andrew.cheng@mail.mcgill.ca

Ben Adlam, Jeffrey Pennington
Google Research, Brain Team
adlam@google.com, jpennin@google.com

MS67

Stochastic Optimization in Relative Scale

We propose a new concept of relatively inexact stochastic subgradients and present novel first-order methods that can use these objects to approximately solve convex optimization problems in relative scale. An important example where such inexact subgradients naturally arise is given by the Power or Lanczos algorithms for computing an approximate leading eigenvector of a symmetric positive semidefinite matrix. Using these algorithms as subroutines in our methods, we get new optimization schemes that can provably solve certain large-scale Semidefinite Programming problems with relative accuracy guarantees by using only matrix-vector products. One interesting application of our algorithms is the approximate solution of the Max-Cut problem.

Anton Rodomanov
UCLouvain
anton.rodomanov@uclouvain.be

Yurii Nesterov
CORE/INMA
Catholic University of Louvain (UCL)
yurii.nesterov@uclouvain.be

MS68

SAPD+: An Accelerated Stochastic Method for Nonconvex-Concave Minimax Problems

We propose a new stochastic method SAPD+ for solving nonconvex-concave minimax problems of the form minmax $L(x,y) = f(x) + \Phi(x,y) - g(y)$, where f,g are closed convex and $\Phi(x,y)$ is a smooth function that is weakly convex in x, (strongly) concave in y. For both strongly concave and merely concave settings, SAPD+ achieves the best known oracle complexities of $\mathcal{O}(L\kappa\epsilon^{-4})$ and $\mathcal{O}(L^3\epsilon^{-6})$, respectively, without assuming compactness of the problem domain, where κ is the condition number and L is the Lipschitz constant. We also propose SAPD+ with variance reduction, which enjoys the best known oracle complexity of $\mathcal{O}(L\kappa^2\epsilon^{-3})$ for weakly convex-strongly concave setting. We demonstrate the efficiency of SAPD+ on a distributionally robust learning problem with a weakly convex cost and also on a multi-class classification problem in deep learning.

Xuan Zhang, Necdet S. Aybat
Penn State University
xxz358@psu.edu, nsa10@psu.edu

Mert Gurbuzbalaban
Rutgers University
mert.gurbuzbalaban@gmail.com

MS68

Beyond the Golden Ratio for Variational Inequality Algorithms

We revisit the golden ratio algorithm for solving variational inequalities and min-max problems which has the distinctive feature of adapting the step sizes to the local Lipschitz constants. First, we establish the equivalence of this algorithm with popular VI methods such as optimistic gradient descent-ascent (OGDA), by eliminating the link between the golden ratio and the algorithm. Second, we improve the adaptive version of the algorithm, by removing the maximum step size hyperparameter (an artifact from the analysis) and proving a tighter complexity bound. Third, we adjust it to nonmonotone problems with weak Minty solutions, and show superior performance.

Axel Böhm
University of Vienna
axel.boehm@univie.ac.at

MS68

Solving Smooth Function Constrained Optimization Is As Almost Easy As Unconstrained Optimization

Consider applying first-order methods to solve the smooth convex constrained optimization problem of the form $\min_{x \in X} F(x)$. For a simple closed convex set $X \subseteq \mathbb{R}^n$ which is easy to project onto, Nesterov proposed the Accelerated Gradient Descent (AGD) method to solve the constrained problem as efficiently as an unconstrained problem in terms of the number of gradient computations of F (i.e., oracle complexity). For a more complicated $\mathcal{X}$ described by function constraints, i.e., $\mathcal{X} = \{x \in X : g(x) \leq 0\}$, where the projection onto $\mathcal{X}$ is not possible, it is an open question whether the function constrained problem can be solved as efficiently as an unconstrained problem in terms of the number of gradient computations for F and g. In this paper, we provide an affirmative answer to the question by proposing a single-loop Accelerated Constrained Gradient Descent (ACGD) method. The ACGD method modifies the AGD method by changing the descent step to a constrained descent step, which adds only a few linear constraints to the prox mapping. It enjoys almost the same oracle complexity as the optimal one for minimizing the optimal Lagrangian function, i.e., the Lagrangian multiplier λ being fixed to the optimal multiplier λ^*. These upper oracle complexity bounds are shown to be unimprovable under a certain optimality regime with new lower oracle complexity bounds.

Zhe (Jimmy) Zhang, Guanghui Lan
Georgia Institute of Technology
jimmy_zhang@gatech.edu, george.lan@isye.gatech.edu

MS69

The Inexact Cyclic Block Proximal Gradient Method and Inexact Proximal Maps

In this talk, we expand the Cyclic Block Proximal Gradient method for block separable, composite minimization to allow for inexactly computed gradients and proximal maps. The resultant algorithm, the Inexact Cyclic Block Proximal Gradient (I-CBPG) method, shares the same convergence rate as its exactly computed analogue, provided how the allowable errors decrease. The foundation of our convergence analysis is our proposed d-Second Prox Theorem, which contains a tight relationship between inexact proximal map evaluations and d-subgradients. Further, we highlight numerical experiments that showcase the practical computational advantage of I-CBPG for certain fixed tolerances of approximation error and for a dynamically decreasing error tolerance regime in particular.

Leandro F. Maia, David H. Gutman
Texas Tech University
Leandro.Maia@ttu.edu, david.gutman@ttu.edu

Ryan Christopher Hughes
Addx Corporation
ryan.christopher.hughes@gmail.com

MS69

Rates of convergence of first-order methods via growth properties

We show that the convergence rates of several kinds of first-order methods range from sublinear to linear depending on a suitable *growth property* of the problem. When the method is affine invariant, as in the case of the conditional gradient method, the convergence results and growth property are also affine invariant.

Javier Pena
Carnegie Mellon University
jfp@andrew.cmu.edu

MS69

Proximal Bundle Methods for Weakly Convex Composite Optimization Problems

This talk discusses the iteration-complexity of a prox bundle method for solving weakly convex (and hence possibly nonconvex) nonsmooth composite optimization problems.

Honghao Zhang
Georgia Tech
zhanghonghao2021@gmail.com

Renato Monteiro
Georgia Institute of Technology
rm88@gatech.edu

MS70

Incorporating Dantzig-Wolfe Decomposition Into Branch and Bound by Cutting Planes

Dantzig-Wolfe (DW) decomposition is a well-known technique in mixed integer programming (MIP) for decomposing and convexifying constraints to obtain potentially strong dual bounds. We investigate Fenchel cuts that can be derived using the DW decomposition algorithm and show that these cuts can provide the same dual bounds as DW decomposition. We show that these cuts, in essence, decompose the objective function cut one can simply write using the DW bound. Compared to the objective function cut, these Fenchel cuts lead to a formulation with lower dual degeneracy, and consequently a better computational performance under the standard branch-and-cut framework in the original space. We also discuss how to strengthen these cuts to improve the computational performance further. We test our approach on the Multiple Knapsack Assignment Problem and show that the proposed cuts are helpful in accelerating the solution time without the need to implement branch-and-price.

Oktay Gunluk
Cornell University
ong5@cornell.edu

Rui Chen, Andrea Lodi
Cornell Tech
rui.chen@cornell.edu, andrea.lodi@cornell.edu

MS70

A Covering Decomposition Algorithm for Power Grid Cyber-Network Segmentation

We present a trilevel interdiction model for optimally segmenting the Supervisory Control and Data Acquisition (SCADA) network controlling an electric power grid. In this formulation, we decide how to partition nodes of the SCADA network in order to minimize the shedding of load from a worst-case cyberattack, assuming that the grid operator has the opportunity for recourse mitigating attack damage. The model is unique in that it couples the cyber and physical networks, using the physical operation of the grid to inform the cyber network segmentation decisions. This feature leads to two violations of assumptions made in much of the prior literature on trilevel interdiction: first, the network designer does not have the ability to make any components of the SCADA or physical networks invulnerable to attack. Instead, the designer makes some attacks more expensive for the attacker. Second, it is possible for the network designer (first player) to make certain attacker (second player) solutions infeasible because he can make them exceed the attacker's budget. In this talk, we present a solution procedure for our formulation that is an adaptation of a covering decomposition algorithm for bilevel interdiction. We show through an empirical study on grids with up to 2,000 buses that this is the first method capable of solving the network segmentation model on realistically sized power grids.

Emma Johnson
Sandia National Laboratories
esjohn@sandia.gov

Santanu Dey
Georgia Institute of Technology
santanu.dey@isye.gatech.edu

Jonathan Eckstein
Rutgers University
jeckstei@business.rutgers.edu

Cynthia Phillips
Sandia National Laboratories
caphill@sandia.gov

John D. Siirola
Sandia National Laboratories
Exploratory Simulation Technologies
jdsiiro@sandia.gov

MS70

ALM and ADMM for Block-Structured MILPs

This paper introduces two decomposition-based methods for two-block mixed-integer linear programs (MILPs), which aim to take advantage of separable structures of the original problem by solving a sequence of lower-dimensional MILPs. The first method is based on the ℓ_1-augmented Lagrangian method (ALM), and the second one is based on a modified alternating direction method of multipliers (ADMM). In the presence of certain block-angular structures, both methods create parallel subproblems in one block of variables, and add nonconvex cuts to update the other block; they converge to globally optimal solutions of the original MILP under proper conditions. Numerical comparisons with recent decomposition methods demonstrate advantages of the proposed methods in either solution time or quality, often in both.

Kaizhao Sun
DAMO Academy
Alibaba Group US
kaizhao.s@alibaba-inc.com

Mou Sun
DAMO Academy, Alibaba Group
mou.sunm@alibaba-inc.com

Wotao Yin
Alibaba Group US/DAMO Academy, U.S.
wotao.yin@gmail.com

MS71

Distributionally Robust Decentralized Volt/Var Control For Distribution Grids

We propose a decentralized Volt/Var control method for distribution grids with high penetration of distributed energy resources (DERs). A two-stage distributionally robust optimization (DRO) problem is proposed to determine how to better prepare slow-timescale control measures as well as neighboring local control systems for the worst-case probability distribution of DER outputs. We present a cutting plane method for solving the DRO problem that leverages a discretization scheme for the boundary of the support set and parallel computation. We conduct numerical experiments on various test systems and demonstrate the benefit of the proposed control method and the solution method.

Geunyeong Byeon
School of Computing and Augmented Intelligence
Arizona State University
geunyeong.byeon@asu.edu

MS71

Mitigating the Impacts of Uncertain Geomagnetic Disturbances on Electric Grids: A Distributionally Robust Optimization Approach

Severe geomagnetic disturbances (GMDs) increase the magnitude of the electric field on the Earths surface (E-field) and drive geomagnetically-induced currents (GICs) along the transmission lines in electric grids. These additional currents can pose severe risks to the grid, leading to system unreliability. Several mitigation actions (e.g., placing GIC blocking devices) exist that can reduce the harmful GIC effects on the grids. Making such decisions can be challenging, however, because the magnitude and direction of the E-field are uncertain and non-stationary. In this paper, we model uncertain E-fields using the distributionally robust optimization (DRO) approach that determines optimal transmission grid operations such that the worst-case expectation of the system cost is minimized. We also capture the effect of GICs on the nonlinear AC power flow equations. For solution approaches, we develop an accelerated column-and-constraint generation (CCG) algorithm by exploiting a special structure of the support set of uncertain parameters representing the E-field. Extensive numerical experiments based uiuc-150 system, designed for GMD studies, demonstrates (i) the computational performance of the accelerated CCG algorithm, (ii) the superior performance of distributionally robust grid operations that satisfy nonlinear, nonconvex AC power flow equations and GIC constraints, in comparison with standard stochastic programming-based methods during the out-of-sample testing.

Minseok Ryu
Argonne National Laboratory
mryu@anl.gov

Harsha Nagarajan, Russell Bent
Los Alamos National Laboratory
harsha@lanl.gov, rbent@lanl.gov

MS71

Improving Power Grid Resiliency with Bi-Objective Stochastic Integer Optimization

Designing a power grid that is both efficient on average and resilient to extreme weather events is a critical challenge. Traditional stochastic programming approaches to this are highly sensitive to sampling error due to the presence of low probability events with very high impacts. Stochastic programming also fails to account for system goals changing under extreme conditions. For example, when the power grid faces extreme weather the goal shifts to minimizing load shed, with little concern for cost. We present a bi-objective modeling approach that addresses these issues and illustrate it in the context of capacity planning in the electric grid. By having an objective that explicitly focuses on load shed in extreme temperature scenarios, we achieve better solutions with smaller sample sizes. We also show the importance of spatial correlation in temperature samples.

Ramsey Rossman
Industrial and Systems Engineering
University of Wisconsin-Madison
rossmann2@wisc.edu

Mihai Anitescu
Argonne National Laboratory
Mathematics and Computer Science Division
anitescu@mcs.anl.gov

Julie Bessac
Argonne National Laboratory
jbessac@anl.gov

Michael C. Ferris
University of Wisconsin-Madison
ferris@cs.wisc.edu

Mitchell Krock
Rutgers University
mk1867@stat.rutgers.edu

James Luedtke, Line A. Roald
University of Wisconsin-Madison
jim.luedtke@wisc.edu, lineroald@gmail.com

MS72

A Geometric Analysis of Neural Collapse with Normalized Features

We provide the first global optimization landscape analysis of Neural Collapse–an intriguing empirical phenomenon that arises in the last-layer classifiers and features of neural networks during the terminal phase of training. As recently reported by Papyan et al., this phenomenon implies that (i) the class means and the last-layer classifiers all collapse to the vertices of a Simplex Equiangular Tight Frame (ETF) up to scaling, and (ii) cross-example within-class variability of last-layer activations collapses to zero. We study the problem based on a simplified unconstrained feature model, which isolates the topmost layers from the classifier of the neural network. In this context, we show that the classical cross-entropy loss with weight decay has a benign global landscape, in the sense that the only global minimizers are the Simplex ETFs while all other critical points are strict saddles whose Hessian exhibit negative curvature directions. Our analysis of the simplified model not only explains what kind of features are learned in the last layer, but also shows why they can be efficiently optimized, matching the empirical observations in practical deep network architectures. These findings provide important practical implications.

Qing Qu
University of Michigan
EECS
qingqu1006@gmail.com

MS72

On the Complexity of Approximate Stationarity Concepts in Non-Smooth Optimization

Non-smooth non-convex optimization problems pose many challenges to the definition and computation of stationarity concepts. Although the field of variational analysis has over the years developed various stationarity concepts for Lipschitz functions and provided many beautiful theoretical tools for studying them, the computational complexity of these concepts remains largely open. In this talk, we first show that with a standard first-order oracle, there is an algorithm with dimension-free finite-time complexity for computing Goldstein approximate stationary (GAS) points of Lipschitz functions. Then, we discuss how such an algorithm can be used to compute near-approximate stationary (NAS) points—a tighter stationarity concept than GAS—of certain Clarke irregular Lipschitz functions. The latter result goes beyond the existing complexity result for computing NAS points of weakly convex functions and applies to objective functions that arise in, e.g., ρ-margin SVM and 2-layer ReLU neural networks. The talk is based on joint work with Lai Tian and Kaiwen Zhou.

Anthony Man-Cho So, Lai Tian, Kaiwen Zhou
The Chinese University of Hong Kong, Hong Kong
manchoso@se.cuhk.edu.hk, tianlai@se.cuhk.edu.hk, kwzhou@cse.cuhk.edu.hk

MS72

A Non-Asymptotic Framework for Approximate Message Passing in Spiked Models.

Approximate message passing (AMP) emerges as an effective iterative paradigm for solving high-dimensional statistical problems. However, prior AMP theory — which focused mostly on high-dimensional asymptotics — fell short of predicting the AMP dynamics when the number of iterations surpasses $o(\log n/\log\log n)$ (with n the problem dimension). To address this inadequacy, this paper develops a non-asymptotic framework for understanding AMP in spiked matrix estimation. Built upon a new decomposition of AMP updates and controllable residual terms, we lay out an analysis recipe to characterize the finite-sample behavior of AMP in the presence of an independent initialization, which is further generalized to allow for spectral initialization. As two concrete consequences of the proposed analysis recipe: (i) when solving Z_2 synchronization, we predict the behavior of spectrally initialized AMP for up to $O(n/\text{poly}\log n)$ iterations, showing that the algorithm succeeds without the need of a subsequent refinement stage (as conjectured recently by Celentano et al.); (ii) we characterize the non-asymptotic behavior of AMP in sparse PCA (in the spiked Wigner model) for a broad range of signal-to-noise ratio.

Yuting Wei
University of Pennsylvania
ytwei@wharton.upenn.edu

MS73

Quantile Inverse Optimization

Inverse linear programming (LP) has received increasing attention because of its potential to infer efficient optimization formulations that can closely replicate the behavior of a complex system. However, inversely inferred parameters and corresponding forward solutions from the existing inverse LP methods can be highly sensitive to noise, errors, and uncertainty in the input data, limiting their applicability in data-driven settings. We introduce the notion of inverse and forward stability in inverse LP and propose a novel inverse LP method that determines a set of objective functions that are stable under data imperfection and generate forward solutions close to the relevant subset of the data. We formulate the inverse model as a large-scale mixed-integer program (MIP) and elucidate its connection to biclique problems, which we exploit to develop efficient algorithms that solve much smaller MIPs instead to construct a solution to the original problem. We numerically evaluate the stability of the proposed method and demonstrate its use in the diet recommendation and transshipment applications.

Taewoo Lee
Department of Industrial Engineering
University of Pittsburgh
taewoo.lee@pitt.edu

Zahed Shahmorad
UTHealth School of Public Health
zahed.shahmoradi@gmail.com

MS73

Inverse Optimization and Low-Rank Plus Sparse Matrix and Tensor Decompositions in Cancer Radiotherapy Optimization

In the first part of this talk, we present the applications of inverse optimization in radiotherapy cancer treatment planning. We also generalize the inverse optimization for multi-objective linear programming where we are looking for the least problem modifications to make a given feasible solution a weak efficient solution. This is a natural extension of inverse optimization for single-objective linear programming with regular ”optimality” replaced by the ”Pareto optimality”. The second part of this talk deals with the applications of matrix/tensor decompositions in radiotherapy cancer treatment planning. Radiotherapy planning gives rise to a matrix, referred to as influence matrix, mapping the machine parameters to the radiation dose delivered to a patient. The influence matrix, which is the main ingredient of radiotherapy optimization, is often very dense and large and therefore computationally prohibitive. We show how matrix/tensor decomposition, in particular low-rank plus sparse decomposition, can be employed to significantly improve the computational efficiency at a negligible cost of plan quality. We also propose a novel algorithm for the low-rank plus sparse matrix decomposition which exploits the special structures of the data arising in radiotherapy applications.

Masoud Zarepisheh
Department of Medical Physics
Memorial Sloan Kettering Cancer Center
zarepism@mskcc.org

Mojtaba Tefagh
Department of Mathematical Sciences
Sharif University of Technology, Tehran, Iran
mtefagh@sharif.edu

MS74

Sensitivity Analysis for Value Functions and Its Application to Bilevel Programs

In this talk, we will study sensitivity analysis of value functions and optimality conditions of bilevel programs. First, for the sensitivity analysis, based on a recent work of the speaker on parametric nonlinear programs, we will further study the directional sensitivity analysis of value functions for parametric set-constrained problems, which include many classical problems as special cases and can be nonsmooth and nonconvex. In particular, we will derive sufficient conditions for the directional Lipschitz continuity, formulae of the directional derivative and upper estimates for the directional limiting / Clarke subdifferential of value functions. Finally, based on the recent development on directional constraint qualifications and directional optimality conditions, using the directional differential properties of value functions, we will derive sharp optimality conditions for general bilevel programs.

Kuang Bai
Department of Applied Mathematics
The Hong Kong Polytechnic University
kuang.bai@polyu.edu.hk

Jane Ye
University of Victoria
Dept of Math & Statistics
janeye@uvic.ca

MS74

Global Optimization of Bilevel Programs and (Generalized) Semi-Infinite Programs - An Open-Source Platform

Over the last 15 years, multiple discretization-based algorithms for bilevel and (generalized) semi-infinite programs have been published. Their computational performance is usually evaluated on a set of test problems. However, evaluating their relative performance is difficult because (i) the comparisons are performed using vastly different platforms and implementations, which impedes comparison of performance measures across publications, (ii) there is no unified large set of test problems that is used consistently across the field, and (iii) many algorithms use algorithmic parameters making the comparison difficult because of the influence of these hyperparameters. Our contribution to remedying these issues is threefold: (i) we present an open-source C++ library containing a collection of algorithm solvers for hierarchical problems. The algorithm solvers for each problem class and the used sub-solvers are easily interchangeable. The results for different algorithm solvers become comparable due to a common platform (in terms of programming language and supported functions) and the usage of the same sub-solvers through a common interface. (ii) we compile an benchmark library consisting of unified existing benchmark libraries and application-oriented test problems. (iii) we utilize the benchmark library to compare the sensitivity of different algorithms to the hyperparameters and the performance of the algorithms for best case hyperparameters for the benchmark library.

Daniel Jungen
Process Systems Engineering (AVT.SVT)
RWTH Aachen University
daniel.jungen@avt.rwth-aachen.de

Aron Zingler, Hatim Djelassi
RWTH Aachen University
aron.zingler@avt.rwth-aachen.de,
hatim.djelassi@rwth-aachen.de

Alexander Mitsos
RWTH Aachen University, Process Systems Engineering
Forschungszenturm Jülich, IEK-10
amitsos@alum.mit.edu

MS74

On a Computationally Ill-Behaved Bilevel Problem with a Continuous and Nonconvex Lower Level

It is well known that bilevel optimization problems are hard to solve both in theory and practice. In this short note, we highlight a further computational difficulty when it comes to solving bilevel problems with continuous but nonconvex lower levels. Even if the lower-level problem is solved to ε-feasibility regarding its nonlinear constraints for an arbitrarily small but positive ε, the obtained bilevel solution as well as its objective value may be arbitrarily far away from the actual bilevel solution and its actual objective value. This result even holds for bilevel problems for which the nonconvex lower level is uniquely solvable, for which the strict complementarity condition holds, and for which the convex constraint set satisfies Slater's constraint qualification for all feasible upper-level decisions. Since the consideration of ε-feasibility cannot be avoided when solving nonconvex problems to global optimality, our result shows that computational bilevel optimization with continuous and nonconvex lower levels needs to be done with great care. Finally, we show that the nonlinearities in the lower level are the key reason for the observed bad behavior by proving that this behavior cannot appear for linear bilevel problems.

Yasmine Beck
Department of Mathematics
Trier University
yasmine.beck@uni-trier.de

Daniel Bienstock
Columbia University IEOR and APAM Departments
IEOR Department
dano@columbia.edu

Martin Schmidt
Universität Trier
martin.schmidt@uni-trier.de

Johannes Thürauf
Department of Mathematics
Trier University
johannes.thuerauf@uni-trier.de

MS75

Projected Solutions of Generalized Quasivariational Problems in Banach Spaces

In this talk, we focus on the analysis of generalized quasi-variational inequalities with non-self map by considering the concept of projected solution introduced by Aussel et al. We present some new existence results in Banach spaces: no monotonicity assumptions are required on the principal operator which is assumed to be norm-to-weak* upper semicontinuous with nonempty weak*-compact convex values. We show the applicability of our techniques by proposing the study of a quasiconvex quasioptimization problem by using the adjusted normal cone operator. [D. Aussel, A. Sultana, V. Vetrivel, On the existence of projected solutions of quasi-variational inequalities and generalized Nash equilibrium problems, J. Optim. Theory Appl. 170 (2016) 818-837.]

Marco Castellani, Massimiliano Giuli
University of LAquila
marco.castellani@univaq.it, massimiliano.giuli@univaq.it

Monica Milasi
University of Messina
mmilasi@unime.it

Domenico Scopelliti
University of Brescia
domenico.scopelliti@unibs.it

MS75

Bilevel Problems under the Bayesian Approach: Existence of Solutions and Some Computational Aspects

In 1996, Mallozzi and Morgan proposed a new model for Stackelberg games which we refer to as the Bayesian approach. In the model, the leader has only partial information about how followers select their reaction among possibly multiple optimal ones. This partial information is modeled as a decision-dependent distribution, the so-called belief of the leader. In this work, we formalize the setting of this approach for bilevel games admitting multiple leaders and we provide new results on existence of solutions. We pay particular attention to the fundamental case of linear bilevel problems and to two classes of beliefs, namely absolutely continuous beliefs, and vertex-supported beliefs. Finally, we provide some numerical experiments.

Anton Svensson
Universidad de O'Higgins
anton.svensson@uoh.cl

MS76

ROC++: Robust Optimization in C++

We propose ROC++, an open-source C++-based platform for automatic robust optimization, applicable to a wide array of single- and multi-stage robust problems with both exogenous and endogenous uncertain parameters, that is easy to both use and extend. It also applies to certain classes of stochastic programs involving continuously distributed uncertain parameters and endogenous uncertainty. Our platform naturally extends existing off-the-shelf deterministic optimization platforms and offers ROPy, a Python interface in the form of a callable library, and the ROB file format for storing and sharing robust problems. We showcase the modeling power of ROC++ on several decision-making problems of practical interest. Our platform can help streamline the modeling and solution of stochastic and robust optimization problems for both researchers and practitioners. It comes with detailed documentation to facilitate its use and expansion. The latest version of ROC++ can be downloaded from https://sites.google.com/usc.edu/robust-opt-cpp/.

Qing Jin, Phebe Vayanos
University of Southern California
qingjin@usc.edu, phebe.vayanos@usc.edu

George Elissaios
N/A
n/a

MS76

Embedding Implicit Function Solvers in Nonlinear Optimization Algorithms

Implicit functions have long been used to formulate nonlinear local optimization problems. Applications have typically been either to scalar-valued functions such as Helmholtz equations of state (e.g. [Eslick et al, Applied Energy, 2022]) or the entire system of equality constraints, as in sequential PDE-constrained optimization [Heinkenschloss, Rice University Tech. Report, 2008]. However, recent work has demonstrated convergence reliability benefits with a modest computational penalty of implicit function formulations between these two extremes [Parker et al, Comp. Chem. Eng., 2022]. Here, we present the software that has made these implicit function formulations possible, some details of our implementation, and benchmarks on a small suite of test problems. Our implicit function formulation leverages PyNumero, a Pyomo extension for the implementation of nonlinear local optimization algorithms, to define parametric square systems of equations for fast re-solves during the optimization algorithm. An important implementation detail is the exploitation of sparsity in derivative evaluations. We implement both sparse and dense implicit function derivatives and compare runtimes between the implementations. Finally, we benchmark our implementation on a small suite of test problems from chemical engineering, comparing runtimes and convergence reliability between a full-space formulation and several configurations of implicit function formulations.

Robert B. Parker
Los Alamos National Laboratory
rbparker@lanl.gov

Bethany Nicholson
Sandia National Laboratories
blnicho@sandia.gov

John D. Siirola
Sandia National Laboratories
Exploratory Simulation Technologies
jdsiiro@sandia.gov

Carl Laird, Larry Biegler
Carnegie Mellon University
claird@andrew.cmu.edu, biegler@cmu.edu

MS76

OMLT: Optimization & Machine Learning Toolkit

This talk presents the optimization and machine learning toolkit (github.com/cog-imperial/OMLT, OMLT 1.0), an open-source software package enabling optimization over high-level representations of neural networks (NNs) and gradient-boosted trees (GBTs) [Ceccon et al., 2022]. Optimizing over trained surrogate models allows integration of NNs or GBTs into larger decision-making problems. Computer science applications include maximizing a neural acquisition function [Volpp et al., 2020] and verifying neural networks [Botoeva et al., 2020]. Engineering applications can utilize black-box optimization or hybridize mechanistic, model-based optimization with data-driven surrogates [Bhosekar and Ierapetritou, 2018]. OMLT 1.0 supports GBTs through an ONNX (github.com/onnx/onnx) interface and NNs through both ONNX and Keras [Chollet et al., 2015.] interfaces. OMLT transforms pre-trained machine learning models into Pyomo [Bynum et al., 2021] to encode optimization formulations. The literature often presents different optimization formulations as competitors, e.g. our partition-based formulation competes with big-M for ReLU [Tsay et al., 2021]. In OMLT, competing optimization formulations become alternative choices for users. We demonstrate applications of OMLT by applying it to case studies including neural network verification, autothermal reformer optimization, and Bayesian optimization.

Francesco Ceccon
Imperial College London
francesco@ceccon.me

Jordan Jalving
Pasteur Labs & ISI
jhjalving@gmail.com

Joshua Haddad
Sandia National Labs
joshuahaddad@ufl.edu

Alexander Thebelt
Imperial College London
alexander.thebelt18@imperial.ac.uk

Calvin Tsay
Imperial College, London
c.tsay@imperial.ac.uk

Carl Laird
Sandia National Laboratories
cdlaird@sandia.gov

Ruth Misener
Imperial College London
r.misener@imperial.ac.uk

MS77

Learning for Optimization Under Uncertainty

We propose a data-driven method to automatically learn the uncertainty sets in robust optimization. Our technique relies on differentiating the solution of robust optimization problems with respect to the parameters of the uncertainty set. By applying gradient-based methods, we resize and reshape the uncertainty sets to better conform to the underlying distribution. Our approach is very flexible it can learn a wide variety of uncertainty sets while preserving tractability and out of sample guarantees. Numerical experiments in portfolio optimization, optimal control, and inventory management show that our method outperforms traditional approaches in robust optimization in terms of out of sample performance and execution time.

Irina Y. Wang, Cole Becker
Princeton University
iywang@princeton.edu, cb35@princeton.edu

Bart Van Parys
Massachusetts Institute of Technology
vanparys@mit.edu

Bartolomeo Stellato
Princeton University
bstellato@princeton.edu

MS77

On the Interface of Robust Optimization and Algorithmic Fairness

Algorithms are now routinely used to make consequential decisions that affect human lives. While algorithms empower us to harness all information hidden in vast data, they may inadvertently amplify existing biases in the available datasets. This concern has sparked increasing interest in fair machine learning, which aims to quantify and mitigate algorithmic discrimination. We propose a distributionally robust logistic regression model with an unfairness penalty that prevents discrimination with respect to sensitive attributes. This model is equivalent to a tractable convex optimization problem if a Wasserstein ball centered at the empirical distribution on the training data is used to model distributional uncertainty. We also derive linear programming-based confidence bounds on the level of the unfairness of any pre-trained classifier by leveraging techniques from optimal uncertainty quantification over Wasserstein balls. Finally, we use ideas from the theory of optimal transport to propose a statistical hypothesis test for detecting unfair classifiers. The test statistic quantifies the distance of the empirical distribution supported on the test samples to the manifold distributions that render a pre-trained classifier fair. We develop a rigorous hypothesis testing mechanism for assessing the equalized loss fairness of any pre-trained logistic classifier.

Bahar Taskesen
EPFL
bahar.taskesen@epfl.ch

Viet Anh Nguyen
Chinese University of Hong Kong
nguyen@se.cuhk.edu.hk

Jose H. Blanchet
Stanford University
jose.blanchet@stanford.edu

Daniel Kuhn
École Polytechnique Fédérale de Lausanne

daniel.kuhn@epfl.ch

MS77

Mean Robust Optimization

Robust optimization is a tractable and expressive technique for decision-making under uncertainty, but it can lead to overly conservative decisions when pessimistic assumptions are made on the uncertain parameters. Wasserstein distributionally robust optimization can reduce conservatism by being data-driven, but it often leads to very large problems with prohibitive solution times. We introduce mean robust optimization, a general framework that combines the best of both worlds by providing a trade-off between computational effort and conservatism. We propose uncertainty sets constructed based on clustered data rather than on observed data points directly thereby significantly reducing problem size. By varying the number of clusters, our method bridges between robust and Wasserstein distributionally robust optimization. We show finite-sample performance guarantees and explicitly control the potential additional pessimism introduced by any clustering procedure. In addition, we prove conditions for which, when the uncertainty enters linearly in the constraints, clustering does not affect the optimal solution. We illustrate the efficiency and performance preservation of our method on several numerical examples, obtaining multiple orders of magnitude speedups in solution time with little-to-no effect on the solution quality.

Irina Y. Wang, Cole Becker
Princeton University
iywang@princeton.edu, cb35@princeton.edu

Bart Van Parys
Massachusetts Institute of Technology
vanparys@mit.edu

Bartolomeo Stellato
Princeton University
bstellato@princeton.edu

MS78

Equivariant and Coordinate-Free Machine Learning

Units equivariance is the exact symmetry that follows from the requirement that relationships among measured quantities of physics relevance must obey self-consistent dimensional scalings. Here, we employ dimensional analysis and ideas from equivariant machine learning to provide a two stage learning procedure for units-equivariant machine learning. For a given learning task, we first construct a dimensionless version of its inputs using classic results from dimensional analysis, and then perform inference in the dimensionless space. Our approach can be used to impose units equivariance across a broad range of machine learning methods which are equivariant to rotations and other groups. We discuss the in-sample and out-of-sample prediction accuracy gains one can obtain in contexts like symbolic regression and emulation, where symmetry is important. We illustrate our approach with simple numerical examples involving dynamical systems in physics and ecology.

Soledad Villar
John Hopkins University
svillar3@jhu.edu

George Kevrekidis
Johns Hopkins University
gkevrek1@jhu.edu

MS78

Undecidability of Weighted Graph Homomorphism Density Inequalities

Many problems and conjectures in extremal combinatorics concern polynomial inequalities between homomorphism densities of graphs, where we allow edges to have real weights. Using the theory of graph limits, we can equivalently evaluate polynomial expressions in homomorphism densities on ¡em¿kernels¡/em¿ W, i.e., symmetric, bounded, and measurable functions W from $[0,1]^2 \to \mathbb{R}$. In 2011, Hatami and Norin proved a fundamental result that it is undecidable to determine the validity of polynomial inequalities in homomorphism densities for graphons (i.e., the case where the range of W is $[0,1]$, which corresponds to unweighted graphs, or equivalently, to graphs with edge weights between 0 and 1). The corresponding problem for more general sets of kernels, e.g., for all kernels or for kernels with range $[-1,1]$, remains open. For any $a > 0$, we show undecidability of polynomial inequalities for any set of kernels which contains all kernels with range $\{0,a\}$. This result also answers a question raised by Lovsz about finding computationally effective certificates for the validity of homomorphism density inequalities in kernels.

Greg Blekherman
Georgia Institute of Technology
gblekherman3@gatech.edu

Annie Raymond
University of Massachusetts Amherst
raymond@math.umass.edu

Fan Wei
Princeton University
fanw@princeton.edu

MS78

Optimal self-concordant barriers for quantum relative entropies

Quantum relative entropies are jointly convex functions of two positive definite matrices (of the same size) that generalize the Kullback-Leibler divergence and arise naturally in quantum information theory. Self-concordant barriers are well-behaved barrier functions on a convex cones that are a key ingredient in efficient interior point methods for solving conic optimisation problems with respect to that cone. In this talk, I'll discuss recent work that establishes the self-concordance of natural barrier functions for the epigraphs of various quantum relative entropies and divergences. Furthermore we show that these barriers have optimal barrier parameter (a measure of complexity of the barrier). The techniques we develop extend to give optimal self-concordant barriers for various closed convex cones related to the noncommutative perspectives of operator concave functions.

Hamza Fawzi
University of Cambridge
h.fawzi@damtp.cam.ac.uk

James Saunderson
Department of Electrical and Computer Systems Engineering
Monash University
james.saunderson@monash.edu

MS79

Multifidelity Deep Neural Operators for Efficient Learning of Partial Differential Equations with Application to Fast Inverse Design of Nanoscale Heat Transport

Deep neural operators can learn operators mapping between infinite-dimensional function spaces via deep neural networks and have become an emerging paradigm of scientific machine learning. However, training neural operators usually requires a large amount of high-fidelity data, which is often difficult to obtain in real engineering problems. Here we address this challenge by using multifidelity learning, i.e., learning from multifidelity data sets. We develop a multifidelity neural operator based on a deep operator network (DeepONet). A multifidelity DeepONet includes two standard DeepONets coupled by residual learning and input augmentation. Multifidelity DeepONet significantly reduces the required amount of high-fidelity data and achieves one order of magnitude smaller error when using the same amount of high-fidelity data. We apply a multifidelity DeepONet to learn the phonon Boltzmann transport equation (BTE), a framework to compute nanoscale heat transport. By combining a trained multifidelity DeepONet with genetic algorithm or topology optimization, we demonstrate a fast solver for the inverse design of BTE problems.

Lu Lu
University of Pennsylvania
lulu1@seas.upenn.edu

MS79

The Evolutionary Deep Operator Neural Network with Sav(edonns) and Its Application on Solving Gradient-Flow Problem

Energy-Dissipative Evolutionary Deep Operator Neural Network is an operator learning neural network. It is designed to seek numerical solutions for a class of partial differential equations instead of a single partial differential equation, such as partial differential equations with different parameters or different initial conditions. In order to preserve essential physical properties of PDEs, such as the Energy Dissipation Law, we adopt a scalar auxiliary variable approach to generate the minimization problem. It introduces modified energy and enables unconditional energy dissipation law at the discrete level. By taking the parameter as a function of the time t variable, this network can predict the accurate solution at any other time by feeding data only at the initial state. The data needed can be generated by the initial conditions, which are readily available. In order to validate the accuracy and efficiency of our neural networks, we provide numerical simulations of several partial differential equations, including heat equations, parametric heat equations, and Allen-Cahn equations.

Jiahao Zhang, Shiheng Zhang, Guang Lin
Purdue University
jiahzhang2@outlook.com, zhan3722@purdue.edu, guanglin@purdue.edu

MS80

Performance estimation of first-order methods applied to convex optimization problems involving linear mappings

In this work, we consider convex optimization problems involving linear mappings, such as those whose objective function includes compositions of the type $f(Ax)$, or which feature linear constraints such as $Ax = b$. First-order methods designed to tackle these problems will typically exploit their specific structure and will need to compute at each iteration products of iterates by matrices A and A^T. Our goal is to identify the worst-case behavior of such first-order methods, based on the Performance Estimation Problem (PEP) methodology. We derive interpolation conditions that allow the use of linear operators A and A^T within the first-order algorithms to be studied using PEP. We cover both the symmetric and nonsymmetric cases and allow bounds on the spectrum of these operators (lower and upper bounds on the eigenvalues in the symmetric case, maximum singular value in the nonsymmetric case). As a byproduct we also obtain interpolation conditions for the class of convex quadratic functions. Using these techniques, we compute several tight worst-case convergence rates, including that of the gradient method applied to the minimization of $f(Ax)$ and that of the Chambolle-Pock algorithm.

Nizar Bousselmi
UCLouvain, Belgium
nizar.bousselmi@uclouvain.be

Julien Hendrickx
Université catholique de Louvain
julien.hendrickx@uclouvain.be

Francois Glineur
UCLouvain - CORE - INMA
francois.glineur@uclouvain.be

MS80

Continuous-Time Analysis of AGM Via Conservation Laws in Dilated Coordinate Systems

We analyze continuous-time models of accelerated gradient methods through deriving conservation laws in dilated coordinate systems. Namely, instead of analyzing the dynamics of $X(t)$, we analyze the dynamics of

$W(t) = t^{\alpha}(X(t) - X_c)$ for some α and X_c and derive a conserved quantity, analogous to physical energy, in this dilated coordinate system. Through this methodology, we recover many known continuous-time analyses in a streamlined manner and obtain novel continuous-time analyses for OGM-G, an acceleration mechanism for efficiently reducing gradient magnitude that is distinct from that of Nesterov. Finally, we show that a semi-second-order symplectic Euler discretization in the dilated coordinate system leads to an $\mathcal{O}(1/k^2)$ rate on the standard setup of smooth convex minimization, without any further assumptions such as infinite differentiability.

Jaewook J. Suh
Seoul National University, South Korea
jacksuhkr@snu.ac.kr

Gyumin Roh
Seoul National University
South Korea
rkm0959@snu.ac.kr

Ernest K. Ryu
Seoul National University, South Korea
ernestryu@snu.ac.kr

MS80

Automated tight Lyapunov analysis for first-order methods

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

Manu Upadhyaya, Sebastian Banert
Lund University
manu.upadhyaya@control.lth.se,
sebastian.banert@control.lth.se

Adrien Taylor
Inria (France)
Adrien.taylor@inria.fr

Pontus Giselsson
Lund University, Sweden
pontusg@control.lth.se

MS81

Differentiable Optimization over Polytopes

We discuss a new technique for differentiable optimization with polytope constraints.

Daniel McKenzie
Colorado School of Mines
dmckenzie@mines.edu

Howard Heaton
Typal Research LLC
research+docs@typal.llc

MS82

Leveraging "Partial" Smoothness for Faster Convergence in Nonsmooth Optimization

First-order methods in nonsmooth optimization are often described as "slow." I will present a (locally) accelerated first-order method that violates this perception. The method converges linearly on "generic" nonsmooth optimization problems. The key insight is that nonsmooth functions are often "partially" smooth in useful ways.

Damek Davis
Cornell University
ORIE
dsd95@cornell.edu

Liwei Jiang
Cornell University
lj282@cornell.edu

MS82

Identifiability, the KL Property in Metric Spaces, and Subgradient Curves

Identifiability, and the closely related idea of partial smoothness, unify classical active set methods and more general notions of solution structure. Diverse optimization algorithms generate iterates in discrete time that are eventually confined to identifiable sets. We present two fresh perspectives on identifiability. The first distills the notion to a simple metric property, applicable not just in Euclidean settings but to optimization over manifolds and beyond; the second reveals analogous continuous-time behavior for subgradient descent curves. The Kurdyka-Lojasiewicz property typically governs convergence in both discrete and continuous time: we explore its interplay with identifiability.

Adrian Lewis
Cornell University
ORIE
adrian.lewis@cornell.edu

Tonghua Tian
Cornell University

tt543@cornell.edu

MS83

Sparse Regression and PCA via Polynomial Roots by Kevin Shu

Sparse regression and sparse PCA are fundamental problems of applied mathematics. We will introduce a heuristic approach to these problems which relates solutions to these problems to the zeros of sparse versions of the determinant polynomial. We will make connections between these ideas, hyperbolic polynomials and determinantal point processes. We will also describe a fast algorithm based on this approach that matches other popular heuristics in time and performance on real instances of these problems.

Kevin Shu
Georgia Institute of Technology
kshu8@gatech.edu

MS83

Reducing Non-Negativity Over General Semialgebraic Sets to Non-Negativity Over Simple Sets

A non-negativity certificate (NNC) is a way to write a polynomial so that its non-negativity on a semialgebraic set becomes evident. Positivstellenstze (Pstze) guarantee the existence of NNCs. Both, NNCs and Pstze underlie powerful algorithmic techniques for optimization. We propose a universal approach to derive new Pstze for general semialgebraic sets from ones developed for simpler sets, such as a box, a simplex, or the non-negative orthant. We provide several results illustrating the approach. First, by considering Handelman's Positivstellensatz (Psatz) over a box, we construct non-SOS Schmdgen-type Pstze over any compact semialgebraic set. That is, a family of Pstze that follow the structure of the fundamental Schmdgen's Psatz, but where instead of SOS polynomials, any class of polynomials containing the non-negative constants can be used, such as SONC, DSOS/SDSOS, hyperbolic or sums of AM/GM polynomials. Second, by considering the simplex as the simple set, we derive a sparse Psatz over general compact sets, which does not require any structural assumptions, such as the running intersection property. Finally, by considering Plya's Psatz over the non-negative orthant, we derive a new non-SOS Psatz over unbounded sets which satisfy some generic conditions.

Olga Kuryatnikova
Erasmus University Rotterdam
kuryatnikova@ese.eur.nl

Juan C. Vera
Tilburg School of Economics and Management
Tilburg University
j.c.veralizcano@tilburguniversity.edu

Luis Zuluaga
Lehigh University
luis.zuluaga@lehigh.edu

MS84

Softmax Policy Gradient Methods Can Take Exponential Time to Converge

The softmax policy gradient (PG) method, which performs gradient ascent under softmax policy parameterization, is arguably one of the de facto implementations of policy optimization in modern reinforcement learning. For ?-discounted infinite-horizon tabular Markov decision processes (MDPs), remarkable progress has recently been achieved towards establishing global convergence of softmax PG methods in finding a near-optimal policy. However, prior results fall short of delineating clear dependencies of convergence rates on salient parameters such as the cardinality of the state space $\mathcal{S}$ and the effective horizon $\frac{1}{1-\gamma}$, both of which could be excessively large. In this paper, we deliver a pessimistic message regarding the iteration complexity of softmax PG methods, despite assuming access to exact gradient computation. Specifically, we demonstrate that the softmax PG method with stepsize η can take

$$\frac{1}{\eta}|\mathcal{S}|^{2^{\Omega(\frac{1}{1-\gamma})}} \text{ iterations}$$

to converge, even in the presence of a benign policy initialization and an initial state distribution amenable to exploration (so that the distribution mismatch coefficient is not exceedingly large). This is accomplished by characterizing the algorithmic dynamics over a carefully-constructed MDP containing only three actions. Our lower bound hints at the necessity of carefully adjusting update rules or enforcing proper regularization in accelerating PG methods.

Gen Li, Yuting Wei
University of Pennsylvania
ligen@wharton.upenn.edu, ytwei@wharton.upenn.edu

Yuejie Chi
Carnegie Mellon University
yuejiechi@cmu.edu

Yuxin Chen
University of Pennsylvania
yuxinc@wharton.upenn.edu

MS84

Global Convergence Conditions for Policy Gradient Methods with Low-dimensional Features

We use examples and analyses to demonstrate the following novel results. First, we show that there exist problems in one-state Markov decision processes (MDPs), where standard policy gradient (PG) methods with low-dimensional features converge to globally optimal policy while reward is not linearly realizable. Second, we show that approximation error is not sufficient to tell whether global convergence is achievable or not for standard PG. Third, we show that global convergence conditions are algorithm dependent concepts. Fourth, we provide analyses and evidences for global convergence conditions using PG and natural PG methods.

Jincheng Mei
Google
jcmei@google.com

MS84

$O(T^{-1})$ Convergence of Optimistic-Follow-the-Regularized-Leader in Two-Player Zero-Sum Markov Games

We prove that optimistic-follow-the-regularized-leader (OFTRL), together with smooth value updates, finds an $O(T^{-1})$-approximate Nash equilibrium in T iterations for two-player zero-sum Markov games with full information. This improves the $\tilde{O}(T^{-5/6})$ convergence rate recently shown in the paper Zhang et al (2022). The refined analysis hinges on two essential ingredients. First, the sum of the regrets of the two players, though not necessarily non-negative as in normal-form games, is approximately non-negative in Markov games. This property allows us to bound the second-order path lengths of the learning dynamics. Second, we prove a tighter algebraic inequality regarding the weights deployed by OFTRL that shaves an extra $\log T$ factor. This crucial improvement enables the inductive analysis that leads to the final $O(T^{-1})$ rate.

Yuepeng Yang, Cong Ma
University of Chicago
yuepengyang@uchicago.edu, congm@uchicago.edu

MS85

Stall Economy: The Value of Mobility in Retail on Wheels

Urban open space emerges as a new territory to embrace retail innovations. Selling products in public spaces with wheeled stalls can potentially become ubiquitous in our future cities. Transition into such a "stall economy" paradigm is being spurred by the rapidly advancing self-driving technologies. Motivated by this transformation, this paper provides models, theory, and insights of spatial queueing systems in which one server moves around to meet mobile customers/machines and in which the "last 100 meters" is expensive. Specifically, we study two service modes i) on-demand, first-come-first-served, and ii) spatially and temporally pooling customer demands. In particular, for the on-demand mode, we propose and solve a "Rendezvous Problem" to analytically characterize the spatial distribution of the stall-customer meeting locations. We also propose a stylized joint truck-stall routing model to capture the inventory replenishment operations. Our main finding is that the stall economy potentially profits more than stationary retail, not only because of the mobility of stalls for providing proximity to customers, but also because of its operational flexibilities that allow for avoiding the "last 100 meters" and pooling demands. In a broader sense, this work looks toward an expanded scope of future retail empowered by self-driving technologies.

Junyu Cao
UT Austin
junyu.cao@mccombs.utexas.edu

Wei Qi
Tsinghua University
qiw@tsinghua.edu.cn

MS85

Provably Good Region Partitioning for On-Time Last-Mile Delivery

On-time last-mile delivery is expanding rapidly as people expect faster delivery of goods ranging from grocery to medicines. Managing on-time delivery systems is challenging because of the underlying uncertainties and combinatorial nature of the routing decision. In practice, the efficiency of such systems also hinges on the driver's familiarity with the local neighborhood. This paper studies the optimal region partitioning policy to minimize the expected delivery time of customer orders in a stochastic and dynamic setting. We allow both the order locations and on-site service times to be random and generally distributed. This policy assigns every driver to a subregion, hence making sure drivers will only be dispatched to their own territories. We characterize the structure of the optimal partitioning policy and show its expected on-time performance converges to that of the flexible dispatching policy in heavy traffic. The optimal characterization features two insightful conditions that are critical to the on-time performance of last-mile delivery systems. We then develop partitioning algorithms with performance guarantees, leveraging ham sandwich cuts and 3-partitions from discrete geometry. This algorithmic development can be of independent interest for other logistics problems. We demonstrate the efficiency of the proposed region partitioning policy via numerical experiments using synthetic and real-world data sets.

Sheng Liu
University of Toronto
shg.liu@rotman.utoronto.ca

MS86

An Inexact Trust-Region Algorithm for Nonsmooth Nonconvex Optimization

In this talk, we develop a new trust-region method to minimize the sum of a smooth nonconvex function and a nonsmooth convex function. Our method is unique in that it permits and systematically controls the use of inexact objective function and derivative evaluations. We prove global convergence of our method in Hilbert space and analyze the worst-case complexity to reach a prescribed tolerance. Our method employs the proximal mapping of the nonsmooth objective function and is simple to implement. Moreover, when using a quadratic Taylor model, our algorithm is a matrix-free proximal Newton-type method that permits indefinite Hessians. We demonstrate the efficiency of our algorithm on various examples from PDE-constrained optimization.

Drew P. Kouri
Optimization and Uncertainty Quantification
Sandia National Laboratories

dpkouri@sandia.gov

MS86

Gnflow: Fast Reinforcement Learning with Automatic Hyperparameter Tuning

Reinforcement learning (RL) seeks to develop an optimal policy through which an agent can interact with their environment to maximize their cumulative reward. In high dimensions, the policy is given by a function approximator, which is learned by solving a complicated stochastic optimization problem. In this talk, we propose a new algorithm, Gauss-Newton Flow (GNflow), to make RL policy learning easier. GNflow incorporates second-order information to accelerate training and leverages recently-developed iterative sampling methods to automatically select hyperparameters and train more robustly. We will support our method empirically using some classic RL control examples.

Elizabeth Newman
Emory University
elizabeth.newman@emory.edu

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

Bas Peters
University of British Columbia
bpeters@eos.ubc.ca

MS86

Implicit Learning of Nash Equilibria for Contextual Games

We study the problem of predicting the outcome of a contextual game, given only the context, and assuming that the player's cost functions are unknown. We use the implicit networks framework to phrase this as a learning problem using historical data consisting of pairs of context and game outcomes. Using several "tricks" (e.g. Davis-Yin operator splitting, constraint decoupling) we improve the efficiency of this scheme to the extent that it can be readily applied to large games with complicated constraint sets. Numerical experiments show the efficacy of this approach on a collection of real-world traffic routing problems.

Samy Wu Fung
Colorado School of Mines
swufung@mines.edu

MS87

Branch-and-Price for Multistage Stochastic Mixed-Integer Nonlinear Programs with Discrete State Variables

Stochastic programming provides a natural framework for modeling sequential optimization problems under uncertainty; however, the efficient solution of large-scale multistage stochastic programs remains a challenge, especially in the presence of discrete decisions and nonlinearities as well as long planning horizons and large numbers of scenarios. In this work, we consider multistage stochastic mixed-integer nonlinear programs (MINLPs) with discrete state variables, which exhibit a decomposable structure that allows its solution using a branch-and-price approach. Following a Dantzig-Wolfe reformulation, we apply column generation such that each pricing subproblem is an MINLP of much smaller size, making it more amenable to exact MINLP solvers. In addition, the mutual independence of the resulting subproblems allows for parallel computation, which significantly reduces the solution time. Several case studies of practical relevance are used to demonstrate the effectiveness of the proposed decomposition algorithm over solving the full-space model.

Qi Zhang, Tushar Rathi
University of Minnesota
qizh@umn.edu, rathi023@umn.edu

Angela Flores-Quiroz
University of Chile
angflore@ing.uchile.cl

MS87

Markov Chain-Based Policies for Multi-Stage Stochastic Integer Linear Programming

We introduce a novel aggregation framework to address multi-stage stochastic programs with mixed-integer state variables and continuous local variables (MSILPs). Our aggregation framework imposes additional structure to the integer state variables by leveraging the information of the underlying stochastic process, which is modeled as a Markov chain (MC). We present a novel branch-and-cut algorithm integrated with stochastic dual dynamic programming as an exact solution method to the aggregated MSILP, which can also be used in an approximation form to obtain dual bounds and implementable feasible solutions. Moreover, we apply two-stage linear decision rule (2SLDR) approximations and propose MC-based variants to obtain high-quality decision policies with significantly reduced computational effort. We test the proposed methodologies in a novel MSILP model for hurricane disaster relief logistics planning. Our empirical evaluation compares the effectiveness of the various proposed approaches and analyzes the trade-offs between policy flexibility, solution quality, and computational effort. Specifically, the 2SLDR approximation yields provable near-optimal solutions for our test instances supported by the proposed bounding procedure.

Margarita Castro
Departamento Ingeniería Industrial y de Sistemas
Pontificia Universidad Católica de Chile
margarita.castro@ing.puc.cl

Merve Bodur
University of Toronto
Department of Mechanical and Industrial Engineering
bodur@mie.utoronto.ca

Yongjia Song
Clemson University

yongjis@clemson.edu

MS87

On the Value of Multistage Risk-Averse Stochastic Facility Location with or without Prioritization

We consider a multiperiod stochastic capacitated facility location problem under uncertain demand and budget. We formulate a multistage stochastic integer program to dynamically locate facilities in each period and compare it with a two-stage approach that determines the facility locations upfront. When the budget is also uncertain, a popular modeling framework is to prioritize the candidate sites. In the two-stage model, the priority list is decided in advance and fixed through all periods, while in the multistage model, the priority list can change adaptively. Using expected conditional risk measures, we derive tight lower bounds for the gaps between the optimal objective values of risk-averse multistage and two-stage models in both settings with and without prioritization. Moreover, we propose approximation algorithms to efficiently solve risk-averse problems without prioritization, which are asymptotically optimal under an expanding market assumption. We also design a set of super-valid inequalities for risk-averse problems with prioritization to reduce computational time. We conduct numerical studies to demonstrate the tightness of the analytical bounds and efficacy of the approximation algorithms and prioritization cuts. We find that the gaps between risk-averse multistage and two-stage models increase as the variations of the uncertain parameters increase, and stagewise-dependent scenario trees attain much higher gaps than the stagewise-independent ones.

Xian Yu
Department of Integrated Systems Engineering
The Ohio State University
yu.3610@osu.edu

Siqian Shen
Department of Industrial and Operations Engineering
University of Michigan
siqian@umich.edu

MS88

Constraint programming models for optimal qubit assignment and swap-based routing

Due to the connectivity limitations of gate model quantum devices, logical quantum circuits must first be compiled to target hardware before they can be executed. Typically, this compilation process involves the efficient insertion of swap gates into the logical circuit, often increasing the depth of the circuit, achieved by solving a so-called qubit assignment and routing problem. Recently, a number of integer linear programming (ILP) models have been proposed for solving the qubit assignment and routing problem. These models encode the objective function and constraints of the problem, and leverage the use of automated solver technology to find hardware-compliant quantum circuits. In this work, we propose constraint programming (CP) models for the same problem, and conduct an empirical analysis of our approaches, comparing their performance against ILP for both linear and two-dimensional lattice quantum devices. Our empirical analysis indicate the promise of CP technology for this important application, demonstrating strong performance with respect to both solution quality and runtime.

Kyle Booth
Amazon Quantum Solutions Lab
kybooth@amazon.com

MS89

Blessing of Nonconvexity in Factorized Models

Factorized models, from low-rank matrix recovery to deep neural networks, play a central role in many modern machine learning problems. Despite their widespread applications, problems based on factorized models are deemed difficult to solve in their worst case due to their inherent nonconvexity. Our talk is inspired by the recent observations in the optimization and machine learning communities that many realistic and practical instances of factorized models are far from their worst case scenarios. We study a natural nonconvex and nonsmooth formulation of two prototypical factorized models, namely low-rank matrix factorization and deep linear regression. On the negative side, we show that these problems do not have a benign landscape: with high probability, there always exists a true solution that is not a global minimum of the loss function. However, on the positive side, we show that a simple subgradient method with small initialization is oblivious to such problematic solutions; instead, it converges to a balanced solution that is not only close to the ground truth but also enjoys a flat local landscape.

Salar Fattahi, Jianhao Ma
University of Michigan
fattahi@umich.edu, jianhao@umich.edu

MS89

A Randomized Algorithm for Nonconvex Minimization with Inexact Evaluations and Complexity Guarantees

We consider minimization of a smooth nonconvex function with inexact oracle access to gradient and Hessian (but not function value) to achieve $(\varepsilon_g, \varepsilon_H)$-approximate second-order optimality. A novel feature of our method is that if an approximate direction of negative curvature is chosen as the step, we choose the direction of the move (positive or negative) with equal probability. We also use relative inexactness measures on gradient and Hessian and relax the coupling between the first- and second-order tolerances ε_g and ε_H. Our convergence analysis includes both an expectation bound based on martingale analysis and a high-probability bound based on concentration inequalities. We apply our algorithm to empirical risk minimization problems and obtain gradient sample complexity that improves over previous works.

Shuyao Li
University of Wisconsin-Madison
shuyao.li@wisc.edu

Stephen Wright
University of Wisconsin-Madison

Dept. of Computer Sciences
swright@cs.wisc.edu

MS89

Nonconvexity of Sparsity Promoting Functions

The essence of sparse modeling is to adopt proper bases or transforms for multiscale representation of data and then to impose sparsity regularization of data under a basis or a transform, by designing sparsity promoting functions (SPFs). Sparsity promoting functions serve as a key ingredient for the sparse modeling. In this talk, we first briefly review some existing SPFs. We particularly focus on SPFs that are the ratio of the L1 and L2 norms and its variants. We then present algorithms for the applications of these SPFs in compressive sensing.

Lixin Shen
Mathematics Department, Syracuse University
lshen03@syr.edu

MS90

On Nonsmooth Mean Field Games with Control and State Constraints

The talk is devoted to noncooperative games with constraints on the controls and the states, and their limiting behaviour for infinitely many players. At first, we are interested with analysing an N-player differential game problem, in a control-based framework, in which all players interact through a mean field term. This model consists of minimising a nonsmooth cost functional for each individual player subject to a time-independent differential equation. We assume couplings to appear in the objectives and the controlled dynamics. Also, we provide an existence result of Nash equilibria and first-order optimality conditions. Next, we address the state constraints and the nonsmoothness of the objectives by using Moreau–Yosida and Huber–type regularisations, respectively. Then, we investigate the limit to infinitely many players and show existence of solution (mean field equilibrium) via a fixed-point approach. Afterwards, we make the connection between equilibrium of the resulting constrained mean field game and equilibria in the finite N-player game, namely approximate Nash equilibria.

Sarah Essadi, Michael Hintermueller
Weierstrass Institute for Applied Analysis and Stochastics
essadi@wias-berlin.de, hintermueller@wias-berlin.de

MS90

Randomized Methods for Nonsmooth and Nonconvex Federated Optimization

Federated learning (FL) is a decentralized optimization framework that allows for training from a multitude of private and possibly, heterogenous datasets. In this talk, we consider the development of zeroth-order FL methods for nonsmooth and hierarchical stochastic optimization problems. To this end, first we devise new randomized zeroth-order local stochastic gradient descent (SGD) methods where we leverage a randomized smoothing technique to approximate the subgradients of the nonsmooth local objectives. We provide convergence theory for the strongly convex, convex, and nonconvex cases. In the second part, motivated by the need for addressing hyperparameter optimization in FL, we consider federated bilevel stochastic optimization problems. We devise a class of randomized smoothing-enabled zeroth-order local-SGD methods and derive new performance guarantees. Preliminary numerical results will be presented.

Yuyang Qiu
Rutgers University
yq117@scarletmail.rutgers.edu

Uday V. Shanbhag
Pennsylvania State University
udaybag@psu.edu

Farzad Yousefian
Rutgers University
farzad.yousefian@rutgers.edu

MS90

Structural Estimation of Markov Decision Processes in High-Dimensional State Space with Finite-Time Guarantees

We consider the task of estimating a structural model of dynamic decisions by a human agent based upon the observable history of implemented actions and visited states. This problem has an inherent nested structure: in the inner problem, an optimal policy for a given reward function is identified while in the outer problem, a measure of fit is maximized. Several approaches have been proposed to alleviate the computational burden of this nested-loop structure, but these methods still suffer from high complexity when the state space is either discrete with large cardinality or continuous in high dimensions. Other approaches in the inverse reinforcement learning (IRL) literature emphasize policy estimation at the expense of reduced reward estimation accuracy. In this paper we propose a single-loop estimation algorithm with finite time guarantees that is equipped to deal with high-dimensional state spaces without compromising reward estimation accuracy. In the proposed algorithm, each policy improvement step is followed by a stochastic gradient step for likelihood maximization. We show that the proposed algorithm converges to a stationary solution with a finite-time guarantee. Further, if the reward is parameterized linearly, we show that the algorithm approximates the maximum likelihood estimator sublinearly. Finally, by using MuJoCo and their transfer settings, we show the proposed algorithm achieves superior performance compared with IRL and imitation learning benchmarks.

Siliang Zeng
University of Minnesota
zeng0176@umn.edu

MS91

Modernization of a Nonlinear Least-Squares Algorithm Operated for Short-Term Electricity De-

mand Forecast in Qubec

An accurate forecast of electricity consumption is at the hearth of the operations of Hydro-Qubec, the main supplier of electricity in Qubec. Short-term demand is estimated with a parametric model whose parameters are calibrated by nonlinear least squares, using a Fortran 77 library. While this model has been successfully applied for more than thirty years, its maintenance is increasingly more complicated, posing the need to modernize it. We present a reimplementation of the least-squares optimization method in Julia, which offers more reliabilty and readability compared to the original library. We discuss the underlying mathematical aspects of the algorithm and compare the results and performance of our implementation to the Fortran 77 algorithm version. In order to do so, we performed several numerical experiments on test problems using models currently operated by Hydro-Qubec. The preliminary result show very good agreement with respect to the intermediate results and the final ouptuts.

Pierre Borie
University of Montreal
pierre.borie@umontreal.ca

Alain Marcotte
Hydro-Québec
marcotte.alain3@hydroquebec.com

Stéphane Dellacherie
Hydro-Québec, Prévision des apports et de la demande
dellacherie.stephane@hydroquebec.com

Fabian Bastin
University of Montreal
bastin@iro.umontreal.ca

MS91

Applications of Interior Point Methods: From Sparse Approximations to Discrete Optimal Transport

A variety of problems in modern applications of optimization require a selection of a 'sparse' solution, a vector with preferably few nonzero entries. Such problems may originate from very different applications in computational statistics, signal or image processing or compressed sensing, finance, machine learning and discrete optimal transport, to mention just a few. Sparse approximation problems are often solved with dedicated and highly specialised first-order methods of optimization. In this talk I will argue that these problems may be very efficiently solved by interior point methods. This is joint work with: Valentina De Simone, Daniela di Serafino, Spyros Pougkakiotis and Marco Viola

Jacek Gondzio
The University of Edinburgh, United Kingdom
J.Gondzio@ed.ac.uk

MS91

Stochastic Inexact Newton Methods for Nonlinear Least-Squares and Systems

We address the solution of large-scale nonlinear least-squares problems and nonlinear systems using linesearch Inexact Newton methods with random models. The procedures proposed belong to the class of stochastic Gauss-Newton type methods for the least-squares problems and to the class of stochastic Newton methods for the nonlinear systems. We discuss both the construction of random models that can greatly reduce the computational effort of the solvers while retaining convergence guarantees, and the computation of trial steps via iterative linear solvers. The approximation of the derivatives is performed by sampling on matrix multiplication or by sampling for sparsification. Numerical validation is presented.

Benedetta Morini
Dipartimento di Ingegneria Industriale
Universita' di Firenze
benedetta.morini@unifi.it

Stefania Bellavia
Universita di Firenze
stefania.bellavia@unifi.it

Greta Malaspina
Università di Firenze, Italia
greta.malaspina@unifi.it

MS92

Packing and Covering in Directed Graphs: Progress Towards Woodall's Conjecture

In any directed graph, the minimum size of a dijoin is equal to the maximum number of disjoint dicuts. This is the celebrated Luccesi-Younger theorem and is even known to hold in a weighted setting. However, by interchanging the roles of dicuts and dijoins, we obtain the much more mysterious Woodall's conjecture, which asks: If the minimum size of a dicut is k, are there always k pairwise disjoint dijoins? This beautiful min-max conjecture has remained open for 50 years with little recent progress. It is known to be false in the weighted setting, and is still open in various special cases, such as when $k = 3$, or when the digraph is planar. In this talk, I will discuss recent progress towards Woodall's conjecture. We show that for the purposes of proving (or disproving) Woodall's conjecture, it suffices to consider only 'bipartite' digraphs satisfying certain degree conditions. We use this to uncover connections to matroid theory and prove, among other results, that in any optimal fractional packing, at least one dijoin can be chosen to have value 1.

Ahmad Abdi
London School of Economics and Political Science
a.abdi1@lse.ac.uk

Gerard Cornuejols, Michael Zlatin
Tepper School of Business
Carnegie Mellon University

gc0v@andrew.cmu.edu, mzlatin@andrew.cmu.edu

MS94

Noisy SGD: Stability, Privacy, and Mixing

This talk investigates the stability of (Noisy) Stochastic Gradient Descent. Our stability results lead to the resolution of problems in several different fields spanning optimization, sampling, and differentially private machine learning. In this talk, we focus on two implications of this growing body of work. First, we resolve the mixing time of the Langevin Algorithm for log-concave sampling. Second, we resolve the differential privacy of Noisy-SGD in the fundamental setting of convex optimization.

Jason Altschuler
NYU
jason.altschuler@gmail.com

MS94

Recent Progress on Graphical Designs

Graphical designs are quadrature rules for graphs inspired by spherical designs. These subsets of graph vertices have beautiful connections to extremal combinatorics, coding theory, polytopes, random walks, graph signal processing, and more. This talk will explain the motivation for and definition of graphical designs and explain some recent results on the topic.

Catherine Babecki
Department of Mathematics
University of Washington
cbabecki@uw.edu

MS94

Mirror Langevin Monte Carlo: the Case Under Isoperimetry

Motivated by the connection between sampling and optimization, we study a mirror descent analogue of Langevin dynamics and analyze three different discretization schemes, giving nonasymptotic convergence rate under functional inequalities such as Log-Sobolev in the corresponding metric. Compared to the Euclidean setting, the result reveals intricate relationship between the underlying geometry and the target distribution and suggests that care might need to be taken in order for the discretized algorithm to achieve vanishing bias with diminishing stepsize for sampling from potentials under weaker smoothness/convexity regularity conditions.

Qijia Jiang
Lawrence Berkeley National Lab
qjiang2@alumni.stanford.edu

MS95

Data-Driven Parameter Optimization for Some Inverse Problems with Sparsity-Based Priors

In recent years, novel ideas have been applied to several inverse problems in combination with machine learning approaches, to improve the inversion by optimally choosing different parameters of interest. A fruitful approach in this sense is bilevel optimization, where the inverse problems are considered as lower-level constraints, while on the upper-level a loss function based on a training set is used. When confronted with inverse problems with sparsity-based regularizers, however, the bilevel optimization problem structure becomes quite involved to be analyzed, as classical nonlinear or bilevel programming results cannot be directly utilized. In this talk, I will discuss on a strategy to overcome these difficulties, leading to a reformulation of the bilevel problems as mathematical programs with complementarity constraints. This enables to obtain sharp first-order optimality conditions, but at the price of lifting the problems to a higher dimension. Some ideas on how to reduce the dimension of the problems back will also be presented, together with the different challenges that these problems pose, when dealing with large training sets.

Juan Carlos De los Reyes
MODEMAT, Escuela Politecnica Nacional
Quito, Ecuador
juan.delosreyes@epn.edu.ec

MS95

Randomized Reconstruction in Medical Imaging

Medical image reconstruction is a critical tool in nuclear medicine but with computational challenges that arise from the size of imaging data, costs due to repeated projection operations between the image and measured data spaces, and other issues. Thus, there has been a significant interest in accelerating these algorithms. In this work we investigate the use of modern stochastic optimisation methodologies on reconstruction tasks in CT and PET on phantom and clinical data. We study the use of differentiable and non-differentiable priors, variance reduction, acceleration, the impact of number of subsets, preconditioners, stepsizes, and other optimisation parameters on the quality of reconstructed solutions and convergence properties.

Zeljko Kereta
University College London
z.kereta@ucl.ac.uk

MS95

An Arnoldi-based preconditioner for iterated Tikhonov regularization

Many problems in science and engineering give rise to linear systems of equations that are commonly referred to as large-scale linear discrete ill-posed problems. These problems arise for instance, from the discretization of Fredholm integral equations of the first kind. The matrices that define these problems are typically severely ill-conditioned and may be rank deficient. Because of this, the solution of linear discrete ill-posed problems may not exist or be extremely sensitive to perturbations caused by error in the available data. These difficulties can be reduced by applying Tikhonov regularization. We describe a novel ”approximate Tikhonov regularization method” based on constructing a low-rank approximation to the matrix in the linear discrete ill-posed problem by carrying out a few steps of the Arnoldi process. The iterative method so defined is transpose-free. Our work is inspired by Donatelli and Hanke whose approximate Tikhonov regularization method seeks to approximate a severely ill-conditioned block-Toeplitz matrix with Toeplitz-blocks by a block-circulant matrix with circulant-blocks. Computed examples illustrate the performance of our proposed iterative regularization method.

Alessandro Buccini
University of Cagliari
alessandro.buccini@unica.it

Lucas Onisk
Emory University
lonisk@emory.edu

Lothar Reichel
Kent State University
reichel@math.kent.edu

MS96

An Exact Method for Nonlinear Network Flow Interdiction Problems

We study network flow interdiction problems with nonlinear and nonconvex flow models. The resulting model is a max-min bilevel optimization problem in which the follower's problem is nonlinear and nonconvex. In this game, the leader attacks a limited number of arcs with the goal to maximize the load shed and the follower aims at minimizing the load shed by solving a transport problem in the interdicted network. We develop an exact algorithm consisting of lower and upper bounding schemes that computes an optimal interdiction under the assumption that the interdicted network remains weakly connected. The main challenge consists of computing valid upper bounds for the maximal load shed, whereas lower bounds can directly be derived from the follower's problem. To compute an upper bound, we propose solving a specific bilevel problem, which is derived from restricting the flexibility of the follower when adjusting the load flow. This bilevel problem still has a nonlinear and nonconvex follower's problem, for which we then prove necessary and sufficient optimality conditions. Consequently, we obtain equivalent single-level reformulations of the specific bilevel model to compute upper bounds. Our numerical results show the applicability of this exact approach using the example of gas networks.

Johannes Thürauf
Department of Mathematics
Trier University
johannes.thuerauf@uni-trier.de

Martin Schmidt
Universität Trier
martin.schmidt@uni-trier.de

MS96

Preconditioned Proximal-Type Methods for Bilevel Optimisation

Bilevel problems have been traditionally solved through either treating the inner problem as a constraint, and solving the resulting Karush-Kuhn-Tucker conditions using a Newton-type solver; or by trivialising the inner problem to its solution mapping. The latter approach in principle requires near-exact solution of the inner problem for each outer iterate. Moreover, an adjoint equation typically needs to be solved to calculate the differential of the inner problem solution mapping. Recently, intermediate approaches have surfaced that solve the inner problem and occasionally the adjoint as well to a low precision, and still obtain some form of convergence. In this talk, we discuss the linear convergence of methods based on taking interleaved steps of proximal-type methods on both the inner and outer problem, and computationally cheap steps for the adjoint. We demonstrate numerical performance on imaging applications. This is joint work with Ensio Suonper (University of Helsinki).

Tuomo Valkonen
University of Helsinki & Escuela Politécnica Nacional
tuomov@iki.fi

MS96

Levenberg-Marquardt Method for Bilevel Optimization

We consider the bilevel optimization problem in a very general setting, where the follower (or lower-level player) is free to choose whatever optimal solution that they want if their optimal solution set is not necessarily unique under some choices of the leader (or upper-level player). As it is well-known in the literature, this framework gives rise to the optimistic and pessimistic bilevel optimization problems, as tractable decision-making setups. In this talk, we consider both the optimistic and pessimistic scenarios of the problem and use the lower-level optimal value function reformulation to transform the followers problem into inequality constraints. We then discuss approaches to write necessary optimality conditions of these problems, which are amenable to tractable numerical methods. It turns out that in each case, these optimality conditions can be rewritten as overdetermined systems of equations. We will therefore discuss smoothing and nonsmooth-types Levenberg-Marquardt methods to solve these equations and report on the very promising numerical results obtained.

Alain B. Zemkoho
School of Mathematics
University of Birmingham
a.b.zemkoho@soton.ac.uk

MS97

Bayesian Filtering Schemes to Estimate Time-Varying Parameters in Dynamical Systems

Estimating and quantifying uncertainty in unknown system parameters from limited data remains a challenging inverse problem in many real-world applications. While many methods focus on estimating constant parameters, these problems can also include time-varying parameters with unknown dynamics that often cannot be directly observed. In this talk, we present novel Bayesian filtering schemes to estimate unknown time-varying parameters in nonstationary inverse problems arising from deterministic dynamical systems, and we demonstrate the capabilities of these approaches with several computed examples.

Andrea Arnold
Worcester Polytechnic Institute
anarnold@wpi.edu

MS97

Bayesian sparse optimization for large scale dictionary learning

In recent years dictionary learning methods have been used to solve inverse problems to bypass the evaluation of the forward model in the traditional optimization algorithms. When the dictionaries are large, and a sparse representation of the data in terms of the atoms is desired, computationally efficient sparse optimization algorithms are needed to improve the performance. Some of these methods may replace the original dictionary by a reduced one. Since reduced dictionaries can represent the data only up to a model reduction error, Bayesian methods for estimating modeling errors may significantly improve the representation accuracy without having to resort to the full dictionary. In this talk, we will outline how to use using Bayesian hierarchical models and modeling error methods in the context of sparse dictionary learning. Computed experiments will illustrate the effectiveness of the approach.

Daniela Calvetti
Case Western Reserve Univ
Department of Mathematics, Applied Mathematics and Statistic
dxc57@case.edu

Erkki Somersalo, Alberto Bocchinfuso
Case Western Reserve University
ejs49@case.edu, axb1094@case.edu

MS97

Regularization by Inexact Krylov Methods with Applications to Semi-Blind Deblurring

In this talk I will present a new class of algorithms for separable nonlinear inverse problems based on inexact Krylov methods. In particular, I will focus on semi-blind deblurring applications, where we are interested in recovering an approximation of the original image and of a small number of parameters defining the blur. Classical methods in this setting involve solving a sequence of ill-posed and computationally expensive linear problems. For example, this can be done using Krylov methods, iterative solvers with intrinsic regularization properties. In this talk I will introduce a new class of algorithms based on inexact Krylov methods, where the inexactness stems from the uncertainty in the parameters defining the blur, which are computed throughout the iterations. After giving a brief overview of the theoretical properties of these methods, as well as strategies to monitor the amount of inexactness that can be tolerated, the performance of the algorithms will be shown through numerical examples.

Malena Sabate Landman
Emory University
msabat3@emory.edu

MS98

Recent Advances in SCIP and its Separation Routines

We discuss the recent developments in the SCIP optimization solver and its plugins. We focus on the separation plugin and discuss new implementations in it, such as the lift-and-project cuts for mixed integer problems. We conclude with the computational performance of these advancements.

Suresh Bolusani
Zuse Institute Berlin
bolusani@zib.de

MS98

Whats new in FICO Xpress 9.1 Solver

In this talk, we present the algorithmic enhancements of the latest release of FICO Xpress, with a special focus on the newly introduced global solver for mixed-integer nonlinear problems.

Gregor Hendel
FICO Xpress Optimization
GregorHendel@fico.com

MS99

Probabilistic Robust Optimization with a Real-World Application on Hospital Optimization

Classical approaches in robust optimization (RO) involve out-of-the-box uncertainty sets such as norm, budget or polyhedral sets. Apart from leveraging available data to cross-validate hyperparameters, out-of-the-box uncertainty sets seldom incorporate distributional information, so they inherently do not capture the real-world behavior of uncertain parameters. Distributionally Robust Optimization (DRO), by contrast, satisfies constraints in expectation among a class of underlying distributions. We propose a new type of uncertainty set motivated by information theory, which incorporates the probability distributions directly. Initial computational tests with our approach yield reductions in the maximum and mean violation as compared to classical RO and DRO. We apply this novel uncertainty set in a real-world setting of the Hartford Hospital's Bone and Joint Institute to optimize the patient census while considering the uncertainty in the rest times of patients after surgery. We generate a timetable for surgeons that reduces the monthly census by up to 10%, outperforming both out-of-the-box uncertainty sets and DRO.

Dimitris Bertsimas
Massachusetts Insitute of Technology
dbertsim@mit.edu

Benjamin Boucher
Massachusetts Institute of Technology

bboucher@mit.edu

MS99

An Rpt Approach to Robust Nonlinear Optimization

We propose to use a reformulation perspectification technique based approach to derive computationally tractable safe approximations for a class of robust constraints which are neither convex nor concave in the uncertain parameters. We extend our approach to two-stage adaptive robust optimization problems in which we impose convex and concave decision rules unaddressed by previous work to obtain approximate solutions. For several types of non-linear decision rules we prove that imposing these decision rules to the two-stage adjustable robust optimization formulation leads to a solution that is at least as tight as imposing an affine decision rule.

Danique de Moor, Dick den Hertog
University of Amsterdam
daniquedemoor@hotmail.com, d.denhertog@uva.nl

MS99

Extremal Stochastic Models with IID Driving Sequences

Stochastic models typically assume that the probability distributions of the driving random variables are fully known. However, in practice, these distributions are often only partially known, with limited moment information available. Distributionally robust analysis evaluates worst-case model performance under partial information, typically utilizing duality techniques from conic optimization. For stochastic models with multiple independent and identically distributed (i.i.d.) random variables, applying these optimization methods can prove challenging, if not impossible, due to the nonconvex nature of the underlying optimization problem. To address this challenge, we present in this talk novel findings based on a simple insight that can aid in the analysis of stochastic models with i.i.d. driving sequences as input. Our approach establishes sharp bounds for the higher-order cumulants of the GI/G/1 steady-state waiting time when only mean-variance information is available. Further, our method can be applied more broadly to other stochastic models.

Wouter van Eekelen
Tilburg University
w.j.e.c.vaneekelen@tilburguniversity.edu

MS100

Tight Semidefinite Relaxations for Sign-Indefinite Qcqps with Bipartite Structures

We propose a new method to determine the tightness of the standard semidefinite programming (SDP) relaxation of nonconvex quadratically constrained quadratic programs (QCQPs) whose aggregated sparsity pattern graphs are bipartite. This method is based on sufficient conditions for a solution matrix of the SDP relaxation to be included in the rank-1 cone, and it can detect the tightness by solving feasibility systems made from data and the sparsity patterns. A sufficient condition of the tightness for sign-definite QCQPs, in which the set of elements at the same index in all the data matrices is either all nonnegative or all nonpositive, has been recently proposed. In this talk, to show that our condition covers a wider class of QCQPs than above condition, we also propose a conversion method from sign-definite QCQPs with no specific sparsity to ones with bipartite sparsity whose SDP relaxations are tight. In addition, we provide instances of QCQPs to show our method can detect the tightness of sign-indefinite QCQPs. Finally, several applications of our method including numerical experiments are discussed.

Godai Azuma
Tokyo Institute of Technology
Department of Mathematical and Computing Science
azuma.g.aa@m.titech.ac.jp

Mituhiro Fukuda
University of São Paulo
Department of Computer Science
mituhiro@is.titech.ac.jp

Sunyoung Kim
Ewha W. University
skim@ewha.ac.kr

Makoto Yamashita
Tokyo Institute of Technology
Dept. of Mathematical and Computing Science
makoto.yamashita@c.titech.ac.jp

MS100

Constant Rank Constraint Qualification for Nonlinear Second-Order Cone Programming

In [R. Andreani, G. Haeser, L. M. Mito, H. Ramirez C., Weak notions of nondegeneracy in nonlinear semidefinite programming, arXiv:2012.14810, 2020] the classical notion of nondegeneracy (or transversality) and Robinsons constraint qualification have been revisited in the context of nonlinear semidefinite programming exploiting the structure of the problem, namely, its eigendecomposition. This allows formulating the conditions equivalently in terms of (positive) linear independence of significantly smaller sets of vectors. In this paper we extend these ideas to the context of nonlinear second-order cone programming. For instance, for an m-dimensional second-order cone, instead of stating nondegeneracy at the vertex as the linear independence of m derivative vectors, we do it in terms of several statements of linear independence of 2 derivative vectors. This allows embedding the structure of the second-order cone into the formulation of nondegeneracy and, by extension, Robinsons constraint qualification as well. This point of view to be crucial in defining significantly weaker constraint qualifications such as the constant rank constraint qualification and the constant positive linear dependence condition. Also, these conditions are shown to be sufficient for guaranteeing global convergence of several algorithms, while still implying metric subregularity and without requiring boundedness of the set of Lagrange multipliers.

Gabriel Haeser
University of São Paulo

FAPESP
ghaeser@ime.usp.br

Roberto Andreani
University of Campinas
andreani@unicamp.br

Leonardo M. Mito
University of São Paulo
FAPESP
leonardommito@gmail.com

Héctor Ramírez
Centro de Modelamiento, Matemático-Universidad de Chile
hramirez@dim.uchile.cl

Thiago Silveira
University of São Paulo
thiagops@ime.usp.br

MS100

Convex Conic Reformulations of Geometric Nonconvex Conic Optimization Problems for a Class of Quadratic Optimization Problems

We discuss a generalization of the results in "A geometric analysis of a class of non convex conic programs", SIAM J. on Optim., 30:1251-1273, 2020. In the paper, convex conic reformulation of a geometric nonconvex conic optimization problem (COP) was studied as a unified framework for completely positive programming (CPP) reformulation of a class of nonconvex quadratic optimization problems (QOPs). Their geometric nonconvex COP is represented in a very simple form, yet captures the essentials of the convexification of general nonconvex COPs. In this talk, we extend the framework by generalizing some assumptions and investigating strong duality. We show that a larger class of QOPs can be reformulated as CPP. Specifically, a class of QOPs with complementarity conditions is discussed for the reformulation.

Sunyoung Kim
Ewha W. University
skim@ewha.ac.kr

Masakazu Kojima
Tokyo Institute of Technology
kojima@is.titech.ac.jp

MS101

Derivative-Informed Neural Operator for PDE-Constrained Optimization under Uncertainty

We present a novel framework for solving PDE-constrained optimization problems under uncertainty (OUU) problems using derivative-informed neural operators. OUU problems are often orders of magnitude more expensive to solve compared to their deterministic counterparts because of the evaluation of the statistical/risk measure by stochastic integration. To tackle this computational challenge, we propose a reduced-basis-based encoder-decoder network to approximate the PDE mapping from the input spaces of the uncertain parameter and the optimization variable to the output state, where the derivative information of the PDE is used to construct the input and output reduced bases. In addition, we incorporate the PDE derivatives into a Sobolev-type training loss to ensure that the neural operator has accurate derivatives with respect to the optimization variable, which enables derivative-based optimization algorithms to solve the OUU problem. We demonstrate its performance and computational speed-ups over a suite of numerical examples.

Dingcheng Luo
Oden Institute for Computational Engineering and Sciences,
The University of Texas at Austin, Austin, TX, USA
dc.luo@utexas.edu

Thomas O'Leary-Roseberry
The University of Texas at Austin
Oden Institute for Computational and Engineering Sciences
tom@oden.utexas.edu

Peng Chen
Georgia Institute of Technology
peng@cc.gatech.edu

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

MS101

Reinforcement Learning for Adaptive Control of PDE-Constrained Environments

PDE-constrained optimization plays an integral role in the decision-making process for a broad range of computational science and engineering applications. In many cases, decision makers can develop reliable response strategies using established optimization methods and high-fidelity simulation models. However, these conventional methods typically work on a case-by-case basis, requiring problem/event-specific knowledge of the exact system state/dynamics before any calculations can be carried out. As a result, these methods can be impractical for time-critical scenarios due to their reliance on expensive, last-minute model queries. In this talk, we explore the potential for using reinforcement learning (RL) to leverage offline calculations and develop flexible control policies capable of adapting to distinct problem realizations in real-time. We demonstrate how this can be accomplished using proximal policy optimization to train an RL agent on a diverse range of randomized problem configurations. More specifically, we consider a family of PDE-constrained optimization problems consisting of advection-diffusion systems on an unstructured mesh with variations in the velocity field and forcing term between each problem realization. We show that the resulting model provides accurate, near-optimal control strategies for the complete family of problem scenarios with minimal computational requirements at run-time.

Nick Winovich, Bart G. Van Bloemen Waanders
Sandia National Laboratories

nwinovi@sandia.gov, bartv@sandia.gov

MS102

Primal-Dual Proximal Optimization Methods with Bregman Divergences

We discuss Bregman distance extensions of the primal-dual three-operator (PD3O) and Condat-Vu proximal algorithms. When used with standard proximal operators these algorithms include several important methods as special cases. Extensions to generalized Bregman distances are attractive if the complexity per iteration can be reduced by matching the Bregman distance to the structure in the problem. As an example, we apply the proposed method to the centering problem in sparse semidefinite programming. The logarithmic barrier function for the cone of positive semidefinite completable sparse matrices is used as a distance-generating kernel. For this distance, the complexity of evaluating the Bregman proximal operator is shown to be roughly proportional to the cost of a sparse Cholesky factorization. This is much cheaper than the standard proximal operator with Euclidean distances, which requires an eigenvalue decomposition.

Xin Jiang
UCLA
xjiang@lehigh.edu

MS102

A Stochastic Bregman Primal-Dual Splitting Algorithm for Composite Optimization

We study a stochastic first order primal-dual method for solving convex-concave saddle point problems over real reflexive Banach spaces using Bregman divergences and relative smoothness assumptions. We show ergodic convergence in expectation of the Lagrangian optimality gap with a rate of O(1/k) and that every almost sure weak cluster point of the ergodic sequence is a saddle point in expectation, under mild assumptions. Our framework is general and does not need strong convexity of the entropies inducing the Bregman divergences in the algorithm. Numerical applications are considered including entropically regularized Wasserstein barycenter problems and regularized inverse problems on the simplex.

Antonio J. Silveti-Falls
Université Paris-Saclay
Gif-sur-Yvette, France
tonys.falls@gmail.com

MS102

Three Operator Splitting with Subgradients, Stochastic Gradients, and Adaptive Learning Rates

Three Operator Splitting (TOS) can minimize the sum of multiple convex functions effectively when an efficient gradient oracle or proximal operator is available for each term. This requirement often fails in machine learning applications: (i) instead of full gradients only stochastic gradients may be available; and (ii) instead of proximal operators, using subgradients to handle complex penalty functions may be more efficient and realistic. Motivated by these concerns, we analyze three potentially valuable extensions of TOS. The first two permit using subgradients and stochastic gradients, and are shown to ensure a $\mathcal{O}(1/\sqrt{t})$ convergence rate. The third extension AdapTOS endows TOS with adaptive step-sizes. For the important setting of optimizing a convex loss over the intersection of convex sets AdapTOS attains universal convergence rates, i.e., the rate adapts to the unknown smoothness degree of the objective. We compare our proposed methods with competing methods on various applications.

Alp Yurtsever
Umeå University
alp.yurtsever@umu.se

Alex Gu, Suvrit Sra
Massachusetts Institute of Technology
gua@mit.edu, suvrit@mit.edu

MS103

Advances in Cyclic Block Coordinate Methods: Gradient Extrapolation, Acceleration, and Variance Reduction

In this talk, I will present new results advancing the theory of cyclic block coordinate methods, enabled by a novel Lipschitz condition w.r.t. a Mahalanobis norm. The newly introduced Lipschitz condition is motivated by addressing block coordinate updates in variational inequality (equilibrium) problems, where traditional block Lipcshitz conditions appear insufficient. I will first present results that address generalized variational inequality problems using gradient extrapolation and cyclic block coordinate updates. I will then discuss further developments of this theory, including variance reduction, gradient norm guarantees in smooth nonconvex minimization settings, and acceleration in composite (smooth + nonsmooth) convex settings.

Jelena Diakonikolas
University of Wisconsin-Madison
jelena@cs.wisc.edu

MS103

Low Rank Approximation for Faster Optimization

Low rank structure is pervasive in real-world datasets. This talk shows how to accelerate the solution of fundamental computational problems, including eigenvalue decomposition, linear system solves, composite convex optimization, and stochastic optimization (including deep learning), by exploiting this low rank structure. We present a simple method based on randomized numerical linear algebra for efficiently computing approximate top eigendecompositions, which can be used to replace large matrices (such as Hessians and constraint matrices) with low rank surrogates that are faster to apply and invert. The resulting solvers for linear systems (NystromPCG), composite convex optimization (NysADMM), and deep learning (SketchySGD) demonstrate strong theoretical and numerical support, outperforming state-of-the-art methods in terms of speed and robustness to hyperparameters.

Zachary Frangella
Stanford University, U.S.

zfran@stanford.edu

Pratik Rathore
Stanford University
pratikr@stanford.edu

Shipu Zhao
Amazon
sz533@cornell.edu

Madeleine Udell
Stanford University
udell@stanford.edu

MS103

Dynamics of Sgd with Stochastic Polyak Stepsizes: Truly Adaptive Variants and Convergence to Exact Solution

In this talk, we study the dynamics and the convergence properties of SGD equipped with new variants of the stochastic Polyak stepsize (SPS) and provide solutions to the two main drawbacks of the original SPS. That is, (i) it requires a priori knowledge of the optimal mini-batch losses, which are not available when the interpolation condition is not satisfied (e.g., regularized objectives), and (ii) it guarantees convergence only to a neighborhood of the solution. We first show that a simple modification of the original SPS that uses lower bounds instead of the optimal function values can directly solve issue (i). On the other hand, solving issue (ii) turns out to be more challenging and leads us to valuable insights into the methods behavior. We show that if interpolation is not satisfied, the correlation between SPS and stochastic gradients introduces a bias, which effectively distorts the expectation of the gradient signal near minimizers, leading to non-convergence - even if the stepsize is scaled down during training. To fix this issue, we propose DecSPS, a novel modification of SPS, which guarantees convergence to the exact minimizer - without a priori knowledge of the problem parameters. For strongly-convex optimization problems, DecSPS is the first stochastic adaptive optimization method that converges to the exact solution without restrictive assumptions like bounded iterates/gradients.

Nicolas Loizou
Johns Hopkins University
nloizou@jhu.edu

MS104

SAT- and IP-Based Causal Discovery Methods for Models with Latents and Cycles

I will review some of the extant methods for causal structure discovery based on SAT-solvers, and then present some new results where we have used integer programming to identify causal structure in the presence of unmeasured confounding and feedback cycles. This work contributes to an effort to develop discovery algorithms to identify causal structure from observational data for a broad variety of precise, but weak, background assumptions, while retaining computational tractability.

Frederick Eberhardt
California Institute of Technology
fde@caltech.edu

MS104

Linear Causal Disentanglement via Interventions

Causal disentanglement seeks a representation of data involving latent variables that relate to one another via a causal model. A representation is identifiable if both the latent model and the transformation from latent to observed variables are unique. In this paper, we study observed variables that are a linear transformation of a linear latent causal model. Data from interventions are necessary for identifiability: if one latent variable is missing an intervention, we show that there exist distinct models that cannot be distinguished. Conversely, we show that a single intervention on each latent variable is sufficient for identifiability. Our proof uses a generalization of the RQ decomposition of a matrix that replaces the usual orthogonal and upper triangular conditions with analogues depending on a partial order on the rows of the matrix, with partial order determined by a latent causal model. We corroborate our theoretical results with a method for causal disentanglement that accurately recovers a latent causal model.

Chandler Squires
Massachusetts Institute of Technology
csquires@mit.edu

Anna Seigal
Harvard University
aseigal@seas.harvard.edu

Caroline Uhler
Massachusetts Institute of Technology
cuhler@mit.edu

MS104

Characterization and Greedy Learning of Gaussian Structural Causal Models under Unknown Interventions

We consider the problem of recovering the causal structure underlying observations from different experimental conditions when the targets of the interventions in each experiment are unknown. We assume a linear structural causal model with additive Gaussian noise and consider interventions that perturb their targets while maintaining the causal relationships in the system. Different models may entail the same distributions, offering competing causal explanations for the given observations. We fully characterize this equivalence class and offer identifiability results, which we use to derive a greedy algorithm called GnIES to recover the equivalence class of the data-generating model without knowledge of the intervention targets. In addition, we develop a novel procedure to generate semi-synthetic datasets with known causal ground truth but distributions closely resembling those of a real dataset of choice. We leverage this procedure and evaluate the performance of GnIES on synthetic, real, and semi-synthetic datasets. Despite the strong Gaussian distributional assumption, GnIES is robust to an array of model violations and competitive in recovering the causal structure in small- to large-sample settings. We provide, in the Python packages gnies and sempller, implementations of GnIES and our semi-synthetic data generation procedure.

Armeen Taeb
University of Washington
ataeb@uw.edu

Juan Gamella, Peter Buehlmann
ETH Zurich
juan.gamella@stat.math.ethz.ch,
peter.buehlmann@stat.math.ethz.ch

Christina Heinze-Deml
Apple Health AI, U.S.
heinzedeml@gmail.com

MS105

Iteratively Reweighted Krylov Subspace Methods for Sparse Reconstruction

This talk presents a couple of new iteratively reweighted least squares (IRLS) algorithms applied to compute sparse solutions of large-scale linear discrete ill-posed problems through l2-l1 regularization. The highlight of the new algorithms is the use of flexible Krylov-Tikhonov methods to partially solve each least squares problem within the IRLS scheme, leading to an overall more efficient scheme when compared to well-established IRLS solvers. The performance of these algorithms is shown through a variety of numerical examples modelling image deblurring and computed tomography.

Silvia Gazzola
University of Bath
s.gazzola@bath.ac.uk

James Nagy, Malena Sabate Landman
Emory University
jnagy@emory.edu, msabat3@emory.edu

Julianne Chung
Department of Mathematics
Emory University
julianne.mei-lynn.chung@emory.edu

MS105

Fantastic Iteratively Reweighted Algorithms and Where to Find Them

Iteratively reweighted methods, or in particular iteratively reweighted least-squares (IRLS), have been a powerful algorithmic paradigm for structured (non-)convex and (non-)smooth optimization problems. Well, but where to find them? We can find them in 1937, when Weiszfeld proposed an IRLS method for the now famous Fermat-Weber problem of computing the geometric median of three points; in sparse reconstruction and robust regression, computer vision (e.g., multi-image matching), deep networks (e.g., dropout), matrix and tensor decomposition, and more. To make it even easier: find them in this mini-symposium. This talk provides a background for iteratively reweighted methods and serves as an overview of the subsequent talks in the mini-symposium. More specifically, it aims to cover (1) some historical aspects, (2) an introduction to IRLS with several applications and connections to other fields, and (3) basic computational considerations and theoretical guarantees.

Liangzu Peng
Johns Hopkins University
lpeng25@jhu.edu

Christian Kümmerle
University of North Carolina at Charlotte
kuemmerle@uncc.edu

Rene Vidal
Mathematical Institute for Data Science
Johns Hopkins University
rvidal@jhu.edu

MS105

Fast and Robust Iteratively Reweighted Methods for Multi-Image Matching

Previous partial permutation synchronization (PPS) algorithms, which are commonly used for multi-image matching, often involve computation-intensive and memory-demanding matrix operations. These operations become intractable for large scale structure-from-motion datasets. For pure permutation synchronization, the recent Cycle-Edge Message Passing (CEMP) framework suggests a memory-efficient and fast solution. Here we overcome the restriction of CEMP to compact groups and propose an improved algorithm, CEMP-Partial, for estimating the corruption levels of the observed partial permutations. It allows us to subsequently implement a nonconvex weighted projected power method without the need of spectral initialization. The resulting new PPS algorithm, MatchFAME (Fast, Accurate and Memory-Efficient Matching), only involves sparse matrix operations, and thus enjoys lower time and space complexities in comparison to previous PPS algorithms. We prove that under adversarial corruption, though without additive noise and with certain assumptions, CEMP-Partial is able to exactly classify corrupted and clean partial permutations. We demonstrate the state-of-the-art accuracy, speed and memory efficiency of our method on both synthetic and real datasets.

Yunpeng Shi
Princeton University
yunpengs@princeton.edu

MS106

Global convergence of the gradient method for functions definable in o-minimal structures

We consider the gradient method with variable step size for minimizing functions that are definable in o-minimal structures on the real field and differentiable with locally Lipschitz gradients. We prove that global convergence holds if continuous gradient trajectories are bounded, with the minimum gradient norm vanishing at the rate o(1/k) if the step sizes are greater than a positive constant. If additionally the gradient is continuously differentiable, all saddle points are strict, and the step sizes are constant, then convergence to a local minimum holds almost surely over any bounded set of initial points.

Cedric Josz
Columbia University
cj2638@columbia.edu

MS106

Partial Smoothness of Set-Valued Operators

Over the past two decades, non-smooth optimization has witnessed a tremendous success in several fields cross science and engineering. Though with non-smooth objectives, first-order algorithms eventually exhibit behaviors analogous to smooth optimization methods. Such observations can be well explained by the framework of "partial smoothness". In this talk, I will introduce an extension of partial smoothness which goes beyond functions. Under the setting of set-valued operator, partial smoothness will be constructed. The benefits of this extension includes geometric interpretation of finite activity identification, upper bounds on identification steps and what happens when non-degeneracy condition fails.

Jingwei Liang
Institute of Natural Sciences
Shanghai Jiaotong University
jingwei.liang@sjtu.edu.cn

MS106

Stochastic Subgradient Descent Avoids Active Strict Saddles of Weakly Convex, Semialgebraic Functions

In non-smooth stochastic optimization, we establish the non-convergence of the stochastic subgradient descent (SGD) to the critical points recently called active strict saddles by Davis and Drusvyatskiy. Such points lie on a manifold M where the function f has a direction of second-order negative curvature. Off this manifold, the norm of the Clarke subdifferential of f is lower-bounded. We require two conditions on f. The first assumption is a Verdier stratification condition, which is a refinement of the popular Whitney stratification. It allows us to establish a strengthened version of the projection formula of Bolte et. al. for Whitney stratifiable functions, and which is of independent interest. The second assumption, termed the angle condition, allows to control the distance of the iterates to M. When f is weakly convex, our assumptions are generic. Consequently, generically in the class of definable weakly convex functions, SGD converges to a local minimizer.

Pascal Bianchi
Télécom ParisTech
pascal.bianchi@telecom-paris.fr

Walid Hachem
CNRS / LIGM Université Gustave Eiffel
walid.hachem@univ-eiffel.fr

Sholom Schechtman
Télécom SudParis
sholom.schechtman@telecom-sudparis.eu

MS107

Global solution of nonconvex variational problems via sparse polynomial optimization

This talk will discuss recent results related to the application of sparse polynomial optimization to integral variational problems. I will show that the global minimizers of certain such problems can be approximated with arbitrary accuracy if one discretizes the variational problem into a sparse polynomial optimization problem, and then relaxes the latter into a hierarchy of structured semidefinite programs (SDPs). The solutions of these SDPs converge to the global optimizer of the original variational problem when the latter is unique and the SDPs satisfies a so-called "running intersection property". I will also present preliminary analysis suggesting that this condition can be removed in certain cases, which dramatically reduces the computational complexity of the SDP hierarchy. Remaining challenges and opportunities for analytical and computational developments will be discussed if time permits.

Giovanni Fantuzzi
FAU Erlangen-Nürnberg
giovanni.fantuzzi@fau.de

Federico Fuentes
Institute for Mathematical and Computational Engineering
Pontificia Universidad Católica de Chile, Santiago
federico.fuentes@uc.cl

MS107

Using polynomial optimization to studying nonlinear dynamics

A family of methods has been developed over the last 20 years in which polynomial optimization computations are used to study dynamical systems governed by differential equations. These methods are fairly mature for ODE systems of moderate dimension, and extensions to PDEs are underway. All of these methods rely on constructing an auxiliary function (mapping phase space to the real numbers) that satisfies certain inequalities; different inequality constraints imply different statements about the dynamics. The inequality constraints can be strengthened into conditions that are computationally tractable by polynomial optimization, usually sum-of-squares constraints. Statements about dynamical systems that can be computed in this way include bounds on properties of an attractor or on transient behavior, estimates of basins of attraction, design of nonlinear controls, and more. This talk will provide an introduction to this family of methods in the ODE context. I will also describe challenges presented by these methods that call for advances in computational polynomial optimization.

David Goluskin
University of Victoria

goluskin@uvic.ca

MS107

Loss functions for finite sets

This talk discusses loss functions for finite sets. For a given finite set S, we give sum-of-square type loss functions of minimum degree. When S is the vertex set of a standard simplex, we show such loss functions have no spurious minimizers (i.e., every local minimizer is a global one). Up to transformations, we give similar loss functions without spurious minimizers for general finite sets. When S is approximately given by a sample set T, we show how to get loss functions by solving a quadratic optimization problem. Numerical experiments and applications are given to show the efficiency of these loss functions.

Jiawang Nie, Suhan Zhong
University of California, San Diego
njw@math.ucsd.edu, suzhong@tamu.edu

MS108

Distributionally Robust Machine Learning with Kernel Machines

Distributionally robust optimization (DRO) has recently been applied to robust machine learning (ML) under *distribution shifts*, i.e., train-test distribution mismatches. To address some of the shortcomings of Wasserstein-DRO for ML problems, we first introduce the *kernel-DRO* framework based on the *maximum-mean-discrepancy* (MMD) geometry, which is tailored for learning nonlinear models. Going beyond joint DRO formulations, we give special attention to a particularly severe type of distribution shift induced by *causal confounding*, a prevalent problem in causal inference and reinforcement learning. Confounding biases can be eliminated by phrasing the estimation problem in terms of *conditional moment restrictions* (CMR). To address such problems, we introduce the *empirical likelihood* (EL) framework, which in its classical form bears the interpretation of a dual formulation of the f-divergence based DRO framework. We provide extensions of EL beyond f-divergences and demonstrate our methodology in both theory and state-of-the-art ML experimental results.

Heiner Kremer, Yassine Nemmour, Bernhard Schoelkopf
Max Planck Institute for Intelligent Systems
heiner.kremer@tuebingen.mpg.de,
ynemmour@tuebingen.mpg.de, bs@tuebingen.mpg.de

Jia-Jie Zhu
Weierstrass Institute for Applied Analysis and Stochastics
zhu@wias-berlin.de

MS108

Data-Driven Tolerated Regret

We study a two-stage expected regret minimization problem with an uncertain technology matrix and an uncertain right-hand side vector. As the expectation is a linear operator, this problem is readily solved by the sample-average-approximation solution when we treat the empirical distribution as the true data-generating distribution. As this premise however is likely untrue, we propose to avoid the post-decision disappointment by minimizing the expected regret in excess of the tolerated regret threshold under any distribution of the uncertainty relative to the Wasserstein distance between this and the empirical distribution. We show that our problem admits a safe tractable approximation which is always feasible and works with a variety of regrets including but not limited to ex-post (which benchmarks each decision against a fully adaptive comparator) and ex-ante (which benchmarks each decision against a static comparator) regrets. Finally, we showcase the benefits of our approach via several numerical studies from economics and operations management.

Yilin Xue
National University of Singapore
yilin.xue@u.nus.edu

Napat Rujeerapaiboon
NUS
napat.rujeerapaiboon@nus.edu.sg

MS108

When Does Simple Estimate-Then-Optimize Beats Integrated Estimation-Optimization? A Stochastic Dominance Perspective

In data-driven stochastic optimization, the model parameter in the underlying distribution needs to be calibrated from data in addition to the optimization task. Recent literature suggests the integration of the estimation and optimization processes, by selecting the model parameter that gives rise to the best empirical objective value performance. While such an integrated approach can be readily shown to outperform simple "estimate then optimize" when the model is misspecified, we argue that, when the model class is rich enough to cover the ground truth, the performance ordering between the two approaches is reversed for nonlinear problems under mild regularity: Simple "estimate then optimize" outperforms the integrated approach, in the strong sense of stochastic dominance of the asymptotic optimality gap (i.e., not only mean or moments but the entire asymptotic distribution of optimality gap, subject to the data noise, is always better). We also demonstrate experimental findings to support our theory.

Adam Elmachtoub, Henry Lam, Haofeng Zhang, Yunfan Zhao
Columbia University
adam@ieor.columbia.edu, henry.lam@columbia.edu,
hz2553@columbia.edu, yz3685@columbia.edu

MS109

A Framework for Incorporating Strategic Asset and Product Decisions in Integrated Planning and Scheduling

Integrating different levels of decision-making has been a long-standing problem in supply chain planning. In this work, we focus on integration of strategic and operational level decision-making while exploring a wider array of strategic choices that often arise in real-world process industries. In particular, we consider a multi-product, multi-asset batch facility along with associated inventory and logistical capacities. At the strategic level, we consider decisions including (1) introduction of new products, (2) introduction of assets and storage tanks, (3) expansion of production capability of assets, and (4) long-term production targets and product mix based on demand projection. We decompose the problem into a strategic planning model that is solved at the master level and the feasibility of the underlying operational model is checked in the sub-level. According to the result of the sub-level we add optimality or feasibility cuts to the master level and iterate until decisions in both levels converge. We propose novel constraints to account for the number of inventory turns targeted by supply chain organizations. We also propose novel production target constraints that the cyclic schedule in the operational level must satisfy on average. Through industrial-scale case studies, we show that this approach can significantly reduce the risk of strategic projects failing, leading to improved long-term profit.

Satyajith Amaran, Maria Paz Ochoa
The Dow Chemical Company
samaran@dow.com, mpochoa@dow.com

MS109

Online Algorithms for Inbound Optimization

When sellers send shipments to Amazon crossdocks, a variety of costs are incurred, including: shipment, congestion, and downstream placement and fulfillment costs. This work considers the problem of the online assignment of sellers to crossdocks with the objective of minimizing cumulative costs. To this end, we propose a congestion-aware hybrid Model-Predictive Control + dynamic dual update algorithm that improves upon state-of-the-art techniques for inbound optimization.

Chamsi Hssaine
USC
Amazon
ch822@cornell.edu

Garrett van Ryzin, Alberto Vera
Amazon
ryzing@amazon.com, albvra@amazon.com

Yash Kanoria
Amazon
Columbia DRO
ykanoria@amazon.com

MS109

Interpretable Learning of Supply Chain Planner Preferences Via Inverse Optimization

Mathematical optimization is increasingly playing an important role in supply chain planning and operations. A key challenge in the use of optimization-based tools are the uncertainties involved in day-to-day planning such as demand swings and production disruptions. In many cases, a two-stage approach is used, wherein deterministic optimization based on forecasts is initially used to develop a base plan, which is then modified by human planners when different uncertainties manifest over time. The decision-making strategies of humans, however, are often inconsistent and opaque. This necessitates the development of a tool that can benchmark the implemented plans against the ones considered ideal by the business. We consider the development of such a tool by fitting an interpretable model on historical planning data. The interpretability of this model is desired because ultimately, we aim to decipher the factors that led to a particular planning decision. Therefore, in this work, we use an inverse optimization-based approach to learning an inherently interpretable optimization model from historical data. The inverse optimization problem is formulated as a bilevel optimization problem with multiple mixed-integer linear optimization problems in its lower level and we apply a cutting plane algorithm to solve this problem. We highlight the real-world utility of our approach through case studies based on realistic supply chain systems.

Rishabh Gupta, Qi Zhang
University of Minnesota
rishabh@umn.edu, qizh@umn.edu

MS110

Unifying Width-Reduced Methods for Quasi-Self-Concordant Optimization

Width reduction is a technique that has proven immensely useful in the context of maximum flow [Christiano et al., STOC11] and, more recently, p-norm regression [Adil et al., SODA19], in terms of improving the iteration complexity from $m^{1/2}$ to $m^{1/3}$, where m is the number of rows of the design matrix, and where each iteration amounts to a linear system solve. However, a considerable drawback is that these methods require both problem-specific potentials and individually tailored analyses. In this talk, I will talk about the first unified approach to achieving $m^{1/3}$-type rates. Notably, our method goes beyond these previously considered problems to more broadly capture quasi-self-concordant losses, a class which has recently generated much interest and includes the well-studied problem of logistic regression, among others. In order to do so, we develop a unified width reduction method for carefully handling these losses based on a more general set of potentials. Additionally, we directly achieve $m^{1/3}$ - type rates in the constrained setting without the need for any explicit acceleration schemes, thus naturally complementing recent work based on an accelerated ball-oracle approach [Carmon et al., NeurIPS20].

Deeksha Adil
Institute for Theoretical Studies, ETH Zurich
deeksha@cs.toronto.edu

MS110

Advances in Ball Acceleration Methods

Much of the recent work in theoretical convex optimization arises through the study of accelerated iterative methods (in the sense of Nesterov) which can leverage problem structure beyond standard smoothness. One natural form of such structure is the ability to implement a ball oracle: i.e, the ability to efficiently minimize the objective within small Euclidean balls. In this talk, I will discuss several recent advances in the theory of convex optimization, which leverage the notion of ball oracles to achieve near-optimal complexity guarantees. I will begin by giving a high-level overview of a near-optimal method for minimizing convex functions admitting a ball oracle. I will follow by discussing the implementation of such ball oracles in a variety of settings, such as quasi-self concordant convex minimization, the minimization of the maximum of Lipschitz functions, and in differentially-private stochastic convex optimization.

Arun Jambulapati
Stanford University
jmblpati@uw.edu

MS110

Fast Convex Optimization Via Time Scale and Averaging of the Steepest Descent

In a Hilbert setting, we develop a gradient-based dynamic approach for fast solving convex optimization problems. By applying time scaling, averaging, and perturbation techniques to the continuous steepest descent (SD), we obtain high-resolution ODEs of the Nesterov and Ravine methods. These dynamics involve asymptotically vanishing viscous damping and Hessian driven damping (either in explicit or implicit form). Mathematical analysis does not require developing a Lyapunov analysis for inertial systems. We simply exploit classical convergence results for (SD) and its external perturbation version, then use tools of differential and integral calculus, including Jensen's inequality. By way of illustration of the flexibility of the method we consider the case where the initial dynamics is the regularized Newton method. Then, we show that the technique can be naturally extended to the case of a monotone cocoercive operator. Our approach leads to parallel algorithmic results, which we study in the case of proximal algorithms. If time permits, we will discuss the closed-loop control approach.

Dang-Khoa Nguyen
University of Vienna
dang-khoa.nguyen@univie.ac.at

Hedy Attouch
Université Montpellier II, France
hedy.attouch@umontpellier.fr

Radu Ioan Bot
University of Vienna
radu.bot@univie.ac.at

MS111

Exploiting Gpu/simd Architectures for Solving Linear-Quadratic Mpc Problems

We report numerical results on solving constrained linear-quadratic model predictive control (MPC) problems by exploiting graphics processing units (GPUs). The presented method reduces the MPC problem by eliminating the state variables and applies a condensed-space interior-point method to remove the inequality constraints in the KKT system. The final condensed matrix is positive definite and can be efficiently factorized in parallel on GPU/SIMD architectures. In addition, the size of the condensed matrix depends only on the number of controls in the problem, rendering the method particularly effective when the problem has many states but few inputs and moderate horizon length. Our numerical results for PDE-constrained problems show that the approach is an order of magnitude faster than a standard CPU implementation. We also provide an open-source Julia framework that facilitates modeling (DynamicNLPModels.jl) and solution (MadNLP.jl) of MPC problems on GPUs.

David Cole
University of Wisconsin-Madison
dlcole3@wisc.edu

Sungho Shin
Argonne National Laboratory
sshin@anl.gov

François Pacaud
Argonne National Laboratory
9700 S. Cass Avenue Lemont, IL 60439
fpacaud@anl.gov

Victor M. Zavala
University of Wisconsin-Madison
victor.zavala@wisc.edu

Mihai Anitescu
Argonne National Laboratory
Mathematics and Computer Science Division
anitescu@mcs.anl.gov

MS111

Proportional Integral Projected Gradient Method for Real-Time Optimal Control

Proportional-integral projected gradient method (PIPG) is a first-order primal-dual algorithm for conic optimization. It allows library-free, matrix-free, and easily-verifiable solver implementation for real-time and embedded applications. It achieves automatic primal and dual infeasibility detection, optimal global convergence rates in theory, and orders-of-magnitudes faster computation in practice. It exploits the sparsity structure of conic constraints via parallel matrix operations and the geometric structure of constraint sets via efficient projection. Unlike most off-the-shelf methods, PIPG allows easy warm-starts and enjoys a light computational overhead because it avoids the cumbersome canonical transformation of standard conic programs. PIPG not only enables versatile convex optimization but also boosts the performance of popular sequential convex programming (SCP) methods for nonconvex optimization, such as PTR, SCvx and GuSTO. Moreover, the recently introduced SeCO framework specializes SCP algorithms to exploit features of PIPG for nonconvex optimal control problems with real-time capability. We will present the theoretical results on PIPG and demonstrate its real-time performance in MPC and in nonconvex trajectory optimization.

Purnanand Elango, Abhinav Kamath
University of Washington
pelango@uw.edu, agkamath@uw.edu

Yue Yu
The University of Texas at Austin
yueyu@utexas.edu

Behcet Acikmese
University of Washington
behcet@uw.edu

MS111

Homogeneous Formulation of Convex Quadratic Programs for Infeasibility Detection

Convex Quadratic Programs (QPs) have come to play a central role in the computation of control action for constrained dynamical systems. In this paper, we present a novel Homogeneous QP (HQP) formulation which is obtained by embedding the original QP in a larger space. The key properties of the HQP are: (i) is always feasible, (ii) an optimal solution to QP can be readily obtained from a solution to HQP, and (iii) infeasibility of QP corresponds to a particular solution of HQP. An immediate consequence is that all the existing algorithms for QP are now also capable of robustly detecting infeasibility. In particular, we present an Infeasible Interior Point Method (IIPM) for the HQP and show polynomial iteration complexity when applied to HQP. A key distinction with prior IPM approaches is that we do not need to solve second-order cone programs. Numerical experiments on the formulation are provided using existing codes.

Arvind Raghunathan
Mitsubishi Electric Research Laboratories
raghunathan@merl.com

MS112

Optimal scaling quantum linear-systems solver via discrete adiabatic theorem

Recently, several approaches to solving linear systems on a quantum computer have been formulated in terms of the quantum adiabatic theorem for a continuously varying Hamiltonian. Such approaches enabled near-linear scaling in the condition number κ of the linear system, without requiring a complicated variable-time amplitude amplification procedure. However, the most efficient of those procedures is still asymptotically sub-optimal by a factor of $\log(\kappa)$. Here, we prove a rigorous form of the adiabatic theorem that bounds the error in terms of the spectral gap for intrinsically discrete time evolutions. In combination with the qubitized quantum walk, our discrete adiabatic theorem gives a speedup for all adiabatic algorithms. Here, we used this combination to develop a quantum algorithm for solving linear systems that is asymptotically optimal, in the sense that the complexity is strictly linear in κ, matching a known lower bound on the complexity. Our $\mathcal{O}(\kappa \log(1/\epsilon))$ complexity is also optimal in terms of the combined scaling in κ and the precision ϵ. Compared to existing suboptimal methods, our algorithm is simpler and easier to implement. This talk is based on [Costa et al., Optimal scaling quantum linear-systems solver via discrete adiabatic theorem, 2022].

Dong An
University of Maryland
dongan@umd.edu

MS112

Accurately Solving Linear Systems with Quantum Oracles

Quantum linear system algorithms (QLSA), w.r.t dimension, have the potential to solve linear systems (LSs) faster than classical methods. However, to extract the classical solution, a quantum tomography algorithm (QTA) is needed, which increases both error and time complexity. To accurately and efficiently solve LSs using QLSA and QTA algorithms, we propose an Iterative Refinement method (IRM) which uses limited-precision quantum oracles iteratively to improve dependence on precision to logarithmic. The IRM is broadly applicable. We discuss its application in Quantum Interior Point Methods (QIPM) and discuss how the proposed IRM accelerates QIPMs.

Mohammadhossein Mohammadisiahroudi
Industrial & System Engineering Dep, Lehigh University
625 Montclair Ave, Bethlehem, PA, USA, 18015
mom219@lehigh.edu

Brandon Augustino
Lehigh University
bra216@lehigh.edu

Tamas Terlaky
Lehigh University
Department of industrial and Systems Engineering
terlaky@lehigh.edu

Giacomo Nannicini
University of Southern California
g.nannicini@usc.edu

MS112

Inexact Feasible Quantum Interior Point Methods for Linear Optimization

We apply Quantum Linear System Algorithms (QLSAs) to Newton systems within Interior Point Methods (IPMs) to gain quantum speedup in solving Linear Optimization problems (LOPs). Due to their inexact nature, QLSAs can be applied only to inexact variants of IPMs, which are inexact infeasible methods due to the inexact nature of their computations. We propose Inexact-Feasible IPMs (IF-IPM) for solving LOPs, using novel Newton systems to generate inexact but feasible steps. We show that this method enjoys the to-date best iteration complexity. Further, we explore how QLSAs can be used efficiently in iterative refinement schemes to find an exact optimal solution without excessive calls to QLSAs. Finally, we experiment with the proposed IF-IPMs efficiency using IBM's QISKIT

environment.

Tamas Terlaky
Lehigh University
Department of industrial and Systems Engineering
terlaky@lehigh.edu

Mohammadhossein Mohammadisiahroudi
Industrial & System Engineering Dep, Lehigh University
625 Montclair Ave, Bethlehem, PA, USA, 18015
mom219@lehigh.edu

Ramin Fakhimi
Industrial & System Engineering Dep, Lehigh University
raf318@lehigh.edu

Zeguan Wu
Lehigh University, PA, USA
zew220@lehigh.edu

MS113

Inexact Proximal-Methods for Constrained Optimization

We discuss several recent advances in proximal-methods for sparse optimization. We first present a new subproblem solver that allows for inexact solutions of the proximal-gradient subproblem to be computed. We then present inexact stopping conditions that allow us to derive complexity and support identification results. We also introduce a proximal method for solving deterministic optimization problems, as well as to minimize regularized stochastic objective functions subject to deterministic constraints.

Daniel Robinson
Lehigh University
daniel.p.robinson@gmail.edu

Frank E. Curtis
Industrial and Systems Engineering
Lehigh University
frank.e.curtis@gmail.com

Yutong Dai
Lehigh University
yud319@lehigh.edu

MS113

Trust Region Algorithms Under Noise

Classical trust region methods were designed to solve problems in which function and derivative information are exact. This paper considers the case when there are errors (or noise) in the function and derivative information and proposes modifications of the trust region method to make solving robust under noise. These new algorithms requires information about the size/standard deviation of the errors in the function evaluations and incurs no additional computational expense. It is shown that, when certain regularity assumptions are satisfied in the objective function, the iterates of the algorithm visit a neighborhood of stationarity infinitely often, assuming errors in the problem are bounded. It is also shown that, after visiting the above neighborhood for the first time, the iterates cannot stray too far from it. Numerical results illustrate how the classical trust region algorithm may fail in the presence of noise, and how the proposed algorithm ensures steady progress towards stationarity in these cases.

Shigeng Sun
Northwestern University
shigengsun2024@u.northwestern.edu

Jorge Nocedal
Department of Industrial Engineering and Management Sciences
Northwestern University
j-nocedal@northwestern.edu

MS113

Statistical Preconditioning for Distributed Empirical Risk Minimization

We consider the setting of distributed empirical risk minimization where multiple machines compute the gradients in parallel and a centralized server updates the model parameters. In order to reduce the number of communications required to reach a given accuracy, we propose a preconditioned accelerated gradient method where the preconditioning is done by solving a local optimization problem over a subsampled dataset at the server. The convergence rate of the method depends on the square root of the relative condition number between the global and local loss functions. We estimate the relative condition number for linear prediction models by studying uniform concentration of the Hessians over a bounded domain, which allows us to derive improved convergence rates for existing preconditioned gradient methods and our accelerated method. Experiments on real-world datasets illustrate the benefits of acceleration in the ill-conditioned regime.

Hadrien Hendrikx
EPFL
hadrien.hendrikx@epfl.ch

Lin Xiao
Facebook
lin.xiao@gmail.com

Sebastien A. Bubeck
Microsoft Research
sebubeck@microsoft.com

Francis Bach
INRIA - École Normale Supérieure
francis.bach@inria.fr

Laurent Massoulie
INRIA
laurent.massoulie@inria.fr

MS114

A Stochastic Subgradient Method for Distributionally Robust Non-Smooth Non-Convex Learning

We consider a distributionally robust formulation of stochastic optimization problems arising in statistical learning, where robustness is with respect to uncertainty in the underlying data distribution. Our formulation builds on risk-averse optimization techniques and the theory of coherent risk measures. It uses semi-deviation risk for quantifying uncertainty, allowing us to compute robust solutions against perturbations in the population data distribution while leading to a two-level stochastic optimization problem. We consider a broad class of generalized differentiable loss functions that can be non-convex and non-smooth, involving upward and downward cusps and we develop an efficient stochastic subgradient method for distributionally robust problems with such functions. We prove that it converges to a point satisfying the optimality conditions. To our knowledge, this is the first method with rigorous convergence guarantees in the context of generalized differentiable non-convex and non-smooth distributionally robust stochastic optimization. Our method allows for control of the desired level of robustness with little extra computational cost compared to population risk minimization with stochastic gradient methods. For the class of weakly convex objectives, we also develop tight iteration complexity bounds. Finally, we illustrate the performance of our algorithm on real datasets arising in convex and non-convex supervised learning problems.

Mert Gurbuzbalaban, Andrzej Ruszczynski, Landi Zhu
Rutgers University
mert.gurbuzbalaban@gmail.com,
rusz@business.rutgers.edu,
lz401@scarletmail.rutgers.edu

MS114

Simultaneous Learning of Fitting Surface and Noise Distribution from Large Data Sets

In this talk we address the problem of learning surfaces from large data sets subject to (possible large amounts of) noise. More precisely, we assume that the surface is described by the zero set of a function belonging to the set spanned by a given basis. Moreover, it is assumed that the noise is iid and its distribution is known up to some parameters. An example of this is noise uniformly distributed over a symmetric interval, but the bounds are not known a priori. Under these two assumptions, we show that the estimation of the function whose zero set describes the surface can be computed by determining the null space of a matrix M that can be efficiently approximated from the available data. Also, we show that the noise parameters can be determined by searching for the values that result in the matrix M mentioned above being singular. Moreover, we show that, as the number of data points tend to infinity, we recuperate the desired surface and the true noise parameters. Numerical results are provided showing the effectiveness of the proposed approach.

Constantino Lagoa
Penn State University
Pennsylvania State University
lagoa@psu.edu

Omar Sleem
The Pennsylvania State University
oms46@psu.edu

MS114

Zeroth-Order Methods for Constrained Minimization of Expectation-Valued Lipschitz Continuous Functions

We consider the minimization of an L_0-Lipschitz continuous and expectation-valued function, denoted by f and defined as $f(x) \triangleq \mathbb{E}[\tilde{f}(x,\omega)]$, over a closed and convex set $\mathcal{X}$ with a view towards obtaining both asymptotics as well as rate and complexity guarantees for computing an approximate stationary point (in a Clarke sense). We adopt a smoothing-based approach reliant on minimizing f_η where $f_\eta(x) \triangleq \mathbb{E}_u[f(x+\eta u)] + \mathbb{I}_{\eta,\mathcal{X}}$, where u is a random variable defined on a unit sphere, $\eta > 0$, and $\mathbb{I}_{\eta,\mathcal{X}}$ is the Moreau-smoothed indicator of the set $\mathcal{X}$. In fact, it is observed that a stationary point of the η-smoothed problem is a 2η-stationary point for the original problem in the Clarke sense. In such a setting, we derive a suitable residual function that provides a metric for stationarity for the smoothed problem. Time permitting, we present the rate statements and guarantees for a stochastic quasi-Newton framework reliant on utilizing sampled function evaluations.

Luke W. Marrinan
Pennsylvania State University
lwm5431@psu.edu

MS115

Block Coordinate Descent for Smooth Nonconvex Constrained Minimization

At each iteration of a Block Coordinate Descent method one minimizes an approximation of the objective function with respect to a generally small set of variables subject to constraints in which these variables are involved. The unconstrained case and the case in which the constraints are simple were analyzed in the recent literature. In this paper we address the problem in which block constraints are not simple and, moreover, the case in which they are not defined by global sets of equations and inequations. A general algorithm that minimizes quadratic models with quadratric regularization over blocks of variables is defined and convergence and complexity are proved. In particular, given tolerances $\delta > 0$ and $\varepsilon > 0$ for feasibility/complementarity and optimality, respectively, it is shown that a measure of $(\delta, 0)$-criticality tends to zero; and the the number of iterations and functional evaluations required to achieve (δ, ε)-criticality is $O(\varepsilon^{-2})$. Numerical experiments in which the proposed method is used to solve a continuous version of the traveling salesman problem are presented.

Ernesto G. Birgin
Department of Computer Science - IME
University of São Paulo
egbirgin@ime.usp.br

José Mario Martínez
University of Campinas

martinez@ime.unicamp.br

MS115

Fast Convergence of Greedy 2-Coordinate Updates for Optimizing with An Equality Constraint

We consider minimizing a smooth function subject to an equality constraint. We analyze a greedy 2-coordinate update algorithm, and prove that greedy coordinate selection leads to faster convergence than random selection (under a Polyak-Lojasiewicz assumption). Our simple analysis exploits an equivalence between the greedy 2-coordinate update and equality-constrained steepest descent in the 1-norm. Unlike previous 2-coordinate analyses, our convergence rate is dimension independent.

Mark Schmidt
University of British Columbia
schmidtm@cs.ubc.ca

MS115

An Adaptive Superfast Inexact Proximal Augmented Lagrangian Method for Smooth Nonconvex Composite Optimization Problems

This talk presents an adaptive superfast proximal augmented Lagrangian (AS-PAL) method for solving linearly-constrained smooth nonconvex composite optimization problems. Each iteration of AS-PAL inexactly solves a possibly nonconvex proximal augmented Lagrangian (AL) subproblem obtained by an aggressive/adaptive choice of prox stepsize with the aim of substantially improving its computational performance followed by a full Lagrangian multiplier update. A major advantage of AS-PAL compared to other AL methods is that it requires no knowledge of parameters (e.g., size of constraint matrix, objective function curvatures, etc) associated with the optimization problem, due to its adaptive nature not only in choosing the prox stepsize but also in using a crucial adaptive accelerated composite gradient variant to solve the proximal AL subproblems. The speed and efficiency of AS-PAL is demonstrated through extensive computational experiments showing that it can solve many instances more than ten times faster than other state-of-the-art penalty and AL methods, particularly when high accuracy is required.

Arnesh M Sujanani
Georgia Institute of Technology, U.S.
asujanani6@gatech.edu

Renato Monteiro
Georgia Institute of Technology
rm88@gatech.edu

MS116

Supermodular Extension of Vizing's Edge-Coloring Theorem

Kőnig's edge-coloring theorem for bipartite graphs and Vizing's edge-coloring theorem for general graphs are celebrated results in graph theory and combinatorial optimization. Schrijver generalized Kőnig's theorem to a framework defined with a pair of intersecting supermodular functions. The result is called the supermodular coloring theorem. In this talk, we present a common generalization of Vizing's theorem and a weaker version of the supermodular coloring theorem. To describe this theorem, we introduce strongly triple-intersecting supermodular functions, which are extensions of intersecting supermodular functions. We also provide an alternative proof of Gupta's edge-coloring theorem using a special case of the supermodular version of Vizing's theorem.

Ryuhei Mizutani
The University of Tokyo
ryuhei_mizutani@mist.i.u-tokyo.ac.jp

MS116

Pairwise Disjoint Perfect Matchings in r-Graphs

For $0 \leq t \leq r$, let $m(t,r)$ be the maximum number s such that every t-edge-connected r-graph has s pairwise disjoint perfect matchings. There are only a few values of $m(t,r)$ known, for instance $m(3,3) = m(4,r) = 1$, and $m(t,r) \leq r-2$ for all $t \neq 5$. We show that $m(t,r) \leq r-3$ if r is even. We prove that $m(2l,r) \leq 3l-6$ for every $l \geq 3$ and $r \geq 2l$. Furthermore, we relate statements on pairwise disjoint perfect matchings of 5-edge-connected 5-regular graphs to well-known conjectures for cubic graphs, such as the Fan-Raspaud Conjecture, the Berge-Fulkerson Conjecture and the 5-Cycle Double Cover Conjecture. This is joint work with Yulai Ma, Davide Mattiollo and Isaak H. Wolf.

Eckhard Steffen
Paderborn University
eckhard.steffen@uni-paderborn.de

MS117

From Complementarity to Risk-averse stochastic equilibria: models and algorithms

We present a mechanism for describing and solving collections of optimization problems that are linked by equilibrium conditions. We show how to incorporate stochastic information into these systems and give examples of their use and possible extensions to hierarchical modeling. We outline several algorithms and investigate their computational efficiency. Stochastic equilibria can be used to model many data driven applications, including many dynamic models of competition. In this model, players solve risk averse optimization problems that are coupled by a shared scenario tree. Players are linked by equilibrium conditions at nodes of that tree. Applications to energy planning and their interactions with environmental concerns will be outlined.

Michael C. Ferris
University of Wisconsin-Madison
ferris@cs.wisc.edu

MS117

Algorithms for Composite Nonsmooth Nonconvex Optimization

Approximations of optimization problems arise in computational procedures and sensitivity analysis. The resulting effect on solutions can be significant, with even small approximations of components of a problem translating into large errors in the solutions. We specify conditions under which approximations are well behaved in the sense of minimizers, stationary points, and level-sets and this leads to a framework of consistent approximations. The framework is developed for a broad class of composite problems, which are neither convex nor smooth. We demonstrate the framework using examples from stochastic optimization, neural-network based machine learning, distributionally robust optimization, penalty and augmented Lagrangian methods, interior-point methods, homotopy methods, smoothing methods, extended nonlinear programming, difference-of-convex programming, and multi-objective optimization. An enhanced proximal method illustrates the algorithmic possibilities. A quantitative analysis supplements the development by furnishing rates of convergence.

Johannes O. Royset
Operations Research Department
Naval Postgraduate School
joroyset@nps.edu

MS118

Semidefinite Relaxations Beyond Polynomials

Semidefinite programming has been used to great effect to derive tractable relaxations for (noncommutative) polynomial optimization problems. In this talk I will discuss ways to derive relaxations for nonconvex and nonpolynomial problems involving the entropy function. The talk will be illustrated with problems from the area of quantum information theory.

Hamza Fawzi
University of Cambridge
h.fawzi@damtp.cam.ac.uk

MS118

Invariants of SDP Exactness in Quadratic Programming

In this talk I will consider the Shor relaxation of quadratic programs by fixing a feasible set and considering the space of objective functions for which the Shor relaxation is exact. I first give conditions under which this region is invariant under the choice of generators defining the feasible set. I then will describe this region when the feasible set is invariant under the action of a subgroup of $GL_n(\mathbb{R})$. If time permits, I will conclude by applying these results to quadratic binary programs by giving an explicit description of objective functions where the Shor relaxation is exact and discuss algorithmic implications of this insight.

Julia Lindberg
University of Wisconsin-Madison, USA
jrlindberg@wisc.edu

MS118

Exploiting Geometric Structure in Matrix-Valued Optimization

Matrix-valued optimization tasks arise in many machine learning applications. Often, exploiting non-Euclidean structure in such problems can give rise to algorithms that are computationally superior to standard nonlinear programming approaches. In this talk, we consider the problem of optimizing Riemannian difference of convex functions. Several classical optimization problems involve objective functions of this form, including barycenter problems, the computation of HPD matrix square roots, as well as the computation of Brascamp-Lieb constants. The latter is of central importance in many areas of Mathematics and Computer Science through connections to maximum likelihood estimators in Gaussian models, Tylers M-estimator of scatter matrices and Operator Scaling. We discuss a class of CCCP-style algorithms for solving Riemannian difference of convex functions, where the geometric structure of the problem gives rise to an efficiently solvable fixed-point iteration. We present a detailed convergence analysis for the proposed approach and illustrate its advantages on the problem of computing Brascamp-Lieb constants. This talk is based on joint work with Suvrit Sra.

Melanie Weber
Harvard University
mweber@seas.harvard.edu

MS119

Bayesian Inference with Projected Densities

Constraints are a natural form of prior information in Bayesian inference. In some applications, the parameters of interest are known to lie on the boundary of some set with some probability, but existing models based on truncated priors such as a truncated normal distribution result in posterior distributions that assign zero probability to the boundary. To address this issue, we construct a posterior distribution that assigns positive probability to the boundary of the constraint set by imposing an implicit prior that corresponds to a projection of posterior mass onto the constraint set. We apply the method to Bayesian linear inverse problems and show that samples from the posterior distribution can be obtained by solving a sequence of perturbed constrained least-squares problems. Finally, we derive a Gibbs sampler for a hierarchical extension of the model in the special case where the constraint set is a polyhedral cone.

Martin S. Andersen
Technical University of Denmark
Department of Applied Mathematics and Computer Science
mskan@dtu.dk

Jasper M. Everink
Technical University of Denmark
jmev@dtu.dk

Yiqiu Dong
Technical University of Denmark
Department of Applied Mathematics and Computer Science

yido@dtu.dk

MS119

Second Order Neural Odes for Inverse Imaging Problems

Image segmentation is a fundamental task in image analysis and clinical practice. The current state-of-the-art techniques are based on U-shape type encoder-decoder networks with skip connections, called U-Net. Despite the powerful performance reported by existing U-Net type networks, they suffer from several major limitations. Issues include the hard coding of the receptive field size, compromising the performance and computational cost, as well as the fact that they do not account for inherent noise in the data. They have problems associated with discrete layers, and do not offer any theoretical underpinnings. In this work we introduce continuous U-Net, a novel family of networks for image segmentation. Firstly, continuous U-Net is a continuous deep neural network that introduces new dynamic blocks modeled by second order ordinary differential equations. Secondly, we provide theoretical guarantees for our network demonstrating faster convergence, higher robustness, and less sensitivity to noise. Thirdly, we derive qualitative measures to tailor-made segmentation tasks. We demonstrate, through extensive numerical and visual results, that our model performs existing U-Net blocks for several medical image segmentation benchmarking datasets.

Chun-Wun Cheng
City University
cwcheng82-c@my.cityu.edu.hk

Christina Runkel, Lihao Liu
University of Cambridge, United Kingdom
cr661@cam.ac.uk, ll610@cam.ac.uk

Raymond Chan
City University of Hong Kong
raymond.chan@cityu.edu.hk

Carola-Bibiane Schönlieb
University of Cambridge
cbs31@cam.ac.uk

Angelica I. Aviles-Rivero
University of Cambridge
Department of Pure Mathematics and Mathematical Statistics
ai323@cam.ac.uk

MS119

Efficient Derivative-Free Bayesian Inference for Large-Scale Inverse Problems

We consider Bayesian inference for large scale inverse problems, where computational challenges arise from the need for repeated evaluations of an expensive forward model. This renders most of Markov chain Monte Carlo methods infeasible, since they typically require $O(10^4)$ model runs. Moreover, the forward model are often given as a black box and may be impractical to differentiate. Therefore derivative-free algorithms are highly desirable. We propose a framework, which is built on Kalman methodology and Fisher-Rao gradient flow, to efficiently calibrate and provide uncertainty estimations of such models with noisy observation data. Theoretical guarantees for linear inverse problems are provided. The effectiveness of the framework is demonstrated on several numerical experiments, including proof-of-concept linear/nonlinear examples and two applications: learning of permeability parameters in subsurface flow; and learning subgrid-scale parameters in a general circulation model.

Daniel Huang
California Institute of Technology
dzhuang@caltech.edu

Jiaoyang Huang
University of Pennsylvania
huangjy@wharton.upenn.edu

Sebastian Reich
Universität Potsdam, Germany
sereich@uni-potsdam.de

Andrew Stuart
Computing + Mathematical Sciences
California Institute of Technology
astuart@caltech.edu

MS120

Exact Methods for Discrete Bilevel Problems with a Gamma-Robust Lower Level

In bilevel optimization, it is usually assumed that the leader and the follower act perfectly rational. However, many applications involve players who face some kind of bounded rationality. We study discrete bilevel problems with a follower who faces lower-level data uncertainty as one specific instance of bounded rationality. To this end, we adopt a Γ-robust approach, present tailored reformulations, and provide generic branch-and-cut frameworks. Specifically, we study interdiction problems with a Γ-robust follower for which we derive problem-tailored cuts. Finally, we computationally assess the performance of the proposed methods using the example of Γ-robust knapsack interdiction problems.

Yasmine Beck
Department of Mathematics
Trier University
yasmine.beck@uni-trier.de

Ivana Ljubic
University of Vienna
ljubic@essec.edu

Martin Schmidt
Universität Trier
martin.schmidt@uni-trier.de

MS120

Using disjunctive cuts in a branch-and-cut proce-

dure to solve convex integer bilevel programs

We present a new branch-and-cut method for solving integer bilevel problems with a convex objective function and convex constraints both in the upper and in the lower level. To this end, we generalize the idea of using disjunctive cuts from the literature to separate bilevel-infeasible points. We obtain the cuts by solving a cut generating problem that can be decomposed into a series of smaller subproblems, most of which can be solved in parallel. We prove the correctness of the method and present its applicability by some first numerical results.

Andreas Horländer
Department of Mathematics
Trier University
horlaender@uni-trier.de

Ivana Ljubic
University of Vienna
ljubic@essec.edu

Martin Schmidt
Universität Trier
martin.schmidt@uni-trier.de

MS120

Learning the Followers Objective Function in Sequential Bilevel Games

Bilevel problems model the hierarchical interaction between two decision makers in which the follower's problem appears as a constraint in the leader's problem. In many practical situations, some parts of the follower's problem may not be explicitly known by the leader. In the talk we consider lower-level problems with quadratic objective functions with only partially known coefficients. We propose a method to learn the unknown objective function that is based on the multiplicative weights update algorithm. We prove the convergence of the method and show its applicability using the examples of bilevel knapsack and pricing problems.

Ioana Molan
Department of Mathematics
Trier University
molan@uni-trier.de

Martin Schmidt
Universität Trier
martin.schmidt@uni-trier.de

Johannes Thürauf
Department of Mathematics
Trier University
johannes.thuerauf@uni-trier.de

MS121

Normalizing Flows for Uncertainty Quantification in Seismic Imaging

Uncertainty quantification is becoming increasingly important in geoscience. One of the key stumbling blocks to its wider incorporation into geophysical imaging is the cost of estimating probability distributions. One way to address this is to use machine learning methods that learn a mapping from one distribution into another such as normalizing flows. Normalizing flows define a sequence of invertible and differentiable transformations of probability distributions to map samples from an unknown (usually complicated) probability distribution(posterior) to a simple known distribution by optimizing a set of parameters defined by two neural networks, then the posterior is estimated by transferring the simple probability distribution through the learned mappings. In this way we can draw samples from the posterior without as many expensive forward simulations. We will show applications of normalizing flows to seismic interpolation and to full-waveform inversion and will discuss how we can exploit normalizing flows to translate uncertanties throughout the seismic imaging and processing workflow. We will also discuss methods of better characterizing the resulting uncertainties, particularly as they apply to distributions that are far from Gaussian.

Alison Malcolm
Memorial University of Newfoundland
amalcolm@mun.ca

Rajiv Kumar
Schlumberger
rkumar19@slb.com

Changxiao Sun
Chinese Academy of Sciences
sunchangxiao@apm.ac.cn

MS121

The Lippmann Schwinger Lanczos algorithm for inverse scattering problems

We combine data-driven reduced order models with the Lippmann- Schwinger integral equation to produce a direct nonlinear inversion method. The ROM is viewed as a Galerkin projection and is sparse due to Lanczos orthogonalization. Embedding into the continuous problem, a data-driven internal solution is produced. This internal solution is then used in the Lippmann-Schwinger equation, in a direct or iterative framework. The approach also allows us to process more general transfer functions, i.e., to remove the main limitation of the earlier versions of the ROM based inversion algorithms. We also describe how the generation of internal solutions simplifies in the time domain, and give examples of its use given mono static data, targeting synthetic aperture radar.

Shari Moskow
Drexel University
Department of Mathematics
moskow@math.drexel.edu

MS121

Phase and absorption contrast imaging using intensity measurements

There are many inverse problems in machine learning, such as regression, classification and distribution estimation. In these problems, many data are in a high-dimensional space but exhibit low-dimensional structures. In mathematics, these data can be modeled as random samples on a low-dimensional manifold. In this talk, I will present mathematical and statistical theories of deep neural networks for solving inverse problems in machine learning, where data are randomly sampled on a low-dimensional manifold. The sample complexity crucially depends on the intrinsic dimension of the manifold instead of the ambient dimension of the data.

Chrysoula Tsogka
University of California, Merced
ctsogka@ucmerced.edu

Miguel Moscoso
Universidad Carlos III de Madrid
moscoso@math.uc3m.es

Alexei Novikov
Penn State University
Mathematics
anovikov@math.psu.edu

George C. Papanicolaou
Stanford University
Department of Mathematics
papanico@math.stanford.edu

MS122

BilevelJuMP.jl: Modeling and Solving Bilevel Optimization in Julia

In this talk, we present BilevelJuMP.jl, a JuMP extension that allows users to model and solve bilevel optimization problems. BilevelJuMP relies on the JuMP ecosystem to automatically dualize any conic convex optimization problem and use it to model bilevel problems by building KKT conditions of the lower-level problem. The package has implementations of multiple formulations for complementarity constraints such as SOS1, Fortuny-Amat, quadratic programming, and actual complementarity constraints. Since the package is built as a JuMP extension, it is automatically integrated with all solvers available to JuMP models. These implementations can help practitioners to save time and think more about the modeling aspect of bilevel problems than worry about the programming aspect of each reformulation strategy. We show how the packages were developed, relying on Julia the JuMP infra-structure, and we present some power systems examples of bilevel problems.

Guilherme Bodin
PSR
guilhermebodin@psr-inc.com

MS122

JuMP 1.0: Recent Improvements to a Modeling Language for Mathematical Optimization

JuMP is an algebraic modeling language embedded in the Julia programming language. JuMP allows users to model optimization problems of a variety of kinds, including linear programming, integer programming, conic optimization, semidefinite programming, and nonlinear programming, and handles the low-level details of communicating with solvers. After nearly 10 years in development, JuMP 1.0 was released in March, 2022. In this short communication, we highlight the improvements to JuMP from recent releases up to and including 1.0.

Miles Lubin
Hudson River Trading
miles.lubin@gmail.com

Oscar Dowson
o.dowson@gmail.com

Joaquim Dias Garcia
PUC-Rio
joaquimdgarcia@gmail.com

Joey Huchette
Google Research
jhuchette@google.com

Benoit Legat
Massachusetts Institute of Technology
blegat@mit.edu

Juan Pablo Vielma
Google
juan.pablo.vielma@gmail.com

MS122

Recent Advances with Graph-Based Optimization and Plasmo.jl

Plasmo.jl was originally released in 2018 as a graph-based modeling language for optimization. The package built on JuMP to facilitate modular model building capabilities and provided a data structure to hook-in problem-level decomposition algorithms such Lagrangian and Benders decompositions as well as linear-algebra decomposition techniques such as Schur complement decomposition used in interior point solvers. Since its initial release, Plasmo.jl has refined its underlying abstraction and implementation into what we now call an optigraph. In an optigraph, optimization models live over nodes containing local variables and constraints and are connected through edges by means of linking (or hyper) constraints. An optigraph is a flexible modeling paradigm to analyze complex structures of optimization problems and it helps to exploit such structures to perform different model processing tasks. For instance, Plasmo.jl supports partitioning over various graph and hypergraph representations to discover suitable decompositions for solvers. Plasmo.jl also provides graph analysis techniques such as querying node neighborhoods and incident edges which can be used to develop structured algorithms that exploit graph properties. In this talk, we present the underlying concepts and implementation pieces of Plasmo.jl. We then discuss current efforts towards a unified interface for decomposition solvers and distributed modeling approaches.

Jordan Jalving

Pasteur Labs & ISI
jhjalving@gmail.com

Sungho Shin
Argonne National Laboratory
sshin@anl.gov

MS123

Distributionally Robust Optimization with Independent Random Variables applied to Random Graphs

We present a Distributionally Robust Optimization (DRO) approach for complex networks that only depends on the mean, range and dispersion of the degree distribution. This requires DRO methods for functions of independent random variables. This approach requires solving semi-infinite linear programs, which proves to be very challenging. We therefore resort to solving two relaxations that allow for correlation between the random variables. Both relaxations turn out to give worst-case distributions which are independent and hence also solve the original semi-infinite program. In this way we establish tight bounds for the clustering coefficient and the number of cliques in random graphs and identify extremal random graphs with specific three-point degree distributions that attain these tight bounds. DRO for independent random variables is largely open and unexplored, and we present several open research directions.

Judith Brugman, Johan S.H. van Leeuwaarden
Tilburg University
J.M.Brugman@tilburguniversity.edu,
j.s.h.vanleeuwaarden@tilburguniversity.edu

Clara Stegehuis
Universiteit Twente
c.stegehuis@utwente.nl

MS123

Integer Programming Games and Combinatorial Uncertainty

Nash equilibria enlighten the structure of rational behavior in multi-agent decision-making, that is, decision-making that accounts for the uncertainty stemming from interactions. Nash equilibria are stable solutions, meaning that no single agent has the incentive to defect from them profitably. This talk explores the connections between Nash equilibria and combinatorial uncertainty. We focus on Integer Programming Games (IPGs), a class of simultaneous and non-cooperative games where each player solves a parametrized integer program. IPGs generalize any finite game (e.g., bimatrix or normal-form), and provide a different perspective to frame the combinatorial uncertainty that affects decision-making. We review the theory and practice currently available to compute Nash equilibria, and we propose a connection with robust (combinatorial) optimization.

Gabriele Dragotto
Princeton University
gdragotto@princeton.edu

MS123

A New Dual-Based Cutting Plane Algorithm for Nonlinear Adjustable Robust Optimization

This talk explores a class of nonlinear Adjustable Robust Optimization (ARO) problems, containing here-and-now and wait-and-see variables, with uncertainty in the objective function and constraints. By applying Fenchels duality on the wait-and-see variables, we obtain an equivalent dual reformulation, which is a nonlinear static robust optimization problem. Using the dual formulation, we provide conditions under which the ARO problem is convex on the here-and-now decision. Furthermore, since the dual formulation contains a non-concave maximization on the uncertain parameter, we use perspective relaxation and an alternating method to handle the non-concavity. By employing the perspective relaxation, we obtain an upper bound, which we show is the same as the static relaxation of the considered problem. Moreover, invoking the alternating method, we design a new dual-based cutting plane algorithm that is able to find a reasonable lower bound for the optimal objective value of the considered nonlinear ARO model. The numerical experiments reveal the abilities of our cutting plane algorithm in producing locally robust solutions with an acceptable optimality gap.

Abbas Khademi
Tehran University
abbaskhademi@ut.ac.ir

Ahmadreza Marandi
Tilburg University
a.marandi@tue.nl

Majid Soleimani
Tehran University
majid soleimani damaneh ¡soleimani@khayam.ut.ac.ir

MS124

Post-Processing Algorithm with Chubanov Method for Symmetric Cone Programs

We propose a new post-processing algorithm for symmetric cone programs based on Chubanov's method. Kanoh & Yoshise (2021) reported that their variant of Chubanov's method could find more accurate solutions to strongly feasible ill-conditioned problems, i.e., problems with feasible solutions only near the boundaries of the cone, than the interior point method. Our algorithm is devised to return a more accurate solution using the output solution from the interior point method and the variant of Chuabonvs method proposed by Kanoh & Yoshise (2021). We conducted numerical experiments with SDPLIB instances. The results show that our algorithm outputs a solution with a very small value of DIMACS error, the measure of accuracy as an optimal solution, compared to the solution that the interior point method returned.

Shinichi Kanoh
University of Tsukuba
kanoh-sin-one@outlook.jp

Akiko Yoshise
University of Tsukuba
Graduate School of Systems and Information Engineering
yoshise@sk.tsukuba.ac.jp

MS124

Interior Point Methods for Nonlinear Optimization on Riemannian Manifolds

We extend the classical primal-dual interior point methods from the Euclidean setting to the Riemannian one. Our method, named the Riemannian interior point method (RIPM), is for solving Riemannian constrained optimization problems. Under the standard assumptions, we establish locally superlinear and quadratic convergence of RIPM. Moreover, we show the global convergence of RIPM with classical line search. These are generalizations of the classical frame of primal-dual interior point methods for nonlinear programming proposed by El-Bakry et al. in 1996.

Zhijian Lai, Akiko Yoshise
University of Tsukuba
Graduate School of Systems and Information Engineering
s2130117@s.tsukuba.ac.jp, yoshise@sk.tsukuba.ac.jp

MS124

An Efficient Dual SPG Method for Log-Det SDP with Hidden Clustering Structure

We propose an extension of the DSPG algorithm for efficiently solving the dual problem of a sparse Gaussian graphical model with a hidden clustering structure proposed by Lin et al. (2020). By exploiting the projection of the new variable based on the properties of the complicated constraints, we prove that the proposed method can generate a sequence that converges to an optimal solution or find an optimal solution in a finite number of iterations. Furthermore, we will also show that each iteration of the proposed method can be computed at a lower cost than applying the DSPG algorithm directly, by using the pool-adjacent-violators (PAV) algorithm. Lastly, we discuss about the result of numerical experiments on synthetic data.

Charles Namchaisiri, Liu Tianxiang
Tokyo Institute of Technology
namchaisiri.c.aa@m.titech.ac.jp, liu.t.af@m.titech.ac.jp

Makoto Yamashita
Tokyo Institute of Technology
Dept. of Mathematical and Computing Science
makoto.yamashita@c.titech.ac.jp

MS125

Accelerated Physical Models for PDE-Constrained Inverse Design

Inverse design of large-scale structures and of average properties are pervasive challenges in engineering, in particular in optics, because they involve optimizing high-dimensional non-convex objective functions where a single function evaluation uses significant time and computing resources. These challenges limit the number and scope of such designs. In this talk, we present methodologies to accelerate PDE-constrained inverse design and break these limits. We highlight two levels of acceleration. The first relies on physical knowledge to build models that exploit approximations or problem formulations for efficiency. The second relies on scientific machine learning surrogates to speed up reusable building blocks by orders of magnitude. We will illustrate these methods on large-scale metasurface designs for incoherent light and polychromatic applications.

Raphaël Pestourie
Massachusetts Institute of Technology
raphael.pestourie@gmail.com

Youssef Mroueh
IBM Research
mroueh@us.ibm.com

Zhaoyi Li
Harvard University
zhaoyili@g.harvard.edu

Das Payel
IBM Research
daspa@us.ibm.com

Boubacar Kanté
UC Berkeley
bkante@berkeley.edu

Federico Capasso
Harvard University
capasso@fas.harvard.edu

Steven Johnson
MIT
stevenj@math.mit.edu

MS126

Communication Acceleration of Local Gradient Methods Via An Accelerated Primal-Dual Algorithm with Inexact Prox

Inspired by a recent breakthrough of Mishchenko et al. [2022], who for the first time showed that local gradient steps can lead to provable communication acceleration, we propose an alternative algorithm which obtains the same communication acceleration as their method (ProxSkip). Our approach is very different, however: it is based on the celebrated method of Chambolle and Pock [2011], with several nontrivial modifications: i) we allow for an inexact computation of the prox operator of a certain smooth strongly convex function via a suitable gradient-based method (e.g., GD or Fast GD), ii) we perform a careful modification of the dual update step in order to retain linear convergence. Our general results offer the new state-of-the-art rates for the class of strongly convex-concave saddle-point problems with bilinear coupling characterized by the absence of smoothness in the dual function. When applied to federated learning, we obtain a theoretically better alternative to ProxSkip: our method requires fewer local steps $\mathcal{O}(\kappa^{1/3})$ or $\mathcal{O}(\kappa^{1/4})$, compared to

$\mathcal{O}(\kappa^{1/2})$of ProxSkip), and performs a deterministic number of local steps instead. Like ProxSkip, our method can be applied to optimization over a connected network, and we obtain theoretical improvements here as well.

Abdurakhmon Sadiev
KAUST
Thuwal, Saudi Arabia
abdurakhmon.sadiev@kaust.edu.sa

Peter Richtárik
King Abdullah University of Science and Technology
peter.richtarik@kaust.edu.sa

Dmitry Kovalev
Université catholique de Louvain
dakovalev1@gmail.com

MS126

A Regularized Variance-Reduced Modified Extragradient Method for Stochastic Hierarchical Games

The theory of learning in games has so far focused mainly on games with simultaneous moves. Recently, researchers in machine learning have started investigating learning dynamics in games involving hierarchical decision-making. We consider an N-player hierarchical game in which the i-th player's objective comprises of an expectation-valued term, parametrized by rival decisions, and a hierarchical term. Such a framework allows for capturing a broad range of stochastic hierarchical optimization problems, Stackelberg equilibrium problems, and leader-follower games. We develop an iteratively regularized and smoothed variance-reduced modified extragradient framework for learning hierarchical equilibria in a stochastic setting. We equip our analysis with rate statements, complexity guarantees, and almost-sure convergence claims. We then extend these statements to settings where the lower-level problem is solved inexactly and provide the corresponding rate and complexity statements.

Mathias Staudigl
Maastricht University
Mathias.staudigl@gmail.com

Shishing Cui
Beijing Institute of Technology
cuishisheng05@gmail.com

Uday V. Shanbhag
Pennsylvania State University
udaybag@psu.edu

MS126

On the Improved Conditions for Some Primal-Dual Algorithms

The convex minimization of $f(x) + g(x) + h(Ax)$ over R^n with differentiable f and linear operator A has been well-studied in the literature. By considering the primal-dual optimality of the problem, many algorithms are proposed from different perspectives, such as monotone operator schemes and fixed point theory. In this talk, we show that the convergence conditions for some primal-dual algorithms can be relaxed, and some relaxations are optimal. Specifically, we show that the product of primal and dual stepsizes for PAPC and Chambool-Pock can be increased up to 1/3, and this relaxation is optimal. We will also show relaxed conditions for other primal-dual algorithms.

Ming Yan
The Chinese University of Hong Kong
Shenzhen, China
yanming@cuhk.edu.cn

MS127

Single-Call Stochastic Extragradient Methods: Improved Analysis under Weaker Conditions

Single-call stochastic extragradient methods, like stochastic past extragradient SPEG and stochastic optimistic gradient SOG, have gained much interest in recent years and are one of the most efficient algorithms for solving large-scale min-max optimization and variational inequalities problems (VIP). It has found numerous applications in generative adversarial networks (GAN), reinforcement learning, and robust learning. However, despite their undoubted popularity, current convergence analyses of SPEG and SOG require a bounded variance assumption. Moreover, several important questions regarding the convergence properties of these methods are still open, including mini-batching, efficient step-size selection, and convergence guarantees under different sampling strategies. In this talk, we address these questions and provide convergence guarantees for two large classes of structured non-monotone VIPs: (i) quasi-strongly monotone problems (a generalization of strongly monotone problems) and (ii) Minty variational inequalities (a generalization of monotone problems). We introduce the expected residual condition and explain its benefits over previously used growth conditions, expected co-coercivity, or bounded variance assumptions. Equipped with this condition, we provide theoretical guarantees for the convergence of single-call extragradient methods for different step-size selections, including constant, decreasing, and stepsize-switching rules.

Sayantan Choudhury
Johns Hopkins University
schoudh8@jhu.edu

Eduard Gorbunov
MBZUAI/Université de Montréal
edgorbunov73@gmail.com

Nicolas Loizou
Johns Hopkins University
nloizou@jhu.edu

MS127

Advances in Differentially Private Stochastic Saddle Points

Motivated by various data analysis problems, including synthetic data generation, group fairness in statistical learning, and other equilibrium problems, the approximation of saddle points under differential privacy has received recent attention. In this talk, I will introduce this problem as well as the main known convergence rates, and show how to exploit geometric structure to obtain faster rates.

Cristobal Guzman
PUC-Chile
crguzmanp@mat.uc.cl

Tomás González
Pontificia Universidad Católica de Chile
tsgonzalez@uc.cl

Fabian Pedregosa
Google Research, Brain Team
pedregosa@google.com

Courtney Paquette
McGill University and Google Brain Team
cypaquette@google.com

MS127

On the Optimization of Variational Inequalities: Insights from Continuous-Time

Solving for equilibria is an underlying problem of many instances ranging from robustifications of standard minimization, GANs, and MARL, to potential applications in economics, social or market dynamics, and finance. Variational Inequalities (VIs) formalize this general problem class. While different methods for solving VIs perform differently in various setups, these have an identical continuous-time ordinary differential equation (ODE) when derived naively. Thus, the ODE perspective—which is proved powerful in analyzing optimization methods—can not be used to address the central challenges of solving VIs such as convergence to spurious limit cycles. We propose a principled way of deriving High-Resolution Differential Equations (HRDEs) that generalizes the standard ODE derivation and is similar to a framework studied in fluid dynamics. We derive the HRDEs for several widely used VI methods. While the HRDEs differ among most of the methods, they yet reveal some non-evident connections and give a novel interpretation of the optimistic gradient descent method. Using tools from dynamical systems, we show that their convergence can easily be verified for bilinear games. We show that the HRDE of Optimistic Gradient Descent Ascent (OGDA) exhibits last-iterate convergence for general monotone variational inequalities. Finally, we provide rates of convergence for the best iterate of the discrete OGDA method, relying solely on the first-order smoothness of the monotone operator.

Tatjana Chavdarova
University of California, Berkeley
tatjana.chavdarova@berkeley.edu

Emmanouil Zampetakis
UC Berkeley
mzampet@berkeley.edu

MS128

Structure Learning with Continuous Unconstrained Optimization

Structure learning of directed acyclic graphs (DAGs) is a challenging task partly owing to the large search space of possible graphs. Recently, continuous optimization approaches for structure learning have received considerable attention, which typically solve a continuous constrained optimization problem with an algebraic characterization of DAG constraint. In this talk, we discuss how the convergence property of existing formulation with hard DAG constraint may lead to optimization difficulties such as ill-conditioning issue. We then introduce an unconstrained formulation based on soft sparsity and DAG constraints, which is guaranteed to learn a DAG equivalent to the ground truth under mild assumptions.

Ignavier Ng
Carnegie Mellon University
ignavierng@cmu.edu

MS128

Learning Latent Causal Graphs Via Mixture Oracles

We study the problem of reconstructing a causal graphical model from data in the presence of latent variables. The main problem of interest is recovering the causal structure over the latent variables while allowing for general, potentially nonlinear dependencies. In many practical problems, the dependence between raw observations (e.g. pixels in an image) is much less relevant than the dependence between certain high-level, latent features (e.g. concepts or objects), and this is the setting of interest. We provide conditions under which both the latent representations and the underlying latent causal model are identifiable by a reduction to a mixture oracle. These results highlight an intriguing connection between the well-studied problem of learning the order of a mixture model and the problem of learning the bipartite structure between observables and unobservables. The proof is constructive, and leads to several algorithms for explicitly reconstructing the full graphical model. We discuss efficient algorithms and provide experiments illustrating the algorithms in practice.

Pradeep Ravikumar
Carnegie Mellon
pkr@andrew.cmu.edu

MS128

Optimization-Based Causal Estimation from Heterogenous Environments

We present an optimization approach to causal estimation. Given data that contains covariates and an outcome, which covariates are causes of the outcome, and what is the strength of the causality? In classical machine learning (ML), the goal of optimization is to maximize predictive accuracy. However, some covariates might exhibit a non-causal association with the outcome. Such spurious associations provide predictive power for classical ML, but they prevent us from causally interpreting the result. This paper proposes CoCo, an optimization algorithm that bridges the gap between pure prediction and causal inference. CoCo leverages the recently-proposed idea of environments, datasets of covariates/responses where the causal relationships remain invariant but where the distribution of the covariates changes from environment to environment. Given datasets from multiple environmentsand ones that exhibit sufficient heterogeneityCoCo maximizes an objective for which the only solution is the causal solution. We describe the theoretical foundations of this approach and demonstrate its effectiveness on simulated and real datasets. Compared to classical ML and existing methods, CoCo provides more accurate estimates of the causal model.

Mingzhang Yin
University of Florida
mingzhang.yin@warrington.ufl.edu

Yixin Wang
University of Michigan
yixinw@umich.edu

David Blei
Columbia University
david.blei@columbia.edu

MS129

Alternating Iteratively Reweighted Least Squares Algorithms for Matrix and Tensor Decomposition

Over the last few years Iterative reweighted least square (IRLS) algorithms have been widely applied in sparse recovery and low-rank matrix estimation problems providing computational efficiency and fast convergence rates. In this talk, I will focus on iterative reweighed least squares algorithms for alternating minimization problems. First, I will present an alternating IRLS algorithm for the low-rank matrix factorization problem via minimizing a variational form of the Schatten-p quasi-norm. The so-called alternating IRLS (AIRLS) algorithm consists of closed-form computationally efficient updates for matrix factors. The proposed AIRLS algorithm converges fast and provides rank-revealing matrix factorizations. Second, I will extend these ideas to a tensor decomposition problem. Specifically, I will focus on the recently proposed block-term tensor (BTD) decomposition model. I will present a hierarchical IRLS (HIRLS) algorithm, which allows us to simultaneously reveal the number of BTD terms and the individual ranks of the BTD factors by promoting column-sparsity in a hierarchical fashion. Finally, I will present simulated and real-data experimental results showing the merits of the proposed IRLS algorithms and their merits in matrix/tensor completion and hyperspectral image denoising problems.

Paris Giampouras
Johns Hopkins University
parisg@jhu.edu

MS129

Iteratively Reweighted Least Squares Recovery on Tensor Tree Networks

One fundamental approach to matrix recovery, being a predecessor to tensor recovery, traces back to the affine rank minimization problem. While there are various surrogate approaches within that setting, we emphasize here that the *asymptotic minimization* of the well-known, so called *log-det* objective functions always yields the desired, minimal rank matrices within the given, affine set; whereas such may or may not recover an a-priorly sought for ground truth. Concerning the commonly applied method of iteratively reweighted least squares (IRLS-0), one thus remains with two concerns. How problematic are local minima inherent to the log-det approach truly; and opposingly, how influential instead is the numerical realization. With higher dimensions in mind, based on the concept of matricization, affine sum-of-ranks minimization then generalizes the setting from matrices to tensors. While convergence properties are directly transferable, we demonstrate that in numerical experiments, the corresponding IRLS-0 method can be exhausted in order to observe the theoretical phase transition for generic tensor recoverability. In large-scale applications in turn, alternating, reweighted optimization on tensor tree networks allows to avoid exponential computational complexity, without substantial loss of approximation quality.

Sebastian Kraemer
RWTH Aachen
kraemer@igpm.rwth-aachen.de

MS129

The Flip Side of the Reweighted Coin: Duality of Adaptive Dropout and Regularization

Among the most successful methods for sparsifying deep neural networks are those that adaptively mask the network weights throughout training. By examining this masking, or dropout, in the linear case, we uncover a duality between such adaptive methods and regularization through the so-called "η-trick" that casts both as iteratively reweighted optimizations. We show that any dropout strategy that adapts to the weights in a monotonic way corresponds to an effective subquadratic regularization penalty, and therefore typically leads to sparse solutions. We obtain the effective penalties for several popular sparsification strategies, which are remarkably similar to classical penalties commonly used in sparse optimization.

Daniel LeJeune
Stanford University
daniel@dlej.net

MS130

The Cost of Nonconvexity in Deterministic Nonsmooth Optimization

We study how nonconvexity affects the difficulty of nonsmooth optimization. A 2020 breakthrough by Zhang et al. analyzed the complexity of an intuitively-appealing randomized algorithm. Despite negative results in general (Jordan et al.), a simple deterministic version succeeds on difference-of-convex objectives, with complexity related to negative ingredients in the objective's distributional second derivative.

Siyu Kong
Columbia University
sk3333@cornell.edu

Adrian Lewis
Cornell University
ORIE
adrian.lewis@cornell.edu

MS130

Global stability of first-order methods with constant step size for coercive tame functions

We consider first-order methods with constant step size for minimizing locally Lipschitz coercive functions that are tame in o-minimal structures on the real field. We prove that if the method is approximated by parametrized subgradient trajectories, then the iterates eventually remain in a neighborhood of a connected component of the set of critical points. Under suitable regularity assumptions, this result applies to the subgradient method with momentum, the stochastic subgradient method with random reshuffling, the incremental gradient method with momentum, and the random-permutations cyclic coordinate descent method.

Cedric Josz, Lexiao Lai
Columbia University
cj2638@columbia.edu, ll3352@columbia.edu

MS130

Error Bounds for the Generalized Power Cone and Applications in Algebraic Structure

Error bounds are a requisite for trusting or distrusting solutions in an informed way, which are difficult to obtain in the absence of constraint qualifications and known succinct expressions of projections. As an example of "nonsymmetric cones", the generalized power cone has found its applications in many areas and attracted more and more attention. In this talk, using a recently developed framework of one-step facial residual functions, we build error bounds for the generalized power cones without any constraint qualifications. We also show that our error bounds are tight in the sense of that framework. Besides their utility for understanding solution reliability, the error bounds we discover have additional applications to the algebraic structure of the underlying cone. In particular, we use the error bounds and facial structures of the generalized power cone to establish the automorphism group of the generalized power cone and compute its dimension, which are both unknown before our work. With this, we identify a set of generalized power cones that are self-dual, irreducible, nonhomogeneous, and perfect.

Ying Lin
The Hong Kong Polytechnic University
ying.lin@connect.polyu.hk

Scott B. Lindstrom
Centre for Optimisation and Decision Science
Curtin University
scott.lindstrom@curtin.edu.au

Bruno Lourenço
The Institute of Statistical Mathematics
bruno@ism.ac.jp

Ting Kei Pong
Department of Applied Mathematics
The Hong Kong Polytechnic University, Hong Kong
tk.pong@polyu.edu.hk

MS131

High-Dimensional Statistical Estimation via Sum-of-Squares

One exciting new development of the past decade is the evolution of the sum-of-squares method for algorithm design for high-dimensional statistical estimation. This paradigm can be viewed as a principled approach to generating and analyzing semidefinite programming relaxations for statistical estimation problems by thinking of the duals as proofs of statistical identifiability – i.e., proof that the input data uniquely identifies the unknown target parameters. I will give an overview of the sum-of-squares method for statistical estimation. Specifically, I will discuss how strengthening (via semidefinite certificates) of basic analytic properties of probability distributions such as subgaussian tails, hypercontractive moments, and anti-concentration resolve the central open question in algorithmic robust statistics of robustly learning a high-dimensional mixture of Gaussians.

Pravesh Kothari
CMU
praveshk@cs.cmu.edu

MS131

Geometric random walks and sampling from spectrahedra

We present algorithmic, complexity, and implementation results on the problem of sampling points from polytopes (the feasible regions of linear programs) and spectrahedra (the feasible region of a semidefinite programs). To sample from log-concave distributions, our main tool is geometric random walks that we realize using primitive geometric operations, that in turn exploit the algebraic and geometric properties of convex bodies and, in particular, the (polynomial) eigenvalue problem. We demonstrate how we can use sampling to compute volumes and solve semidefinite programs. If time permits, we also present applications in systems biology and computational finance. The talk is based on joint works with Apostolos Chalkis and Vissarion Fisikopoulos.

Elias Tsigaridas
Inria Paris
elias.tsigaridas@inria.fr

MS131

Data-scarce identification of learning dynamics via sum-of-squares optimization

Understanding how populations of infinitesimal anonymous agents learn to adapt their behavior over time as a result of repeated strategic interactions is a fundamental problem in science and engineering. In this work, we develop a theoretical and algorithmic framework for identifying the dynamics that govern agent behavior using only samples from a short run of a single system trajectory. In contrast to the modern data-driven paradigm for model discovery, our framework is applicable to settings where observational data is scarce. Our computational approach uses sum-of-squares optimization to solve a side-information-assisted polynomial regression problem, where we compensate for the absence of data by incorporating side-information constraints modeling a wide range of agent behaviors. Our experimental results demonstrate that the dynamics recovered by our method are indistinguishable from the ground truth dynamics with respect to important benchmarks, including the equilibrium selection problem in congestion games and the exact identification of chaotic dynamics. Based on joint work with Georgios Pilliouras and Joseph Sakos.

Antonios Varvitsiotis
Singapore University of Technology and Design
avarvits@gmail.com

MS132

Distributionally Robust Offline Reinforcement Learning with Optimal Transport Distance

For offline reinforcement learning, the environment that generates the offline data general differs from the environment that the learned policy deploys. To hedge against such distributional shift, we adapt distributionally robust offline reinforcement learning. To mitigate the unseen worst cases, we adopt Wasserstein and Sinkhorn distance instead of the f-divergence to construct the uncertainty set. We further propose to combine Q learning with efficient gradient-based methods to solve the problem.

Yifan Hu
EPFL
yifan.hu@epfl.ch

Andreas Krause
ETH Zurich
krausea@ethz.ch

Daniel Kuhn
École Polytechnique Fédérale de Lausanne
daniel.kuhn@epfl.ch

MS132

Bayesian Risk Markov Decision Processes

We consider finite-horizon Markov Decision Processes where parameters, such as transition probabilities, are unknown and estimated from data. The popular distributionally robust approach to addressing the parameter uncertainty can sometimes be overly conservative. In this paper, we propose a new formulation, Bayesian risk Markov Decision Process (BR-MDP), to address parameter uncertainty in MDPs, where a risk functional is applied in nested form to the expected total cost with respect to the Bayesian posterior distribution of the unknown parameters. The proposed formulation provides more flexible risk attitudes towards parameter uncertainty and takes into account the availability of data in future times stages. To solve the proposed formulation with the conditional value-at-risk (CVaR) risk functional, we propose an efficient approximation algorithm by deriving an analytical approximation of the value function and utilizing the convexity of CVaR. We demonstrate the empirical performance of the BR-MDP formulation and proposed algorithms on a gambler's betting problem and an inventory control problem.

Yifan Lin, Yuxuan Ren, Enlu Zhou
Georgia Institute of Technology
yifanlin@gatech.edu, yren79@gatech.edu,
enlu.zhou@isye.gatech.edu

MS132

Finite Sample Guarantee for Distributionally Robust Q Learning

We consider a reinforcement learning setting in which the deployment environment is different from the training environment. Applying a robust Markov decision processes formulation, we extend the distributionally robust Q-learning framework studied in Liu et. al. 2021. Further, we improve the design and analysis of their multi-level Monte Carlo estimator. Assuming access to a simulator, we prove that the worst-case expected sample complexity of our algorithm to learn the optimal robust Q-function within an ϵ error in the sup norm is upper bounded by $\tilde{O}(|S||A|(1-\gamma)^{-5}\epsilon^{-2})$. This is the first sample complexity result for the model-free robust RL problem. Simulation studies further validate our theoretical results.

Shengbo Wang
Stanford University
shengbo.wang@stanford.edu

MS133

Joint Assortment and Inventory Planning Under Markov Chain Choice Model

We study a joint assortment and inventory optimization problem faced by an online retailer who needs to decide on both the assortment and the inventories of a set of N substitutable products before the start of a selling season of length T to maximize the expected profit. We model dynamic stock-out based substitution by proposing a Markov Chain choice model and present a near-optimal algorithm for the problem achieving a O(v(NT)) regret with respect to an LP upper bound. Our algorithm balances between expected revenue and inventory costs by identifying a subset of products that can pool demand without significantly cannibalizing the revenue in the presence of dynamic substitution. We conduct computational experiments that show that our algorithm empirically outperforms natural approaches both on synthetic and realistic instances.

Omar El Housni
Cornell University
oe46@cornell.edu

Omar Mouchtaki
Columbia University
om2316@columbia.edu

Guillermo Gallego

HKUST
ggallego@ust.hk

Vineet Goyal
Columbia University
vgoyal@ieor.columbia.edu

Salal Humair
Amazon
salal@amazon.com

Sangjo Kim
SUFE
sangjo@sufe.edu.cn

Ali Sadighian
Uber
ali.sadighian@gmail.com

Jingchen Wu
Amazon
jcwu@amazon.com

MS133

Fulfillment Decisions And Route Selection For Online Retailing

Order fulfillment for online retailing includes deciding from where to source the inventory for each order and how to route the order from the source to the customer. This research focuses on the routing decision for a context in which the retailer has multiple routing options. These options can include third party carriers, as well as the retailer's own transportation resources, which may be capacity constrained. To make these decisions in real time, we propose computing shadow prices on the constrained transportation resources, in order to have a way to compare the (imputed) cost of the routing options. We illustrate the approach and the trade-offs with a simple fulfillment network and then report on results from a more extensive computational experiment with a larger fulfillment network. This research is joint with Pinyi Chen from MIT, and with industry collaborators.

Stephen Graves
Massachusetts Institute of Technology
sgraves@mit.edu

MS133

The Benefits of Delay to Online Decision-Making

Real-time decisions are usually irrevocable in many contexts of online decision-making. One common practice is delaying real-time decisions so that the decision-maker can gather more information to make better decisions (for example, in online retailing, there is typically a time delay between when an online order is received and when it gets picked and assembled for shipping). However, decisions cannot be delayed forever. In this paper, we study this fundamental trade-off and aim to theoretically characterize the benefits of delaying real-time decisions. We provide such a theoretical foundation for a broad family of online decision- making problems by proving that the gap between the proposed online algorithm with delay and the offline optimal hindsight policy decays exponentially fast in the length of delay. We also conduct extensive numerical experiments on the benefits of delay, using both synthetic and real data that is publicly available. Both our theoretical and empirical results suggest that a little delay is all we need.

Yaqi Xie
University of Chicago Booth School of Business
yaqi.xie@chicagobooth.edu

Will Ma
Columbia University
wm2428@gsb.columbia.edu

Linwei Xin
University of Chicago Booth School of Business
linwei.xin@chicagobooth.edu

MS134

A Superlinearly Convergent Subgradient Method for Sharp Semismooth Problems

Subgradient methods comprise a fundamental class of nonsmooth optimization algorithms. Classical results show that certain subgradient methods converge sublinearly for general Lipschitz convex functions and converge linearly for convex functions that grow sharply away from solutions. Recent work has moreover extended these results to certain nonconvex problems. In this work we seek to improve the complexity of these algorithms, asking: is it possible to design a superlinearly convergent subgradient method? We provide a positive answer to this question for a broad class of sharp semismooth functions.

Vasileios Charisopoulos
Cornell University
vc333@cornell.edu

Damek Davis
Cornell University
ORIE
dsd95@cornell.edu

MS134

Lifted Primal-Dual Method for Bilinearly Coupled Smooth Minimax Optimization

We study the bilinearly coupled minimax problem: $\min_x \max_y f(x) + y^\top Ax - h(y)$, where f and h are both strongly convex smooth functions and admit first-order gradient oracles. Surprisingly, no known first-order algorithms have hitherto achieved the lower complexity bound of $\Omega((\sqrt{\frac{L_x}{\mu_x}} + \frac{\|A\|}{\sqrt{\mu_x \mu_y}} + \sqrt{\frac{L_y}{\mu_y}}) \log(\frac{1}{\varepsilon}))$ for solving this problem up to an ε primal-dual gap in the general parameter regime, where L_x, L_y, μ_x, μ_y are the corresponding smoothness and strongly convexity constants. We close this gap by devising the first optimal algorithm, the Lifted Primal-Dual (LPD) method. Our method lifts the objective into an extended form that allows both the smooth terms and the bilinear term to be handled optimally and seamlessly with the same primal-dual framework. Besides optimality, our method yields a desirably simple single-loop algorithm that uses only one gradient oracle call per iteration. Moreover, when f is just convex, the same algorithm applied to a smoothed objective achieves the nearly optimal iteration complexity. We also provide a direct single-loop algorithm, using the LPD method, that achieves the iteration complexity of $O(\sqrt{\frac{L_x}{\varepsilon}} + \frac{\|A\|}{\sqrt{\mu_y \varepsilon}} + \sqrt{\frac{L_y}{\varepsilon}})$.

Kiran K. Thekumparampil
Amazon
kirankoshy@gmail.com

Niao He
ETH Zurich
niao.he@inf.ethz.ch

Sewoong Oh
University of Washington
sewoong@cs.washington.edu

MS134

Gradient-Free Methods for Deterministic and Stochastic Nonsmooth Nonconvex Optimization

Nonsmooth nonconvex optimization problems broadly emerge in machine learning and business decision making, whereas two core challenges impede the development of efficient solution methods with finite-time convergence guarantee: the lack of computationally tractable optimality criterion and the lack of computationally powerful oracles. The contributions of this paper are two-fold. First, we establish the relationship between the celebrated Goldstein subdifferential and uniform smoothing, thereby providing the basis and intuition for the design of gradient-free methods that guarantee the finite-time convergence to a set of Goldstein stationary points. Second, we propose the gradient-free method (GFM) and stochastic GFM for solving a class of nonsmooth nonconvex optimization problems and prove that both of them can return a (δ, ϵ)-Goldstein stationary point of a Lipschitz function f at an expected convergence rate at $O(d^{3/2}\delta^{-1}\epsilon^{-4})$ where d is the problem dimension. Two-phase versions of GFM and SGFM are also proposed and proven to achieve improved large-deviation results. Finally, we demonstrate the effectiveness of 2-SGFM on training ReLU neural networks with the Minst dataset.

Zeyu Zheng
University of California, Berkeley
zyzheng@berkeley.edu

Tianyi Lin
UC Berkeley
darren_lin@berkeley.edu

Michael I. Jordan
University of California, Berkeley
jordan@cs.berkeley.edu

MS135

Adaptively Optimizing Experiments: Past, Present, and Future

Using data to select, calibrate, and validate mathematical models is a concerstone of science and engineering. Yet for many partitioners rely on classical “black-box” design of experiments (DoE) methods such as partial factorial or response surface to postulate data collection campaigns. These methods, while popular, are cumbersome and data inefficient. In this talk, we survey recent advances in adaptive design of experiments and discuss their relationship to dynamic optimization and optimal control. Using several engineering examples, we argue optimization-based methods such as model-based design of experiments (MBDoE) and Bayesian optimization (BO) with Gaussian process regression (GPR) surrogates are more data efficient than classical DoE. Finally, we share our recent software package, Pyomo.DoE (Wang & Dowling, AIChE Journal, 2022, https://doi.org/10.1002/aic.17813), that facilitates MBDoE for large-scale partial differential-algebraic equation systems defined in the Python/Pyomo modeling environment.

Alexander Dowling, Jialu Wang, Kyla D. Jones
University of Notre Dame
adowling@nd.edu, jwang44@nd.edu, kjones29@nd.edu

Ke Wang
University of Notre-Dame
kwang5@nd.edu

MS135

Collaborative Decision-Making under Adversarial and Information Constraints

Two major challenges in modern distributed computing paradigms such as federated learning are communication-efficiency and robustness to adversarial agents. While several recent works have explored these themes for supervised learning, little to nothing is known about sequential decision-making (e.g., reinforcement learning and bandits) in this regard. Motivated by this gap, we initiate the study of these challenges by focusing on the linear stochastic bandit formalism. First, we consider a collaborative linear bandit setup where a fraction of the agents are adversarial. For this setting, we design new algorithms that balance the exploration-exploitation trade-off via carefully constructed robust confidence intervals. When the fraction of adversaries is small, we show that our proposed algorithms provably benefit from collaboration, despite arbitrary adversarial behavior. We also prove a fundamental lower bound, revealing that our regret guarantees are minimax optimal in all parameters. Next, we formulate a novel linear bandit problem involving a communication channel of finite capacity. We develop an adaptive encoding mechanism that exploits statistical concentration bounds, and prove that for a d dimensional model, a channel capacity of $O(d)$ bits suffices to achieve order-optimal regret. Overall, our work takes a significant step towards paving the way for multi-agent sequential decision-making subject to adversarial feedback and information constraints.

Aritra Mitra
University of Pennsylvania

amitra20@seas.upenn.edu

MS135

Nonconvex Constrained Distributed Optimization for Plantwide Model Predictive Control

Model predictive control (MPC) on a plantwide scale involves large optimization problems. For flexibility as well as computational speed, distributed optimization algorithms can be used, which decomposes large-scale systems into subsystems and coordinate the subsystems by communicating their primal or primal-dual values. In the presence of nonconvex constraints, however, the large number of iterations in distributed algorithms makes them prohibitive to implement in real time. Here we discuss several algorithms, including two from our research, for distributed plantwide MPC. (1) ELLADA, a two-layer iterative algorithm with two modifications. First, the tolerances of the subproblems are made adaptive to the progress of the iterations, thus saving computation by inexact solutions in intermediate iterations. Second, an Anderson acceleration scheme is applied on inner iterations, which practically accelerates the convergence with a pseudo-Newtonian construction. (2) The Lyapunov envelope approach, a primal algorithm under a 1- and 2-norm regularized objective. The regularization terms are based on the incrementally dissipative properties of the subsystems, thus retaining closed-loop stability even when the iterations are early-terminated prior to convergence. We also discuss potential future directions, focusing on integrating distributed optimization algorithms with automatic decomposition methods of large systems based on network analysis.

Wentao Tang
North Carolina State University
wtang23@ncsu.edu

Prodromos Daoutidis
University of Minnesota
Dept. of Chemical Engineering and Materials Science
daout001@umn.edu

MS136

Infinite Quantum Signal Processing

The quantum singular value transformation (QSVT) [Gilyen, Su, Low, Wiebe, STOC 2019] provides a unified viewpoint of a large class of practically useful quantum algorithms. At the heart of QSVT is a new polynomial representation, called quantum signal processing (QSP). QSP represents a degree-d polynomial using products of matrices in SU(2), parameterized by (d+1) real numbers called the phase factors. When the polynomial of interest is obtained by truncating an infinite polynomial series, a natural question is whether the phase factors have a well defined limit as $d \to \infty$. In this talk, we will show that there exists a consistent choice of the parameterization so that the limit is well defined. This generalizes QSP to represent a large class of non-polynomial functions, and this construction is referred to as infinite QSP (iQSP). We present a very simple algorithm for finding such infinitely long phase factors with provable performance guarantees. We will also show a surprising connection between the regularity of the target function and the structural properties of the phase factors.

Yulong Dong
University of California, Berkeley
dongyl@berkeley.edu

Lin Lin
University of California, Berkeley
Lawrence Berkeley National Laboratory
linlin@math.berkeley.edu

Hongkang Ni
Stanford University, U.S.
hongkang@stanford.edu

Jiasu Wang
University of California, Berkeley
jiasu@berkeley.edu

MS136

Quantum Hamiltonian Descent

Gradient descent is a fundamental algorithm in the theory and practice of continuous optimization. Identifying its counterpart in quantum computing is not only theoretically interesting but also could lead to practical quantum algorithms. A conventional approach to quantum speedups in optimization relies on the quantum acceleration of intermediate steps of corresponding classical algorithms while keeping the quality of their solutions. We propose Quantum Hamiltonian Descent (QHD) as a truly quantum counterpart of classical gradient methods, derived from the path integral of classical dynamical systems corresponding to the continuous-time limit of classical gradient descent algorithms. We establish QHDs convergence to the global optimum in both convex and non-convex settings. More importantly, QHD is described as a time-dependent Hamiltonian evolution that can be efficiently simulated on both digital and analog quantum computers. By embedding QHD into the so-called Quantum Ising Machine (including D-Wave and others), we empirically observe that the D-Wave-implemented QHD outperforms a selection of state-of-the-art gradient-based classical solvers and the standard quantum adiabatic algorithm, based on the time-to-solution metric, on non-convex quadratic programming instances up to 75 dimensions. Finally, we propose a three-phase picture to explain the behavior of QHD, especially its difference from the quantum adiabatic algorithm.

Jiaqi Leng, Ethan Hickman, Joseph Li, Xiaodi Wu
University of Maryland, College Park, USA
jiaqil@umd.edu, ethanh@umd.edu, jli0108@umd.edu, xiaodiwu@umd.edu

MS136

The Design of Analog Quantum Algorithms

The conventional design of quantum algorithms is centered around the abstraction of quantum circuits and relies on a digital mindset for algorithmic applications. We propose another paradigm where the design of quantum algorithms is directly built on the abstraction of continuous-time Hamiltonian evolution. Namely, we treat quantum machines as programmable and efficient solvers for the Schrodinger equation and directly develop applications on top of that. We showcase that a few leading quantum applications (e.g., continuous-time quantum walk and quantum PDE solvers) can be re-derived in this way, which share the same features (e.g., efficiency, input/output assumptions) as their digital counterparts but could be likely implemented on near-term analog quantum simulators with minimal overheads.

Xiaodi Wu, Jiaqi Leng, Joseph Li
University of Maryland, College Park, USA
xwu@cs.umd.edu, jiaqil@umd.edu, jli0108@umd.edu

MS137

Algorithms for Deterministically Constrained Stochastic Optimization

In this talk, we discuss stochastic optimization algorithms for deterministically constrained problems based on the sequential quadratic optimization (commonly known as SQP) methodology. We present the rationale behind our proposed techniques, convergence in expectation and complexity guarantees for our algorithms, and the results of preliminary numerical experiments that we have performed.

Albert S. Berahas
Department of Industrial and Operations Engineering
University of Michigan
albertberahas@gmail.com

Jiahao Shi
University of Michigan
jiahaos@umich.edu

Baoyu Zhou
University of Chicago
zbaoyu@uchicago.edu

Raghu Bollapragada
Operations Research and Industrial Engineering
The University of Texas at Austin
raghu.bollapragada@utexas.edu

Frank E. Curtis
Industrial and Systems Engineering
Lehigh University
frank.e.curtis@gmail.com

Michael O'Neill
Lehigh University
mikeoneill@unc.edu

Michael O'Neill
University of North Carolina at Chapel Hill
mikeoneill@unc.edu

Daniel Robinson
Lehigh University
daniel.p.robinson@gmail.edu

MS137

Learning-Rate-Free Learning by D-Adaptation

We describe a new hyper-parameter free method which asymptotically achieves the optimal rate of convergence for minimizing convex Lipschitz functions, with no backtracking, line searches or other additional computation per step. Our approach is the first parameter-free method for this class without additional multiplicative log factors in the convergence rate. We present extensive experiments for SGD and Adam variants of our method, where the method automatically matches hand-tuned learning rates across more than a dozen diverse machine learning problems, including large-scale vision and language problems.

Aaron Defazio
Meta
aaron.defazio@gmail

MS138

Incremental Gradient Methods for Optimization Problems with Variational Inequality Constraints

We consider minimizing a sum of agent-specific nondifferentiable merely convex functions over the solution set of a variational inequality (VI) problem in that each agent is associated with a local monotone mapping. This problem finds an application in computation of the best equilibrium in nonlinear complementarity problems arising in transportation networks. We develop an iteratively regularized incremental gradient method where at each iteration, agents communicate over a cycle graph to update their solution iterates using their local information about the objective and the mapping. The proposed method is single-timescale in the sense that it does not involve any excessive hard-to-project computation per iteration. We derive non-asymptotic agent-wise convergence rates for the suboptimality of the global objective function and infeasibility of the VI constraints measured by a suitably defined dual gap function. The proposed method appears to be the first fully iterative scheme equipped with iteration complexity that can address distributed optimization problems with VI constraints over cycle graphs. Preliminary numerical experiments for a transportation network problem and a support vector machine model are presented.

Harshal Kaushik
Bayer
harshaldkaushik@gmail.com

Sepideh Samadi, Farzad Yousefian
Rutgers University
ss4303@scarletmail.rutgers.edu,
farzad.yousefian@rutgers.edu

MS138

An Augmented Lagrangian Framework for Stochastic Optimization Problems with Nonsmooth and Compositional Constraints

We consider resolving the convex stochastic optimization problem where the objective and constraints are convex but possibly complicated by uncertainty and compositional interactions. We present a smoothed augmented Lagrangian framework for resolving such problems where the Lagrangian subproblems are L_k-smooth and L_k is a known constant. By employing variance-reduced schemes for getting approximate solutions of such problems and a suitably defined dual update, rate statements are established for the primal sub-optimality and infeasibility. In addition, overall complexity guarantees are provided in both expectation-valued and compositional regimes.

Peixuan Zhang, Uday V. Shanbhag
Pennsylvania State University
pqz5090@psu.edu, udaybag@psu.edu

Shuogang Yang
Hong Kong University of Science and Technology
yangsg@ust.hk

Ethan Xingyuan Fang
Duke University
ethanfangxy@gmail.com

MS138

Data-driven Piecewise Affine Decision Rule Method for Stochastic Programming with Covariate Information

In this talk, we study a class of stochastic programming problems that minimize the conditionally expected cost given a new covariate observation. With scenario pairs of the random variable and the covariate information, we formulate a data-driven piecewise affine decision rule (PADR) framework for approximately solving the contextual decision-making problem. We provide the non-asymptotic consistency of the data-driven PADR-based framework by establishing the approximation accuracy of piecewise affine functions. To solve the PADR-based empirical risk minimization problem with the coupled non-convex and nondifferentiable structure, we develop an enhanced stochastic majorization minimization algorithm with the nonasymptotic convergence rate analysis in terms of directional stationarity. Numerical results in both unconstrained and constrained problems with various nonlinear generating models imply the superiority of the proposed method compared with the state-of-the-art data-driven methods, especially when the data is small and the true generating model is sparse.

Junyi Liu, Yiyang Zhang, Xiaobo Zhao
Tsinghua University
junyiliu@mail.tsinghua.edu.cn,
zyy21@mails.tsinghua.edu.cn, xbzhao@tsinghua.edu.cn

MS139

Quasi-Newton Subspace Methods for Nonconvex Uncontrained Optimization

This talk discusses subspace properties of subproblems that appear in trust-region and cubic regularization methods for smooth optimization problems. When the Hessian of the objective function is approximated by suitable quasi-Newton matrices it is possible to identify a low dimensional subspace where any solution of the subproblem belongs. These subspace properties allow the development of subspace versions of the referred methods which are suitable to large-scale problems. Convergence and complexity results are established for the subspace methods. The potential advantages of the proposed methods are illustrated by numerical results.

Geovani Grapiglia
Université Catholique de Louvain
geovani.grapiglia@uclouvain.be

MS139

On the Solution of the Cubic Regularization Model in Nonlinear Optimization

The adaptive regularization algorithm (AR2) is a variant of the second-order trust region scheme in which the step length is controlled not explicitly but implicitly. This is achieved by adding a cubic term in the quadratic Taylor model to penalize long steps and updating the strength of the penalization [Cartis,Gould,Toint, SIAM series, 2022]. AR2 enjoys optimal evaluation complexity and is an efficient solver for unconstrained minimization. The main computational cost at each iteration is the solution of the cubic regularization subproblem. Several approaches are available in the literature which are tailored either for the large-scale case or for the case where matrix factorizations are feasible. In this talk we focus on both cases and explore the use of new generation Krylov subspace methods to enhance the inexact model minimization.

Stefania Bellavia
Universita di Firenze
stefania.bellavia@unifi.it

Davide Palitta
Universita' di Bologna
davide.palitta@unibo.it

Margherita Porcelli
Dipartimento di Matematica
Universita' di Bologna
margherita.porcelli@unibo.it

Valeria Simoncini
Universita' di Bologna
valeria.simoncini@unibo.it

MS140

Scheduling Servers with Stochastic Bilinear Rewards

In this paper, we study scheduling in multi-class, multi-server queueing systems with stochastic rewards of job-server assignments following a bilinear model in feature vectors characterizing jobs and servers. A bilinear model allows capturing pairwise interactions of features of jobs and servers. Our goal is regret minimization for the objective of maximizing cumulative reward of job-server assignments over a time horizon against an oracle policy that has complete information about system parameters, while maintaining queueing system stable and allowing for different job priorities. We study a scheduling algorithm based on weighted proportionally fair allocation criteria augmented with marginal costs for reward maximization, along with a linear bandit algorithm for estimating rewards of job-server assignments. We show that our algorithm has a sub-linear regret, as well as a sub-linear bound on the mean queue length, in the time horizon. We also show stability conditions for distributed iterative algorithms for computing allocations, which is of interest in large-scale system applications. We demonstrate the efficiency of our algorithms by numerical experiments using both synthetic randomly generated data and a real-world cluster computing data-trace.

Jung-hun Kim
Department of Industrial & Systems Engineering
KAIST
junghunkim@kaist.ac.kr

Milan Vojnovic
London School of Economics
m.vojnovic@lse.ac.uk

MS140

Scheduling Jobs with Stochastic Holding Costs

We study a single-server scheduling problem for the objective of minimizing the expected cumulative holding cost incurred by jobs, where parameters defining stochastic job holding costs are unknown to the scheduler. We consider a general setting allowing for different job classes, where jobs of the same class have statistically identical holding costs and service times, with an arbitrary number of jobs across classes. In each time step, the server can process a job and observes random holding costs of the jobs that are yet to be completed. We consider a learning-based $c\mu$ rule scheduling which starts with a preemption period of fixed duration, serving as a learning phase, and having gathered data about jobs, it switches to nonpreemptive scheduling. Our algorithms are designed to handle instances with large and small gaps in mean job holding costs and achieve near-optimal performance guarantees. The performance of algorithms is evaluated by regret, where the benchmark is the minimum possible total holding cost attained by the $c\mu$ rule scheduling policy when the parameters of jobs are known. We show regret lower bounds and algorithms that achieve nearly matching regret upper bounds. Our numerical results demonstrate the efficacy of our algorithms and show that our regret analysis is nearly tight.

Dabeen Lee
Korea Advanced Institute of Science and Technology (KAIST)
dabeenl@kaist.ac.kr

Milan Vojnovic
London School of Economics
m.vojnovic@lse.ac.uk

MS141

Stochastic Optimization under Distributional Drift

We consider the problem of minimizing a convex function that is evolving according to unknown and possibly stochastic dynamics, which may depend jointly on time and on the decision variable itself. Such problems abound in the machine learning and signal processing literature, under the names of concept drift, stochastic tracking, and performative prediction. We provide novel non-asymptotic convergence guarantees for stochastic algorithms with iterate averaging, focusing on bounds valid both in expectation and with high probability. The efficiency estimates we obtain clearly decouple the contributions of optimization error, gradient noise, and time drift. Notably, we show that the tracking efficiency of the proximal stochastic gradient method depends only logarithmically on the initialization quality, when equipped with a step-decay schedule. Numerical experiments illustrate our results.

Zaid Harchaoui, Joshua Cutler, Dmitriy Drusvyatskiy
University of Washington
zaid@uw.edu, jocutler@uw.edu, ddrusv@uw.edu

MS141

Bounding Optimality Gap in Stochastic Optimization via Bagging: Statistical Efficiency and Stability

We present a statistical method to estimate the optimal value, and the optimality gap of a given solution, for stochastic optimization. Our method is based on bootstrap aggregating, or bagging, resampled sample average approximation (SAA). We show how our approach leads to valid statistical confidence bounds for possibly non-smooth optimization with enhanced statistical efficiency and stability. In particular, our approach offers variance reduction over batching methods, and more stable standard error estimates than methods directly based on central limit theorems. The theoretical underpinning of our method relies on a new view of SAA as a kernel in an infinite-order symmetric statistic that can be approximated via bagging.

Henry Lam, Huajie Qian
Columbia University
henry.lam@columbia.edu, hq2157@columbia.edu

MS142

Optimal Convex and Nonconvex Regularizers for a Data Source

The use of regularization functionals to promote structure in solving inverse problems has had a long history in applied mathematics, with a recent surge of interest in the use of both convex and nonconvex functionals. The structure of such functionals largely lies in computational and modeling considerations, but there is a lack of an overarching understanding of when convexity or nonconvexity should be preferred for a given data distribution. As a step towards understanding this, we consider the following question: for a given data source, what is the optimal regularizer induced by a star body? To answer this, a variational optimization problem over the space of star bodies is proposed to search for the optimal regularizer. This class of sets covers convex bodies and also some nonconvex geometries. We analyze the structure of population risk minimizers by connecting minimization of the objective to dual mixed volumes. These results are also shown to be robust, as we establish convergence of empirical risk minimizers to the population risk minimizer with increasing data. This talk is based on joint work with Oscar Leong, Yong Sheng Soh, and Venkat Chandrasekaran.

Eliza W. O'Reilly
California Institute of Technology
eoreilly@caltech.edu

MS142

Data Processing Problems with Symmetries

In this talk, we discuss some data analytical problems in which symmetries or invariances in the data arise. We study these problems from a convex geometric perspective in which the space of feasible solutions modeling signals form an orbitope. We briefly discuss the algebraic and geometric properties of such an object along with its implications from a signal processing perspective.

Yong Sheng Soh
National University of Singapore
matsys@nus.edu.sg

MS142

Preconditioned Gradient Descent for Overparameterized Nonconvex Burer-Monteiro Factorization with Global Optimality Certification

We consider using gradient descent to minimize the nonconvex function $f(X) = \phi(XX^T)$ over an $n \times r$ factor matrix X, in which ϕ is an underlying smooth convex cost function defined over $n \times n$ matrices. While only a second-order stationary point X can be provably found in reasonable time, if X is additionally *rank deficient*, then its rank deficiency certifies it as being globally optimal. This way of certifying global optimality necessarily requires the search rank r of the current iterate X to be *overparameterized* with respect to the rank $r^\star$ of the global minimizer $X^\star$. Unfortunately, overparameterization significantly slows down the convergence of gradient descent, from a linear rate with $r = r^\star$ to a sublinear rate when $r > r^\star$, even when ϕ is strongly convex. In this paper, we propose an inexpensive preconditioner that restores the convergence rate of gradient descent back to linear in the overparameterized case, while also making it agnostic to possible ill-conditioning in the global minimizer $X^\star$.

Gavin Zhang
University of Illinois at Urbana-Champaign
jialun2@illinois.edu

Salar Fattahi
University of Berkeley
fattahi@berkeley.edu

Richard Zhang
University of Illinois at Urbana-Champaign
ryz@illinois.edu

MS143

Exploiting Low-Dimensional Structure in Bayesian Inverse Problems Governed by a Nonlinear Stokes Ice Sheet Flow Model

Solving large-scale Bayesian inverse problems governed by complex forward models informed by large observational data suffers from the twin difficulties of the high dimensionality of the uncertain parameters and computationally expensive forward models. When first and second-order (Hessian) derivatives of the underlying optimization problem are available, inexact Newton in conjunction with a Krylov space iterative solver and a suitable preconditioner is an effective means of solving these problems. However, large observational data has the effect of increasing the numerical rank of the Hessian operator and hence the computational cost per Newton iteration. In this talk, we exploit the local sensitivity of model predictions to parameters which suggests that the Hessian has off-diagonal low-rank structure, and so therefore build hierarchical matrix Hessian approximations. We illustrate our approach with a model ice sheet inverse problem governed by the nonlinear Stokes equation for which the basal sliding coefficient field is inferred from the surface ice flow velocity. The results show that for problems with sufficiently-informative data, hierarchical matrices provide a more computationally efficient approximation than the related global low-rank approximation.

Tucker Hartland, Noemi Petra
University of California, Merced
hartland1@llnl.gov, npetra@ucmerced.edu

MS144

Making Bilevel Machine Learning Fast, Scalable and Flexible

Recent tremendous successes of deep learning highly rely on the revolutions of a broad range of learning technologies. Many of these technologies, such as hyperparameter optimization and meta-learning, require bilevel optimization structures. With the increasing scale of datasets and models, large-scale bilevel optimization emerges recently as an exciting but challenging research topic in deep learning. In this talk, I will first discuss some limitations of existing bilevel methods on the scalability with big data and deep neural networks. I will then present two novel algorithms, inspired by Neumann Series (NS) and Evolution Strategy (ES) in stochastic optimization, for addressing these challenges. I also present the theoretical guarantee of the computational efficiency to support our design principles, where some analysis tools can be of independent interest.

Kaiyi Ji
University at Buffalo
kaiyiji@buffalo.edu

MS144

Implicit Gradient Algorithms for Constrained

Bilevel Optimization in Machine Learning

In this talk, we will discuss several applications of bilevel optimization (BLO) in machine learning. Specifically, we will focus on BLO problems with constraints in the lower-level (LL) objective. For such problems, the standard implicit gradient-based approaches do not work in general. In our talk, we will discuss specific cases of constrained BLO and discuss approaches that allow us to develop (approximate) implicit gradient-based algorithms for such problems. Finally, we will demonstrate the performance of the proposed algorithms on signal processing and machine learning applications.

Prashant Khanduri
Wayne State University
khanduri.prashant@wayne.edu

MS144

Algorithmic Foundation of Deep X-risk Optimization

In this talk, I will present our recent research efforts for deep X-risk optimization, which is applicable for solving a variety of compositional measures. I will first present some background about optimization for machine learning and then introduce you a broad family of compositional measures/objectives called X-risk. Then I will talk about the algorithms for optimizing different X-risks and their theoretical complexities. Finally, I will briefly talk about their applications in machine learning and AI.

Tianbao Yang
Texas A&M University
tianbao-yang@tamu.edu

MS145

Modern Challenges in Large-Scale and High Dimensional Data Analysis

Rapidly-growing fields such as data science, uncertainty quantification, and machine learning rely on fast and accurate methods for inverse problems. Three emerging challenges on obtaining relevant solutions to large-scale and data-intensive inverse problems are ill-posedness of the problem, large dimensionality of the parameters, and the complexity of the model constraints. Tackling the immediate challenges that arise from growing model complexities and data-intensive studies, state-of-the-art methods can easily exceed their limits of applicability. In this talk we discuss efficient methods for computing solutions to dynamic inverse problems, where both the quantities of interest and the forward operator typically change at time instances. We consider large-scale ill-posed problems that are made more challenging by their dynamic nature and, possibly, by the limited amount of available data per measurement step. In the first part of the talk, to remedy these difficulties, we apply efficient regularization methods that enforce simultaneous regularization in space and time and achieve this with low computational cost and enhanced accuracy. In the remainder of the talk, we focus on designing spatiotemporal Bayesian Besov priors for computing the MAP estimate in large-scale dynamic inverse problems. Numerical examples from a wide range of applications, such as biomedical and tomographic reconstruction are used to illustrate the effectiveness of the described approaches.

Mirjeta Pasha
Arizona State University
mpasha3@asu.edu

MS145

Optimal parameter design for multi-energy computed tomography technique

Multi-energy computed tomography (MECT) is an x-ray transmission imaging technique that uses the energy dependence of x-ray photon attenuation to determine the elemental composition of an object of interest. Mathematically, forward MECT measurements are modeled by a nonlinear integral transform for which no analytical inversion is available. In this talk, I will present some of our recent results on the global uniqueness and the stability of reconstructions in MECT. These analyses are useful for designing optimal scan parameters, which will be demonstrated using numerical simulations. This is joint work with G. Bal, R. Gong, and E. Sidky.

Fatma Terzioglu
Department of Mathematics
North Carolina State University
fterzioglu@ncsu.edu

MS146

Epigraphs of Spectral Functions As Exotic Cones

We will discuss subclasses of cones defined from epigraphs and perspectives of spectral functions that admit simple self-concordant barrier functions. We implement these cones in the solver Hypatia. We will show that many advanced oracles that make Hypatia's interior point algorithm more efficient are available in analytic form, which reduces the gap in the efficiency of oracles between standard and exotic cones.

Lea Kapelevich
The MathWorks
lkapelev@mathworks.com

Chris Coey
MIT Operations Research Center
coey.chris@gmail.com

Juan Pablo Vielma
Google
juan.pablo.vielma@gmail.com

MS146

ProxSDP.jl : A Semidefinite Programming Solver Written in Julia/jump

ProxSDP is an open-source semidefinite programming (SDP) solver that exploits the low-rank structure inherent to several SDP problems. In this talk we will focus on the following: 1) Presenting the main ideas behind the approximate proximal point algorithm underlying ProxSDP.

2) Using ProxSDP as an example of how one can leverage the Julia and JuMP environment to develop state-of-art optimization software.

Mario Souto
Equilibrium Energy
mariohsouto@gmail.com

MS147

Optimization Problems with Eigenvalue Constraints

In this work, we consider optimization problems including a spectral constraint $\lambda(x) \in C$, where $\lambda(x)$ is the "eigenvector" of a vector x and C is a convex set of simple structure (i.e., projection is easily computable). The eigenvector is considered in an algebraic space like Jordan algebra so that a spectral decomposition is available. While the constraint is nonconvex, we argue that, if the objective function has a specific form, the optimization problems with spectral constraint can be solved via convex optimization over C. We use this fact to investigate some iterative methods such as the projected gradient method and their convergence for more general problems. We demonstrate some examples and numerical results what our optimization model can handle.

Masaru Ito
Nihon University
ito.masaru@nihon-u.ac.jp

Lorenco Bruno
The Institute of Statistical Mathematics
bruno@ism.ac.jp

MS147

Benders Decomposition for Integer Conic Programming: a Computational Report

We consider to apply Benders decomposition for conic programming having integer variables with difficult constraints. It is well-known that Farkas' lemma for conic programming requires an assumption of relative interior feasible points, and we need a similar assumption for Benders decomposition algorithm to work. A similar model and results were already proposed by Vizbári (1996), but our formulation is slightly different. We test the efficiency of Benders decomposition for conic programming by numerical experiments. In the experiments, we focus on second-order cones for closed convex cones in the problems. Test problems are randomly generated SOCP (second-order cone programming) with integer variables. Detailed results will be shown in the talk.

Shinichi Kogawa
3. The University of Electro-Communications
u-41kgw@uec.ac.jp

Masakazu Muramatsu
The University of Electro-Communications
masakazumuramatsu@uec.ac.jp

MS148

A Nearest Neighbor Method for Continuous Simulation Optimization

We propose a version of the nearest neighbor method in machine learning that operates as a global optimization algorithm for continuous simulation optimization. The proposed algorithm is an adaptation of the single observation search algorithm (SOSA) that incorporates the upper confidence bound approach into its sampling process and adopts the quadratic function approximation for its function value estimation. The upper confidence bound component balances exploration and exploitation of SOSA, enhancing the robustness of the sampling process against the random errors. The quadratic function approximation offers a more accurate generalization of the function values within the neighborhood of an optimal solution, increasing the accuracy of the optimal value estimation. The numerical experiments illustrates the behavior of the proposed method when it is applied to well-known test functions.

Seksan Kiatsupaibul
Chulalongkorn University
seksan@cbs.chula.ac.th

Pariyakorn Maneekul
University of Washington
parim@uw.edu

Zelda B. Zabinsky
University of Washington
Industrial and Systems Engineering
zelda@u.washington.edu

MS148

Set-Based Optimization for Multiple Objectives Utilizing Multi-Fidelity Models

Multi-objective and multi-fidelity modeling have a wide range of literature available. Traditional multi-fidelity methods typically assume the high-fidelity model represents ground truth. In reality, there are many applications in which the truth is unknown, such as a system that is yet to be implemented. Even high-fidelity models must incorporate some simplifying assumptions and/or contain noise. I remove this assumption of ground truth and instead analyze solution correlation between models, as highly correlated solutions among different models reflect accuracy. Additionally, the assumption of a ground truth in traditional multi-fidelity modeling leads to methods that focus sampling on selecting the best high-fidelity function evaluations via knowledge derived from low-fidelity models. I instead develop an importance sampling approach that considers objective value, variance/uncertainty, relative accuracy/trustworthiness, and computational cost to sample from all models. Complex systems often contain multiple metrics for optimization, leading to the necessitation of multi-objective optimization methods. In this research, I develop a set-based multi-objective optimization algorithm that utilizes this novel multi-fidelity approach to identify the Pareto-optimal set of solutions.

Danielle F. Morey
University of Washington
dmorey43@uw.edu

Cherry Wakayama
Naval Information Warfare Center Pacific
cherry.y.wakayama.civ@us.navy.mil

Zelda B. Zabinsky
University of Washington
Industrial and Systems Engineering
zelda@u.washington.edu

MS148

Gaussian Process Models for Guiding the Global Optimization of Noisy Functions

Many science applications give rise to optimization problems with computationally expensive, black-box, and noisy objective functions. Surrogate models are often employed to deal with the computational expense and black-box nature of the problems. However, few are suited for noisy functions. We present an optimization method that combines global and local searches to address these types of problems. Here, we use a Gaussian process model to identify interesting regions of the parameter space, and we employ Implicit filtering to thoroughly sample these regions. We show the performance of this method on a set of test problems and an application from quantum computing.

Juliane Mueller
National Renewable Energy Lab
juliane.mueller@nrel.gov

MS149

Only Tails Matter: Average-Case Universality and Robustness in the Convex Regime

The recently developed average-case analysis of optimization methods allows a more fine-grained and representative convergence analysis than usual worst-case results. In exchange, this analysis requires a more precise hypothesis over the data-generating process, namely assuming knowledge of the expected spectral distribution (ESD) of the random matrix associated with the problem. This work shows that the concentration of eigenvalues near the edges of the ESD determines a problem's asymptotic average complexity. This a priori information on this concentration is a more grounded assumption than complete knowledge of the ESD. This approximate concentration is effectively a middle ground between the coarseness of the worst-case scenario convergence and the restrictive previous average-case analysis. We also introduce the Generalized Chebyshev method, asymptotically optimal under a hypothesis on this concentration and globally optimal when the ESD follows a Beta distribution. We compare its performance to classical optimization algorithms, such as gradient descent or Nesterov's scheme, and we show that, in the average-case context, Nesterov's method is universally nearly optimal asymptotically.

Damien Scieur
Samsung SAIL Montreal
damien.scieur@gmail.com

MS150

Federated Minimax Optimization

In this paper, we consider nonconvex minimax optimization, which is gaining prominence in many modern machine learning applications such as GANs. Large-scale edge-based collection of training data in these applications calls for communication-efficient distributed optimization algorithms, such as those used in federated learning, to process the data. In this paper, we analyze Local stochastic gradient descent ascent (SGDA), the local-update version of the SGDA algorithm. SGDA is the core algorithm used in minimax optimization, but it is not well-understood in a distributed setting. We prove that Local SGDA has *order-optimal* sample complexity for several classes of nonconvex-concave and nonconvex-nonconcave minimax problems, and also enjoys *linear speedup* with respect to the number of clients. We provide a novel and tighter analysis, which improves the convergence and communication guarantees in the existing literature. For nonconvex-PL and nonconvex-one-point-concave functions, we improve the existing complexity results for centralized minimax problems. Furthermore, we propose a momentum-based local-update algorithm, which has the same convergence guarantees, but outperforms Local SGDA as demonstrated in our experiments.

Pranay Sharma
Carnegie Mellon University
pranaysh@andrew.cmu.edu

MS150

Compressed Proximal Stochastic Gradient for Decentralized Composite Optimization

In this talk, I will present a compressed proximal stochastic gradient by gradient tracking for solving decentralized composite optimization. To guarantee convergence, existing methods without compression need to either take a large batch of samples for each update or adopt a momentum variance-reduction technique. Numerically, we observe that on many machine learning problems, these methods often do not generalize as well as a standard proximal stochastic gradient method with a small batch. However, the convergence rate result of the latter is unknown. We adopt the technique of the Moreau envelope and establish its convergence rate result. Furthermore, in order to reduce the communication, we incorporate a compression technique in the method and show a similar convergence result can be shown.

Yangyang Xu
Rensselaer Polytechnic Institute
xuy21@rpi.edu

MS150

Advances in Stochastic Variational Inequalities and Applications in Federated Learning

Recently Mishchenko et al. (2022) proposed and analyzed

ProxSkip, a provably efficient method for minimizing the sum of a smooth (f) and an expensive nonsmooth proximable (R) function (i.e. $\min_{x\in\mathbb{R}^d} f(x)+R(x)$). The main advantage of ProxSkip, is that in the federated learning (FL) setting, offers provably an effective acceleration of communication complexity. In this talk, we extend this approach to the more general regularized variational inequality problems (VIP). In particular, we propose ProxSkip-VIP algorithm, which generalizes the original ProxSkip framework of Mishchenko et al. (2022) to VIP, and we provide convergence guarantees for a class of structured non-monotone problems. In the federated learning setting, we explain how our approach achieves acceleration in terms of the communication complexity over existing state-of-the-art FL algorithms. We will also discuss other extensions in FL based on this framework.

Siqi Zhang, Sayantan Choudhury
Johns Hopkins University
szhan207@jhu.edu, schoudh8@jhu.edu

Sebastian U Stitch
CISPA Helmholtz Center for Information Security
stich@cispa.de

Nicolas Loizou
Johns Hopkins University
nloizou@jhu.edu

MS151

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Most causal discovery procedures assume that there are no latent confounders in the system, which is often violated in real-world problems. In this talk, we consider a challenging scenario for causal structure identification, where some variables are latent and they form a hierarchical graph structure to generate the measured variables; the children of latent variables may still be latent and only leaf nodes are measured, and moreover, there can be multiple paths between every pair of variables (i.e., it is beyond tree structure). We propose an estimation procedure that can efficiently locate latent variables, determine their cardinalities, and identify the latent hierarchical structure, by leveraging rank deficiency constraints over the measured variables. We show that the proposed algorithm can find the correct Markov equivalence class of the whole graph asymptotically under proper restrictions on the graph structure.

Biwei Huang
University of California, San Diego
bih007@ucsd.edu

MS151

Reframed GES with a Neural Conditional Dependence Measure

In a nonparametric setting, the causal structure is often identifiable only up to Markov equivalence, and for the purpose of causal inference, it is useful to learn a graphical representation of the Markov equivalence class (MEC). In this paper, we revisit the Greedy Equivalence Search (GES) algorithm, which is widely cited as a score-based algorithm for learning the MEC of the underlying causal structure. We observe that in order to make the GES algorithm consistent in a nonparametric setting, it is not necessary to design a scoring metric that evaluates graphs. Instead, it suffices to plug in a consistent estimator of a measure of conditional dependence to guide the search. We therefore present a reframing of the GES algorithm, which is more flexible than the standard score-based version and readily lends itself to the nonparametric setting with a general measure of conditional dependence. In addition, we propose a neural conditional dependence (NCD) measure, which utilizes the expressive power of deep neural networks to characterize conditional independence in a nonparametric manner. We establish the optimality of the reframed GES algorithm under standard assumptions and the consistency of using our NCD estimator to decide conditional independence. Together these results justify the proposed approach. Experimental results demonstrate the effectiveness of our method in causal discovery, as well as the advantages of using our NCD measure over kernel-based measures.

Xinwei Shen
ETH Zurich
xinwei.shen@stat.math.ethz.ch

MS151

Consistent and efficient mixed integer programming for learning directed acyclic graphs

Learning directed acyclic graphs (DAGs) is computationally and statistically challenging. We cast the problem as a mixed-integer program with an objective function composed of a convex quadratic loss function and a regularization penalty subject to linear constraints. The optimal solution to this mathematical program is known to have desirable statistical properties under certain conditions. However, the state-of-the-art optimization solvers are not able to obtain provably optimal solutions to the existing mathematical formulations for medium-size problems within reasonable computational time. To address this difficulty, we tackle the problem from both computational and statistical perspectives. Computationally, we propose an efficient mixed-integer quadratic optimization (MIQO) model, the layered network formulation. In addition to offering improvements compared with the existing approaches, the new formulation can also take advantage of easily obtainable super structures, such as the moral graph, to reduce the number of possible DAGs. Statistically, we propose an early stopping criterion to terminate the branch-and-bound process in order to obtain a near-optimal solution to the mixed-integer program, and establish the consistency of this approximate solution.

Ali Shojaie
University of Washington
ashojaie@uw.edu

Simge Kucukyavuz
Northwestern University

simge@northwestern.edu

MS152

Subgradient methods avoid strict saddle point in nonsmooth optimization

This talk seeks to answer whether subgradient-type methods avoid saddle points of nonsmooth functions. Our main result shows that for a ”generic” class of semialgebraic functions, stochastically perturbed (projected)subgradient methods converge only to local minimizers. As a by-product of our proof technique, we extend the classical asymptotic normality guarantee to stochastic subgradient methods.

Damek Davis
Cornell University
ORIE
dsd95@cornell.edu

Dmitriy Drusvyatskiy
University of Washington
ddrusv@uw.edu

Liwei Jiang
Cornell University
lj282@cornell.edu

MS152

Subgradient Sampling for Nonsmooth Nonconvex Minimization

We focus on a stochastic minimization problem in the nonsmooth and nonconvex setting which applies for instance to the training of deep learning models. A popular way in the machine learning community to deal with this problem is to use stochastic gradient descent (SGD). This method combines both subgradient sampling and backpropagation, which is an efficient implementation of the chain-rule formula on nonsmooth compositions. Due to the incompatibility of these operations in the nonsmooth world, the SGD can generate spurious critical points in the optimization landscape which does not guarantee the convergence of the objective function to a meaningful value. We will explain in this talk how the model of Conservative Gradients is compatible with subgradient sampling and backpropagation, allowing to obtain convergence results for nonsmooth SGD. By means of definable geometry, we will emphasize that functions used in machine learning are locally endown with geometric properties of piecewise affine functions. In this setting, chain-ruling nonsmooth functions, and sampling subgradients output conservative gradients, but also, spurious critical points are hardly attained when performing SGD in practice.

Jérôme Bolte
TSE / Université Toulouse I Capitole
jerome.bolte@ut-capitole.fr

Tam Le
TSE, Université Toulouse 1 Capitole
ngoc-tam.le@ut-capitole.fr

Edouard Pauwels
IRIT, Université Toulouse 3
edouard.pauwels@irit.fr

MS152

Convergence of Random Reshuffling under the Kurdyka-Lojasiewicz Inequality

In this talk, we present novel convergence properties of the random reshuffling (RR) method for smooth nonconvex optimization problems with a finite-sum structure. Though this method is widely utilized in practice, e.g., in the training of neural networks, its convergence behavior is only understood in several limiting cases. Based on the Kurdyka-Lojasiewicz (KL) inequality, we verify that the whole sequence of iterates generated by RR is convergent and converges to a single stationary point of problem in an almost sure sense. Corresponding rates of convergence are derived which depend on the KL exponent and suitably selected diminishing step sizes. When the KL exponent lies in $[0, \frac{1}{2}]$, the convergence is at a rate of $\mathcal{O}(t^{-1})$ with t counting the number of iterations. The standard KL inequality-based convergence framework only applies to algorithms with a certain descent property. We conduct a novel analysis for the non-descent RR method with diminishing step sizes based on the KL inequality which generalizes the standard KL framework and techniques.

Junwen Qiu, Xiao Li
The Chinese University of Hong Kong, Shenzhen
junwenqiu@link.cuhk.edu.cn, lixiao@cuhk.edu.cn

Andre Milzarek
The Chinese University of Hong Kong, Shenzhen
School of Data Science
andremilzarek@cuhk.edu.cn

MS153

Interior Point Method for SDPs Using the Factor-Width Cone

In this talk we revisit the class of factor-width cones and discuss how they may be leveraged to approximate semidefinite programming problems. We present an interior point algorithm over these cones which converges to the optimal solution of the SDP, thereby generalizing a result by Sznaier and Roig-Solvas. The underlying idea of the proposed method is an iterative rescaling of the data matrices and was inspired by Ahmadi and Hall. This is joint work with Etienne de Klerk.

Felix Kirschner
Tilburg University
f.c.kirschner@tilburguniversity.edu

MS153

Polynomial Optimization and Low-Rank SDPs

Polynomial optimization is an important class of nonconvex optimization problems, which is closely related to measure theory, the theory of polynomial nonnegativity in mathematics, and also arises in diverse areas including power networks, computer vision, combination optimization, neutral networks, quantum information. In this talk, I will introduce the popular tool for globally solving polynomial optimization – the Moment-SOS hierarchy, and discuss how to reduce its complexity by exploiting sparsity. To efficiently solve large-scale low-rank SDPs arising from the Moment-SOS hierarchy, we propose an augmented Lagrange framework based on Riemann manifold optimization.

Jie Wang
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
wangjie212@amss.ac.cn

MS154

Optimal Distributional Robustness for Offline Reinforcement Learning

We introduce a new modelling framework for the Offline Off-policy Reinforcement Learning problem when having access to a single trajectory of correlated data. The presented method combines a generalized information projection which is induced by the correlation structure of the underlying Markov Decision Process with a specifically tailored distributionally robust optimization problem. Our approach is purely data-driven relying solely on the behavioral data without assuming any expert knowledge. We show that the presented framework achieves optimal robustness guarantees among a wide range of popular approaches.

Mengmeng Li
EPFL
mengmeng.li@epfl.ch

Tobias Sutter
Swiss Federal Institute of Technology in Lausanne
tobias.sutter@uni-konstanz.de

Daniel Kuhn
École Polytechnique Fédérale de Lausanne
daniel.kuhn@epfl.ch

MS154

Numerically Solving High-Dimensional Reflected Brownian Motion Control Based on Deep Reinforcement Learning

Reflected Brownian Motion (RBM) is essential in stochastic modeling and operations management. Despite its modeling power, it is well-known that the RBM control problems are challenging, even numerically. The recent progress of deep learning and reinforcement learning inspires us to solve using neural network approximations. In this paper, we study the discount-reward and ergodic RBM drift-control problems based on neural network parametrization of the value and gradient functions. Our method is an analogue of Q-learning in reinforcement learning and could be seen as an off-policy method because it only uses simulated paths under one particular control. We then extend our method to deal with singular control. Finally, We demonstrate the performance of our methods in two standard operations management models: make-to-order production with thin streams and dynamic pricing in high-dimensional regime.

Nian Si
The University of Chicago
niansi@chicagobooth.edu

Baris Ata
ChicagoBooth
baris.ata@chicagobooth.edu

Michael Harrison
Stanford GSB
mike.harrison@stanford.edu

MS155

Balancing Social Optimality and Nash Equilibria for Bid Recommendations in Online Ad Auctions: A Mean-field Optimization Approach

Mean-field Game and Mean-field Control have been widely utilized in various application fields to analyze multi-agent games. However, they fall short in addressing complex problems, such as bid recommendations in online ad auctions, due to the interplay of competition and cooperation. Recommended bids should consider not only the competition among advertisers for online ad impressions, but also the interests of the publisher, shoppers, and advertisers in terms of ad revenue, user experience, and sales. The ad publisher aims to optimally balance the interests of all parties, including their own, while considering the existing competition. This cannot be achieved through traditional social optimality or equilibrium concepts alone. To address this challenge, we propose an optimization framework that balances the trade-off between social optimality and equilibria. We further establish convergence guarantees and study our framework s asymptotic relationship with the classical equilibrium selection problem. Our extensive experiments in a simulated ad auction environment show that our proposed framework efficiently estimates mean-field equilibria bidding policies and effectively balances the optimization of multiple objectives, demonstrating its empirical efficacy.

Xin Guo
Amazon and University of California, Berkeley, U.S.
xnguo@amazon.com

Lihong Li, Sareh Nabi, Rabih Salhab
Amazon, U.S.
llh@amazon.com, sareh@amazon.com,
rabihsal@amazon.com

Junzi Zhang
The Citadel, U.S.
junziz@stanford.edu

MS155

Scalable Approximately Optimal Policies for Multi-Item Stochastic Inventory Problems

We analyze an inventory system wit I items sharing a limited storage capacity or inventory budget in each of T periods of the planning horizon. Demands for the different items follow a general multivariate Normal distribution allowing for arbitrary correlation structures. Inventories may be adjusted by placing orders which arrive after a given lead time, or by salvaging part of the inventory. The capacity constraints are modeled as chance constraints that impose an upper bound on the overflow probability in each period. We propose a heuristic policy that is asymptotically optimal in limit of a large number of items I when demands are correlated only among items within a common product line. The complexity of computing the policy grows as I2̂ T(̂3/2). We also propose another policy that we recommend for moderate values of I. An extensive numerical study involving more than 28,000 instances with up to 40 items, shows that the average gap between the upper bound and lower bound is 1.05%, with 98.2% exhibiting a gap smaller than 5%. Empirically, we observe that the runtime of the inventory reduction algorithm grows linearly with I. From a managerial perspective, we provide a scalable methodology to identify near optimal procurement strategies for a problem that is central to most brick-and-mortar and online retailers and distributors. Additionally, this methodology can be used to guide capacity planning as well as assortment decisions.

Garud N. Iyengar
Columbia University
Dept of Indust Eng & Oper Rsch
garud@ieor.columbia.edu

MS155

Cost Optimization in Physical Store Supply Chain

In daily operations of physical retail stores (e.g. grocery), store managers face tight capacity constraints and thus continuously have to make trade-offs between cost of capacity and labor cost. In this paper, we investigate a two-echelon distribution network, where the distribution centers (DC) are used to supply customer facing physical retail stores. More specifically, we focus on the cost impact of ship-pack decisions for the flows between DCs and stores. Ship-packs could be master-packs, inner-packs or eaches. We developed a mathematical model that makes optimal ship-pack decisions for each item by considering the total cost impact on the supply chain, i.e. on the DCs as well as on the physical retail stores. We show that by optimally deciding the ship-pack for each item, overall supply chain costs could be significantly reduced. Finally, most store managers tend to argue that the item sales velocity is the single most dominant factor affecting its ship-pack decision. Our results also demonstrate that this is not quite true, and the inventory replenishment strategy and store shelf capacity for each item are also critical factors.

Di Wu
Amazon
diw@amazon.com

Tolga Cezik
Amazon.com
cezikm@amazon.com

MS156

A Second-Order Descent Method with Active-Set Prediction for Group Sparse Optimization

In this paper, we propose a second-order algorithm for the solution of finite and infinite dimensional group sparse optimization problems. Group sparse optimization has gained a lot of attention in the last years due to several important classification problems requiring group sparse solutions. The most prominent application example is the group LASSO problem, which consists in minimizing a least-squares fitting term together with the group sparsity l_1/l_2 norm. The method is built upon the steepest descent directions of the nonsmooth problem, which are further modified by using second-order information. A prediction step is also proposed for faster identification of the strong active set. A general convergence result is proved, and the active set behavior is analyzed. The work ends with comparative computational experiments to test the performance of the devised algorithm.

Sofía López
Escuela Politécnica Nacional del Ecuador
sofia.lopezo@epn.edu.ec

Juan Carlos De los Reyes
MODEMAT, Escuela Politecnica Nacional
Quito, Ecuador
juan.delosreyes@epn.edu.ec

Pedro Merino
MODEMAT, Escuela Politecnica Nacional Quito, Ecuador
pedro.merino@epn.edu.ec

MS156

Information in High-Dimensional Stochastic Control

In economics and many disciplines of engineering, the available ‘information’ (i.e., consumer/market data, sensor measurements) plays an essential role in the design of models and control mechanisms. The rapidly developing field of mean field games and control theory builds a framework to analyze models consisting of many interacting agents. 'Information structures' in these models can dictate how the agents cooperate as well as how they determine their strategies from available data. In the 'mean field limit' as the number of agents goes to infinity, the limiting mean field game is known to not distinguish between information structures when the interactions are sufficiently 'long range'. We investigate further this phenomenon to better understand the role that information does play in these problems. Using a functional central limit theorem, we find that the ‘fluctuations’ from the mean field limit are determined by the solution of a linear-quadratic-Gaussian stochastic control problem. We also consider machine learning based algorithms that leverage this emergent structure to obtain good approximations to the optimal control for high dimensional problems.

Aaron Z. Palmer
University of British Columbia
azp@math.ucla.edu

MS156

Energy Dissipative Adaptive Gradient Descent Method

We propose a new optimization algorithm for a general non-convex function, which possesses unconditionally energy stable property. The optimization problem is equivalent to a gradient flow problem in some case, where the gradient flow can be derived by a free energy functional. We introduce a new modified energy based on a relaxation form of the original energy. The algorithm will keep energy dissipative regardless of the learning rate. Despite it, the selection of the learning rate can be determined by an energy indicator. Assuming the objective function to be smooth and convex, we can show the convergence and obtain the first order convergence rate of the algorithm.

Shiheng Zhang
Purdue University
zhan3722@purdue.edu

MS157

Exponential Decay of Sensitivity and Domain Decomposition with Overlap for Dynamic and Other Graph-Indexed Optimization

We prove that certain classes of graph-indexed optimization (GIO) problems exhibit exponential decay of sensitivity with respect to perturbation in the data. GIOs include dynamic optimization (for the linear graph), or distributed, including network control (for the mesh/network-time product graph). This feature allows for very efficient approximation, solutions, or policies based on domain decomposition with overlap relative to centralized or monolithic approaches. In particular, we prove that the proper efficiency metric increases exponentially fast with the overlap size. Immediate consequences of such behavior are that distributed control policies with overlap approach the performance of centralized policies exponentially fast and that Schwarz-type algorithms exhibit, in addition to exceptional parallelism, a linear rate of convergence that tends exponentially fast to zero.

Sungho Shin
Argonne National Laboratory
sshin@anl.gov

Victor Zavala
University of Wisconsin Madison
zavalatejeda@wisc.edu

Mihai Anitescu
Argonne National Laboratory
Mathematics and Computer Science Division
anitescu@mcs.anl.gov

MS157

Modeling and Solution of Large-Scale Block-Structured Nonlinear Optimization Problems in Julia

Many large-scale nonlinear programs (NLPs) exhibit block structures; prototypical examples include optimal control problems and multi-stage stochastic programs. Several decomposition-based algorithms have been proposed in the literature to solve large-scale instances of these problems in a scalable manner. Although these algorithms have been shown to be effective in practice, there are currently no readily usable modeling and solution tools that implement these methods. The implementation of these methods with existing modeling and solution tools is complex as it requires ad-hoc and intrusive manipulation of problem data structures. In this work, we introduce BlockNLPModels.jl and BlockNLPAlgorithms.jl, open-source software frameworks implemented in Julia Language that facilitate modeling and solution of block-structured NLPs. BlockNLPModels.jl provides AbstractBlockNLPModel, a minimal abstraction for block-structured NLPs. The models represented in AbstractBlockNLPModel format can be solved in a scalable manner by BlockNLPAlgorithms.jl, which provides parallel implementations of dual decomposition and alternating direction method of multipliers (ADMM). Our framework allows the modular construction of block-structured models and offers the flexibility for the user to specify tailored algorithms for individual subproblems. We showcase the capability of BlockNLPModels.jl and BlockNLPAlgorithms.jl with large-scale PDE-constrained control problems.

Rishabh Gupta
University of Minnesota
rishabh@umn.edu

François Pacaud, Sungho Shin
Argonne National Laboratory
francois.pacaud@protonmail.com, sshin@anl.gov

Anirudh Subramanyam
Pennsylvania State University
subramanyam@psu.edu

MS157

STOCS: Simultaneous Trajectory Optimization and Contact Selection for Contact-Rich Manipulation

Contact-implicit trajectory optimization is an effective method to plan complex trajectories for various contact-rich systems including manipulation and locomotion. These methods formulate contact as complementarity constraints and require solving a mathematical program with complementarity constraints (MPCC). However, MPCC solve times increase steeply with the number of variables and complementarity constraints, which limits their applicability to problems with low geometric complexity. This paper introduces the simultaneous trajectory optimization and contact selection (STOCS) method that embeds the detection of salient contact points and contact times inside trajectory optimization. Because the number of active contact points is usually small, this approach minimize the number of MPCC variables and constraints, which makes solving manipulation trajectories for objects with complex, non-convex geometry computationally tractable. The proposed approach is validated on pivoting and sliding problems in simulation and on a 6 DoF manipulator arm.

Mengchao Zhang
University of Illinois, Urbana-Champaign
mz17@illinois.edu

Yuki Shirai
University of California, Los Angeles
yukishirai4869@g.ucla.edu

Devesh Jha, Arvind Raghunathan
Mitsubishi Electric Research Laboratories
jha@merl.com, raghunathan@merl.com

Kris Hauser
University of Illinois, Urbana-Champaign
kkhauser@illinois.edu

MS158

Role of Non-Newtonian Geometry of Curved Space in Optimization

We illustrate this role in multiple cases of optimization occurring in nature or man-made problems which may be convex or non-convex. Nature solves optimization problems, since laws of nature are nothing but optimality conditions. While Newtons theory of gravitation was defined in flat Euclidean space, Einstein showed that space-time is actually a curved manifold with metric governed by his famous p.d.e. This theory has correctly explained many observed phenomenon in astrophysics and structure of the universe. In man-made convex optimization problems such as Linear Programming and Tensor Optimization (which includes semi-definite programming), interior-point methodology exploits underlying Non-Newtonian geometry of interior-point space, leading to efficient methods for large-scale problems. This role is further amplified for non-convex problems by defining appropriate metric and affine-connection of curved space satisfying certain p.d.e., so that the function to be optimized becomes a generalized polynomial, i.e. having Taylor series expansion w.r.t. arc-length parameter of geodesics in the curved space which terminates after finite number of terms.

Narendra Karmarkar
Consultant to the Indian Government, India
narendrakarmarkar@yahoo.com

MS158

Tensor Optimization and Applications

First example of applying tensor optimization to combinatorial problems was shown in IPCO 1992: pages 406-420. We improve and strengthen those results in several ways and obtain computational results on three problems graph partitioning, satisfiability and analysis of counterexamples related to Hilberts 17th problem. For this we created a mixed symbolic numeric model formulation package which facilitates definition of objective function, equality and inequality constraints and definition of new dependent variables. For discrete problems certain inequalities valid at candidate solutions are dynamically incorporated in the iterations of the continuous optimization algorithm based on underlying non-Newtonian geometry of the interior-point space. For graph partitioning we obtain optimal solutions including proof of optimality. For satisfiability problem we either find the satisfiable assignment or construct and output proof of unsatisfiability. For Hilberts 17th problem we analyse concrete examples whose non-negativity has been established to be not provable using sums of the squares expressions valid in RN. However, for these counterexamples, we construct non-negativity proofs by computationally constructing sums of squares expressions valid on certain sub-varieties of RN The same modeling package mentioned above is used to post process the solver output into symbolic proofs of optimality or infeasibility.

Tanmaya Karmarkar
Computer Science, I. K. Barber Faculty of Science
University of British Columbia
tanmayak@student.ubc.ca

MS158

Searching For Hypergraphs Using Reinforcement Learning

We investigate the use of deep reinforcement learning methods for finding optimal combinatorial structures. In particular, we extend the methods of Wagner to hypergraphs, aiming to find examples that challenge the Fredman and Khachiyan duality checking algorithm. In doing this, we highlight some natural optimization problems on pairs of transversal hypergraphs, where we present some preliminary extremal results.

Parsa Salimi
University of Waterloo
p2salimi@uwaterloo.ca

Tamon Stephen
Simon Fraser University
tamon@sfu.ca

MS159

Balancing Communication and Computation in Gradient Tracking Algorithms for Decentralized Optimization

In this talk, we present a unified framework for the various gradient tracking algorithms present in literature. We introduce flexibility with respect to communication structure and number of communication and computation steps performed each iterations over this framework. We theoretically differentiate among the different gradient tracking algorithms. We provide convergence conditions and theoretically show improvement with the introduced flexibility. We illustrate the improvements on quadratic functions and binary classification problems.

Shagun Gupta

University of Texas at Austin
shagungupta@utexas.edu

MS159

Huber Loss-based Penalty Approach to Problems with Linear Constraints*

We consider an optimization problem with many linear inequalities constraints. To deal with a large number of constraints, we provide a penalty reformulation of the problem, where the penalty is a variant of the one-sided Huber loss function. We study the infeasibility properties of the solutions of penalized problems for nonconvex and convex objective functions, as the penalty parameters vary with time. Then, we propose a random incremental penalty method for solving convex problem, and investigate its convergence properties for convex and strongly convex objective functions. We show that the iterates of the method converge to a solution of the original problem almost surely and in expectation for suitable choices of the penalty parameters and the stepsize. Also, we establish convergence rate of the method for appropriately defined weighted averages of the iterates for the expected function values. We establish convergence rate results which, to the best of our knowledge, are the first results on the convergence rate for the penalty-based incremental subgradient method with time-varying penalty parameters.

Angelia Nedich
Arizona State University
angelia.nedich@asu.edu

Tatiana Tatarenko
TU Darmstadt
tatiana.tatarenko@rmr.tu-darmstadt.de

MS159

Analyzing Inexact Hypgergradients for Bilevel Learning

Estimating hyperparameters has been a long-standing problem in machine learning. We consider the case where the task at hand is modelled as the solution to an optimization problem. In this case the exact gradient with respect to the hyperparameters cannot be computed and approximate strategies are required. We provide an analysis of two classes of methods based on inexact automatic differentiation or approximate implicit differentiation. Our analysis reveals that these two strategies are actually tightly connected, and we derive a priori and a posteriori estimates for both methods which can be used to bound computations and gain further insights what their accuracy actually depends on.

Matthias Ehrhardt
Institute for Mathematical Innovation
University of Bath
m.ehrhardt@bath.ac.uk

Lindon Roberts, Lindon Roberts
University of Sydney
lindon.roberts@sydney.edu.au, lindon.roberts@sydney.edu.au

MS160

On the conservative fields of optimal value functions

Properties of optimal value functions are critical for the design of numerical algorithms for min-max, two-stage and bilevel optimization problems. In this talk, we address the first-order variational properties of the value functions for bi-parametrized nonlinear optimization problems, where both the data in the objective and constraints are varying coefficients. We characterize the conservative fields of such optimal value functions without requiring the uniqueness of the optimal solutions or multipliers. The results provide the theoretical justification of (stochastic) gradient descent methods for solving complicated hierarchical optimization problems with varying constrained sets.

Ying Cui
University of Minnesota
yingcui@umn.edu

MS160

Complexity analysis for a zeroth-order framework for the optimization of value-functions under uncertainty.

Value-function optimization emerges in a range of settings. Such problems are particularly challenging when the value-function corresponds to a stochastic optimization problem. Few schemes, if any, exist for contending with such problems in constrained regimes. We present an inexact zeroth-order scheme for resolving such problems and provide the associated rate and complexity guarantees.

Aswin Kannan
Humboldt University
aswin.kannan@hu-berlin.de

Farzad Yousefian
Rutgers University
farzad.yousefian@rutgers.edu

Uday V. Shanbhag
Pennsylvania State University
udaybag@psu.edu

MS160

Surrogate Schemes for Computation of Nash Equilibria under Uncertainty

Surrogate schemes have proved useful in resolving a broad range of distributed and nonconvex optimization problems. We consider a class of Nash equilibrium problems complicated by uncertainty. Synchronous surrogate schemes are presented and the associated rate and complexity guarantees for inexact counterparts are provided.

Zhuoyu Xiao, Uday V. Shanbhag
Pennsylvania State University

zvx5148@psu.edu, udaybag@psu.edu

MS161

A Logistic Regression and Linear Programming Approach for Multi-Skill Staffing Optimization in Call Centers

We study a staffing optimization problem in multi-skill call centers. The objective is to minimize the total cost of agents under some quality of service (QoS) constraints. The key challenge lies in the fact that the QoS functions have no closed-form and need to be approximated by simulation. We propose a new way to approximate the QoS functions by logistic functions and design a new algorithm that combines logistic regression, cut generations, and logistic-based local search to efficiently find good staffing solutions. We report computational results using examples up to 65 call types and 89 agent groups showing that our approach performs well in practice, in terms of solution quality and computing time.

Thuy Anh Ta
Phenikaa University
anh.tathuy@phenikaa-uni.edu.vn

Tien Mai
Singapore Management University
atmai@smu.edu.sg

Fabian Bastin
University of Montreal
bastin@iro.umontreal.ca

Pierre L'Ecuyer
Université de Montréal
lecuyer@iro.umontreal.ca

MS161

New Results on Penalty Decomposition Methods for Constrained Optimization

The Penalty Decomposition (PD) approach is a well consolidated optimization strategy devised for some classes of problems with structured, difficult constraints. In this talk, we discuss some novel results concerning the possibility of adopting this approach to tackle a wider, general class of constrained optimization problems, both in inexact and in exact fashion. The proposed algorithms have been theoretically analyzed, so that convergence to feasible, first-order stationary solutions have been proved in the general setting. Moreover, we show the results of thorough numerical experiments highlighting the computational potential of the method.

Matteo Lapucci
DINFO, Università degli Studi di Firenze
matteo.lapucci@unifi.it

Christian Kanzow
University of Wuerzburg
kanzow@mathematik.uni-wuerzburg.de

Marco Sciandrone
Università la Sapienza, Roma
sciandrone@diag.uniroma1.it

MS162

Low-Rank Binary Matrix Completion via Integer Programing over GF(2)

Given a data matrix X, the goal in low-rank approximation is to find two rank-r matrices such that UV approximates X. Along with recovering missing entries in X, low-rank approximation also helps in finding factors and patterns in the data. This problem is well studied over reals but has received less attention in the case of binary data. Many machine learning applications, e.g., classification, community detection and others, involve data with binary values, and discovering their latent structure is often useful. In this work, we focus on finding rank-r binary matrices U and V such that UV approximates the given partially observed binary data matrix X. We define the vector product with respect to the finite field GF(2) which follows modulo-2 arithmetic. We derive two new integer programming-based formulations: a disjunctive formulation making use of the representation of the parity polytope and a compact formulation arising from a new class of inequalities. We study theoretical and computational comparison of the proposed formulations.

Akhilesh Soni
UW-Madison
soni6@wisc.edu

Jeff Linderoth
University of Wisconsin-Madison
Dept. of Industrial and Sys. Eng
linderoth@wisc.edu

James Luedtke
University of Wisconsin-Madison
Department of Industrial and Systems Engineering
jrluedt1@wisc.edu

Daniel L. Alarcon
University of Wisconsin-Madison
pimentelalar@wisc.edu

MS163

Risk Constraints in Stochastic Dynamic Optimization

We consider multi-stage stochastic optimization problems with time-consistent constraints reflecting risk-aversion. Integration of such constraints in a sequential decision problems frequently results in a time-inconsistent evaluations and inconsistency in decision policy. We propose a framework for inclusion of risk-evaluation and constraint at any stage of the decision process and analyze the resulting multi-stage problems. We derive optimality condition and establish some relations to other models of risk.

Darinka Dentcheva
Department of Mathematical Sciences
Stevens Institute of Technology

darinka.dentcheva@stevens.edu

MS163

Mini-Batch Risk Forms

Risk forms are real functionals of two arguments: a bounded measurable function on a Polish space and a probability measure on that space. They are convenient mathematical structures adapting the coherent risk measures to the situation of variable reference probability measure. We introduce a new class of risk forms, which we call mini-batch forms. We construct them by using a random empirical probability measure as the second argument and by post-composition with the expected value operator. We prove that coherent and law-invariant risk forms generate mini-batch risk forms which are well-defined on the space of integrable random variables, and we derive their dual representation. We demonstrate how unbiased stochastic subgradients of such risk forms can be constructed. Then, we consider pre-compositions of mini-batch risk forms with nonsmooth and nonconvex functions, which are differentiable in a generalized way, and we derive generalized subgradients and unbiased stochastic subgradients of such compositions. Finally, we study the dependence of risk forms and mini-batch risk forms on perturbation of the probability measure and establish quantitative stability in terms of optimal transport metrics. For mini-batch risk forms involving functions on a finite-dimensional space, we obtain finite-sample expected error estimates.

Andrzej Ruszczynski
Rutgers University
rusz@business.rutgers.edu

Darinka Dentcheva
Department of Mathematical Sciences
Stevens Institute of Technology
darinka.dentcheva@stevens.edu

MS164

Efficiency and novelty in dynamic tomography

Tomography is probably the most well-known inverse problem with countless applications motivating an equally wide variety of different models and priors. One particular example is tomography of moving targets or dynamic tomography which is of growing interest to mathematicians and engineers studying inverse problems and designing novel algorithms, and the many users of computed tomography in medicine, research and industry studying some dynamic phenomena. Since any major movement of the target leads to errors, artifacts and ill-posedness, the solutions are often tailored to the specific problem at hand and use novel and specialized models, priors and objective functions. This can greatly increases the complexity of the problem and make the algorithms tricky to apply in practice. Whereas in applications we are often used to high resolution 3D volumes produced by lightning fast algorithms. In this presentation I will briefly discuss balancing efficiency and novelty for at least some dynamic tomography problems with realistic data.

Tommi Heikkilä
University of Helsinki (Finland)
tommi.heikkila@helsinki.fi

MS164

A Projected Nesterov-Kaczmarz Approach to Stellar Population Distribution Reconstruction in Extragalactic Archaeology

We consider the problem of reconstructing a galaxys stellar population distribution function from spectroscopy measurements. These quantities can be connected via the single-stellar population spectrum, resulting in a very large-scale integral equation with a system structure. To solve this problem, we propose a projected Nesterov-Kaczmarz reconstruction (PNKR) method, which efficiently leverages the system structure and incorporates physical prior information such as smoothness and non-negativity constraints.

Fabian Hinterer
Institute fro Industrial Mathematics
Kepler University Linz
fabian.hinterer@indmath.uni-linz.ac.at

MS165

Solving A Class of Stochastic Bilevel Optimization Problems is Nearly as Easy as Solving Single-level Problems

Stochastic bilevel optimization, including stochastic compositional and min-max optimization as special cases, is gaining popularity in many machine learning applications. While the three problems share a nested structure, existing works often treat them separately, thus developing problem-specific algorithms and analyses. Among various exciting developments, simple (projected) SGD-type updates (potentially on multiple variables) are still prevalent in solving this class of nested problems. Still, they are believed to have a slower convergence rate than single-level problems. This talk unifies several (projected) SGD-type updates for stochastic bilevel problems into a single SGD approach that we term Alternating Implicit Stochastic gradient dEscenT (AISET) method. By leveraging the smoothness of the problem, this talk reveals that for a growing class of stochastic bilevel optimization problems, AISET matches SGD for solving single-level nonconvex optimization problems in terms of sample complexity.

Tianyi Chen
RPI
chentianyi19@gmail.com

MS165

INTERACT: Achieving Low Sample and Communication Complexities in Decentralized Bilevel Learning over Networks

In recent years, decentralized bilevel optimization problems have received increasing attention in the networking and machine learning communities. However, for decentralized bilevel optimization over networks with limited computation and communication capabilities, how to achieve low sample and communication complexities are two fundamental challenges. In this work, we make the first attempt to investigate the class of decentralized bilevel optimization problems with nonconvex and strongly-convex structure corresponding to the outer and inner subproblems, respectively. Our main contributions in this paper are two-fold: i) We first propose a deterministic algorithm called INTERACT that requires low sample and communication complexities to solve the bilevel optimization problem. To relax the need for full gradient evaluations in each iteration, we propose a stochastic variance-reduced version of INTERACT, which improves the sample complexity while achieving the same communication complexity as the deterministic algorithm.

Jia Liu
Ohio State University
liu.1736@osu.edu

MS165

A Conditional Gradient-based Method for Simple Bilevel Optimization with Convex Lower-level Problem

In this talk, we study a class of bilevel optimization problems, also known as simple bilevel optimization, where we minimize a smooth objective function over the optimal solution set of another convex constrained optimization problem. Several iterative methods have been developed for tackling this class of problems. Alas, their convergence guarantees are either asymptotic for the upper-level objective, or the convergence rates are slow and sub-optimal. To address this issue, in this paper, we introduce a novel bilevel optimization method that locally approximates the solution set of the lower-level problem via a cutting plane, and then runs a conditional gradient update to decrease the upper-level objective. When the upper-level objective is convex, we show that our method requires $O(\max 1/\epsilon_f, 1/\epsilon_g)$ iterations to find a solution that is ϵ_f-optimal for the upper-level objective and ϵ_g-optimal for the lower-level objective. Moreover, when the upper-level objective is non-convex, our method requires $O(\max 1/\epsilon_f^2, 1/(\epsilon_f \epsilon_g))$ iterations to find an (ϵ_f, ϵ_g)-optimal solution. We also prove stronger convergence guarantees under the Hölderian error bound assumption on the lower-level problem. To the best of our knowledge, our method achieves the best-known iteration complexity for the considered class of bilevel problems.

Aryan Mokhtari
Assistant Proessor UT Austin
mokhtari@austin.utexas.edu

MS166

Photonics Design Via Optimization

In the photonic design problem, an engineer chooses certain parameters of a device (for example, the material composition) so that it best matches some desired behavior. These design problems can often be stated as mathematical optimization problems, where the desired behavior is captured by an objective function that the designer seeks to minimize. In many cases, these design problems are nonconvex and computationally difficult to solve. In this talk, we show how to model these design problems and use heuristics to generate approximately optimal designs. We then present a method for computing bounds on the best possible performance that any device can achieve for a given objective function, using convex optimization. All examples are accompanied by open source Julia code.

Theo Diamandis
MIT
tdiamand@mit.edu

MS166

A Decision - Emulation Approach to the Simulation of Power Systems Operations

This presentation introduces PowerSimulations.jl v1.0, a Julia-based BSD-licensed power system operations simulation tool developed as a flexible and open-source software for quasi-static power systems simulations including Production Cost Models (PCM). PowerSimulations.jl tackles the issues of developing a simulation model in a modular way providing tools for the formulation of decision models and emulation models that can be solved independently or in an interconnected fashion. Existing operations simulations tools have focused on single-period problems or multi-period problems like UC or OPF, thus avoiding the complexities of inter-problem information flow. Despite the widespread usage of PCM in research, and industry decision-making, the definitions and algorithmic structure that enables PCM simulations is poorly represented in literature. Given that the alternatives to developing a large-scale operations simulation result in complex software projects or depend on commercial tools, the underlying optimization models used in the simulation tend to be rigidly defined. The lack of flexibility characterizing the models and operational sequences influences the type of simulations that can be conducted and in turn the scientific inquiries PowerSimulations.jl uses JuMP.jl to achieve two objectives: 1) enable a scientific approach to the simulation of large-scale power system operations; and 2) reduce model-limited choice when framing operations simulation experiments.

José Daniel Lara
National Renewable Energy Laboratory
josedaniel.lara@nrel.gov

MS167

Convex Lifting for Efficiently Escaping Saddle Points and Spurious Local Minima of Burer-Monteiro Decomposition for Low-Rank Matrix and Semidefinite Optimization

This work proposes a rapid global solver, MF-Global, for the nonconvex Burer-Monteiro (BM) decomposition form of matrix optimization problems with a regularizer to induce low-rank solutions. Through convex lifting steps, our method efficiently escapes saddle points and spurious local minima ubiquitous in noisy real-world data and is guaranteed to always converge to the global optima. Moreover, the proposed approach adaptively adjusts the rank for the BM decomposition and provably identifies the optimal rank for the decomposition problem automatically in the course of optimization through tools of manifold identification, and thus it also spends significantly less time on parameter tuning than existing matrix-factorization (MF) methods, which require an exhaustive search for this optimal rank. On the other hand, when compared to methods for solving the lifted convex form only, MF-Global leads to significantly faster convergence and much shorter running time. Experiments on real-world large-scale recommendation system problems and Euclidean distance metric embedding problems confirm that MF-Global can indeed effectively escapes spurious local solutions at which existing MF approaches are stuck, and is magnitudes faster than state-of-the-art algorithms for the lifted convex form.

Ching-Pei Lee
Institute of Statistical Mathematics in Tokyo, Japan
leechingpei@gmail.com

Ling Liang, Tianyun Tang
National University of Singapore
e0220121@u.nus.edu, ttang@u.nus.edu

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

MS167

On Solving Large-Scale Semidefinite Relaxations in Geometric Computer Vision

Geometric computer vision concerns the estimation of geometric models from sensor data and priors. Such estimation is often obtained by solving nonconvex and combinatorial polynomial optimization problems (POPs). Recently, it has been found that semidefinite programming (SDP) relaxations coming from the second-order Lasserre's hierarchy are empirically almost always exact for solving these computer vision problems. However, these relaxations are notoriously difficult to solve due to their scale and degeneracy. We propose a new algorithmic framework that blends local search using the nonconvex POP into global descent using the convex SDP. In particular, we first design a globally convergent inexact projected gradient method (iPGM) for solving the SDP that serves as the backbone. We then accelerate iPGM by taking long, but safeguarded, rank-one steps generated by fast nonlinear programming. We show that the new framework, named STRIDE, is still globally convergent for solving the SDP. To solve the iPGM subproblem of projecting a given point onto the SDP feasible set, we design a two-phase algorithm with phase one using a symmetric Gauss-Seidel based accelerated proximal gradient method (sGS-APG) to generate a good initial point, and phase two using a modified limited-memory BFGS (L-BFGS) method to obtain an accurate solution. STRIDE demonstrates state-of-the-art efficiency, scalability, and robustness in solving SDP relaxations of computer vision problems to high accuracy.

Heng Yang
Harvard University
hankyang@seas.harvard.edu

Ling Liang
National University of Singapore
Singapore
liang.ling@u.nus.edu

Luca Carlone
MIT
lcarlone@mit.edu

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

MS167

Adaptive Sieving: An Efficient Dimension Reduction Technique for Sparse Optimization Problems

In this talk, we will introduce a novel dimension reduction technique, called adaptive sieving (AS), which can accelerate algorithms for solving large-scale sparse optimization problems. We will talk about the finite convergence property of the AS with inexact solutions of the reduced subproblems generated by the AS technique. As important applications, we apply the AS technique to important statistical learning models, such as lasso, exclusive lasso, group lasso, convex clustering, and so on.

Yancheng Yuan
Hong Kong PolyU
yancheng.yuan@polyu.edu.hk

Meixia Lin
SUTD
meixia_lin@sutd.edu.sg

Defeng Sun
The Hong Kong Polytechnic University, Hong Kong
defeng.sun@polyu.edu.hk

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

MS168

Composite Bayesian Optimization for Efficient and Scalable Adaptive Experimentation

Many experimentation tasks can be cast as optimization problems with expensive or time-consuming to evaluate objective functions. Bayesian optimization has emerged as a powerful tool for tackling such problems. However, several critical applications, such as drug discovery and materials design, are out of the reach of standard approaches. In this talk, I will describe recent advances in Bayesian optimization that aim to address this challenge. In particular, I will focus on how we can exploit the composite structure of experimentation tasks to improve the efficiency and scalability of Bayesian optimization methods.

Raul Astudillo
California Institute of Technology
rastudil@caltech.edu

Peter I. Frazier
Dept. of Operations Research and Information Engineering
Cornell University
pf98@cornell.edu

MS168

Bayesian Optimization for Complex Engineering Problems

Systems in automotive, bio-pharma, aerospace, energy, have become increasingly complex, and simulation represents a standard tool to evaluate their performance with the purpose of the analysis being optimization, control, certification. Black-box optimization, that can embed simulation to perform a wide range of analyses, has attracted a lot of attention from the science and engineering communities. This talk discusses Bayesian optimization for analysis of Cyber Physical Systems. We first focus on control and acceleration of the explore/exploit process for the falsification of safety requirements without exploiting any property. I n the second part of the talk, we present algorithms that, in some form, exploit the structure of the problem at hand. Part-X is a family of partitioning-informed Bayesian optimizers that can identify regions in which the system can present safety concerns (bugs in the case a software is analyzed). In this sense, the algorithm learns the structure of the robustness function used to find falsification. We also produce a global estimate of the falsification volume. The algorithm min-BO identifies faults in systems with complex requirements that can be decomposed into a set of simpler requirements that need to be simultaneously satisfied by the system (conjunctive requirements). Finally, we show the basic ideas behind the design of algorithms that can exploit, when available, instrumented source code for the CPS to verify.

Giulia Pedrielli
Arizona State University
giulia.pedrielli@asu.edu

MS168

Role of Estimation in Global Optimization

Solving a black-box optimization problem with noisy function evaluations (e.g., simulation) often involves repeating computationally expensive function evaluations at points encountered during the algorithm. Striking a balance between exploring new design points and better estimating the function values of previously considered points is critical for computationally efficient algorithms. We describe an extension to Hesitant Adaptive Search, called Hesitant Adaptive Search with Estimation (HAS-E), that embeds a confidence interval on the estimate of the current best objective function value into the algorithm. We present a finite-time analysis of algorithm performance that combines estimation with a sampling distribution. The analysis relates the number of replications used in the estimation of the objective function to the overall performance of the algorithm. Under certain conditions, an upper bound on function evaluations is shown to be cubic in dimension. The analysis of HAS-E suggests that computational effort is better expended on sampling improving points than refining estimates of objective function values during the progress of an adaptive search algorithm. This suggests that a single observation approach such as the Single Observation Search Algorithm (SOSA) could be efficient for high dimensional problems, as long as the exploration scheme satisfies the conditions of HAS-E.

Zelda B. Zabinsky
University of Washington
Industrial and Systems Engineering
zelda@u.washington.edu

MS169

Frugal Splitting Operators: Representation, Minimal Lifting and Convergence

We consider frugal splitting operators for finite sum monotone inclusion problems, i.e., splitting operators that use exactly one direct or resolvent evaluation of each operator of the sum. A novel representation of these operators in terms of what we call a generalized primal-dual resolvent is presented. This representation reveals a number of new results regarding lifting numbers, existence of solution maps, and parallelizability of the forward and backward evaluations. We show that the minimal lifting is $n-1-f$ where n is the number of monotone operators and f is the number of direct evaluations in the splitting. Furthermore, we show that this lifting number is only achievable as long as the first and last evaluations are resolvent evaluations. In the case of frugal resolvent splitting operators, these results are the same as the results of Ryu and Malitsky–Tam. The representation also enables a unified convergence analysis and we present a generally applicable theorem for the convergence and Fejér monotonicity of fixed point iterations of frugal splitting operators with cocoercive direct evaluations. We conclude by constructing a new convergent and parallelizable frugal splitting operator with minimal lifting.

Sebastian Banert, Martin Morin
Lund University
sebastian.banert@control.lth.se,
martin.morin@control.lth.se

Pontus Giselsson
Lund University, Sweden
pontusg@control.lth.se

MS169

Branch-and-Bound Performance Estimation Programming: A Unified Methodology for Constructing Optimal Optimization Methods

We present the Branch-and-Bound Performance Estimation Programming (BnB-PEP), a unified methodology for constructing optimal first-order methods for convex and nonconvex optimization. BnB-PEP poses the problem of finding the optimal optimization method as a nonconvex but practically tractable quadratically constrained quadratic optimization problem and solves it to certifiable global optimality using a customized branch-and-bound algorithm. By directly confronting the nonconvexity, BnB-PEP offers significantly more flexibility and removes the many limitations of the prior methodologies. Our customized branch-and-bound algorithm, through exploiting specific problem structures, outperforms the latest off-the-shelf implementations by orders of magnitude, accelerating the solution time from hours to seconds and weeks to minutes. We apply BnB-PEP to several setups for which the prior methodologies do not apply and obtain methods with bounds that improve upon prior state-of-the-art results. Finally, we use the BnB-PEP methodology to find proofs with potential function structures, thereby systematically generating analytical convergence proofs.

Shuvomoy DasGupta
MIT
Cambridge, USA
sdgupta@mit.edu

MS170

Passive and Active Multi-Task Representation Learning

Representation learning has been widely used in many applications. In this talk, I will present our work which uncovers when and why representation learning provably improves the sample efficiency, from a statistical learning point of view. Furthermore, I will talk about how to actively select the most relevant task to boost the performance.

Simon S. Du
University of Washington
ssdu@cs.washington.edu

MS170

Representation Learning in Two-Layer Neural Networks

We study the first gradient step on the first-layer parameters of a width-N two-layer neural network, where the training objective is the empirical MSE loss. In the proportional asymptotic limit, and an idealized student-teacher setting, we show that the first gradient update contains a rank-1 "spike", which results in the alignment between the first-layer weights and the linear component of the teacher model. When the learning rate is small, based on a recently established Stein-CLT in the feature space, we prove a Gaussian equivalence property for the trained feature map; this allows us to show that the learned kernel improves upon the initial random features model, but cannot defeat the best linear model on the input. Whereas for sufficiently large learning rate, we prove that trained features can go beyond this "linear regime" and outperform a wide range of random features models.

Murat Erdogdu
University of Toronto
erdogdu@cs.toronto.edu

MS170

Flecs: A Federated Learning Second-Order Framework Via Compression and Sketching

Inspired by the recent work FedNL (Safaryan et al, FedNL: Making Newton-Type Methods Applicable to Federated Learning), we propose a new communication efficient second-order framework for Federated learning, namely FLECS. The proposed method reduces the high-memory requirements of FedNL by the usage of an L-SR1 type update for the Hessian approximation which is stored on the central server. A low dimensional 'sketch' of the Hessian is all that is needed by each device to generate an update, so that memory costs as well as number of Hessian-vector products for the agent are low. Biased and unbiased compressions are utilized to make communication costs also low. Convergence guarantees for FLECS are provided in both the strongly convex, and nonconvex cases, and local linear convergence is also established under strong convexity. Numerical experiments confirm the practical benefits of this new FLECS algorithm.

Artem Agafonov, Dmitry Kamzolov
Mohamed bin Zayed University of Artificial Intelligence
artem.agafonov@mbzuai.ac.ae,
kamzolov.dmitry.ae@gmail.com

Rachael Tappenden
Edinburgh
rachael.tappenden@canterbury.ac.nz

Alexander Gasnikov
Moscow Institute of Physics and Technology
gasnikov@yandex.ru

Martin Takac
Mohamed bin Zayed University of Artificial Intelligence
martin.takac@mbzuai.ac.ae

MS171

Structure Learning with Global-Local Prior-Penalty Dual

High-dimensional data with complex dependence structure is routinely observed in many areas of science and engineering, and the problem of sparse precision matrix estimation is one such methodological problem, fundamental for network estimation. Although there exist both Bayesian and frequentist approaches, it is difficult to obtain good Bayesian and frequentist properties under the same prior-penalty dual, complicating justification. In this talk, I will briefly review recent developments in precision matrix estimation using global-local shrinkage priors, the state-of-the-art Bayesian tool for sparse signal recovery. We will propose possible solutions that lead to a prior-penalty dual that offers fully Bayesian uncertainty quantification as well as computationally efficient point estimates, and discuss possible extensions.

Jyotishka Dutta
Virginia Tech
jyotishka@vt.edu

MS171

Nonparametric Learning of Stochastic Differential Equations

Inherent randomness in many temporal systems necessitates modeling of their dynamics by appropriate stochastic differential equations. Existing estimation schemes for such systems only cover the finite-dimensional inference problems where the underlying assumption is that the stochastic models are completely known up to some finite number of real-valued parameters. While this simplifying assumption works in certain cases, a vast majority of realistic systems often require estimation of certain unknown functions from the given data. These are however infinite-dimensional learning or inference problems which rely on minimization of the likelihood functional over a function space. The first part of the talk will focus on some novel results for such infinite-dimensional optimization problems, which, in particular, generalize the well-known Representer Theorem. In the second part, we will discuss how this framework can be used to develop Bayesian algorithms for the above learning problems. The talk is based on a joint work with J. Zhou and R. Mitra.

Arnab Ganguly
Louisiana State University
aganguly@math.lsu.edu

MS171

Statistical Learning and Multiscale Approximations in Chemical Reaction Networks

The talk will focus on a particular area of applied mathematics studying chemical reaction networks (CRNs) that describe creation, annihilation, combination or binding, and changes in the physical state of a collection of chemical species. Many prominent examples of intracellular dynamics, genetic switches, and dynamics of popula- tion interactions can be modelled by CRNs, where the interacting particles exhibit vastly different intrinsic scales in terms of abundance, or the reactions operate at different time scales varying over many orders of magnitude. The traditional deterministic approach to multiscale approximations used in such situations employs singular perturbation theory, often invoking Tikhonovs theorem and Fenichel theory. In this talk, I will take a stochastic viewpoint and introduce, with the help of a number of examples, a probabilistic technique to derive multiscale approximations. I will then describe how such approximations could be used for the purpose of optimization and statistical learning. The talk will be fairly nontechnical, and no prior knowledge biology is required.

Wasiur R. KhudaBukhsh
University of Nottingham
wasiur.khudabukhsh@nottingham.ac.uk

MS172

Introducing Loraine - a Solver for Low-Rank POP Relaxations

The aim of this talk is to introduce a new code for the solution of large-and-sparse linear Semidefinite Programs (SDPs) with low-rank solutions and/or low-rank data. We propose to use a preconditioned conjugate gradient method within an interior-point SDP algorithm and an efficient preconditioner fully utilizing the low-rank information. The efficiency will be demonstrated by numerical experiments using, among others, Lasserre relaxations of the MAXCUT problems and the sensor network localization problems. The code is available in Matlab and Julia, admits input in several standard SDP input formats and can be used not only for low-rank problems but for any linear SDP.

Michal Kocvara
School of Mathematics
University of Birmingham
m.kocvara@bham.ac.uk

MS172

Low-Rank Nonconvex Solver for Sum-of-Squares

This talk studies the problem of solving a univariate Sum-of-Squares program using a low-rank non-convex formulation instead of the classical convex relaxation. This formulation corresponds to the rank-r Burer-Monteiro factorization of the semidefinite program (SDP) encoding the sum of squares decomposition. We show that, if r is at least 2 then the objective function has no spurious second-order critical points, showing that all local optima are also global optima. This is in contrast to previous work showing that for general SDPs, in addition to genericity conditions, r has to be roughly the square root of the number of constraints (the degree of the polynomial) for there to be no spurious second-order critical points. We also show that for a specific scalar product, the gradient of the objective function can be computed in nearly linear time using fast Fourier transforms. Experimentally we demonstrate that this method has very fast convergence using first-order optimization algorithms with near-linear scaling to million-degree polynomials.

Benoit Legat
Massachusetts Institute of Technology
blegat@mit.edu

MS173

Sharper Rates for Separable Minimax and Finite Sum Optimization

We design accelerated algorithms with improved rates for several fundamental classes of optimization problems, including separable minimax optimization problems, finite-sum minimization problems, and separable finite-sum minimax optimization problems. Our algorithms all build upon techniques related to the analysis of primal-dual extragradient methods via relative Lipschitzness proposed recently by Cohen, Sidford, and Tian 21. Our work demonstrates the power of understanding fundamental optimization tasks from a primal-dual perspective.

Yujia Jin, Aaron Sidford, Kevin Tian
Stanford University
yujiajin@stanford.edu, sidford@stanford.edu, tiankevin@microsoft.com

MS173

Discrete Optimal Transport with Independent Marginals Is #P-Hard

We study the computational complexity of the optimal transport problem that evaluates the Wasser11 stein distance between the distributions of two K-dimensional discrete random vectors. The best known algorithms for this problem run in polynomial time in the maximum of the number of atoms of the two distributions. However, if the components of either random vector are independent, then this number can be exponential in K even though the size of the problem description scales linearly with K. We prove that the described optimal transport problem is #P-hard even if all components of the first random vector are independent uniform Bernoulli random variables, while the second random vector has merely two atoms, and even if only approximate solutions are sought. We also develop a dynamic programming-type algorithm that approximates the Wasserstein distance in pseudo-polynomial time when the components of the first random vector follow arbitrary independent discrete distributions, and we identify special problem instances that can be solved exactly in strongly polynomial time.

Karthik Natarajan
Engineering Systems and Design
Singapore University of Technology and Design
karthik_natarajan@sutd.edu.sg

Bahar Taskesen
EPFL
bahar.taskesen@epfl.ch

Soroosh Shafieezadeh-Abadeh
Carnegie Mellon University
sshafiee@andrew.cmu.edu

Daniel Kuhn
École Polytechnique Fédérale de Lausanne
daniel.kuhn@epfl.ch

MS173

A Max-Min-Max Algorithm for Robust Optimization

Robust optimization is a popular paradigm for decision making under uncertainty and finds applications in a wide range of areas. Existing approaches for solving robust optimization, such as the reformulation approach and the cutting-plane method, do not scale well with the problem size. The purpose of the talk is to present a new algorithm for solving large-scale robust optimization problems. More specifically, based on a max-min-max perspective, a tailor-made first-order algorithm for robust optimization has been devised. The algorithm operates directly on the functions and uncertainty sets, through their (sub-)gradient and projection oracles, respectively. Therefore, it can effectively exploit problem structures (e.g., low-rankness or sparsity of matrices, strong convexity or smoothness of functions and sets) to speed up the algorithm. Theoretically, we prove under mild conditions that the algorithm enjoys an iteration complexity of $O(\epsilon^{-1})$. Experiment results on large-scale instances will also be presented to illustrate the remarkable performance of the proposed algorithm.

Man Chung Yue
Musketeers Foundation Institute of Data Science
The University of Hong Kong
mcyue@hku.hk

Kai Tu
The University of Hong Kong
mailto:kaitu_02@163.com

Zhi Chen
City University of Hong Kong
chenzhi.james@gmail.com

MS174

Liquefied Natural Gas Shipping Logistics and Inventory Management

Liquefied natural gas (LNG) is natural gas that has been cooled to a liquid state which is about 600 times smaller than its volume in its gaseous state. This allows economical transportation over long distances. Every year planners are required to come up with an annual delivery plan while satisfying operational and contractual requirements. While operating the delivery plan, planners need to re-optimize the operating schedule to recover from disruptions due to unexpected vessel failures or weather conditions and production forecast changes, exploit changing market conditions such as spot trading opportunities and their impact on the LNG supply chain. At ExxonMobil we have developed an optimization software that maximizes profitability in LNG shipping routing planning and inventory management (MIRP) by reducing shipping cost while maximizing revenue in an illiquid market and ensuring operability. One mathematical optimization model for in LNG shipping routing planning and inventory management is founded upon the concept of a time-expanded multi-commodity network flow formulation. LNG-MIRP for real-world business cases require tailored-solution approaches such as matheuristics employing local neighborhood searches.

Mustafa Kilinc
ExxonMobil
mustafa.kilinc@exxonmobil.com

Jose Santiago Rodriguez Osorio
ExxonMobil, Spring, TX
jose.s.osorio@exxonmobil.com

Kevin C. Furman, Sebastian Terrazas Moreno
ExxonMobil
kevin.c.furman@exxonmobil.com, sebastian.t.moreno@exxonmobil.com

MS174

Parallel Local Search Heuristics For Lng Maritime Inventory Routing Problems

Natural gas is an essential resource towards transitioning to a sustainable energy market. To transport natural gas around the world, from producer countries to consumer countries, the natural gas is liquefied and transported over long distances. In order to optimize LNG shipping operations, it is necessary to solve a large-scale maritime inventory routing problem (MIRP). Given the rapid growth of the LNG supply chain, LNG-MIRPs require tailored-solution approaches that can overcome the course of dimensionality of large-scale MIRPs. To address these challenges, ExxonMobil has developed a best-in-class LNG shipping optimization software with the goal to optimize LNG scheduling economics to reduce cost, ensure operability, mitigate and quickly recover from disruptions, and capture value opportunities. In this talk, we will present deterministic and non-deterministic parallel large-neighborhood search algorithms implemented as part of the software. We characterize the performance of the algorithms using metrics that describe accuracy, speed-up, and processor utilization, and we will present results over a set of real-world LNG-MIRPs.

Jose S. Rodriguez
ExxonMobil Upstream Research Company
jose.s.osorio@exxonmobil.com

Mustafa Kilinc, Sebastian Terrazas Moreno, Kevin C. Furman
ExxonMobil
mustafa.kilinc@exxonmobil.com, sebastian.t.moreno@exxonmobil.com,
kevin.c.furman@exxonmobil.com

MS174

Dimensionality Reduction Heuristics for Lng Maritime Inventory Routing Problems

Natural gas is one of the most environmentally friendly sources of energy widely available. It plays an important role in the generation of electric power in combination with renewable sources. Liquefied Natural Gas (LNG) is steadily becoming a common mode for commercializing natural gas. The logistics of LNG shipping operations can be modeled as Maritime Inventory Routing problems (MIRPs). This type of problems can be complicated by long time horizons, resulting in MIRPs of high dimensionality. ExxonMobil has developed a best-in-class LNG shipping optimization software with the goal to optimize LNG scheduling economics to reduce cost, ensure operability, mitigate and quickly recover from disruptions, and capture value opportunities. An important functionality in the toolset of this software is the capability to define groups of LNG Vessels. These groups help reduce the dimensionality of the MIRP problem and can reduce computational time significantly. In this talk we will present the Vessel Class heuristic and discuss computational results over a set of real-world LNG-MIRPs.

Sebastian Terrazas Moreno
ExxonMobil
sebastian.t.moreno@exxonmobil.com

MS175

Optimal Boundary Control of the Thin Film Equation

In this work, we consider the problem of controlling the fluid film profile of a liquid that is flowing down a vertical fiber. The fluid dynamics is governed by a fourth-order partial differential equation (PDE) derived from the Navier-Stokes equation. Classical theory suggests that such flows are subject to the Rayleigh-Plateau instability that leads the film to rupture into droplets. In this work, we show that control of the film profile can be achieved by dynamically altering the input flux to the fluid system that appears as a boundary condition of the PDE. We use the optimal control methodology to compute the control function. This method entails solving a minimization of a given cost function over a time horizon. We formally derive the optimal control conditions, and numerically verify that subject to the domain length constraint, the thin film equation can be controlled to generate a desired film profile with a single point of actuation. Specifically, we show that the system can be driven to both constant film profiles and traveling waves of certain speeds.

Shiba Biswal
University of California, Los Angeles
sbiswal@math.ucla.edu

MS175

Adaptive Covers for Ball Mapper

Ball Mapper, introduced by Dlotko in 2019, takes as input a point-cloud X and produces as output a one-dimensional graph that captures the underlying topology of X. Unlike the traditional Mapper algorithm introduced by Singh, Mmoli, and Carlsson in 2007, Ball Mapper does not use a lens function to determine an open cover of X; instead, this cover is created by taking as input a parameter ε and generating an ε-net of balls on the points of X. To take the guesswork out of parameter selection, we use the idea of the Adaptive Mapper algorithm introduced by Chalapathi, Zhou, and Wang in 2021 to iteratively split the ε-balls in the cover. We then introduce a method to merge open sets in the split cover, resulting in a new open cover that produces a graph more robust to the initial choice of ε.

Sarah Percival
Michigan State University
perciva9@msu.edu

MS175

Project and Forget: Solving Large-Scale Metric Constrained Problems

Given a set of distances amongst points, determining what metric representation is most consistent with the input distances or the metric that captures the relevant geometric features of the data is a key step in many machine learning algorithms. In this paper, we focus on metric constrained problems, a class of optimization problems with metric constraints. In particular, we identify three types of metric constrained problems: metric nearness Brickell et al.(2008), weighted correlation clustering on general graphs Bansal et al.(2004), and metric learning Bellet et al.(2013); Davis et al.(2007). Because of the large number of constraints in these problems, however, researchers have been forced to restrict either the kinds of metrics learned or the size of the problem that can be solved. We provide an algorithm, PROJECT AND FORGET, that uses Bregman projections with cutting planes, to solve metric constrained problems with many (possibly exponentially) inequality constraints. We also prove that our algorithm converges to the global optimal solution. Additionally, we show that the optimality error (L2 distance of the current iterate to the optimal) asymptotically decays at an exponential rate. We show that using our method we can solve large problem instances of three types of metric constrained problems, out-performing all state of the art methods with respect to CPU times and problem sizes.

Rishi Sonthalia
University of California Los Angeles
rsonthal@math.ucla.edu

MS176

Dynamics-Driven Meta Algorithms for Operator Splitting Methods

Operator splitting methods like ADMM and its dual (DouglasRachford) may be described by iterative application of an operator whose fixed points allow for fast recovery of locally optimal solutions. Under light-weight assumptions, stability is equivalent to existence of a Lyapunov function that describes the associated discrete time dynamical system admitted by the repeated application of that operator. Critically, the Lyapunov function encodes structural information about both the problem and the operator. Lyapunov functions are usually hard to find, but if a practitioner had a priori knowledge—or a reasonable guess—about one's structure, they could equivalently tackle the problem by seeking to minimize the Lyapunov function directly. We introduce a class of methods that does this. Interestingly, for certain feasibility problems, circumcentered-reflection method (CRM) is an extant example therefrom. However, CRM may not lend itself well to primal/dual adaptation, for reasons we show. Motivated by the discovery of our new class, we experimentally demonstrate the success of one of its other members, implemented in a primal/dual framework. We also sample its performance for various other problems. This talk is based on the article: Scott B. Lindstrom, "Computable Centering Methods for Spiraling Algorithms and their Duals, with Motivations from the Theory of Lyapunov Functions,' *Computational Optimization and Applications*, (2022) in press.

Scott B. Lindstrom
Centre for Optimisation and Decision Science
Curtin University
scott.lindstrom@curtin.edu.au

MS176

A Convex QCQP Optimization Model for the Vertical Alignment Problem in Road Design

A new convex Quadratically-Constrained Quadratic Programming model (QCQP) is proposed. It is compared to a Mixed Integer Linear Programming (MILP) model. The QCQP model can be viewed as the limit of this MILP model. Numerical results show that for roads with less than 100 stations, the QCQP model has similar computation time as the MILP model. However, the QCQP model significantly outperforms the MILP model for other roads, in some cases finding a global optimum in minutes while the MILP model fails to find any solution in hours. The models are implemented in MATLAB using YALMIP and CPLEX.

Yves Lucet
UBC
yves.lucet@ubc.ca

MS176

Multi-Precision Quadratic Regularization Algorithm for Unconstrained Optimization with Rounding Error Monitoring

We introduce MPR2, a multi-precision version of the quadratic regularization algorithm and study its convergence properties in finite-precision arithmetic. MPR2 is a steepest descent algorithm that controls the step size via a regularization parameter, and allows to evaluate the objective function and the gradient with different floating point formats (e.g. half, single, double precision). MPR2 chooses suitable formats for objective function and gradient evaluation at every iteration in order to limit computational effort while ensuring convergence. We perform a comprehensive convergence analysis that takes all the rounding errors that occur during the algorithm execution into account. From this analysis, we derive convergence conditions that depend on the unit round-off. MPR2 enforces these conditions by monitoring the rounding errors and choosing floating point format with suitable unit round-off. The behaviour of the algorithm is illustrated and discussed over a test problem set.

Dominique H. Monnet
Department of Mathematics and Industrial Engineering
Polytechnique Montréal
dominique.monnet@polymtl.ca

Dominique Orban
GERAD and Dept. Mathematics and Industrial Engineering
Ecole Polytechnique de Montreal
dominique.orban@gerad.ca

MS177

Non-Asymptotic Superlinear Convergence of Standard Quasi-Newton Methods

In this talk, we discuss the non-asymptotic superlinear convergence rate of the Broyden class of quasi-Newton algorithms which includes the Davidon-Fletcher-Powell (DFP) method and the Broyden-Fletcher-Goldfar-Shanno (BFGS) method. Specifically, we present a finite-time (non-asymptotic) convergence analysis for Broyden quasi-Newton algorithms under the assumptions that the objective function is strongly convex, its gradient is Lipschitz continuous, and its Hessian is Lipschitz continuous at the optimal solution. We show that in a local neighborhood of the optimal solution, the iterates generated by both DFP and BFGS converge to the optimal solution at a superlinear rate of $(1/k)^{k/2}$, where k is the number of iterations. We also extend our results to the case that the objective function is self-concordant. Numerical experiments on several datasets confirm the tightness of our explicit convergence rate bounds. We further compare our established convergence results with other concurrent results on the non-asymptotic superlinear convergence rate of quasi-Newton methods.

Qiujiang Jin
University of Texas at Austin
qiujiang@austin.utexas.edu

Aryan Mokhtari
Assistant Proessor UT Austin
mokhtari@austin.utexas.edu

MS177

Fast Projection Onto Convex Smooth Constraints

The Euclidean projection onto a convex set is an important problem that arises in numerous constrained optimization tasks. Unfortunately, in many cases, computing projections is computationally demanding. In this work, we focus on projection problems where the constraints are smooth and the number of constraints is significantly smaller than the dimension. The runtime of existing approaches to solving such problems is either cubic in the dimension or polynomial in the inverse of the target accuracy. Conversely, we propose a simple and efficient primal-dual approach, with a runtime that scales only linearly with the dimension, and only logarithmically in the inverse of the target accuracy. We empirically demonstrate its performance, and compare it with standard baselines.

Ilnura Usmanova
Technion
ilnurau@control.ee.ethz.ch

MS178

Reformulations and Rotational Transformations for Convex Minlp

This talk focuses on how reformulations and rotational transformations can be utilized to solve mixed-integer nonlinear problems (MINLPs) with quadratic structures more efficiently with outer approximation type algorithms. By quadratic structures, we refer to either quadratic constraints or nonlinear constraints containing quadratic terms. In this talk we focus on MINLPs with convex nonlinear constraints. The main idea behind the reformulations and rotational transformations is to extract low dimensional convex separable expressions that can be delt with more efficiently. More efficiently in the sense that we can construct a tighter polyhedral outer approximation, and derive strong valid inequalities (cutting planes). We will for example describe how the SHOT solver utilizes eigendecomposition to perform a lifted reformulation for convex mixed-integer problems with non-separable quadratic expressions. An eigenvalue decomposition is first performed on the non-diagonal matrices defining quadratic expressions in the problem, and is used for transforming the quadratic expressions into convex additively separable constraints. The resulting additively separable constraints are then further lifted into a form where SHOT generates polyhedral outer approximations of convex quadratic univariate functions. The reformulations have been integrated into SHOT's automatic problem reformulation functionality.

Jan Kronqvist
KTH Royal Institute of Technology
jankr@kth.se

MS178

Convexification Techniques For Signomial Functions in Mixed-Integer Nonlinear Programming

The topic of this presentation is how to utilize single-variable exponential and power transformations in combination with linearization techniques to solve mixed-integer nonlinear optimization (also called mixed-integer nonlinear programming – MINLP) problems containing signomial functions. Signomial functions are sums of terms, where each term is a product of power functions. This general function class include common nonconvexities in optimization problems such as bilinear and trilinear terms. The lifting reformulations convexify the nonconvex signomial terms in an extended variable space by moving the nonconvexities to nonlinear equality constraints connecting the original variables with the auxiliary transformation variables. In a second step, the nonlinear equality constraints are approximated using piecewise linear functions (PLFs) and as these are updated with more breakpoints, the solution to the resulting overestimated convex MINLP problem converges to that of the original nonconvex problem. Finally, it will be explained how the signomial transformations have been integrated into the Supporting Hyperplane Optimization Toolkit (SHOT), which is a state-of-the-art open-source solver for convex MINLP.

Andreas Lundell
Abo Akademi University
andreas.lundell@abo.fi

MS180

Local Algebro-Geometric Properties for Complexity Estimates in Polynomial Optimization

In Polynomial Optimization, the Lasserre-Parrilo hierarchies are used to compute lower approximations of the global minimum of a polynomial objective funciton f over a basic closed semialgebraic set S. In this talk, we describe local algebro-geometric parameters attached to f and the description of S that are used to study convergence properties of these hierarchies to the minimum of f. In particular, we investigate: - asymptotic convergence rates, using Lojasiewicz inequalities associated to f and the description of S; - finite converge properties (in particular flat truncation), using radicality and regularity properties of an extended quadratic module defining the minimizers of f on S. Based on joint works with B. Mourrain and A. Parusinski.

Lorenzo Baldi
INRIA, Sophia Antipolis
lorenzo.baldi@inria.fr

Bernard Mourrain
INRIA Sophia Antipolis
Bernard.Mourrain@inria.fr

Adam Parusinski
Université Côte d'Azur
adam.parusinski@univ-cotedazur.fr

MS180

Polyhedral Homotopy Method for Nash Equilibrium Problem

In this talk, we discuss the problem of finding generalized Nash equilibria (GNE) from the viewpoint of sparse polynomials. To obtain optimality conditions for GNE, we consider the Karush-Kuhn-Tucker (KKT) system using the Lagrange multiplier. We discuss that if all objectives and constraints polynomials are generic, the number of solutions of the KKT system equals its mixed volume. So the polyhedral homotopy method can be optimal for finding GNEs. Lastly, results for numerical experiments will be given.

Kisun Lee
University of California, San Diego, U.S.
kil004@ucsd.edu

Xindong Tang
The Hong Kong Polytechnic University
xindong.tang@polyu.edu.hk

MS180

How do exponential size solutions arise in Semidefinite Programming?

A curious pathology of SDPs is illustrated by a famous example of Khachiyan: feasible solutions of SDPs may need exponential space to even write down. Understanding such large solutions is a key to solve one of the most important open problems in optimization theory: can we decide feasibility of SDPs in polynomial time? We first address the question: how common are such large solutions in SDPs? We prove that they are quite common: a linear change of variables transforms every strictly feasible SDP into a Khachiyan type SDP in which the leading variables are large. As to "how large, that depends on the singularity degree, a ubiquitous parameter of SDPs. Finally, we give a partial "yes answer to the question: can we represent exponential size solutions in a compact fashion, in polynomial space?

Gabor Pataki
University of North Carolina, at Chapel Hill
gabor@unc.edu

MS181

The Variable Projected Augmented Lagrangian Method

Inference by means of mathematical modeling from a collection of observations remains a crucial tool for scientific discovery and is ubiquitous in application areas such as signal compression, imaging restoration, and supervised machine learning. With ever-increasing model complexities and growing data size, new specially designed methods are urgently needed to recover meaningful quantities of interest. We consider the broad spectrum of linear inverse problems where the aim is to reconstruct quantities of interest with a sparse representation on some vector space; often solved using the (generalized) least absolute shrinkage and selection operator (lasso). The associated optimization problems have received significant attention, in particular in the early 2000s, because of their connection to compressed sensing and the reconstruction of solutions with favorable sparsity properties using augmented Lagrangians, alternating directions and splitting methods. We provide a new perspective on the underlying l1 regularized inverse problem by exploring the generalized lasso problem through variable projection methods. We arrive at our proposed variable projected augmented Lagrangian (vpal) method. We demonstrate on various imaging problems the computational efficiency of various imaging problems.

Matthias Chung
Emory University, U.S.
matthias.chung@emory.edu

Rosemary A. Renaut
Arizona State University
School of Mathematical and Statistical Sciences
renaut@asu.edu

MS181

Exploiting Mixed Precision Arithmetic in Image Reconstruction

Although some work has been done to exploit mixed precision computations for inverse problems, most previous work focuses on particular applications, such as image registration and image reconstruction. Extended (high) precision has been proposed for inverse problems to control significant digits to avoid the influence of rounding errors. In this talk we consider a different perspective: because we cannot expect to precisely know data, we develop and analyze solvers for general inverse problems that can take advantage of low precision speed of modern computer architectures, and which can be used in a variety of imaging applications.

James G. Nagy
Emory University
Department of Math and Computer Science
nagy@mathcs.emory.edu

MS181

Multiresolution Image Domain Sketching for Inverse Imaging Problems

An important challenge in high-dimensional inverse problems such as Computed Tomography (CT) is developing iterative solvers to find the accurate solution of regularized optimization with reduced computational cost. In this work we propose an efficient method to solve such a problem via saddle point optimization and constructing a family of deterministic multiresolution operators that we call image domain sketches. We develop a stochastic gradient algorithm (ImaSk) to solve the saddle point problem that uses at each iteration operators at different resolutions selected through a uniform or non-uniform discrete probability distribution. We demonstrate that the algorithm is converging for strongly convex regularization functions. Numerical simulations on CT show that the proposed method is effective in reducing the computational time to reach the modelled solution compared to the full resolution-based solvers.

Alessandro Perelli
University of Dundee
aperelli001@dundee.ac.uk

Carola-Bibiane Schönlieb
University of Cambridge
cbs31@cam.ac.uk

Matthias Ehrhardt
University of Bath
me549@bath.ac.uk

MS182

On Unboundedness and Infeasibility of Linear Bilevel Optimization Problems

Bilevel optimization problems are known to be challenging to solve in practice. In particular, the feasible set of a bilevel problem is in general non-convex, even for linear bilevel problems. We are concerned with better understanding the feasible set of linear bilevel programs. Specifically we seek to develop means to identify when a bilevel problem is unbounded or infeasible. In this presentation, we show that extending the well-known High Point Relaxation with lower-level dual information can be relevant to detect when a bilevel is infeasible due to unboundedness of its lower-level problem. Moreover, we present a new linear model to detect that the bilevel is unbounded in the case when that unboundedness originates from the upper-level variables alone. Furthermore, we derive sufficient conditions to guarantee bilevel boundedness. Finally, we highlight issues arising from redundancy in bilevel problems, and discuss how it differs from redundancy for its single-level relaxations.

Bárbara Rodrigues, Miguel F. Anjos
University of Edinburgh
b.c.rodrigues@sms.ed.ac.uk, anjos@stanfordalumni.org

MS182

Finding Feasible Operation Modes for Power Distribution Systems by Discrete Adjustments in Robust Optimization

In recent years the amount of distributed generators (DG), such as local wind farms or solar cells, in power distribution systems (PDS) has increased rapidly. However, the power generation of these DGs is affected by uncertainties, e.g. stemming from weather conditions. These uncertainties may threaten the grid safety and thus smart inverters were developed to stabilize the voltage. In this talk we propose a mathematical framework to optimize the operation mode of the smart inverters. The framework boils down to a robust optimization problem, where it is possible to adjust some of the variables to the realized uncertainty in the capacity of the DGs. Particularly, one obtains a three-level optimization problem with discrete decisions on the lowest level. In order to address this problem, we formulate a mixed-integer program, that provides feasible solutions for this very challenging problem.

Jan Rolfes, Jan Kronqvist
KTH Royal Institute of Technology
jrolfes@kth.se, jankr@kth.se

MS182

A Lagrange Multiplier Expression Method for Bilevel Polynomial Optimization

Bilevel polynomial optimization is a challenging problem. In this talk, Ill introduce a Lagrange multiplier expression method to solve bilevel polynomial optimization globally by using polynomial optimization relaxations. For induced polynomial optimization problems, Moment-sum-of-squares relaxations are applied to solve them efficiently.

Jiawang Nie
University of California, San Diego
njw@math.ucsd.edu

Li Wang
Department of Mathematics
University of Texas at Arlington
li.wang@uta.edu

Jane J. Ye
University of Victoria
janeye@uvic.ca

Suhan Zhong
University of California, San Diego
suzhong@tamu.edu

MS183

Learning to Branch in Mixed-Integer Programming Solvers

The expressive and computationally inexpensive bipartite Graph Neural Networks (GNN) have been shown to be an important component of deep learning based Mixed-Integer Linear Program (MILP) solvers. Recent works have demonstrated the effectiveness of such GNNs in replacing the branching (variable selection) heuristic in branch-and-bound (B&B) solvers. These GNNs are trained, offline and on a collection of MILPs, to imitate a very good but computationally expensive branching heuristic, strong branching. Given that B&B results in a tree of sub-MILPs, we ask (a) whether there are strong dependencies exhibited by the target heuristic among the neighboring nodes of the B&B tree, and (b) if so, whether we can incorporate them in our training procedure. Specifically, we find that with the strong branching heuristic, a child node's best choice was often the parent's second-best choice. We call this the "lookback" phenomenon. Surprisingly, the typical branching GNN of Gasse et al. (2019) often misses this simple "answer". To imitate the target behavior more closely by incorporating the lookback phenomenon in GNNs, we propose two methods: (a) target smoothing for the standard cross-entropy loss function, and (b) adding a Parent-as-

Target (PAT) Lookback regularizer term. Finally, we propose a model selection framework to incorporate harder-to-formulate objectives such as solving time in the final models.

Prateek Gupta
University of Oxford
pgupta@robots.ox.ac.uk

MS183

Unsupervised Learning for Combinatorial Optimization with Gradients

Using machine learning to solve combinatorial optimization (CO) problems is challenging, especially when the data is unlabeled. This work proposes an unsupervised learning framework for CO problems. Our framework follows a standard relaxation-plus-rounding approach and adopts neural networks to parameterize the relaxed solutions so that simple back-propagation can train the model end-to-end. Our key contribution is the observation that if the relaxed objective satisfies entry-wise concavity, a low optimization loss guarantees the quality of the final integral solutions. This observation significantly broadens the applicability of the previous framework inspired by Erdos' probabilistic method. In particular, this observation can guide the design of objective models in applications where the objectives are not given explicitly while requiring being modeled in prior. We evaluate our framework by solving a synthetic graph optimization problem, and two real-world applications including resource allocation in circuit design and approximate computing. Our framework largely outperforms the baselines based on nave relaxation, reinforcement learning, and Gumbel-softmax tricks.

Pan Li
Department of Computer Science
Purdue University
panli@purdue.edu

MS184

Approximation Algorithms for Stochastic Combinatorial Optimization

Stochastic problems in combinatorial optimization are often computationally difficult, so approximation algorithms have been proposed for many such problems. I will give a short survey of some of these results and techniques, focusing on problems in network design, scheduling, and routing.

Anupam Gupta
Carnegie Mellon University
anupamg@cs.cmu.edu

MS184

On the (Unweighted) Unit-Cost Stochastic Score Classification Problem

The (Unweighted) Unit-Cost Stocastic Score Classification Problem is as follows. There are n binary-valued tests, numbered 1 through n, that can be performed on a patient. Test i has probability p_i of having outcome 1, and probability $1 - p_i$ of having outcome 0. Each test can only be performed once and the outcomes of the n tests are independent. The range of integers from 0 to n is divided into intervals, called "score classes". If all n tests are performed on a patient, and exactly k of the tests have outcome 1, then the patient is assigned to the score class (interval) which contains the number k. It may be possible to determine the score class to which a patient will be assigned before performing all n tests, if the patient's score class will be the same regardless of the outcomes of the tests not yet performed. Given the probabilities p_i associated with the n tests, the problem is to determine the optimal order in which to sequentially perform the tests, so as to minimize the expected number of tests performed before the patient's score class can be determined. (This problem is essentially equivalent to the "Stochastic Boolean Function Evaluation" problem for Symmetric Boolean Functions, with unit costs.) We survey results on the (Unweighted) Unit-Cost Stochastic Score Evaluation Problem and describe open problems.

Lisa Hellerstein
NYU
lisa.hellerstein@nyu.edu

MS184

Stochastic Minimum Norm Combinatorial Optimization

We undertake a systematic study of designing approximation algorithms for a wide class of stochastic-optimization problems with norm-based objective functions. We introduce the model of stochastic minimum-norm combinatorial optimization, wherein the costs involved are random variables with given distributions, and we are given a monotone, symmetric norm f. Each feasible solution induces a random multidimensional cost vector whose entries are independent random variables, and the goal is to find a solution that minimizes the expected f-norm of the induced cost vector. This is a very broad class of objectives, containing all ℓ_p norms, as also Top-ℓ norms (sum of ℓ largest coordinates in absolute value), which enjoys various closure properties. Our chief contribution is a framework for designing approximation algorithms for stochastic minimum-norm optimization, with two key components: (i) A reduction showing that one can control the expected f-norm by simultaneously controlling a (small) collection of expected Top-ℓ norms; and (ii) Showing how to tackle the minimization of a single expected Top-ℓ-norm by leveraging techniques used to deal with minimizing the expected maximum, circumventing the difficulties posed by the non-separable nature of Top-ℓ norms. We apply our framework to obtain strong approximation guarantees for two concrete, well-motivated problem settings: stochastic load balancing and stochastic spanning tree.

Chaitanya Swamy, Sharat Ibrahimpur
University of Waterloo
cswamy@uwaterloo.ca, sharat.ibrahimpur@uwaterloo.ca

MS185

Learning Classification Trees Robust to Distribu-

tion Shifts

We consider the problem of learning classification trees that are robust to distribution shifts between training and testing/deployment data. This problem arises frequently in high stakes settings such as public health and social work where data is often collected using self-reported surveys which are highly sensitive to e.g., the framing of the questions, the time when and place where the survey is conducted, and the level of comfort the interviewee has in sharing information with the interviewer. We propose a method for learning optimal robust classification trees based on mixed-integer robust optimization technology. In particular, we demonstrate that the problem of learning an optimal robust tree can be cast as a single-stage mixed-integer robust optimization problem with a highly nonlinear and discontinuous objective. We reformulate this problem equivalently as a two-stage linear robust optimization problem for which we devise a tailored solution procedure based on constraint generation. We evaluate the performance of our approach on numerous publicly available datasets, and compare the performance to a regularized, non-robust optimal tree. We show an increase of up to 14.16% in worst-case accuracy and of up to 4.72% in average-case accuracy across several datasets and distribution shifts from using our robust solution in comparison to the non-robust one.

Nathan Justin, Sina Aghaei
University of Southern California
njustin@usc.edu, saghaei@usc.edu

Andres Gomez
University of Southern California
University of Southern California
gomezand@usc.edu

Phebe Vayanos
University of Southern California
phebe.vayanos@usc.edu

MS186

Low-rank Spectral Optimization via Gauge Duality

Various applications in signal processing and machine learning give rise to highly structured spectral optimization problems characterized by low-rank solutions. Two important examples that motivate this work are optimization problems from phase retrieval and from blind deconvolution, which are designed to yield rank-1 solutions. An algorithm is described that is based on solving a certain constrained eigenvalue optimization problem that corresponds to the gauge dual which, unlike the more typical Lagrange dual, has an especially simple constraint. The dominant cost at each iteration is the computation of rightmost eigenpairs of a Hermitian operator. A range of numerical examples illustrate the scalability of the approach. (Joint work with Ives Macdo.)

Michael Friedlander
University of British Columbia, Canada
michael.friedlander@ubc.ca

MS186

Sparse Matrix Factorization from An Optimization Point of View

Approximating a dense matrix by a product of sparse factors is a fundamental problem for many signal processing and machine learning tasks. It can be formulated as a constrained optimization problem and decomposed into two subproblems: finding the positions of the non-zero coefficients in the sparse factors, and determining their values. While the first step is usually seen as the most challenging one due to its combinatorial nature, this talk focuses on the second step, referred to as sparse matrix approximation with fixed support (FSMF). We show the NP-hardness of FSMF, we present a non-trivial family of support constraints making the FSMF problem practically tractable with a fast gradient-free algorithm, and we investigate the landscape of the FSMF optimization formulation, proving the absence of spurious local valleys and spurious local minima.

Tung Q. Le
ENS Lyon
quoc-tung.le@ens-lyon.fr

Elisa Riccietti
Ecole Normale Supérieure (ENS) Lyon
LIP
elisa.riccietti@ens-lyon.fr

Remi Gribonval
INRIA Rennes - Bretagne Atlantique
remi.gribonval@inria.fr

MS186

Nonnegative Matrix Factorization in the l1 norm for Sparse Data

In this work, we consider the nonnegative matrix factorization (NMF) problem using the component-wise ℓ_1-norm to measure the error. Although this model has been previously considered, no algorithm has been designed that scales well with sparse input matrices. We propose, for the first time, a method that scales linearly in the number of non-zero entries of the input matrix. Interestingly, we explain why this is only possible for NMF, that is, the nonnegativity constraints provide us with a unique opportunity to achieve such a speed up. The algorithm is based on a block coordinate descent scheme as such methods have proved to be the most effective for NMF. This allows us to apply ℓ_1-NMF to large sparse data sets, such as document data sets for which using the ℓ_1-norm is arguably a more reasonable model.

Arnaud Vandaele, Nicolas Gillis
Université de Mons
Department of Mathematics and Operational Research
arnaud.vandaele@umons.ac.be, nico-

las.gillis@umons.ac.be

MS187

Gradient-based Local Search Algorithms for Black-box Quantile Optimization

We consider quantile optimization problems under a general black-box setting. We propose two new iterative multi-time scale stochastic approximation type of algorithms. The first algorithm uses an appropriately modified finite-difference-based gradient estimator that requires 2d+1 samples of the black-box function per iteration of the algorithm, where d is the number of decision variables. The second algorithm employs a simultaneous-perturbation-based gradient estimator that uses only three samples for each iteration regardless of problem dimension. Under appropriate conditions, we show the almost sure convergence of both algorithms and establish their rates of convergence. Numerical results are also reported to illustrate and compare the performance of the algorithms with alternative methods.

Jiaqiao Hu, Meichen Song
Stony Brook University, New York
jiaqiao.hu.1@stonybrook.edu,
meichen.song@stonybrook.edu

Michael Fu
University of Maryland,
mfu@umd.edu

MS187

Meta-Control: Paradigm of Rule-Based Control

We present an algorithmic technology for solving optimization problems described by rules, as opposed to traditional mathematical expressions. The algorithmic approach converts the rules to a constrained variational model through construction of a Lagrangian. In place of a traditional objective function, the approach constructs an ideal Lagrangian from rules. The gauge optimization transforms the constrained Lagrangian to be as close as possible to the ideal Lagrangian. The approach is based on invariance theory, due to Rund-Trautman and Noether, that defines the gauge. The gauge allows us to propagate the solution in the variational domain in finite time. The computation is linear in the number of decision variables. The variational domain has stochastic characteristics which capture uncertainty due to incompleteness of the rules and noisy sensory data. The gauge transformation leads to a Finsler geodesic turnpike between the current state and a desired future state. An inverse transformation gives the optimal solution in the rule-based domain. Two applications will be briefly discussed; nanotechnology and power systems. Optimization is an essential component for applications with dynamic rule-based systems.

Wolf Kohn
CrowdPoint
wolf.kohn3@gmail.com

Zelda B. Zabinsky
University of Washington
Industrial and Systems Engineering
zelda@u.washington.edu

MS187

Probabilistic Branch and Bound with Importance Sampling for Noisy Black-box Optimization

Probabilistic Branch and Bound (PBnB) approximates a level set of a noisy black-box function by branching subregions and statistically classifying them as maintained (inside the level set) or pruned (no intersection with the level set). We propose to use supervised machine learning models as surrogate models to guide an importance sampling distribution for improved efficiency. To allocate the samples efficiently, we consider variations of importance sampling probabilities, constructed to be proportional to the incumbent solution, model uncertainty, and the combination of both. Model regularization is also adopted to reduce the size of the surrogate model representing high dimensional problems. We present variations of PBnB with importance sampling and model regularization, and derive finite-time performance analyses in terms of incorrect pruning and maintaining of subregions of the solution space. Numerical experiments on benchmark problems are presented.

Pariyakorn Maneekul
University of Washington
parim@uw.edu

Zelda B. Zabinsky
University of Washington
Industrial and Systems Engineering
zelda@u.washington.edu

MS188

DouglasRachford Algorithm for Control-constrained Minimum-energy Control Problems

The DouglasRachford algorithm has been applied to many optimization problems due to its simplicity and efficiency but the application of this algorithm to optimal control is less common. In this talk we utilize this method to solve control-constrained linear-quadratic optimal control problems. Instead of the traditional approach where we discretize the problem and solve it using large-scale finite-dimensional numerical optimization techniques we split the problem in two and use projection methods to find a point in the intersection of the solution sets of these two subproblems hence giving the solution to the original problem. In a 2022 preprint we provide numerical results and comparisons for various other projection methods as well as the DouglasRachford algorithm.

Bethany Caldwell
University of South Australia, Australia
bethany.caldwell@mymail.unisa.edu.au

Regina Burachik
University of Southern Australia
regina.burachik@unisa.edu.ar

Yalcin Kaya

University of South Australia
yalcin.kaya@unisa.edu.au

MS188

On the Splitting Algorithm with Alternated Inertia

In this talk, we propose a simple modification of the operator splitting algorithm for finding the zero of the sum of monotone operators in a real Hilbert space. Our proposed algorithm is based on alternating inertial steps. We show some attractive properties of the proposed algorithm. Finally, we illustrate our results with numerical examples.

Shin-ya Matsushita
Akita Prefectural University
matsushita@akita-pu.ac.jp

MS188

A Proximal Subgradient Algorithm with Extrapolation for Structured Nonconvex Nonsmooth Problems

In this work, we consider a class of structured nonconvex nonsmooth optimization problems, in which the objective function is formed by the sum of a possibly nonsmooth nonconvex function and a differentiable function with Lipschitz continuous gradient, subtracted by a weakly convex function. This general framework allows us to tackle problems involving nonconvex loss functions and problems with specific nonconvex constraints, and it has many applications such as signal recovery, compressed sensing, and optimal power flow distribution. We develop a proximal subgradient algorithm with extrapolation for solving these problems with guaranteed subsequential convergence to a stationary point. The convergence of the whole sequence generated by our algorithm is also established under the widely used Kurdyka–Łojasiewicz property. To illustrate the promising numerical performance of the proposed algorithm, we conduct numerical experiments on two important nonconvex models. These include a compressed sensing problem with a nonconvex regularization and an optimal power flow problem with distributed energy resources.

Tan N. Pham
Federation University Australia
pntan.iac@gmail.com

Minh N. Dao
RMIT University
minh.dao@rmit.edu.au

Rakibuzzaman Shah, Nargiz Sultanova
Federation University Australia
m.shah@federation.edu.au, n.sultanova@federation.edu.au

Guoyin Li
The University of New South Wales
Sydney
g.li@unsw.edu.au

Syed Islam
Federation University Australia
s.islam@federation.edu.au

MS189

A Single-Loop Subgradient Method for Weakly Convex Non-Smooth Constrained Optimization

It is known that a proximal subgradient method can find a nearly ϵ-stationary point of a weakly convex non-smooth function in $O(\epsilon^{-4})$ iterations (Davis and Drusvyatskiy, 2018). However, the proximal subgradient method requires that the feasible set has a simple structure that allows an easy projection. In this work, we study the switching subgradient method (Polyak, 1967), which is a simple single-loop subgradient method that can be applied to weakly convex non-smooth optimization with nonlinear non-smooth convex constraints. Through the notion of an ϵ-feasible Moreau envelop, we show that the switching subgradient method finds a nearly ϵ-stationary point of the constrained problem in $O(\epsilon^{-4})$ iterations.

Qihang Lin, Yankun Huang
University of Iowa
qihang-lin@uiowa.edu, yankun-huang@uiowa.edu

MS189

First-Order Methods for Convex Optimization and Monotone Inclusions under Local Lipschitz Conditions

In this talk, we first consider convex optimization whose smooth components have a locally Lipschitz continuous gradient and propose a first-order method for finding an epsilon-KKT solution. We then consider monotone inclusion in which the point-valued operator is locally Lipschitz continuous and propose a primal-dual extrapolation method for finding an epsilon-residual solution. All the proposed methods are parameter free with a verifiable termination criterion and also enjoy a nearly optimal complexity. This is joint work with Sanyou Mei (University of Minnesota).

Zhaosong Lu
University of Minnesota
zhaosong@umn.edu

MS189

New Root-Finding Algorithms for Nonlinear Equations and Applications

In this talk, we first discuss a connection between the Halpern fixed-point iteration and Nesterovs accelerated schemes for a root-finding problem involving a co-coercive operator. We also study this connection for different recent schemes, including extra anchored gradient method and its variants. We show how the convergence results from one scheme can be transferred to another. Next, we develop a randomized block-coordinate algorithm for solving the above root-finding problem, which is different from existing randomized coordinate methods in optimization. Finally, we consider the applications of this randomized coordinate scheme to monotone inclusions and finite-sum monotone inclusions. The latter one can be applied to a federated learning setting.

Quoc Tran-Dinh
Department of Statistics and Operations Research
UNC-Chapel Hill, USA
quoctd@email.unc.edu

MS190

Convergence of Stochastic Gradient Descent for Analytic Target Functions

In this talk we discuss almost sure convergence of Stochastic Gradient Descent in discrete (and continuous) time for a given twice continuously-differentiable target function F. In a first step we give assumptions on the step-sizes and perturbation size to ensure convergence of the target value F and the gradient of F assuming that ∇F is locally Hlder-continuous. This result entails convergence of the iterates itself in the case where F does not possess a continuum of critical points. In a general non-convex setting with F possibly containing a rich set of critical points, convergence of the process itself is sometimes taken for granted, but actually is a non-trivial issue as there are solutions to the gradient flow ODE for smooth target functions that stay in a compact set but do not converge. Using the Lojasiewicz-inequality we give sharp bounds on the step-sizes and the size of the perturbation in order to guarantee convergence of the SGD scheme for analytic target functions. Also, we derive the convergence rate of the function value under the assumptions that F satisfies a particular Lojasiewicz-inequality with exponent in $[1/2, 1)$. Finally, we compare the discrete and continuous time results and discuss optimality of the assumptions.

Steffen Dereich
WWU Münster
steffen.dereich@uni-muenster.de

Sebastian Kassing
University of Bielefeld
skassing@math.uni-bielefeld.de

MS190

Convergence of a Normal Map-Based Prox-SGD Method for Stochastic Composite Optimization

In this talk, we present a novel stochastic normal map-based algorithm (nor-SGD) for nonconvex composite-type optimization problems and discuss its asymptotic convergence properties. We first analyze the global convergence behavior of nor-SGD and show that every accumulation point of the generated sequence of iterates is a stationary point almost surely and in an expectation sense. The obtained results hold under standard assumptions and extend the more limited convergence guarantees of nonconvex prox-SGD. In addition, based on the Kurdyka-Lojasiewicz (KL) framework and utilizing an adaptive time window mechanism, we establish almost sure convergence of the iterates and derive convergence rates that depend on the KL exponent and the step size dynamics. The techniques studied in this work can be potentially applied to other families of stochastic and simulation-based algorithms.

Andre Milzarek
The Chinese University of Hong Kong, Shenzhen
School of Data Science
andremilzarek@cuhk.edu.cn

Junwen Qiu
The Chinese University of Hong Kong, Shenzhen
junwenqiu@link.cuhk.edu.cn

MS191

Optimal Control and Optimal Design of Electromagnetic Devices

We consider the solution of (1) optimal control problems with the goal to camouflage electromagnetic sources and (2) optimal design problems with the goal to improve the performance of optical devices based on electromagnetic metamaterials. We formulate the problems as optimization problems constrained by Maxwell's equations in the time domain, resulting in extreme-scale discretizations requiring massive high-performance computing resources. We discuss the computational challenges that arise in using gradient-based optimization algorithms, and propose a domain decomposition approach with tunable accuracy and asynchronous scheduling.

Denis Ridzal
Sandia National Laboratories
dridzal@sandia.gov

Kelsey DiPietro
University of Notre Dame
kdipiet@sandia.gov

Tim Wildey
Optimization and Uncertainty Quantification Dept.
Sandia National Laboratories
tmwilde@sandia.gov

MS191

Optimization-Based Implicit Shock Tracking for Unsteady Flows

We introduce a high-order numerical method for approximating solutions of unsteady shock-dominated flows without requiring nonlinear stabilization, e.,g limiting or artificial viscosity, by tracking these features in space-time with the underlying high-order mesh. Central to the framework is a high-order discontinuous Galerkin (DG) discretization of a slab-based space-time formulation of the governing equations and an optimization problem whose solution is the nodal coordinates of a feature-aligned mesh and the corresponding DG approximation to the flow. In this sense, the framework is an implicit tracking method, which distinguishes it from methods that aim to explicitly mesh relevant features. In this talk, we formulate the proposed method as a sequence of PDE-constrained optimization problems, introduce a full-space sequential quadratic programming (SQP) solver, and demonstrate the approach with several canonical unsteady flows in one- and two-dimensions.

Matthew J. Zahr
University of Notre Dame

mzahr@nd.edu

MS192

Fast Optimistic Gradient Descent Ascent Method in Continuous and Discrete Time

In this talk we address continuous in time dynamics as well as numerical algorithms for the problem of approaching the set of zeros of a single-valued monotone and continuous operator V. The starting point of our investigations is a second order dynamical system that combines a vanishing damping term with the time derivative of V along the trajectory. Our method exhibits fast convergence rates of order $o\left(\frac{1}{t\beta(t)}\right)$ for $\|V(z(t))\|$, where $z(\cdot)$ denotes the generated trajectory and also for the restricted gap function. We also prove the weak convergence of the trajectory to a zero of V. Temporal discretizations of the dynamical system generate implicit and explicit numerical algorithms, which can be both seen as accelerated versions of the Optimistic Gradient Descent Ascent (OGDA) method for monotone operators, for which we prove that the generated sequence of iterates $(z_k)_{k\geq 0}$ shares the asymptotic features of the continuous dynamics. All convergence rate statements are last iterate convergence results; in addition to these we prove for both algorithms the convergence of the iterates to a zero of V. Numerical experiments indicate the overwhelming superiority of our explicit numerical algorithm over other methods designed to solve monotone equations governed by monotone and Lipschitz continuous operators.

Radu Ioan Bot, Ernö Robert Csetnek, Dang-Khoa Nguyen
University of Vienna
radu.bot@univie.ac.at, robert.csetnek@univie.ac.at, dang-khoa.nguyen@univie.ac.at

MS192

A splitting algorithm for structured non-convex optimization problems

In this talk, we present a splitting method for handling optimization problems which can be formulated as a sum of a non-smooth function with a certain descent property, a prox-bounded function and a concave function. In addition, we propose a line-search based method which accelerates the convergence of the method to a critical point of the problem. We test the performance of the algorithm and the acceleration technique in some highly non-convex clustering problems.

Francisco J. Aragón-Artacho
University of Alicante
francisco.aragon@ua.es

Pedro Pérez-Aros
Universidad de O'Higgins
Universidad de O'Higgins
pedro.perez@uoh.cl

David Torregrosa-Belen
University of Alicante
david.torregrosa@ua.es

MS192

A mirror inertial forward-reflected-backward splitting: Global convergence and linesearch extension beyond convexity and Lipschitz smoothness

This work investigates a Bregman and inertial extension of the forward-reflected-backward algorithm applied to structured nonconvex minimization problems under relative smoothness. To this end, the proposed algorithm hinges on two key features: taking inertial steps in the dual space, and allowing for possibly negative inertial values. Our analysis begins with studying an associated envelope function that takes inertial terms into account through a novel product space formulation. Such construction substantially differs from similar objects in the literature and could offer new insights for extensions of splitting algorithms. Global convergence and rates are obtained by appealing to the generalized concave Kurdyka-Łojasiewicz (KL) property, which allows us to describe a sharp upper bound on the total length of iterates. Finally, a linesearch extension is given to enhance the proposed method.

Hongjia Ou, Andreas Themelis
Kyushu University
ou.honjia.069@s.kyushu-u.ac.jp,
andreas.themelis@ees.kyushu-u.ac.jp

Xianfu Wang, Ziyuan Wang
University of British Columbia
shawn.wang@ubc.ca, ziyuan.wang@ubc.ca

MS193

Cubic Regularization for Nonlinear Conjugate Gradient Methods

This talk will focus on the application of cubic regularization in the context of conjugate gradient minimization (CGM) methods for nonlinear programming. Using Shanno's reformulation of CGM as a memoryless BFGS method, we derive new formulas for the regularized step direction, which can be evaluated without additional computational effort. The new step directions are shown to improve iteration counts and runtimes and reduce the need to restart the CGM.

Hande Y. Benson
Drexel University
Department of Decision Sciences
benson@drexel.edu

MS193

Black-box optimization for reserve design in biodiversity conservation

Given an allotment of land divided into parcels, government decision-makers, private developers, and conservation biologists can collaborate to select which parcels to protect, in order to accomplish sustainable ecological goals with various constraints. In this paper, we propose a mixed-integer nonlinear optimization model that considers the presence of multiple species on these parcels, subject to predatorprey relationships, crowding effects, and movement across parcels.

Cassidy K. Buhler
Department of Decision Sciences
Drexel University
cb3452@drexel.edu

Hande Y. Benson
Drexel University
Department of Decision Sciences
benson@drexel.edu

MS193

Numerical experiments with training support vector machines

The talk presents numerical experiments for training support vector machines with two nonlinear optimization methods. The first method is an augmented Lagrangian algorithm based on a fast projected gradient method, which uses only the first derivatives. Another method is a nonlinear rescaling algorithm based on Newton steps, which uses the second derivatives and solves linear systems of equations every iteration. The numerical experiments demonstrate that the training times scale up approximately as $O(n^3)$ for both algorithms, where n is the number of training examples. Some arguments explaining this phenomenon are provided.

Igor Griva
George Mason University
igriva@gmu.edu

MS194

Optimal Client Sampling for Federated Learning

It is well understood that client-master communication can be a primary bottleneck in federated learning (FL). In this work, we address this issue with a novel client subsampling scheme, where we restrict the number of clients allowed to communicate their updates back to the master node. In each communication round, all participating clients compute their updates, but only the ones with "important' updates communicate back to the master. We show that importance can be measured using only the norm of the update and give a formula for optimal client participation. This formula minimizes the distance between the full update, where all clients participate, and our limited update, where the number of participating clients is restricted. In addition, we provide a simple algorithm that approximates the optimal formula for client participation, which allows for secure aggregation and stateless clients, and thus does not compromise client privacy. We show both theoretically and empirically that for Distributed SGD (DSGD) and Federated Averaging (FedAvg), the performance of our approach can be close to full participation and superior to the baseline where participating clients are sampled uniformly. Moreover, our approach is orthogonal to and compatible with existing methods for reducing communication overhead, such as local methods and communication compression methods.

Samuel Horváth
Mohamed bin Zayed University of Artificial Intelligence
samuel.horvath@mbzuai.ac.ae

Wenlin Chen
University of Cambridge
Max Planck Institute for Intelligent Systems
https://wenlin-chen.github.io/#contact

Peter Richtarik
King Abdullah University of Science & Technology (KAUST)
Peter.Richtarik@kaust.edu.sa

MS194

Lower Bounds and Nearly Optimal Algorithms in Distributed Learning with Communication Compression

Recent advances in distributed optimization and learning have shown that communication compression is one of the most effective means of reducing communication. While there have been many results for convergence rates with compressed communication, a lower bound is still missing. Analyses of algorithms with communication compression have identified two abstract properties that guarantee convergence: the unbiased property or the contractive property. They can be applied either unidirectionally (compressing messages from worker to server) or bidirectionally. In the smooth and non-convex stochastic regime, this paper establishes a lower bound for distributed algorithms whether using unbiased or contractive compressors in unidirection or bidirection. To close the gap between this lower bound and the best existing upper bound, we further propose an algorithm, NEOLITHIC, that almost reaches our lower bound (except for a logarithm factor) under mild conditions. Our results also show that using contractive compressors in bidirection can yield iterative methods that converge as fast as those using unbiased compressors unidirectionally. We report experimental results that validate our findings.

Xinmeng Huang
University of Pennsylvania
xinmengh@sas.upenn.edu

Yiming Chen
MetaCarbon
yiming@metacarbon.vip

Wotao Yin
Alibaba Group US/DAMO Academy, U.S.
wotao.yin@gmail.com

Kun Yuan
Peking University
kunyuan@pku.edu.cn

MS194

EF21: A New, Simpler, Theoretically Better, and Practically Faster Error Feedback

Error feedback (EF), also known as error compensation, is an immensely popular convergence stabilization mechanism in the context of distributed training of supervised machine learning models enhanced by the use of contractive communication compression mechanisms, such as Top-k. First proposed by Seide et al (2014) as a heuristic, EF resisted any theoretical understanding until recently [Stich et al., 2018, Alistarh et al., 2018]. However, all existing analyses either i) apply to the single node setting only, ii) rely on very strong and often unreasonable assumptions, such global boundedness of the gradients, or iterate-dependent assumptions that cannot be checked a-priori and may not hold in practice. In this work we fix these deficiencies by proposing and analyzing a new EF mechanism, which we call EF21, which consistently and substantially outperforms EF in practice. Our theoretical analysis relies on standard assumptions only, works in the distributed heterogeneous data setting, and leads to better and more meaningful rates. In particular, we prove that EF21 enjoys a fast O(1/T) convergence rate for smooth nonconvex problems, beating the previous bound of $O(1/T^{2/3})$, which was shown a bounded gradients assumption. We further improve this to a fast linear rate for PL functions, which is the first linear convergence result for an EF-type method not relying on unbiased compressors.

Peter Richtarik
King Abdullah University of Science & Technology (KAUST)
Peter.Richtarik@kaust.edu.sa

Igor Sokolov
King Abdullah University of Science and Technology
igor.sokolov.1@kaust.edu.sa

Ilyas Fatkhulin
TUM
ilyas.fatkhullin@ai.ethz.ch

MS196

Certifying the Absence of Spurious Local Minima at Infinity

When searching for global optima of non-convex unconstrained optimization problems, it is desirable that every local minimum be a global minimum. This property of having no spurious local minima is true in various problems of interest nowadays, including principal component analysis, matrix sensing, and linear neural networks. However, since these problems are non-coercive, they may yet have spurious local minima at infinity. The classical tools used to analyze the optimization landscape, namely the gradient and the Hessian, are incapable of detecting spurious local minima at infinity. In this paper, we identify conditions that certify the absence of spurious local minima at infinity, one of which is having bounded subgradient trajectories. We check that they hold in several applications of interest.

Cedric Josz, Xiaopeng Li
Columbia University
cj2638@columbia.edu, xl3040@columbia.edu

MS196

Distributed Data-Parallel Training of Neural Networks At-Scale with Distributed Shampoo

Distributed Shampoo is a promising preconditioned stochastic gradient method within the Adagrad family of algorithms specifically designed for training neural networks. In this talk, I will provide a technical introduction to the distributed Shampoo algorithm and its PyTorch implementation for distributed data-parallel training. Key approximations based on DNN-specific structures enable Shampoo to outperform standard diagonal scaling methods (Adagrad, Adam, AdamW) while remaining computationally tractable. We discuss the tradeoffs between memory, efficiency, and model quality when tuning distributed Shampoo and other block-preconditioned methods for industrial settings, as well as the performance optimizations that are necessary to produce a fast, high-quality implementation. Multi-node and multi-GPU experiments demonstrate Shampoos impact on training large-scale models.

Hao-Jun Shi
Meta Platforms, Inc.
hjmshi@meta.com

Tsung-Hsien Lee
Independent Researcher
tsung.hsien.lee@gmail.com

Jose Gallego-Posado
MILA/Meta Platforms
josegp@meta.com

Shintaro Iwasaki
Meta Platforms
siwasaki@meta.com

Dheevatsa Mudigere
NVIDIA Corporation
dmudigere@nvidia.com

Michael Rabbat
Meta Platforms, Inc.
mikerabbat@fmeta.com

Kaushik Rangadurai
Meta Platforms
krangadu@meta.com

MS196

Nonconvex Sampling Through the Lens of Riemannian Geometry and Langevin MCMC

Sampling is deeply intertwined with optimization, inference, and exploration. A key question in machine learning is the design of sampling algorithms that can leverage the power of geometry. A prime example is Riemannian Langevin MCMC, an algorithm that has enjoyed remarkable empirical success, albeit without a full rigorous theoretical understanding of its convergence. This talk focuses on an exciting new result: the first non-asymptotic convergence guarantee for Riemannian Langevin MCMC for sampling from densities of the form exp(-h(x)) on Riemannian manifolds. Our analysis is based on bounding the discretization error of the geometric Euler-Murayama scheme. Combined with a contraction guarantee for the geometric Langevin diffusion under Kendall-Cranston coupling, we show that the Riemannian Langevin MCMC iterates lie within e-Wasserstein distance of the target distribution after $O(1/e^2)$ steps, which matches the known complexity of the Euclidean case. Our results apply where 'h' can be nonconvex and the manifold may have negative Ricci curvature. We also establish convergence guarantees for Riemannian Langevin MCMC in KL-diverence, assuming a suitable log-Sobolev inequality. The key innovation therein is modeling one step of MCMC as a diffusion process over the tangent space. We show that Riemannian Langevin MCMC achieves e-KL-divergence after $O(1/e)$ steps, again matching the best known complexity of the Euclidean case. We will discuss several applications of our results.

Suvrit Sra, Xiang Cheng
Massachusetts Institute of Technology
suvrit@mit.edu, chengx@mit.edu

Jingzhao Zhang
Tsinghua University
jingzhaoz@mail.tsinghua.edu.cn

Jiehao Liang
Renmin University of China, Beijing, China
liangjiehao127@gmail.com

MS197

Tracking Error Bounds for Smooth Strongly Convex Time-Varying Optimization

This presentation is on first-order online methods applied to smooth strongly convex time-varying problems, focusing first on methods with constant step-size, momentum, and extrapolation-length. Precise results for the tracking iterate error (the limit supremum of the norm of the difference between the optimal solution and the iterates) for online gradient descent are derived. The presentation then considers a general first-order framework, where a universal lower bound on the tracking iterate error is established. Furthermore, a method using "long-steps" is proposed and shown to achieve the lower bound up to a fixed constant. This method is then compared with online gradient descent for specific examples.

Liam Madden
University of British Colombia
liam@ece.ubc.ca

MS197

Regularized Primal-Dual Gradient Method for Time-Varying Constrained Optimization

In this talk, we consider a time-varying constrained optimization problem and study the synthesis and analysis of online regularized primal-dual gradient methods to track a KKT trajectory. The proposed regularized primal-dual gradient method is implemented in a running fashion, in the sense that the underlying optimization problem changes during the execution of the algorithms. In order to study its performance, we first derive its continuous-time limit as a system of differential inclusions. We then study sufficient conditions for tracking a KKT trajectory, and also derive asymptotic bounds for the tracking error (as a function of the time-variability of a KKT trajectory). Further, we also investigate the optimal choice of the parameters of the algorithm. At the end of the talk, we will briefly introduce our recent work of accelerating zeroth-order optimization by borrowing tools from extremum seeking control, and discuss possible extensions to the time-varying setting.

Yujie Tang
Peking University
yujietang@pku.edu.cn

Emiliano Dall'Anese
University of Colorado Boulder
emiliano.dallanese@colorado.edu

Andrey Bernstein
National Renewable Energy Laboratory
andrey.bernstein@nrel.gov

Steven Low
Caltech
slow@caltech.edu

Xin Chen
Massachusetts Institute of Technology
xinch512@mit.edu

Na Li
Harvard University
nali@seas.harvard.edu

MS198

The difference-of-convex algorithm, the Bregman proximal point method, and quantum conditional entropy

The difference-of-convex (DC) algorithm is a conceptually simple method for the minimization of (non)convex functions that are expressed as the difference of two convex functions. The DC algorithm can be reinterpreted as the Bregman proximal point algorithm. This viewpoint allows for the straightforward derivation of convergence guarantees for the DC algorithm. We will focus on the case where the difference-of-convex function is itself convex. In the second part of the talk, we will consider an application in quantum statistical mechanics. We will discuss a simple first-order algorithm for computing the convex conjugate and proximal operator of the conditional entropy of a bipartite quantum system.

Oisín Faust
University of Cambridge
opbf2@cam.ac.uk

MS198

Mirror Descent and Quantum Blahut-Arimoto Algorithms

Convex optimization problems arise naturally when studying the fundamental limits of communication systems, such as computing quantities known as channel capacities. The Blahut-Arimoto algorithm is a well known method to solve for classical channel capacities, and recent works have extended this algorithm to compute various quantum channel capacities. We show that the classical and quantum Blahut-Arimoto algorithms can be interpreted as mirror descent, which is a well studied generalisation of projected gradient descent for non-Euclidean geometries. Using this perspective, we show how existing tools in convex optimization can be used to guarantee bounds on convergence rates, and how the algorithm can be modified to solve a wider range of related problems in quantum information theory.

Kerry He
Monash University
kerry.he1@monash.edu

MS198

Efficient Implementation of Interior-Point Methods for Quantum Relative Entropy

Optimization over the quantum relative entropy(QRE) cone has many applications in quantum information processing, for example, calculating the key rates for quantum key distribution (QKD) protocols. Fawzi and Saunderson recently proved the self-concordance of a barrier function for the QRE cone. Our optimization software package Domain-Driven Solver (DDS) has been using this function, with some limitations on the size of the problems. In this talk, we present new theoretical and computational results to significantly improve the interior-point methods' performance for QRE. We show how duality concepts such as the dual cone and the Legendre-Fenchel conjugate can be used to improve performance. Wefinish the talk with some numerical results using DDS, which let us combine QRE constraints with many other function/set constraints.

Mehdi Karimi
Illinois State University
mkarim3@ilstu.edu

Levent Tuncel
University of Waterloo
Dept. of Combinatorics and Optimization
ltuncel@uwaterloo.ca

MS199

On Nonconvex Quadratic Programming with Ball Constraints

Inspired by the work of Kurt Anstreicher, we study semidefinite relaxations of nonconvex quadratic programming with ball constraints. We construct a new semidefinite relaxation and demonstrate empirically that it is quite strong compared to existing relaxations. The key idea is a simple lifting of the original problem into one higher dimension.

Samuel Burer
University of Iowa
Dept of Management Sciences
samuel-burer@uiowa.edu

MS199

Data and Optimization

Joe Gillis : Wait a minute, haven't I seen you before? I know your face. Norma Desmond : Get out! Or shall I call my solver? Joe Gillis : You're Optimization. You used to be in Data. Optimization used to be big. Norma Desmond : Optimization *is* big. It's the Data that got small.

Jon Lee
University of Michigan
jonxlee@umich.edu

MS200

Mars: A Second-Order Reduction Algorithm for High-Dimensional Sparse Precision Matrices Estimation

Estimation of the precision matrix is of great importance in statistical data analysis and machine learning. However, as the number of parameters scales quadratically with the dimension p, computation becomes very challenging when p is large. In this paper, we propose an adaptive sieving reduction algorithm to generate a solution path for the estimation of precision matrices under the ℓ_1 penalized D-trace loss, with each subproblem being solved by a second-order algorithm. In each iteration of our algorithm, we are able to greatly reduce the number of variables in the problem based on the Karush-Kuhn-Tucker (KKT) conditions and the sparse structure of the estimated precision matrix in the previous iteration. As a result, our algorithm is capable of handling datasets with very high dimensions that may go beyond the capacity of the existing methods. Moreover, for the sub-problem in each iteration, other than solving the primal problem directly, we develop a semismooth Newton augmented Lagrangian algorithm with global linear convergence rate on the dual problem to improve the efficiency. Theoretical properties of our proposed algorithm have been established. In particular, we show that the convergence rate of our algorithm is asymptotically superlinear. The high efficiency and promising performance of our algorithm are illustrated via extensive simulation studies and real data applications, with comparison to several state-of-the-art solvers.

Qian Li, Binyan Jiang
The Hong Kong Polytechnic University
qianxa.li@connect.polyu.hk, by.jiang@polyu.edu.hk

Defeng Sun
The Hong Kong Polytechnic University, Hong Kong
defeng.sun@polyu.edu.hk

MS200

A Feasible Method for General Convex Sdp Problems

In this work, we propose a low rank decomposition (SDPLRE) of a linearly constrained convex SDP problem that uses a squared slack variable to transfer inequality constraints into equality constraints. We prove that a random perturbation to appropriate components of the constant vector of the constraints in (SDPLRE) can eliminate all singular points with probability 1. We also propose a rank-adaptive feasible method to solve a general linearly constrained convex SDP problem with convergence guarantee. Compared with previous work, our method does not require the objective function to be twice differentiable and achieves a better iteration complexity. Numerical experiments are conducted to verify the high efficiency and robustness of our method.

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

Tianyun Tang
National University of Singapore
ttang@u.nus.edu

MS200

Convex and Nonconvex Risk-Based Linear Regression at Scale

The Value-at-Risk (VaR) and the Conditional Value-at-Risk (CVaR) are two popular risk measures to hedge against the uncertainty of data. In this talk, we will provide a computational toolbox for solving high-dimensional sparse linear regression problems under either the VaR or the CVaR measures, the former being nonconvex while the latter being convex. We address the convex CVaR linear regression problem by adopting a semismooth Newton based proximal augmented Lagrangian method. The matrix structures of the Newton systems are carefully explored to reduce the computational cost per iteration. The method is further embedded in a majorization-minimization algorithm as a subroutine to tackle the nonconvex VaR-based regression problem. We shall also discuss an adaptive sieving strategy to iteratively guess and adjust the effective problem dimension, which is particularly useful when a solution path associated with a sequence of tuning parameters is needed. Extensive numerical experiments on both synthetic and real data will be provided to demonstrate the effectiveness of our proposed methods.

Can Wu
The Hong Kong Polytechnic University
math-can.wu@polyu.edu.hk

Ying Cui
University of Minnesota
yingcui@umn.edu

Donghui Li
South China Normal University
lidonghui@m.scnu.edu.cn

Defeng Sun
The Hong Kong Polytechnic University, Hong Kong
defeng.sun@polyu.edu.hk

MS201

Applications of Adaptive Online Algorithms in Optimization

One of the simplest techniques to build stochastic convex optimization algorithms is via the notion of regret from online convex optimization: any online algorithm with a sublinear regret bound can be shown to converge on stochastic problems. In this talk, we will discuss new advances in online optimization and improvements to the conversion between online and stochastic problems that allow for significantly more refined guarantees. In addition to recovering classical results such as acceleration, these techniques allow for algorithms that provably dispense with much of the hyperparameter tuning typical in stochastic optimization. The algorithms developed in this way can thus ensure convergence guarantees that are optimally tuned for the problem at hand.

Ashok Cutkosky
Boston University
ashok@cutkosky.com

MS201

Parameter-Free Online Convex Optimization: Past, Present, and Future

Machine Learning has been described as the fuel of the next industrial revolution. Yet, current state-of-the-art learning algorithms still heavily rely on having a human in the loop in order to work properly. The most common options for the practitioners are to follow their intuitions or to exhaustively evaluate all the possibilities. However, in certain cases, we can do better. Parameter-free online optimization is a class of algorithms that does not require tuning hyperparameters, yet they achieve the theoretical optimal performance. Moreover, they often achieve state-of-the-art performance too. An example would be gradient descent algorithms completely without learning rates. In this talk, I review past and present ideas of this field. Building upon a fundamental idea connecting optimization, gambling, and information theory, I discuss selected applications of parameter-free algorithms to machine learning and statistics. Finally, we conclude with an overview of the future directions of this field.

Francesco Orabona
Boston University
fo@bu.edu

MS201

High Probability Complexity Bounds for Adaptive Optimization Methods with Stochastic Oracles

We consider a simple adaptive optimization framework for continuous optimization where the (explicit or implicit) step size in each iteration is adaptively adjusted by the estimated progress of the algorithm instead of requiring manual tuning or using a pre-specified sequence of step sizes. The framework accommodates a stochastic setting where function value, gradient (and possibly Hessian) estimates are available only through inexact probabilistic oracles. This framework is very general and encompasses stochastic variants of line search, quasi-Newton, cubic regularized Newton methods for unconstrained problems, and stochastic SQP methods for constrained problems. The probabilistic oracles capture multiple standard settings including expected loss minimization in machine learning, zeroth-order (derivative-free) optimization, and low-precision optimization. Moreover, the stochastic function, gradient (and possibly Hessian) estimates are allowed to be biased and possibly arbitrarily bad with some constant probability. Under reasonable conditions on the oracles, we derive high probability tail bounds on the iteration and sample complexity of the algorithms.

Miaolan Xie
Cornell University
mx229@cornell.edu

MS202

Strong Branching for Combinatorial Optimization

In this paper, we attempt an analysis of the performance of the strong-branching rule both from a theoretical and a computational perspective. On the positive side for strong-branching we identify vertex cover as a class of instances where this rule provably works well. In particular, for vertex cover we present an upper bound on the size of the branch-and-bound tree using strong-branching as a function of the additive integrality gap, show how the Nemhauser-Trotter property of persistency which can be used as a pre-solve technique for vertex cover is being recursively and consistently used through-out the strong-branching based branch-and-bound tree, and finally provide an example of a vertex cover instance where not using strong-branching leads to a tree that has at least exponentially more nodes than the branch-and-bound tree based on strong-branching. On the negative side for strong-branching, we identify another class of instances where strong-branching based branch-and-bound tree has exponentially larger tree in comparison to another branch-and-bound tree for solving these instances. On the computational side, we conduct experiments on various types of instances to understand how much larger is the size of the strong-branching based branch-and-bound tree in comparison to the optimal branch-and-bound tree.

Yatharth Dubey
Carnegie Mellon University
yathartd@andrew.cmu.edu

MS202

Compressing Branch-and-Bound Trees

A branch-and-bound (BB) tree certifies a dual bound on the value of an integer program. In this talk, we introduce the tree compression problem (TCP): Given a BB tree T that certifies a dual bound, can we obtain a smaller tree with the same (or stronger) bound by either (1) applying a different disjunction at some node in T or (2) removing leaves from T? We believe such post-hoc analysis of BB trees may assist in identifying helpful general disjunctions in BB algorithms. We initiate our study by considering computational complexity and limitations of TCP. We then conduct experiments to evaluate the compressibility of realistic branch-and-bound trees generated by commonly-used branching strategies, using both an exact and a heuristic compression algorithm.

linson S. Xavier
Argonne National Laboratory
axavier@anl.gov

MS202

Solid Angles of Polyhedral Cones and the Strength of Cutting Planes

Solid angle of a polyhedral cone, which indicates the proportion of space occupied by the cone, is of significant interest in integer programming. In 1969, Gomory introduced the cyclic group relaxation of IP. The facets of the cyclic group polyhedra are useful in generating cutting planes. To predict the importance or relative size of each facet, researchers have used the shooting experiment, which provides an estimate of the solid angle subtended by the facet at the origin. However, the obtained sizes were not always consistent due to randomness. We propose to use the solid angles measure, whose closed formulas were well established in two and three dimensions. For higher dimensions, Aomoto and Ribando showed that the solid angle of a simplicial cone can be computed using a multivariable hypergeometric series, provided that the cone satisfies a certain condition related to positive-definiteness. We provide decomposition methods to ensure that the positive-definite criterion is met. We present the results of our solid angle measures and compare them with those obtained from the shooting experiments. This is a joint work with Allison Fitisone.

Yuan Zhou
University of Kentucky
yuan.zhou@uky.edu

MS203

Stochastic Gradient Succeeds for Bandits

The stochastic policy gradient is a classic algorithm for bandit problems, which is naturally compatible with deep neural networks, in stark contrast to UCB and Thompson sampling, and has seen significant practical success. However, the theoretical understanding of the algorithm, especially the global convergence result has not been previously established. In this talk, we answer the question affirmatively, and will show that the stochastic gradient bandit algorithm converges to a globally optimal policy at an O(1/t) rate, even with a constant step size. The new result is achieved by establishing two novel technical findings: first, the noise of the stochastic updates in the gradient bandit algorithm satisfies a strong growth condition property, where the variance diminishes whenever progress becomes small, implying that additional noise control via diminishing step sizes is unnecessary; second, a form of weak exploration is automatically achieved through the stochastic gradient updates, since they prevent the action probabilities from decaying faster than O(1/t), thus ensuring that every action is visited enough times. These two findings can be used to show that the stochastic gradient update is already sufficient for bandits in the sense that exploration versus exploitation is automatically balanced in a manner that ensures almost sure convergence to a global optimum.

Bo Dai
Google
bodai@google.com

MS204

Using saddle graphs for the approximate synthesis of mechanisms

Approximate kinematic synthesis of mechanisms is performed by optimizing an objective function derived from motion specifications. A primary difficulty of solving such an optimization problem for approximate synthesis is that not all auxillary conditions are known a priori. To overcome this challenge, we propose using a so-called saddle graph that identifies minima of an objective function as vertices and connections between them as edges. Such a graph is interactively presented to a designer, whereby edges are continuously traversed to navigate families of design candidates in between minima. This talk will explain saddle graphs and their computation via homotopy continuation along with several examples demonstrating its use to solve approximate synthesis problems. This is joint work with Aravind Baskar and Mark Plecnik.

Jonathan Hauenstein
University of Notre Dame
Dept. of App. Comp. Math. & Stats.
hauenstein@nd.edu

MS204

Improved Effective Lojasiewicz Inequality and Applications

Given a closed and bounded semi-algebraic set $A \subset \mathbb{R}^n$ and semi-algebraic continuous functions $f, g : A \to \mathbb{R}$, such that $f^{-1}(0) \subset g^{-1}(0)$, there exist N and $c > 0$, such that the inequality (Łojasiewicz inequality) $|g(x)|^N \leq c \cdot |f(x)|$ holds for all $x \in A$. In this talk, we prove a nearly tight upper bound on the Łojasiewicz exponent for semi-algebraic functions. We show that the Łojasiewicz exponent N in this case is bounded by $(8d)^{2(n+7)}$, where d is the maximum degree of polynomials in the given semi-algebraic descriptions of A and the graphs of f and g. Unlike the previous best known bound in this setting, our bound is independent of the cardinalities of the semi-algebraic descriptions of f , g, and A. We exploited this fact to improve the best known error bounds for polynomial and non-linear semi-definite systems.

Saugata Basu
Purdue University
sbasu@math.purdue.edu

Ali Mohammad Nezhad
Department of Math Sciences, Carnegie Mellon University
anezhad@andrew.cmu.edu

MS204

Gale Diagrams for the PSD cone

Studying the facial structure of polyhedra has been a vital part of research both from a combinatorial perspective as well as in optimization through linear programming. One tool developed for understanding the facial structure of polytopes is the Gale diagram. In joint work with Venkat Chandrasekaran and Amy Wiebe, we generalize this to a larger family of convex objects, spectratopes, linear images of the spectraplex. The spectraplex is the analogue of the simplex in the cone of positive semidefinite matrices, the cone of interest in semidefinite programming. We show that our Gale diagrams can be used to understand the facial structure of these new convex bodies.

Isabelle Shankar
Portland State University
Department of Mathematics and Statistics
isabelle.shankar@pdx.edu

MS205

Neural Blind Deconvolution with Poisson Data

Blind Deconvolution problem is a challenging task in several scientific imaging domains, such as Microscopy and Medicine. The Point Spread Function inducing the blur effect on the acquired image can be solely approximately known, or just its mathematical model may be available. Blind deconvolution aims to reconstruct the image when only the recorded data is available. Recently, among the classical variational approaches, Deep Learning techniques have gained interest thanks to their impressive performances. The Deep Image Prior framework has been employed for solving this task, giving rise to the so-called Neural Blind Deconvolution. In this work, we adapt this approach to microscopy images where the predominant noise is of Poisson type, hence signal-dependent. This leads to consider the generalized Kullback-Leibler as loss function and to couple it with regularization terms on both the blur operator and the image, estimated via two different neural networks. Furthermore, an upper bound for the blur kernel, depending on the optical instrument, is added to the problem formulation. A numerical solution is obtained by an alternating Proximal Gradient Descent-Ascent procedure, which results in a Double Deep Image Prior for Poisson noise algorithm. The proposed strategy is evaluated on both synthetic and real-world images, achieving promising results and proving that the correct choice of the loss and regularization functions strongly depends on the application at hand.

Ambra Catozzi
University of Parma
ambra.catozzi@unipr.it

Valeria Ruggiero
Dipartimento di Matematica e Informatica
valeria.ruggiero@unife.it

Alessandro Benfenati
Department of Environmental Science and Policy
University of Milan
alessandro.benfenati@unimi.it

MS205

Accelerated Optimization for Dynamic MRI Re-

construction with Locally Low-Rank Regularizers

Many image reconstruction problems involve models that assume that an image or image sequence satisfies low-rank or locally low-rank properties. These models often involve optimization problems involve nuclear norms or Schatten p-norms. Many machine learning problems like robust PCA also involve such regularizers. For convex problems with a single nuclear norm or Schatten p-norm regularizer, the proximal optimized gradient method (POGM) with adaptive restart is a very effective first-order method that converges faster than FISTA and avoids the extra tuning parameters needed by augmented Lagrangian and ADMM methods. Nevertheless, such first-order proximal methods have worst-case convergence rates that are slower than the asymptotic convergence rates of smooth optimization algorithms like nonlinear CG, or methods like (limited memory) quasi-Newton algorithms that bring in second-order information. Furthermore, such first-order methods seem inapplicable to locally low-rank models that involve regularizers that sum numerous nuclear norms of overlapping patches, because such regularizers are not prox friendly. This work-in-progress explores the use of smooth approximations to nuclear norms to facilitate gradient-based optimization methods for regularizers based on global and local low-rank models.

Jeffrey Fessler, Javier Salazar Cavazos, Rodrigo Lobos, Rajesh Nadakuditi
University of Michigan
fessler@umich.edu, javiersc@umich.edu, rlobos@umich.edu, rajnrao@umich.edu

MS205

Proximal Splitting Methods for Neural Field Based Dynamic-Image Reconstruction

Neural fields are an emerging class of neural networks that act as representations of continuous objects parametrized by a finite number of weights. This parametrization can be exploited to reframe infinite-dimensional optimization problems into finite-dimensional forms. Recently, neural fields have been investigated for solving dynamic image reconstruction problems, which are optimization problems dependent on an imaging operator and regularization. Compared to pixelwise discretization methods, neural fields can represent spatiotemporal images with fewer parameters and demonstrate enhanced image quality. Directly training the weights of the neural field assumes that the imaging operator is compatible with the neural field representation. This work proposes training neural fields for image reconstruction using a proximal splitting method that separates computations involving the imaging operator and updates of the network parameters by performing proximal operations on the neural field based on subsampled gradients of the imaging operator. This splitting reduces updates of the neural fields to supervised learning problems and reduces the memory footprint, thereby allowing for neural field reconstruction of large, high-resolution dynamic images. This method is demonstrated with a 4D photoacoustic tomography inspired problem where the imaging operator is the circular Radon transform and Tikhonov regularization on spatiotemporal derivatives.

Luke Lozenski
Electrical and Systems Engineering
Washington University in St Louis
ljlozenski@wustl.edu

Mark Anastasio
Bioengineering
University of Illinois Urbana-Champaign
maa@illinois.edu

Umberto Villa
Oden Institute for Computational Engineering & Sciences
uvilla@oden.utexas.edu

MS206

Design of Poisoning Attacks on Linear Regression Using Bilevel Optimization

A poisoning attack is one of the attack types commonly studied in the field of adversarial machine learning. The adversary generating poison attacks is assumed to have access to the training process of a machine learning algorithm and aims to prevent the algorithm from functioning properly by injecting manipulative data while the algorithm is trained. In this work, our focus is on poisoning attacks against linear regression models which target to weaken the prediction power of the attacked regression model. We propose a bilevel optimization problem to model this adversarial process between the attacker generating poisoning attacks and the learner that tries to learn the best predictive regression model. We give an alternative single-level optimization problem by using the optimality conditions of the learner's problem. A commercial solver is used to solve the resulting single-level optimization problem where we generate the whole set of poisoning attack samples at once. Besides, an iterative approach that allows to determine only a portion of poisoning attack samples at every iteration is introduced. The proposed attack strategies are shown to be superior to a benchmark algorithm from the literature by carrying out extensive experiments on two realistic publicly available datasets.

Zeynep Suvak, Miguel F. Anjos
University of Edinburgh
zsuvak@exseed.ed.ac.uk, anjos@stanfordalumni.org

Luce Brotcorne
INRIA Lille Nord Europe
luce.brotcorne@inria.fr

Diego Cattaruzza
CNRS, Centrale Lille, INRIA Lille Nord Europe
diego.cattaruzza@centralelille.fr

MS206

Stackelberg Actor-Critic for Equilibrium Bidding Strategy Computation

We consider a resource allocation problem where in the first stage multiple strategic conventional and renewable generators with local constraints submit bids in the electricity market to maximize their profits, while in the second stage the market operator allocates the resource and determines the market clearing price under a set of global possibly nonconvex constraints taking into account the uncertainty of the renewable production. We propose a novel Stackelberg actor-critic reinforcement learning algorithm that sets up the problem in multi-dimensional continuous state and action spaces, enabling the generators to receive accurate feedback regarding the impact of their bidding strategies on the stochastic market clearing outcome, and update their bidding strategies by exploiting the entire action domain, also accounting for the effect of global nonconvex constraints and uncertainty of the supply. Then, we provide a local convergence guarantee for the Stackelberg actor-critic algorithm to a local Stackelberg equilibrium together with numerical results, showing that the algorithm can be effectively used to solve the allocation problem in high dimension with high share of renewables.

Hélène Le Cadre
Inria Lille
helene.le-cadre@inria.fr

Luce Brotcorne
INRIA Lille Nord Europe
luce.brotcorne@inria.fr

Pierre Gruet
EDF R&D
pierre.gruet@edf.fr

Alejandro Joffre
CMM, Universidad de Chile
ajofre@dim.uchile.cl

Francesco Morri
Inria Lille
francesco.morri@inria.fr

MS206

Network design under demand uncertainty

In this work, we present a robust network design problem (RNDP) with uncertain demands. We consider selfish users satisfying a Wardrop equilibrium. This behavior is incorporated through a bilevel problem approach, where the lower decision level determines the behavior of users while a network planner decides to reduce the costs of the network at the upper-level decision problem. We provide lower and upper bounds to the optimal value of the RNDP in the case of polyhedral uncertainty and convex cost function. In our simulations we show how the RNDP is able to provide solutions that performs close the average case while reducing significantly the cost in the worst-case scenario.

Fernando Ordóñez
Universidad de Chile
fordon@dii.uchile.cl

Victor Bucarey
Universidad de O'Higgins
victor.bucarey@uoh.cl

Azucena Orellana
Simpliroute
azucena.orellana@simplit-solutions.com

MS207

Perspectives on Plug-and-Play: Optimization, Equilibrium, and Applications

Plug-and-Play (PnP) methods make use of black-box denoisers, including CNNs, in place of proximal maps when using ADMM to solve regularized inversion problems. While the resulting algorithm converges to a high-quality fixed point in many applications, the PnP algorithm no longer solves an optimization problem. To explain this gap, we describe the connections between optimization, PnP methods, and an equilibrium formulation that generalizes the optimization form of regularized inversion. We also describe some applications that use neural networks in novel ways to produce state-of-the-art results for multiple imaging modalities.

Gregery Buzzard
Purdue University
buzzard@purdue.edu

MS207

Data-Driven Market Design in Web3

Indecipherable black boxes are common in machine learning (ML), but applications increasingly require explainable artificial intelligence (XAI). The core of XAI is to establish transparent and interpretable data-driven algorithms. This talk presents new problems arising from matching markets with blockchains, which may be designed with optimization models, and relevant tools from the learning to optimize (L2O) literature.

Howard Heaton
Typal LLC
research@typal.llc

MS207

Symbolic Learning to Optimize

Recent studies on Learning to Optimize (L2O) suggest a promising path to automating and accelerating the optimization procedure for complicated tasks. Existing L2O models parameterize optimization rules by neural networks, and learn those numerical rules via meta-training. However, they face two common pitfalls: (1) scalability: the numerical rules represented by neural networks create extra memory overhead for applying L2O models, and limits their applicability to optimizing larger tasks; (2) interpretability: it is unclear what each L2O model has learned in its black-box optimization rule, nor is it straightforward to compare different L2O models in an explainable way. To avoid both pitfalls, this paper proves the concept that we can kill two birds by one stone, by introducing the powerful tool of symbolic regression to L2O. In this paper, we establish a holistic symbolic representation and analysis framework for L2O, which yields a series of insights for learnable optimizers. Leveraging our findings, we further propose a lightweight L2O model that can be meta-trained on large-scale problems and outperformed human-designed and tuned optimizers. Our work is set to supply a brand-new perspective to L2O research.

Zhangyang Wang, Wenqing Zheng
University of Texas at Austin
atlaswang@utexas.edu, w.zheng@utexas.edu

MS208

Prophet Inequalities and Posted Prices for Combinatorial Auctions

The prophet inequality [Krengel, Sucheston, Garling, 1978] is a famous result from stopping theory. One is presented a random sequence of rewards. Whenever a reward is presented, one can choose to pick this reward and to stop the sequence or to discard this reward and to continue. This setting has the intriguing interpretation of selling one item among multiple buyers. As optimal strategies are based on thresholds, they naturally correspond to posting a price for the item. In this talk, we will discuss how this result generalizes to combinatorial auctions, in which there are m (heterogeneous) items. Depending on the assumptions, we will be able to show that despite the uncertainty we can guarantee a good fraction of the optimal social welfare, which we would be obtaining by optimally allocating the items with perfect knowledge of the buyers values and unlimited computational power.

Thomas Kesselheim
University of Bonn
thomas.kesselheim@uni-bonn.de

Paul Dütting
Google Research
duetting@google.com

Brendan Lucier
Microsoft Research
brlucier@microsoft.com

Alexander Braun
University of Bonn
alexander.braun@uni-bonn.de

MS208

Tight Guarantees for Multi-Unit Prophet Inequalities and Online Stochastic Knapsack

Prophet inequalities are a useful tool for designing online allocation procedures and comparing their performance to the optimal offline allocation. In the basic setting of k-unit prophet inequalities, the well-known procedure of Alaei (2011) with its celebrated performance guarantee of $1 - \frac{1}{\sqrt{k+3}}$ has found widespread adoption in mechanism design and general online allocation problems in online advertising, healthcare scheduling, and revenue management. Despite being commonly used to derive approximately-optimal algorithms for multi-resource allocation problems that suffer from the curse of dimensionality, the tightness of Alaei's procedure for a given k has remained unknown. In this paper we resolve this question, characterizing the optimal procedure and tight bound, and consequently improving the best-known guarantee for k-unit prophet inequalities for all $k > 1$. We also consider the more general online stochastic knapsack problem where each individual allocation can consume an arbitrary fraction of the initial capacity. Here we introduce a new "best-fit' procedure with a performance guarantee of $\frac{1}{3+e^{-2}} \approx 0.319$, which we also show is tight with respect to the standard LP relaxation. This improves the previously best-known guarantee of 0.2 for online knapsack.

Jiashuo Jiang
NYU Stern School of Business
jj2398@stern.nyu.edu

Will Ma
Columbia University
wm2428@gsb.columbia.edu

Jiawei Zhang
NYU Stern
jz31@stern.nyu.edu

MS208

Bandit Algorithms for Multi-Stage Stochastic Optimization

A standard model in Stochastic/Bayesian Optimization is that we are given probability distributions of n independent random variables and the goal is to design a policy (a multi-stage algorithm) to optimize the expected objective value. Examples include Prophet Inequalities, Pandora's Box, Stochastic Probing, and Fixed-Price Auctions. In this work we want to relax the assumption that the underlying distributions are known to the algorithm. Inspired by the field of Multi-Armed Bandits, we study stochastic problems that are played over T days: on each day t our algorithm plays some policy x_t and receives a partial feedback on the performance of x_t for the underlying unknown-but-fixed distributions. The goal is to minimize the expected regret, which is the difference in the expected value of the optimal algorithm that knows the underlying distributions vs. the average value of our algorithm (over T days) that learns the distributions from partial feedback. We will design near-optimal algorithms for Prophet, Pandora's Box, and Sequential Posted Pricing problems in this setting. This is based on joint work with Thomas Kesselheim, Yifan Wang, and Khashayar Gatmiry.

Khashayar Gatmiry
MIT
khashayargatmiry@gmail.com

Thomas Kesselheim
University of Bonn
thomas.kesselheim@uni-bonn.de

Sahil Singla, Yifan Wang
Georgia Tech
ssingla@gatech.edu, ywang3782@gatech.edu

MS209

Robust Assortment Revenue Optimization and

Satisficing

We consider an assortment optimization problem where the retailer needs to choose a set of products to offer to customers from the range of available products. By choosing the assortment, which will affect customers' purchase choice, the retailer aims to obtain a high expected revenue. We study the robust setting in the sense that we do not have exact knowledge of the distribution of customers' utility from each product. Specifically, we introduce two distributionally robust assortment formulations, robust assortment revenue optimization and robust assortment revenue satisficing. While the former uses a pre-specified ambiguity set to characterize the scope of the probability distributions of customers' utilities, the latter uses a target-driven approach to take all probability distributions into account. By using the multinomial logit model as the reference choice model for both formulations, we show that the optimal assortments exhibit a revenue-ordered property, i.e., products in the optimal assortment have higher revenue than those not in the assortment. When the assortment optimization problems have a cardinality constraint, we develop efficient methods to find optimal solutions. Theoretically, we show that by comparison with the revenue optimization approach, the revenue satisficing approach can achieve the target revenue with a higher probability and has a lower computational complexity.

Bin Hu
Chinese University of Hong Kong
bhu@se.cuhk.edu.hk

Qingwei Jin
Zhejiang University
qingweijin@zju.edu.cn

Daniel Zhuoyu Long
Chinese University of Hong Kong
zylong@cuhk.edu.hk

MS209

Optimized Dimensionality Reduction for Moment-Based Distributionally Robust Optimization

We propose an optimized dimensional reduction method for distributionally robust optimization (DRO) problems with moment-based ambiguity sets. First, we focus on the semidefinite programming (SDP) formulations of DRO problems and prove the low-rank property of these SDP problems. We transform the high-dimensional SDP problem equivalently into a bilinear SDP problem with a very small size by using low-dimensional random variables to represent the original random variables. In addition, we develop two algorithms to solve the bilinear SDP problems iteratively, where we only need to solve a low-dimensional SDP problem in each iteration. We extend the method to solve the moment-based DRO problems with chance constraints and general SDP problems. We show the effectiveness of our proposed approaches in solving two practical problems: (i) a multiproduct newsvendor problem and (ii) a multidimensional continuous knapsack problem with chance constraints. The results demonstrate that our approaches can obtain the (near-)optimal solution much more efficiently than commercial solvers and benchmark approaches.

Kai Pan
Hong Kong Polytechnic University
kai.pan@polyu.edu.hk

Shiyi Jiang
The Hong Kong Polytechnic University
shiyi.jiang@connect.polyu.hk

Jianqiang Cheng
University of Arizona
jqcheng@arizona.edu

Zuo-Jun Shen
University of California- Berkeley
maxshen@berkeley.edu

MS209

Entropic Regularization for (Wasserstein) Robust Optimization

In the first part, I will talk about distributionally robust optimization with Sinkhorn distance—a variant of Wasserstein distance based on entropic regularization. We derive convex programming dual reformulation for a general nominal distribution. Compared with Wasserstein DRO, it is computationally tractable for a larger class of loss functions, and its worst-case distribution is more reasonable. To solve the dual reformulation, we propose an efficient stochastic mirror descent method with biased gradient oracles. It finds a δ-optimal solution with computation cost $\tilde{O}(\delta^{-3})$ and memory cost $\tilde{O}(\delta^{-2})$, and the computation cost further improves to $\tilde{O}(\delta^{-2})$ when the loss function is smooth. Second, I will talk about how entropic regularization can help with the adversarial robust training. Finally, I will reformulate Sinkhorn DRO as a special bi-level formulation, which motivates new optimization algorithms.

Jie Wang
Georgia Institute of Technology
jwang3163@gatech.edu

Rui Gao
The University of Texas at Austin
Department of Information, Risk, and Operations Management
rui.gao@mccombs.utexas.edu

Yao Xie
Georgia Institute of Technology
yao.xie@isye.gatech.edu

MS210

Hierarchical Matrix Approximations of Hessians in Pde-Constrained Optimization

Hessian operators arising in inverse problems and PDE-constrained optimization play a critical role in delivering efficient and dimension-independent convergence at scale. However, the Hessian is a formally dense operator that is intractable to form explicitly for practical large scale problems. Low rank approximations of the Hessians are effective, but only in limited regimes. Hierarchical matrix representations promise to overcome the high complexity of dense representations and provide effective data structures and matrix operations that have only log-linear complexity, and provide high-quality approximations for broader classes of problems. In this talk, we describe algorithms for constructing and updating hierarchical matrix approximations of Hessians, and illustrate them on a number of representative inverse problems involving time-dependent diffusion, advection-dominated transport, frequency domain acoustic wave propagation, and low frequency Maxwell equations. The algorithms have been implemented in H2Opus, a performance-oriented package that supports a broad variety of hierarchical matrix operations on CPUs and GPUs.

George M. Turkiyyah
KAUST
George.Turkiyyah@kaust.edu.sa

MS210

Bounded Simplex-Structured Matrix Factorization: Applications and Algorithm

We propose a new low-rank matrix factorization model dubbed bounded simplex-structured matrix factorization (BSSMF). Given an input matrix X and a factorization rank r, BSSMF looks for a matrix W with r columns and a matrix H with r rows such that $X \approx WH$ where the entries in each column of W are bounded, that is, they belong to given intervals, and the columns of H belong to the probability simplex, that is, H is column stochastic. BSSMF generalizes nonnegative matrix factorization (NMF), and simplex-structured matrix factorization (SSMF). BSSMF is particularly well suited when the entries of the input matrix X belong to a given interval; for example when the rows of X represent images, or X is a rating matrix such as in the Netflix and MovieLens data sets where the entries of X belong to the interval $[1, 5]$. The simplex-structured matrix H not only leads to an easily understandable decomposition providing a soft clustering of the columns of X, but implies that the entries of each column of WH belong to the same intervals as the columns of W. We propose a fast algorithm and identifiability conditions for BSSMF, that is, we provide conditions under which BSSMF admits a unique decomposition, up to trivial ambiguities.

Olivier Vu Thanh
Université de Mons
olivier.vuthanh@umons.ac.be

Nicolas Gillis, Fabian Lecron
Université de Mons
Department of Mathematics and Operational Research
nicolas.gillis@umons.ac.be, fabian.lecron@umons.ac.be

MS211

Some Sharp Linear Interpolation Error Bounds

Linear interpolation is one of the most basic techniques for approximating functions and plays an essential role in derivative-free optimization. We studied the function approximation error of linear interpolation and extrapolation. We present several upper bounds along with the conditions under which they are sharp. All results are under the assumptions that the function has Lipschitz continuous gradient and is interpolated on an affinely independent sample set. Errors for quadratic functions and errors of bivariate linear extrapolation are analyzed in depth.

Liyuan Cao
Peking University
caoliyuan@bicmr.pku.edu.cn

Zaiwen Wen
Peking University
Beijing International Center For Mathematical Research
wenzw@pku.edu.cn

Ya-Xiang Yuan
Chinese Academy of Sciences
Inst of Comp Math & Sci/Eng, AMSS
yyx@lsec.cc.ac.cn

MS211

Adapting the Centred Simplex Gradient to Compensate for Misaligned Sample Points

The centred simplex gradient (CSG) is a popular gradient approximation technique in derivative-free optimization. Its computation requires a perfectly symmetric set of sample points and is known to provide an accuracy of $\mathcal{O}(\Delta^2)$ where Δ is the radius of the sampling set. In this talk, we consider the situation where the set of sample points is not perfectly symmetric. By adapting the formula for the CSG to compensate for the misaligned points, we define a new *Adapted-CSG*. We study the error bounds and the numerical stability of the Adapted-CSG.

Warren Hare, Yiwen Chen
University of British Columbia, Okanagan Campus
warren.hare@ubc.ca, yiwen.chen@alumni.ubc.ca

MS211

Glassbox Optimization of a Model of a Blackbox in a Hypersphere.

We present a new method for solving a broad class of bound-constrained composite minimization problems. This approach works for both smooth and nonsmooth solvers. This method is specifically designed for problems that are a chosen mapping of the outputs from a computationally expensive simulation or function. We demonstrate a method (GOOMBAH) that uses more difficult trust-region subproblems than are typically analyzed. In general, there are no guarantees that these subproblems can be (approximately) solved in finite time. However, by safeguarding this algorithm method with another algorithm, we obtain a provably convergent method that performs exceptionally well in practice, particularly when function evaluations are expensive.

Jeffrey Larson, Matt Menickelly
Argonne National Laboratory

jmlarson@anl.gov, mmenickelly@anl.gov

MS212

Inertial Proximal Block Coordinate Method for a Class of Sum-of-Ratios Problems

In this work, we consider a class of nonsmooth sum-of-ratios fractional optimization problems with block structure. This model class is ubiquitous and encompasses several important nonsmooth optimization problems in the literature. We first propose an inertial proximal block coordinate method for solving this class of problems by exploiting the underlying structure. The global convergence of our method is guaranteed under the Kurdyka–Łojasiewicz (KL) property and some mild assumptions. We then identify the explicit exponents of the KL property for three important structured fractional optimization problems. In particular, for the sparse generalized eigenvalue problem with either cardinality regularization or sparsity constraint, we show that the KL exponents are 1/2, and so, the proposed method exhibits linear convergence rate.

Minh N. Dao
RMIT University
minh.dao@rmit.edu.au

MS212

Progressive Hedging for Stochastic Integer Programming - Advances in Theory and Practise

Motivated by recent literature demonstrating the surprising effectiveness of the heuristic implementations of progressive hedging (PH) to stochastic mixed-integer programming (SMIP) problems, we provide theoretical support for the inclusion of integer variables, filling this gap between theory and practice. We provide greater insight into the following observed phenomena of PH as applied to SMIP where the intended optimal is found or at least feasible convergence is observed. We provide an analysis of a modified PH algorithm from a different viewpoint, drawing on the interleaving of (split) proximal point methods (including PH), Gauss-Seidel methods, and the utilisation of variational analysis tools. Through this analysis, we provide conditions for convergence to a feasible solution that is observed in practise. In terms of convergence analysis, we contribute insight into the convergence of proximal point like methods in the presence of integer variables via the introduction of the notion of persistent local minima and also contribute to an enhanced Gauss-Seidel convergence analysis that accommodates the variation of the objective function. We provide a practical implementation of a modified PH and demonstrate its convergent behaviour with computational experiments in line with the provided analysis.

Andrew C. Eberhard
RMIT University
andy.eberhard@rmit.edu.au

MS212

Local Maximal Monotonicity and Variational Convexity in Banach Spaces

The talk discusses two notions of locally maximal monotone set-valued mappings (see [Pennanen T, Local convergence of the proximal point algorithm and multiplier methods without monotonicity, Math. Oper. Res, 2002] and [Rockafellar RT, Variational Analysis, Springer, 1998]) on general Banach spaces (both strong and non-strong) and their relationships. To deal efficiently with these notions, a local counterpart Minty theorem is constructed in reflexive spaces. On the other hand, using the Frchet and mixed limiting coderivative (from [Mordukhovich BS, Variational Analysis and Generalized Differentiation, in I: Basic Theory, Springer, 2006], we establish full characterizations of the mentioned local concepts. From the obtained characterizations, a sum rule for locally maximal monotone operators is built. Moreover, we carry out a coderivative characterization for variational convexity of functions (see [Rockafellar RT, Variational convexity and local monotonicity of subgradient mappings, Vietnam J. Math., 2019]) in infinite-dimensional spaces.

Khoa Vu
Wayne State University
khoavu@wayne.edu

Boris Mordukhovich
Wayne State University
Department of Mathematics
boris@math.wayne.edu

Pham Duy Khanh
Ho Chi Minh City University of Education
khanhpd@hcmue.edu.vn

Thanh Phat Vo
Wayne State University
phatvt@wayne.edu

MS213

High-Dimensional Inference with Stochastic Gradient Descent

Stochastic approximation algorithms like SGD are widely used in modern data science. In order to perform statistical inference with such algorithms, it is important to not only obtain point estimates but also to obtain confidence regions quantifying the associated uncertainty. In this talk, I will discuss high-dimensional central limit theorems for linear functionals of online SGD iterates for solving overparameterized linear regression. Such results could be used for example to obtain prediction confidence intervals or entry-wise estimation confidence intervals. Our results hold in particular when the dimensionality grows polynomially/exponentially in terms of the number of iterations of SGD (or equivalently the number of observations used).

Krishnakumar Balasubramanian
University of California, Davis, U.S.
kbala@ucdavis.edu

MS213

Stochastic Multi-Level Nested Composition Opti-

mization

Nested composition optimization problems whose objective functions is a composition of exceptions, have received much attention due to their emerging applications over the past few years. Obtaining unbiased gradient estimators of the objective function is the main difficulty for this class of structured problems. In this talk, I will present a few settings under which we can provide efficient algorithms to find an approximate stationary point of the problem. We show that the sample complexities of these methods match that of the single level stochastic optimization problems and hence are optimal in terms of dependence on the target accuracy and only polynomially depend on the number of levels in the objective function. In addition, these methods can be easily applied to the online stochastic setting as they do not require any batch of samples to converge.

Krishnakumar Balasubramanian
University of California, Davis, U.S.
kbala@ucdavis.edu

Saeed Ghadimi
Department of Management Sciences
University of Waterloo
sghadimi@uwaterloo.ca

MS213

On the Performance of Gradient Tracking with Local Update(s)

We study the decentralized optimization problem where a network of n agents seeks to minimize the average of a set of heterogeneous non-convex cost functions distributedly. State-of-the-art decentralized algorithms like Exact Diffusion (ED) and Gradient Tracking (GT) involve communicating every iteration. However, communication is expensive, resource intensive, and slow. In this work, we analyze a locally updated GT method (LU-GT), where agents perform local recursions before interacting with their neighbors. While local updates have been shown to reduce communication overhead in practice, their theoretical influence has not been fully characterized. We show LU-GT has the same communication complexity as the Federated Learning setting but allows arbitrary network topologies. In addition, we prove that the number of local updates does not degrade the quality of the solution achieved by LU-GT. Numerical examples reveal that local updates can lower communication costs in certain regimes (e.g., well-connected graphs).

Edward Duc Hien Nguyen
Rice University
en18@rice.edu

Sulaiman Alghunaim
Kuwait University
sulaiman.alghunaim@ku.edu.kw

Kun Yuan
Peking University
kunyuan@pku.edu.cn

César Uribe
Rice University
cauribe@rice.edu

MS214

Almost sure convergence rates for stochastic gradient methods

We study stochastic gradient descent (SGD) and the stochastic heavy ball method (SHB, otherwise known as the momentum method) for the general stochastic approximation problem. For SGD, in the convex and smooth setting, we provide the first almost sure asymptotic convergence rates for a weighted average of the iterates. More precisely, we show that the convergence rate of the function values is arbitrarily close to $o(1/\sqrt{k})$, and is exactly $o(1/k)$ in the so-called overparametrized case. We show that these results still hold when using stochastic line search and stochastic Polyak stepsizes, thereby giving the first proof of convergence of these methods in the non-overparametrized regime. Using a substantially different analysis, we show that these rates hold for SHB as well, but at the last iterate. This distinction is important because it is the last iterate of SGD and SHB which is used in practice. We also show that the last iterate of SHB converges to a minimizer almost surely. Additionally, we prove that the function values of the deterministic HB converge at a $o(1/k)$ rate, which is faster than the previously known $O(1/k)$.

Othmane Sebbouh
ENS Paris
othmane.sebbouh@ens.fr

MS214

Stochastic Localization Methods for Convex Discrete Optimization Via Simulation

We develop and analyze a set of new sequential simulation-optimization algorithms for large-scale multi-dimensional discrete optimization via simulation problems with a convexity structure. The proposed algorithms are targeted to identify a solution that is close to the optimal solution given any precision level with any given probability. Utilizing the convexity structure, our algorithm design does not need to scan all the choices of the decision variable, but instead sequentially draws a subset of feasible decisions and uses them to "localize' near-optimal solutions to an adaptively shrinking region. To show the power of the localization operation, we first consider one-dimensional large-scale problems. We propose the shrinking uniform sampling algorithm, which is proved to achieve the optimal cost asymptotically. For multi-dimensional problems, we combine the idea of localization with subgradient information and propose a framework to design stochastic cutting-plane methods. In addition, utilizing the discrete nature of the problems, we propose a dimension reduction algorithm, which does not require prior information about the Lipschitz constant of the objective function and its simulation costs are upper bounded by a value that is independent of the Lipschitz constant. We implement the proposed algorithms on simulation optimization problems, and demonstrate better performances especially for large-scale examples.

Haixiang Zhang
UC Berkeley
haixiang_zhang@berkeley.edu

Zeyu Zheng
University of California, Berkeley
zyzheng@berkeley.edu

Javad Lavaei
University of California Berkeley
lavaei@berkeley.edu

MS215

Adaptive Randomized Sketching for Dynamic Nonsmooth Optimization

Dynamic optimization problems arise in many applications including optimal flow control, full waveform inversion, and medical imaging, where they are plagued by significant computational challenges. For example, memory is often a limiting factor on the size of problems one can solve since the evaluation of derivatives requires the entire state trajectory. Additionally, many applications employ nonsmooth regularizers such as the L^1-norm or the total variation as well as auxiliary constraints on the optimization variables. In this paper, we introduce a novel trust-region algorithm for minimizing the sum of a smooth, nonconvex function and a nonsmooth, convex function that addresses these two challenges. Our algorithm employs randomized sketching to store a compressed version of the state trajectory for use in derivative computations. By allowing the trust-region algorithm to adaptively learn the rank of the state sketch, we arrive at a provably convergent method with near optimal memory requirements. We demonstrate the efficacy of our method on a parabolic PDE-constrained optimization problem with measure-valued control variables.

Robert Baraldi
Sandia National Laboratories
Optimization & Uncertainty Quantification
rjbaral@sandia.gov

Evelyn Herberg
Interdisciplinary Center for Scientific Computing
evelyn.herberg@iwr.uni-heidelberg.de

Drew P. Kouri
Optimization and Uncertainty Quantification
Sandia National Laboratories
dpkouri@sandia.gov

Harbir Antil
George Mason University
hantil@gmu.edu

MS215

An Optimal Control Framework for Neuromorphic Imaging

Event or Neuromorphic cameras are novel biologically inspired sensors that record data based on the change in light intensity at each pixel asynchronously. They have a temporal resolution of microseconds. This is useful for scenes with fast moving objects that can cause motion blur in traditional cameras, which record the average light intensity over an exposure time for each pixel synchronously. We present a bilevel optimization based variational framework for neuromorphic imaging.

David Sayre
George Mason University
Dsayre100@gmail.com

MS215

Numerical Analysis of Optimal Control Problem Governed by Transient Stokes Equations with State Constraints Pointwise in Time

We consider the following optimal control problem:

$$\text{Minimize } \frac{1}{2}\int_0^T \int_\Omega (u(t,x) - u_d(t,x))^2 \, dx\, dt + \frac{\alpha}{2}\int_0^T \int_\Omega q(t,x)^2 \, dx\, dt$$

subject to transient Stokes equations

$$\begin{aligned} \partial_t u - \Delta u + \nabla p = q & \quad \text{in } (0,T)\times\Omega \\ \nabla\cdot u = 0 & \quad \text{in } (0,T)\times\Omega \\ u = 0 & \quad \text{on } (0,T)\times\partial\Omega \\ u(0) = u_0 & \quad \text{in } \Omega \end{aligned}$$

and to state constraints formulated pointwise in time, i.e.

$$\int_\Omega u(t,x)\cdot w(x)\, dx \le b \quad \text{for all } t\in[0,T]$$

with a given $u_d \in L^2(I; L^2(\Omega)^d)$ and $w \in L^2(\Omega)^d$. The domain $\Omega \subset \mathbb{R}^d$, $d = 2,3$ is assumed to be polygonal/polyhedral and convex. The optimality system for this problem involves a Lagrange multiplies from the space of regular Borel measures $\mu \in \mathcal{M}([0,T])$, which affects the regularity of the solution. We discretize the problem with inf-sup stable finite elements in space and with a discontinuous Galerkin method in time. For this discretization we provide quasi-optimal error estimates for the state and the control. The analysis is based on recently established error estimates for the Stokes system.

Boris Vexler
Technical University of Munich
Centre for Mathematical Sciences, M17
vexler@tum.de

Dmitriy Leykekhman
Department of Mathematics
University of Connecticut
dmitriy.leykekhman@uconn.edu

MS216

Maximal Monotonicity of Multi-marginal Subdifferentials

Motivated by multi-marginal optimal transport, we discuss the monotonicity of c-splitting sets, the multi-marginal version of classical subdifferentials. A natural question under investigation is whether maximal monotonicity of subdifferentials still holds in the multi-marginal case for classical cost functions. We discuss recent progress in the study of this problem.

Sedi Bartz
University of Massachusetts Lowell
USA
sedi_bartz@uml.edu

MS216

Extended Relative Interiors and Applications to Convex Generalized Differentiation

It is well-known that the relative interior plays a crucial role in many aspects of convex analysis and optimization in finite dimension. This is the motivation for introducing several notions of extended relative interiors in infinite dimensions. In this talk we will present a number of extended relative interiors and discuss their relationships. We will also present new results on using extended relative interiors for developing convex generalized differentiation for nonsmooth functions and set-valued mappings in infinite dimensions.

Mau Nam Nguyen
Fariborz Maseeh Department of Mathematics and Statistics
Portland State University
mnn3@pdx.edu

MS216

The Splitting Algorithms by Ryu, by Malitsky-Tam, and by Campoy Applied to Normal Cones of Linear Subspaces Converge Strongly to the Projection onto the Intersection

Finding a zero of a sum of maximally monotone operators is a fundamental problem in modern optimization and nonsmooth analysis. Assuming that the resolvents of the operators are available, this problem can be tackled with the Douglas-Rachford algorithm. However, when dealing with three or more operators, one must work in a product space with as many factors as there are operators. In groundbreaking recent work by Ryu and by Malitsky and Tam, it was shown that the number of factors can be reduced by one. A similar reduction was achieved recently by Campoy through a clever reformulation originally proposed by Kruger. All three splitting methods guarantee weak convergence to some solution of the underlying sum problem; strong convergence holds in the presence of uniform monotonicity. In this paper, we provide a case study when the operators involved are normal cone operators of subspaces and the solution set is thus the intersection of the subspaces. Even though these operators lack strict convexity, we show that striking conclusions are available in this case: strong (instead of weak) convergence and the solution obtained is (not arbitrary but) the projection onto the intersection. Numerical experiments to illustrate our results are also provided.

Shambhavi Singh
University of British Columbia Okanagan
Canada
sambha@student.ubc.ca

MS217

Recursive McCormick Linearization of Multilinear Programs

Linear programming (LP) relaxations are widely employed in exact solution methods for multilinear programs (MLP). One example is the family of Recursive McCormick Linearization (RML) strategies, where bilinear products are substituted for artificial variables, which deliver a relaxation of the original problem when introduced together with concave and convex envelopes. In this work, we introduce the first systematic approach for identifying RMLs, in which we focus on the identification of linear relaxation with a small number of artificial variables and with strong LP bounds. We present a novel mechanism for representing all the possible RMLs, which we use to design an exact mixed-integer programming (MIP) formulation for the identification of minimum-size RMLs; we show that this problem is NP-hard in general, whereas a special case is fixed-parameter tractable. Moreover, we explore the structural properties of our formulation to derive an exact MIP model that identifies RMLs of a given size with the best possible relaxation bound. Our numerical results on a collection of benchmark instances indicate that our algorithms outperform the RML strategy implemented in state-of-the-art global optimization solvers.

Carlos H. Cardonha
University of Connecticut
carlos.cardonha@uconn.edu

MS217

Learning to Accelerate the Global Optimization of Quadratically-Constrained Quadratic Programs

Many real-world applications involve the repeated solution of the same underlying quadratically-constrained quadratic program (QCQP) with slightly varying model parameters. Examples include the pooling problem with varying input qualities and the cost-efficient operation of the power grid with varying loads and renewable sources. These hard problems are solved using global optimization software whose heuristics are tuned to work well on average over a diverse test library and may perform sub-optimally for instances from a specific application. We use machine learning (ML) to accelerate the guaranteed global minimization of nonconvex QCQPs. Specifically, we consider partitioning-based mixed-integer programming relaxations for nonconvex QCQPs and propose the problem of strong partitioning to optimally partition variable domains without sacrificing global optimality. We design a local optimization method for solving this challenging max-min strong partitioning problem and replace this expensive benchmark strategy with an ML approximation for homogeneous families of QCQPs. We present a detailed computational study on randomly generated families of QCQPs using the open-source global solver Alpine. Our numerical experiments demonstrate that strong partitioning and its ML approximation significantly reduce Alpines solution time by factors of 3.5 - 16.5 and 2 - 4.5 on average and by maximum factors of 15 - 700 and 10 - 200, respectively, over the different QCQP families.

Rohit Kannan
Los Alamos National Laboratory
rohitk@alum.mit.edu

Deepjyoti Deka
Theory Division and the Center for Nonlinear Studies
Los Alamos National Laboratory
deepjyoti@lanl.gov

Harsha Nagarajan
Los Alamos National Laboratory
harsha@lanl.gov

MS217

A Novel Algorithm for Constructing Tight Quadratic Underestimators for Global Optimization

Global optimization algorithms extensively utilize polyhedral outer approximations of convex relaxations to determine lower bounds. Noting recent improvements in algorithms for solving quadratically constrained quadratic programming problems, we introduce quadratic non-linearity into outer approximation methods at the prospect of computing tighter bounds. We present an algorithm to construct quadratic outer approximators for twice differentiable convex functions and non-convex difference of convex functions. Using function and optimization problem examples extracted from global optimization benchmark libraries, we demonstrate the quality of the quadratic underestimators and evaluate the root-node lower bounds they provide.

William Strahl, Chrysanthos Gounaris
Carnegie Mellon University
wstrahl@andrew.cmu.edu, gounaris@cmu.edu

Nikolaos Sahinidis
Georgia Institute of Technology
nikos@gatech.edu

Arvind Raghunathan
Mitsubishi Electric Research Laboratories
raghunathan@merl.com

MS218

Sharp Bounds for Federated Averaging

Federated Averaging (FedAvg), also known as Local SGD, is one of the most popular algorithms in Federated Learning (FL). In this work, we resolve the question of the convergence rate of FedAvg in the most canonical setting (convex and smooth objective, homogeneous distributions, and bounded covariance). We provide a lower bound for FedAvg that matches the existing upper bound. Additionally, we establish a lower bound in a heterogeneous setting that nearly matches the existing upper bound. While our lower bounds show the limitations of FedAvg, under an additional assumption of third-order smoothness, we prove more optimistic state-of-the-art convergence results in both convex and non-convex settings. Our analysis stems from a notion we call iterate bias, which is defined by the deviation of the expectation of the SGD trajectory from the noiseless gradient descent trajectory with the same initialization.

Margalit Glasgow, Honglin Yuan, Tengyu Ma
Stanford University
margalit.r.glasgow@gmail.com, hongl.yuan@gmail.com, tengyuma@stanford.edu

MS218

Towards Optimal Communication Complexity in Distributed Non-Convex Optimization

We study the problem of distributed stochastic non-convex optimization with intermittent communication. We consider the full participation setting where M machines work in parallel over R communication rounds and the partial participation setting where M machines are sampled independently every round from some meta-distribution over machines. We propose and analyze a new algorithm that improves existing methods by requiring fewer and lighter variance reduction operations. We also present lower bounds, showing our algorithm is either *optimal* or *almost optimal* in most settings. Numerical experiments demonstrate the superior performance of our algorithm.

Kumar Kshitij Patel
Toyota Technological Institute
kkpatel@ttic.edu

Lingxiao Wang
Toyota Technological Institute at Chicago
lingxw@ttic.edu

Blake Woodworth
SIERRA, INRIA
blakewoodworth@gmail.com

Brian Bullins
Purdue University
bbullins@purdue.edu

Nati Srebro
Toyota Technological Institute at Chicago
nati@ttic.edu

MS218

Minibatch Vs Local Sgd with Shuffling: Tight Convergence Bounds and Beyond

In distributed learning, local SGD (also known as federated averaging) and its simple baseline minibatch SGD are widely studied optimization methods. Most existing analyses of these methods assume independent and unbiased gradient estimates obtained via with-replacement sampling. In contrast, we study shuffling-based variants: minibatch and local Random Reshuffling, which draw stochastic gradients without replacement and are thus closer to practice. For smooth functions satisfying the Polyak-Lojasiewicz condition, we obtain convergence bounds (in the large epoch regime) which show that these shuffling-based variants converge faster than their with-replacement counterparts. Moreover, we prove matching lower bounds showing that our convergence analysis is tight. Finally, we propose an algorithmic modification called synchronized shuffling that leads to convergence rates faster than our lower bounds in near-homogeneous settings.

Chulhee Yun
Korea Advanced Institute of Science & Technology (KAIST)
chulhee.yun@kaist.ac.kr

Shashank Rajput
University of Wisconsin-Madison
rajput.shashank11@gmail.com

Suvrit Sra
Massachusetts Institute of Technology
suvrit@mit.edu

MS220

Neural networks with linear threshold activations: structure and algorithms

In this talk I will present new results on neural networks with linear threshold activation functions. The class of functions that are representable by such neural networks can be fully characterized, and two hidden layers are necessary and sufficient to represent any function in the class. This is a surprising result in the light of recent exact representability investigations for neural networks using other popular activation functions like rectified linear units. I will discuss precise bounds on the sizes of the neural network, and I present an algorithm to solve the empirical risk minimization (ERM) problem to global optimality for these neural networks with a fixed architecture. The algorithms running time is polynomial in the size of the data sample, if the input dimension and the size of the network architecture are considered fixed constants. Finally I will present a new type architecture, the shortcut linear threshold networks, a strict superclass of the rectified linear units (ReLU) neural networks, which has several desirable theoretical properties. In particular, the ERM problem can also be solved to global optimality with a similar algorithm. Joint work: Sammy Khalife, Hongyu Cheng, Amitabh Basu

Sammy Khalife
john hopkins university
skhalif4@jhu.edu

MS220

Global Complexity Bound of a Proximal ADMM for Linearly-Constrained Non-Separable Nonconvex Composite Programming

This talk presents a dampened proximal alternating direction method of multipliers (DP.ADMM) for solving linearly-constrained nonconvex optimization problems where the smooth part of the objective function is nonseparable. Each iteration of DP.ADMM consists of: (i) a sequence of partial proximal augmented Lagrangian (AL) updates, (ii) an under-relaxed Lagrange multiplier update, and (iii) a novel test to check whether the penalty parameter of the AL function should be updated. Under some mild regularity conditions, we describe how DP.ADMM obtains a first-order stationary point of the constrained problem in $\mathcal{O}(\varepsilon^{-3})$ iterations for a given numerical tolerance $\varepsilon > 0$. One of the main novelties of the method is that the convergence of the method is obtained without requiring any rank assumptions on the constraint matrices.

William Kong
Oak Ridge National Labs
wwkong92@gmail.com

Renato Monteiro
Georgia Institute of Technology
rm88@gatech.edu

MS220

Second-Order Convergence of an Affine Scaling Polyhedral Active Set Algorithm (PASA)

Polyhedral Active Set Algorithm (PASA) provides an algorithm framework based on active set strategies to solve polyhedral constrained smooth nonlinear optimization. In this talk, we will show how the second-order Hessian information combined with affine-scaling techniques could be used to accelerate the convergence speed as well as ensure the convergence to a second-order stationary point. We will discuss these results from both theoretical convergence and practical implementation point of view. Some preliminary results show the second-order Hessian based PASA has significant numerical performance improvement over the original gradient-based PASA.

Hongchao Zhang
Department of Mathematics and CCT
Louisiana State University
hozhang@math.lsu.edu

MS221

A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization

We present and discuss numerical algorithms for the solution of dynamic inverse problems in which for each time point, a time-dependent linear forward operator mapping the space of measures to a time-dependent Hilbert space has to be inverted. These problems are regularized with dynamic optimal-transport energies that base on the continuity equation as well as convex functionals of BenamouBrenier-type [ESAIM:M2AN 54(6):23512382, 2020]. For the purpose of deriving properties of the solutions as well as numerical algorithms, we present sparsity results for general inverse problems that are connected with the extremal points of the BenamouBrenier energy subject to the continuity equation. For the latter, it is proven that the extremal points are realized by point masses moving along curves with Sobolev regularity [Bull. LMS 53(5):14361452, 2021]. This result will be employed in numerical optimization algorithms of generalized conditional gradient type. We present instances of this algorithm that are tailored towards dynamic inverse problems associated with point tracking. Finally, the application and numerical performance of the method is demonstrated for sparse dynamic superresolution [FOCM, 2022].

Kristian Bredies
University of Graz
Institute of Mathematics and Scientific Computing
kristian.bredies@uni-graz.at

Carioni Marcello
University of Twente
m.c.carioni@utwente.nl

Silvio Fanzon, Francisco Romero
University of Graz
silvio.fanzon@uni-graz.at, francisco.romero-hinrichsen@uni-graz.at

MS221

Online Optimization for Dynamic Electrical Impedance Tomography

Online optimization generally studies the convergence of optimization methods as more data is introduced into the problem; think of deep learning as more training samples become available. We adapt the idea to dynamic inverse problems that naturally evolve in time. We introduce an improved primal-dual online method specifically suited to these problems, and demonstrate its performance on dynamic monitoring of electrical impedance tomography.

Neil Kristofer D. Dizon, Jyrki Jauhiainen
University of Helsinki
neil.dizon@helsinki.fi, jyrki.h.jauhiainen@helsinki.fi

Tuomo Valkonen
University of Helsinki & Escuela Politécnica Nacional
tuomov@iki.fi

MS221

Centralized and Distributed Methods for Sparse Time-Varying Optimization

Sparse optimization consists in learning models with a reduced number of parameters, through the solution of suitable optimization problems. In the last years, substantial attention has been devoted to the tracking of time-varying sparse models. This has widespread applications, ranging from dynamic compressed sensing to parsimonious time-varying system identification. In contrast to static sparse optimization, tracking time-varying sparse systems calls for fast optimization algorithms, that promptly detect and identify time variations. For this purpose, centralized and distributed iterative algorithms have been proposed and analyzed in terms of dynamic regret, in an online learning perspective. These methodologies have been developed for convex sparse optimization problems, in particular for elastic net regularized regression. However, more recent work envisages also non-convex regularization, which in some cases is more efficient to obtain sparse models. The goal of this talk is to illustrate the state-of-the-art approaches to sparse time-varying optimization, both in centralized and distributed settings.

Sophie M. Fosson
Politecnico di Torino
sophie.fosson@polito.it

MS222

Facial Reduction for Semidefinite Relaxations of Combinatorial Optimization Problems

We propose an efficient implementation of FR algorithms for SDP relaxations of combinatorial optimization problems. Our innovative approach is to leverage the combinatorial and the polyhedral structures behind its SDP relaxations. The key idea is that the feasible set of the semidefinite program is a relaxation for the lifted feasible set of the associated problem. Thus, it is possible to connect certain geometric properties between these feasible sets. Our strategy is to derive these relevant properties from the underlying CO problem directly. As a result, this allows us to facially reduce its semidefinite relaxation in a more efficient and effective manner.

Hao Hu
Clemson University
hhu2@clemson.edu

Boshi Yang
Clemon University
boshiy@clemson.edu

MS222

Semidefinite Relaxations for the Cutwidth Minimization Problem

The Cutwidth Minimization Problem is a linear ordering problem on graphs, whose applications include natural language processing, information retrieval, graph drawing, and network migration scheduling. Finding the minimum cutwidth of a graph is an NP-hard problem, for which several heuristics and metaheuristics have been designed in order to compute upper bounds. Regarding lower bounds, integer linear programming models have been proposed, some with very good results on sparse graphs. However, even the best approaches see their performances decrease with the density and the size of the instances. In this talk, we introduce a new semidefinite relaxation for the Cutwidth Minimization Problem and investigate ways to efficiently strengthen it, using different sets of valid inequalities. We develop a cutting-plane algorithm based on these improved semidefinite relaxations, allowing us to significantly improve lower bounds for dense graphs. These methods can then be used for other challenging problems related to finding linear orderings minimizing other graph parameters, such as the pathwidth or the bandwidth.

Diane Puges
Alpen-Adria-Universität Klagenfurt
diane.puges@aau.at

MS222

Computing the Edge-Expansion using Semidefinite Programming

The edge expansion is an NP-hard to compute graph constant and gives us information about the connectivity of a graph. It is the minimum ratio of the number of edges joining two sets and the size of the smaller set over all possible non-trivial bipartitions of the vertices. The edge expansion of a connected graph is small if there is a bottleneck between two large parts of the graph. Because of this fact, it is for example used in clustering or network design. There are some heuristics to find a bipartition, like the well-known spectral clustering. This fractional optimization problem does not fit into the typical setting like other graph constants as the maximum cut which are NP-hard to compute. We propose different strategies to compute the edge expansion efficiently. One is to divide the problem into subproblems of an easier-to-handle type. Another one is to apply Dinkelbach's algorithm for fractional programming. Furthermore, we investigate the conjecture of Mihail and Vazirani, stating that the edge epxansion of the graph from a 0/1-polytope is at least 1, using our techniques.

Melanie Siebenhofer
University of Klagenfurt
melanie.siebenhofer@aau.at

Akshay Gupte
School of Mathematics
University of Edinburgh
akshay.gupte@ed.ac.uk

Angelika Wiegele
Alpen-Adria-Universit{ä}t Klagenfurt
Institut f. Mathematik
angelika.wiegele@aau.at

MS223

New Results for the Quadratic Assignment Problem

In this talk I build upon the work of Anstreicher, Goux, Linderoth, and myself (2002) for solving large quadratic assignment problems (QAPs). In particular, I highlight elements of the research which still persist as important trends in optimization even two decades later. We present new branching techniques which provide advantages over those used in previous approaches, and outline new ideas for future progress on the QAP.

Nathan Brixius
Amazon
nathan.brixius@gmail.com

MS223

Completely Positive Reformulations and Cutting Plane Algorithms for Mixed Integer Quadratic Programs

In seminal work, Burer proposed a reformulation of a 0-1 integer program with a (nonconvex) quadratic objective as a (convex) completely positive program (CPP), demonstrating that under mild conditions, the optimal solution to the convex program is a convex combination of optimal solutions to the MIQP. Many authors, including Anstreicher, have proposed solving the convex dual of Burer's reformulation using a cutting-plane approach that relies on solving an NP-Hard optimization problem to find cutting planes. We will show some connections between existing approaches before proposing a new completely positive formulation, where the cone arises from the completely positive cone of the nonnegative orthant intersected with a lifting of the linear equalities. We discuss some potential advantages of the new formulation, derive its dual cone, and propose cutting-plane approaches for its solution. We will present computational results comparing all formulations and separation approaches on a variety of practical instances.

Jeff Linderoth
University of Wisconsin-Madison
Dept. of Industrial and Sys. Eng
linderoth@wisc.edu

Arvind Raghunathan
Mitsubishi Electric Research Laboratories
raghunathan@merl.com

MS224

Learning Rate Grafting: Disentangling Adaptivity from Learning Rate Schedules

In the empirical science of training large neural networks, the learning rate schedule is a notoriously challenging-to-tune hyperparameter, which can depend on all other properties (architecture, optimizer, batch size, dataset, regularization, ...) of the problem. In this work, we probe the entanglements between the optimizer and the learning rate schedule. We propose the technique of optimizer grafting, which allows for the transfer of the overall implicit step size schedule from a tuned optimizer to a new optimizer, preserving empirical performance. This provides a robust plug-and-play baseline for optimizer comparisons, leading to reductions to the computational cost of optimizer hyperparameter search. Using grafting, we discover a non-adaptive learning rate correction to SGD which allows it to train a BERT model to state-of-the-art performance. Besides providing a resource-saving tool for practitioners, the invariances discovered via grafting shed light on the successes and failure modes of optimizers in deep learning.

Naman Agarwal
Google AI Princeton
naman33k@gmail.com

MS224

DoG Is SGD's Best Friend

We develop an SGD variant that does not require learning-rate tuning, which we call DoG. Theoretically, we prove that DoG enjoys strong parameter-free convergence guarantees for stochastic quasar-convex optimization, as long as its iterates stay bounded. Furthermore, we prove such bounded iterate guarantee for a small modification of the DoG iterates. Empirically, we show that DoG is competitive with carefully-tuned SGD on multiple neural network training tasks.

Yair Carmon
Computer Science
Tel Aviv University

ycarmon@cs.tau.ac.il

MS224

Making SGD Parameter-Free

We develop an algorithm for parameter-free stochastic convex optimization (SCO) whose rate of convergence is only a double-logarithmic factor larger than the optimal rate for the corresponding known-parameter setting. In contrast, the best previously known rates for parameter-free SCO are based on online parameter-free regret bounds, which contain unavoidable excess logarithmic terms compared to their known-parameter counterparts. Our algorithm is conceptually simple, has high-probability guarantees, and is also partially adaptive to unknown gradient norms, smoothness, and strong convexity. At the heart of our results is a novel parameter-free certificate for SGD step size choice, and a time-uniform concentration result that assumes no a-priori bounds on SGD iterates.

Oliver H. Hinder
University of Pittsburgh
ohinder@pitt.edu

MS225

Discretization Methods in MINLP

We study mixed-integer programming (MIP) relaxation techniques for the solution of nonconvex mixed-integer quadratically constrained quadratic programs (MIQCQPs). We present two MIP relaxation methods for nonconvex continuous variable products that enhance existing approaches. One is based on a separable reformulation, while the other extends the well-known MIP relaxation normalized multiparametric disaggregation technique (NMDT). In addition, we introduce a logarithmic MIP relaxation for univariate quadratic terms, called "sawtooth relaxation". We combine the latter with the separable reformulation to derive MIP relaxations of MIQCQPs. We provide a comprehensive theoretical analysis of these techniques, and perform a broad computational study to demonstrate the effectiveness of the enhanced MIP relaxations in terms producing tight dual bounds for MIQCQPs.

Andreas Barmann
Friedrich-Alexander-Universitat Erlangen-Nurnberg
andreas.baermann@math.unio-erlangen.de

Ben Beach
Grado Department of Industrial and Systems Engineering
Virginia Tech
bben6@vt.edu

Robert Burlacu
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Mathematik, EDOM
robert.burlacu@fau.de

Lukas Hagar
Friedrich-Alexander-Universitat Erlangen-Nurnberg
lukas.hager@fau.de

Robert Hildebrand
Virginia Tech
rhil@vt.edu

MS225

New Bounds for the Integer Carathodory Rank

Given a pointed rational n-dimensional cone C, the integer Carathodory rank $CR(C)$ is defined as the smallest integer k such that any integer vector in C can be expressed as a non-negative integer combination of at most k elements from the Hilbert basis of C. Cook, Fonlupt and Schrijver proved the bound $CR(C) \leq 2n-1$, which Sebő was able to improved by one. In this talk, I present new parametric and asymptotic bounds. In particular in an asymptotic setting, where we only considers 'most' integer vectors in C, we are able to improve the upper bound significantly. This talk is based on joint work with I.Aliev, M.Henk, M.Hogan and S.Kuhlmann.

Timm Oertel
Friedrich-Alexander-Universität
timm.oertel@fau.edu

MS226

Efficient Importance Scenario Generation for Optimization with Rare Events

Scalable algorithmic approaches for solving optimization formulations involving tail risks and rare events have remained elusive, primarily due to the large number of scenarios needed to witness the relevant rare scenarios in the corresponding scenario approximations. Attempts to alleviate this difficulty, via variance reduction or scenario aggregation, typically suffer from being extremely specialised to the optimization model and the distribution at hand. In this talk, we exhibit an algorithmic procedure for generating importance scenarios which overcome these challenges by producing widely-applicable scenario approximations with low variance. The variance reduction is substantial enough to result in an exponential reduction in the the number of scenarios needed for approximating objectives or constraints formulated in terms of rare events. The challenge of selecting a suitable distribution for generating scenarios for scenario approximation is automated here by means of a transformation that implicitly learns and replicates the concentration properties observed in less rare samples. The proposed procedure is the first to attain asymptotically optimal variance reduction for a wide variety of optimization objectives, including those modeled with algorithmic feature-mapping tools such as kernels, neural networks, etc.

Anand Deo
Singapore University of Technology and Design
deo.anand.a@gmail.com

Karthyek Murthy
Singapore University of Technology and Design (SUTD)
karthyek_murthy@sutd.edu.sg

MS226

Combating the Curse of Circularity: Adaptive Im-

portance Sampling for Rare-Event Optimization

In solving stochastic optimization problems that involve rare events, running crude Monte Carlo can be prohibitively inefficient. To address this issue, importance sampling (IS) can be employed to drive down the sampling error to a desirable level. However, selecting a good IS requires knowledge of the solution to the problem at hand, which is the goal to begin with and thus forms a circular challenge. We investigate the use of adaptive IS to untie this circularity. Our procedure sequentially updates the IS to reach the optimal sampler and the optimal solution simultaneously, and can be embedded in both sample average approximation and stochastic approximation-type algorithms. Our theoretical analysis reveals that the resulting solution achieves, in a minimax sense, the best possible asymptotic variance. We illustrate our effectiveness via numerical experiments, and also discuss some connections to rare-event gradient estimation and cross-entropy methods if time permits.

Yuanlu Bai
Columbia University, U.S.
yb2436@columbia.edu

Michael Fu
University of Maryland,
mfu@umd.edu

Shengyi He
Columbia University, U.S.
sh3972@columbia.edu

Guangxin Jiang
Harbin Institute of Technology
gxjiang@hit.edu.cn

Henry Lam
Columbia University
henry.lam@columbia.edu

MS226

Automatic Outlier Removal and Detection via Optimal Transport

We consider the task of simultaneous optimal outlier removal and supervised learning prediction of machine learning models. We propose an optimal transport-based formulation for automatic outlier removal. Optimal transport methods can be used to transfer mass from a probability distribution (call it a source) into another distribution (refer to as target). The source is typically the observed empirical distribution and, in our case, the target is a model which permits the best predictions when optimal transport is used to rectify the data under a budget constraint. This formulation enables the application of stochastic gradient descent strategies to find optimal outlier removals. We study global convergence algorithms for these types of formulations.

Jose H. Blanchet, Jiajin Li, Xuhui Zhang
Stanford University
jose.blanchet@stanford.edu, jiajinli@stanford.edu, xuhui.zhang@stanford.edu

MS227

Manifold-Free Riemannian Optimization

Riemannian optimization is a principled framework for solving optimization problems where the desired optimum is constrained to a smooth manifold. Algorithms designed in this framework usually require some geometrical description of the manifold, which typically includes tangent spaces, retractions, and cost function gradients. However, in many cases, only a subset (or none at all) of these elements can be accessed due to a lack of information or intractability. In this talk, we present a novel approach that allows us to perform approximate Riemannian optimization in such cases, where the constraining manifold is a submanifold of Euclidean space. At the bare minimum, our method requires only a noiseless sample set of the cost function and the intrinsic dimension of the constraining manifold. Using the samples, and utilizing the Manifold-MLS framework for manifold learning [Sober and Levin, Manifold Approximation by Moving Least-Squares Projection (MMLS), 2020] and for approximation of functions over manifolds [Sober, Aizenbud, and Levin, Approximation of functions over manifolds: A Moving Least-Squares approach, 2021], we construct approximations of the missing components allowing zero-order optimization with respect to the cost function and the constraints. We analyze the global convergence of a Riemannian gradient-based method based on our approximations and empirically demonstrate the strength of this method.

Boris Shustin
Tel Aviv University
Israel
borisshy@mail.tau.ac.il

Haim Avron
Tel Aviv University
haimav@tauex.tau.ac.il

Barak Sober
Hebrew University of Jerusalem, Israel
Barak.Sober@mail.huji.ac.il

MS227

Continuation Methods for Riemannian Optimization

Numerical continuation in the context of optimization can be used to mitigate convergence issues due to a poor initial guess. In this work, we extend this idea to Riemannian optimization problems, that is, the minimization of a target function on a Riemannian manifold. For this purpose, a suitable homotopy is constructed between the original problem and a problem that admits an easy solution. We develop and analyze a path-following numerical continuation algorithm on manifolds for solving the resulting parameter-dependent problem. To illustrate our developments, we consider two typical applications of Riemannian optimization: the computation of the Karcher mean and low-rank matrix completion. We demonstrate that numerical continuation can yield improvements for challenging instances of both problems.

Axel Séguin
Swiss Federal Institute of Technology Lausanne (EPFL)
axel.seguin@epfl.ch

Daniel Kressner
EPFL Lausanne
Mathicse
daniel.kressner@epfl.ch

MS227

Gauss-Southwell type descent methods for low-rank matrix optimization

We consider gradient-related methods for low-rank matrix optimization with a smooth strongly convex cost function. The methods operate on single factors and share aspects of both alternating and Riemannian optimization. We compare two possible choices for the search directions based on Gauss-Southwell type selection rules: one using the gradient of a factorized non-convex formulation, the other using the Riemannian gradient. Both methods provide convergence guarantees for the gradient that are analogous to the unconstrained case.

André Uschmajew
MPI MiS Leipzig
andre.uschmajew@uni-a.de

Bart Vandereycken
University of Geneva
bart.vandereycken@unige.ch

Guillaume Olikier
Université catholique de Louvain
guillaume.olikier@uclouvain.be

MS228

Sparsity-Based Nonlinear Reconstruction of Optical Parameters in Two-Photon Photoacoustic Computed Tomography

We present a new nonlinear optimization approach for the sparse reconstruction of single-photon absorption and two-photon absorption coefficients in the photoacoustic computed tomography (PACT). This framework comprises of minimizing an objective functional involving a least squares fit of the interior pressure field data corresponding to two boundary source functions, where the absorption coefficients and the photon density are related through a semi-linear elliptic partial differential equation (PDE) arising in photoacoustic tomography. Further, the objective functional consists of an L^1 regularization term that promotes sparsity patterns in absorption coefficients. The motivation for this framework primarily comes from some recent works related to solving inverse problems in acousto-electric tomography and current density impedance tomography. We provide a new proof of existence and uniqueness of a solution to the semi-linear PDE. Further, a proximal method, involving a Picard solver for the semi-linear PDE and its adjoint, is used to solve the optimization problem. Several numerical experiments are presented to demonstrate the effectiveness of the proposed framework.

Madhu Gupta
University of Texas at Arlington
mgupta25@gmu.edu

MS228

Unifying Branch-and-Bound Approaches to Solve ℓ_0-Penalized Problems

L0-penalized problems can be addressed using Branch-and-Bound approaches at the cost of a slight perturbation of the problem. More precisely, this perturbation allows constructing bounded relaxations at each node of the Branch-and-Bound tree. It is often a Big-M constraint or an additional Ridge penalty. So far, Branch-and-Bound algorithms were developed depending on the chosen perturbation. In this presentation, we present new theoretical results allowing to devise a generic Branch-and-Bound algorithm with only mild assumptions on the perturbation choice. This may lead to other perturbation strategies that have not yet been explored. In addition, we show how acceleration procedures recently proposed for special choices of perturbation can be generalized to our more generic framework. We illustrate our presentation with examples and numerical experiments.

Theo Guyard
Inria and INSA Rennes, France
theo.guyard@insa-rennes.fr

MS228

Relaxation Approaches for Nonlinear Sparse Optimization Problems

Recently, applications such as compressed sensing, image and signal processing, machine learning , but also economic applications such as sparse portfolio selection ask for lower dimensional (sparse) solutions of high dimensional optimization problems. Hence, there is an increasing interest in structured (sparse) solutions of linear and nonlinear optimization problems. One way to induce such sparse optimal solutions is to add a suitable (penalty) term to the objective function. In this talk, we will consider nonlinear optimization problems that involve an ℓ_0-norm term to induce sparsity of the solution vector. We will discuss reformulations based on complementarity type constraints which will further be used to define constraint qualifications and optimality conditions for the nonlinear sparse optimization problems. Next, we present suitable relaxation approaches and analyze their convergence properties. Finally, we finish the talk with some preliminary numerical test that compare the different approaches based on our reformulations.

Sonja Steffensen
RWTH Aachen University
Dept. of Mathematics
steffensen@igpm.rwth-aachen.de

MS229

Modeling the Collapsed k-Core Problem via Bilevel

Programming

In social network analysis, the size of the k-core, i.e., the maximal induced subgraph of the network with minimum degree at least k, is frequently adopted as a typical metric to evaluate the cohesiveness of a community. In this talk, we focus on the Collapsed k-Core Problem, which seeks to find a subset of b users, namely the most critical users of the network, the removal of which results in the smallest possible k-core. On the one hand, we model the Collapsed k-Core Problem as a natural deletion-round-indexed Integer Linear formulation. On the other hand, we provide two bilevel programs for the problem, which differ in the way in which the k-core identification problem is formulated at the lower level, as well as in the solution approach. The performances of the proposed formulations are compared on a set of instances generated from several network data sets available in the literature. We further compare our approaches with the an existing state-of-the-art solver for mixed-integer bilevel problems (Fischetti et al. "A New General-Purpose Algorithm for Mixed-Integer Bilevel Linear Programs", Operations Research, 2017).

Martina Cerulli
ESSEC
b00799089@essec.edu

Domenico Serra, Carmine Sorgente
University of Salerno
dserra@unisa.it, csorgente@unisa.it

Claudia Archetti
ESSEC Business School
archetti@essec.edu

Ivana Ljubic
University of Vienna
ljubic@essec.edu

MS229

Poisoning Attacks Against Linear Regression Models with Categorical Variables: a Mixed-Integer Nonlinear Bilevel Optimization Approach

Machine Learning (ML) models have become a very powerful tool to extract information from large datasets and use it to make predictions and automated decisions. However, ML models can be vulnerable to external attacks, causing them to underperform or to deviate from their expected task. It is possible to anticipate and prepare for such events by designing strong attacks, which are later used for creating and testing defence strategies. One way to attack ML models is by injecting them with poisoning data to mislead the algorithm during the training phase. These problems are usually solved using gradient-based methods that do not optimize discrete categorical variables. We present a poisoning attack for a ridge regression model containing both numerical and categorical variables that explicitly models and optimizes categorical variables. We formulate the problem as a bilevel optimization problem that is quadratic mixed-integer on the upper-level and unconstrained quadratic on the lower-level (MIQP-QP). We reformulate this problem using the KarushKuhnTucker (KKT) conditions of the lower level, resulting in a mixed-integer nonlinear problem (MINLP). We then use sensitivity analysis of linear regression to perform bound tightening of bilinear terms. Then, the model is solved using a nonlinear optimization solver. This presentation outlines the mathematical formulation of the problem, the proposed bound tightening procedure, and presents computational experiments.

Monse Guedes Ayala
School of Mathematics
University of Edinburgh
s1553571@sms.ed.ac.uk

Miguel F. Anjos
University of Edinburgh
anjos@stanfordalumni.org

Lars Schewe
The University of Edinburgh, School of Mathematics, UK
lars.schewe@ed.ac.uk

Zeynep Suvak
University of Edinburgh
zsuvak@exseed.ed.ac.uk

MS229

A Rank-Based Reward Between a Principal and a Field of Agents: Application to Energy Savings

We consider a problem where a Principal aims to design a reward function to a field of agents (heterogeneous population). We focus on a specific type of reward: the agents compete with each other through their rank within the population in order to obtain the best reward. We first explicit the equilibrium for the mean-field game played by the agents, and then characterize the optimal reward in the homogeneous setting. For more general case, we develop a numerical approach, which we apply to our case study: the market of Energy Saving Certificates.

Clemence Alasseur
EDF
clemence.alasseur@edf.fr

Erhan Bayraktar
University of Michigan
Department of Mathematics
erhan@umich.edu

Roxana Dumitrescu
King's College London
roxana.dumitrescu@kcl.ac.uk

Quentin Jacquet
EDF
quentin.jacquet@edf.fr

MS230

Learned Robust Pca: A Scalable Deep Unfolding Approach for High-Dimensional Outlier Detection

Robust principal component analysis (RPCA) is a critical tool in modern machine learning, which detects outliers in the task of low-rank matrix reconstruction. In this paper, we propose a scalable and learnable non-convex approach for high-dimensional RPCA problems, which we call Learned Robust PCA (LRPCA). LRPCA is highly efficient, and its free parameters can be effectively learned to optimize via deep unfolding. Moreover, we extend deep unfolding from finite iterations to infinite iterations via a novel feedforward-recurrent-mixed neural network model. We establish the recovery guarantee of LRPCA under mild assumptions for RPCA. Numerical experiments show that LRPCA outperforms the state-of-the-art RPCA algorithms, such as ScaledGD and AltProj, on both synthetic datasets and real-world applications.

Hanqin Cai
University of Central Florida
hqcai@ucf.edu

MS230

A Journey Through Algorithm Unrolling for Unsupervised Inverse Problems

Inverse problems are ubiquitous in observational science such as imaging, neurosciences or astrophysics. They consist in recovering a signal given noisy observations through a measurement operator. To solve such problems, Machine learning approaches have been proposed based on algorithm unrolling. With such techniques, classical optimization algorithms used to solve inverse problems can be seen as differentiable procedure with parameters that can be learned. In this talk, I will focus on the understudied case where no ground truth is available for the problem, i.e., the signal itself can never be observed, only its measurememts. First, I will present results that aim at understanding what unrolled algorithms learn when used to solve an optimization problem. In particular, results from [Ablin et al., 2019, Learning step sizes for unfolded sparse coding] show that learned algorithms cannot go faster than the original algorithm assymptotically, only improve the first iterations. This shows that unrolling should mostly be used with low number of iterations. Then, I will show how algorithm unrolling can be used in conjonction with dictionaries to learn a data-driven structure on the inverse problem solution. For this problem also, we show in [Malzieux et al., 2022, Understanding approximate and unrolled dictionary learning for pattern recovery] that using a few unrolled iterations can provide benefit for learning dictionary while too many iterations can be harmful.

Thomas Moreau
INRIA
thomas.moreau@inria.fr

MS230

Structured Learning for Safe and Efficient Operation of Energy Systems

The optimal power flow (OPF) problem is a fundamental problem in power systems operations and planning. With high penetration of uncertain renewable resources in power systems, DCOPF needs to be solved repeatedly for a large amount of scenarios, which can be computationally challenging. As an alternative to iterative solvers, neural networks are often trained and used to solve OPF. These approaches can offer orders of magnitude reduction in computational time. A central challenge is to ensure that the learned solutions are feasible. Using the geometry of the optimization problem and by designing a clever filtering layer, we show how to guaranteed feasibly of a linearized version of the OPF problem without projections.

Baosen Zhang
University of Washington
zhangbao@uw.edu

MS231

The Power of Adaptivity for Stochastic Submodular Cover

We study the stochastic submodular cover problem, where one needs to select a subset of stochastic items to cover a submodular function at minimum expected cost. Solutions in this setting correspond to sequential decision processes that select items one by one adaptively, depending on prior observations. While such adaptive solutions achieve the best objective, they are inherently sequential, which is undesirable in many applications. We show how to obtain solutions that approximate fully-adaptive solutions using only a few rounds of adaptivity. We study both independent and correlated settings, proving smooth tradeoffs between the number of adaptive rounds and the solution quality. We also present experimental results demonstrating that a few rounds of adaptivity suffice to obtain high-quality solutions in practice.

Rohan Ghuge
University of Michigan
rghuge@umich.edu

Anupam Gupta
Carnegie Mellon University
anupamg@cs.cmu.edu

Viswanath Nagarajan
University of Michigan
viswa@umich.edu

MS231

Optimization with Explorable Uncertainty

In the traditional frameworks for optimization under uncertainty, an algorithm has to accept the incompleteness of input data. Clearly, more information or even knowing the exact data would allow for significantly better solutions. How much more information suffices for obtaining a certain solution quality? Which information shall be retrieved? Explorable uncertainty refers to settings where certain numerical input parameters are initially unknown, but can be obtained at a certain cost by using queries. An algorithm can make queries (adaptively) until it has obtained sufficient information to solve a given problem. The challenge lies in balancing the cost for querying and the impact on the solution quality. In this talk, I will give a short overview on recent work on explorable uncertainty for problems such as finding a minimum spanning tree in a graph of uncertain edge cost, selecting the minimum in a set of unknown values and finding a subset of minimum total value. I will discuss both, an adversarial online model and a stochastic model for revealing the true value.

Nicole Megow
Universität Bremen, Germany
nicole.megow@uni-bremen.de

MS231

When Matching Meets Batching: Optimal Multi-Stage Algorithms and Applications

In several applications of real-time matching of demand to supply in online marketplaces — for example matching delivery requests to dispatching centers in Amazon or allocating video-ads to users in YouTube — the platform allows for some latency (or there is an inherent allowance for latency) in order to batch the demand and improve the efficiency of the resulting matching. Motivated by these scenarios, I investigate the optimal trade-off between batching and inefficiency in the context of designing robust online allocations in this talk. In particular, I consider K-stage variants of the classic vertex weighted bipartite b-matching and AdWords problems in the adversarial setting, where online vertices arrive stage-wise and in K batchesin contrast to online arrival. Our main result for both problems is an optimal $(1-(1-1/K)^K)$-competitive competitive (fractional) matching algorithm, improving the classic $(1 - 1/e)$ competitive ratio bound known for the online variants of these problems (Mehta et al., 2007; Aggarwal et al., 2011). Our main technique at high-level is developing algorithmic tools to vary the trade-off between "greedyness" and "hedging" of the matching algorithm across stages. We rely on a particular family of convex-programming based matchings that distribute the demand in a specifically balanced way among supply in different stages, while carefully modifying the balancedness of the resulting matching across stages.

Rad Niazadeh
Chicago Booth School of Business
rad.niazadeh@chicagobooth.edu

MS232

Towards a First-Order Algorithmic Framework for Wasserstein Distributionally Robust Risk Minimization

Wasserstein distance-based distributionally robust optimization (DRO) has received much attention lately due to its ability to provide a robustness interpretation of various learning models. In recent years, it has been shown that various DRO formulations of learning models admit tractable convex reformulations. However, most existing works propose to solve these convex reformulations by standard off-the-shelf solvers (e.g., IPOPT, MOSEK). Nevertheless, these solvers often rely on general-purpose interior-point or nonlinear programming based methods, which do not scale well with problem size. Such a state of affairs severely limits the applicability of the DRO approach in large-scale settings. On the other hand, there are very few works that attempt to develop fast iterative methods to solve these DRO problems, which typically possess complicated structures. In this talk, we take a first step towards resolving the above difficulty by developing a first-order algorithmic framework for tackling a class of Wasserstein distance-based distributionally robust risk minimization problems in machine learning. To ensure the efficiency, the hidden structures in the equivalent finite convex ormulation are carefully identified, splitted and exploited. Thus, the updates can be computed in a highly effective manner. Finally, we show extensive experiments on both synthetic and real-world datasets that our methods are orders of magnitude faster than the state-of-the-art solver.

Jiajin Li
Stanford University
jiajinli@stanford.edu

Caihua Chen
Nanjing University
chchen@nju.edu.cn

Anthony So
The Chinese University of Hong Kong
manchoso@se.cuhk.edu.hk

MS232

A Stochastic Conjugate Subgradient Algorithm for Nonsmooth Optimization Problems

Stochastic Programming (SP) focuses on solving optimization problems with uncertainty described by random variables. While first-order methods dominate the field because of their low computational requirements in any one iteration, higher-order methods can provide faster convergence rate especially through coordinated choice between successive directions and step-size. In this presentation, we propose a stochastic algorithm which will accommodate conjugate subgradients of non-smooth functions leading to only marginally more computations per iteration while providing faster descent. This method also preserves the decomposition of subgradients by data points, as is common in SP. This approach combining conjugate directions with subgradients leads to an algorithm which goes beyond first order approximations. The convergence result can be established, and our experiments reveal that the new algorithm converges faster than classic first order method.

Di Zhang, Sen Suvrajeet
University of Southern California
dzhang22@usc.edu, suvrajes@usc.edu

MS232

On Computational and Statistical Challenges of Solving Nonconvex Minimax Optimization Problems

This talk studies the computational and statistical challenges of adversarial learning through solving min-max optimization problems. In the first part, we focus on the computational challenges. While many algorithms are developed for solving min-max optimization problems, many of them involve multiple loop iterations. Consequently, they are challenging to implement in practice. We focus on single-loop algorithms and discuss smoothed gradient descent-ascent (smoothed GDA) for solving nonconvex minimax problems. With a novel proximal-primal-dual analysis, we prove that our algorithm converges for nonconvex-concave problems with $\mathcal{O}(\infty/\epsilon^{\triangle})$ iteration complexity and achieves the benchmark iteration complexity $\mathcal{O}(\infty/\epsilon^{\in})$ for minimizing a pointwise maximum of a class of nonconvex functions. In the second part, we focus on the generalization bound analysis of algorithms for solving stochastic min-max optimization problems. We discuss appropriateness of metrics to derive generalization bounds under. The generalization bounds are used to analyze algorithms for adversarial training and generative adversarial networks (GAN). For adversarial training, we develop a new algorithm called smoothed GDMAX, which improves the generalization bound by introducing a novel smoothing technique to GDMAX (a novel variation of a technique used before). For GAN training, we study why gradient descent-ascent is better than the GDMAX algorithms in practice.

Jiawei Zhang
Massachusetts Institute of Technology
jwzhang@mit.edu

Asu Ozdaglar
MIT
asuman@mit.edu

Kaiqing Zhang
University of Maryland
kaiqing@umd.edu

Sarath Pattathi
MIT
sarathp@mit.edu

MS233

Revisiting Spectral Bundle Methods: Primal-Dual (Sub)linear Convergence Rates

The spectral bundle method proposed by Helmberg and Rendl is well-established for solving large-scale semidefinite programs (SDP) thanks to its low per iteration computational complexity and strong practical performance. In this talk, we revisit this classic method showing it achieves sublinear convergence rates in terms of both primal and dual SDPs under merely strong duality, complementing previous guarantees on primal-dual convergence. Moreover, we show the method speeds up to linear convergence if (1) structurally, the SDP admits strict complementarity, and (2) algorithmically, the bundle method captures the rank of the optimal solutions. Such complementary and low rank structure is prevalent in many modern and classical applications. The linear convergent result is established via an eigenvalue approximation lemma which might be of independent interest. Numerically, we confirm our theoretical findings that the spectral bundle method, for modern and classical applications, speeds up under these conditions.

Lijun Ding
Cornell University
lijunbrianding@gmail.com

MS233

Memory-Efficient Approximation Algorithm for Max-Cut, Max-K-Cut and Correlation Clustering

Semidefinite programs (SDPs) are a special class of convex optimization problems that have a wide range of applications in areas such as control theory, statistical modeling, and combinatorial optimization. We discuss three fundamental graph partitioning problems Max-Cut, Max-k-Cut, and the Max-Agree variant of correlation clustering. Given a graph $G = (V, E)$, Max-k-Cut aims to divide the nodes into k partitions such that the weight of the edges crossing the partitions is maximized. Max-Cut is a special case of Max-k-Cut with $k = 2$. While Max-Agree requires additional information about the 'similarity' of 'dissimilarity' of each pair of nodes $(i, j) \in E$ and seeks to cluster the nodes of the graph such that the sum of the 'similar' edges within the cluster and 'dissimilar' edges across the clusters is maximized. For large-scale instances of problems, the memory required to solve SDPs becomes a key computational bottleneck. We discuss the application of a recent Gaussian sampling-based technique to the three combinatorial optimization problems. We show that by applying our approach to the three problems, we achieve nearly the same approximation guarantees as the best-known results, while the memory used by our method (in addition to the memory required to store the problem instance) is $\mathcal{O}(|V|)$ for Max-Cut and $\mathcal{O}(|V| + |E|)$ for Max-k-Cut and Max-Agree that uses $\mathcal{O}(|V| + |E|)$.

Nimita Shinde
Brown University
nimitashinde25@gmail.com

MS233

Accelerated First-Order Methods for a Class of Semidefinite Programs

We discuss a new storage-optimal first-order method for solving a special class of semidefinite programs (SDPs) to high accuracy. The class of SDPs that we consider, the exact QMP-like SDPs is characterized by low-rank solutions, a priori knowledge of the restriction of the SDP solution to a small subspace, and standard regularity assumptions such as strict complementarity. Crucially, we show how to use a certificate of strict complementarity to construct a low-dimensional strongly convex minimax problem whose optimizer coincides with a factorization of the SDP optimizer. From an algorithmic standpoint, we show how to construct the necessary certificate and how to solve the minimax problem efficiently. Based on joint work with Fatma Kilin-Karzan

Alex L. Wang
Centrum Wiskunde & Informatica
Purdue University
wang5984@purdue.edu

Fatma Kilinc-Karzan
Tepper School of Business
Carnegie Mellon University

fkilinc@andrew.cmu.edu

MS234

Approximating Hessians

In this talk, how to approximate a full or partial Hessian using a set of sample points is discussed. Error bounds are presented. These error bounds provide information on how to choose the set of sample points to obtain a certain degree of accuracy for some, or all entries of a Hessian.

Gabriel Jarry-Bolduc
University of British Columbia
gabjarry@gmail.com

Warren Hare
University of British Columbia, Okanagan Campus
warren.hare@ubc.ca

Chayne Planiden
University of British Columbia Okanagan
chayne.planiden@ubc.ca

MS234

SOLAR: A solar thermal power plant simulator for blackbox optimization benchmarking

This work presents SOLAR, a collection of optimization problems provided as a benchmarking tool for blackbox optimization solvers. Each problem optimizes the design of a concentrated solar power plant defined as a blackbox numerical model. The type of variables, dimensionality, and number and type of constraints are different across problems. Optimization may be single or biobjective. The solar plant model considers several subsystems: a heliostats field, a central cavity receiver, a molten salt thermal energy storage, a steam generator and an idealized power block. Benchmark optimization results are provided using the NOMAD software package.

Sébastien Le Digabel
Ecole Polytechnique de Montreal
Sebastien.Le.Digabel@gerad.ca

MS234

Derivative-Free Variance-Reduced Jacobian Sketching

We consider the setting of derivative-free finite-sum minimization. Motivated in particular by problems in nuclear model calibration, we suppose that each summand function is computationally expensive, and that moreover, the summand functions can be evaluated independently. In this setting, we propose and demonstrate a novel method inspired by variance reduction methods in machine learning; such a method permits us to more judiciously select a subset of summand functions to model on each iteration according to a particular probability distribution. We then combine this methodology with sketching methods, enabling judicious sketches of (a model of) the problem Jacobian. Numerical results demonstrating the efficiency of our Jacobian sketching method will be presented.

Matt Menickelly
Argonne National Laboratory
mmenickelly@anl.gov

Stefan Wild
Argonne National Laboratory
LANS, Mathematics and Computer Science
wild@lbl.gov

MS235

Accelerated Gradient Algorithm with Dry-like Friction and Nonmonotone Line Search for Nonconvex Optimization Problems

In this work, we propose an accelerated gradient algorithm for the problem of minimizing a differentiable (possibly nonconvex) function in Hilbert spaces. We first extend the dry friction property for convex functions to what we call a *dry-like friction property* in a nonconvex setting, and then employ a line search technique to adaptively update parameters at each iteration. Depending on the choice of parameters, the proposed algorithm exhibits subsequential convergence to a critical point or full sequential convergence to an "approximate' critical point of the objective function. We also establish the full sequential convergence to a critical point under the Kurdyka-Lojasiewicz (KL) property of a merit function. Thanks to the parameters' flexibility, our algorithm can reduce to a number of existing inertial gradient algorithms with Hessian damping and dry friction. By exploiting variational properties of the Moreau envelope, the proposed algorithm is adapted to address weakly convex nonsmooth optimization problems. In particular, we extend the result on KL exponent for the Moreau envelope of a convex KL function to a broad class of KL functions that are not necessarily convex nor continuous.

Lien Nguyen
RMIT University
nguyenthuylien228@gmail.com

MS235

Higher Order Dynamical Systems for Monotone Inclusion and Variational Inequalities

We consider a third order dynamical system for finding a zero of the sum of two generalized monotone operators. Under standard conditions, we prove the existence and uniqueness of strong global solution of the proposed dynamical systems. The global exponential convergence of the trajectories is established under strong monotonicity of the sum and Lipschitz continuity of the single valued operator. Discrete version of the proposed dynamical system leads to a relaxed forward-backward algorithms with double momentum and the linear convergence is obtained under suitable conditions on parameters. We also discuss the application of the obtained results to strongly pseudomonotone variational inequality.

Vuong Phan
University of Southampton

t.v.phan@soton.ac.uk

MS236

Benefits of Attention Mechanism for Context-aware Learning

Prompt-tuning is an emerging strategy to adapt large language models (LLM) to downstream tasks by learning a (soft-)prompt parameter from data. Despite its success in LLMs, there is limited theoretical understanding of the power of prompt-tuning and the role of the attention mechanism in prompting. In this talk, we explore prompt-tuning for one-layer attention architectures and study contextual mixture-models where each input token belongs to a context-relevant or -irrelevant set. We isolate the role of prompt-tuning through a self-contained prompt-attention model. Our contributions are as follows: (1) We show that softmax-prompt-attention is provably more expressive than softmax-self-attention and linear-prompt-attention under our contextual data model. (2) We analyze the initial trajectory of gradient descent and show that it learns the prompt and prediction head with near-optimal sample complexity and demonstrate how prompt can provably attend to sparse context-relevant tokens. (3) Assuming a known prompt but an unknown prediction head, we characterize the exact finite sample performance of prompt-attention which reveals the fundamental performance limits and the precise benefit of the context information. We also provide experiments that verify our theoretical insights on real datasets and demonstrate how prompt-tuning enables the model to attend to context-relevant information.

Samet Oymak
UC Riverside
oymak@ece.ucr.edu

MS237

Trade-off among Infeasibility, Efficiency and Accuracy for Gromov-Wasserstein Computation

In this talk, we study the design and analysis of a class of efficient algorithms for computing the Gromov-Wasserstein (GW) distance tailored to large-scale graph learning tasks. Armed with the Luo-Tseng error bound condition, two proposed algorithms, called Bregman Alternating Projected Gradient (BAPG) and hybrid Bregman Proximal Gradient (hBPG) enjoy the convergence guarantees. Upon task-specific properties, our analysis further provides novel theoretical insights to guide how to select the best-fit method. As a result, we are able to provide comprehensive experiments to validate the effectiveness of our methods on a host of tasks, including graph alignment, graph partition, and shape matching. In terms of both wall-clock time and modeling performance, the proposed methods achieve state-of-the-art results.

Jiajin Li
Stanford University
jiajinli@stanford.edu

Jianheng Tang
HKUST
jtangbf@connect.ust.hk

Lemin Kong
The Chinese University of Hong Kong
lkong@se.cuhk.edu.hk

Huikang Liu
Shanghai University of Finance and Economics
hkliu2014@gmail.com

Jia Li
HKUST
jialee@ust.hk

Anthony So
The Chinese University of Hong Kong
manchoso@se.cuhk.edu.hk

Jose Blanchet
Stanford University
jose.blanchet@stanford.edu

MS237

Communication-Efficient Second-Order Optimization Methods

Developing scalable second-order optimization methods for machine learning has become increasingly important with the need to accurately process the nearly limitless quantity of data generated from applications such as sensor networks, social media, and computational science. Second-order optimization methods can potentially provide stronger convergence guarantees compared to the first-order variants, however, they are not widely used in practice due to excessive data communication. Communication, i.e. the cost of moving data in the memory hierarchy of a single processor or processors over a network, is expensive and its costs already greatly exceed arithmetic costs, and the gap is growing over time following technological trends. In this talk, I will present examples of how second order optimization methods are reformulated and their implementations optimized to be communication-efficient. As a result the methods are more scalable and practical for applications such as in machine learning.

Maryam Mehri Dehnavi
University of Toronto
mmehride@cs.toronto.edu

MS237

Nonconvex and Combinatorial Optimization for Clustering

This talk is devoted to optimization problems arising from data clustering. We first study the scenario where the data come from a mixture of two Gaussians that share a common but unknown covariance matrix. We derive a Max-Cut program based on maximum likelihood estimation and prove its optimal statistical accuracy and sample complexity. We also prove the sub-optimal sample complexity of sum-of-square relaxations, depicting a statistical-computational trade-off. Finally, we achieve the best of both worlds in a slightly perturbed version of the problem where the Gaussian distribution is changed to a leptokurtic elliptical distribution. The key is constructing an objective function for clustering with a benign landscape: all local minima are statistically optimal, and all saddle points are strict.

Kaizheng Wang
Columbia University
kw2934@columbia.edu

Damek Davis
Cornell University
ORIE
dsd95@cornell.edu

Mateo Diaz
California Institute of Technology,U.S.
mateodd@caltech.edu

Yuling Yan
Princeton University
yulingy@princeton.edu

MS238

Cell Tracking Using Uncertainty Quantification For the Radiative Transfer Equation

We are interested in cell tracking using radiative transfer equation. This problem will be formulated as regularized optimal control problem with respect to a possibly high-dimensional partial differential equations. The equation may include uncertainty in the geometry as well as possible sparse controls. Approximation methods using gPC and/or moment methods in the angular direction will be applied. Numerical results will be shown.

Jan Friedrich, Michael Herty
RWTH Aachen University
friedrich@igpm.rwth-aachen.de, herty@igpm.rwth-aachen.de

MS238

Multilevel Diffusion: Infinite Dimensional Score-Based Diffusion Models for Image Generation

Score-based diffusion models (SBDM) have recently emerged as state-of-the-art approaches for image generation. Existing SBDMs are typically formulated in a finite-dimensional setting, where images are considered as tensors of a finite size. This papers develops SBDMs in the infinite-dimensional setting, that is, we model the training data as functions supported on a rectangular domain. Besides the quest for generating images at ever higher resolution our primary motivation is to create a well-posed infinite-dimensional learning problem so that we can discretize it consistently on multiple resolution levels. We thereby hope to obtain diffusion models that generalize across different resolution levels and improve the efficiency of the training process. We demonstrate how to overcome the shortcomings of current SBDM approaches in the infinite-dimensional setting.

Nicole Yang
Emory University
tianjiao.yang@emory.edu

Paul Hagemann
TU Berlin
hagemann@math.tu-berlin.de

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

Gabriele Steidl
TU Berlin
steidl@math.tu-berlin.de

MS238

Distributed Gradient Tracking Methods for Computing the Best Nash Equilibrium

In noncooperative Nash games, equilibria are known to be inefficient. This is exemplified by the Prisoner's Dilemma and was provably shown in the 1980s. Since then, understanding the quality of Nash equilibrium (NE) received considerable attention, leading to the emergence of inefficiency measures including Price of Anarchy and Price of Stability. Of these, the latter is characterized in terms of the best NE. Traditionally, computing the best NE is done through two-loop schemes that are criticized by their computational inefficiency and lack of provable guarantees. Our goal lies in the development of amongst the first single-timescale distributed optimization methods over networks for computing the best NE. The main contributions are as follows. Employing a regularization-based relaxation approach and distributed gradient tracking techniques, we devise two iteratively regularized gradient tracking algorithms. The first method addresses computing the best NE over directed networks, while the second method addresses a stochastic variant of this problem over undirected networks. For both methods, we establish the consensus and derive new convergence rate statements for suboptimality (relative to the Tikhonov trajectory) and consensus violation of the generated iterates. We provide preliminary numerical results on a networked Nash-Cournot game as well as a traffic equilibrium model where we compare the performance of the proposed methods with that of other few existing methods.

Farzad Yousefian
Rutgers University
farzad.yousefian@rutgers.edu

MS239

Making Continuous Functions Lipschitz

Let $\langle X, \tau \rangle$ be a metrizable topological space and let $\langle Y, \rho \rangle$ be a metric space. Let Ω be a family of bounded continuous functions from X to Y. We show that the family is Lipschitzian with respect to some compatible metric on X if and only if the family can be written as a countable union of pointwise equicontinuous subfamilies. From this, we easily characterize those families of continuous functions between metrizable spaces that are Lipschitzian with respect to appropriately chosen metrics on the domain and target space.

Gerald Beer
California State University, Los Angeles
gbeer@cslanet.calstatela.edu

Zvi Artstein
The Weizmann Institute of Science
Artstein@weizmann.ac.il

MS239

Quasidensity, maximal monotonicity locally and local maximal monotonicity

Let E be a nonzero real Banach space. 'Quasidensity' is a concept that can be applied to subsets of $E \times E^*$ (or equivalently to multifunctions from E into E^*). Every closed quasidense monotone set is maximally monotone, but there exist maximally monotone sets that are not quasidense. The subdifferential of a proper, convex lower semicontinuous function on E is quasidense. The subdifferentials of certain nonconvex functions are also quasidense. (This follows from joint work with Xianfu Wang.) The closed monotone quasidense sets satisfy a sum theorem and a dual sum theorem. We know of ten conditions equivalent to the statement that a closed monotone set be quasidense, but quasidensity seems to be the only one of the ten that extends easily to nonmonotone sets. We also discuss multifunctions of 'type (FPV)' = 'maximal monotone locally' and 'type (FP)' = 'locally maximal monotone'.

Stephen Simons
Department of Mathematics
University of California Santa Barbara
stesim38@gmail.com

MS239

Uniformly Monotone Operators on bounded Sets

Properties of monotone operators that are uniformly monotone on bounded sets will be examined. In particular, self-dual properties of these operators that parallel the convex function case are established using the reflected resolvent. The self-dual properties tie closely with classes of nonexpansive mappings that possess nice fixed point properties, in particular, natural generalizations of contractive mappings and strongly nonexpansive mappings are central to the self-dual properties.

Jon Vanderwerff
Department of Mathematics
La Sierra University
jvanderw@lasierra.edu

Walaa M. Moursi
University of Waterloo
walaa.moursi@uwaterloo.ca

MS240

Global Optimization Using Maching Learning

Many approaches for addressing Global Optimization problems typically rely on relaxations of nonlinear constraints over specific mathematical primitives. This is restricting in applications with constraints that are black-box, implicit or consist of more general primitives. Trying to address such limitations, Bertsimas and Ozturk (2022) proposed OCTHaGOn as a way of solving black-box global optimization problems by approximating the nonlinear constraints using hyperplane-based Decision-Trees and then using those trees to construct a unified MIO approximation of the original problem. We provide significant extensions to this approach, by (i) approximating the original problem using a much richer family of MIO-representable ML models besides Decision Trees, (ii) proposing adaptive sampling procedures for more accurate ML-based constraint approximations, (iii) utilizing robust optimization to account for the uncertainty of the sample-dependent training of the ML models and (iv) leveraging a family of relaxations to address the infeasibilities of the final MIO approximation. We show the improvements resulting from those enhancements through a wide range of Global Optimization benchmarks. We demonstrate the promise of the enhanced approach in finding globally optimal solutions, and compare it with well-established global optimizers such as BARON.

Dimitris Bertsimas
Massachusetts Insitute of Technology
dbertsim@mit.edu

Georgios Margaritis
MIT
geomar@mit.edu

MS240

An Extension of the Reformulation-Linearization Technique to Nonlinear Optimization

We introduce a novel Reformulation-Perspectification Technique (RPT) to obtain convex approximations of nonconvex continuous optimization problems. RPT consists of two steps, those are, a reformulation step and a perspectification step. The reformulation step generates redundant nonconvex constraints from pairwise multiplication of the existing constraints. The perspectification step then convexifies the nonconvex components by using perspective functions. The proposed RPT extends the existing Reformulation-Linearization Technique (RLT) in two ways. First, it can multiply constraints that are not linear or not quadratic, and thereby obtain tighter approximations than RLT. Second, it can also handle more types of nonconvexity than RLT. We demonstrate the applicability of RPT by extensively analyzing all 15 possibilities of pairwise multiplication of the five basic cone constraints (linear cone, second-order cone, power cone, exponential cone, semi-definite cone). We show that many well-known RLT based results can also be obtained and extended by applying RPT. Numerical experiments on dike height optimization and convex maximization problems demonstrate the effectiveness of the proposed approach.

Dimitris Bertsimas
Massachusetts Insitute of Technology
dbertsim@mit.edu

Danique de Moor, Dick den Hertog
University of Amsterdam
daniquedemoor@hotmail.com, d.denhertog@uva.nl

Theodoros Koukouvinos
MIT
tkoukouv@mit.edu

Jianzhe Zhen
N.A.
trevorzhen@gmail.com

MS241

Escaping Saddle Points with Compressed Sgd

Stochastic gradient descent (SGD) is a prevalent optimization technique for large-scale distributed machine learning. While SGD computation can be efficiently divided between multiple machines, the communication typically becomes a bottleneck in the distributed setting. Gradient compression methods can be used to alleviate this problem, and a recent line of work shows that SGD augmented with gradient compression converges to an ϵ-first-order stationary point. In this work, we extend these results to convergence to an ϵ-*second*-order stationary point (ϵ-SOSP). In addition, we show that, when the stochastic gradient is not Lipschitz, compressed SGD with Random-k compressor converges to an ϵ-SOSP with the same number of iterations as uncompressed SGD, while improving the total communication by a factor of $\tilde{\Theta}(\sqrt{d}\epsilon^{-3/4})$, where d is the dimension of the optimization problem. We present additional results for the cases when the compressor is arbitrary and when the stochastic gradient is Lipschitz.

Dmitrii Avdiukhin
Indiana University
dimonbv@gmail.com

Grigory Yaroslavtsev
Indiana University, Bloomington, U.S.
grigory@grigory.us

MS241

Communication-Efficient Second-Order Methods for Distributed Optimization

Despite their high computation and communication costs, Newton-type methods remain an appealing option for distributed training due to their robustness against ill-conditioned convex problems. We study communication compression and aggregation mechanisms for curvature information in order to reduce these costs while preserving theoretically superior local convergence guarantees. We prove that the recently developed class of three point compressors of Richtarik et al. [2022] for gradient communication can be generalized to Hessian communication as well. This result opens up a variety of communication strategies, such as contractive compression and lazy aggregation, available to our disposal to compress prohibitively costly curvature information. Moreover, we discovered several new 3PC mechanisms, such as adaptive thresholding and Bernoulli aggregation, which require reduced communication and occasional Hessian computations. Furthermore, we extend and analyze our approach to bidirectional communication compression and partial device participation setups to cater to the practical considerations of applications in federated learning. For all our methods, we derive fast condition-number-independent local linear and/or superlinear convergence rates. Finally, with extensive numerical evaluations on convex optimization problems, we illustrate that our designed schemes achieve state-of-the-art communication complexity compared to several key baselines using second-order information.

Rustem Islamov
Institut Polytechnique de Paris
islamov.ri@phystech.edu

Xun Qian
JD Explore Academy
xun.qian@kaust.edu.sa

Slavomír Hanzely
King Abdullah University of Science and Technology
slavomir.hanzely@kaust.edu.sa

Mher Safaryan
Institute of Science and Technology Austria
mher.safaryan@gmail.com

Peter Richtarik
King Abdullah University of Science & Technology (KAUST)
Peter.Richtarik@kaust.edu.sa

MS241

Personalized Asynchronous Federated and Decentralized Learning

We study the personalized federated and decentralized learning problem under asynchronous updates. In this problem, each client seeks to obtain a personalized model that simultaneously outperforms local and global models. We consider two optimization-based frameworks for personalization: (i) Model-Agnostic Meta-Learning (MAML) and (ii) Moreau Envelope (ME). MAML involves learning a joint model adapted for each client through fine-tuning, whereas ME requires a bi-level optimization problem with implicit gradients to enforce personalization via regularized losses. We focus on improving the scalability of personalized federated learning by removing the synchronous communication assumption. Moreover, we extend the studied function class by removing boundedness assumptions on the gradient norm. Our main technical contribution is a unified proof for asynchronous federated learning with bounded staleness that we apply to MAML and ME personalization frameworks. For the smooth and non-convex functions class, we show the convergence of our method to a first-order stationary point. We illustrate the performance of our method and its tolerance to staleness through experiments for classification tasks over heterogeneous datasets.

Taha Toghani
Rice University
mttoghani@gmail.com

Soomin Lee
Yahoo! Research
soominl@yahooinc.com

César A. Uribe
Rice University
cauribe@rice.edu

MS242

Structured Machine Learning for Spatial Environmental Data

Spatial generalized linear mixed-models are widely used for analyses of geospatial data. We consider the setting where the linearity assumption is inappropriate and consider flexible machine learning algorithms like random forests or deep neural networks. We propose well-principled extensions of these methods for non-linear modeling of spatial data. We still model the spatial correlation using Gaussian Process, as is traditional, and account for this dependence within the optimization procedure of these machine learning algorithms. The basic principle is guided by how ordinary least squares extends to generalized least squares for linear models to account for dependence. We demonstrate how the same extension can be done for these machine learning approaches like random forests and neural networks. We provide extensive theoretical and empirical support for the methods and show how they fare better than nave or brute-force approaches to use machine learning algorithms for spatially correlated data. We demonstrate the RandomForestsGLS R-package that implements this extension for random forests.

Abhi Datta
Johns Hopkins University, U.S.
abhidatta@jhu.edu

MS242

Continuous Monitoring of Parkinson Tremor in Real-Life: a Prototype-Based Approach

Parkinsons disease (PD) is the fastest-growing neurological disorder worldwide, with an expected doubling of patients to 12 million by 2040. Unobtrusive wearable sensors offer potentially critical improvements by allowing us to capture objectively and continuously how PD patients function in their daily life. As such, the use of wearable sensors to monitor PD is increasing, both in the (remote) care setting to provide guidance to healthcare providers, and in clinical trials to provide more sensitive outcome measures to detect treatment effects of potentially disease-modifying therapies. To enable passive monitoring of tremor using wearable sensors, we need models that can reliably detect and quantify different tremor episodes in the highly variable context of daily life. To address these challenges and develop a robust tremor model, we build on the idea of prototype-based learning, which we use to embed domain expertise in the decision boundaries of our model. The proposed framework uses radial basis function (RBF) layers to ensure local stationarity in the region of the selected prototypes, and to provide an interpretable representation of the model posterior. We propose a novel mechanism for training hierarchical Bayesian neural network architectures which minimizes the curvature of the loss function, as well as, incorporating the cause interpretation of prototypes to achieve desirable guarantees.

Yordan P. Raykov
University of Nottingham
yordan.raykov@nottingham.ac.uk

MS242

Network Inference Via Process Motifs for Lagged Correlation in Linear Stochastic Processes

A major challenge for causal inference from time-series data is the trade-off between computational feasibility and accuracy. Motivated by process motifs for lagged covariance in an autoregressive model with slow mean-reversion, we propose to infer networks of causal relations via pairwise edge measures (PEMs) that one can easily compute from lagged correlation matrices, and we formulate two PEMs that respectively correct for confounding factors and for reverse causation. To demonstrate the performance of our PEMs, we consider network inference from simulations of linear stochastic processes, and we show that our proposed PEMs can infer networks accurately and efficiently. Specifically, for autocorrelated time-series data, our approach achieves accuracies higher than or similar to Granger causality, transfer entropy, and convergent crossmapping—but with much shorter computation time than possible with any of these methods. Our fast and accurate PEMs are easy-to-implement methods for network inference with a clear theoretical underpinning. They provide promising alternatives to current paradigms for the inference of linear models from time-series data, including Granger causality, vector-autoregression, and sparse inverse covariance estimation.

Alice C. Schwarze
Dartmouth College
Alice.C.Schwarze@dartmouth.edu

MS243

Anisotropic Proximal Gradient

In this talk we present a novel algorithm for nonconvex composite minimization which can be interpreted in terms of dual space nonlinear preconditioning for the classical proximal gradient method. The proposed scheme can be applied to additive composite minimization problems whose smooth part exhibits an anisotropic descent inequality relative to a reference function. In the convex case this is a dual characterization of relative strong convexity in the Bregman sense. It is proved that the anisotropic descent property is closed under pointwise average if the dual Bregman distance is jointly convex and, more specifically, closed under pointwise conic combinations for the KL-divergence. We analyze the methods asymptotic convergence and prove its linear convergence under an anisotropic proximal gradient dominance condition. This is implied by anisotropic strong convexity, a recent dual characterization of relative smoothness in the Bregman sense. Applications are discussed including exponentially regularized LPs and logistic regression with nonsmooth regularization. In the LP case the method can be specialized to the Sinkhorn algorithm for regularized optimal transport and a classical parallel update algorithm for AdaBoost. Complementary to their existing primal interpretations in terms of entropic subspace projections this provides a new dual interpretation in terms of forward-backward splitting with entropic preconditioning.

Emanuel Laude
Katholieke Universiteit Leuven
emanuel.laude@esat.kuleuven.be

Panagiotis Patrinos
KU Leuven
panos.patrinos@esat.kuleuven.be

MS243

Explicit Convergence Rates under Consistent Error Bounds and Beyond

In this talk, we introduce the notion of consistent error bound functions which provides a unifying framework for error bounds for multiple convex sets. This framework goes beyond the classical Lipschitzian and Holderian error bounds and includes logarithmic and entropic error bounds found in the exponential cone. Our main result is to obtain explicit convergence rates of a wide class of algorithms including projection algorithms for convex feasibility problems. In particular, Karamata theory in regular variations is first used as a powerful tool to analyze the convergence rates in optimization.

Tianxiang Liu
Tokyo Institute of Technology
liu.t.af@m.titech.ac.jp

Bruno Lourenço
The Institute of Statistical Mathematics
bruno@ism.ac.jp

MS243

First Order Methods for Minmax Optimisation

The point-wise maximum of finitely many smooth functions is ubiquitous in convex optimisation, since many non-smooth functions can be written in this form. However, since this maximum is non-differentiable, standard first order methods are unavailable. In this talk, we discuss how to exploit the differentiability of the functions inside the maximum by reformulating the non-smooth convex minimisation problem as a smooth saddle point problem. Even if each function has a globally Lipschitz continuous gradient, this is not preserved in the saddle formulation, and so we apply several methods for variational inequalities which only assume local Lipschitz continuity. In particular, the celebrated adaptive Golden RAtio ALgorithm (aGRAAL), which approximates the inverse of a local Lipschitz constant fully explicitly, ie, without backtracking. We therefore compare the aGRAAL against other established methods. Time permitting, we will discuss several applications, including relay location for wireless sensor networks

Daniel Uteda
The University of Melbourne
duteda@student.unimelb.edu.au

MS244

Modeling Combinatorial Disjunctive Constraints Via Junction Trees and Applications in Robotics

In this work, we study the independent-branching (IB) framework of combinatorial disjunctive constraints (CDCs) and identify a class of pairwise IB-representable CDCs: combinatorial disjunctive constraints admitting junction trees. For this class of constraints, the existence of junction trees can be recognized in polynomial time. We also present a polynomial-time heuristic algorithm for the minimum biclique cover problem on the associated conflict graphs to build small and strong mixed-integer programming (MIP) formulations. Furthermore, we provide a novel ideal extended formulation of any combinatorial disjunctive constraints with fewer auxiliary binary variables. In this talk, we also cover applications of this framework including building MIP formulations of planar obstacle avoidances and piecewise linear relaxations of univariate nonlinear functions such as sine and cosine appearing in rotation matrices.

Bochuan Lyu, Illya V. Hicks
Rice University
bl46@rice.edu, ivhicks@rice.edu

Joey Huchette
Google Research
jhuchette@google.com

MS244

Theoretical Advances in Solving Markov Decision Processes

Markov Decision Processes (MDP) are a fundamental mathematical model for reasoning about uncertainty and a prominent structured optimization problem. Over the past 5 years there have been great strides in provably computing approximately optimal policies for MDPs in a variety of settings. In this talk I will survey these advances touching upon a variety of optimization tools of broader utility.

Aaron Sidford
Stanford University
sidford@stanford.edu

MS244

Discrete and Continuous Optimization for Motion Planning and Control: Applications and Recent Progress

Control problems that blend discrete and continuous variables arise in a number of application areas, from robotics, to power systems, autonomous vehicles, and energy grids. This talk will survey modern approaches to such problems, highlighting the interplay between applications, theory, and practice. Of particular interest will be problem settings in which discrete and continuous decisions cannot be decoupled, and instead must be made jointly. In these settings, there is a need for tractable, optimization-friendly problem abstractions. The talk will discuss the requirements for such abstractions and close by introducing recent attempts to fill this gap.

Jack Umenberger
MIT
umnbrgr@mit.edu

MS245

Efficiently Solving QUBO Problems

Efficient solution approaches for quadratic unconstrained binary optimization (QUBO) problems are of great interest for several reasons. Firstly, the well-known maximum-cut problem admits a natural formulation as a QUBO problem. Secondly, other optimization problems like the graph bisection problem, the maximum independent set problem, linearly constrained binary quadratic problems, and many more can be reformulated as an instance of QUBO. Since linear programming based approaches perform poorly on dense problem instances, all state-of-the-art approaches involve semidefinite relaxations in this case. In this talk, we present a branch-and-cut solver for QUBO problems based on semidefinite programming. It uses the so-called *mixing method*, a low-rank coordinate descent method, as its main tool to tackle the occurring semidefinite programs. We discuss some new ideas implemented in the solver and provide numerical results showing that it outperforms the other existing semidefinite approaches. Moreover, it was a paradigm in recent years to tighten semidefinite relaxations by considering more and more valid inequalities. In contrast to this, we demonstrate that weaker relaxations can still be competitive when used in a branch-and-bound approach, even for medium-sized instances.

Jan Schwiddessen
Alpen-Adria-Universität Klagenfurt
jan.schwiddessen@aau.at

MS245

An SDP-SOS Based MAX-SAT Solver

We consider semidefinite programming (SDP) approaches for solving the well-known Satisfiability (SAT) problem and the related maximum-satisfiability (MAX-SAT) problem. SAT is a well-known binary decision problem stemming from computer science with a wide range of applications. It is widely known that SDP is well-suited to approximate (MAX-)2-SAT instances, a class of (MAX-)SAT instances similar to Max-Cut. Our work shows the potential of SDP on (MAX-)3-SAT instances as well, by being competitive with some of the best solvers in the yearly MAX-SAT competition. These results are obtained by a tailored SDP solver and parser, in combination with a branch and cut scheme. Moreover, we elegantly demonstrate a new connection between two previous approaches of combining SDP and SAT. This is joint work with Renata Sotirov.

Lennart Sinjorgo, Renata Sotirov
Tilburg University
l.m.sinjorgo@tilburguniversity.edu, r.sotirov@uvt.nl

MS245

Semidefinite Representable Reformulations for Two Variants of the Trust-Region Subproblem

Motivated by encouraging numerical results in the literature, in this talk we consider two specific variants of the trust-region subproblem and provide exact semidefinite representable reformulations. The first is over the intersection of two balls; the second is over the intersection of a ball and a special second-order conic representable set. The reformulations are based on partitions of the feasible regions into sub-regions with known lifted convex hulls.

Sarah Kelly
Clemson University
skelly5@g.clemson.edu

Yuyuan "Lance" Ouyang
Department of Mathematical and Statistical Sciences
Clemson University
yuyuano@clemson.edu

Boshi Yang
Clemon University
boshiy@clemson.edu

MS246

Some Unsolved Problems in Mathematical Programming

In this talk I will describe several open problems that are related to my own research. The topics include algorithmic complexity, convex representations of nonconvex problems and computational challenges. Although these problems remain open they all seem to be nearly within reach and are, in my opinion, excellent candidates for further research.

Kurt M. Anstreicher
University of Iowa
Department of Management Sciences
kurt-anstreicher@uiowa.edu

MS246

Single-Line Drawings via Linear Optimization

We present a survey of approaches to using linear optimization to design single-line drawings. Examples include TSP Art (continuous line drawings obtained by finding high-quality solutions to carefully constructed instances of the Traveling Salesperson Problem), Figurative Tours, Knights Tours, and Easily-Traversed Eulerian Trails.

Robert Bosch
Oberlin College
robert.a.bosch@oberlin.edu

MS246

Applications of Integer Programming and Reinforcement Learning in Data Center Planning and Operations

In this talk we present, in real-world practices, a class of optimization problems in large data center capacity planning and operation, that has significant impacts on resource utilization efficiency and system reliability. We describe how hierarchy of integer programing models were designed to solve large-scale problems and how reinforcement learning methods were used to solve long-term strategic questions stochasticity in supply and demand. We also describe our engineering efforts in model diagnosis, such model infeasibility caused by data inaccuracies.

Hongbo Dong
Meta
hongbodong@fb.com

MS247

Worst-Case Complexity for the Minimization of Strict Saddle Functions on Manifolds

Many problems in data science can be expressed as the minimization of a strict saddle function. Strict saddle functions are in general nonconvex, but have the property that the Hessian has a negative eigenvalue at every saddle point. Using this information on the landscape, we design an efficient and specific algorithm for the minimisation of strict saddle functions. We derive guarantees on the number of iterations that have to be performed in the worst-case before an approximate minimizer of the function is reached. We show how our algorithm improves on the classical complexity results of nonconvex optimization.

Florentin Goyens
The Alan Turing Institute for Data Science
goyensflorentin@gmail.com

MS247

A Consistently Adaptive Trust-Region Method

Adaptive trust-region methods attempt to maintain strong convergence guarantees without depending on conservative estimates of problem properties such as Lipschitz constants. However, on close inspection, one can show existing adaptive trust-region methods have theoretical guarantees with severely suboptimal dependence on problem properties such as the Lipschitz constant of the Hessian. For example, TRACE developed by Curtis et al. obtains a $O(\Delta_f L^{3/2}\epsilon^{-3/2}) + \tilde{O}(1)$ iteration bound where L is the Lipschitz constant of the Hessian. Compared with the optimal $O(\Delta_f L^{1/2}\epsilon^{-3/2})$ bound this is suboptimal with respect to L. We present the first adaptive trust-region method which circumvents this issue and requires at most $O(\Delta_f L^{1/2}\epsilon^{-3/2}) + \tilde{O}(1)$ iterations to find an ϵ-approximate stationary point, matching the optimal iteration bound up to an additive logarithmic term. Our method is a simple variant of a classic trust-region method and in our experiments performs competitively with both ARC and a classical trust-region method.

Fadi Hamad
University of Pittsburgh
fah33@pitt.edu

MS247

Quartic Polynomial Sub-Problem Solutions in Tensor Methods for Nonconvex Optimization

There has been growing interest in high-order tensor methods for nonconvex optimization in machine learning as these methods provide better/optimal worst-case evaluation complexity, stability to parameter tuning, and robustness to problem conditioning. The well-known pth-order adaptive regularization (ARp) method relies crucially on repeatedly minimising a nonconvex multivariate Taylor-based polynomial sub-problem. It remains an open question to find efficient techniques to minimise such a sub-problem for $p \geq 3$. We will present a second-order method (SQO) for the AR3 (ARp with $p = 3$) sub-problem. SQO approximates the special-structure quartic polynomial sub-problem from above and below by using second-order models that can be minimised efficiently and globally. We prove that SQO finds a local minimiser of a quartic polynomial, but in practice, due to its construction, it can find a much lower minimum than cubic regularization approaches. This encourages us to continue our quest for algorithmic techniques that find approximately global solutions for such polynomials. This work is joint with Coralia Cartis.

Wenqi Zhu
Oxford University, United Kingdom
wenqi.zhu@maths.ox.ac.uk

MS248

Dyadic Linear Programming

A finite vector is dyadic if each of its entries is a dyadic rational number, i.e. if it has an exact and finite floating point representation. We study the problem of finding a dyadic optimal solution to a linear program, if one exists: We characterize when a dyadic optimal solution exists, and in case it does, we provide a polynomial time algorithm for solving the problem. We also study some related problems.

Ahmad Abdi
London School of Economics and Political Science
a.abdi1@lse.ac.uk

Bertrand Guenin
University of Waterloo
bguenin@math.uwaterloo.ca

Levent Tuncel
University of Waterloo
Dept. of Combinatorics and Optimization
ltuncel@uwaterloo.ca

MS248

A Randomized Approximation Algorithm for the Weighted Fractional Cut-Covering Problem

The fractional cut-covering problem is the linear programming relaxation for the problem of covering each edge by a set of cuts to meet given edge weights. We define a semidefinite programming relaxation of fractional cut covering problem whose approximate optimal solutions may be rounded into a fractional cut cover via a randomized polynomial-time algorithm. These results arise from the tight connection between fractional cut covering problem and the maximum cut problem. By clarifying this relationship via antiblocking duality and gauge duality, not only do we obtain dual results to the celebrated work of Goemans and Williamson on maxcut, but also tie the approximation constants for these two problems to each other, obtain new optimality certificates, and relate both problems to geometric representation of graphs.

Nathan Benedetto Proenca
University of Waterloo
nathan.benedettoproenca@uwaterloo.ca

Marcel de Carli Silva
University of Sao Paulo Brazil
mksilva@ime.usp.br

Cristiane Maria Sato
Federal University of ABC, Brazil
cmsato@gmail.com

Levent Tuncel
University of Waterloo
Dept. of Combinatorics and Optimization
ltuncel@uwaterloo.ca

MS248

On the Column Number and Excluded Submatrices for Delta-Modular Matrices

Totally unimodular constraint matrices have been used for decades to model many problems in combinatorial optimization; these matrices are particularly useful as then integer programs can be solved as linear ones. Recent results, in particular one due to Artmann et al., has spurred the study of more general Delta-modular constraint matrices defined to have minors within a fixed range. In this talk, we explore a simple property of Delta-modular matrices: "how many distinct columns they can have?". We give an overview of this question and provide recent results, and in particular we identify some excluded sub matrices that can be used to address this question. Finally, we demonstrate how this column number can be used in integer programming.

Joseph Paat
Sauder School of Business, UBC
joseph.paat@sauder.ubc.ca

MS249

The Small-Noise Limit of the Most Likely Element is the Most Likely Element of the Small-Noise Limit

The characterization of rare event probabilities in infinite dimensional or path spaces is complicated, owing to the fact that there is no natural notion of a uniform measure on these spaces. The Onsager-Machlup (OM) function can serve as a density on infinite dimensional spaces, and has been seen as the Lagrangian for the "most likely element", and could be used to characterize rare event probabilities. On the other hand, small-noise limits are often used to establish large deviations (LD) principles that characterize rare event probabilities in another sense. In this setting, the Freidlin-Wentzell (FW) function is the LD rate function, and has also been identified as a Lagrangian for the "most likely element". This leads to a conundrum - what is the relationship between these two functionals? In this talk, I will present our recent work on the relationship between the OM and FW functions for measures equivalent to arbitrary infinite dimensional Gaussian measures. We show both pointwise and Γ-convergence of the Onsager-Machlup function under the small-noise limit to the Freidlin-Wentzell function - and give an expression for both. That is, we show that the small-noise limit of the most likely element is the most likely element in the small noise limit in this setting. Examples of measures include the law of solutions to stochastic differential equations or the law of an infinite system of random algebraic equations. This is joint work with Zachary Selk.

Harsha Honnappa
Purdue University
honnappa@purdue.edu

MS249

Network Cascade Vulnerability using Constrained Bayesian Optimization

Measures of power grid vulnerability are often assessed by the amount of damage an adversary can exact on the network. However, the cascading impact of such attacks is often overlooked, even though cascades are one of the primary causes of large-scale blackouts. This paper explores modifications of transmission line protection settings as candidates for adversarial attacks, which can remain undetectable as long as the network equilibrium state remains unaltered. This forms the basis of a black-box function in a Bayesian optimization procedure, where the objective is to find protection settings that maximize network degradation due to cascading. Notably, our proposed method is agnostic to the choice of the cascade simulator and its underlying assumptions. Extensive experiments reveal that, against conventional wisdom, maximally misconfiguring the protection settings of all network lines does not cause the most cascading. More surprisingly, even when the degree of misconfiguration is limited due to resource constraints, it is still possible to find settings that produce cascades comparable in severity to instances where there are no resource constraints.

Albert Lam
Argonne National Laboratory
albert3033@gmail.com

Anirudh Subramanyam
Pennsylvania State University
subramanyam@psu.edu

MS250

Optimization on Manifolds via Graph Gaussian Processes

We integrate manifold learning techniques within a Gaussian process upper confidence bound algorithm to optimize an objective function on a manifold. Our approach is motivated by applications where a full representation of the manifold is not available and querying the objective is expensive. We rely on a point cloud of manifold samples to define a graph Gaussian process surrogate model for the objective. Query points are sequentially chosen using the posterior distribution of the surrogate model given all previous queries. We establish regret bounds in terms of the number of queries and the size of the point cloud. Several numerical examples complement the theory and illustrate the performance of our method.

Hwanwoo Kim, Daniel Sanz-Alonso
University of Chicago
hwkim@uchicago.edu, sanzalonso@uchicago.edu

Ruiyi Yang
Princeton University
ry8311@princeton.edu

MS250

Tensor Decompositions Using Stochastic and Deterministic Optimization

Low-rank tensor decompositions in general, and canonical polyadic (CP) tensor decompositions specifically, are now ubiquitous in the area of multi-way data analysis. The task of fitting a low-rank non-negative CP model to Poisson-distributed count data via tensor maximum likelihood estimation is often formulated as a nonlinear, nonconvex optimization problem. Yet several popular algorithms use local search methods to approximate the global maximum likelihood estimator from local minima. Recent research in developing efficient local methods reflects an emergent dichotomy in numerical linear algebra: deterministic versus randomized algorithms. CP Alternating Poisson Regression (CPAPR) is a deterministic CP tensor decomposition method that alternates over a sequence of convex Poisson loss subproblems iteratively. Generalized CP (GCP) tensor decompositions extend previous work on CP decompositions to incorporate general loss functions and stochastic optimization methods. Previously we showed that CPAPR is performant and can compute accurate maximum likelihood estimator approximations with a higher probability than GCP. Here we explore a hybrid method for computing CP tensor decompositions that leverages GCP for scalability and CPAPR for performance and accuracy. Our approach is inspired by Adaptive Simulated Annealing for global optimization and allows for fine-grain parameter tuning as well as adaptive updates to algorithm parameters.

Jeremy Myers
The College of WIlliam and Mary
jmmyers01@email.wm.edu

Daniel M. Dunlavy
Sandia National Laboratories
Computer Science and Informatics
dmdunla@sandia.gov

MS251

Quadratic Envelopes and ℓ_0; How to Find the Oracle Solution

We introduce a new operator called the quadratic envelope, used to regularize discontinuous penalties such as the l0 norm. The quadratic envelope enjoys several desirable properties, such as removing local minimizers whilst not moving the global minimizer, under suitable conditions. When applied to the l0-norm we obtain the so called Minimax Concave Penalty (MCP), which has been extensively studied in prior works by different authors. We extend the known theory, in particular showing concrete conditions ensuring that the global minimizer of the arising functional is the so called oracle solution. This result is much stronger than previous contributions demonstrating that the MCP-solution enjoy “oracle properties’ asymptotically, i.e. as the dimensions approach infinity. We also discuss how to alter the l0 norm in ways that removes the problem of high cardinality local minimizers where algorithms have a tendency to get stuck.

Marcus Carlsson
Lund University, Lund, Sweden
marcus.carlsson@math.lu.se

MS251

Hidden Convexity in the ℓ_0 Pseudonorm

The so-called ℓ_0 pseudonorm counts the number of nonzero components of a vector. In this talk, we review a series of recent results on a class of Capra (Constant Along Primal Rays) conjugacies that reveal hidden convexity in the ℓ_0 pseudonorm . First, we present the Euclidean Capra-conjugacy. We show that it is suitable to analyze the ℓ_0 pseudonorm, as this latter is ”convex” in the sense of generalized convexity (equal to its biconjugate). We immediately derive a convex factorization property (the ℓ_0 pseudonorm coincides, on the unit sphere, with a convex lsc function) and variational formulations for the ℓ_0 pseudonorm. We present mathematical expressions of the Capra-subdifferential of the ℓ_0 pseudonorm, and graphical representations. In a second part, we provide different extensions. We introduce the class of Capra-conjugacies defined by means of norms — especially strictly-orthant monotonic norms (including the Euclidean norm) — or, more generally, of 1-homogeneous nonnegative functions. We show that such Capra-conjugacies are suitable to analyze, not only the ℓ_0 pseudonorm, but provide convex lower bounds for 0-homogeneous functions. We will also point out how to tackle the rank matrix function. Finally, we discuss how the theory could open the way for possible algorithms in sparse optimization problems.

Michel De Lara
Ecole des Ponts ParisTech, France
michel.delara@enpc.fr

Jean-Philippe Chancelier
CERMICS, Ecole des Ponts ParisTech
jean-philippe.chancelier@enpc.fr

MS252

Sifting Through the Noise: Universal First-Order

Methods for Stochastic Variational Inequalities

We examine a flexible algorithmic framework for solving monotone variational inequalities in the presence of randomness and uncertainty. The proposed template encompasses a wide range of popular first-order methods, including dual averaging, dual extrapolation and optimistic gradient algorithms both adaptive and non-adaptive. Our first result is that the algorithm achieves the optimal rates of convergence for cocoercive problems when the profile of the randomness is known to the optimizer: $O(1/\sqrt{T})$ for absolute noise profiles, and $O(1/T)$ for relative ones. Subsequently, we drop all prior knowledge requirements (the absolute/relative variance of the randomness affecting the problem, the operators cocoercivity constant, etc.), and we analyze an adaptive instance of the method that gracefully interpolates between the above rates i.e., it achieves $O(1/T)$ and $O(1/T)$ in the absolute and relative cases, respectively. To our knowledge, this is the first universality result of its kind in the literature and, somewhat surprisingly, it shows that an extra-gradient proxy step is not required to achieve optimal rates.

Kimon Antonakopolous
EPFL
kimon.antonakopoulos@epfl.ch

MS252

RandProx: Primal-Dual Optimization Algorithms with Randomized Proximal Updates

Proximal splitting algorithms are well suited to solving large-scale nonsmooth optimization problems. We propose Randprox, a new generic primal-dual algorithm, in which the dual update is randomized. For instance, some randomly chosen dual variables, instead of all, are updated at each iteration. Or, the proximity operator of a function is called with some small probability only. Thanks to a new variance-reduction mechanism, the algorithm converges to an exact minimizer. We derive linear convergence results in presence of strong convexity, which are new even in the deterministic case. Some randomized algorithms of the literature are recovered as particular cases (e.g., Point-SAGA). But our randomization technique is general and encompasses many unbiased mechanisms beyond sampling and probabilistic updates, including compression. It has long been known for stochastic-gradient-type methods that randomness helps getting faster algorithms. Our work opens the door to benefiting from randomness in the more general primaldual setting as well.

Laurent Condat
King Abdullah University of Science and Technology (KAUST)
laurent.condat@kaust.edu.sa

Peter Richtarik
King Abdullah University of Science & Technology (KAUST)
Peter.Richtarik@kaust.edu.sa

MS252

Convergence of Chambolle-Pock and Douglas-Rachford Splitting in the Absence of Monotonicity

Splitting methods have seen widespread use in a variety of applications in optimization and for solving variational inequalities. Despite this popularity, convergence results of these methods have been largely limited to the convex/monotone setting. In this work, we study two such methods in the absence of the monotonicity assumption and without requiring smoothness, namely DouglasRachford splitting (DRS) and the Chambolle-Pock (CP) algorithm. To this end, we introduce the concept of semimonotonicity and provide sufficient conditions for global convergence of DRS and CP involving the sum of two semimonotone operators. Most notably, it is shown that DRS and CP converge even when the sum of the involved operators (or of their inverses) is non-monotone. Our analysis relies on establishing a connection between both methods and the preconditioned proximal point algorithm applied to a corresponding underlying operator satisfying an oblique weak Minty assumption.

Panagiotis Patrinos, Brecht Evens, Pieter Pas, Puya Latafat
KU Leuven
panos.patrinos@esat.kuleuven.be, brecht.evens@kuleuven.be, pieter.pas@kuleuven.be, puya.latafat@kuleuven.be

MS253

Cooperation and Cost Sharing Problems in Supply Networks

Across several contexts such as supply chain security or traceability, costly actions by firms (e.g., decision to invest in security or decision to adopt traceability technologies) can yield payoffs to other firms in the supply chain. Such positive externalities imply that network-wide cooperative strategies can yield improvements over firms independently choosing individually rational actions. However, cooperation can be hindered by disagreements over cost-sharing arrangements. A priori, it is not clear whether there will always exist a stable and fair sharing of costs that can sustain network-wide cooperation. Furthermore, networked firms typically have visibility and mechanisms to cooperate and monitor with only immediate partners. Extended multi-tier supply chains are often associated with a potential loss in visibility over firms further away in the network. Thus, it is also unclear whether one can find suitable mechanisms to implement cost-sharing arrangements that circumvent coordination across firms that are not immediate or direct partners. In this talk, we review two recent applications (supply network security and traceability) and develop a general framework to identify implementable cost sharing mechanisms that can sustain network-wide cooperative actions.

Sanjith Gopalakrishnan
Desautels Faculty of Management
McGill University
sanjith.gopalakrishnan@mcgill.ca

Sriram Sankaranarayanan
IIM Ahmedabad

srirams@iima.ac.in

MS254

Comparing Solution Paths of Sparse Quadratic Minimization with a Stieltjes Matrix

This paper studies several solution paths of sparse quadratic minimization problems as a function of the weighing parameter of the bi-objective of estimation loss versus solution sparsity.

Ziyu He
Department of Industrial and Systems Engineering
University of Southern California
ziyuhe@usc.edu

MS254

Novel Algorithms for Nonconvex Second-order Optimization with Complexity Guarantees

Second-order optimization has recently experienced significant developments, leading to numerous fruitful applications in science and engineering. In particular, recent research has shown that a second-order stationary point (SOSP) of a nonconvex optimization problem is often a globally optimal solution for instances that arise in areas such as machine learning and statistics. Therefore, developing efficient algorithms for computing such points is pivotal for advancing those areas. Our research introduces new algorithms with substantial theoretical improvements for solving two types of nonconvex constrained optimization problems and conducts numerical studies to show the practical advantages of the proposed methods over the state-of-the-art methods. In our first work, we develop a novel augmented Lagrangian (AL) method for finding an SOSP of a nonconvex equality constrained optimization problem. The proposed AL method improves upon the best-known complexity guarantees. Moreover, the computational speed of our AL method is vastly faster than the competing ones for solving a classical statistical problem. In our second work, we design efficient algorithms to find an SOSP of a general conic constrained nonconvex optimization problem. For the first time, we introduce a notion of an SOSP for general conic constrained optimization and propose an efficient algorithm to find it. Numerical results demonstrate the significant potential superiority of our barrier-AL method over the state-of-the-art method in terms of solution quality.

Chuan He, Zhaosong Lu
University of Minnesota
he000233@umn.edu, zhaosong@umn.edu

MS254

An Inexact Augmented Lagrangian Algorithm for Training Leaky Relu Neural Network with Group Sparsity

The leaky ReLU network with a group sparse regularization term has been widely used in the recent years. However, training such a network yields a nonsmooth nonconvex optimization problem and there exists a lack of approaches to compute a stationary point deterministically. In this paper, we first resolve the multi-layer composite term in the original optimization problem by introducing auxiliary variables and additional constraints. We show the new model has a nonempty and bounded solution set and its feasible set satisfies the Mangasarian-Fromovitz constraint qualification. Moreover, we show the relationship between the new model and the original problem. Remarkably, we propose an inexact augmented Lagrangian algorithm for solving the new model and show the convergence of the algorithm to a KKT point. Numerical experiments demonstrate that our algorithm is more efficient for training sparse leaky ReLU neural networks than some well-known algorithms.

Wei Liu
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
liuwei175@lsec.cc.ac.cn

MS255

A Definition of Non-Stationary Bandits

The subject of non-stationary bandit learning has attracted much recent attention. However, non-stationary bandits lack a formal definition. Loosely speaking, non-stationary bandits have typically been characterized in the literature as those for which the reward distribution changes over time. We demonstrate that this informal definition is ambiguous. Further, a widely-used notion of regret—the dynamic regret—is motivated by this ambiguous definition and thus problematic. In particular, even for an optimal agent, dynamic regret can suggest poor performance. The ambiguous definition also motivates a measure of the degree of non-stationarity experienced by a bandit, which often overestimates and gives rise to extremely loose regret bounds. The primary contribution of this paper is a formal definition that resolves ambiguity. This definition motivates a new notion of regret, an alternative measure of the degree of non-stationarity, and a regret analysis that leads to tighter bounds for non-stationary bandit learning. The regret analysis applies to any bandit, stationary or non-stationary, and any agent.

Yueyang Liu, Benjamin Van Roy, Kuang Xu
Stanford University
yueyl@stanford.edu, bvr@stanford.edu,
kuangxu@stanford.edu

MS255

Epi-Convergent Data-Driven Approximation: A Unified View of Pessmestic and Optimistic

In modern decision-making, there is often a need to handle a stochastic programming problem with an uncertain distribution. To address this challenge, the standard approach is to use a data-driven approximation of the unknown distribution through the simple sample average approximation. However, this method can still result in a failure to reach the optimal solution in the case of nonsmooth optimization problems. To overcome this issue, many studies further examine the behavior of the objective function within the ambiguity set and construct an epi-convergent data-driven approximation. In this paper, we bring together Distributionally Robust Optimization, which focuses on a conservative approach, and Rockafellian relaxation, a recently proposed optimistic framework, to create asymptotic exact approximations. We demonstrate that Distributionally Robust Optimization can be reduced to a Rockafellian relaxation, and as a result, we achieve benefits on both sides. Firstly, we provide a novel epi-convergent proof with weaker assumptions for the distributionally robust data-driven approximation. Secondly, one can design efficient optimization procedures for Rockafellian if it has an equivalent distributionally robust optimization formulation.

Yiping Lu, Jiajin Li, Jose H. Blanchet
Stanford University
yplu@stanford.edu, jiajinli@stanford.edu, jose.blanchet@stanford.edu

MS255

Randomization of Spectral Risk Measure and Distributional Robustness

In this talk, we consider a situation where a decision maker's risk preference can be described by a spectral risk measure (SRM) but there is not a single SRM which can be used to represent the DM's preferences consistently. Consequently we propose to randomize SMR by introducing a random parameter in risk spectrum. The randomized SRM (RSMR) allows one to describe the DM's preferences at different states with different SRMs. When the distribution of the random parameter is known, i.e., the randomness of the DM's preference can be described by a probability distribution, we introduce a new risk measure which is the mean value of the RSMR. In the case when the distribution is unknown, we propose a distributionally robust formulation of RSMR. The RSRM paradigm provides a new framework for interpreting Kusuoka's representation theorem and addressing inconsistency issues arising from observation/measurement errors or erroneous responses in preference elicitation process. We discuss in detail computational schemes for solving optimization problems based on RSMR and distributionally robust RSMR.

Huifu Xu
The Chinese University of Hong Kong
hfxu@se.cuhk.edu.hk

Manlan Li
Xiangtan University
liml@smail.xtu.edu.cn

Xiaojiao Tong
Hunan First Normal University
dysftxj@hnfnu.edu.cn

MS256

Using semidefinite programming relaxations in branch and cut methods for MIS problems

Maximum independent set (MIS) problems on graphs with as few as one thousand nodes can be surprisingly challenging to solve using integer linear programming branch-and-cut solvers. Semidefinite programming relaxations can provide stronger bounds than LP relaxations and are an intriguing alternative. However, there are significant issues in incorporating semidefinite programming relaxations within a branch-and-cut solver. The choice of SDP relaxation and cuts affects the quality of the bounds and the computational effort required to obtain the bounds. The choice of using an interior point solver versus a first order method such as ADMM involves a trade-off between more accurate subproblem solutions and faster, more memory efficient approximate solutions. There are challenges in dealing with subproblems that might not satisfy Slater's condition or strict duality. Warm starting the solution of a subproblem must also be considered. Approaches to these issues will be compared and evaluated using computational experiments. Results on some challenging and previously unsolved problems will be presented.

Brian Borchers
Department of Mathematics
New Mexico Tech
borchers@nmt.edu

MS257

Erratum, counterexample and an additional revealing poll step for a result of "Analysis of direct searches for discontinuous functions"

This talk provides a counterexample to a theorem announced in the last part of the paper "Analysis of direct searches for discontinuous functions", Mathematical Programming Vol. 133, pp. 299–325, 2012. The counterexample involves a one-dimensional objective function f which satisfies all the assumptions required by the theorem but contradicts some of its conclusions. A corollary of this theorem is also affected by this counterexample. This talk also discusses the exactitude of each statements in the theorem in general. Finally this talk introduces a modification of the directional direct search method (dDSM) that allows to recover the properties broken by the counterexample. The main flaw revealed by the counterexample is the possibility that a dDSM generates a sequence of trial points (xk) converging to a point x* at which f is discontinuous and whose objective function value is strictly less than the limit of the sequence (f(xk)). Moreover the dDSM generates no trial point in only one of the branches of f near x*.

Pierre-Yves Bouchet
Polytechnique Montréal
pierre-yves.bouchet@polymtl.ca

Charles Audet
École Polytechnique de Montréal - GERAD
Charles.Audet@gerad.ca

Loïc Bourdin
Université de Limoges
loic.bourdin@unilim.fr

MS257

Convergence Analysis of Evolution Strategies with Covariance Matrix Adaptation (CMA-ES) via

Markov Chain Stability Analysis

In black-box optimization, Evolution Strategies (ES) constitute a class of stochastic algorithms which, unlike gradient-based methods, do not use the derivatives of the objective function f. Instead, only the f-value ranks of different candidates are used. Covariance Matrix Adaptation Evolution Strategies (CMA-ES) have the particularity to even approximate the inverse Hessian of the objective function without using its derivatives. It is widely used in practice. We discuss in this presentation the convergence analysis of CMA-ES by establishing the stability of an underlying Markov chain. This approach has already been successful to prove the stability of other Evolution Strategy variants, in particular when only a step-size but no covariance matrix is adapted. We explain how this approach extends to CMA-ES.

Armand Gissler
INRIA
armand.gissler@inria.fr

MS257

Stochastic Bilevel Multi-Objective Optimization

In this work, we propose different stochastic gradient-based algorithms for bilevel multi-objective problems with conflicting objectives at the upper level, lower level, or both. Such problems have never been studied from a stochastic approximation viewpoint despite the recent advances in stochastic methods for single-level, bilevel, and multi-objective optimization. A case where past methodology is not applicable is bilevel optimization with a multi-objective lower level problem, where different approaches can be followed depending on the interpretation of the lower level optimality. An approach previously introduced is called optimistic and consists of minimizing the upper level function over all non-dominated lower level solutions. We introduce two other approaches, one risk neutral and the other risk averse, where one needs to sample from such a Pareto set, for which the closed-form expression may be unknown. Bilevel stochastic optimization has recently been adopted to solve fair machine learning (ML) problems, where the goal is to minimize the prediction error on validation data by training an ML model to avoid discriminatory predictions against people with sensitive attributes. To ensure accurate and fair prediction outcomes in real-life decision-making applications, accuracy and fairness loss functions must be jointly considered, thus leading to bilevel problems where at least one of the two optimization levels has more than one objective.

Luis Nunes Vicente, Tommaso Giovannelli, Griffin Kent
Lehigh University
lnv218@lehigh.edu, tog220@lehigh.edu, gdk220@lehigh.edu

MS258

Graph and Distributed Extensions of the Douglas-Rachford Splitting with Applications

In this talk, we present several graph-based extensions of the Douglas-Rachford splitting method to solve monotone inclusion problems involving the sum of N maximal monotone operators. The resulting schemes can be understood as unconditionally stable frugal resolvent splitting methods with minimal lifting, as well as instances of the (degenerate) Preconditioned Proximal Point method, which provides robust convergence guarantees. Further, we show that the proposed methods admit efficient distributed implementations, which respect the prescribed communication constraint. We conclude with an application to a congested optimal transport problem and to distributed Support Vector Machines, which show an interesting dependence on the underlying graph topology and highly competitive performances with state-of-the-art distributed optimization approaches.

Enis Chenchene
Institute of Mathematics and Scientific Computing
University of Graz, Heinrichstraße 36, 8010 Graz, Austria
enis.chenchene@uni-graz.at

Kristian Bredies
University of Graz
Institute of Mathematics and Scientific Computing
kristian.bredies@uni-graz.at

Emanuele Naldi
TU Braunschweig
e.naldi@tu-braunschweig.de

MS258

Convergence Analysis of Davis-Yin Splitting Via Scaled Relative Graphs

Davis-Yin splitting (DYS) has found a wide range of applications in optimization, but its linear rates of convergence have not been studied extensively. The scaled relative graph (SRG) simplifies the convergence analysis of operator splitting methods by mapping the action of the operator onto the complex plane, but the prior SRG theory did not fully apply to the DYS operator. In this work, we formalize an SRG theory for the DYS operator and use it to obtain tighter contraction factors.

Jongmin Lee
Department of mathematical science
Seoul National University
dlwhd2000@snu.ac.kr

Ernest K. Ryu
Seoul National University, South Korea
ernestryu@snu.ac.kr

Soheun Yi
Department of mathematical science
Seoul National University
lsdluis@snu.ac.kr

MS258

Adaptive Stepsizes for Primal-Dual Douglas-Rachford Splitting

In this talk we discuss the possibility of finding a minimum of $f(x) + g(Kx)$ for two proper, convex and lower semicontinuous functions and a matrix K, using a non-stationary version of the Douglas-Rachford algorithm. The presented Douglas-Rachford method, applied to primal-dual optimality conditions, uses not only one primal but also a second dual stepsize sequence. We allow both the primal and the dual stepsize sequences to vary independently from each other and present the outline of a variable metric proximal point algorithm-based proof. The talk continues with the introduction of an eigenvalue based heuristic which motivates an adaptive way of choosing both stepsize sequences in order to increase the convergence speed of the algorithm.

Jannis Marquardt
Institute for Partial Differential Equations
TU Braunschweig
j.marquardt@tu-braunschweig.de

Emanuele Naldi, Dirk Lorenz
TU Braunschweig
e.naldi@tu-braunschweig.de, d.lorenz@tu-braunschweig.de

MS259

Accelerated Dynamics with Dry Friction Via Time Scaling and Averaging of Doubly Nonlinear Evolution Equations

In a Hilbert framework, with the aim of convex differentiable optimization, we analyze the long-time behavior of inertial dynamics with dry friction. We rely on the general acceleration method recently developed by Attouch, Bot and Nguyen, which consists of applying the time scaling and then averaging method to a first order in time continuous differential equation. In our approach, we start from a doubly nonlinear first-order evolution equation involving two potentials: one is the differentiable function f to be minimized, which acts on the state of the system via its gradient, and the other is the dry friction potential $\varphi(x) = r\|x\|$ which acts on the velocity vector via its subdifferential. We so obtain a second-order in time evolution system involving dry friction, asymptotically vanishing viscous damping (directly related to Nesterov's accelerated gradient method), and a damping driven by the Hessian in the implicit form. The mathematical analysis does not require developing a Lyapunov analysis for inertial systems. We obtain fast convergence rates for both the system and its dual, which governs the evolution of the gradients, using Riemannian gradient structure.

Samir Adly
Université de Limoges, Laboratoire XLIM
France
samir.adly@unilim.fr

Hedy Attouch
Université Montpellier II, France
hedy.attouch@umontpellier.fr

MS259

Automatic FISTA Restart

FISTA is a widely used first-order method in the field of large convex optimization problems. By introducing inertia, this scheme ensures a fast convergence of order $\mathcal{O}\left(1/k^2\right)$ for convex functions. For functions that satisfy stronger assumptions than convexity such as strong convexity, inertia can cause an oscillatory behavior of the iterates slowing down the convergence. This behavior can be avoided if parameters related to the geometry of the function of interest are known, which is rarely the case in practice. We focus on functions exhibiting quadratic growth around their minimizers and we introduce a restart scheme for FISTA that ensures fast convergence without requiring prior knowledge of the geometry of the function (i.e. of the growth parameter).

Hippolyte Labarriere
INSA Toulouse, France
labarrie@insa-toulouse.fr

Jean-Francois Aujol
IMB, CNRS, Université Bordeaux 1
jean-Francois.Aujol@math.u-bordeaux.fr

Charles Dossal, Aude Rondepierre
INSA Toulouse
charles.dossal@insa-toulouse.fr, aude.rondepierre@insa-toulouse.fr

MS259

An Sde Perspective on Stochastic Convex Optimization

we analyze the global and local behavior of gradient-like flows under stochastic errors towards the aim of solving convex optimization problems with noisy gradient input. We first study the unconstrained differentiable convex case, using a stochastic differential equation where the drift term is minus the gradient of the objective function and the diffusion term is either bounded or square-integrable. In this context, under Lipschitz continuity of the gradient, our first main result shows almost sure convergence of the objective and the trajectory process towards a minimizer of the objective function. We also provide a comprehensive complexity analysis by establishing several new pointwise and ergodic convergence rates in expectation for the convex, strongly convex, and (local) Łojasiewicz case. The latter, which involves local analysis, is challenging and requires non-trivial arguments from measure theory. Then, we extend our study to the constrained case and more generally to certain nonsmooth situations. We show that several of our results have natural extensions obtained by replacing the gradient of the objective function by a cocoercive monotone operator. This makes it possible to obtain similar convergence results for optimization problems with an additively "smooth + non-smooth" convex structure. Finally, we consider another extension of our results to non-smooth optimization which is based on the Moreau envelope.

Rodrigo Maulen
ENSICAEN
rodrigo.maulen@ensicaen.fr

Jalal Fadili
Normandie Univ, ENSICAEN, CNRS
Jalal.Fadili@ensicaen.fr

Hedy Attouch

Université Montpellier II, France
hedy.attouch@umontpellier.fr

MS260

High-Dimensional Limit Theorems for Sgd: Effective Dynamics and Critical Scaling

We study the scaling limits of stochastic gradient descent (SGD) with constant step-size in the high-dimensional regime. We prove limit theorems for the trajectories of summary statistics (i.e., finite-dimensional functions) of SGD as the dimension goes to infinity. Our approach allows one to choose the summary statistics that are tracked, the initialization, and the step-size. It yields both ballistic (ODE) and diffusive (SDE) limits, with the limit depending dramatically on the former choices. We find a critical scaling regime for the step-size below which this "effective dynamics" matches gradient flow for the population loss, but at which, a new correction term appears which changes the phase diagram. About the fixed points of this effective dynamics, the corresponding diffusive limits can be quite complex and even degenerate. We demonstrate our approach on popular examples including estimation for spiked matrix and tensor models and classification via two-layer networks for binary and XOR-type Gaussian mixture models. These examples exhibit surprising phenomena including multimodal timescales to convergence as well as convergence to sub-optimal solutions with probability bounded away from zero from random (e.g., Gaussian) initializations. At the same time, we demonstrate the benefit of overparametrization by showing that the latter probability goes to zero as the second layer width grows.

Gerard Ben Arous
New York University
Courant Institute
benarous@cims.nyu.edu

Reza Gheissari
Northwestern University
gheissari@northwestern.edu

Aukosh Jagannath
University of Waterloo
a.jagannath@uwaterloo.ca

MS260

Constrained Stochastic Nonconvex Optimization with State-Dependent Markov Data

We study stochastic optimization algorithms for constrained nonconvex stochastic optimization problems with Markovian data. In particular, we focus on the case when the transition kernel of the Markov chain is state-dependent. Such stochastic optimization problems arise in various machine learning problems including strategic classification and reinforcement learning. For this problem, we study both projection-based and projection-free algorithms. In both cases, we establish that the number of calls to the stochastic first-order oracle to obtain an appropriately defined ϵ-stationary point is of the order $O(1/\epsilon^2.5)$. In the projection-free setting we additionally establish that the number of calls to the linear minimization oracle is of order $O(1/\epsilon^5.5)$. We also empirically demonstrate the performance of our algorithm on the problem of strategic classification with neural networks.

Abhishek Roy
University of California, San Diego
a2roy@ucsd.edu

MS260

Policy Gradient Methods Find the Nash Equilibrium in N-Player General-Sum Linear-Quadratic Games

Policy optimization algorithms have achieved substantial empirical successes in addressing a variety of non-cooperative multi-agent problems. However, there have been few results from a theoretical perspective showing why such a class of reinforcement learning algorithms performs well with the presence of competition among agents. In this talk, we explore the natural policy gradient method for a class of N-agent general-sum linear-quadratic games. We provide a global linear convergence guarantee for this approach in the setting of finite time horizon and stochastic dynamics when there is a certain level of noise in the system. The noise can either come from the underlying dynamics or carefully designed explorations from the agents. We illustrate the performance of our algorithm with two examples: a bargaining game with two players and a toy example in high dimensions.

Renyuan Xu
University of Southern California
renyuanx@usc.edu

MS261

How Averaged is the Composition of Two Linear Projections?

Projection operators are fundamental algorithmic operators in Analysis and Optimization. It is well known that these operators are firmly nonexpansive; however, their composition is generally only averaged and no longer firmly nonexpansive. In this note, we introduce the modulus of averagedness and provide an exact result for the composition of two linear projection operators. As a consequence, we deduce that the Ogura-Yamada bound for the modulus of the composition is sharp.

Heinz Bauschke
The University of British Columbia, Kelowna, BC Canada
Mathematics
heinz.bauschke@ubc.ca

Theo Bendit
University of British Columbia, Canada
theo.bendit@gmail.com

Walaa M. Moursi
University of Waterloo
walaa.moursi@uwaterloo.ca

MS261

The Range of the DouglasRachford Operator in

Infinite-dimensional Hilbert Spaces

The DouglasRachford algorithm is one of the most prominent splitting algorithms for solving convex optimization problems. Recently, the method has been successful in finding a generalized solution (provided that one exists) for optimization problems in the inconsistent case, i.e., when a solution does not exist. The convergence analysis of the inconsistent case hinges on the study of the range of the displacement operator associated with the DouglasRachford splitting operator and the corresponding minimal displacement vector. In this paper, we provide a formula for the range of the DouglasRachford splitting operator in (possibly) infinite-dimensional Hilbert spaces under mild assumptions on the underlying operators. Our new results complement known results in finite-dimensional Hilbert spaces. Several examples illustrate and tighten our conclusions.

Walaa M. Moursi
University of Waterloo
walaa.moursi@uwaterloo.ca

MS262

A Stochastic Newton Algorithm for Distributed Convex Optimization

We propose and analyze a stochastic Newton algorithm for homogeneous distributed stochastic convex optimization, where each machine can calculate stochastic gradients of the same population objective, as well as stochastic Hessian-vector products (products of an independent unbiased estimator of the Hessian of the population objective with arbitrary vectors), with many such stochastic computations performed between rounds of communication. We show that our method can reduce the number, and frequency, of required communication rounds compared to existing methods without hurting performance, by proving convergence guarantees for quasi-self-concordant objectives (e.g., logistic regression), alongside empirical evidence.

Brian Bullins
Purdue University
bbullins@purdue.edu

Kumar Kshitij Patel
Toyota Technological Institute
kkpatel@ttic.edu

Ohad Shamir
Weizmann Institute of Science
ohad.shamir@weizmann.ac.il

Nati Srebro
Toyota Technological Institute at Chicago
nati@ttic.edu

Blake Woodworth
SIERRA, INRIA
blakewoodworth@gmail.com

MS262

Fl_PyTorch: Optimization Research Simulator for Federated Learning

Federated Learning (FL) has emerged as a promising technique for edge devices to collaboratively learn a shared machine learning model while keeping training data locally on the device, thereby removing the need to store and access the full data in the cloud. However, FL is difficult to implement, test and deploy in practice considering heterogeneity in common edge device settings, making it fundamentally hard for researchers to efficiently prototype and test their optimization algorithms. In this work, our aim is to alleviate this problem by introducing FL_PyTorch : a suite of open-source software written in python that builds on top of one the most popular research Deep Learning (DL) framework PyTorch. We built FL_PyTorch as a research simulator for FL to enable fast development, prototyping and experimenting with new and existing FL optimization algorithms. Our system supports abstractions that provide researchers with a sufficient level of flexibility to experiment with existing and novel approaches to advance the state-of-the-art. Furthermore, FL_PyTorch is a simple to use console system, allows to run several clients simultaneously using local CPUs or GPU(s), and even remote compute devices without the need for any distributed implementation provided by the user. FL_PyTorch also offers a Graphical User Interface. For new methods, researchers only provide the centralized implementation of their algorithm.

Konstantin P. Burlachenko
King Abdullah University of Science and Technology
konstantin.burlachenko@kaust.edu.sa

Samuel Horváth
Mohamed bin Zayed University of Artificial Intelligence
samuel.horvath@mbzuai.ac.ae

Peter Richtárik
King Abdullah University of Science and Technology (KAUST)
richtarik@gmail.com

MS262

Shifted Compression Framework for Distributed Learning

Communication is one of the key bottlenecks in the distributed training of large-scale ML models, and lossy compression of exchanged information, such as stochastic gradients or models, is one of the most effective instruments to alleviate this issue. Among the most studied techniques is the class of unbiased compression operators with variance bounded by a multiple of the square norm of the vector we wish to compress. By design, this variance may remain high, and only diminishes if the input vector approaches zero. However, unless the model being trained is overparameterized, there is no a-priori reason for the vectors we wish to compress to approach zero during the iterations of classical methods such as distributed compressed SGD, which has adverse effects on the convergence. Due to this issue, several more elaborate and seemingly very different algorithms have been proposed recently, with the goal of circumventing this problem. These methods are based on the idea of compressing the difference between the vector we would normally wish to compress and some auxiliary changing vector. In this work we develop a unified framework for studying such methods, conceptually, and theoretically. Our framework incorporates methods compressing both gradients and models, using unbiased and biased compressors, and sheds light on the construction of the auxiliary vectors. Furthermore, our approach can lead to the improvement of several existing algorithms, and can produce new methods.

Egor Shulgin
King Abdullah University of Science and Technology
egor.shulgin@kaust.edu.sa

Peter Richtarik
King Abdullah University of Science & Technology (KAUST)
Peter.Richtarik@kaust.edu.sa

MS263

Phenomenological Modeling and Analysis of Neurodegeneration

Neurodegeneration is a chronic disease with significant mortality. The underlying causes of many neurodegenerative disorders are unknown, and there is putative value in phenomenological models having interpretable mechanisms of action and effects. In this talk I give a brief introduction to a simple class of neural architectures that naturally recapitulate some key dynamics of neurodegeneration and describe application of uncertainty quantification to elucidate the model input-output behavior. The talk is concluded with some testable biological hypotheses.

Caleb Bastian
Princeton University
cbastian@princeton.edu

MS263

A Hyperparameter-Free, Fast and Efficient Framework to Detect Clusters from Limited Samples Based on Ultra High-Dimensional Features

Clustering is a challenging problem in machine learning in which one attempts to group N objects into K 0 groups based on P features measured on each object. In this article, we examine the case where N ? P and K 0 is not known. Clustering in such high dimensional, small sample size settings has numerous applications. Whereas most existing clustering algorithms either require the number of clusters to be known a priori or are sensitive to the choice of tuning parameters, our method does not require the prior specification of K 0 or any tuning parameters. Our method is based on a simple transformation of the Gram matrix. If the correlation between features decays as the number of features grows, we show that the transformed feature vectors concentrate tightly around their respective cluster expectations in a low-dimensional space. This result simplifies the detection and visualization of the unknown cluster configuration. We illustrate the algorithm by applying it to 32 benchmarked microarray datasets. Compared to 21 other commonly used clustering methods, we find that the proposed algorithm is faster and twice as accurate in determining the best cluster configuration.

Shahina Rahman
Amazon
orashahi@amazon.com

MS263

Understanding and controlling pandemics Analysis and optimization of spreading processes

The modern world comprises interlinked networks of contacts between individuals, computing devices and social groups, where infectious diseases, information and opinions propagate through their edges in a probabilistic or deterministic manner via interactions between constituents. The spread of information, opinions and marketing material can be modeled and analyzed similarly to epidemic spreading. To contain and mitigate the spread one would like to implement effective prevention and mitigation policies and deploy vaccines in a way that minimizes the spread. This is a difficult problem and becomes even harder in the presence of infectious but asymptomatic states. In the world of marketing and opinion setting, winners are those who maximize the impact by deploying resource to the most influential available nodes at the right time, occasionally in competition (or collaboration) with adversarial (supportive) processes. These can represent opinion formation by political parties (competitive) or diseases that increase the susceptibility for mutual infections (collaborative). I will explain the modelling of spreading processes and present a probabilistic framework for impact maximization/minimization, addressing questions of vaccine (budget) deployment and spreading maximization in single and competitive/collaborative processes. I will also present the analysis for epidemic spreading processes with infectious but asymptomatic states and the effectiveness of mitigation measures.

David Saad
Aston University
d.saad@aston.ac.uk

MS264

Multiplicative Updates for Online Convex Optimization over Symmetric Cones

We study online convex optimization where the possible actions are trace-one elements in a symmetric cone, generalizing the extensively-studied experts (simplex) setup and its quantum (density matrix) counterpart. Symmetric cones provide a unifying framework for some of the most important optimization models, including linear, second-order cone, and semidefinite optimization. Using tools from the field of Euclidean Jordan Algebras, we introduce the Symmetric-Cone Multiplicative Weights Update (SCMWU), a projection-free algorithm for online optimization over an arbitrary symmetric cone. We show that SCMWU is equivalent to Follow-the-Regularized-Leader and Mirror Descent algorithms with symmetric-cone negative entropy as regularizer. Using this structural result we show that SCMWU is a no-regret algorithm, and verify our theoretical results with extensive experiments. Our results unify and generalize the analysis for the Multiplicative

Weights Update method over the simplex and the Matrix Multiplicative Weights Update method over the set of density matrices. Based on joint work with Ilayda Canyakmaz, Georgios Piliouras, and Antonios Varvitsiotis.

Wayne Lin
Singapore University of Technology and Design
wayne_lin@mymail.sutd.edu.sg

MS264

Facial Residual Functions, Error Bounds and Applications

In this talk, we overview some recent error bound results obtained under the framework of facial residual functions. This includes new error bounds for exponential cones, p-cones and others. We will briefly illustrate the applications of such results in proving convergence rates of algorithms and determining automorphism groups of cones.

Scott B. Lindstrom
Centre for Optimisation and Decision Science
Curtin University
scott.lindstrom@curtin.edu.au

Bruno Lourenço
The Institute of Statistical Mathematics
bruno@ism.ac.jp

Ting Kei Pong
Department of Applied Mathematics
The Hong Kong Polytechnic University, Hong Kong
tk.pong@polyu.edu.hk

MS264

Away-Step Frank-Wolfe for Minimizing Logarithmically-Homogeneous Self-Concordant Barriers

We propose an away-step Frank-Wolfe method for minimizing logarithmically-homogeneous self-concordant barriers over a polytope. We show this method has global linear convergence. Under certain strict-complementarity conditions, we show that this method generates iterates that will eventually land on the minimal optimal face, and hence it has (potentially) much faster local linear convergence. Numerical experiments on various applications showcase the superior practical performance of this method compared to existing ones.

Renbo Zhao
MIT
Operations Research Center
renboz@mit.edu

MS265

Shortest Paths in Graphs of Convex Sets

Given a graph, the shortest-path problem requires finding a sequence of edges of minimum cost connecting a source vertex to a target vertex. In this talk we introduce a generalization of this classical problem in which the position of each vertex in the graph is a continuous decision variable, constrained to lie in a corresponding convex set, and the cost of an edge is a convex function of the positions of the vertices it connects. Problems of this form arise naturally in motion planning of autonomous vehicles, robot navigation, and even optimal control of hybrid dynamical systems. The price for such a wide applicability is the complexity of this problem, which is easily seen to be NP-hard. We discuss this novel problem formulation along with different solution approaches, including a strong mixed-integer convex formulation based on perspective functions. This formulation has a very tight convex relaxation and makes it possible to efficiently find globally-optimal paths in large graphs and in high-dimensional spaces.

Tobia Marcucci, Jack Umenberger
MIT
tobiam@mit.edu, umnbrgr@mit.edu

Pablo A. Parrilo
Massachusetts Institute of Technology
parrilo@MIT.EDU

Russ Tedrake
MIT
russt@mit.edu

MS265

Computing tighter bounds on the n-queens constant via Newtons method

In recent work Simkin shows that bounds on an exponent occurring in the famous n-queens problem can be evaluated by solving convex optimization problems, allowing him to find bounds far tighter than previously known. In this talk we use Simkin's formulation, a sharper bound developed by Knuth, and a Newton method that scales to large problem instances, to find even sharper bounds.

Parth Nobel
Stanford University
ptnobel@stanford.edu

Stephen Boyd
Stanford University
Department of Electrical Engineering
boyd@stanford.edu

MS265

Learning to Control with Untrusted Advice

Making use of modern black-box ML tools such as deep reinforcement learning is potentially transformational for the control of safety-critical systems. However, such machine-learned algorithms typically do not have formal guarantees on their worst-case performance, stability, or safety and are typically difficult to make use of in distributed, networked settings. So, while their performance may improve upon traditional approaches in typical cases, they may perform arbitrarily worse in scenarios where the training examples are not representative due to, e.g., distribution shift, or in situations where global information is unavailable to local controllers. Thus, a challenging open question emerges:

Is it possible to provide guarantees that allow black-box ML tools to be used in safety-critical applications? In this work, we provide new algorithms for combining model-free and model-based approaches to yield ML-based control of non-linear systems with formal guarantees on performance, stability, safety, and sample complexity.

Adam Wierman
Caltech
adamw@caltech.edu

MS266

Flag Sums of Squares for Sidorenko's Conjecture

Sidorenko's conjecture can be formulated as "Of all the graphs with edge density p, the graph with edges chosen uniformly at random (with probability p) contains the fewest bipartite subgraphs." This conjecture, first formulated in 1991 by Sidorenko, has received considerable attention over the last decades, and yet remains open in the general case. We provide a new approach based on Razborov's flag algebra method. While it was recently shown [Blekherman, Raymond, Singh, Thomas, 2020] that flag-sums-of-squares are not enough to prove even small, known cases of the conjecture, we provide a strenghtening of the hierarchy, based on derivatives of subgraph density functions, which can handle some of these cases.

Daniel Brosch
Alpen-Adria-Universität Klagenfurt
daniel.brosch@aau.at

MS266

On Integrality in SDP Formulations of Combinatorial Optimization Problems

It is well-known that if we add integrality constraints to the semidefinite programming (SDP) relaxation of the max-cut problem, the resulting integer SDP (ISDP) is an exact formulation of the problem. In this work we show similar results for a wide variety of combinatorial optimization problems for which SDP relaxations have been derived. We show that the generic vector- and matrix-lifting SDPs for binary quadratic problems are exact under the presence of integrality constraints. Moreover, combinatorial problems that allow a quadratic matrix programming formulation can be modelled as an ISDP. In particular, this leads to a compact ISDP formulation for the quadratic assignment problem. We also show that several structured problems allow for novel compact ISDP formulations based on algebraic notions. Although solving ISDPs is still practically challenging, these formulations induce new bounds that are at least as good as their continuous counterparts via Lagrangian duality. Being able to solve the Lagrangian dual problems via a projected subgradient approach, we show that this leads to bounds for the max-cut and stable set problem that are stronger than their standard continuous SDP bounds.

Frank de Meijer, Renata Sotirov
Tilburg University
f.j.j.demeijer@uvt.nl, r.sotirov@uvt.nl

MS268

Adaptive Sampling Augmented Lagrangian Methods for Stochastic Constrained Optimization

In this talk, we consider a stochastic augmented Lagrangian framework for solving constrained optimization problems where the objective function is stochastic and the constraints are deterministic. We use adaptive sampling methods to solve the inner sub-problems at each iteration and establish global convergence under expected inexactness criteria. We also develop worst-case sample complexity results and demonstrate the methods performance on portfolio optimization, optimal design problems, and constrained machine learning problems.

Cem Karamanli
The University of Texas at Austin
cem.karamanli@utexas.edu

Raghu Bollapragada
Operations Research and Industrial Engineering
The University of Texas at Austin
raghu.bollapragada@utexas.edu

Brendan Keith
Brown University
brendan_keith@brown.edu

Boyan Lazarov
Lawrence Livermore National Laboratory
lazarov2@llnl.gov

Socratis Petrides
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
petrides1@llnl.gov

Jingyi Wang
Lawrence Livermore National Laboratory
7000 East Avenue, Livermore, CA 94550
wang125@llnl.gov

MS268

SGD under Interpolation: Convergence, Line-Search, and Acceleration.

One of the striking aspects of modern machine learning methods is their ability to exactly fit, or *interpolate*, their training data. This phenomenon occurs most widely in over-parameterized neural networks, but is also satisfied by flexible kernel regression and in separable classification problems. A multitude of recent works have shown that stochastic gradient methods achieve the fast convergent rates of their deterministic counterparts in this so-called "interpolation setting". However, these works make heterogeneous assumptions whose connection to interpolation is not always clear. Moreover, most convergence rates in the literature are not tight with the best deterministic analyses. In this talk, we address these issues by developing a general notion of interpolation based on first-order oracles. We then use our framework to analyze stochastic gradient descent (SGD) with a fixed step-size, with a stochastic Armijo line-search, and with Nesterov acceleration. Our proofs show faster convergence under a wider range of parameters than existing results and reduce to the best-known rates in the deterministic setting. For Nesterov acceleration, the improvement is comparable to dividing by the square-root of the condition number and addresses criticism that previous analyses could be slower than SGD.

Aaron Mishkin
Stanford University
amishkin@cs.stanford.edu

MS268

Towards Noise-Adaptive, Problem-Adaptive (Accelerated) Stochastic Gradient Descent

We aim to make stochastic gradient descent (SGD) adaptive to (i) the noise σ^2 in the stochastic gradients and (ii) problem-dependent constants. When minimizing smooth, strongly-convex functions with condition number κ, we prove that T iterations of SGD with exponentially decreasing step-sizes and knowledge of the smoothness can achieve an $\tilde{O}\left(\exp\left(\frac{-T}{\kappa}\right)+\frac{\sigma^2}{T}\right)$ rate, without knowing σ^2. In order to be adaptive to the smoothness, we use a stochastic line-search (SLS) and show (via upper and lower-bounds) that SGD with SLS converges at the desired rate, but only to a neighbourhood of the solution. On the other hand, we prove that SGD with an offline estimate of the smoothness converges to the minimizer. However, its rate is slowed down proportional to the estimation error. Next, we prove that SGD with Nesterov acceleration and exponential step-sizes (referred to as ASGD) can achieve the near-optimal $\tilde{O}\left(\exp\left(\frac{-T}{\sqrt{\kappa}}\right)+\frac{\sigma^2}{T}\right)$ rate, without knowledge of σ^2. When used with offline estimates of the smoothness and strong-convexity, ASGD still converges to the solution, albeit at a slower rate. Finally, we empirically demonstrate the effectiveness of exponential step-sizes coupled with a novel variant of SLS.

Sharan Vaswani
University of Alberta
vaswani.sharan@gmail.com

Benjamin Dubois-Taine
Universite Paris-Saclay
benj.duboistaine@gmail.com

Reza Babanezhad
Samsung SAIL Montreal
babanezhad@gmail.com

MS269

Strict Monotone Diameters of Lattice Polytopes

I will discuss recent work in which I show that for any d-dimensional lattice polytope P in $[0, k]^n$ and generic linear objective function there is a monotone path on P of length at most 2dk from the unique minimum to the unique maximum. In the language of Ziegler's strict monotone Hirsch conjecture, this result implies the strict monotone diameter satisfies the same best known asymptotic upper bound as the combinatorial diameter for lattice polytopes.

Alexander E. Black
UC Davis
aeblack@ucdavis.edu

MS269

Minimum Cost Adaptive Submodular Cover

We consider the problem of minimum cost cover of adaptive-submodular functions, and provide a $4(\ln Q + 1)$-approximation algorithm, where Q is the goal value. This bound is nearly the best possible as the problem does not admit any approximation ratio better than $\ln Q$ (unless $P = NP$). Our result is the first $O(\ln Q)$-approximation algorithm for this problem. Previously, $O(\ln Q)$-approximation algorithms were only known assuming either independent items or unit-cost items. Furthermore, our result easily extends to the setting where one wants to simultaneously cover *multiple* adaptive-submodular functions: we obtain the first approximation algorithm for this generalization.

Viswanath Nagarajan, Yubing Cui
University of Michigan
viswa@umich.edu, yubingc@umich.edu

MS269

The Subspace Flatness Conjecture and Faster Integer Programming

In a seminal paper, Kannan and Lovász (1988) considered a quantity $\mu_{KL}(\Lambda, K)$ which denotes the best volume-based lower bound on the covering radius $\mu(\Lambda, K)$ of a convex body K with respect to a lattice Λ. Kannan and Lovász proved that $\mu(\Lambda, K) \leq n \cdot \mu_{KL}(\Lambda, K)$ and the Subspace Flatness Conjecture by Dadush (2012) claims a $O(\log n)$ factor suffices, which would match the lower bound from the work of Kannan and Lovász. We settle this conjecture up to a constant in the exponent by proving that $\mu(\Lambda, K) \leq O(\log^7(n)) \cdot \mu_{KL}(\Lambda, K)$. Our proof is based on the Reverse Minkowski Theorem due to Regev and Stephens-Davidowitz (2017). Following the work of Dadush (2012, 2019), we obtain a $(\log n)^{O(n)}$-time randomized algorithm to solve integer programs in n variables. Another implication of our main result is a near-optimal flatness constant of $O(n \log^8(n))$. This is joint work with Victor Reis.

Thomas Rothvoss
University of Washington
rothvoss@uw.edu

MS270

Entropic Regularization for Frank-Wolfe-Type Methods in Probability Space

We revisit a recent class of Frank-Wolfe-type (FW) algorithms for the optimization of functionals in the space of probabilities in which Wasserstein balls are defined iteratively as trust regions in which we minimize the first-variation approximation to the underlying objective function. The first variation (also known as the influence function) may be well-defined, even if the Wasserstein gradient is ill-posed – this occurs, for example, if the influence function is non-smooth. This situation arises, for instance, when studying variational representations for rare event estimation. We study the complexity of FW methods when applying entropic regularization techniques in order to address these ill-posed situations.

Kyriakos Lotidis, Jiajin Li, Jose H. Blanchet
Stanford University
klotidis@stanford.edu, jiajinli@stanford.edu, jose.blanchet@stanford.edu

MS270

Efficient Computation of Approximate Global Minimizers of Probabilistically Constrained Optimization Problems: A Minkowski Functional Approach

We consider optimization problems with probabilistic constraints, given by $\mathbb{P}\{\,\zeta \mid \zeta \in K(x)\,\}$. Suppose $K(x) \triangleq \{\,\zeta \in \mathcal{K} \mid c(x,\zeta) \geq 0\}$. Efficient algorithms for computing approximate global minimizers for such problems are largely unavailable, motivating the current research. We consider settings where ζ is uniformly distributed on a convex compact set $\mathcal{K}$ and $c(x,\zeta)$ is defined as either $c(x,\zeta) \triangleq 1-\left|\zeta^T x\right|^m$ and $m \geq 0$ (Setting A) or $c(x,\zeta) \triangleq Tx - \zeta$ (Setting B). By leveraging recent findings in the context of non-Gaussian integrals of positively homogenous functions, we show that $\mathbb{P}\{\,\zeta \mid \zeta \in K(x)\,\}$ can be expressed as the expectation of a suitably defined continuous function $F(\bullet,\xi)$ with respect to an appropriately defined Gaussian density (or its variant), i.e. $\mathbb{E}_{\tilde{p}}\left[\,F(x,\xi)\right]$. Aided by an observation in convex analysis, we develop convex compositional representations of such a problem. We present a stochastic approximation framework for resolving such a problem and provide rate and complexity guarantees. Preliminary numerics demonstrate the efficiency of the proposed schemes.

Uday V. Shanbhag
Pennsylvania State University
udaybag@psu.edu

Constantino Lagoa
Penn State University
Pennsylvania State University
lagoa@psu.edu

Peixuan Zhang
Pennsylvania State University
pqz5090@psu.edu

MS270

Stochastic Optimization Over the Space of Measures: Application to Emergency Response

Motivated by problems in resource logistics, we consider the question of optimizing a convex functional J over the space of signed measures on R^2 and subject to a "budget constraint." Assuming J is equipped with a von Mises derivative, we present a key diagnostic that establishes the necessary and sufficient conditions for optimality, followed by a recursion we call measure gradient descent (MGD) in analogy with the classical gradient descent scheme in Euclidean space. We show that the functional value sequence at the iterates generated by MGD converges to the minimum of J. For settings where the influence function underlying the von Mises derivative can be estimated in an unbiased fashion, MGD leads to stochastic MGD, in analogy with stochastic gradient descent.

Di Yu
Purdue University, Department of Statistics
yu1128@purdue.edu

Shane Henderson
Cornell University
School of Operations Research and Industrial Engineering
sgh9@cornell.edu

Raghu Pasupathy
Purdue University
pasupath@purdue.edu

MS271

A Fast Solver for Endpoint Geodesic Problem on the Stiefel Manifold under the Canonical Metric

In this paper, we consider the endpoint geodesic problem under canonical metric on Stiefel manifold $\mathrm{St}_{n,k}$, the set of k ordered orthonormal vectors in $\mathbb{R}^n$, that seeks a Riemannian geodesic joining the given points. This is a weaker version of the Riemannian logarithm problem seeking the shortest geodesic. An optimization formulation based on a Riemannian submersion structure is proposed and a method is derived to construct a descent direction that leads to local quadratic convergence. Excessive numerical experiments are done to this problem and the proposed solver is shown to have superior efficiency and robustness comparing to the state-of-the-art solver proposed in [Ralf Zimmermann and Knut Huper, Computing the Riemannian Logarithm on the Stiefel Manifold: Metrics, Methods, and Performance, Algorithm 4]. The numerical results also shed some light on the corresponding Riemannian logarithm problem for any given points on Stiefel manifold, which is still an open question.

Zhifeng Deng, Kyle Gallivan
Florida State University
zd16d@fsu.edu, kgallivan@fsu.edu

Pierre-Antoine Absil
Université Catholique de Louvain
ICTEAM Institute
pa.absil@uclouvain.be

Wen Huang
Xiamen University
wen.huang@xmu.edu.cn

MS271

A Riemannian ADMM

We consider a class of Riemannian optimization problems where the objective is the sum of a smooth function and a nonsmooth function, considered in the ambient space. This class of problems finds important applications in machine learning and statistics such as the sparse principal component analysis, sparse spectral clustering, and orthogonal dictionary learning. We propose a Riemannian alternating direction method of multipliers (ADMM) to solve this class of problems. Our algorithm adopts easily computable steps in each iteration. The iteration complexity of the proposed algorithm for obtaining an ϵ-stationary point is analyzed under mild assumptions. To the best of our knowledge, this is the first Riemannian ADMM with provable convergence guarantee for solving Riemannian optimization problem with nonsmooth objective. Numerical experiments are conducted to demonstrate the advantage of the proposed method.

Shiqian Ma
Rice University
sqma@rice.edu

MS271

Rotation Group Synchronization via Quotient Manifold

Rotation group synchronization is a significant inverse problem and has attracted intense attention from numerous application fields such as graph realization, computer vision, and robotics. In this talk, we focus on the least squares estimator of rotation group synchronization with general additive noise, which is a nonconvex optimization problem with manifold constraints. Departing from the standard approach of utilizing the geometry of the ambient Euclidean space to study phase/orthogonal group synchronization, we adopt an intrinsic Riemannian approach to study rotation group synchronization. Benefiting from a quotient geometric view, we prove the negative definite condition of quotient Riemannian Hessian around the optimal solution of orthogonal group synchronization problem. Consequently, the Riemannian local error bound property holds and can be applied to analyze the convergence properties of various Riemannian algorithms. On the other hand, improved estimation results of the spectral and least squares estimator are derived, which guarantee the tightness of orthogonal group synchronization for solving rotation group version under certain noise level. The sequential convergence guarantee of the Riemannian (quotient) gradient method for solving orthogonal/rotation group synchronization problem is studied and we derive its linear convergence rate to the optimal solution with the spectral initialization. All results are deterministic without any probabilistic model.

Linglingzhi Zhu
The Chinese University of Hong Kong
llzzhu@se.cuhk.edu.hk

Chong Li
Department of Mathematics
Zhejiang University
cli@zju.edu.cn

Anthony Man-Cho So
The Chinese University of Hong Kong, Hong Kong
manchoso@se.cuhk.edu.hk

MS272

A Two-Species Model for Pathological Tau Spreading in Alzheimer's Disease

Image-driven biophysical biomarkers have the potential to help with disease staging and patient stratification. Here we focus on biomarkers (parameters and predictions of a biophysical model) provided by Tau-PET images. In this talk, we will introduce the qualitative inconsistencies between tau observations and the popular Fisher-Kolmogorov (FK) model predictions. To address the problem, we will present a two-species model, which we term Heterodimer Fisher-Kolmogorov model (HFK) to represent the normal tau (healthy and non- observable) and abnormal tau (observable) progression in AD patients. We conduct clinical experiments on publicly available datasets focusing on AD patients, and assess the sensitivity of our model regarding different parcellations.

Zheyu Wen
University of Texas at Austin
zheyw@utexas.edu

George Biros
The University of Texas et Austin
Oden Institute
biros@oden.utexas.edu

MS272

A Scalable Variational Approach for Solving Data-Consistent Stochastic Inverse Problems

In this presentation, we discuss a scalable approach to approximate the solution of linear and nonlinear stochastic inverse problems using a recently developed approach based on push-forward probability measures and Bayes rule. Previous approaches utilized Monte Carlo sampling over the input stochastic space and kernel density estimation to approximate push-forward densities which can result in slow convergence and/or poor scaling with the dimensionality of the data. In this paper, we develop an approach that avoids the need for random sampling or density estimation. We adapt a randomized dimension reduction technique developed for the classical Bayesian maximum a posteriori (MAP) point method to develop a method for the push-forward based approach which scales with the data, state, and parameter dimensions. We also extend the push-forward based approach to the infinite dimensional setting where the model input parameters may be associated with a mesh-based approximation of a random field. We show that from a deterministic point of view, the maximum of the updated density for the push-forward based method predicts the mean of the data, and that it removes the regularization in the directions that are informed by the map/data. We demonstrate the proposed approach using a linear inverse problem governed by an advection-diffusion equation and a nonlinear inverse problem governed by a hyperelasticity equation.

Tim Wildey
Optimization and Uncertainty Quantification Dept.

Sandia National Laboratories
tmwilde@sandia.gov

Tan Bui-Thanh
University of Texas
tanbui@oden.utexas.edu

Troy Butler
University of Colorado Denver
butler.troy.d@gmail.com

MS273

A Probabilistic Game-Theoretic Vehicle Traffic Routing Approach

We consider a routing problem for self-interested vehicles via probabilistic decision strategies. Thanks to a novel approximation of the road latency functions and a nonlinear variable transformation, we model the problem as an aggregative game. Then we provide a characterization of the approximation error and a novel monotonicity condition for the special class of games that contains that of the considered routing problem. Next, we propose a semi-decentralized algorithm for computing a routing decision as a generalized Nash equilibrium and show the advantages of the proposed solution via numerical simulations.

Emilio Benenati
TU Delft
e.benenati@tudelft.nl

Sergio Grammatico
ETH
grammatico.sergio@gmail.com

MS273

Accelerated Minimax Algorithms Have Merging Paths

Recently, several new accelerated methods in minimax optimization and fixed point iterations have been discovered. Interestingly, their acceleration relies on the mechanism of anchoring, which retracts the iterates toward the initial point, and is distinct from Nesterov's momentum-based acceleration. We introduce these accelerated algorithms and show that they exhibit what we call the merging path (MP) property; the trajectories of these algorithms merge quickly. Using this novel MP property, we establish point convergence of existing accelerated minimax algorithms and present new efficient algorithms for the strongly-convex-strongly-concave setup and for the prox-grad setup.

Taeho Yoon, TaeHo Yon
Seoul National University
tetrzim@gmail.com, tetrzim@snu.ac.kr

MS274

Programming Soft Materials by Inverse Design

Soft Materials are ubiquitous in everyday life and are crucial in many different forms of revolutionary technologies. One property of Soft Materials is their ability to self-assemble into intricate structures from a finite set of building blocks with continuously tunable parameters. This giant design space of building blocks is a double-sided sword: on one side it provides researchers infinite possibilities to design building blocks for targeted functions, while on the other side it might take for ever to search the design space. Here, we propose a new inverse design method that captures both interaction and geometry of building blocks. By enabling rigid body functionalities in JAX-MD, an end-to-end differentiable molecular dynamics engine, we can create soft materials model with components that are simple enough to design yet powerful enough to capture complex materials properties. In this talk, I will discuss the implementation of the methods alongside examples to showcase its potential applications.

Chrisy Xiyu Du
University of Hawai‘i at Manoa
xiyudu@hawaii.edu

MS274

Statistical Mechanics of Wicked Design Problems

Design of complex engineering systems frequently pursues multiple objectives governed by different stakeholders, requires solutions for incomplete problem specification, and lacks natural stopping conditions. Together these constraints pose design as a Wicked Problem that cannot be solved via traditional optimization-based techniques. In this work we generalize the optimization approach with the Systems Physics framework grounded in maximal entropy statistical mechanics. Statistical mechanics combines the information about the cost function landscape with accounting for the entropy of solutions. For example Naval Engineering arrangement problems, Systems Physics uncovers a host of new phenomena invisible to optimization, including abrupt phase transitions in design space, variable robustness of competing architecture classes, and tight interplay of spatial and topological structure of design.

Andrei A. Klishin
University of Washington
University of Michigan
aklishin@uw.edu

MS274

Data-Driven Sparse Sensor Placement for Reconstruction

TBA

Krithika Manohar
University of Washington
kmanohar@uw.edu

MS275

A Combinatorial Algorithm for Computing the Entire Sequence of the Maximum Degree of Minors of a Generic Partitioned Polynomial Matrix with 2×2 Submatrices

In this talk, we consider the problem of computing the entire sequence of the maximum degree of minors of a block-structured symbolic matrix (a generic partitioned polynomial matrix) $A = (A_{\alpha\beta} x_{\alpha\beta} t^{d_{\alpha\beta}})$, where $A_{\alpha\beta}$ is a 2×2 matrix over a field $\mathbf{F}$, $x_{\alpha\beta}$ is an indeterminate, and $d_{\alpha\beta}$ is an integer for $\alpha = 1, 2, \ldots, \mu$ and $\beta = 1, 2, \ldots, \nu$, and t is an additional indeterminate. This problem can be viewed as an algebraic generalization of the maximum weight bipartite matching problem. The main result of this study is a combinatorial $O(\mu\nu \min\{\mu, \nu\}^2)$-time algorithm for computing the entire sequence of the maximum degree of minors of a (2×2)-type generic partitioned polynomial matrix of size $2\mu \times 2\nu$. We also present a minimax theorem, which can be used as a good characterization (NP $\cap$ co-NP characterization) for the computation of the maximum degree of minors of order k. Our results generalize the classical primal-dual algorithm (the Hungarian method) and minimax formula (Egerváry's theorem) for the maximum weight bipartite matching problem.

Yuni Iwamasa
Kyoto University
iwamasa@i.kyoto-u.ac.jp

MS275

Algebraic Algorithms for Fractional Linear Matroid Parity Via Non-Commutative Rank

Matrix representations are a powerful tool for designing efficient algorithms for combinatorial optimization problems such as matching, and linear matroid intersection and parity. In my talk, we initiate the study of matrix representations using the concept of non-commutative rank (nc-rank), which has recently attracted attention in the research of Edmonds' problem. We reveal that the nc-rank of the matrix representation of linear matroid parity corresponds to the optimal value of fractional linear matroid parity: a half-integral relaxation of linear matroid parity. Based on our representation, we present an algebraic algorithm for the fractional linear matroid parity problem by building a new technique to incorporate the search-to-decision reduction into the half-integral problem represented via the nc-rank. We further present a faster divide-and-conquer algorithm for finding a maximum fractional matroid matching and an algebraic algorithm for finding a dual optimal solution. They together lead to an algebraic algorithm for the weighted fractional linear matroid parity problem. Our algorithms are significantly simpler and faster than the existing algorithms.

Taihei Oki
The University of Tokyo
oki@mist.i.u-tokyo.ac.jp

MS275

Shrunk Subspaces Via Operator Sinkhorn Iteration

A recent breakthrough in Edmonds' problem showed that the noncommutative rank can be computed in deterministic polynomial time, and various algorithms for it were devised. However, only quite complicated algorithms are known for finding a so-called shrunk subspace, which acts as a dual certificate for the value of the noncommutative rank. In particular, the operator Sinkhorn algorithm, perhaps the simplest algorithm to compute the noncommutative rank with operator scaling, does not find a shrunk subspace. Finding a shrunk subspace plays a key role in applications, such as separation in the Brascamp-Lieb polytope, one-parameter subgroups in the null-cone membership problem, and primal-dual algorithms for matroid intersection and fractional matroid matching. In this talk, we provide a simple Sinkhorn-style algorithm to find the smallest shrunk subspace over the complex field in deterministic polynomial time. To this end, we introduce a generalization of the operator scaling problem, where the spectra of the marginals must be majorized by specified vectors. Then we design an efficient Sinkhorn-style algorithm for the generalized operator scaling problem. Applying this to the shrunk subspace problem, we show that a sufficiently long run of the algorithm also finds an approximate shrunk subspace close to the minimum exact shrunk subspace. Finally, we show that the approximate shrunk subspace can be rounded if it is sufficiently close.

Cole Franks, Tasuku Soma
MIT
franks@mit.edu, tasuku@mit.edu

Michel Goemans
Massachusetts Institute of Technology
goemans@math.mit.edu

MS276

Decision-Aware Data Aggregation

When one has to solve many stochastic optimization problems where one only has limited data per problem, the most natural approach is data aggregation. For example, in setting prices for many distinct products each with limited sales history, a common heuristic is first to cluster the products (e.g. with k-means clustering) and then treat all products within a cluster identically (set a single price). While such data aggregation heuristics abound, they typically follow an “estimate-then-optimize” paradigm – i.e., the aggregation step is agnostic to the form of the downstream optimization problem. We propose novel decision-aware algorithms for data aggregation that adapt to the downstream optimization problem when aggregating. We show these variants are tractable and improve upon estimate-then-optimize approaches. We complement these empirical results with theoretical analysis establishing best-in-class performance for these approaches as the number of problems grows large, even when the amount of data remains small and constant.

Vishal Gupta
University of Southern California
guptavis@usc.edu

MS276

Integrated Conditional Estimation-Optimization

Many real-world optimization problems involve uncertain parameters with probability distributions that can be estimated using contextual feature information. In contrast to the standard approach of first estimating the distribution of uncertain parameters and then optimizing the objective based on the estimation, we propose an integrated conditional estimation-optimization (ICEO) framework that estimates the underlying conditional distribution of the random parameter while considering the structure of the optimization problem. We directly model the relationship between the conditional distribution of the random parameter and the contextual features, and then estimate the probabilistic model with an objective that aligns with the downstream optimization problem. We show that our ICEO approach is asymptotically consistent under moderate regularity conditions and further provide finite performance guarantees in the form of generalization bounds. Computationally, performing estimation with the ICEO approach is a non-convex and often non-differentiable optimization problem. We propose a general methodology for approximating the potentially non-differentiable mapping by a differentiable one, which greatly improves the performance of gradient-based algorithms applied to the non-convex problem. Numerical experiments are also conducted to show the empirical success of our approach in different situations including with limited data samples and model mismatches.

Paul Grigas
University of California- Berkeley
pgrigas@berkeley.edu

Meng Qi
Cornell University
mq56@cornell.edu

Zuo-Jun Shen
University of California- Berkeley
maxshen@berkeley.edu

MS276

Decision-making with Side Information: A Causal Transport Robust Approach

We consider stochastic optimization with side information where, prior to decision-making, covariate data are available to inform better decisions. In particular, we propose to consider a distributionally robust formulation based on causal transport distance. Compared with divergence and Wasserstein metric, the causal transport distance is better at capturing the information structure revealed from the conditional distribution of random problem parameters given the covariate values. We derive a dual reformulation for evaluating the worst-case expected cost and show that the worst-case distribution in a causal transport distance ball has a similar conditional information structure as the nominal distribution. When optimizing over affine decision rules, we identify cases where the overall problem can be solved by convex programming. When optimizing over all (non-parametric) decision rules, we identify a new class of robust optimal decision rules when the cost function is convex with respect to a one-dimensional decision variable.

Luhao Zhang
University of Texas, Austin
luhaozhang@utexas.edu

MS277

Under-Counted Tensor Completion with Neural Side Information Learner: Recoverability and Algorithm

Systematical *unknown* under-counting effects are observed in data collected across many disciplines, e.g., ecology and epidemiology. *Under-counted tensor completion* (UC-TC) is therefore well-motivated. However, theory and methods for UC-TC have been largely elusive. The closest prior work is an under-counted matrix completion (UC-MC) approach by Fu *et al.* There, the ground-truth (fully counted) data entries were modeled as Poisson parameters with low matrix rank characterizations; and a linear function-based side information learner was used to assist modeling the under-counting effect. Nonetheless, the over-simplified linear side information learner struggles to accurately reflect the complex relation between the side information and the under-counting effect. In addition, there is a lack of understanding of the effectiveness of the UC-MC method. This work puts forth a nontrivial generalization of the UC-MC framework for UC-TC tasks. A low-rank Poisson tensor model with an expressive unknown *nonlinear* side information extractor is proposed for modeling the generative process of under-counted multi-aspect data. A joint low-rank tensor completion and neural network learning algorithm is devised for efficiently handling our UC-TC formulation. Moreover, recoverability analysis is provided to support our design. The recoverability result stands as the first of its kind, to our best knowledge.

Xiao Fu, Shahana Ibrahim
Oregon State University
xiao.fu@oregonstate.edu, ibrahish@oregonstate.edu

MS277

Weighted Factor Match Score - a Tensor Kernel

When the data samples to be classified are given as tensors, it is beneficial to exploit the multidimensional structure in order to preserve structure and consequently, improved accuracy. A kernelized support vector machine that acts on tensor data (also called support tensor machine, STM) makes use of kernels that are designed specifically for tensors or even for low-rank tensor decompositions. Recently, the DuSK kernel has exhibited superior performance on data given in the CP format. In previous work, we generalized the DuSK kernel to the TT format, thus allowing for more efficient computation of the kernel and often also better accuracy with more stable model. Here, we compare and investigate the strengths of different kernels both on simulated and real world data, resulting in a new kernel that is based on the factor match score and further improves accuracy and efficiency.

Kirandeep Kour
Max Planck Institute for Dynamics of Complex Technical Systems, Magdbeurg
kour@mpi-magdeburg.mpg.de

Sergey Dolgov

University of Bath
Department of Mathematical Sciences
s.dolgov@bath.ac.uk

Peter Benner
Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
benner@mpi-magdeburg.mpg.de

Martin Stoll
Chemnitz University of Technology
martin.stoll@mathematik.tu-chemnitz.de

Max Pfeffer
TU Chemnitz
max.pfeffer@math.tu-chemnitz.de

MS277

A Non-Negative Probabilistic Tensor Decomposition and Applications in Cancer Genomics

The tumor microenvironment (TME) is a complex milieu around the tumor, whereby cancer cells interact with stromal, immune, vascular, and extracellular components. The TME is being increasingly recognized as a key determinant of tumor growth, disease progression, and response to therapies. We build a generalizable and robust tensor-based framework capable of integrating dissociated single-cell and spatially resolved RNA-seq data for a comprehensive analysis of the TME. Tensors are a generalization of matrices to higher dimensions. Tensor methods are known to be able to successfully incorporate data from multiple sources and perform a joint analysis of heterogeneous high-dimensional data sets. The methodologies developed as part of this effort will advance our understanding of the TME in multiple directions. These include cellular heterogeneity within the TME, crosstalks between cells, and tumor-intrinsic pathways stimulating tumor growth and immune evasion.

Neriman Tokcan
Broad Institute of MIT and Harvard
neriman.tokcan@umb.edu

MS278

Elastic constant determination using DFO analysis of resonant ultrasound spectroscopic data

Resonant Ultrasound Spectroscopy (RUS) has been a popular and well-studied technique for measuring elastic constants of crystalline materials ever since the pioneering work of Migliori some 30 years ago. RUS measurements are performed using minimal-contact source and receiver transducers, which excite and record the resonant response of the sample as a function of frequency. For a well-characterized sample geometry and a sufficient number of measured lowest-order resonances, a very accurate determination of elastic constants can be achieved. The resulting (ideal) optimization problem is convex and involves anywhere from two to twenty-one decision variables depending on the crystal symmetry of the sample. A significant difficulty sometimes arises from the fact that the nature of the experiment and sample geometry can result in some missing or hidden data. We introduce a modified objective that is robust to missing data but also renders the optimization problem nonsmooth and non-convex. We employ an instance of Mesh Adaptive Direct Search (EAD-MADS) which includes primary directions drawn from a sequence of uniform hypersphere partitions and positive basis construction using a QR construction or Householder transformation. Simple box constraints are enforced using an extreme barrier approach. We present results for samples of relatively simple crystal symmetry and geometry using data obtained from other researchers.

Tom Asaki
Washington State University
tasaki@wsu.edu

MS278

Scalable Bayesian optimization for noisy problems

Stochastic simulators may exhibit low signal-to-noise ratios, hence requiring large budget of evaluations to find accurate solutions. For deterministic and high signal to noise ratio problems, Bayesian optimization is a sample efficient technique. In particular, recent works have shown the interest of combining it with trust region methods to help convergence. Here we revisit this combination for noisier problems, when the number of evaluations may reach millions or even billions. We build upon local heteroscedastic Gaussian process modeling to adapt to both noise and nonstationarity, while adapting the degree of replication at each design to scale with respect to the number of observations. In particular, our method is able to identify solutions with an arbitrary precision. We illustrate our method on several synthetic test cases, as well as some more realistic ones.

Mickael Binois
Inria
mickael.binois@inria.fr

MS278

Exploiting problem structures in multiobjective optimization

In this talk, we introduce our response surface modeling and optimization framework ParMOO for building custom multiobjective optimization solvers. We will describe how ParMOO enables users to exploit a broad range of structures that commonly arise in multiobjective simulation optimization problems, in order to accelerate convergence. We will then explore how exploiting these problem structure affects performance with surprising results, and discuss possible explanations.

Tyler H. Chang
Argonne National Laboratory
tchang@anl.gov

Stefan Wild
Argonne National Laboratory
LANS, Mathematics and Computer Science

wild@lbl.gov

MS279

The Geometry of Random Kaczmarz: Problems and Opportunities

I will discuss a couple of recent results that provide a rather clear picture of the geometry underlying the Random Kaczmarz method. This allows us to discuss a couple of exciting open problems that merge geometry and probability in a nice way.

Stefan Steinerberger
University of Washington
steinerb@uw.edu

MS279

Faster Randomized Block Sparse Kaczmarz by Averaging

The standard randomized sparse Kaczmarz (RSK) method is an algorithm to compute sparse solutions of linear systems of equations and uses sequential updates, and thus, does not take advantage of parallel computations. In this work, we introduce a parallel (mini batch) version of RSK based on averaging several Kaczmarz steps. Naturally, this method allows for parallelization and we show that it can also leverage large over-relaxation. We prove linear expected convergence and show that, given that parallel computations can be exploited, the method provably provides faster convergence than the standard method. This method can also be viewed as a variant of the linearized Bregman algorithm, a randomized dual block coordinate descent update, a stochastic mirror descent update, or a relaxed version of RSK and we recover the standard RSK method when the batch size is equal to one. We also provide estimates for inconsistent systems and show that the iterates convergence to an error in the order of the noise level. Finally, numerical examples illustrate the benefits of the new algorithm

Lionel Tondji
TU Braunschweig
l.ngoupeyou-tondji@tu-braunschweig.de

MS279

On the Fast Convergence of Minibatch Heavy Ball Momentum

Simple stochastic momentum methods are widely used in machine learning optimization, but their good practical performance is at odds with an absence of theoretical guarantees of acceleration in the literature. In this work, we aim to close the gap between theory and practice by showing that stochastic heavy ball momentum, which can be interpreted as a randomized Kaczmarz algorithm with momentum, retains the fast linear rate of (deterministic) heavy ball momentum on quadratic optimization problems, at least when minibatching with a sufficiently large batch size is used. The analysis relies on carefully decomposing the momentum transition matrix, and using new spectral norm concentration bounds for products of independent random matrices. We provide numerical experiments to demonstrate that our bounds are reasonably sharp.

Rachel Ward
Department of Mathematics
University of Texas at Austin
rward@math.utexas.edu

MS280

Polynomial Convergence of Diffusion Models Under Minimal Data Assumptions

We provide theoretical convergence guarantees for score-based generative models (SGMs) such as denoising diffusion probabilistic models (DDPMs), which constitute the backbone of large-scale real-world generative models such as DALLE 2. Our main result is that, assuming accurate score estimates, such SGMs can efficiently sample from essentially any realistic data distribution. In contrast to prior works, our results (1) hold for an L2-accurate score estimate (rather than L8-accurate); (2) do not require restrictive functional inequality conditions that preclude substantial non-log-concavity; (3) scale polynomially in all relevant problem parameters; and (4) match state-of-the-art complexity guarantees for discretization of the Langevin diffusion, provided that the score error is sufficiently small. We view this as strong theoretical justification for the empirical success of SGMs. We also examine SGMs based on the critically damped Langevin diffusion (CLD). Contrary to conventional wisdom, we provide evidence that the use of the CLD does not reduce the complexity of SGMs.

Sitan Chen
MIT
sitan@seas.harvard.edu

Sinho Chewi
Massachusetts Institute of Technology
schewi@mit.edu

Jerry Li
Microsoft Research
jerrl@microsoft.com

Yuanzhi Li
CMU
yuanzhil@andrew.cmu.edu

Adil Salim
Microsoft Research
adilsalim@microsoft.com

Anru Zhang
Duke University
anru.zhang@duke.edu

MS280

Convergence of Stein Variational Gradient Descent under a Weaker Smoothness Condition

Stein Variational Gradient Descent (SVGD) is an important alternative to the Langevin-type algorithms for sampling from probability distributions of the form p(x) ?

exp(-V (x)). In the existing theory of Langevin-type algorithms and SVGD, the potential function V is often assumed to be L-smooth. However, this restrictive condition excludes a large class of potential functions such as polynomials of degree greater than 2. Our paper studies the convergence of the SVGD algorithm for distributions with (L0, L1)-smooth potentials. This relaxed smoothness assumption was introduced by Zhang et al. [2019a] for the analysis of gradient clipping algorithms. With the help of trajectory-independent auxiliary conditions, we provide a descent lemma establishing that the algorithm decreases the KL divergence at each iteration and prove a complexity bound for SVGD in the population limit in terms of the Stein Fisher information.

Avetik Karagulyan, Lukang Sun, Peter Richtarik
King Abdullah University of Science & Technology (KAUST)
auavetik@gmail.com, lukang.sun@kaust.edu.sa, Peter.Richtarik@kaust.edu.sa

MS280

Convergence of Score-Based Generative Modeling

Score-based generative modeling (SGM) is a highly successful approach for learning a probability distribution from data and generating further samples, based on learning the score function (gradient of log-pdf) and then using it to simulate a stochastic differential equation that transforms white noise into the data distribution. We develop a convergence theory for SGM's which applies to any distribution with bounded 2nd moment, with polynomial dependence on all parameters and no reliance on smoothness or functional inequalities. We give an error of ϵ in KL-divergence to the variance-δ perturbation of the data distribution with $\tilde{O}(d^2 \log^2(1/\delta)/\epsilon^2)$ steps, and TV guarantees under smoothness of the data distribution. Our analysis builds on the Girsanov framework by incorporating the high-probability smoothing effect of the forward diffusion and quantifies the effect of step size schedule. Based on the papers:

1. Holden Lee, Jianfeng Lu, Yixin Tan. Convergence for score-based generative modeling with polynomial complexity. https://arxiv.org/abs/2206.06227
2. Holden Lee, Jianfeng Lu, Yixin Tan. Convergence of score-based generative modeling for general data distributions. https://arxiv.org/abs/2209.12381
3. Hongrui Chen, Holden Lee, Jianfeng Lu. Improved Analysis of Score-based Generative Modeling: User-Friendly Bounds under Minimal Smoothness Assumptions. https://arxiv.org/abs/2211.01916

Holden Lee
Johns Hopkins University
hlee283@jhu.edu

MS281

Efficient Discretizations for Trajectory Optimization

Spectral collocation is a popular way of discretizing trajectory optimization problems. The approach – which is also called "pseudospectral" – requires that the equations of motion hold exactly at finitely many times. These collocation times are chosen with the infinite-dimensional problem in mind: an order n polynomial that interpolates an analytic function f at n such times will converge geometrically to f as $n \to \infty$. In this talk, we will go further. We will share practical methods that efficiently solve trajectory optimization problems by leveraging more of their infinite-dimensional structure. This perspective motivates a right preconditioner for spectral collocation. We will also discuss the connections between this structure and the optimization itself, including adaptive and inexact algorithms that we find effective.

Aurya Javeed
Sandia National Laboratories
asjavee@sandia.gov

MS281

Optimal Control of a Quasi-Variational Sweeping Process

The talk addresses the study of a class of evolutionary quasi-variational inequalities of the parabolic type arising in the formation and growth models of granular and cohensionless materials. Such models and their mathematical descriptions are highly challenging and require powerful tools of their analysis and implementation. We formulate a spacetime continuous optimal control problem for a basic model of this type, develop several regularization and approximation procedures, and establish the existence of optimal solutions for the time-continuous and space-discrete problem. Viewing a version of this problem as a controlled quasi-variational sweeping process leads us to deriving necessary optimality conditions for the fully discrete problem by using the advanced machinery of variational analysis and generalized differentiation.

Nguyen-Truc-Dao Nguyen
University of Michigan
nntdao@umich.edu

Harbir Antil
George Mason University
hantil@gmu.edu

Rafael Arndt
George Mason University, U.S.
tarndt@gmu.edu

Boris S. Mordukhovich
Wayne State University, U.S.
aa1086@wayne.edu

Carlos Rautenberg
George Mason University, U.S.
crautenb@gmu.edu

MS281

Birkhoff Theory for Fast Dynamic Optimization

In 2007, SIAM News ran a page 1 article that announced, "Pseudospectral Optimal Control Theory Makes Debut

Flight, Saves NASA $1M in Under Three Hours. Since this historic flight, dynamic optimization methods based on pseudospectral theory have exploded in popularity and have formed the backbone of a vast number of software products developed around the world. In a sharp departure from this highly successful Lagrange collocation method, we propose a new starting point based on "universal" Birkhoff interpolants. The Birkhoff approach offers a substantial computational upgrade to the Lagrange theory in at least three ways: (1) the Birkhoff-specific primal-dual computations are isolated to a linear system even for nonlinear, nonconvex problems, (2) dualization and discretization operations are commutable under mild hypotheses, and (3) the condition number of a key linear system is completely flattened when the grid is selected from a family of Gegenbauer node points. Combined with recent advances in spectral methods, the Birkhoff approach holds the potential to generate real-time solutions to dynamic optimization problems governed by ordinary differential equations. The problem formulation scales gracefully with the addition of algebraic path constraints and can handle various boundary conditions of practical interest. Illustrative numerical examples from aerospace engineering will be presented to demonstrate various features of the new theory.

Isaac M. Ross
Naval Postgraduate School
imross@nps.edu

MS282

Operator Splitting Based Newton-type Method for Constrained Optimization

We present a super linearly convergent Newton-type method for solving constrained optimization problems. The algorithm is constructed based on operator splitting, by iteratively taking a Newton step for the minimization problem, followed by a Newton-type step for imposing the constrains. Both smooth and semismooth, equality and inequality constrains and smooth and semismooth objectives can be handled. The analysis of the convergence rate is done by employing a general framework, centered around a generalization of semismoothness called Newton differentiability, that can yield convergence rate results for a general class of Newton-type methods.

Titus Pinta
University of Göttingen
t.pinta@math.uni-goettingen.de

MS282

A First Order Method for Nonsmooth Convex Bi-Level Optimization Problems

In this talk, I will discuss the Bi-Sub-Gradient (Bi-SG) method, which is a generalization of the classical sub-gradient method to the setting of convex bi-level optimization problems. This is a first-order method that is very easy to implement in the sense that it requires only a computation of the associated proximal mapping or a sub-gradient of the outer non-smooth objective function, in addition to a proximal gradient step on the inner optimization problem. We show, under very mild assumptions, that Bi-SG tackles bi-level optimization problems and achieves sub-linear rates both in terms of the inner and outer objective functions. Moreover, if the outer objective function is additionally strongly convex (still could be non-smooth), the outer rate can be improved to a linear rate.

Shoham Sabach
Technion - Israel Institute of Technology
Israel
ssabach@technion.ac.il

MS282

Proximal Based Methods: Recent Advances

Proximal based methods are nowadays extensively used and provide a powerful and fundamental paradigm for efficiently solving many classes of well structured optimization problems. We will discuss some of the recent developments on proximal based algorithms and their variants, highlighting extension and progress on their theoretical convergence/complexity analysis, as well as some of the involved difficulties.

Marc Teboulle
Tel Aviv University
teboulle@post.tau.ac.il

MS283

On the Equivalence of Metric Regularity and Strong Metric Regularity in Generalized Equations

Strong metric regularity of generalized equations, introduced by Robinson in the early 1980s, plays a crucial role in local convergence analysis of important numerical algorithms including the Newton method. It was shown by Dontchev and Rockafellar that this property is equivalent to metric regularity for a particular class of generalized equations, defined by the normal cone to a polyhedral convex set. In this talk, we discuss how the same equivalence can be obtained for a large class of generalized equations at their nondegenerate solutions. Our approach relies upon a second-order variational property, called strict proto-differentiability, which can be characterized via a relative interior condition for important classes of functions in variational analysis. This talk is based on joint works with Nguyen T. V. Hang.

Ebrahim Sarabi
Miami University
sarabim@miamioh.edu

MS283

Variational Convexity of Functions and Variational Sufficiency in Optimization

The talk introduces the study, characterizations, and applications of variational convexity of functions, the property that has been recently introduced by Rockafellar together with its strong counterpart. First we show that these variational properties of an extended-real-valued function are equivalent to, respectively, the conventional (local) convexity and strong convexity of its Moreau envelope. Then we derive new characterizations of both variational convexity and variational strong convexity of general functions via their second-order subdifferentials (generalized Hessians), which are coderivatives of subgradient mappings. We also study relationships of these notions with local minimizers and tilt-stable local minimizers. The obtained results are used for characterizing related notions of variational and strong variational sufficiency in composite optimization with applications to nonlinear programming.

Thanh Phat Vo
Wayne State University
phatvt@wayne.edu

Boris Mordukhovich
Wayne State University
Department of Mathematics
boris@math.wayne.edu

Pham Duy Khanh
Ho Chi Minh City University of Education
khanhpd@hcmue.edu.vn

MS284

An Inexact Restoration Direct MultiSearch Filter Approach to Biobjective Constrained Optimization

In practical applications, it is common to have several conflicting objective functions to optimize. Frequently, these functions are of black-box type, preventing the use of multiobjective derivative-based optimization techniques. Direct Multisearch (DMS) is a multiobjective derivative-free optimization class of methods, with a well-established convergence analysis and competitive computational implementations, often successfully used for benchmark of new algorithms and in practical applications. From the theoretical point of view, DMS is developed for continuous optimization with general constraints, making use of an extreme barrier approach, only evaluating feasible points. In this work, we propose the integration of a filter approach in DMS with an inexact feasibility restoration step, to address constrained biobjective optimization problems. The violations of the nonlinear constraints are aggregated and are treated as an additional objective to be minimized. Every time that the algorithm selects an infeasible point as iteration center, an inexact restoration step is performed, in an attempt of restoring feasibility. We will describe the proposed algorithmic structure in detail, provide results on the theoretical properties of the method, and report numerical experiments that state the competitiveness of this approach to address nonlinear constraints.

Everton Silva
NOVA University of Lisbon
NOVA Math
ejo.silva@campus.fct.unl.pt

Ana Luisa Custodio
NOVA University of Lisbon
CMA
alcustodio@fct.unl.pt

MS284

Robust Optimization of Uncertain Multiobjective Problems by a Set Approach

Set-valued optimization using the set approach is a research topic of high interest due to its practical relevance and numerous interdependencies to other fields of optimization. An important example are robust approaches to uncertain multiobjective optimization problems which result in such set optimization problems. However, it is a very difficult task to solve these optimization problems even for specific cases. In this talk, we present parametric multiobjective optimization problems for which the optimal solutions are strongly related to the optimal solutions of the set optimization problems. This corresponds to the well-known idea of scalarization in multiobjective optimization. By using such an approach together with epigraphical reformulations, we can approximate the solution set of uncertain multiobjective problems with desired accuracy.

Gabriele Eichfelder
Institute of Mathematics
Technische Universität Ilmenau
Gabriele.Eichfelder@tu-ilmenau.de

Ernest Quintana
Technical University of Ilmenau
ernest.quintana-aparicio@tu-ilmenau.de

MS284

Novel Approaches for Multi-Objective Cardinality-Constrained Optimization Problems

In this work, we consider multi-objective optimization problems with a cardinality constraint on the vector of variables and, possibly, additional bound constraints. For this class of problems, we propose two novel approaches: the first one aims to find a single Pareto optimal solution, while the second one to approximate the whole Pareto front of the problem at hand. For both methods, we conduct a rigorous theoretical analysis, proving their properties of convergence to points satisfying necessary conditions for global optimality. Some preliminary numerical experiments have been carried out, showing the effectiveness of our approaches with respect to other state-of-the-art methods from the literature.

Pierluigi Mansueto
GoI (Global Optimization Laboratory)
University of Florence
pierluigi.mansueto@unifi.it

Matteo Lapucci
DINFO, Università degli Studi di Firenze
matteo.lapucci@unifi.it

MS285

ProxSkip and Its Variance-Reduced Version for Federated Learning

We study distributed optimization methods based on the

local training (LT) paradigm: achieving communication efficiency by performing richer local gradient-based training on the clients before parameter averaging. Looking back at the progress of the field, we *identify 5 generations of LT methods*: 1) heuristic, 2) homogeneous, 3) sublinear, 4) linear, and 5) accelerated. The 5th generation, initiated by the ProxSkip method and its analysis, is characterized by the first theoretical confirmation that LT is a communication acceleration mechanism. Additionally, we contribute to the 5th generation of LT methods by showing that it is possible to enhance them further using *variance reduction*. While all previous theoretical results for LT methods ignore the cost of local work altogether, and are framed purely in terms of the number of communication rounds, we show that our methods can be substantially faster in terms of the *total training cost* than the state-of-the-art method ProxSkip in theory and practice in the regime when local computation is sufficiently expensive. We characterize this threshold theoretically, and confirm our theoretical predictions with empirical results.

Grigory Malinovsky
King Abdullah University of Science and Technology
grigorii.malinovskii@kaust.edu.sa

Konstantin Mishchenko
INRIA
konsta.mish@gmail.com

Peter Richtarik
King Abdullah University of Science & Technology (KAUST)
Peter.Richtarik@kaust.edu.sa

Sebastian U. Stich
CISPA Helmholtz Center for Information Security, Germany
sebastianustich@gmail.com

Kai Yi
King Abdullah University of Science and Technology (KAUST)
kai.yi@kaust.edu.sa

MS285

Differentially Private Algorithm for Constrained Federated Learning

Differentially private federated learning (FL) is a privacy-preserving machine learning (ML) that enables learning from the distributed data owned by multiple agents without the need for transferring the data into a central server while guaranteeing differential privacy (DP) against inference attacks. In practice, the agents often need to impose constraints on the model parameters of the training models for explaining decisions suggested by ML models, promoting fair decisions, and observing some physical laws. For such constrained FL models, we develop an inexact alternating direction method of multipliers algorithm with the objective perturbation method that injects the noise into the loss function for DP. In particular, the perturbation method enables the resulting parameters to satisfy the constraints and the outperformance as compared with the state-of-the-art algorithms based on the output perturbation. We further enhance the training algorithm by reducing the communication costs of multiple local update technique. We show the privacy and convergence analyses of the proposed algorithm for convex constrained FL. Using datasets for image classification, we demonstrate that our algorithm significantly reduces the testing error compared with the existing DP algorithms while achieving the same level of data privacy.

Minseok Ryu
Argonne National Laboratory
mryu@anl.gov

MS285

Efficient Algorithms for Non-Convex Federated Learning on Heterogeneous Data

Federated learning is a paradigm for machine learning where multiple data holders (clients) collaborate to train a model on their combined dataset. Clients only share partially trained models and other statistics computed from their dataset, while their raw data remains local and private. Statistical heterogeneity between clients pose severe optimization difficulties in FL: data-heterogeneity can slow down the training and reduce accuracy of the trained models. While this training difficulty can be mitigated for convex learning tasks control mechanisms, these techniques do not yet work satisfactory on non-convex problems that are ubiquitous in data-science and deep-learning. In this work we discuss why methods designed for the convex domain cannot transfer their properties to the non-convex regime. We further present efficient solutions that outperform the state of the art FedAvg on non-convex heterogeneous data. We demonstrate the effectiveness of the proposed scheme on standard computer vision benchmarks.

Sebastian U. Stich
CISPA Helmholtz Center for Information Security, Germany
sebastianustich@gmail.com

MS287

On the Centralization of the Circumcentered-Reflection Method

In this talk, we present the first circumcenter iteration scheme that does not employ a product space reformulation for finding a point in the intersection of two closed convex sets. We introduce a so-called centralized version of the circumcentered-reflection method (CRM). Developed with the aim of accelerating classical projection algorithms, CRM is successful for tracking a common point of a finite number of affine sets. In the case of general convex sets, CRM was shown to possibly diverge if Pierras product space reformulation is not used. In this work, we prove that there exists an easily reachable region consisting of what we refer to as centralized points, where pure circumcenter steps possess properties yielding convergence. The resulting algorithm is called centralized CRM (cCRM). In addition to having global convergence, cCRM converges linearly under an error bound condition, and superlinearly if the two target sets are so that their intersection have nonempty interior and their boundaries are locally differentiable manifolds. We also present numerical experiments with successful results.

Roger Behling
Fundação Getûlio Vargas
Universidade Federal de Santa Catarina
roger.behling@ufsc.br

MS287

Second-Order Optimization with Lazy Hessians

We analyze Newton's method with lazy Hessian updates for solving general possibly non-convex optimization problems. We propose to reuse a previously seen Hessian for several iterations while computing new gradients at each step of the method. This significantly reduces the overall arithmetical complexity of second-order optimization schemes. By using the cubic regularization technique, we establish fast global convergence of our method to a second-order stationary point, while the Hessian does not need to be updated each iteration. For convex problems, we justify global and local superlinear rates for lazy Newton steps with quadratic regularization, which is easier to compute. The optimal frequency for updating the Hessian is once every d iterations, where d is the dimension of the problem. This provably improves the total arithmetical complexity of second-order algorithms by a factor $\sqrt{d}$.

Nikita Doikov
École Polytechnique Fédérale de Lausanne
nikita.doikov@epfl.ch

El Mahdi Chayti, Martin Jaggi
EPFL
el-mahdi.chayti@epfl.ch, martin.jaggi@epfl.ch

MS287

A Unified Analysis of Descent Sequences in Weakly Convex Optimization, Including Convergence Rates for Bundle Methods

We present a framework for analyzing the convergence and the respective rate of convergence of a class of descent algorithms, when the objective function is weakly convex. The framework is general, combining the possibility of explicit iterations (based on the gradient or a subgradient at the current iterate), implicit iterations (using a subgradient at the next iteration, like in the proximal schemes), as well as iterations when the associated subgradient is specially constructed at an auxiliary point (this is the case of descent steps in bundle methods). Under a subdifferential-based error bound on the distance to critical points, linear rates of convergence are established. Our analysis applies, among other techniques, to prox-descent for decomposable functions, the proximal-gradient method for a sum of functions, redistributed bundle methods, and a class of algorithms that can be cast in the feasible descent framework for constrained optimization.

Paulo J. S. Silva, Felipe Atenas
University of Campinas
pjssilva@ime.unicamp.br, f262900@dac.unicamp.br

Claudia Sagastizábal
Univeristy of Campinas
sagastiz@unicamp.br

Mikhail V. Solodov
IMPA-Instituto de Matematica Pura e Aplicada, Brazil
solodov@impa.br

MS288

A Primal-Dual Interior Method Based on An Iterative Trust-Region Solver

Implementations of interior-point methods for large-scale nonlinearly constrained optimization generally avoid the use of iterative methods for solving the primal-dual Karush-Kuhn-Tucker (KKT) system in favor of methods based on a sparse symmetric indefinite factorization. For nonconvex problems the need for some procedure to ensure that progress can be made with the resulting direction makes it difficult to apply iterative solvers. In this talk we present a new method for formulating the symmetric indefinite KKT system that provides a well-conditioned trust-region subproblem. In addition, a new iterative method is proposed for solving the trust-region subproblem based on the Lanczos process. Numerical results from the CUTEst test collection demonstrate that the method improves the overall performance of the interior-point method, as well as reducing the computational work required to solve the trust-region subproblem compared to the More-Sorensen approach.

Philip E. Gill, Alexander Guldemond
University of California, San Diego
Department of Mathematics
pgill@ucsd.edu, aguldemo@ucsd.edu

Vyacheslav Kungurtsev
Czech Technical University in Prague
Department of Computer Science
vyacheslav.kungurtsev@fel.cvut.cz

Daniel Robinson
Lehigh University
daniel.p.robinson@gmail.edu

MS288

Recent Developments in the Polyhedral Active Set Algorithm

The Polyhedral Active Set Algorithm (PASA) is designed to optimize a general nonlinear function over a polyhedron. Phase one of the algorithm is a nonmonotone gradient projection algorithm, while phase two is an active set algorithm that explores faces of the constraint polyhedron. Two different implementations of phase two have been developed; one implementation generates iterates by a limited memory nonlinear conjugate gradient method, while the other implementation employs a quadratic model for the objective and iterates that approximate minimizers of the quadratic. A unified algorithm is now developed which dynamically chooses between the implementations of phase two in an effort to optimize performance.

William Hager
University of Florida
Department of Mathematics
hager@ufl.edu

Hongchao Zhang
Department of Mathematics and CCT
Louisiana State University
hozhang@math.lsu.edu

MS288

Communication-Efficient Hessian-Based Federated Optimization

Second-order methods are underutilized in federated learning due to the cost of computing and communicating the Hessian or Jacobian matrices. This issue is exacerbated by the growing sizes of deep learning models and presents a bottleneck in applications requiring second-order information such as Newton's method, bilevel programming, and implicit deep learning. The primary contribution of this work is Hessian-based algorithms in which each client only stores and communicates vectors. To this end, we identify the decentralized computation of inverse matrix-vector product (IMVP) as the central challenge and develop efficient solutions that interpret IMVP as the solution of a system of linear equations. To solve IMVP with vector costs, we leverage Neumann series approximation and propose a novel method called FedIMVP. Our method achieves a significant improvement in convergence rate and communication rounds over the prior art. Finally, we apply our FedIMVP solver to develop an efficient and provable federated stochastic Newton algorithm. To our knowledge, this represents two major improvements over existing literature on Newton-like methods: 1) it operates with vector costs, and 2) benefits from local training. Under deterministic samples, this also leads to an efficient federated Newton algorithm with quadratic convergence. Our evaluations show the benefit of our algorithms (for FedIMVP and FedNewton) and shed light on the fundamental communication/computation tradeoffs.

Davoud Tarzanagh
University of Pennsylvania
tarzanaq@umich.edu

MS289

Solving Cut-Generating Linear Programs Via Machine Learning

Cut-generating linear programs (CGLPs) play a key role as a separation oracle to produce valid inequalities for the feasible region of mixed-integer programs. Running CGLPs at nodes of the branch-and-bound tree, however, is computationally cumbersome due to the large number of node candidates and the lack of a priori knowledge on which nodes admit useful cutting planes. In this paper, we propose a novel framework based on machine learning to approximate the optimal value of the CGLP, which is the deciding factor in generating cutting planes. Translating the CGLP as an indicator function of the objective function vector, we show that it can be approximated through conventional data classification techniques. We provide a systematic procedure to efficiently generate training data sets for the corresponding classification problem based on the CGLP structure. Preliminary computational experiments on benchmark instances will be presented.

Danial Davarnia
Iowa State University
davarnia@iastate.edu

MS289

Convexification Techniques for 0-1 Linear Fractional Programs

In this work, we use convexification techniques to solve 0-1 linear fractional programs. We reduce, via a projective transformation, the convexification of 0-1 linear fractional programs to that of boolean quadratic programs. This allows us to build connections between classes of inequalities of the boolean quadratic polytope with the convex hull of substructures of 0-1 linear fractional programs. This gives us strengthened relaxations for 0-1 linear fractional programs. We also present a preliminary computational study to demonstrate the performance of our relaxations.

Siyue Liu
Carnegie Mellon University
siyueliu@andrew.cmu.edu

Taotao He
Shanghai Jiao Tong University
hetaotao@sjtu.edu.cn

MS289

Approximating Nonlinear Integer Programs with Monomial Orders

We use monomial orders to optimise a submodular or subadditive function over integer points in some compact set. We give approximation results and a characterisation of the convex hull for monomial orders satisfying some properties, and also propose an algorithm to generate feasible solutions that works well on benchmark test instances for integer LPs and integer QPs.

Akshay Gupte, Yiran Zhu
School of Mathematics
University of Edinburgh
akshay.gupte@ed.ac.uk, y.zhu-95@sms.ed.ac.uk

MS290

Robust Statistics for Scalable Solutions in Stochastic Process Optimization

Optimization problems commonly consist in minimizing objective functions where every data point contributes to their evaluation throughout the procedure, whether it be in a standard or stochastic optimization setting. Important examples from the domain of Statistics or Computer Science are likelihood-based methods or general empirical loss minimization problems in which the objective functions (e.g. negative log-likelihoods) are evaluated for each data point and therefore often require computationally-efficient optimization tricks to overcome bottlenecks such as high-dimensional matrix inversion. In this context, one may also be interested in assigning less weight to certain data points (outliers) which can greatly and negatively affect the final solution. However in doing so these weights can also introduce additional bias by steering the optimization away from the solution which minimizes the expected value of the objective function. In this work we put forward an alternative approach to simplify optimization procedures by changing the objective functions using informative moments that target the same solutions as the original functions and can easily be made resistant towards problematic data points. More specifically, we present this type of approach for robust estimation and inference on large scale stochastic processes and use this example to discuss benefits and extensions of switching optimization problems to moment-based criteria.

Roberto Molinari
Auburn University
Department of Mathematics and Statistics
robmolinari@auburn.edu

Stephane Guerrier
University of Geneva
stephane.guerrier@unige.ch

Haotian Xu
Penn State University
haotian.xu2@icloud.com

Maria-Pia Victoria-Feser
University of Geneva
maria-pia.victoriafeser@unige.ch

MS290

A General Approach to Solve Shape-Constrained Distributional Optimization via Importance-Weighted Sample Average Approximation

Shape-constrained optimization arises in a wide range of problems including distributionally robust optimization (DRO) that has surging popularity in recent years. In the DRO literature, these problems are usually solved via reduction into moment-constrained problems using the Choquet representation or its variants, whose tractability is determined in a case-by-case manner. In this work, we develop a general method to solve shape-constrained DRO by integrating sample average approximation with importance sampling, which reduces these infinite-dimensional problems into finite-dimensional linear programs. Our approach handles shape-constrained problems that are beyond the reach of previous Choquet-based reformulation, with vanishing and readily quantifiable optimality gaps. Our theoretical underpinning builds on empirical process theory and reveals how shape constraints play an important role in guaranteeing desirable consistency and convergence rates.

Dashi Singham
Naval Postgraduate School
dsingham@nps.edu

Henry Lam, Zhenyuan Liu
Columbia University
henry.lam@columbia.edu, zl2817@columbia.edu

MS291

Accelerated Riemannian Optimization

Global full acceleration in Riemannian first-order methods under geodesic convexity and smoothness has been studied recently in several works. The aim is to obtain an algorithm with rates of convergence as close as possible to the ones of Nesterov's accelerated gradient descent, and studying under which assumptions this is possible, when we can obtain a fast implementation, how the rates should depend on the curvature, etc. It is known through a lower bound that the rates must contain a geometric term depending on the curvature and the initial distance to the minimizer, which is not present in the Euclidean case. I will present some Riemannian optimization methods that are accelerated and some generic techniques for achieving Riemannian acceleration. For example, an accelerated Riemannian inexact proximal point algorithm and subroutines to implement the inexact prox steps. These techniques lead to general algorithms that work under mild assumptions, in contrast to previous works, that either assumptively require the iterates to be contained in some specific set without any mechanism for enforcing this condition or that are limited to restricted settings, like optimization over a small local neighborhood of the minimizer.

David Martinez-Rubio
Zuse Institute Berlin
david_martirubio@hotmail.com

MS291

A Riemannian Proximal Newton Method

In recent years, the proximal gradient method and its invariants have been generalized to Riemannian manifolds for solving optimization problems in the form of $f+g$, where f is continuously differentiable and g may be nonsmooth. In this talk, we generalize the proximal Newton-type method to the Riemannian setting. It is shown that the direct generalization by replacing the Euclidean gradient and Euclidean Hessian by the Riemannian gradient and Riemannian Hessian does not yield a satisfactory result in the sense that the asymptotic convergence rate is only linear. We propose a new approach that utilizes the Weingarten and semismooth analysis. The local superlinear convergence is given under reasonable assumptions. Numerical experiments for problems on the unit sphere and oblique manifold are used to demonstrate the performance of the proposed method. Joint work with Wen Huang, Rujun Jiang, Simon Vary, Pierre-Antoine Absil

Wutao Si
UCLouvain
wutao.si@uclouvain.be

MS291

**Solving graph equipartition SDPs on an algebraic

variety

Riemannian optimization is a powerful method for solving low-rank SDP problems whose feasible sets are manifolds. In this talk, we focus on solving the graph balanced cut SDP whose feasible set is not a manifold. The constraints of the decomposed problem involve an algebraic variety with conducive geometric properties which we analyse. We also develop a method to escape from a non-optimal singular point on this variety. This allows us to use Riemannian optimization techniques together with the previous method to solve the SDP problem very efficiently with certified global optimality.

Tianyun Tang
National University of Singapore
ttang@u.nus.edu

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

MS293

An Algorithmic Solution to the Blotto Game Using Multi-Marginal Couplings

We describe an efficient algorithm to compute solutions for the general two-player Blotto game on n battlefields with heterogeneous values. While explicit constructions for such solutions have been limited to specific, largely symmetric or homogeneous, setups, this algorithmic resolution covers the most general situation to date: value-asymmetric game with asymmetric budget. The proposed algorithm rests on recent theoretical advances regarding Sinkhorn iterations for matrix and tensor scaling. An important case which had been out of reach of previous attempts is that of heterogeneous but symmetric battlefield values with asymmetric budget. In this case, the Blotto game is constant-sum so optimal solutions exist, and our algorithm samples from an ϵ-optimal solution in time $\tilde{O}(n^2 + \epsilon^{-4})$, independently of budgets and battlefield values. In the case of asymmetric values where optimal solutions need not exist but Nash equilibria do, our algorithm samples from an ϵ-Nash equilibrium with similar complexity but where implicit constants depend on various parameters of the game such as battlefield values.

Vianney Perchet
CMLA Research Center for Applied Maths
ENS Paris Saclay
vianney.perchet@gmail.com

MS293

Estimation of Unknown Payoff Parameters in Large Network Games

We consider games where a large number of agents interact according to a network sampled from a random graph model, represented by a graphon. By exploiting previous results on convergence of such large network games to graphon games, we examine a procedure for estimating unknown payoff parameters, from observations of equilibrium actions, without the need for exact network information. We prove smoothness and local convexity of the optimization problem involved in computing the proposed estimator. Additionally, under a notion of graphon parameter identifiability, we show that the optimal estimator is globally unique and derive sufficient conditions for the optimization problem to be globally convex. We present several examples of identifiable homogeneous and heterogeneous parameters in different classes of network games with numerical simulations to validate the proposed estimator.

Francesca Parise, Kiran Rokade
Cornell University
fp264@cornell.edu, rokade.kiran94@gmail.com

MS293

Decentralized Learning in Online Queuing Systems

Motivated by packet routing in computer networks and resource allocation in radio networks, online queuing systems are composed of queues receiving packets at different rates. Repeatedly, they send packets to servers, each of them treating only at most one packet at a time. In the centralized case, the number of accumulated packets remains bounded (i.e., the system is *stable*) as long as the ratio between service rates and arrival rates is larger than 1. In the decentralized case, individual no-regret strategies ensures stability when this ratio is larger than 2. Yet, myopically minimizing regret disregards the long term effects due to the carryover of packets to further rounds. On the other hand, minimizing long term costs leads to stable Nash equilibria as soon as the ratio exceeds $\frac{e}{e-1}$. Stability with decentralized learning strategies with a ratio below 2 was a major remaining question. We first argue that for ratios up to 2, cooperation is required for stability of learning strategies, as selfish minimization of policy regret, a *patient* notion of regret, might indeed still be unstable in this case. We therefore consider cooperative queues and propose the first learning decentralized algorithm guaranteeing stability of the system as long as the ratio of rates is larger than 1, thus reaching performances comparable to centralized strategies.

Flore Sentenac
CREST
flore.sentenac@gmail.com

MS294

Higher-Order Learning in Large-Population Games

In this talk, we discuss principles and algorithms for decision making in population games where a large number of decision-making agents select strategies to engage in strategic interactions with one another. Rather than compute and adopt the optimal strategy selection based on a known cost function, the agents are required to learn such strategy selection based on instantaneous rewards they receive over repeated interactions. Using the evolutionary dynamics formalism, I explain principled approaches to define an approximate population decision model from individual agents' strategy selection rules, and how the notion of passivity from dynamical system theory is used to examine convergence of the agents' repeated strategy selection to the Nash equilibrium, where no agent can better off by changing its strategy unilaterally. We also discuss how higher-order learning can improve robustness, especially when the agents' strategy selection is subject to time delays. Multi-robot task allocation will be presented to illustrate an application in engineering domains.

Shinkyu Park
King Abdullah University of Science and Technology
shinkyu.park@kaust.edu.sa

MS294

Escaping Saddle Points in Zeroth-Order Optimization: Two Function Evaluations Suffice

Two-point zeroth order methods are important in many applications of zeroth-order optimization arising in robotics, wind farms, power systems, online optimization, and adversariable robustness to black-box attacks in deep neural networks, where the problem can be high-dimensional and/or time-varying. Furthermore, such problems may be nonconvex and contain saddle points. While existing works have shown that zeroth-order methods utilizing $\Omega(d)$ function valuations per iteration (with d denoting the problem dimension) can escape saddle points efficiently, it remains an open question if zeroth-order methods based on two-point estimators can escape saddle points. In this paper, we show that by adding an appropriate isotropic perturbation at each iteration, a zeroth-order algorithm based on $2m$ (for any $1 \leq m \leq d$) function evaluations per iteration can not only find ϵ-second order stationary points polynomially fast, but do so using only $\tilde{O}\left(\frac{d}{\epsilon^{2.5}}\right)$ function evaluations.

Zhaolin Ren
Harvard University
zhaolinren@g.harvard.edu

Yujie Tang
Peking University
yujietang@pku.edu.cn

Na Li
Harvard University
nali@seas.harvard.edu

MS294

Deep Networks and the Multiple Manifold Problem

Data in science and engineering often exhibit nonlinear, low-dimensional structure, due to the physical laws that govern data generation. In this talk, we study how deep neural networks interact with structured data: + When can we guarantee to fit and generalize? + How do the resources (depth, width, data) required depend on the complexity of the data? + How can we leverage physical prior knowledge to reduce these resource requirements? Our main mathematical result is a guarantee of generalization for a model classification problem involving data on low-dimensional manifolds we prove that for networks of polynomial width, with polynomially many samples, randomly initialized gradient descent rapidly converges to a solution which correctly labels *every* point on the two manifolds. To our knowledge this is the first such result for deep networks on data which are not linearly separable. We highlight intuitions about the roles of depth, width, sample complexity, and the geometry of feature representations, which may be useful in analyzing other problems involving low-dimensional structure (e.g., model discovery). We illustrate these ideas through applied problems in astrophysics and computer vision. In these settings, we further suggest how incorporating physical prior knowledge can reduce the resources (architecture, data) required for learning, leading to more efficient and interpretable learning architectures.

John Wright
Columbia University
jw2966@columbia.edu

MS295

Reduction of Matroids and Its Application to Optimization Problems

A matroid N is called a reduction of another matroid M if each independent set of N is also independent in M. The idea of reducing a matroid to a simpler one goes back to the late 60's, when Crapo and Rota [H. H. Crapo and G. C. Rota, Combinatorial Geometries, MIT press Cambridge, 1968] introduced the notion of weak maps. Weak maps were further investigated by Lucas [D. Lucas. Properties of rank preserving weak maps, Bulletin of the American Mathematical Society, 1974] who characterized rank-preserving weak maps for linear matroids. In this talk, we concentrate on the case when the reduced matroid is ought to be a partition matroid. There is a one-to-one correspondence between such reductions of a matroid and its rainbow circuit-free colorings. Indeed, taking the partition classes of any reduction to a partition matroid as color classes gives a rainbow circuit-free coloring. Vica versa, if we pick one element from each color class of any rainbow circuit-free coloring, then the resulting set is independent – in other words, the partition matroid defined by the color classes provides a reduction of the matroid. We discuss several applications of reductions, starting from packing common independent sets in the intersection of two matroids, through characterizing binary matroids, up to proving a matroidal extension of Sperners lemma.

Kristóf Bérczi
Eötvös Loránd University
Hungary
kristof.berczi@ttk.elte.hu

Tamás Schwarcz
Eötvös Loránd University
tamas.schwarcz@ttk.elte.hu

MS295

Optimal General Factor Problem and Jump System Intersection

In the optimal general factor problem, we are given a graph $G = (V, E)$ and a set $B(v)$ of integers for each $v \in V$, and the objective is to find an edge subset $F \subseteq E$ of maximum cardinality subject to $d_F(v) \in B(v)$ for $v \in V$. A recent crucial work by Dudycz and Paluch shows that this problem can be solved in polynomial time if each $B(v)$ has no gap of length more than one. While their algorithm is simple, its correctness proof is quite complicated. In this talk, we formulate the optimal general factor problem as the jump system intersection, and reveal when the algorithm by Dudycz and Paluch can be applied to this abstract form of the problem. By using this abstraction, we give another correctness proof of the algorithm, which is much simpler than the original one.

Yusuke Kobayashi
Kyoto University
yusuke@kurims.kyoto-u.ac.jp

MS296

On the Power of Static Assignments Policies for Robust Facility Location Problems

We consider a two-stage robust facility location problem on a metric under an uncertain demand. The decision-maker needs to decide on the (integral) units of supply for each facility in the first stage to satisfy an uncertain second-stage demand, such that the sum of first stage supply cost and the worst-case cost of satisfying the second-stage demand over all scenarios is minimized. The second-stage decisions are only assignment decisions without the possibility of adding recourse supply capacity. This makes our model different from existing work on two-stage robust facility location and set covering problems. We consider an implicit model of uncertainty with an exponential number of demand scenarios specified by an upper bound k on the number of second-stage clients. In an optimal solution, the second-stage assignment decisions depend on the scenario; surprisingly, we show that restricting to a fixed (static) fractional assignment for each potential client irrespective of the scenario gives us an $O(\log k/\log\log k)$-approximation for the problem. Moreover, the best such static assignment can be computed efficiently giving us the desired guarantee.

Omar El Housni
Cornell University
oe46@cornell.edu

Vineet Goyal
Columbia University
vgoyal@ieor.columbia.edu

David B. Shmoys
Cornell University
david.shmoys@cornell.edu

MS296

Scalable Nonlinear Decision Rules Via Nonparametric Liftings

Decision rules provide approximate solutions by restricting decisions to simple functions of uncertainties in dynamic robust optimization. In this paper, we consider a nonparametric lifting framework where the uncertainty space is lifted to higher dimensions to obtain nonlinear decision rules. Current lifting-based approaches require pre-determined functions and are parametric. We propose two nonparametric liftings, which derive the nonlinear functions by leveraging the uncertainty set structure and problem coefficients. Both methods integrate the benefits from lifting and nonparametric approaches, and hence, provide scalable decision rules with performance bounds. More specifically, the set-driven lifting is constructed by finding polyhedrons within uncertainty sets, inducing piecewise-linear decision rules with performance bounds. The dynamics-driven lifting, on the other hand, is constructed by extracting geometric information and accounting for problem coefficients. This is achieved by using linear decision rules of the original problem, also enabling to quantify lower bounds of objective improvements over linear decision rules. Numerical comparisons with competing methods demonstrate superior computational scalability and comparable performance in objectives. These observations are magnified in multistage problems with extended time horizons, suggesting practical applicability of the proposed nonparametric liftings in large-scale dynamic robust optimization.

Eojin Han
Southern Methodist University
eojinh@smu.edu

Omid Nohadani
Benefits Science Technologies
onohadani@gmail.com

MS296

Inexact Bundle Method for Two-Stage Stochastic Programming

We propose an inexact bundle method for solving two-stage stochastic programming problems that arise from important application areas including revenue management and power system. We consider the setting that it is intractable to compute true objective function and gradient information, and instead only estimates of objective function and gradient values are available. Under common assumptions, the algorithm generates a sequence of iterates converging to a neighborhood of optimality, where the radius of neighborhood depends on the level of inexactness from stochastic objective function estimates. The numerical results demonstrate empirical performance of our proposed algorithm.

Baoyu Zhou, Haihao Lu
University of Chicago
zbaoyu@uchicago.edu, haihao.lu@chicagobooth.edu

John Birge
University of Chicago
Booth School of Business
john.birge@chicagobooth.edu

MS297

Scalable Symmetric Tucker Decomposition via Projected Gradient Descent

We study the best low-rank Tucker decomposition of symmetric tensors, advocating a straightforward projected gradient descent (PGD) method and the higher-order eigenvalue decomposition (HOEVD) approximation for its computation. The main application of interest is in decomposing higher-order multivariate moments, which are symmetric tensors. We develop scalable adaptations of the basic PGD and HOEVD methods to decompose sample moment tensors. With the help of implicit and streaming techniques, we evade the overhead cost of building and storing the moment tensor. Such reductions make computing the Tucker decomposition realizable for large data instances in high dimensions. Numerical experiments demonstrate the efficiency of the algorithms and the applicability of moment tensor decompositions to real-world datasets. Lastly, we study the convergence on the Grassmannian manifold, and prove that the update sequence derived by the PGD solver achieves first and second-order criticality.

Ruhui Jin
University of Wisconsin
rjin@math.wisc.edu

Joe Kileel
University of Texas at Austin, U.S.
jkileel@math.utexas.edu

Tammy Kolda
MathSci
tammy.kolda@mathsci.ai

Rachel Ward
Department of Mathematics
University of Texas at Austin
rward@math.utexas.edu

MS297

Certificates for Well-Posedness of Best Canonical Polyadic Approximation

The canonical polyadic decomposition (CPD) of higher-order tensors plays a major role in data analysis and signal processing by allowing for unique recovery of underlying factors. However, it is well known that the low-rank approximation problem is ill-posed in general. That is, a tensor may fail to have a best rank-R approximation when $R > 1$. In this talk, we switch to the opposite point of view, namely that the CP approximation problem is well-posed, unless the tensor is "too far" from the tensors with rank at most R. This is more an engineering perspective: the problem is well-posed unless the data are too noisy for a proper analysis. We give deterministic bounds under which the existence of a best low-rank approximation is guaranteed. We pay special attention to approximations of tensors that are positive semidefinite, as they appear in many applications. For positive semidefinite tensors of order three, we show that our dedicated bound is sharp and that it can be computed using semidefinite programming.

Lieven De Lathauwer
KU Leuven
lieven.delathauwer@kuleuven.be

Eric Evert
Northwestern University
eric.m.evert@gmail.com

Raphaël Widdershoven
STADIUS - ESAT
KU Leuven
raphael.widdershoven@kuleuven.be

MS298

Stochastic trust-region algorithm in random subspaces

This work proposes a framework for large-scale stochastic derivative-free optimization (DFO) by introducing STARS, a trust-region method which achieves scalability using random models in random low-dimensional affine subspaces. STARS significantly reduces per-iteration costs in terms of function evaluations, thus yielding strong performance on large-scale stochastic DFO problems. The user-determined dimension of these subspaces can be chosen via so-called Johnson–Lindenstrauss transforms, and independently of the dimension of the problem. For convergence purposes, both a particular quality of the subspace and the accuracies of random function estimates and models are required to hold with sufficiently high, but fixed, probabilities. Convergence and expected complexity results of STARS are obtained using martingale theory.

Kwassi Joseph Dzahini
Argonne National Laboratory
kdzahini@anl.gov

Stefan Wild
Argonne National Laboratory
LANS, Mathematics and Computer Science
wild@lbl.gov

MS298

Direct search based on probabilistic descent in reduced subspaces

In this talk, we propose a generic direct-search algorithm in which the polling directions are defined using random subspaces. Complexity guarantees for such an approach are derived thanks to probabilistic properties related to both the subspaces and the directions used within these subspaces. By leveraging results on random subspace embeddings and sketching matrices, we show that better complexity bounds are obtained for randomized instances of our framework. A numerical investigation confirms the benefit of randomization, particularly when done in subspaces, when solving problems of moderately large dimension.

Lindon Roberts, Lindon Roberts
University of Sydney
lindon.roberts@sydney.edu.au,
lindon.roberts@sydney.edu.au

Clement W. Royer
Université Paris Dauphine-PSL

clement.royer@lamsade.dauphine.fr

MS298

PRIMA: Reference Implementation for Powell's methods with Modernization and Amelioration

I will introduce PRIMA, a project that aims to provide the reference implementation of Powell's methods in modern languages, including modern Fortran (F2008 or newer), MATLAB, Python, C++, and probably Julia and R. It will be a faithful implementation, in the sense that the code will be mathematically equivalent to Powells, except for the bug fixes and improvements that we make intentionally. The focus is to implement these methods in a structured and modularized way so that they are easily understandable, maintainable, extendable, and future-proof. The code will have no GOTO (of course) and will use matrix-vector procedures instead of loops whenever possible. I will present the current status of PRIMA, the bugs and issues we have fixed, and the improvements we have achieved.

Zaikun Zhang
Hong Kong Polytechnic University
zaikun.zhang@polyu.edu.hk

MS299

Kaczmarz for Doubly Noisy Linear Systems

Large-scale linear systems, $Ax = b$, frequently arise in practice and demand effective iterative solvers. Often, these systems are noisy due to operational errors or faulty data-collection processes. In the past decade, the randomized Kaczmarz algorithm (RK) was studied extensively as an efficient iterative solver for such systems. However, the convergence study of RKA in the noisy regime is limited and considers measurement noise in the right-hand side vector, b. Unfortunately, in practice, that is not always the case; the coefficient matrix A can also be noisy. In this talk, we motivate and discuss the application of RK to doubly noise linear systems, i.e., linear systems with noise in both the measurements and the measurement matrix.

Anna Ma
University of California - Irvine
anna.ma@uci.edu

MS299

Minimal Error Momentum Kaczmarz

Recently, it was observed that introducing a heavy ball acceleration term in the Kaczmarz method can improve convergence, if the data contains several coherent measurements. So far, this method has only been considered with a fixed momentum parameter β, which needs to be tuned experimentally. Convergence rates have been shown for a certain range for β, which can be much smaller than observed in practice. Also, it appears hard to compare the convergence rate with the one for the vanilla Kaczmarz method. We compute an adaptive momentum parameter which is optimal in the sense that it gives minimal distances to all solutions without any prior knowledge on the set of solutions. We also compute a corresponding adaptive step size which is optimal in the same sense. The resulting methods are hyperparameter-free and converge with at least the same expected linear rate which has been established for the Kaczmarz method. In experiments, we compare the performance of the proposed methods with the Kaczmarz method and its heavy ball acceleration with fixed momentum parameters.

Maximilian Winkler
TU Braunschweig
maximilian.winkler@tu-braunschweig.de

MS300

A Dynamical System View of Langevin-Based Non-Convex Sampling

Non-convex sampling is a key challenge in machine learning, central to non-convex optimization in deep learning as well as to approximate probabilistic inference. Despite its significance, theoretically there remain many important challenges: Existing guarantees (1) typically only hold for the averaged iterates rather than the more desirable last iterates, (2) lack convergence metrics that capture the scales of the variables such as Wasserstein distances, and (3) mainly apply to elementary schemes such as stochastic gradient Langevin dynamics. In this paper, we develop a new framework that lifts the above issues by harnessing several tools from the theory of dynamical systems. Our key result is that, for a large class of state-of-the-art sampling schemes, their last-iterate convergence in Wasserstein distances can be reduced to the study of their continuous-time counterparts, which is much better understood. Coupled with standard assumptions of MCMC sampling, our theory immediately yields the last-iterate Wasserstein convergence of many advanced sampling schemes such as proximal, randomized mid-point, and Runge-Kutta integrators. Beyond existing methods, our framework also motivates more efficient schemes that enjoy the same rigorous guarantees.

Ya-Ping Hsieh
ETH Zurich
yaping.hsieh@inf.ethz.ch

MS300

Sampling with Riemannian Hamiltonian Monte Carlo in a Constrained Space

We demonstrate for the first time that ill-conditioned, non-smooth, constrained distributions in very high dimensions, upwards of 100,000, can be sampled efficiently in practice. Our algorithm incorporates constraints into the Riemannian version of Hamiltonian Monte Carlo and maintains sparsity. This allows us to achieve a mixing rate independent of smoothness and condition numbers. On benchmark data sets in systems biology and linear programming, our algorithm outperforms existing packages by orders of magnitude. In particular, we achieve a 1,000-fold speed-up for sampling from the largest published human metabolic network (RECON3D). Our package has been incorporated into the COBRA toolbox.

Ruoqi Shen
University of Washington, U.S.
shenr3@cs.washington.edu

Yunbum Kook
KAIST
Discrete Mathematics Group, Institute for Basic Science
yb.kook@kaist.ac.kr

Yin Tat Lee
University of Washington
yintat@uw.edu

Santosh Vempala
Georgia Institute of Technology
vempala@gatech.edu

MS300

Convergence of the Inexact Langevin Algorithm with Application to Score-based Generative Models

We study the Inexact Langevin Algorithm (ILA) for sampling using estimated score function when the target distribution satisfies log-Sobolev inequality (LSI), motivated by Score-based Generative Modeling (SGM). We prove a long-term convergence in Kullback-Leibler (KL) divergence under a sufficient assumption that the error of the score estimator has a bounded Moment Generating Function (MGF). Our assumption is weaker than L^∞ (which is too strong to hold in practice) and stronger than L^2 error assumption, which we show not sufficient to guarantee convergence in general. Under the L^∞ error assumption, we additionally prove convergence in Rényi divergence, which is stronger than KL divergence. We then study how to get a provably accurate score estimator which satisfies bounded MGF assumption for LSI target distributions, by using an estimator based on kernel density estimation. Together with the convergence results, we yield the first end-to-end convergence guarantee for ILA in the population level. Last, we generalize our convergence analysis to SGM and derive a complexity guarantee in KL divergence for data satisfying LSI under MGF-accurate score estimator. Joint work with Kaylee Yang.

Andre Wibisono
Yale University
andre.wibisono@yale.edu

MS301

Parallel and Alternating Proximal Methods for Nonsmooth Nonconvex Minimax: A Unified Convergence Analysis

Nonconvex nonsmooth min-max problems are abundant in modern applications. We present proximal based gradient schemes with either parallel or alternating steps. We show that both methods can be analyzed through a single scheme within a unified analysis which relies on expanding a general convergence mechanism used for analyzing nonconvex nonsmooth optimization problems. In contrast to the current literature which focuses on the complexity of obtaining nearly approximate stationary solutions, we prove subsequence convergence to a critical point of the primal objective and global convergence when the latter is semialgebraic. Furthermore, as a by-product, we provide refined complexity results.

Eyal Cohen
Tel Aviv University
Israel
eyalcohen1@mail.tau.ac.il

Marc Teboulle
Tel Aviv University
teboulle@post.tau.ac.il

MS302

Discretization and Reduction Techniques of Infsup and Sip Optimization Problems

Infsup convex optimizations problems in Rn are shown to be discretizable and reducible to smaller infsup problems, each involving at most n+1 functions. The maximum in the initial problem is taken over an arbitrary (possibly infinite) family of convex extended real-valued functions indexed on a Hausdorff compact space, with an upper semicontinuous dependence of this index. Reformulations of Semi-Infinite Programming problems (SIP, for short) into Infsup problems will allow the discretization and reduction to smaller infsup problems, which are in turn written as smaller SIP problems when additional (Slater) conditions are applied. We also provide a relationship between the solutions of the reference optimization problem and its reduced subproblems. Equivalently, we give some new expressions for the subdifferential of the supremum function that are based only on the subdifferential of at most n+1 active functions.

Abderrahim Hantoute
University of Alicante
hantoute@ua.es

MS302

Coderivative Leibniz Rules for Expected Set-Valued Mappings and and Applications to Sensitivity Analysis of Stochastic Systems

In this talk, we summary some results about coderivative calculus of expected-integral multifunctions given in the form

$$\mathrm{E}_\Phi(x) := \int_T \Phi_t(x) d\mu,$$

where $\Phi\colon T \times \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued mapping on a measure space $(T, \mathcal{A}, \mu)$. The calculus rules are used to give applications to sensitivity analysis of random constraint and variational systems related to stochastic optimization and variational inequalities

Pedro Pérez-Aros
Universidad de O'Higgins
Universidad de O'Higgins
pedro.perez@uoh.cl

MS302

Extensions of Constant Rank Qualification Constraint Condition to Nonlinear Conic Programming

We present new constraint qualification conditions for nonlinear conic programming that extend some of the constant rank-type conditions from nonlinear programming. Specifically, we propose a general and geometric approach, based on the study of the faces of the cone, for defining a new extension of this condition to the conic context. We then compare these new conditions with some of the existing ones, including the nondegeneracy condition, Robinsons constraint qualification, and the metric subregularity constraint qualification. The main advantage of the latter is that we are able to recast the strong second-order properties of the constant rank condition in a conic context. In particular, we obtain a second-order necessary optimality condition that is stronger than the classical one obtained under Robinsons constraint qualification, in the sense that it holds for every Lagrange multiplier, even though our condition is independent of Robinsons condition.

Héctor Ramírez
Centro de Modelamiento, Matemático-Universidad de Chile
hramirez@dim.uchile.cl

MS303

Energy Supply Network Optimization in a Multi-objective Setting

Facing the current developments of the climate crisis, it is necessary to consider more than the mere economic perspective in energy supply network planning. While trying to minimize the costs for, e.g., necessary grid expansions, one also wants to avoid overstraining the social consent with the proposed actions as well as to lower the carbon emissions as much as possible. This leads to a classical multiobjective optimization problem in the sense of conflicting objective functions. Due to the presence of choices in the model along with respecting stationary flow equations the resulting optimization problem has a mixed-integer nonconvex structure. In this talk, we present an approach for computing an enclosure of the nondominated set of such problems suitable for larger network instances.

Moritz Link
University of Konstanz
Mathematics und Statistics
moritz.link@uni-konstanz.de

Stefan Volkwein
Universität Konstanz
Department of Mathematics and Statistics
stefan.volkwein@uni-konstanz.de

MS303

A Derivative-Free Approach to Mixed Integer Constrained Multi-Objective Non-Smooth Optimization

In this work, we consider multi-objective optimization problems with both bound constraints on the variables and general nonlinear constraints, where objective and constraint function values can only be obtained by querying a black-box. Furthermore, we consider the case where a subset of the variables can only take integer values. We propose a new linesearch-based solution method and show that it converges to a set of stationary points for the problem. For what concerns the continuous variables, we employ a strategy for the estimation of the Pareto frontier recently proposed in the literature and which takes advantage of dense sequences of search directions. The subset of variables that must assume discrete values are dealt with using primitive directions appropriately modified to take into account the presence of more than one objective functions. Numerical results obtained with the proposed method on a set of test problems and comparison with other solution methods show the viability and efficiency of the proposed approach.

Giampaolo Liuzzi
Dipartimento di Ing, Informatica Automatica e Gestionale
"Sapienza" Università di Roma, Rome, Italy
liuzzi@diag.uniroma1.it

Stefano Lucidi
Dipartimento di Ing, Informatica Automatica e Gestionale
lucidi@diag.uniroma1.it

MS303

A Tailored Multi-Objective Branch-and-Bound Algorithm for the Design-for-Control of Water Networks with Global Bounds

This work investigates multi-objective water distribution network design-for-control (DfC) problems, which consist in simultaneously optimizing design decisions (the installation of new pipes and/or valves) and control variables (valve settings) for the joint minimization of pressure-induced leakage, maximization of resilience and minimization of design cost. We are interested in computing approximations of the Pareto fronts of the resulting nonconvex multi-objective mixed-integer non-linear programs (MOMINLP) with global bounds of ϵ-non-dominance. Following a general multi-objective branch-and-bound framework from the literature, we present tailored node selection, branching, lower bounding, discarding and upper bounding rules inspired by successful branch-and-bound strategies for the single-objective design-for-control of WDNs. The proposed algorithm is applied to the optimal DfC of two case study networks and is shown to outperform alternative (scalarization-based) global methods and heuristic approaches from the literature.

Aly-Joy Ulusoy
Department of Civil and Environmental Engineering
Imperial College London
aly-joy.ulusoy15@imperial.ac.uk

MS304

Q-Gadmm: Quantized Group Admm for Communication Efficient Decentralized Machine Learning

We propose a communication-efficient decentralized machine learning (ML) algorithm, coined *quantized group ADMM (Q-GADMM)*. To reduce the number of communication links, every worker in Q-GADMM communicates only with two neighbors, while updating its model via the group alternating direction method of multipliers (GADMM). Moreover, each worker transmits the quantized difference between its current and its previously quantized model, thereby decreasing the communication payload size. However, due to the lack of centralized entity in decentralized ML, the spatial sparsity and payload compression may incur error propagation, hindering model training convergence. To overcome this issue, we develop a novel stochastic quantization method to adaptively adjust model quantization levels while proving the convergence of Q-GADMM for convex problems. To demonstrate the feasibility of Q-GADMM for non-convex problems, we propose quantized stochastic GADMM (Q-SGADMM) that incorporates deep neural network architectures and stochastic sampling. Simulation results corroborate that Q-GADMM (Q-SGADMM) significantly outperforms GADMM in terms of communication efficiency while achieving the same accuracy and convergence speed for a linear regression task. Similarly, for an image classification task using DNN, Q-SGADMM achieves significantly less total communication cost with identical accuracy and convergence speed compared to stochastic GADMM (SGADMM).

Anis Elgabli
University of Oulu
anis.elgabli@oulu.fi

MS304

Totally Asynchronous Federated Learning

Synchronous Federated Learning (FL) is known to suffer from slow convergence rates owing to the presence of straggler nodes. Conversely, asynchronous FL methods alleviate this issue by allowing the server to update the global model without waiting for an update from every node. Existing asynchronous FL methods typically rely on aggregation schemes where incoming information from clients is stored in a buffer and accessed periodically or weighted according to staleness criteria. In this work we adopt an alternative, totally asynchronous approach, where nodes randomly and independently query the server with different probabilities; the client updates are then appropriately aggregated and stored in a buffer which is randomly accessed by the server in order to update the global model. For strongly convex functions with Lipschitz gradients, the proposed method provably converges (in expectation) to the optimal solution with a linear rate. Our theoretical analysis is supported by our numerical results on a regularized logistic regression problem.

Charikleia Iakovidou
Argonne National Laboratory
ciakovidou@anl.gov

MS304

Adaptive Client Sampling in Federated Learning Via Online Learning with Bandit Feedback

Due to the high cost of communication, federated learning (FL) systems need to sample a subset of clients that are involved in each round of training. As a result, client sampling plays an important role in FL systems as it affects the convergence rate of optimization algorithms used to train machine learning models. We cast client sampling as an online learning task with bandit feedback, which we solve with an online stochastic mirror descent (OSMD) algorithm designed to minimize the sampling variance. We then theoretically show how our sampling method can improve the convergence speed of optimization algorithms. To handle the tuning parameters in OSMD that depend on the unknown problem parameters, we use the online ensemble method and doubling trick. We prove a dynamic regret bound relative to any sampling sequence. The regret bound depends on the total variation of the comparator sequence, which naturally captures the intrinsic difficulty of the problem. These theoretical contributions are new and the proof technique is of independent interest. Through both synthetic and real data experiments, we illustrate advantages of the proposed client sampling algorithm over the widely used uniform sampling and existing online learning based sampling strategies. The proposed adaptive sampling procedure is applicable beyond the FL problem studied here and can be used to improve the performance of stochastic optimization procedures such as stochastic gradient and coordinate descent.

Mladen Kolar
University of Chicago Booth School of Business
mkolar@chicagobooth.edu

MS305

Optimal Low-Rank Matrix Completion: Semidefinite Relaxations and Eigenvector Disjunctions

Low-rank matrix completion consists of computing a matrix of minimal complexity that recovers a given set of observations as accurately as possible, and has numerous applications such as product recommendation. Unfortunately, existing methods for solving low-rank matrix completion are heuristics that, while highly scalable and often identifying high-quality solutions, do not possess any optimality guarantees. We reexamine matrix completion with an optimality-oriented eye, by reformulating low-rank problems as convex problems over the non-convex set of projection matrices and implementing a disjunctive branch-and-bound scheme that solves them to certifiable optimality. Further, we derive a novel and often tight class of convex relaxations by decomposing a low-rank matrix as a sum of rank-one matrices and incentivizing, via a Shor relaxation, that each two-by-two minor in each rank-one matrix has determinant zero. In numerical experiments, our new convex relaxations decrease the optimality gap by two orders of magnitude compared to existing attempts. Moreover, we showcase the performance of our disjunctive branch-and-bound scheme and demonstrate that it solves matrix completion problems over 150x150 matrices to certifiable optimality in hours, constituting an order of magnitude improvement on the state-of-the-art for certifiably optimal methods.

Ryan G. Cory-Wright
IBM Research, USA
and Imperial College London
r.cory-wright@imperial.ac.uk

MS305

Activated Benders Decomposition for Day-Ahead

Paratransit Itinerary Planning

To address the high costs and complexity of paratransit operations, we optimize day-ahead driver itineraries while accounting for uncertainty from trip cancellations and driver no-shows. We formulate a two-stage stochastic optimization model, with a vehicle shareability representation of routing operations to define a tight recourse model. This formulation involves exponentially many variables and constraints with column-dependent rowsa notoriously challenging class of problems. We develop an activated Benders decomposition approach that exploits linking relationships between the first-stage and second-stage problems to (i) accelerate the generation of Benders cuts by lifting the solution of a restricted subproblem into global optimality and feasibility cuts; and to (ii) strengthen the Benders cuts with locally Pareto-optimal cuts. Using data from a major paratransit platform, we show that our algorithm scales to real-world instances, outperforming several benchmarks in terms of computational times, solution quality, and solution guarantees. From a practical standpoint, the SIPPAR model mitigates operating costs by strategically adding slack to driver itineraries in order to create flexibility and robustness against operating disruptions.

Kayla Cummings
Massachusetts Institute of Technology, U.S.
kaylac@mit.edu

MS305

Branch-and-Price for Prescriptive Contagion Analytics

Motivated by vaccine allocation problems, this paper formalizes prescriptive contagion analytics problems where a centralized decision-maker allocates shared resources across multiple segments of a population, each governed by contagion dynamics. These problems involve mixed-integer non-convex optimization models with constraints governed by ordinary differential equations. We develop a branch-and-price methodology based on: (i) a set partitioning reformulation; (ii) a column generation decomposition; (iii) a novel state clustering algorithm for discrete-decision continuous-state dynamic programming; and (iv) a novel tri-partite branching scheme to circumvent nonlinearities. Extensive experiments show that the algorithm scales to large and otherwise-intractable instances, significantly outperforming state-of-the-art benchmarks. Our methodology provides a novel decision-making tool to support resource allocation in contagion systems. In particular, its application can increase the effectiveness of vaccination campaigns by 50-70%, resulting in 17,000 extra saved lives over 12 weeks in a situation mirroring the COVID-19 pandemic.

Alexandre Jacquillat
MIT Sloan School of Management
alexjacq@mit.edu

MS306

Adaptive Sampling Sequential Quadratic Programming for Stochastic Constrained Optimization

This talk presents a methodology for using varying sample sizes in sequential quadratic programming (SQP) methods for solving equality-constrained stochastic optimization problems. We develop adaptive criteria for controlling the sample size, and the accuracy in the solutions of the SQP subproblems based on the variance estimates obtained as the optimization progresses. We establish global convergence results and sample complexity results for the proposed method and demonstrate the performance of the method on a subset of the CUTE problems and constrained classification tasks.

Albert S. Berahas
Department of Industrial and Operations Engineering
University of Michigan
albertberahas@gmail.com

Raghu Bollapragada
Operations Research and Industrial Engineering
The University of Texas at Austin
raghu.bollapragada@utexas.edu

Baoyu Zhou
University of Chicago
zbaoyu@uchicago.edu

MS306

Adaptive Stochastic Algorithms for Nonlinearly Constrained Optimization

I will discuss the interesting features that drive the convergence guarantees for a set of adaptive stochastic algorithms that my collaborators and I have proposed for solving nonlinearly constrained optimization problems. These algorithms are of the sequential quadratic optimization and interior-point varieties, and they operate in the fully stochastic regime in which we prove convergence-in-expectation and almost-sure-convergence guarantees.

Frank E. Curtis
Industrial and Systems Engineering
Lehigh University
frank.e.curtis@gmail.com

Michael O'Neill
Lehigh University
mikeoneill@unc.edu

Michael O'Neill
University of North Carolina at Chapel Hill
mikeoneill@unc.edu

Daniel Robinson
Lehigh University
daniel.p.robinson@gmail.com

Baoyu Zhou
University of Chicago
zbaoyu@uchicago.edu

MS306

Adaptive Finite-Difference Interval Estimation for

Noisy Derivative-Free Optimization

A common approach for minimizing a smooth nonlinear function is to employ finite-difference approximations to the gradient. While this can be easily performed when no error is present within the function evaluations, when the function is noisy, the optimal choice requires information about the noise level and higher-order derivatives of the function, which is often unavailable. Given the noise level of the function, we propose a bisection search for finding a finite-difference interval for any finite-difference scheme that balances the truncation error, which arises from the error in the Taylor series approximation, and the measurement error, which results from noise in the function evaluation. Our procedure produces reliable estimates of the finite-difference interval at low cost without explicitly approximating higher-order derivatives. We show its numerical reliability and accuracy on a set of test problems. When combined with L-BFGS, we obtain a robust method for minimizing noisy black-box functions, as illustrated on a subset of unconstrained CUTEst problems with synthetically added noise.

Hao-Jun Shi
Meta
hjmshi@meta.com

Yuchen Xie, Melody Xuan
Northwestern University
yxie@u.northwestern.edu, qxuan@u.northwestern.edu

Jorge Nocedal
Department of Industrial Engineering and Management Sciences
Northwestern University
j-nocedal@northwestern.edu

MS307

An Efficient Proximal Sparse Smoothing Newton Method with Low Complexity for Wasserstein Barycenter Problems

The Wasserstein barycenter problem has many important applications in statistics, machine learning, and image processing. When the supports of the barycenter are prespecified, we can model it as a linear programming (LP) problem, which is highly degenerate. The existing state-of-the-art solvers, such as Gurobi, are challenging to solve this LP when the problem scale is large. In this talk, we propose a proximal sparse smoothing Newton (PSSN) method to solve this LP. We discuss the global convergence and local convergence rate of PSSN under some mild conditions. A key advantage of PSSN is that it can make full use of the data sparsity and the solution sparsity. As a result, the PSSN enjoys a low computational complexity per iteration for solving the Wasserstein barycenter problem. Extensive numerical results on both the synthetic and real data sets demonstrate the superior performance of the PSSN for solving large-scale Wasserstein barycenter problems.

Guojun Zhang
The Hong Kong Polytechnic University
guojun.zhang@connect.polyu.hk

MS308

Weighted Trust-Region Methods with Hessian Estimates

For unconstrained optimization problems in which only gradients of the objective function are available, we develop a trust-region method that updates a symmetric factorization of the Hessian approximation. Typically, the main computational challenge in trust-region methods is the solution of the trust-region subproblem, to determine the next search direction. In order to overcome this challenge, the factored form of the Hessian approximation can be used to define the trust-region subproblem in terms of weighted norms, which significantly reduce the computational cost of a subproblem. We develop two solvers based on a weighted ℓ_2 and a weighted ℓ_∞ norm. The methods are tested on a vast set of unconstrained optimization problems from CUTEst, on which they are competitive with state-of-the-art line-search implementations.

Johannes J. Brust
University of California, San Diego
Mathematics Department
jjbrust@ucsd.edu

Philip E. Gill
University of California, San Diego
Department of Mathematics
pgill@ucsd.edu

MS308

Methods for Bound-Constrained Nonsmooth Regularized Optimization

We introduce a variant of the proximal gradient method in which the quadratic term is diagonal but may be indefinite, and is safeguarded by a trust region. We derive the method as a special case of the proximal quasi-Newton trust-region method of Aravkin, Baraldi and Orban (2022). In certain cases, we provide closed-form solution of the step computation so that no subproblem solver is required. Our analysis expands upon that of Aravkin, Baraldi and Orban (2022) in two ways: (i) we generalize the trust-region approach to problems with bound constraints, and (ii) we provide a worst-case iteration complexity result in situations where the Hessian approximation may grow unbounded. We also discuss a primal-dual interior-point method for such problems, its practical implementation and worst-case iteration complexity. We provide efficient open-source implementations of our methods in the Julia language, in which Hessians approximations are given by diagonal quasi-Newton updates. We illustrate the performance of our methods on unconstrained and bound-constrained problems.

Dominique Orban
GERAD and Dept. Mathematics and Industrial Engineering
Ecole Polytechnique de Montreal

dominique.orban@gerad.ca

MS308

Recent Developments in Quasi-Newton Methods for Numerical Optimization

Quasi-Newton methods form the basis of many effective methods for unconstrained and constrained optimization. Quasi-Newton methods require only the first-derivatives of the problem to be provided and update an estimate of the Hessian matrix of second derivatives to reflect new approximate curvature information found during each iteration. In the years following the publication of the Davidon-Fletcher-Powell (DFP) method in 1963 the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update emerged as the best update formula for use in unconstrained minimization. More recently, a number of quasi-Newton methods have been proposed that are intended to improve on the efficiency and reliability of the BFGS method. Unfortunately, there is no known analytical means of determining the relative performance of these methods on a general nonlinear function, and there is no accepted standard set of test problems that may be used to verify that results reported in the literature are comparable. In this talk we will discuss ongoing work to provide a thorough derivation, implementation, and numerical comparison of these methods in a systematic and consistent way. We will look in detail at several modifications, discuss their relative benefits, and review relevant numerical results.

Jeb Runnoe
University of California, San Diego, U.S.
jrunnoe@ucsd.edu

Philip E. Gill
University of California, San Diego
Department of Mathematics
pgill@ucsd.edu

MS309

Cluster Explanation Via Polyhedral Descriptions

Clustering is an unsupervised learning problem that aims to partition unlabelled data points into groups with similar features. Traditional clustering algorithms provide limited insight into the groups they find as their main focus is accuracy and not the interpretability of the group assignments. This has spurred a recent line of work on explainable machine learning for clustering. In this paper we focus on the cluster description problem where, given a dataset and its partition into clusters, the task is to explain the clusters. We introduce a new approach to explain clusters by constructing polyhedra around each cluster while minimizing either the complexity of the resulting polyhedra or the number of features used in the description. We formulate the cluster description problem as an integer program and present a column generation approach to search over an exponential number of candidate half-spaces that can be used to build the polyhedra. To deal with large datasets, we introduce a novel grouping scheme that first forms smaller groups of data points and then builds the polyhedra around the grouped data, a strategy which outperforms simply sub-sampling data. Compared to state of the art cluster description algorithms, our approach is able to achieve competitive interpretability with improved description accuracy.

Connor Lawless
Cornell University
cal379@cornell.edu

MS309

Employing Pairwise Similarities in Machine Learning, and Its Resilience to Noise: Hochbaums Normalized Cut with Confidence Weights

We present a new technique for addressing biased or noisy training data. The method adapts HNC (Hochbaums Normalized Cut) which is a semi-supervised classification approach that relies on pairwise similarities between samples and, in that sense, is less reliant on labels. As a machine learning technique, HNC combines the effect of pairwise similarities between all samples, including the labeled samples, with constraining the labeled data to belong to the labeled class. In addition, HNC is solved, or determine a classification, using a very efficient parametric minimum cut procedure. HNC was shown to be competitive with many well-known classification methods. The new method introduced here is *Confidence HNC*, which is HNC where the labeled samples belong to their label class with limited confidence level. It permits a labeled sample to be classified differently from its label if the effect of similarity to other samples, whether labeled or unlabeled, overrides the penalty of such misclassification. In an extensive experimental study we compare Confidence HNC with leading algorithms for coping with noisy labels, for both real and synthetic data sets. We evaluate the classification accuracy of the method as well as its capability to detect noise in comparison to a variant of k-nearest neighbor classifier, a deep learning method Co-teaching+ and a recent noise filtering technique Confident learning, demonstrating the dominant performance of Confidence HNC.

Dorit Hochbaum
Industrial Engineering and Operations Research
University of California, Berkeley
hochbaum@ieor.berkeley.edu

Tor Nitayanont
UC Berkeley
torpong_nitayanont@berkeley.edu

MS309

Robust Counterfactual Explanations

Counterfactual explanations play an important role in detecting bias and improving the explainability of data-driven classification models. A counterfactual explanation (CE) is a minimal perturbed data point for which the decision of the model changes. Most of the existing methods can only provide one CE, which may not be achievable for the user. In this work we derive an iterative method to calculate robust CEs, i.e. CEs that remain valid even after the features are slightly perturbed. To this end, our method provides a whole region of CEs allowing the user to choose a suitable recourse to obtain a desired outcome. We use algorithmic ideas from robust optimization and prove convergence results for the most common machine learning methods including logistic regression, decision trees, random forests, and neural networks. Our experiments show that our method can efficiently generate globally optimal robust CEs for a variety of common data sets and classification models.

Donato Maragno, Tabea E. Röber
University of Amsterdam
d.maragno@uva.nl, t.e.rober@uva.nl

MS310

Exact Confidence Regions in Stochastic Optimization and Systems of Nonlinear Equations

We consider the question of constructing confidence regions in two contexts: (i) stochastic optimization problems with a smooth convex objective and observable first-order oracle; and (ii) stochastic nonlinear equations observable with a stochastic zeroth-order oracle. Relying on results from M-Estimation, we first show that the solutions to the sample-path problems associated with (i) and (ii) satisfy a *strong Bahadur Representation*. We then exploit such strong regularity to devise a simple but computationally intensive resampling technique called *batching* toward constructing exact confidence regions on the optimal value and solution to (i), and on the solution to (ii). The confidence region procedure is easy to implement as we illustrate with a number of examples.

Yi Chu
Purdue University, Department of Statistics
yi.chu317@gmail.com

Raghu Pasupathy
Purdue University
pasupath@purdue.edu

MS310

Stochastic Gradient Descent with Adaptive Data: Applications to Online Learning

We consider a stochastic optimization problem where the data is generated sequentially via Markov chains whose dynamics are determined by the decision variables and the loss function is defined with respect to the invariant measures of the Markov chains. Examples include finding the optimal price to maximize the long-run average revenue minus waiting cost in queueing systems; finding the optimal order-up-to level to minimize the long-run average holding cost plus backlog cost in inventory systems, etc. We study the convergence of the corresponding stochastic gradient descent algorithm with adaptive data. Our results establish an O(1/sqrt(T)) convergence rate when the loss function is convex and an O(1/T) convergence rate when the loss function is strongly convex under reasonably general conditions on the ergodicity of the Markov chains and smoothness of the loss function. We demonstrate how to apply our results to several online-learning problems in operations research and a policy gradient algorithm in reinforcement learning.

Jing Dong
Columbia University
jing.dong@gsb.columbia.edu

Xin Tong
National University of Singapore
xin.t.tong@nus.edu.sg

MS310

Stochastic Approximation for Estimating the Price of Stability in Stochastic Nash Games

The goal of this talk is to approximate the Price of Stability (PoS) for stochastic Nash games using stochastic approximation (SA) schemes. PoS is amongst the most popular metrics in game theory and provides an avenue for estimating the efficiency of Nash games. In particular, knowing the value of PoS can help significantly with designing efficient networked systems, including transportation networks and power market mechanisms. Motivated by the lack of efficient methods for computing the PoS, first we consider stochastic optimization problems with a nonsmooth and merely convex objective function and a merely monotone stochastic variational inequality (SVI) constraint. This problem appears in the numerator of the PoS. We develop a randomized block-coordinate stochastic extra-(sub) gradient method where we employ a novel iterative penalization scheme to account for the mapping of the SVI in each of the two gradient updates of the algorithm. Second, we develop an SA-based scheme for approximating the PoS and derive lower and upper bounds on the approximation error. To validate our theoretical findings, we provide some preliminary simulation results on a networked stochastic Nash Cournot competition.

Afrooz Jalilzadeh
University of Arizona
afrooz@arizona.edu

Farzad Yousefian
Rutgers University
farzad.yousefian@rutgers.edu

Mohammadjavad Ebrahimi
Rutgers University
School of Business
me586@scarletmail.rutgers.edu

MS311

Local stochastic gradient descent over Riemannian manifolds

This work considers minimizing a finite-sum objective function over a Riemannian manifold in a distributed setting in presence of parallel workers. We develop a local stochastic gradient descent (SGD) framework that performs local SGD updates in parallel on different workers and merge the sequence once in a while. Theoretical convergence results in different optimization and communication settings will be presented.

Sam Davanloo
Ohio State University

davanloo.1@osu.edu

MS311

Tangent Subspace Descent on Reductive Homogeneous Spaces

The tangent subspace descent method (TSD) extends the coordinate descent algorithm to manifold domains. The key insight underlying TSD is to draw an analogy between coordinate blocks in Euclidean space and tangent subspaces of a manifold. The core principle behind ensuring convergence of TSD for smooth functions is the appropriate choice of subspace at each iteration. In this talk, we will show that it is always possible to appropriately pick such subspaces on the broad class of manifolds known as reductive homogeneous spaces. This class includes Grassmannians, flag manifolds, and the positive definite manifold when endowed with an appropriate geometry. As a result of our developments we derive new and efficient methods for large-scale optimization on these domains.

David H. Gutman
Texas Tech University
david.gutman@ttu.edu

Daniel Grady
Wichita State University
daniel.grady@wichita.edu

Nam Ho-Nguyen
Discipline of Business Analytics
The University of Sydney
nam.ho-nguyen@sydney.edu.au

MS311

Fast convergence to non-isolated minima: four equivalent conditions for C2 functions

In optimization, non-isolated minima are detrimental to fast local convergence, in the sense that they are incompatible with a positive definite Hessian and many classical results break in their presence. In the literature there exist several alternative local conditions that are consistent with singular minima, notably the Polyak–Łojasiewicz inequality, the quadratic growth, and error bounds. We show that for C^2 cost functions all these conditions equivalently shape the local non-convex landscape of the cost function around a local minimum. Specifically, they imply a Morse–Bott property, meaning that the solution set is locally a submanifold, and the Hessian is positive definite in the normal spaces. Building on this, we study the question of super-linear convergence rates to singular minima. While the Newton method breaks down, we find that the adaptive regularization algorithm (using cubics) maintains a quadratic convergence rate, even with inexact subproblem solutions. The picture is less clear for the class of trust-region algorithms.

Quentin Rebjock
EPFL
quentin.rebjock@epfl.ch

Nicolas Boumal
EPFL
Institute of Mathematics
nicolas.boumal@epfl.ch

MS312

Lattice-Type Structure Optimization Using Component-Wise Reduced Order Models

Lattice-like structures can provide a combination of high stiffness with light weight that is useful in many applications, but a resolved finite element mesh of such structures results in a computationally expensive discretization. This computational expense may be particularly burdensome in many-query applications, such as optimization. We develop a stress-constrained topology optimization method for lattice-like structures that uses component-wise reduced order models as a cheap surrogate, providing accurate computation of stress fields while greatly reducing run time relative to a full order model. We demonstrate the ability of our method to produce large reductions in mass while respecting a constraint on the maximum stress in a pair of test problems. The ROM methodology provides a speedup of about 150x in forward solves compared to full order static condensation and provides a relative error of less than 5% in the relaxed stress.

Sean McBane
University of Texas at Austin
sean@ices.utexas.edu

Youngsoo Choi
Lawrence Livermore National Laboratory
choi15@llnl.gov

Karen E. Willcox
University of Texas at Austin, U.S.
kwillcox@utexas.edu

MS312

Learning the Regularization Parameter for Inverse Problems

One popular approach to solving linear inverse problems is via variational regularisation, wherein the data-fit is perturbed by a regulariser term, in order to encourage a priori information of the solution in the attained reconstruction. One must also choose the regularisation parameter, a value which determines how much the regulariser term contributes to the problem. A correct parameter value is critical, as one wants to overcome the ill-posedness of the problem, yet still attain a meaningful reconstruction. This work will cover bilevel learning, a framework in which one is able to learn appropriate parameter values via a machine learning approach.

Seb Scott
University of Bath
ss2767@bath.ac.uk

Matthias Ehrhardt
Institute for Mathematical Innovation
University of Bath
m.ehrhardt@bath.ac.uk

Silvia Gazzola
University of Bath
s.gazzola@bath.ac.uk

MS313

Model-Informed Generative Adversarial Network (MI-GAN) for Optimal Power Flow (opf) Problem

The optimal power flow (OPF) problem, as a critical component of power system operations, becomes increasingly difficult to solve due to the variability, intermittency, and unpredictability of renewable energy brought to the power system. Although traditional optimization techniques, such as stochastic and robust optimization approaches, could be used to address the OPF problem in the face of renewable energy uncertainty, their effectiveness in dealing with large-scale problems remains limited. As a result, deep learning techniques, such as neural networks, have recently been developed to improve computational efficiency in solving large-scale OPF problems. However, the feasibility and optimality of the solution may not be guaranteed. In this paper, we propose an optimization model-informed generative adversarial network (MI-GAN) framework to solve OPF under uncertainty. The main contributions are summarized into three aspects: (1) to ensure feasibility and improve optimality of generated solutions, three important layers are proposed: feasibility filter layer, comparison layer, and gradient-guided layer; (2) in the GAN-based framework, an efficient model-informed selector incorporating these three new layers is established; and (3) a new recursive iteration algorithm is also proposed to improve solution optimality. The numerical results on IEEE test systems show that the proposed method is very effective and promising.

Chenang Liu
Oklahoma State University
chenang.liu@okstate.edu

MS313

Expert-Guided Policy Optimization Via Metric-Aware Trust Region: Reinforcement Learning Approach for Whole-Building Hvac Control

We propose expert-guided Wasserstein policy optimization (WPO-IL), a reinforcement learning framework integrating human guidance into reinforcement learning (RL) for the efficient control of the heating, ventilation and air conditioning system in buildings. Trust-region methods based on Kullback-Leibler divergence are pervasively used to stabilize policy optimization in RL. In this paper, we exploit more flexible metrics and examine a natural extension of policy optimization with Wasserstein trust regions. Instead of restricting the policy to a parametric distribution class, we directly optimize the policy distribution and derive close-form policy updates based on the Lagrangian duality. A practical WPO-IL algorithm is proposed based on the closed-form policy updates and expert guardian in the loop. Experiments on HVAC control demonstrate the performance improvement of WPO-IL over state-of-art RL.

Jun Song
University of Washington
juns113@uw.edu

MS313

Frequency-Secured Stochastic Unit Commitment

The increasing share of renewables induces the reduction of rotational inertia and primary frequency response (PFR) capacities in the power system, for which unit commitment and PFR reserve should be carefully decided to guarantee frequency security. In this talk, we propose a frequency-secured stochastic unit commitment model. A continuous-time smooth trajectory is used to describe the frequency variation. The case studies demonstrate the effectiveness of the proposed method in guaranteeing frequency security.

Bo Zhou
Department of Industrial and Operations Engineering
University of Michigan at Ann Arbor
bozum@umich.edu

Ruiwei Jiang
Industrial and Operations Engineering
University of Michigan
ruiwei@umich.edu

Siqian Shen
Department of Industrial and Operations Engineering
University of Michigan
siqian@umich.edu

MS314

Optimal Transport Networks in Biology

TBA

Eleni Katifori
University of Pennsylvania
katifori@sas.upenn.edu

MS314

Optimization in Robot Navigation for Environmental Sensing

Autonomous mobile robots are versatile platforms in a wide range of environmental sensing applications. Their sensing efficacy and efficiency are heavily dependent on their navigation trajectories. Motion planning for mobile robots often leads to multi-objective optimization with non-convex, nonlinear constraints. For sensing in many environments, such as the ocean and space, the mobility of the robots is subject to strong environmental forces, further complicating the underlying sensing and optimization problems. Research progress in two representative applications will be reported. An energy-optimal path planning method will be introduced to allow sparse reconstruction of a fluid flow with mobile sensors. The optimization landscape encodes the underlying transport characteristics of the flow field and the information gain of candidate trajectories for flow sensing. For an Earth-sensing satellite constellation, we will show how the optimization landscape of the planning problem rises from the interplay among sensing objectives, orbital mechanics, and inter-agent coordination.

Zhuoyuan Song
University of Hawai'i at Manoa
zsong@hawaii.edu

MS314

Data Driven Exploration of Planetary Surfaces under Constrained Trajectories

Some environments, like unvisited planetary surfaces or regions of the ocean, have never been explored and thus, there is no a priori knowledge to leverage. Instead, a vehicle must sample upon arrival, process this data, and either send this information back to a teleoperator or autonomously decide where to go next. Teleoperation is suboptimal in that human intuition can be imprecise and cannot be mathematically guaranteed to yield an optimal result. Given a surface environment, a mobile agent will map the distribution of a scalar variable without any prior information, with a degree of confidence of model convergence, and while minimizing distance traveled. Gaussian processes have been popularly incorporated into active learning exploration strategies due to their ability to incorporate information into a parameter-free model and quantify a confidence metric with every model prediction. We evaluate the performance of this informative path planning algorithm in mapping an environment on three different surfaces: parabola, Townsend, and hydration value across a lunar crater from LAMP data. We quantify model variance, model root-mean-squared error, distance, and surface global minimum identification of the various methods in exploring a range of surfaces with no a priori knowledge. The results show that the information-driven methods significantly outperform naive exploration methods in minimizing model error and distance with potential of convergence.

Frances Zhu
University of Hawai'i at Manoa
zhuf@hawaii.edu

MS315

Best-Worst Probability Bounds in Pert Networks and Combinatorial Optimization

We compute the tightest possible bounds on the probability that the optimal value of a combinatorial optimization problem in maximization form with a random objective exceeds a given number, assuming only knowledge of the marginal distributions of the objective coefficient vector. The problem draws its motivations from project networks where the activity durations are random. The bounds proposed are valid across all joint distributions with the given marginals. We analyze the complexity of computing the bounds assuming discrete marginals and identify instances when the bounds are computable in polynomial time. For 0/1 V-polytope representations (where the extreme points are explicitly enumerated), we show that the tightest upper bound is weakly NP-hard to compute by providing a pseudopolynomial time algorithm. On the other hand, the tightest lower bound is shown to be strongly NP-hard to compute for 0/1 V-polytopes representations, by restricting attention to Bernoulli random variables. For compact 0/1 H-polytopes, in project management networks, we show that the tightest upper bound is weakly NP-hard to compute by providing a pseudopolynomial time algorithm. The results complement existing results in the literature for computing the probability with independent random variables.

Divya Padmanabhan
Indian Institute of Technology Goa
divya@iitgoa.ac.in

Selin D. Ahipasaoglu
University of Southampton
University of Southampton
sda1u20@soton.ac.uk

Arjun Ramachandra
Singapore University of Technology and Design
arjun_ramachandra@sutd.edu.sg

Karthik Natarajan
Engineering Systems and Design
Singapore University of Technology and Design
karthik_natarajan@sutd.edu.sg

MS315

Regret Minimization for Robust and Satisficing Monopolists

We study a robust monopoly pricing problem where a seller aspires to sell an item to a buyer. We assume that the seller, unaware of the buyers willingness to pay, ambitiously optimizes over a space of all individually rational and incentive compatible mechanisms with a regret-type objective criterion. Using robust optimization, a mechanism that minimizes the worst-case regret has been previously derived. In this paper, we alternatively adopt robust satisficing which minimizes the excess regret that is above the predetermined target level. We analytically show that the optimal mechanism involves the seller offering a menu of lotteries that charges a buyer-dependent participation fee and allocate the item with a buyer-dependent probability. Then, we consider two additional variants of the problem where the seller restricts her attention to a class of only deterministic posted price mechanisms and where the seller is relieved from specifying the target regret in advance. Finally, we determine a randomized posted price mechanism that is readily implementable and equivalent to the optimal mechanism, compute its statistics, and quantify the strength of the entailed randomization. Besides, we compare the proposed mechanism with a robust benchmark and numerically find that the former is predominantly superior to the latter in terms of the expected regret and the expected revenue when the coefficient of variation of the buyers value is under a hundred percent.

Napat Rujeerapaiboon
NUS
napat.rujeerapaiboon@nus.edu.sg

Yize Wei, Yilin Xue
National University of Singapore

yize.wei@u.nus.edu, yilin.xue@u.nus.edu

MS316

Time-Varying Semidefinite Programming: Path Following a Burer-Monteiro Factorization

We present an online algorithm for time-varying semidefinite programs (TV-SDPs), based on the tracking of the solution trajectory of a low-rank matrix factorization, also known as the Burer-Monteiro factorization, in a path-following procedure. There, a predictor-corrector algorithm solves a sequence of linearized systems. This requires the introduction of a horizontal space constraint to ensure the local injectivity of the low-rank factorization. The method produces a sequence of approximate solutions for the original TV-SDP problem, for which we show that they stay close to the optimal solution path if properly initialized. Numerical experiments for a time-varying Max-Cut SDP relaxation demonstrate the computational advantages of the proposed method for tracking TV-SDPs in terms of runtime compared to off-the-shelf interior point methods.

Antonio Bellon
CTU Prague
belloant@fel.cvut.cz

MS316

Low-Rank Tensor Decompositions with Nonlinear Constraints for Conserving Quantities of Interest in Numerical Simulation Data Modeling

Tensors, or multidimensional data, are a powerful means of representing relationships in multiway data and thus play an important role in many scientific disciplines, including combustion, earth systems, fusion, cyber security, quantum chemistry, and image processing. Many science applications are typically not interested in all data generated in a simulation, experiment, or observation but rather application-specific quantities of interest (QoIs) that are derived from this data. However, tensor decompositions applied to numerical simulation data for compression and reduced-order modeling often focus only on preserving global norms between data and model. We present an approach for computing goal-oriented low-rank tensor decompositions that incorporates problem-specific QoIs and general nonlinear constraints. Results for compression of direct numerical simulation data from a combustion application are presented to demonstrate the utility of this approach.

Daniel M. Dunlavy
Sandia National Laboratories
Computer Science and Informatics
dmdunla@sandia.gov

Hemanth Kolla
Sandia National Laboratories
hnkolla@sandia.gov

Eric Phipps
Sandia National Laboratories
Center for Computing Research
etphipp@sandia.gov

MS316

Maximum Relative Distance Between Real Rank-Two and Rank-One Tensors

It is shown that the relative distance in Frobenius norm of a real symmetric order-d tensor of rank two to its best rank-one approximation is upper bounded by $\sqrt{1-(1-1/d)d-1}$. This is achieved by determining the minimal possible ratio between spectral and Frobenius norm for symmetric tensors of border rank two, which equals $(1-1/d)(d-1)/2$. These bounds are also verified for arbitrary real rank-two tensors by reducing to the symmetric case.

Henrik Eisenmann
RWTH Aachen
eisenmann@igpm.rwth-aachen.de

MS317

Fitting Mathematical Models of Strobe Rockets

Pyrotechnic compositions called strobe produce oscillating flashes of light, and they are used in many fireworks applications. When such compositions are used as skyrocket fuel, the result is a projectile that instead emits a series of chugging noises, much like the hovering of a Vietnam-era helicopter. Our primary goal is to use parameterized differential equations to model the acoustics of strobe rockets as accurately as possible. We first apply both Newtons Law of Cooling and the classical heat equation to model the temperature distribution and then we use apply the ideal gas law to convert to the acoustic wave equation. We present here the mathematical models, a description of the optimization problem and algorithms we can use to find optimal parameters for fitting the model to real acoustic data, and some preliminary numerical results.

Mark A. Abramson
Utah Valley University
mark.abramson@uvu.edu

Wesley D. Smith
Brigham Young University
wsmith@sisna.com

MS317

COBYQA A Derivative-Free Trust-Region SQP Method for Nonlinearly Constrained Optimization

This talk introduces COBYQA, a derivative-free trust-region SQP method for general nonlinear optimization problems. The method builds the trust-region quadratic models using Powell's derivative-free symmetric Broyden update. An important feature of COBYQA is that it always respects bound constraints if any. COBYQA is competitive with NEWUOA and BOBYQA while being able to handle more general problems. On linearly constrained problems, COBYQA outperforms LINCOA if the problems also contain bound constraints that cannot be violated. Most importantly, COBYQA outperforms COBYLA on all types of problems, no matter whether bound constraints (if any) can be violated or not. COBYQA is implemented in Python and publicly available at https://www.cobyqa.com/. This is joint work with Zaikun Zhang.

Tom M. Ragonneau
The Hong Kong Polytechnic University
tom.ragonneau@polyu.edu.hk

Zaikun Zhang
Hong Kong Polytechnic University
zaikun.zhang@polyu.edu.hk

MS317

Using weak tail-bound probabilistic conditions for the analysis of stochastic derivative-free optimization methods

In this talk, we use tail bounds to define tailored probabilistic conditions for function estimation that ease the theoretical analysis of stochastic derivative-free optimization methods. We first outline a simplified convergence proof for a basic direct search scheme, using the new tail bound conditions. We then study the trade-off between noise, algorithm parameters and number of samples needed per iteration. Finally, we extend our analyses to a direct search method with nonmonotone line searches.

Francesco Rinaldi
University of Padova
rinaldi@math.unipd.it

Luis Nunes Vicente
Lehigh University
lnv218@lehigh.edu

Damiano Zeffiro
University of Padova
zeffiro@math.unipd.it

MS318

Extended Randomized Kaczmarz Method for Sparse Least Squares and Impulsive Noise Problems

The Extended Randomized Kaczmarz method is a well known iterative scheme which can find the Moore-Penrose inverse solution of a possibly inconsistent linear system and requires only one additional column of the system matrix in each iteration in comparison with the standard randomized Kaczmarz method. Also, the Sparse Randomized Kaczmarz method has been shown to converge linearly to a sparse solution of a consistent linear system. Here, we combine both ideas and propose an Extended Sparse Randomized Kaczmarz method. We show linear expected convergence to a sparse least squares solution in the sense that an extended variant of the regularized basis pursuit problem is solved. Moreover, we generalize the additional step in the method and prove convergence to a more abstract optimization problem. We demonstrate numerically that our method can find sparse least squares solutions of real and complex systems if the noise is concentrated in the complement of the range of the system matrix and that our generalization can handle impulsive noise.

Dirk Lorenz
TU Braunschweig
d.lorenz@tu-braunschweig.de

Schöpfer Frank
Uni Oldenburg
frank.schoepfer@uni-oldenburg.de

Maximilian Winkler, Lionel Tondji
TU Braunschweig
maximilian.winkler@tu-braunschweig.de, l.ngoupeyou-tondji@tu-braunschweig.de

MS318

Random Descent and Adjoint Sampling for Least Squares Problems

The Landweber iteration $v^{k+1} = v^k - \tau A^T(Av^k - b)$ is a well known algorithm to compute solutions of linear systems $Av = b$. It involves both the application of the linear operator A and its adjoint A^T in each iteration. Hence, further knowledge about the linear operator A is needed. In this talk we are interested in the setting, where the linear operator is not given as a matrix, but only a black box model is available. Moreover, we work under the assumption that storing many vectors of either the input or the output size is not feasible due to memory constraints. Without further knowledge about the adjoint and no possibility to evaluate the application of A^T to any input, we propose to replace the adjoint by a simple stochastic process. With this process as an unbiased estimator we show that both the iterates and the residuals still converge in quadratic mean. Moreover we provide rates of convergence under appropriate conditions. However, it is also possible to interpret the Landweber iteration with an estimated adjoint as a descent step in a random direction. We provide optimal and adaptive stepsizes that do not only guarantee convergence of the residuals in this setting, but also lead to at linear convergence in quadratic mean. Lastly, we visualise the results with numerical experiments and compare the algorithm to other adjointfree methods.

Felix Schneppe
Technische Universität Braunschweig
f.schneppe@tu-braunschweig.de

MS318

Asynchronous Parallel Block-Coordinate Forward-Backward Algorithm

We are going to present our work on the convergence properties of a randomized block-coordinate descent algorithm for the minimization of a composite convex objective function, where the block- coordinates are updated in parallel, asynchronously and randomly according to an arbitrary probability distribution. In that work, we prove that the iterates generated by the algorithm form a stochastic quasi-Fejr sequence and thus converge almost surely to a minimizer of the objective function. Moreover, we prove a general sublinear rate of convergence in expectation for the function values and a linear rate of convergence in expectation under an error bound condition of Tseng type. Under the same condition strong convergence of the iterates is provided as well as their linear convergence rate.

Cheik Traoré
Università degli Studi di Genova, Italy
cheik.traore@edu.unige.it

MS319

Wasserstein Gradient Flows with Input Convex Neural Networks

Gradient flows are a powerful tool for optimizing functionals in general metric spaces, including the space of probabilities endowed with the Wasserstein metric. A typical approach to solving this optimization problem relies on its connection to the dynamic formulation of optimal transport and the celebrated Jordan-Kinderlehrer-Otto (JKO) scheme. However, this formulation involves optimization over convex functions, which is challenging, especially in high dimensions. In this work, we propose an approach that relies on the recently introduced input-convex neural networks (ICNN) to parametrize the space of convex functions in order to approximate the JKO scheme, as well as in designing functionals over measures that enjoy convergence guarantees. We derive a computationally efficient implementation of this JKO-ICNN framework and experimentally demonstrate its feasibility and validity in approximating solutions of low-dimensional partial differential equations with known solutions. We also demonstrate its viability in high-dimensional applications through an experiment in controlled generation for molecular discovery. Joint work with David Alvarez-Melis and Yair Schiff; Link: https://arxiv.org/abs/2106.00774

Youssef Mroueh
IBM Research
mroueh@us.ibm.com

MS319

Variational Wasserstein Gradient Flow

Wasserstein gradient flow has emerged as a promising approach to solve optimization problems over the space of probability distributions. A recent trend is to use the well-known JKO scheme in combination with input convex neural networks to numerically implement the proximal step. The most challenging step, in this setup, is to evaluate functions involving density explicitly, such as entropy, in terms of samples. In this talk, I discuss our approach to solve this problem that builds on the recent works with a slight but crucial difference: we propose to utilize a variational formulation of the objective function formulated as maximization over a parametric class of functions. Theoretically, the proposed variational formulation allows the construction of gradient flows directly for empirical distributions with a well-defined and meaningful objective function. Computationally, this approach replaces the computationally expensive step in existing methods, to handle objective functions involving density, with inner loop updates that only require a small batch of samples and scale well with the dimension. The performance and scalability of the proposed method are illustrated with the aid of several numerical experiments involving high-dimensional synthetic and real datasets.

Jiaojiao Fan, Qinsh Zhang
Georgia Institute of Technology
jiaojiao-fan@gatech.edu, qzhang419@gatech.edu

Amirhossein Taghvaei
University of Washington Seattle
amirtag@uw.edu

Yongxin Chen
Georgia tech
yongchen@gatech.edu

MS319

Optimal Neural Network Approximation of Wasserstein Gradient Direction via Convex Optimization

The computation of Wasserstein gradient direction is essential for poste- rior sampling problems and scientific computing. The approximation of the Wasserstein gradient with finite samples requires solving a variational problem. We study the vari- ational problem in the family of two-layer networks with squared-ReLU activations, towards which we derive a semi-definite programming (SDP) relaxation. This SDP can be viewed as an approximation of the Wasserstein gradient in a broader function family including two-layer networks. By solving the convex SDP, we obtain the optimal ap- proximation of the Wasserstein gradient direction in this class of functions. Numerical experiments including PDE-constrained Bayesian inference and parameter estimation in COVID-19 modeling demonstrate the effectiveness of the proposed method.

Yifei Wang
Stanford University, U.S
wangyf18@stanford.edu

MS320

An Adaptive Lagrangian-Based Scheme for Nonconvex Composite Optimization

This talk introduces an adaptive augmented Lagrangian-based method to address the comprehensive class of nonsmooth nonconvex models with a nonlinear functional composite structure in the objective. The method uses an adaptive mechanism for the update of the feasibility penalizing elements, essentially turning the multiplier type method into a simple alternating minimization procedure based on the augmented Lagrangian function, from some iteration onward. This allows to avoid the restrictive, and until now, mandatory, surjectivity-type assumptions on the model, while achieving theoretical guarantees comprising iteration complexity and criticality of the limit point.

Nadav Hallak
Technion - Israel Institute of Technology
Israel

ndvhllk@technion.ac.il

MS320

Adaptive Proximal Gradient Method

In this talk, I will present a new way to make the classical proximal gradient method fully adaptive and at the same time without increasing the iteration cost. We don't need any additional assumptions and even relax the assumption of global Lipschitzness for the differentiable component to a local one. The stepsizes approximate the local curvature of the differentiable function and can increase from iteration to iteration. Compared to the adaptive gradient descent method, this is not a straightforward extension, but requires some new insight.

Yura Malitsky
Linkoping University
y.malitsky@gmail.com

MS320

Detecting Equivalence Between Iterative Algorithms for Optimization

When are two algorithms the same? How can we be sure a recently proposed algorithm is novel, and not a minor twist on an existing method? In this talk, we present a framework for reasoning about equivalence between a broad class of iterative algorithms, with a focus on algorithms designed for convex optimization. We propose several notions of what it means for two algorithms to be equivalent, and provide computationally tractable means to detect equivalence. Our main definition, oracle equivalence, states that two algorithms are equivalent if they result in the same sequence of calls to the function oracles (for suitable initialization). Borrowing from control theory, we use state-space realizations to represent algorithms and characterize algorithm equivalence via transfer functions. Our framework can also identify and characterize some algorithm transformations including permutations of the update equations, repetition of the iteration, and conjugation of some of the function oracles in the algorithm. A software package named Linnaeus implements the framework and makes it easy to find other iterative algorithms that are equivalent to an input algorithm. More broadly, this framework and software advances the goal of making mathematics searchable.

Madeleine Udell
Stanford University
udell@stanford.edu

Shipu Zhao
Amazon
sz533@cornell.edu

Laurent Lessard
Northeastern University
l.lessard@northeastern.edu

MS321

Second-order optimality conditions and calculus for second subderivatives

In this talk, we present second-order necessary and sufficient optimality condition for optimization problems with constraints $g(x) \in C$, where g is a sufficiently smooth mapping and C is a closed set. Clearly, this model covers many prominent mathematical programs. We explain the role of the curvature of set C and how it can be captured by various variational tools. Our ultimate goal is to handle difficult problem classes like bilevel programs or programs with constraints governed by quasi-variational inequalities. In these cases, set C can be written in a unified way as the graph of a certain set-valued mapping. The main challenge thus becomes the second-order variational analysis of sets possessing this structure, which is needed to express the curvature of C in terms of problem data. Concerning sufficient conditions, we capture the curvature of C via the second subderivative of the indication function associated with C. We propose a comprehensive calculus for second subderivatives (chain rule, marginal function rule). As it turns out, the chain rule yields lower estimates, desirable for sufficient conditions, for free. Similar estimates for the marginal function are valid under a rather mild inner calmness* assumption. This enables one to derive second-order sufficient conditions very easily and using minimal assumptions even for the aforementioned difficult problem classes.

Matus Benko
University of Viena, Austria
matusbenko@hotmail.com

Helmut Gfrerer
Johannes Kepler Universitaet Linz
Austria
gfrerer@numa.uni-linz.ac.at

Patrick Mehlitz
BTU Cottbus-Senftenberg
patrick.mehlitz@b-tu.de

Jane Ye
University of Victoria
Dept of Math & Statistics
janeye@uvic.ca

Jin Zhang
Southern University of Science and Technology, Shenzhen
zhangj9@sustech.edu.cn

Jinchuan Zhou
Shandong University of Technology
jinchuanzhou@163.com

MS321

Generalized Newton methods in nonsmooth optimization via second-order subdifferentials

This talk proposes and develops a new Newton-type algorithm to solve subdifferential inclusions defined by subgradients of extended-real-valued prox-regular functions. The proposed algorithm is formulated in terms of the second-order subdifferential of such functions that enjoys extensive calculus rules and can be efficiently computed for broad classes of extended-real-valued functions. Based on this and on metric regularity and subregularity properties of subgradient mappings, we establish verifiable conditions ensuring well-posedness of the proposed algorithm and its local superlinear convergence. The obtained results are also new for the class of equations defined by continuously differentiable functions with Lipschitzian gradients ($\mathcal{C}^{1,1}$ functions), which is the underlying case of our consideration. The developed algorithms for prox-regular functions and its extension to a structured class of composite functions are formulated in terms of proximal mappings and forward-backward envelopes. Besides numerous illustrative examples and comparison with known algorithms for $\mathcal{C}^{1,1}$ functions and generalized equations, the paper presents applications of the proposed algorithms to regularized least square problems arising in statistics, machine learning, and related disciplines

Pham Duy Khanh
Ho Chi Minh City University of Education
khanhpd@hcmue.edu.vn

MS321

Geometric Characterizations of Strong Solutions with Applications to Nuclear Norm Minimization Problems

In this talk, I will present new geometric characterizations of strong minimizers for optimization problems. These characterizations do not involve the calculations of the second subderivatives, the curvatures, or the sigma terms that are usually contained in classic results. This makes our characterizations look more computable and checkable. Consequently, we apply our results to analyzing strong minimizers of nuclear norm minimization problems. Some other new results on solution uniqueness of nuclear norm minimization problems will also be discussed.

Tran Nghia
Oakland University
nttran@oakland.edu

MS322

Approximating Convex Sets and Unbounded Convex Projection

In this work we consider a problem of approximating convex sets. A special case is the convex projection problem, where a feasible set is projected onto a lower dimensional subspace. The bounded cases of these problems were handled in [Shao, Zhao and Cong; Approximation of convex bodies by multiple objective optimization and an application in reachable sets; Optimization, 67(6): 783–796, 2018] and [Kovcov and Rudloff; Convex projection and convex multi-objective optimization; Journal of Global Optimization 83: 301-327, 2021]. There the sets were approximated by (implicitly or explicitly) transforming the problem into a multi-objective problem with one additional objective. In this work we propose an alternative algorithm that does not require this additional objective and, therefore, does not increases the dimension of the image space. Beyond this, we also address the unbounded case. Inspired by the approach of [Wagner, Ulus, Rudloff, Kovcov and Hey; Algorithms to solve unbounded convex vector optimization problems; 2022] for convex vector optimization problems, we provided an algorithm to approximate the recession cone of a convex set. We use this as a first phase of an algorithm to approximate an unbounded convex set. Our algorithms provide us both with inner and outer polyhedral approximations of the target set.

Gabriela Kovacova, Birgit Rudloff
Vienna University of Economics and Business
gabriela.kovacova@wu.ac.at, birgit.rudloff@wu.ac.at

MS322

Fast Multiobjective Gradient Methods with Nesterov Acceleration

In many applications in industry, economics or transport, optimizing several criteria is of interest. For example in transport, one wants to reach a destination as fast as possible with minimal power consumption. As this example shows, the different criteria one seeks to optimize are often contradictory. There is no design that is best for all objectives at once. This insight shifts the focus from finding a single optimal solution to a set of optimal tradeoffs the Pareto set.
A well-known algorithm to compute elements of the Pareto set is the multiobjective steepest descent method. While accelerated first-order methods are popular in singleobjective optimization, these methods are not studied sufficiently for multiobjective optimization problems. A fruitful approach to analyze gradient methods is to interpret them as discretizations of appropriate dynamical systems. The analysis of the continuous dynamics is often simpler and can later on be transferred to the discrete setting.
In this talk, we present contributions in the field of accelerated multiobjective gradient methods. We introduce an inertial gradient-like dynamical system in the multiobjective setting, whose trajectories converge to Pareto optimal solutions. The discretization of this system leads to an accelerated multiobjective algorithm that generates sequences which converge to Pareto optimal solutions. We will discuss the convergence behaviour of this algorithm and present numerical examples.

Konstantin Sonntag, Sebastian Peitz
Paderborn University
konstantin.sonntag@upb.de, sebastian.peitz@upb.de

MS322

A Norm Minimization-Based Convex Vector Optimization Algorithm with Convergence Analysis

We propose an algorithm for solving bounded convex vector optimization problems. The algorithm is based on solving norm-minimizing scalarizations. It starts with an initial outer approximation to the upper image and iterates by computing finer outer approximations. The novelty of this work is to provide the convergence rate of one such algorithm dealing with convex vector optimization problems. To obtain this result, we modify the norm-minimizing scalarization using an additional constraint that enables working with a compact set in the image space. We show that the sequence of polyhedral approximations obtained from the algorithm is a Hausdorff sequence (H-sequence) and provide a lower bound on the convergence rate of the algorithm.

Firdevs Ulus, Cagin Ararat, Muhammad Umer
Bilkent University
firdevs@bilkent.edu.tr, cararat@bilkent.edu.tr, muhammad.umer@bilkent.edu.tr

MS323

Private Federated Learning with Superquantile Aggregation for Heterogeneous Data

We present a federated learning framework that is designed to robustly deliver good predictive performance across individual clients with heterogeneous data. The proposed approach hinges upon a superquantile-based learning objective that captures the tail statistics of the error distribution over heterogeneous clients. We present a stochastic training algorithm which interleaves differentially private client filtering with federated averaging steps. We prove finite time convergence guarantees for the algorithm: $1/\sqrt{T}$ in the nonconvex case in T communication rounds and $\exp(-T/\kappa^{3/2}) + \kappa/T$ in the strongly convex case with local condition number κ. Experimental results on benchmark datasets for federated learning demonstrate that our approach is competitive with classical ones in terms of average error and outperforms them in terms of tail statistics of the error.

Krishna Pillutla
University of Washington
pillutla@cs.washington.edu

Yassine Laguel
Rutgers Business School
yassine.laguel@rutgers.edu

Jérôme Malick
University of Grenoble
jerome.malick@univ-grenoble-alpes.fr

Zaid Harchaoui
University of Washington
zaid@uw.edu

MS323

The Impact of Data Heterogeneity on Personalized Federated Learning

In many federated learning applications, learning one personalized model per client (i.e. local data distribution) can greatly improve performance and provide a more personalized experience to end users. However, such a personalization requires to intelligently take advantage of clients whose data distributions are relatively similar. To what extent can a client use another client's data to improve its performance? In this presentation, I will discuss the impact of data heterogeneity on personalized federated learning, and show that a natural notion of distance between the local data distributions arises when analyzing this problem from the lens of optimization theory. More precisely, I will provide matching lower and upper sample complexities for personalized federated learning, and show that a simple gradient filtering approach is optimal is a worst-case minimax sense. The optimal sample complexity will depend on a class of distances between data distributions known as Integral Probability Metrics (IPMs).

Kevin Scaman
INRIA
kevin.scaman@inria.fr

MS323

On the Unreasonable Effectiveness of Federated Averaging with Heterogeneous Data

Existing theory predicts that data heterogeneity will degrade the performance of the Federated Averaging (FedAvg) algorithm. However, in practice, the simple FedAvg algorithm converges very well. In this talk, we explain the seemingly unreasonable effectiveness of FedAvg that contradicts the previous theoretical predictions. We find that the key assumption of bounded gradient dissimilarity in previous theoretical analyses is too pessimistic to characterize data heterogeneity in practical applications. For a simple quadratic problem, we demonstrate there exist regimes where large gradient dissimilarity does not have any negative impact on the convergence of FedAvg. Motivated by this observation, we propose a new quantity average drift at optimum to measure the effects of data heterogeneity and explicitly use it to present a new theoretical analysis of FedAvg. We show that the average drift at optimum is nearly zero across many real-world federated training tasks, whereas the gradient dissimilarity can be large. And our new analysis suggests FedAvg can have identical convergence rates in homogeneous and heterogeneous data settings, and hence, leads to a better understanding of its empirical success.

Jianyu Wang
Meta AI Research
jianyuw1@andrew.cmu.edu

MS324

Optimization for Helping Vulnerable Groups

The Sustainable Development Goal 3 of the 2030 Agenda for Sustainable Development is to ensure healthy lives and promoting well-being for all at all ages. Therefore, ensure that every person has access to health and social services is one of the greatest challenges of today and in future. Optimization can help to provide solutions that reach universal high-quality health coverage at an affordable cost, in particular for vulnerable groups. In this talk, we present the optimization methodologies, models, algorithms and tools applied to Non-Profit Organizations (NPO) and Non-Governmental Organizations (NGO) in areas as health and social that had led to a positive impact on the society. We describe real applications on home health and social care logistics and personnel scheduling; circular economy of assistive technology; location of the primary health care centers or schools in high growing population areas; food security and distribution; and scheduling special transportation for persons with reduced mobility.

Helena Ramalhinho, Jesica de Armas
Pompeu Fabra University

Barcelona, Spain
helena.ramalhinho@upf.edu, jesica.dearmas@upf.edu

MS325

The Adaptive Fixed Step Stochastic Gradient Method

A folklore *adaptive fixed step* stochastic gradient (SG) algorithm works as follows. Execute SG with a fixed step $\eta > 0$ until the iterates approach stationarity, then re-execute SG starting from last iterate of the previous execution but with the revised fixed step $\eta \leftarrow \beta_0 \eta, \beta_0 < 1$ until the iterates again approach stationarity, and so on. This restart scheme is attractive because (i) it is simple to implement, and (ii) each execution of SG is known to attain stationarity *exponentially fast*. A particular version of this scheme was originally conceived by Georg Pflug in 1983, and enhancements, e.g., SASA (Lang et al., 2019), during 2018–2021 have reported surprisingly positive numerical results. We build on these variations by constructing a mechanism to rigorously detect when each execution of fixed step SG attains stationarity.

Raghu Pasupathy
Purdue University
pasupath@purdue.edu

MS325

Counter Examples for (Stochastic) Gradient Descent

Stochastic Gradient Descent (SGD) is a widely deployed algorithm for solving estimation problems that arise in statistics and learning. Accordingly, SGD has been analyzed from many perspectives to understand its behavior and to ensure its reliability, especially from a global convergence/consistency perspective. Unfortunately, we will show through simple examples that existing global convergence analyses make unrealistic deterministic assumptions, which result in incorrect conclusions or the utilization of inappropriate techniques. To be specific, counter to existing results, we will construct a deterministic example under realistic assumptions for which Gradient Descent (GD) will diverge catastrophically. Then, counter to a popular technique, we will provide a deterministic example for which approximating GD with continuous GD leads to incorrect conclusions about GD. Turning to stochastic assumptions, we show that existing stochastic assumptions are unrealistic for simple machine learning and statistics problems. Thus, we highlight that GD and SGD do not have an appropriate theory for learning problems. Finally, we provide results for the global convergence of GD and SGD that address these gap.

Vivak Patel
Department of Statistics, University of Wisconsin-Madison
vivak.patel@wisc.edu

MS325

On the Efficiency of a Wide Class of Derivative-Free Optimization Methods

Various research communities, ranging from machine learning to engineering design, have adopted distinct derivative-free optimization (DFO) methods. There has, however, not been a concerted effort to understand the relative merits of this disparate class of methods in a controlled setting. We report the performance of interpolation-based methods, finite differences and randomized methods for solving prompt optimization problems in natural language processing and noisy variants of the CUTEst test set.

Melody Xuan
Northwestern University
qxuan@u.northwestern.edu

Jorge Nocedal
Department of Industrial Engineering and Management Sciences
Northwestern University
j-nocedal@northwestern.edu

MS326

On Proximal Augmented Lagrangian Based Decomposition Methods for Dual Block-Angular Convex Composite Conic Programming Problems

We design inexact proximal augmented Lagrangian based decomposition methods for convex composite conic programming problems with dual block-angular structures. Our methods are particularly well suited for convex conic programming problems arising from stochastic programming models. The algorithmic framework is based on the application of the abstract inexact proximal ADMM framework developed in [Chen, Sun, Toh, Math. Prog. 161:237–270] to the dual of the target problem, as well as the application of the recently developed symmetric Gauss-Seidel decomposition theorem for solving a proximal multi-block convex composite quadratic programming problem. The key issues in our algorithmic design are firstly in designing appropriate proximal terms to decompose the dual variable blocks to make the subproblems easier to solve, and secondly to develop novel numerical schemes to solve the decomposed subproblems efficiently. Our inexact augmented Lagrangian based decomposition methods have guaranteed convergence. We present an application of the proposed algorithms to the doubly nonnegative relaxations of uncapacitated facility location problems, as well as to two-stage stochastic optimization problems. We conduct numerous numerical experiments to evaluate the performance of our method against state-of-the-art solvers such as Gurobi and MOSEK. Moreover, our proposed algorithms also compare favourably to the well-known progressive hedging algorithm.

Kuangyu Ding
National University of Singapore
e0444161@u.nus.edu

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

Xin-Yee Lam
National University of Singapore Department of Mathematics
matlxy@nus.edu.sg

MS326

Communication Topology of Decentralized Optimization

Massive data and computing resources are owned by massive parties in the society. Collaboratively utilizing these data and resources have significant economic values but obstructed by considerable obstacles, including lack of trust among distinct parties, risks of leaking privacy, and concerns on the collaboration quality. This talk discusses a promising solution via decentralized optimization, where parties are connected without a central moderator; everyone only communicates model weights/gradients with its direct neighbours. This talk will focus on how the communication topology would influence the behaviour of decentralized systems.

Fengxiang He
JD Explore Academy, JD.com Inc.
fengxiang.f.he@gmail.com

MS326

On Regularized Square-Root Regression Problems: Distributionally Robust Interpretation and Fast Computations

Square-root (loss) regularized models have recently become popular in linear regression due to their nice statistical properties. Moreover, some of these models can be interpreted as the distributionally robust optimization counterparts of the traditional least-squares regularized models. In this paper, we give a unified proof to show that any square-root regularized model whose penalty function being the sum of a simple norm and a seminorm can be interpreted as the distributionally robust optimization (DRO) formulation of the corresponding least-squares problem. In particular, the optimal transport cost in the DRO formulation is given by a certain dual form of the penalty. To solve the resulting square-root regularized model whose loss function and penalty function are both nonsmooth, we design a proximal point dual semismooth Newton algorithm and demonstrate its efficiency when the penalty is the sparse group Lasso penalty or the fused Lasso penalty. Extensive experiments demonstrate that our algorithm is highly efficient for solving the square-root sparse group Lasso problems and the square-root fused Lasso problems.

Chu Thi Mai Hong
National University of Singapore
hongtmchu@u.nus.edu

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

Yangjing Zhang
Chinese Academy of Sciences
yangjing.zhang@amss.ac.cn

MS327

Compressing Deep Neural Networks Can Improve Out-Of-Distribution Robustness

Successful adoption of deep learning (DL) in the wild requires models to be: (1) compact, (2) accurate, and (3) robust to distributional shifts. Unfortunately, efforts towards simultaneously meeting these requirements have mostly been unsuccessful. This raises an important question: Is the inability to create Compact, Accurate, and Robust Deep neural networks (CARDs) fundamental? To answer this question, we perform a large-scale analysis of popular model compression techniques which uncovers several intriguing patterns. Notably, in contrast to traditional pruning approaches (e.g., fine tuning and gradual magnitude pruning), we find that "lottery ticket-style" approaches can surprisingly be used to produce CARDs, including binary-weight CARDs. Specifically, we are able to create extremely compact CARDs that, compared to their larger counterparts, have similar test accuracy and matching (or better) robustness – simply by pruning and (optionally) quantizing. Leveraging the compactness of CARDs, we develop a simple domain-adaptive test-time ensembling approach (CARD-Decks) that uses a gating module to dynamically select appropriate CARDs from the CARD-Deck based on their spectral-similarity with test samples. The proposed approach builds a "winning hand' of CARDs that establishes a new state-of-the-art (on RobustBench) on CIFAR-10-C and CIFAR-100-C accuracies with better memory usage than non-compressed baselines.

James Diffenderfer
Lawrence Livermore National Laboratory
diffenderfer2@llnl.gov

MS327

L-SR1 Adaptive Regularization Using Cubics for Deep Neural Networks

Stochastic gradient descent is commonly used in the field of deep learning due to its computational efficiency and low-storage memory requirements. However, they involve fine-tuning many hyper-parameters. Quasi-Newton methods do not require such tuning and have been implemented effectively in both line-search and trust-region settings. In this work, we propose an adaptive regularized using cubics approach using limited-memory symmetric rank-one updates. We investigate the performance of this approach on auto encoders and feed-forward neural network models and compare with existing first-order stochastic methods.

Roummel F. Marcia, Aditya Ranganath, Mukesh Singhal
University of California, Merced
rmarcia@ucmerced.edu, aranganath@ucmerced.edu, msinghal@ucmerced.edu

MS327

Generalising Quasi-Newton Updates to Higher Orders

At the heart of all quasi-Newton methods is an update rule that enables us to gradually improve the Hessian approximation using the already available gradient evaluations. Theoretical results show that the global performance of optimization algorithms can be improved with higher-order derivatives. This motivates an investigation of generalizations of quasi-Newton update rules to obtain for example third derivatives (which are tensors) from Hessian evaluations. Our generalization is based on the observation that quasi-Newton updates are least-change updates satisfying the secant equation, with different methods using different norms to measure the size of the change. We present a full characterization for least-change updates in weighted Frobenius norms (satisfying an analogue of the secant equation) for derivatives of arbitrary order. Moreover, we establish convergence of the approximations to the true derivative under standard assumptions and explore the quality of the generated approximations in numerical experiments.

Karl Welzel
Mathematical Institute
University of Oxford
karl.welzel@maths.ox.ac.uk

Raphael A. Hauser
Mathematical Institute, Oxford
hauser@maths.ox.ac.uk

MS328

On the Convergence Properties of Infeasible Inexact Proximal Alternating Linearized Minimization

Proximal alternating linearized minimization method (PALM) suits well for solving block-structured optimization problems, which are ubiquitous in real applications. In the cases where subproblems do not have closed-form solutions, e.g., due to complex constraints, infeasible subsolvers are indispensable, giving rise to an infeasible inexact PALM (PALM-I). Numerous efforts have been devoted to analyzing feasible PALM, while little attention has been paid to PALM-I. The usage of PALM-I thus lacks theoretical guarantee. The essential difficulty of analyses consists in the objective value nonmonotonicity induced by the infeasibility. We study in the present work the convergence properties of PALM-I. In particular, we construct a surrogate sequence to surmount the nonmonotonicity issue and devise an implementable inexact criterion. Based upon these, we manage to establish the stationarity of any accumulation point and, moreover, show iterate convergence and provide new asymptotic convergence rates for PALM-I under the assumption of Lojasiewicz property. The prominent advantages of PALM-I on running time are corroborated via numerical experiments on problems arising from quantum physics and 3D anisotropic frictional contact.

Yukuan Hu
AMSS
Chinese Academy of Sciences
ykhu@lsec.cc.ac.cn

Xin Liu
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
liuxin@lsec.cc.ac.cn

MS328

Decentralized Optimization Over the Stiefel Manifold by An Approximate Augmented Lagrangian Function

In this talk, we focus on the decentralized optimization problem over the Stiefel manifold, which is defined on a connected network of d agents. The objective is an average of d local functions, and each function is privately held by an agent and encodes its data. The agents can only communicate with their neighbors in a collaborative effort to solve this problem. In existing methods, multiple rounds of communications are required to guarantee the convergence, giving rise to high communication costs. In contrast, this paper proposes a decentralized algorithm, called DESTINY, which only invokes a single round of communications per iteration. DESTINY combines gradient tracking techniques with a novel approximate augmented Lagrangian function. The global convergence to stationary points is rigorously established. Comprehensive numerical experiments demonstrate that DESTINY has a strong potential to deliver a cutting-edge performance in solving a variety of testing problems.

Lei Wang
AMSS
Chinese Academy of Sciences
wlkings@lsec.cc.ac.cn

MS329

Stochastic Optimization of Water Distribution Network Operation to Provide Power Grid Flexibility

Drinking water distribution networks can be operated as flexible loads in the power grid. We present a stochastic optimization framework that controls the supply pump operation in the water distribution network while satisfying power and water distribution network constraints. We solve for the schedule and real-time adjustments in the pump power consumption given power demand uncertainty. Analytical reformulations are employed, resulting in a tractable problem. Through case studies, we demonstrate the performance of our approach.

Anna Stuhlmacher, Johanna L. Mathieu
University of Michigan
akstuhl@umich.edu, jlmath@umich.edu

MS329

Pre-Trained Solution Methods for Unit Commitment

This study aims to improve the solution methods for the unit commitment problem, a short-term planning problem in the energy industry. In particular, we focus on Dantzig-Wolfe decomposition with a regularised column generation procedure. Firstly, initialisation methods of the column generation procedure based on machine learning techniques are studied. After offline training, for each unit commitment problem, the method outputs dual values which can be used to warmstart the solution method, leading to a significant saving of computational time. Secondly, the column generation procedure is extended to handle incremental generation of columns. Instead of generating columns for all the components in each iteration, our method generates a subset of them and updates the dual variable using the partially updated restricted master problem. Convergence analysis of the method is given under various conditions. These enhancements are tested on large-scale test instances.

Nagisa Sugishita
University of Edinburgh
nsugishi@ed.ac.uk

Andreas Grothey
School of Mathematics
University of Edinburgh
A.Grothey@ed.ac.uk

Ken McKinnon
University of Edinburgh
k.mckinnon@ed.ac.uk

MS329

Flexibility Management in Economic Dispatch with Dynamic Automatic Generation Control

As the installation of electronically interconnected renewable energy resources grows rapidly in power systems, system frequency maintenance and control become challenging problems to maintain the system reliability in bulk power systems. As two of the most important frequency control actions in the control centers of independent system operators (ISOs) and utilities, the interaction between Economic Dispatch (ED) and Automatic Generation Control (AGC) attracts more and more attention. In this talk, we propose a robust optimization based framework to measure the system flexibility by considering the interaction between two hierarchical processes (i.e., ED and AGC). We propose a cutting plane algorithm with the reformulation technique to obtain seven different indices of the system. In addition, we study the impacts of several system factors (i.e., the budget of operational cost, ramping capability, and transmission line capacity) and show numerically how these factors can influence the system flexibility.

Lei Fan
University of Houston
lfan8@central.uh.edu

Chaoyue Zhao
University of Washington
cyzhao@uw.edu

Guangyuan Zhang
RWE Renewables
zhangguangyuan1985@gmail.com

Qiuhua Huang
Utilidata
qiuhua.huang@ieee.org

MS330

Data-Driven Optimization Via the Scenario Approach: a Theory for the User Based on Conditional Risk Assessments

Data-driven methods are gaining increasing ground in the field of decision-making and - today more than ever in the past owing to the widespread diffusion of autonomous systems the reliable adoption of these methods requires to develop solid theories to keep control on the probability that a decision may underperform when applied to a new, out-of-sample, case (this probability is called the risk). In the context of the scenario approach, it has been proven that there exists an observable called the complexity (defined as the number of elements in the data set from which the solution can be reconstructed) that serves as a universal indicator to estimate the risk. In this presentation, we move a further step forward and consider the conditional probability distribution of the risk given the complexity. First, we provide evidence that this quantity is of great importance to the end user and yet - due to fundamental theoretical limitations - no meaningful conditional assessments can be made without calling for some additional knowledge on the uncertainty generation mechanism. We then introduce a new perspective that allows one to incorporate prior knowledge and establish the result that strong conditional results can be established under very mild priors. Besides its intrinsic epistemological value, this result allows for tight, a-posteriori, evaluations of the risk and improves the usability of data-driven optimization methods.

Simone Garatti
Politecnico di Milano
Dipartimento di Elettronica, Informazione e Bioingegneria
simone.garatti@polimi.it

Marco Campi
Università di Brescia
marco.campi@unibs.it

MS330

Robust Qualitative System Identification Through Robust Optimization

Safe implementation of any algorithm requires understanding its performance. In the context of linear system identification, one might know that system is stable, yet, due to noise corrupted data, traditional tools can lead to an identified model that fails to be stable. Such a mismatch in qualitative behaviour can be detrimental. In this talk we leverage the theory of moderate deviations to handle precisely this situation. Given a single trajectory of correlated data, we prove that the corresponding least squares estimator satisfies a moderate deviation principle for a rate function we compute explicitly. Next, we use this rate function to define a data-driven robust optimization problem that captures our desire of system identification subject to the underlying system being stable. Here, the ambiguity set we optimize over is not chosen but inferred directly from the problem. Although the optimization problem at hand is non-convex, we can construct an arbitrarily precise solution via a single linear algebra routine, plus, our method comes equipped with finite-sample statistical guarantees detailling how the probabilistic error between our identified model and the true unknown model decays in sampling size. Summarizing, we present the first statistical analysis of a practical qualitative system identification algorithm and we contribute to the area of optimization over stable matrices. We will also mention nonlinear hyperbolic systems and a link with Lyapunov exponents.

Wouter Jongeneel
EPFL
wouter.jongeneel@epfl.ch

Tobias Sutter
Swiss Federal Institute of Technology in Lausanne
tobias.sutter@uni-konstanz.de

Daniel Kuhn
École Polytechnique Fédérale de Lausanne
daniel.kuhn@epfl.ch

MS330

A Pareto-Dominance Principle for Data-Driven Optimization

We propose a statistically optimal approach to construct data-driven decisions for stochastic optimization problems. Fundamentally, a data-driven decision is simply a function that maps the available training data to a feasible action. It can always be expressed as the minimizer of a surrogate optimization model constructed from the data. The quality of a data-driven decision is measured by its out-of-sample risk. An additional quality measure is its out-of-sample disappointment, which we define as the probability that the out-of-sample risk exceeds the optimal value of the surrogate optimization model. The crux of datadriven optimization is that the data-generating probability measure is unknown. An ideal data-driven decision should therefore minimize the out-of-sample risk simultaneously with respect to every conceivable probability measure (and thus in particular with respect to the unknown true measure). Unfortunately, such ideal data-driven decisions are generally unavailable. This prompts us to seek data-driven decisions that minimize the out-of-sample risk subject to an upper bound on the out-of-sample disappointment - again simultaneously with respect to every conceivable probability measure. We prove that such Pareto-dominant data-driven decisions exist under conditions that allow for interesting applications.

Tobias Sutter
Swiss Federal Institute of Technology in Lausanne
tobias.sutter@uni-konstanz.de

Bart Van Parys
Massachusetts Institute of Technology
vanparys@mit.edu

Daniel Kuhn
École Polytechnique Fédérale de Lausanne
daniel.kuhn@epfl.ch

MS331

Neural Network Certification via Nonconvex Low-Rank Semidefinite Relaxations

It is well known that most neural networks are susceptible to adversarial attacks. In safety critical applications, people often robustly train their neural networks to be resilient towards potential adversarial attacks. However, without a certification that the model is rigorously robust against all possible attacks, the model could still be vulnerable to unseen further attacks. One promising technique for neural network certification is based on semidefinite programming, convex relaxation of the ReLU gate. Unfortunately, due to the large number of neurons in the network, most SDP solvers are not scalable for neural network certification. In this talk, we present a nonconvex relaxation for the ReLU gate, based on applying the nonconvex Burer-Monteiro approach to a semidefinite programming relaxation. Despite nonconvexity, we provide an algorithm to systematically escape saddle points, in order to arrive at a solution that is guaranteed to be near-globally optimal in moderate time and memory complexity. Our experimental results on small to medium size neural networks find that our nonconvex relaxation is significantly tighter than previous convex relaxations, and can consistently overcome the convex relaxation barrier that stymies the existing state-of-the-art.

Hong-Ming Chiu
University of Illinois, Urbana-Champaign
hmchiu2@illinois.edu

Richard Zhang
University of Illinois at Urbana-Champaign
ryz@illinois.edu

MS331

A Newton-Type Method for Sdp

This work proposes a squared smoothing Newton method via the Huber smoothing function for solving semide?nite programming problems (SDPs). We ?rst study the fundamental properties of the matrix-valued mapping de?ned upon the Huber function. Using these results and existing ones in the literature, we then conduct rigorous convergence analysis and establish convergence properties for the proposed algorithm. To evaluate the practical performance of the algorithm, we conduct extensive numerical experiments for solving various classes of SDPs.

Ling Liang
National University of Singapore
Singapore
liang.ling@u.nus.edu

MS331

First-Order Proximal Splitting Methods for SDP Formulation of Euclidean Distance Matrix Optimization

We propose first-order splitting methods for a family of matrix optimization problems involving Euclidean distance matrices (EDMs). In our framework, we first reformulate the EDM optimization problem as a semidefinite program (SDP). The equivalent SDP formulation is then solved via first-order splitting methods, in which the semidefinite conic constraint and linear constraints are handled separately. In many applications, the matrix size remains moderate while the number of linear constraints scales quickly. By using first-order splitting methods, the large-scale linear constraints are handled via simple linear algebra operations. The proposed framework is applied to two machine learning problems, and numerical results validate its scalability.

Chaorui Yao
University of California, Los Angeles
California, US
chaorui@g.ucla.edu

MS332

Positive k-spanning sets and their use in derivative free optimization

Positive spanning sets (PSSs) are sets of vectors spanning $\mathbb{R}^{\ltimes}$ using non-negative linear combinations. They can also be defined as sets properly approximating any given direction in the space and as such they prove very useful in derivative-free optimization algorithms. Optimization methods based on PSSs typically favor those with the best cosine measure, as this implies that the elements of the PSS are well spread in space. However, this metric does not fully account for the structure of the PSS: in particular, it does not reflect the spanning capabilities of a given subset of the PSS vectors. In this talk, we will focus on a particular subclass of PSSs, called positive k-spanning sets. A positive k-spanning set (PkSS) remains positively spanning when some of its elements are removed. After formally defining the positive k-spanning property, we will provide examples of PkSSs and present the classic way to construct them using arguments from polytope theory. We will present a new tool to quantify the quality of a PkSS, called the k-cosine measure. PkSSs are usually built using Gale diagrams. This very useful technique unfortunately does not give any kind of control over the k-cosine measure of the PkSS. To fix that problem, we will introduce an alternative method to build such sets giving a minimal bound on their k-cosine measure. Finally, we will compare the efficiency of DFO algorithms using these PkSSs with that of those using regular PSSs.

Sébastien Kerleau, Clement W. Royer
Université Paris Dauphine-PSL
sebastien.kerleau@dauphine.eu,
clement.royer@lamsade.dauphine.fr

Gabriel Jarry-Bolduc
University of British Columbia
gabjarry@gmail.com

Warren Hare
University of British Columbia, Okanagan Campus
warren.hare@ubc.ca

MS332

Coupling learning of hidden constraints with surrogate optimization

Real industrial studies often give rise to optimization problems involving time-consuming complex simulators that can produce some failures or instabilities for some input sets: for instance, convergence issues of the numerical scheme of partial derivative equations. The set of inputs corresponding to failures is often not known a priori and corresponds to a hidden constraint, also called a crash constraint. Since the observation of a simulation failure might be as costly as a feasible simulation, we seek to learn the feasible region to guide the optimization process in areas without simulation failures. Therefore, we propose to couple a surrogate-based optimization method with a Gaussian process classifier active learning method to learn the crash domain. A dedicated infill criterion is proposed to choose new simulations to, on one side, reduce the uncertainty on the feasible domain and, on the other side, explore promising feasible regions regarding the optimization objective.

Delphine Sinoquet, Morgane Menz
IFP Energies nouvelles
delphine.sinoquet@ifpen.fr, morgane.menz@ifpen.fr

Miguel Munoz Zuniga
IFP Energies Nouvelles
miguel.munoz-zuniga@ifpen.fr

MS332

Accelerating Randomized Adaptive Subspace Trust-Region Algorithms

This talk presents an algorithm for solving deterministic derivative-free optimization problems with thousands of decision variables. We employ a randomized dimension reduction technique based on Johnson-Lindenstrauss transforms and the trust-region framework of Cartis and Roberts. A key ingredient in our work is to extend the affine space within which the algorithm operates to include information learned in the course of the optimization. By adapting this space as well as its dimension, we obtain significant performance benefits on a class of large-scale test problems without sacrificing convergence guarantees.

Kwassi Joseph Dzahini
Argonne National Laboratory
kdzahini@anl.gov

Xiaoqian Liu
University of Texas MD Anderson Cancer Center
xiaoqian.liu1025@gmail.com

Stefan Wild
Argonne National Laboratory
LANS, Mathematics and Computer Science

wild@lbl.gov

MS333

Stochastic Variance-Reduced Majorization-Minimization Algorithms

We consider a class of nonconvex nonsmooth optimization problems whose objective is a sum of two functions. One is the average of a large number of differentiable component functions, while the other is a proper lower semicontinuous function. Such problems usually arise in machine learning, known as regularized empirical risk minimization. However, their nonconvexity and large-sum structure are challenging for designing new algorithms. Indeed, research dealing with these two features is still scarce. In this talk, we introduce three stochastic variance-reduced majorization-minimization (MM) algorithms combining the general MM principle with novel variance reduction techniques SAGA, SVRG, and SARAH. Under standard assumptions, we study the almost surely subsequential convergence of the generated sequence to a stationary point and prove that our algorithms can achieve the best-known complexity bounds in terms of the number of gradient evaluations. We demonstrate the effectiveness of our algorithms on sparse binary classification problems, sparse multi-class logistic regression, and neural networks using several well-known data sets.

Nhat Phan
University of Massachusetts Lowell
duynhat_phan@uml.edu

Sedi Bartz
University of Massachusetts Lowell
USA
sedi_bartz@uml.edu

Nilabja Guha, Hung M. Phan
University of Massachusetts Lowell
nilabja_guha@uml.edu, hung_phan@uml.edu

MS333

Inexact Reduced Gradient Methods for Nonconvex Smooth Optimization

The talk introduces new gradient-type methods with inexact gradient information for finding stationary points of nonconvex continuously differentiable functions on finite-dimensional spaces. A general scheme for inexact reduced gradient (IRG) methods with different stepsize selections is proposed to construct sequences of iterates with stationary accumulation points. Convergence results with convergence rates for the developed IRG methods are established under the Kurdyka-Lojasiewicz property. The conducted numerical experiments confirm the efficiency of the proposed algorithms.

Dat Tran
Wayne State University
tranbadat@wayne.edu

MS334

On the Hausdorff Distance Between a Pareto Set and Its Discretization

We discuss upper bounds on the Hausdorff distances between the efficient set and its discretization in the decision space, and between the Pareto set and its discretization in the objective space, in the context of bi-objective convex quadratic optimization on a compact feasible set. Our results imply that if t is the dispersion of the sampled points in the discretized feasible set, then the Hausdorff distances in both the decision space and the objective space are $O(\sqrt{t})$ as t decreases to zero.

Susan Hunter, Burla Ondes
Purdue University
susanhunter@purdue.edu, bondes@purdue.edu

MS334

Approximate Optimal Multiobjective Model Predictive Control

Model predictive control (MPC) is a control method that approximates the solution of an optimal control problem on an infinite time horizon. In each iteration step an optimal control problem on a finite horizon is solved and, thus, by shifting the horizon the approximate solution of the infinite horizon problem can be constructed from these solutions. A natural extension of MPC is to consider not only one, but multiple objectives. This leads to the formulation of a multiobjective optimal control problem for which we introduce a multiobjective MPC scheme. In this talk, we discuss the system theoretic properties of our proposed algorithm as well as, due to the optimization-based nature of the method, performance results for the objective functions. Finally, we analyze the quality of the solution in terms of ε-efficient solutions in a multiobjective sense. More precisely, we consider a weighted sum approach to make use of well-known single-objective MPC results and then, to obtain an error bound on our algorithm. Application examples and numerical simulations will illustrate our findings.

Lisa Krügel
Universität Bayreuth
lisa.kruegel@uni-bayreuth.de

Lars Grüne
Mathematical Institute
Universitaet Bayreuth
lars.gruene@uni-bayreuth.de

Gabriele Eichfelder
Institute of Mathematics
Technische Universität Ilmenau
Gabriele.Eichfelder@tu-ilmenau.de

MS334

Using Vector Optimization to Compute the Set of

All Nash Equilibria of a Game

Nash equilibria and Pareto optimization are two distinct concepts in multi-criteria decision making. It is well known that the two concepts do not coincide. However, in this work we show that it is possible to characterize the set of all Nash equilibria for any non-cooperative game as the set of all Pareto optimal solutions of a certain vector optimization problem. To accomplish this task, we enlarge the objective function and formulate a non-convex ordering cone under which Nash equilibria are Pareto efficient. These results are first discussed for shared constraint games, but extend easily to generalized Nash games, as well as to vector-valued games. The characterization holds for all non-cooperative games (non-convex, convex, linear). This characterization opens a new way of computing Nash equilibria. It allows to use algorithms from vector optimization to compute resp. to approximate the set of all Nash equilibria, which is in contrast to the classical fixed point iterations that find just a single Nash equilibrium. Examples are given, first in the linear case. Then, the convex case is considered and an algorithm is proposed that computes a subset of the set of epsilon-Nash equilibria such that it contains the set of all (true) Nash equilibria for convex games with either independent convex constraint sets for each player, or polyhedral joint constraints.

Birgit Rudloff
Vienna University of Economics and Business
birgit.rudloff@wu.ac.at

Zachary Feinstein
Stevens Institute of Technology
zfeinste@stevens.edu

Niklas Hey
Vienna University of Economics and Business
niklas.hey@wu.ac.at

MS335

Sequential Optimization of CVaR for MDPs: Existence and Computation of Optimal Policies

We study the problem of Conditional Value at Risk (CVaR) optimization for a finite-state Markov Decision Process (MDP) with total discounted costs and the reduction of this problem to a robust MDP. The CVaR optimization problems for finite and infinite-horizon MDPs can be reformulated as robust MDPs with compact state spaces. States in these robust MDPs are the original states of the problems augmented with tail risk levels, and for each problem the decision maker knows only the initial tail risk levels at the initial state and time. Thus, in order to find an optimal policy following this approach, the Decision Maker needs to solve a robust MDP with incomplete state observations because after the first move the Decision Maker observes states of the system, but the tail risk levels are unknown. This paper shows that for the CVaR optimization problem the corresponding Robust MDP can be solved by using the methods of convex analysis. This paper introduces the algorithm for computing and implementing an optimal CVaR policy by using the value function for the version of this Robust MDP with completely observable tail risk levels at all states. This algorithm and our main results are presented for a more general problem of optimization of a linear combination of CVaR and mean values for possibly different cost functions.

Rui Ding
Stony Brook University
Applied Mathematics and Statistics
rui.ding.1@stonybrook.edu

Eugene A. Feinberg
Stony Brook University
eugene.feinberg@stonybrook.edu

MS335

General Forms of Berge's Maximum Theorem with Applications to Games with Perfect Information and Distributionally Robust Optimization

Berges maximum theorem is the classic fact on the continuity of objective functions and upper semicontinuity of solution multifunction for parametric optimization, which is broadly used in robust optimization. The original formulation of Berge's maximum theorem assumes that the objective function is continuous and decision sets are compact. These assumptions do not hold in many applications, including many inventory control problems. We generalize Berge's maximum theorem in three directions: (i) the objective function may not be continuous, (ii) the sets of feasible decisions may not be compact, and (iii) necessary and sufficient conditions are provided. These results are applied to prove continuity of values and existence of solutions for distributionally robust inventory control problems with moment constraints.

David Kraemer, Eugene A. Feinberg
Stony Brook University
david.kraemer@stonybrook.edu,
eugene.feinberg@stonybrook.edu

Pavlo Kasyanov
National Technical University of Ukraine
kasyanov@i.ua

MS335

Optimizing Entropic Risk in MDPs

Prior work on safe Reinforcement Learning (RL) has studied risk-aversion to randomness in dynamics (aleatory) and to model uncertainty (epistemic) in isolation. We propose and analyze a new framework to jointly model the risk associated with epistemic and aleatory uncertainties in finite-horizon and discounted infinite-horizon MDPs. We call this framework that combines Risk-Averse and Soft-Robust methods RASR. We show that when the risk-aversion is defined using either EVaR or the entropic risk, the optimal policy in RASR can be computed efficiently using a new dynamic program formulation with a time-dependent risk level. As a result, the optimal risk-averse policies are deterministic but time-dependent, even in the infinite-horizon discounted setting. We also show that particular RASR objectives reduce to risk-averse RL with mean posterior transition probabilities. Our empirical results show that our new algorithms consistently mitigate uncertainty as measured by EVaR and other standard risk measures.

Marek Petrik
Department of Computer Science
University of New Hampshire
mpetrik@cs.unh.edu

Jia Lin Hau
University of New Hampshire
jialin.hau@unh.edu

Mohammad Ghavamzadeh
Google Research, US
ghavamza@google.com

MS336

Gathering Systems Design Optimization

Oil and gas field development often involves multiple assets with different reservoir characteristics, and the efficiency of gathering system design across these assets has a significant impact on overall liquid recovery and economic performance. Frequent challenges include how to efficiently allocate and inject gas back into reservoirs to boost liquid production, how to cost-effectively route oil and gas among assets and terminals, and how to reduce flaring. To resolve these challenges altogether, we develop a general framework where the subsurface dynamics are incorporated via surrogate modeling, and the gathering system design considers both integrated asset management and economic efficiency. We use mathematical modeling to maximize the overall economics consisting of liquid recovery revenue, network infrastructure expenses, etc., while satisfying external oil and gas demands that are sometimes uneconomical due to contracts and policies. When economics assumptions are unclear, we break down the objective into multiple steps such as to maximize liquid recovery while satisfying external demands, and then to minimize gathering network infrastructures used for routing. The most significant uncertainties in this problem are external demands and subsurface uncertainties, which can be addressed using stochastic programming. Alternatively, we analyze hundreds of cases and conduct statistical analysis, revealing solution patterns that are robust across all or most scenarios.

Yuzixuan Zhu, Jizhou Li, Yufen Shao
ExxonMobil
yuzixuan.zhu1@exxonmobil.com,
jizhou.li@exxonmobil.com, yufen.shao@exxonmobil.com

MS336

Integrated Network Optimization for Oil Production Using Mixed-Integer Piecewise-Linear Programming

The oil production system typically consists of a complex network of wells, flow lines and surface equipment. To increase the total oil output from the production system, produced gas can be injected into the wellbore via an artificial lift process called gas-lift. The role of injected gas (or gas-lift gas) in a gas-lift operation is to reduce the density of the fluid column, leading to a decrease in well bottom hole pressure and consequently increasing production. In offshore applications, gas-lift gas can also be injected into the subsea flow lines or risers to further increase production. In this paper, we propose a mathematical programming approach whereby piecewise-linear functions are used to approximate the nonlinear multiphase flow behavior in both wells and flow lines within a production network. The resulting optimization model is a mixed-integer linear program (MILP) that determines the optimal operation of the integrated production network, i.e. optimal gas lift injection, choke setting and well-separator routing. The proposed approach realizes both increased oil production (or uplift) and computational time improvement compared to traditional industry tools. Application to an offshore production system show up to 9% increase in oil production and an order magnitude reduction in computation time.

Keith Zorn
ExxonMobil
keith.p.zorn@exxonmobil.com

Emmanuel Ogbe
ExxonMobil, U.S.
emmanuel.ogbe@exxonmobil.com

Jose Santiago Rodriguez Osorio
ExxonMobil, Spring, TX
jose.s.osorio@exxonmobil.com

MS336

Multiperiod Blending With a Twist

Multiperiod blending problems are a natural extension of classic pooling problems and play a critical role in numerous industrial planning and scheduling applications. Traditional multiperiod blending problems model the physical blending of raw materials within a series of blending tanks. This work builds on this traditional setting in two ways. First, we present a model that includes an additional accounting network layer through which a raw materials nonphysical properties are blended and stored over time. Second, we showcase decomposition and discretization methods that are capable of providing high-quality solutions to these large-scale non-convex optimization problems in a reasonable amount of time.

Benjamin Sauk, Rick Zhu, Dimitri Papageorgiou, Youngdae Kim
ExxonMobil
benjamin.f.sauk@exxonmobil.com,
yu.zhu1@exxonmobil.com, dimitri.j.papageorgiou@exxonmobil.com, youngdae.kim@exxonmobil.com

MT1

On the Role of Circuits in Linear Programming

In this minitutorial, we will explain the role minimal linear dependencies of the constraint matrix, aka circuits, play in controlling the strongly polynomial complexity of linear programming. Using the circuit lens, we will show how to extend and improve upon works of Tardos '86 and Vavavis and Ye '96 on solving "well-conditioned' LPs in strongly-polynomial time. Our first focus will be to explain the properties of standard form LPs whose constraint matrix have bounded circuit imbalance measure: this is the largest ratio between the absolute values between the nonzero entries of a support minimal vector in the kernel of the matrix. We will cover proximity theorems for such LPs together with their algorithmic applications. Our second focus will be on interior point methods and the geometry of the central path. In this context, we will show how an IPM with “layered-least squares steps’ (LLS) make combinatorial progress along the central path in two different ways. Firstly, we will use circuit imbalances to show that it breaks up the path into a quadratic number of “short and curvy’ parts linked by ’long and straight’ segments. Secondly, we will explain how equipped with a subspace generalization of LLS steps, the IPM can be analyzed in a very fine-grained manner: the number of iterations it requires is within a polynomial factor of optimal with respect to ANY path-following method using a self-concordant barrier. In a rather unexpected connection, this analysis is enabled by comparing the geometry of barrier based central paths to certain shadow vertex simplex paths, which are canonical examples of circuit paths.

Bento Natura
Georgia Institute of Technology, U.S.
benton.natura@isye.gatech.edu

MT2

On the Role of Circuits in Linear Programming

In this minitutorial, we will explain the role minimal linear dependencies of the constraint matrix, aka circuits, play in controlling the strongly polynomial complexity of linear programming. Using the circuit lens, we will show how to extend and improve upon works of Tardos ’86 and Vavavis and Ye ’96 on solving “well-conditioned’ LPs in strongly-polynomial time. Our first focus will be to explain the properties of standard form LPs whose constraint matrix have bounded circuit imbalance measure: this is the largest ratio between the absolute values between the nonzero entries of a support minimal vector in the kernel of the matrix. We will cover proximity theorems for such LPs together with their algorithmic applications. Our second focus will be on interior point methods and the geometry of the central path. In this context, we will show how an IPM with “layered-least squares steps’ (LLS) make combinatorial progress along the central path in two different ways. Firstly, we will use circuit imbalances to show that it breaks up the path into a quadratic number of “short and curvy’ parts linked by ’long and straight’ segments. Secondly, we will explain how equipped with a subspace generalization of LLS steps, the IPM can be analyzed in a very fine-grained manner: the number of iterations it requires is within a polynomial factor of optimal with respect to ANY path-following method using a self-concordant barrier. In a rather unexpected connection, this analysis is enabled by comparing the geometry of barrier based central paths to certain shadow vertex simplex paths, which are canonical examples of circuit paths.

Bento Natura
Georgia Institute of Technology, U.S.
benton.natura@isye.gatech.edu

MT3

Optimization on Manifolds

To minimize a cost function f on a set M, it helps if f and M have mathematical structure that we can exploit for algorithmic gains. For example, if M happens to be a smooth manifold, we can benefit by using tools from differential geometry and Riemannian geometry. This observation dates back to 1972 (Luenberger). It extends many of our tools from unconstrained optimization to applications with orthonormality constraints, rank constraints, positivity, and also invariance (symmetries). In this tutorial, we visit the basics of differential and Riemannian geometry, tailored for an audience of optimizers who are not yet familiar with those topics. Then, we see how these tools help to design practical algorithms, and we discuss available resources to connect with your research. To learn more about this topic, see also the book ”An introduction to optimization on smooth manifolds”, freely available at https://www.nicolasboumal.net/book.

Nicolas Boumal
EPFL
Institute of Mathematics
nicolas.boumal@epfl.ch

MT4

Optimization on Manifolds

To minimize a cost function f on a set M, it helps if f and M have mathematical structure that we can exploit for algorithmic gains. For example, if M happens to be a smooth manifold, we can benefit by using tools from differential geometry and Riemannian geometry. This observation dates back to 1972 (Luenberger). It extends many of our tools from unconstrained optimization to applications with orthonormality constraints, rank constraints, positivity, and also invariance (symmetries). In this tutorial, we visit the basics of differential and Riemannian geometry, tailored for an audience of optimizers who are not yet familiar with those topics. Then, we see how these tools help to design practical algorithms, and we discuss available resources to connect with your research. To learn more about this topic, see also the book ”An introduction to optimization on smooth manifolds”, freely available at https://www.nicolasboumal.net/book.

Nicolas Boumal
EPFL
Institute of Mathematics
nicolas.boumal@epfl.ch

PP1

AWM Workshop: Multistage Stochastic Mixed-Integer Programming Using DSP

Multistage stochastic mixed-integer programs (MSMIPs) are highly-structured optimization problems that can be represented using scenario trees or scenario lattices. The block-diagonal structure of these models can then be exploited via decomposition of the problem into subproblems that can be solved in parallel, drastically increasing computational efficiency. We provide a framework by which to decompose and solve general MSMIPs using the Julia programming language and DSP (a parallel decomposition MIP solver), and demonstrate the efficacy of our approach using the air conditioner production problem as an example.

Rachael M. Alfant
Rice University
rma10@rice.edu

Kibaek Kim
Argonne National Laboratory
kimk@anl.gov

PP1

AWM Workshop: The Computational Performance of Iterated Linear Optimization

Semidefinite programming is frequently utilized as a relaxation for clustering problems, but its results are heavily dependent on the rounding step implemented to obtain a discrete clustering. We will assess the viability of combining semidefinite programming with a deterministic rounding approach called iterated linear optimization as an alternative clustering method to k-means and spectral clustering. This analysis will include comparisons of the efficiency and accuracy of the algorithms, as well as experiments on various datasets.

Teresa Chambers, Caroline Klivans
Brown University
teresa_chambers@brown.edu, caroline_klivans@brown.edu

PP1

AWM Workshop: Tropicalization of Graph Profiles for Some Classes of Trees

Many important problems in extremal combinatorics can be stated as inequalities of graph homomorphism numbers. For a fixed collection of graphs U, the *tropicalization of the graph profile of* $\mathcal{U}$ essentially records all valid binomial inequalities involving graph homomorphism numbers for graphs in $\mathcal{U}$. Building upon ideas and techniques described by Blekherman and Raymond in 2021, I present progress toward finding the tropicalization for some classes of trees.

Maria Dascalu, Annie Raymond
University of Massachusetts Amherst
mdascalu@umass.edu, raymond@math.umass.edu

PP1

AWM Workshop: Using Bayesian Spectral Reprojection to Resolve the Gibbs Phenomenon

Fourier reconstruction yields spectral accuracy for smooth and periodic functions, but produces the Gibbs phenomenon for non-periodic functions. Spectral reprojection resolves the Gibbs phenomenon by projecting the Fourier partial sum onto Gegenbauer polynomials. Numerical round-off error and the Runge phenomenon can both degrade the quality of the solution, however. Here we use a Bayesian approach to construct a posterior using the given Fourier data and a prior based on properties of orthogonal polynomials.

Tongtong Li
Dartmouth College
tongtong.li@dartmouth.edu

PP1

AWM Workshop: Data Analytics for Membrane Material Innovations

Membrane characterization provides essential information for the scale-up, design, and optimization of new separation systems. We proposed the Diafiltration Apparatus for high-Throughput Analysis(DATA) framework, which enables a 10-times reduction in the time necessary to characterize membranes by integrating experiments, dynamic modeling, and parameter estimation. Other tools from data analytics, including identifiability analysis, Fisher Information Matrix(FIM) based analysis, Model-Based Design of Experiments(MBDoE), and model selection, are applied in DATA successfully to facilitate membrane material innovations.

Xinhong Liu
Department of Chemical and Biomolecular Engineering
University of Notre Dame
xliu27@nd.edu

Alexander Dowling
University of Notre Dame
adowling@nd.edu

William Phillip
Department of Chemical and Biomolecular Engineering
University of Notre Dame
wphillip@nd.edu

PP1

AWM Workshop: Optimization of Control Free Time Sweeping Processes and Applications

This project addresses a free time optimal control problem of Mayer type for sweeping processes in which the perturbation is non-smooth. We develop a constructive finite-difference approximation procedure that allows us to establish necessary optimality conditions for the original continuous-time controlled sweeping process, and then show how these optimality conditions are applied to solve several applications in the real life.

Giovanni Colombo
University Padova
Dipartimento di Matematica
colombo@math.unipd.it

Boris Mordukhovich
Department of Mathematics,
Wayne State University
aa1086@wayne.edu

Nguyen-Truc-Dao Nguyen
University of Michigan
nntdao@umich.edu

Trang Nguyen
Wayne State University
ha2288@wayne.edu

PP1

AWM Workshop: Combined Modeling and Experimental Study of Shape Formation of the *Drosophila* Wing Imaginal Disc

Developmental mechanisms of tissue growth and shape formation are not well understood. Combined experiments and multi-scale computational model of the *Drosophila* wing imaginal disc will be used to show that nonhomogeneous actomyosin patterning defines the local basal curvature and nuclear positioning while cell proliferation enhances the local basal curvature. A coarse-grained stochastic model of actomyosin dynamics will be described to capture the directionality of actomyosin contractile forces and nuclear positioning during cell growth and division.

Jennifer Rangel Ambriz
University of California, Riverside
University of California, Riverside
jrang008@ucr.edu

Nilay Kumar
Department of Chemical and Biomolecular Engineering
University of Notre Dame
nkumar4@nd.edu

Kevin Tsai
Department of Mathematics
University of California, Riverside
ytsai003@ucr.edu

Mayesha Sahir Mim
Department of Chemical and Biomolecular Engineering
University of Notre Dame
mmim@nd.edu

Weitao Chen
University of California, Riverside
weitaoc@ucr.edu

Jeremiah Zartman
Department of Chemical and Biomolecular Engineering
University of Notre Dame
jzartman@nd.edu

Mark Alber
University of California, Riverside
malber@ucr.edu

PP1

AWM Workshop: Numerical Analysis of a Corrected Smagorinsky Model

The classical Smagorinsky model's solution is an approximation to a (resolved) mean velocity. Since it is an eddy viscosity model, it cannot represent a flow of energy from unresolved fluctuations to the (resolved) mean velocity. This model has recently been corrected to incorporate this flow and still be well-posed. Herein, we first develop some basic properties of the corrected model. Next, we perform a complete numerical analysis of two algorithms (Backward Euler and Crank-Nicolson with linear extrapolation) for its approximation. They are tested and proven to be effective. Turbulent flows strain resources, both memory and CPU speed. Variable Time Step Method of DAHLQUIST, LINIGER AND NEVANLINNA (DLN) has greater accuracy and allows larger timesteps which means it requires less memory and fewer FLOPS. DLN can also be implemented adaptively. Hence, we also consider DLN method for the time discretization in Corrected Smagorinsky Model. We test that DLN gives best result.

Farjana Siddiqua
University of Pittsburgh
fas41@pitt.edu

PP1

AWM Workshop: Multicellular Model for Morphogenesis of Pavement Cells in Plant Leaves

The leaf epidermal cells in plants have complex jigsaw-puzzle shapes. The formation of these shapes relies on mechanical, chemical, and cell-to-cell signals at different scales. Thus, these pavement cells are an interesting model system used to study the mechanisms involved in cell morphogenesis. To investigate these mechanisms, a multicellular model that utilizes the local level set method which incorporates biochemical dynamics on moving cell boundaries, cell-cell adhesion, and crosstalk between neighboring cells is used.

Alysha Toomey, Weitao Chen
University of California, Riverside
atoom001@ucr.edu, weitaoc@ucr.edu

PP1

AWM Workshop: Dynamics of a Multi-Layer Fluid System

A long-wave asymptotic model is used to analyze the linear stability of the Navier-Stokes equations. The impact of different parameters on the linear growth of the interfacial disturbances is investigated when disturbances are small. The growth rate of the disturbances is computed by solving an eigenvalue problem, which helps determine where the system is unstable to small amplitude disturbances. Nonlinear evolution equations are also derived to investigate the dynamics when the disturbances are not small.

Awa Traore
The University of Alabama
atraore@crimson.ua.edu

PP1

AWM Workshop: Data Assimilation for Quantum

Nv Diamond Spectroscopy

Nitrogen-vacancy (NV) defect centers in diamond have generated much interest for their uses in quantum information and sensing. Negatively charged centers are used for high spatial-resolution sensing and for quantum information. Despite the rapid NV applications development, our grasp of basic NV properties is incomplete, which is important to understand to fully exploit potential uses. In this work we construct a statistical model for NV spectroscopy and use it in synthetic experiments to solve inverse problems. Our principal application is to develop a primary sensor based on the NV diamond quantum optical properties. This is a significant challenge because the NV diamond structure is sensitive to temperature and pressure as well as magnetic and electric fields, First, using the Hamiltonian for the effects of local strain and the environmental variables, we identify the observable components based on the invertibility of various observation systems. Next, we observe the influence of temperature and pressure on the NV center by solving the Schrdinger Equation and computing the theoretical spectroscopy curve. We assume that the observed photon counts are Poisson random variables with rates proportional to the theoretical spectroscopy. Then, using the Maximum Likelihood Estimation we find the parameter values that maximize the likelihood. Last but not the least we determine the robustness of the model using sensitivity analysis.

Shraddha Verma
George Mason University
srajpal4@gmu.edu

PP1

AWM Workshop: Characterizing the Pareto Optimal Trade-off Between Model-Based Information Content and Measurements Cost

Model-Based Design of Experiments (MBDoE) is a powerful tool for building and validating mathematical models. It leverages science-based models to maximize information gain from experiments while minimizing time and resource costs by optimizing experimental conditions such as temperature, pressure, or batch time. In this work, we consider a related problem: measurement (sensor) selection. When not enough or unsuitable measurements are collected, uncertainty parameters in the science-based mathematical model may remain unidentifiable or MBDoE may suggest uninformative or infeasible optimal experiments. We proposed a novel convex optimization formulation to identify the best set of measurements that maximizes the experimental information content, quantified by the trace or determinant of the Fisher information matrix, subject to a measurement cost budget. This general framework identifies measurement campaigns for multi-response systems and supports arbitrary (positive semi-definite) variance and covariances between responses. We leveraged the continuous-effort design concept to retain a convex optimization problem regardless of model structure. The Pareto trade-off between information content and measurement budget is easily computed using the proposed optimization framework. The model sensitivity matrix can be provided directly by the user or computed by Pyomo.DoE.

Jialu Wang
University of Notre Dame
jwang44@nd.edu