IP1

Rigorous Guarantees for Nonconvex Optimization?

The classical complexity barrier in continuous optimization is between convex (solvable in polynomial time) and nonconvex (intractable in the worst-case setting). However, many problems of interest are not constructed adversarially, but instead arise from statistical models of scientific phenomena. Can we provide rigorous guarantees for such random ensembles of nonconvex optimization problems? In this talk, we survey various positive answers to this question, including optimal results for sparse regression with nonconvex penalties, direct approaches to low-rank matrix recovery, and non-local guarantees for the EM algorithm. All of these results involve natural weakenings of convexity that hold for various classes of nonconvex functions, thus shifting the barrier between tractable and intractable.

Martin Wainwright
University of California
Berkeley
wainwrig@stat.berkeley.edu

Sivaraman Balakrishnan
Carnegie Mellon University
siva@stat.cmu.edu

Yudong Chen
Cornell University
yudong.chen@cornell.edu

Po-Ling Loh
University of Wisconsin
loh@ece.wisc.edu

Bin Yu
University of California, Berkeley
binyu@stat.berkeley.edu

IP2

Innovation in Big Data Analytics

Risk and decision models and predictive analytics have long been cornerstones for advancement of business analytics in industrial, government, and military applications. In particular, multi-source data system modeling and big data analytics and technologies play an increasingly important role in modern business enterprise. Many problems arising in these domains can be formulated into mathematical models and can be analyzed using sophisticated optimization, decision analysis, and computational techniques. In this talk, we will share some of our successes in healthcare, defense, and service sector applications through innovation in predictive and big data analytics.

Eva K. Lee
Georgia Inst of Technology
Sch of Ind & Systems Eng
evakylee@isye.gatech.edu

IP3

Optimizing Today's Communication Networks

Optimization methods have long played a major role in the design and management of communication systems that we use today. In recent years, the communication networks have grown much larger in size and have to provide a multitude of services at increasingly higher rates. These changes have presented new challenges in the design and the management of communication networks. In this talk, we will illustrate the use of modern optimization methods in the provision of large scale heterogeneous communication networks to ensure energy efficiency and network scalability.

Zhi-Quan Luo
Univeristy of Minnesota
Minneapolis, Minnesota
luozq@ece.umn.edu

IP4

Subgradient Methods

Subgradient methods are the most basic of algorithms in convex optimization, applicable quite generally and having simple convergence proofs, modulo assumptions such as orthogonal projections onto convex sets being readily computable. The fundamentals regarding subgradient methods have changed little in decades, perhaps in part due a supposition by researchers that not much else can be said for algorithms relying on hardly any problem structure other than convexity and Lipschitz continuity. We explain that, to the contrary, further advances are possible. We display an elementary algorithmic device which removes the need for Lipschitz continuity, and that replaces orthogonal projections with line searches. Additionally, we give focus to using subgradient methods for the most elemental of general optimization problems, computing a feasible point.

James M. Renegar
Cornell University
School of ORIE
renegar@cornell.edu

IP5

Using Second-order Information in Training Large-scale Machine Learning Models

We will give a broad overview of the recent developments in using deterministic and stochastic second-order information to speed up optimization methods for problems arising in machine learning. Specifically, we will show how such methods tend to perform well in convex setting but often fail to provide improvement over simple methods, such as stochastic gradient descent, when applied to large-scale nonconvex deep learning models. We will discuss the difficulties faced by quasi-Newton methods that rely on stochastic first order information and Hessian-Free methods that use stochastic second order information.

Katya Scheinberg
Lehigh University
kas410@lehigh.edu

IP6

Using Local Measurements to Infer Global Network Properties

Networks are widely used to model relational data in many applications such as social sciences, biology, finance, and marketing. Analyses of these networks can reveal scientific insights, but formulations lead to NP-Hard problems. Heuristics are commonly used without guarantees on solution quality, as even some polynomialtime algorithms become impractical due to the sheer sizes of the problems. To achieve scalability, our works has focused on trying to infer global network properties based on distributions of local patterns, such as triangles or 4-cliques. Such local measurements are scalable and provide uniquely measurable properties of the graph. This talk will show examples of our approach to network modeling and analysis.

Ali Pinar
Sandia National Labs
apinar@sandia.gov

IP7

Recent Progress on Dual Decomposition for Stochastic Integer Programming

We will discuss two methods for efficiently computing the value of the Lagrangian Dual of a Stochastic Integer Program (SIP). First, we discuss a primal-dual algorithm, the well-known progressive hedging method, but unlike previous progressive hedging approaches for SIP, our algorithm can be shown to converge to the optimal Lagrangian dual value. The key improvement in the new algorithm is an inner loop of optimized linearization steps, similar to those taken in the classical Frank-Wolfe method. Second, we will discuss recent our work in improving the performance of classical stochastic (sub)gradient methods for SIP. Enhancements include both sub-sampling and asynchronous versions that run effectively on large-scale, distributed, heterogeneous computing platforms.

Jeffrey Linderoth
University of Wisconsin-Madison
linderoth@wisc.edu

Natashia Boland
Georgia Institute of Technology
natashia.boland@gmail.com

Jeffrey Christiansen
Mathematical Sciences,
School of Science, RMIT University
s3507717@student.rmit.edu.au

Andrew Eberhard
RMIT University
andy.eberhard@rmit.edu.au

Fabricio Oliveira
Mathematical Sciences,
School of Science, RMIT University
fabricio.oliveira@rmit.edu.au

Brian C. Dandurand
Argonne National Laboratory
brian.dandurand@rmit.edu.au

Cong Han Lim
?University of Wisconsin-Madison
conghan@cs.wisc.edu

James Luedtke
University of Wisconsin-Madison
Department of Industrial and Systems Engineering
jrluedt1@wisc.edu

Stephen J. Wright
University of Wisconsin-Madison
swright@cs.wisc.edu

SP1

SIAG/OPT Prize Lecture: Proximal Minimization Algorithms for Nonconvex and Nonsmooth Problems

We introduce a self-contained convergence analysis framework for first order methods in the setting of nonconvex and nonsmooth optimization problems. Our approach builds on the powerful Kurdyka-Lojasiewcz property. It allows for analyzing, under mild assumptions, various classes of nonconvex and nonsmooth problems with semi-algebraic data, a property shared by many optimization models arising in various fundamental data science paradigms. We illustrate our results by deriving a new and simple proximal alternating linearized minimization algorithm (PALM). The versatility of PALM permits to exploit structures and data information relevant to important applications and paves the ways to the derivation of other interesting algorithmic variants.

Shoham Sabach
Technion–Israel Institute of Technology
Faculty of Industrial Engineering and Management
ssabach@gmail.com

Jérôme Bolte
Université Toulouse 1
jerome.bolte@univ-tlse1.fr

Marc Teboulle
Tel Aviv University
teboulle@post.tau.ac.il

CP1

Robust Optimization in Adaptive Radiation Therapy Planning

Uncertainties in radiation therapy can lead to non-negligible deficiencies in treatment quality which can be addressed by robust optimization. Adaptive radiation therapy allows for adaptation during the course of the treatment to uncertainties that cannot be accounted for during the planning process. We first present a simulation framework combining robust planning with adaptive reoptimization strategies based on stochastic minimax optimization for a series of simulated treatments on a one-dimensional phantom subjected to systematic and random uncertainties. The plan applied during the first fractions can handle anticipated uncertainties. The individual uncertainties and their impact on treatment quality is measured at fixed time points. The plan is reoptimized based on the measurements, if necessary. This adapted plan is applied during the subsequent fractions until a next adaptation becomes necessary. In the adaptive strategy proposed, the measured uncertainty scenarios and their assigned probabilities are updated to guide the robust reoptimization. We also discuss other adaptation strategies. The simulations show that robustly optimized plans perform well in presence of uncertainties similar to the anticipated ones. In case of unpredictably large uncertainties, robust adaptive strategies manage to adjust to the unknown probability distribution. We also discuss optimal control strategies applicable in this framework.

Michelle Boeck
KTH Royal Institute of Technology
KTH Royal Institute of Technology

miboeck@kth.se

Anders Forsgren
KTH Royal Institute of Technology
Dept. of Mathematics
andersf@kth.se

Kjell Eriksson, Bjorn Hardemark
RaySearch Laboratories AB
kjell.eriksson@raysearchlabs.com,
bjorn.hardemark@raysearchlabscom

CP1

Explicit Optimization of Plan Quality Measures in Intensity-Modulated Radiotherapy Treatment Planning

The aim of intensity-modulated radiotherapy is to irradiate the tumor while sparing organs and healthy tissue. This is translated into conflicting objective functions of a treatment plan multicriteria optimization model. An optimized plan is approved only if the plan quality is acceptable; if not, the plan undergoes re-optimization with adjusted objective functions. The conventional approach is to use penalty-based functions and is known to necessitate several re-optimizations. We hypothesize that objective functions with a more direct relationship to plan quality metrics would reduce the need for fine-tuning. We investigate the potential of such an explicit approach by approximating the optimization of plan quality indices using mean-tail-doses as objective functions. The result is a linear programming problem. A drawback of using mean-tail-doses is the largely increased problem size, now of the same order as the number of voxels (at least 10^5). However, by using the specific problem structure, a tailored higher-order interior-point method significantly reduces the computational cost. Each iteration only requires factorization of a matrix whose size is proportional to the number of beamlets (of order 10^3 in many cases). We present results of applying the explicit approach to optimize sliding-window treatment plans, and compare these to plans automatically generated using a commercially available algorithm for conventional planning.

Lovisa Engberg
Division of Optimization and Systems Theory
Department of Mathematics, KTH Royal Institute of Technology
loven140@kth.se

Kjell Eriksson
RaySearch Laboratories AB
kjell.eriksson@raysearchlabs.com

Anders Forsgren
KTH Royal Institute of Technology
Dept. of Mathematics
andersf@kth.se

Björn Hårdemark
RaySearch Laboratories, Stockholm
bjorn.hardemark@raysearchlabs.com

CP1

Cluster Analysis of Biological Medicine

Many name-brand biological medicines are currently facing patent expiration, thus creating an opportunity for pharmaceutical companies to develop cost-effective alternatives to their name-brand counterparts. These alternatives are typically referred to as "biosimilars', and have the potential to significantly reduce patient costs by introducing competition into the pharmaceutical marketplace. However, before biosimilars are released to consumers, they must undergo a rigorous comparison with their name brand counterparts to ensure safety and efficacy. Scientists have recently started using Nuclear Magnetic Resonance (NMR) experiments as a basis for this comparison. The result of such an NMR experiment is a highly detailed two-dimensional image, which can reveal the chemical structure underlying the biosimilar in question. Current analysis of NMR protein images relies on a combination of principal component analysis and visual inspection. By formulating the image comparison question as an optimal transport problem, we introduce a quantitative metric for comparing protein structure. Moreover, we employ this metric in a clustering algorithm to group proteins according to their chemical structure.

Ryan M. Evans, Anthony Kearsley
Applied and Computational Mathematics Division
National Institute for Standards and Technology
rmevans@udel.edu, anthony.kearsley@nist.gov

CP1

A Markov Decision Process Approach to Optimizing Cancer Therapy Using Multiple Modalities

There are several different modalities, e.g., surgery, chemotherapy, and radiotherapy, that are currently used to treat cancer. It is common practice to use a combination of these modalities to maximize clinical outcomes, a balance between maximizing tumor damage while minimizing normal tissues toxicity due to treatment. However, multi-modality treatment policies are mostly empirical in current practice, and therefore subject to individual clinicians experiences. We present a novel formulation of optimal multi-modality cancer management using a finite-horizon Markov decision process approach. Specifically, at each decision epoch the clinician chooses the optimal treatment modality based on the patients observed state, which we define as a combination of tumor progression and normal tissue side effect. Treatment modalities are categorized as (1) Type 1, which has a high risk and high reward but can be implemented only once during the course of treatment; (2) Type 2, which has a lower risk and lower reward than Type 1 but may be repeated; and (3) Type 3, no treatment (surveillance), which is the only action that has the possibility of reducing normal tissue side effect at the risk of worsening tumor progression. Numerical simulations using various intuitive reward functions show the structural insights of optimal policies and the potential benefits of using a rigorous model for optimal multi-modality cancer management.

Kelsey Maass
University of Washington
kmaass@uw.edu

Minsun Kim
Department of Radiation Oncology
University of Washington School of Medicine
mk688@uw.edu

CP1

A Parallel Implementation of the Surrogate Management Framework for Optimization in Cardiovas-

cular Surgery

The evolution of surgical methods in cardiovascular disease has traditionally relied on trial & error, outcomes statistics, and surgeon experience. With the advent of high performance computing and patient-specific hemodynamic simulation capabilities, a systematic approach to patient-specific treatment planning, combining simulations and optimization, is now feasible. However, clinically-relevant cost functions are typically expensive to compute and are often non-smooth. In this study, we employ the Surrogate Management Framework (SMF), a derivative free optimization method, in conjunction with black-box hemodynamics simulations for cardiovascular optimization. SMF cycles between an exploratory search step that accelerates convergence and a pattern search based poll step that ensures convergence. To make the SMF more tractable for a time budget, we present a parallel framework for optimization using openDIEL, which allows for multiple simultaneous function evaluations in a high performance computing environment. We compare performance of several competing infill strategies, as well as methods to address ill-conditioning of the surrogate function. We demonstrate the performance of the parallel SMF framework, compared to a standard serial implementation, on a set of benchmark problems and quantify changes in surrogate quality and performance. Finally, we demonstrate an application in shape optimization of a cardiovascular surgery for patients with congenital heart disease.

Aekaansh Verma
Department of Mechanical Engineering,
Stanford University
aekaansh@stanford.edu

Kwai L. Wong
Joint Institute for Computational Science
University of Tennessee/ORNL
kwong@utk.edu

Alison Marsden
School of Medicine
Stanford University, USA
amarsden@stanford.edu

CP2

Robust Convex and Conic Quadratic Optimization with CONCAVE Uncertainty

We derive (computationally) tractable reformulations of the robust counterparts of convex quadratic and conic quadratic constraints with concave uncertainties for a broad range of uncertainty sets. Furthermore, for quadratic problems with linear uncertainty containing a mean vector and covariance matrix, we construct a natural uncertainty set using an asymptotic confidence level and show that its support function is semi-definite representable. Finally, we apply our results to a portfolio choice problem.

Dick Den Hertog, Ahmadreza Marandi, Bertrand Melenberg
Tilburg University
d.denhertog@tilburguniversity.edu, a.marandi@uvt.nl, b.melenberg@uvt.nl

Aharon Ben-Tal
Technion - Israel Institue of Technology
Israel
abental@ie.technion.ac.il

CP2

Applying Robust Optimization to Convex Nonlinear Regression

In many optimization problems, some of the variables are subject to a convex nonlinear constraint, with the parameters of that constraint determined through the solution of a nonlinear regression model. We develop a framework for applying robust optimization to problems of this form, by constructing a piecewise linear approximation to the convex constraint. The parameters of the different linear pieces are connected through a unifying robust constraint. We fix the location of the boundaries between the pieces of the function. Under this assumption, the subproblem for determining the parameters can be solved explicitly as a function of the variables of the original optimization problem. It then follows that the robust version of the original optimization problem can be formulated as a single convex integer programming problem. Various formulations of the robust constraint are considered. The approach is illustrated through application to a problem in humanitarian logistics. In this situation, part of the cost function corresponds to the deprivation costs, which are typically modeled as convex exponential functions of the time a person is deprived of a resource.

Dan Lu
Rensselaer Polytechnic Institute (RPI)
lud@rpi.edu

John E. Mitchell
Rensselaer Polytechnic Institute
110, 8th Street, Troy, NY, 12180
mitchj@rpi.edu

CP2

Robust Convex and Conic Quadratic Optimization with CONVEX Uncertainty

For convex quadratic and conic quadratic constraints with convex uncertainty, it is well-known that the robust counterpart is, in general, intractable. We derive tractable inner and outer approximations of the robust counterpart of such constraints for various types of convex compact uncertainty sets. The approximations are made by replacing the quadratic terms in the uncertain parameters with suitable linear upper and lower bounds and applying our results to a quadratic constraint with linear uncertainty to derive a tractable formulation. Finally, we apply our results to a norm approximation and a regression line problem.

Ahmadreza Marandi
Tilburg University
a.marandi@uvt.nl

Aharon Ben-Tal
Technion - Israel Institue of Technology
Israel
abental@ie.technion.ac.il

Dick Den Hertog
Tilburg University
Faculty of Economics and Business Administration
D.denHertog@uvt.nl

Bertrand Melenberg

Tilburg University
b.melenberg@uvt.nl

CP2

Robust Optimization Approach for Multivariate Sampling Allocation Problems in Stratified Sampling

A central problem in survey sampling is the optimal allocation of samples to entities of the population. The primary goal of the allocation is to minimize the variances of estimated totals of the population which, in practice, often has to be treated as a multivariate allocation problem. The optimal solution in stratified sampling, where the strata are the entities, relies on the stratum specific variances of the (multiple) variables of interest that are assumed to be known from earlier surveys or highly correlated variables. However, these stratum specific variances are generally not known precisely. In order to account for this uncertainty we formulate a robust optimization model for this multivariate stratified allocation problem. Using robust optimization, we seek an optimal allocation which is feasible for any realization of the uncertain variances. A real life problem is considered and solved in order to validate the applicability of the model.

Mohammad Asim Nomani
Research and Training Group on Algorithmic Optimization
Department of Mathematics, University of Trier, Germany
nomani@uni-trier.de

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

Mirjam Dür
Department of Mathematics
University of Trier
duer@uni-trier.de

Ralf Münnich
Economic and Social Statistics Department
University of Trier, Germany
muennich@uni-trier.de

CP2

Distributed Adjustable Robust Optimization

Conventional robust optimization (RO) considers that uncertain parameters belong to some compact (uncertainty) set and all variables must be determined before the realization of uncertain parameters becomes known. All variables represent here and now decisions. Often a subset of variables referred to as adjustable can be chosen after the realization of the uncertain data while others (nonadjustable) must be determined before its realization. The former variables represent wait and see decisions that can be made when revealing part of the uncertain data and that can adjust to the corresponding part of the data. In this case, Adjustable RO (ARO) produces robust counterparts (ARC) that are less conservative than the usual RC though computationally harder. In the non-robust context, decomposition provides a useful method where the structure of the original problem can be exploited in order to split it into subproblems that are repeatedly solved. The major strength of decomposition methods stems from their ability to reformulate the original problem and generate subproblems (coupled through a master problem) that are significantly easier to solve than the original problem and that can be solved concurrently (if subproblems are independent). However, the decomposable structure of the uncertain original problem (OP) may disappear when formulating its ARC. Hence, preserving the decomposability property becomes essential in order to allow distributed solving of ARCs.

Dimitri Papadimitriou
Bell Labs
dimitri.papadimitriou@alcatel-lucent.com

CP2

Adjustable Robust Optimization via Fourier-Motzkin Elimination

We demonstrate how adjustable robust optimization (ARO) problems with fixed recourse can be casted as static robust optimization problems via Fourier-Motzkin elimination (FME). Through the lens of FME, we characterize the structures of the optimal decision rules for a broad class of ARO problems. A scheme based on a blending of classical FME and a simple Linear Programming technique that can efficiently remove redundant constraints, is developed to reformulate ARO problems. This generic reformulation technique enhances the classical approximation scheme via decision rules, and enables us to solve adjustable optimization problems to optimality. We show via numerical experiments that, for small-size ARO problems our novel approach finds the optimal solution. For moderate or large-size instances, we eliminate a subset of the adjustable variables, which improves the solutions from decision rule approximations.

Jianzhe Zhen
Faculty of Economics and Business Administration
Tilburg University
J.Zhen@uvt.nl

Dick Den Hertog
Tilburg University
Faculty of Economics and Business Administration
D.denHertog@uvt.nl

Melvyn Sim
Department of Decision Sciences NUS Business School
National University of Singapore
melvynsim@gmail.com

CP3

A Distributed Jacobi Algorithm for Large-Scale Constrained Convex Optimization

We present an iterative distributed Jacobi algorithm for solving large-scale constrained convex optimization problems that come from distributed model predictive control (DMPC). The centralized problem has convex cost function with respect to affine constraints, assuming there are couplings between subsystems in both the cost function and the constraints. The proposed distributed approach involves a decomposition based on the sparsity structure of the centralized problem, then a cooperative Jacobi parallel iteration is performed, the new iterate is obtained by a convex combination of new local solutions and old iterates. Starting from a feasible solution, the algorithm iteratively reduces the cost function value with guaranteed feasible so-

lutions at every iteration step, thus it is suitable for DMPC applications in which hard constraints are important. We provide conditions for convergence to centralized optimality using the distributed scheme. This proposed approach is applicable to inter-connected systems where the number of subsystems is large, the coupling is sparse, and local communication is available. This work follows the work on Jacobi algorithm for large-scale unconstrained convex optimization in: Bertsekas, D.P. and Tsitsiklis, J.N. (1989). Parallel and Distributed Computation: Numerical Methods. Prentice-Hall, Upper Saddle River, NJ.

Dang Doan
Systems Control and Optimization Laboratory
University of Freiburg
dang.doan@imtek.uni-freiburg.de

CP3

Frchet Differentiability of the Boussinesq Flow with Respect to Domain Variations

In this talk we investigate shape differentiability properties of the solution operator associated with the unsteady incompressible Boussinesq equations and the method of mapping by Murat and Simon. More precisely, by using the method of mapping we transform the Boussinesq equations defined on Ω to a reference domain $\Omega_{ref} = \tau^{-1}(\Omega)$ and show that the corresponding solution operator $\tau \mapsto (v, \theta, p)(\tau)$ is Frchet differentiable in a suitable Banach space setting. To this end, we propose a general analytical framework beyond the implicit function theorem to show the Frchet differentiability of the transformation-to-state mapping conveniently. This framework can be applied to a variety of other shape optimization or optimal control problems and takes care of the usual norm discrepancy needed for nonlinear problems to show differentiability of the state equation and invertibility of the linearized operator. The analysis is carried out for Ω_{ref} being a Lipschitz domain and aims at minimizing the regularity assumptions on the transformation variable τ. This is in parts based on joint work with F. Lindemann, M. Ulbrich and S. Ulbrich.

Michael H. Fischer
TU Darmstadt
mfischer@mathematik.tu-darmstadt.de

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

CP3

Nash Equilibrium Strategy for Uncertain Markov Jump Linear Stochastic Systems with Multiple Decision Makers

Nash equilibrium strategy for uncertain Markov jump linear stochastic systems with multiple decision makers is addressed. First, the stochastic algebraic Riccati inequality is established such that the closed-loop stochastic system is exponentially mean-square stable and has a cost bound. Second, a robust strategy set is established such that Nash equilibrium condition is satisfied. It is shown that the necessary conditions, which are determined from the Karush-Kuhn-Tucker conditions, for the existence of a strategy set are defined by means of solutions of cross-coupled stochastic algebraic Riccati equations. Furthermore, Newton's method is used for solving the CSAREs. In an alternate approach designed to simplify implementation, a recursive technique based on linear matrix inequality as the convex optimization technique is applied. Finally, a numerical example is provided to demonstrate the effectiveness of the method.

Hiroaki Mukaidani
Institute of Engineering, Hiroshima University
mukaida@hiroshima-u.ac.jp

CP4

jInv - A Flexible Julia Package for Parallel PDE Parameter Estimation

jInv is a software package and framework for the solution of large-scale PDE constrained optimization problems, implemented in the Julia programming language. It supports linear and nonlinear PDE models and provides many tools commonly used in parameter estimation problems such as different misfit functions, regularizers, and efficient methods for numerical optimization. Also, it provides easy access to both iterative and direct linear solvers for solving linear PDEs. A main feature of jInv is the provided easy access to parallel and distributed computation supporting a variety of computational architectures: from a single laptop to large clusters of cloud computing engines. Being written in the high-level dynamic language Julia, it is easily extendable and yet fast. We demonstrate the capabilities of jInv using numerical experiments motivated by geophysical imaging.

Patrick Belliveau
University of British Columbia
patrick.t.belliveau@gmail.com

Eldad Haber
Department of Mathematics
The University of British Columbia
haber@math.ubc.ca

Lars Ruthotto
Department of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

Eran Treister
Ben Gurion University of the Negev
erant@bgu.ac.il

CP4

Adaptive, Inexact Optimization with Low-Rank Tensors for Optimal Control of PDEs under Uncertainty

We present an adaptive, inexact optimization method for the solution of optimal control problems with PDEs under uncertainty. These parametrized PDEs are formulated in a suitable tensor Banach space, which gives rise to a discretization with coefficients in tensor form. To avoid the curse of dimensionality these coefficients are represented in a low-rank tensor format (Hierarchical Tucker or Tensor Train). Therefore, the developed methods are tailored to the usage of low-rank tensor arithmetics, which only offer a limited set of operations and require truncation (rounding) in between. The occurring errors are controlled to ensure global convergence of the algorithm.

Sebastian Garreis
TU Muenchen, Germany

garreis@ma.tum.de

Michael Ulbrich
Technische Universitaet Muenchen
Chair of Mathematical Optimization
mulbrich@ma.tum.de

CP4

Adaptive Eigenspace Method for Inverse Scattering Problems in the Frequency Domain

A nonlinear optimization method is proposed for the solution of inverse scattering problems in the frequency domain, when the scattered field is governed by the Helmholtz equation. The time-harmonic inverse medium problem is formulated as a PDE-constrained optimization problem and solved by an inexact truncated Newton-type iteration. Instead of a grid-based discrete representation, the unknown wave speed is projected to a particular finite-dimensional basis of eigenfunctions, which is iteratively adapted during the optimization. Truncating the adaptive eigenspace (AE) basis at a (small and slowly increasing) finite number of eigenfunctions effectively introduces regularization into the inversion and thus avoids the need for standard Tikhonov-type regularization. Both analytical and numerical evidence underpins the accuracy of the AE representation. Numerical experiments demonstrate the efficiency and robustness to missing or noisy data of the resulting adaptive eigenspace inversion (AEI) method.

Marcus J Grote
Department of Mathematics and Computer Science
University of Basel, Switzerland
marcus.grote@unibas.ch

Marie Graff-Kray
Department of Earth, Ocean and Atmospheric Sciences
University of British Columbia, Vancouver BC, Canada
mgraff@eoas.ubc.ca

Uri Nahum
Department of Mathematics and Computer Science
University of Basel, Switzerland
uri.nahum@unibas.ch

CP4

Controlling the Footprint of Droplets via Surface Tension

The development of engineered substrates has progressed to a very advanced level, which allows for control of the shape of sessile droplets on these substrates. Controlling local droplet shape via substrate surface tensions may be useful for directing the growth of bio-films and cell cultures because it can affect the distribution of nutrients as well as the gross shape of the film. In addition, depositing a film of material onto a substrate in a particular pattern could be affected by the droplet shape. Also, droplets can act as lenses, with focal properties controlled by locally modifying substrate tensions. In this talk, we present an optimal control for the shape of droplets on substrates. Specifically, we wish to direct the shape of the droplet-substrate interface (i.e. the liquid-solid interface) by controlling the substrate surface tension through the minimization of an appropriate cost functional. We refer to this as droplet footprint control.

Antoine Laurain
Humboldt-University of Berlin
laurain@ime.usp.br

Shawn W. Walker
Louisiana State University
Department of Mathematics and CCT
walker@math.lsu.edu

CP5

Approximation of Stochastic Equilibrium Problems via Lopsided Convergence

In this talk, a review of an approximation scheme for solving maxinf (minsup) optimization problems is presented, with a special focus on stochastic equilibrium problems. The family of the problems that can be modeled under this setting is appealingly large, ranging from linear and nonlinear complementarity problems, fixed points, variational inequalities, equilibrium problems, as well as optimization under stochastic ambiguity, robust optimization, risk-based optimization, and semi-infinite optimization. The mathematical programming problem is set to find a maxinf (minsup) point of a given bifunction. Given this setting, a solution strategy is proposed using an approximation scheme, based on the application of lopsided convergence theory. The first definition of lopsided convergence was established during the eighties for extended real value bifunctions, then revised by 2010, for finite value functions over product sets. Lately, it was revisited by extending the definition to more general domains and proposing an associated metric, the lopsided distance. Finally, some numerical results are presented for three examples: 1) a general equilibrium model for an exchange economy with uncertainty; 2) a stochastic general equilibrium model with financial markets, and 3) an infrastructure planning for fast electric vehicles charging station problems.

Julio Deride
University of California Davis
julio.deride@gmail.com

CP5

On Stochastic Nonlinear Minimax Optimization

We study the formulation $\min_{x\in\mathcal{X}} g(x)$ where $g(x) = \max_{y\in\mathcal{Y}} h(x,y) = E_\xi[H(x,y,\xi)]$, when $h(x,y)$ is concave in y for fixed x but $g(x)$ is not convex in x. These arise, for instance, in optimizing total profit by maximizing over exogenous revenue g obtained from sales x and minimizing production cost h over supply y, where the random quantity ξ may represent uncertainty in production yield or supply chain unreliability etc. The convergence guarantees of the standard technique of sample average approximation (SAA) of a Lagrangian reformulation, provided for stochastic convex-concave minimax formulations, do not apply in this case. We propose a stochastic first-order (gradient following) recursion for the indefinite outer objective g and combine it with SAA for the concave h. An estimate for the gradient of g is obtained from the concave inner problem via parametric programming, but is biased because of the finite sample approximation of h. We provide conditions under which the bias in the iterations can be controlled to obtain convergence. Conditions under which the algorithm is computationally efficient, i.e. when it achieves the fastest possible convergence rate (in a certain precise sense), are specified.

Soumyadip Ghosh, Mark Squillante
IBM Research

ghoshs@us.ibm.com, mss@us.ibm.com

Ebisa Wollega
IE, Colorado State, Pueblo
ebisa.wollega@csupueblo.edu?

CP5

Stochastic Algorithms for Minimizing Relatively Smooth Functions

Relative smoothness - a notion recently introduced by Lu, Freund and Nesterov - generalizes the standard notion of smoothness typically used in the analysis of gradient type methods. We develop several randomized algorithms for minimizing relatively-smooth functions such as stochastic gradient descent and randomized coordinate descent, thus extending the reach of modern fast stochastic methods to a wider class of functions.

Filip Hanzely, Peter Richtarik
University of Edinburgh
fhanzely@gmail.com, peter.richtarik@ed.ac.uk

CP5

Reconfigurable Optimization

We develop a novel modeling framework for a class of optimization problems in which operational decisions are altered as a system transitions stochastically between world states. Our framework addresses the problem of minimizing reconfiguration costs associated with changing the operational decisions and ensuring good quality solutions in each individual state. We compare this framework with existing modeling frameworks like Markov decision processes, stochastic programming, and robust optimization. We then use our framework to formulate a large real-world power systems problem and develop a Lagrangian-relaxation based algorithm to solve it.

Prateek R. Srivastava, Nediako Dimitrov
University of Texas at Austin
prateekrs@utexas.edu, ned@austin.utexas.edu

David Alderson
Naval Postgraduate School
dlalders@nps.edu

CP5

Stochastic Methods for Stochastic Variational Inequalities

We consider the stochastic variational inequality problem (svi): solve a variational inequality over a convex closed set whose operator is only revealed trough an unbiased stochastic oracle. In particular, svi includes first order conditions of stochastic optimization, saddle-point problems and Nash games. We propose stochastic approximated methods based on cheap first-order methods for large-scale problems assuming only monotonicity and Lipschitz continuity. We weaken significantly the standard assumptions in stochastic approximation: (1) the feasible set and operator may unbounded without strong-monotonicity and (2) we do not require an oracle with uniformly bounded variance (e.g. multiplicative noise). One class of methods are based on extragradient schemes. Under a mini-batch procedure, we obtain rate $O(1/k)$ with near-optimal sample complexity $O(\epsilon^{-2})$ under the setting of (1)-(2), improving the previous rate $O(1/\sqrt{k})$ assuming boundedness or strong-monotonicity or uniformly bounded variance. A second class of methods include randomized alternating projections when the number of constraints is too large or its components are revealed in an online fashion.

Philip Thompson
Center for Mathematical Modeling (CMM)
philipthomp@gmail.com

Alfredo N. Iusem
IMPA, Rio de Janeiro
iusp@impa.br

Alejandro Jofré
Universidad de Chile, Santiago, Chile
ajofre@dim.uchile.cl

Roberto Oliveira
IMPA, Rio de Janeiro
rimfo@impa.br

CP5

Structural Properties of Probability Constraints

A probability function results from a measure acting on a random inequality system depending both on a decision and random vector. Probability constraints requests that such a system holds with high enough probability in order to ensure safety of the decision. In this talk we will discuss recent insights concerning differentiability of probability functions, but too new insights on convexity of feasible sets for probability constraints. The discussed results cover nonlinear systems as well as elliptically distributed random vectors.

Wim Van Ackooij
EDF R&D
Department OSIRIS
wim.van.ackooij@gmail.com

Rene Henrion
Weierstrass Institute
Berlin, Germany
henrion@wias-berlin.de

Jérôme Malick
University of Grenoble
jerome.malick@univ-grenoble-alpes.fr

CP6

On the Relation Between the Cauchy and the Gradient Sampling Method

The minimization of nonsmooth functions appears in many real world applications. Therefore, robust and efficient algorithms for this kind of problem are a necessity. Recently, a method known as Gradient Sampling (GS) has gained attention by its good numerical results and by its intuitive functioning. Although one can view the GS method as a generalization of the steepest descent method, studies exploring this relation are scarce. Since the GS method finds its search direction in a nondeterministic approach, it is not a trivial task to establish a connection between the descent direction taken by the Cauchy method and the Gradient Sampling algorithm in each iteration. In order to have a better understanding of this relation, this study presents a local convergence result for the GS method, showing how the linear rate of the steepest descent method can be obtained for the GS algorithm under special circumstances.

Lucas Eduardo Azevedo Simoes
University of Campinas
simoes.lea@gmail.com

Elias Salomao Helou
University of Sao Paulo
elias@icmc.usp.br

Sandra Augusta Santos
University of Campinas
sandra@ime.unicamp.br

CP6

Weak Subgradient Based Solution Algorithm in Nonconvex Unconstrained Optimization

In this work, we study a weak subgradient based solution algorithm in nonconvex unconstrained optimization. The concept of weak subgradients in nonconvex analysis, is introduced by Azimov and Kasimbeyli, which is one of the natural generalizations of the classic subdifferential of the convex analysis. The weak subdifferentiability of a function at the given point means the existence of a supporting cone (instead of a supporting hyperplane in convex analysis) to the epigraph of the function at the point under consideration. Because of this property the weak subdifferentiability does not require convexity condition. We propose a solution algorithm for unconstrained minimization problem, where no convexity is assumed and at every iteration the weak subgradient of the objective function is used to generate new solution. The paper investigates convergence properties for the sequence generated by the algorithm. The performance of the proposed algorithm is demonstrated on test problems from the literature.

Gulcin Dinc Yalcin, Refail Kasimbeyli
Anadolu University
gdinc@anadolu.edu.tr, rkasimbeyli@anadolu.edu.tr

CP6

Optimality Conditions for Nonsmooth Constrained Optimization Problems

Nonsmoothness arises in many practical optimization problems and can often be recast in so-called abs-normal form where every occurence of nonsmoothness is expressed in terms of the absolute value function. Necessary and sufficient first and second order optimality conditions for this class of unconstrained nonlinear nonsmooth minimization have recently been developed. We discuss illustrative examples and extend the theory to nonsmooth constrained optimization.

Lisa C. Hegerhorst
Leibniz Universität Hannover
Institute for Apllied Mathematics
hegerhorst@ifam.uni-hannover.de

CP6

Clustering in Large Data Sets with the Limited Memory Bundle Method

Clustering is among most important tasks in data mining. This problem in large data sets is challenging for most existing clustering algorithms. Here we introduce a new algorithm based on nonsmooth optimization techniques for solving the minimum sum-of-squares clustering problems in large data sets. The clustering problem is first formulated as a nonsmooth optimization problem. Then the limited memory bundle method [Haarala et.al. *Math. Prog.*, Vol. 109, No. 1, pp. 181–205, 2007] is modified and combined with an incremental approach to solve this problem. The algorithm is evaluated using real world data sets with both large numbers of attributes and large numbers of data points. The new algorithm is also compared with some other optimization based clustering algorithms. The numerical results demonstrate that the new algorithm is both efficient and accurate and it can be used to provide real-time clustering in large data sets.

Napsu Karmitsa
Department of Mathematics
University of Turku
napsu@karmitsa.fi

Adil Baghirov
CIAO, School of ITMS,
The University of Ballarat
a.bagirov@federation.edu.au

Sona Taheri
Federation University Australia
Faculty of Science and Technology
s.taheri@federation.edu.au

CP6

Nonsmooth DC Optimization Algorithm for Clusterwise Linear L1 Regression

The aim of this research is to develop an algorithm for solving the clusterwise linear least absolute deviations regression problem. This problem is formulated as a nonsmooth nonconvex optimization problem. The objective function in the problem is represented as a difference of two convex functions. In this representation the first DC component is expressed as a sum of simple nonsmooth functions. We proposed to smooth the first DC component. Such an approach allows to design algorithms for which strong convergence results can be obtained. Optimality conditions are derived using this representation. An algorithm is designed based on the difference of convex representation and an incremental approach. The proposed algorithm is tested using synthetic, real world small and large scale data sets for regression analysis.

Sona Taheri
Federation University Australia
Faculty of Science and Technology
s.taheri@federation.edu.au

Adil Baghirov
CIAO, School of ITMS,
The University of Ballarat
a.bagirov@federation.edu.au

CP6

An SQP Trust-Region Algorithm for Constrained Optimization Without Derivatives

Constrained optimization without using derivatives is an active area of research as there are several reasons, derivative information can not be retrieved. Derivative-free optimization algorithms modeling the objective and constraint functions by any kind of polynomials have shown to be very efficient. We want to present our SQP trust-region algorithm for constrained optimization without derivatives using different interpolation techniques. The application of variable size models between linear and quadratic degree and under-determined interpolation, where significantly less sample points are needed, is of special interest. Our implementation is compared to other well-known DFO-libraries. Numerical experiments will be carried out on a subset of the CUTEst test problem library and on a real-life application.

Anke Troeltzsch
German Aerospace Center (DLR)
Cologne, Germany
anke.troeltzsch@dlr.de

CP7

Design and Optimization of a Cylindrical Explosive

The design of explosive devices has a wide reaching impact in mining, demolition, and military applications. Traditionally the design of explosives is an expensive and time consuming process; involving extensive machining, building experimental setups, and going to great lengths to ensure the safety of everyone involved. These material and labor costs can be significantly reduced by using computational tools to assist in the design of an explosive. Here a cylindrically contained high explosive was analyzed using Dakota, an optimization toolkit developed at Sandia National Laboratories, coupled with two hydrocodes: KO, a one dimensional Lagrangian formulation, and CTH, an industry standard Eulerian shock physics code. Initial studies were performed with KO to optimize the thickness of a cylinder for varying outer radii; the explosive was modelled as a bulk gas with pressure and velocity taken from a standard STEX test. This resulted in an intuitive parametric surface which was both smooth and convex. These results agreed with a second study of a one dimensional CTH simulation using an HEBURN explosive TNT model. Further investigation showed a degree of similitude in the one dimensional model, and a potential to reduce the system down to a scalar relationship. This provides a foundation to explore and simplify the three dimensional case. The effects of shell geometry will also be explored by parameterizing the outer profile as an input to the optimizer.

Logan E. Beaver, John Borg
Marquette University
logan.beaver@marquette.edu, john.borg@marquette.edu

Geremy Kleiser
Eglin Air Force Base
geremy.kleiser@us.af.mil

CP7

A Stochastic Programming Framework for Multi-Stakeholder Decision-Making and Conflict Resolution

Engineering decision-making is inherently multiobjective and involves multiple stakeholders. As such, it suffers from ambiguity and dimensionality. We propose a decision-making framework to compute compromise solutions that balance conflicting priorities of multiple stakeholders. In our setting, we shape the stakeholder dissatisfaction distribution by minimizing a strongly monotone risk metric, such as conditional value-at-risk (CVaR) and entropic value-at-risk (EVaR). Such risk metrics are often parameterized by a probability level that shapes the tail of the dissatisfaction distribution. The proposed approach allows us to compute a family of compromise solutions and generalizes multi-stakeholder settings previously proposed in the literature. For the CVaR metric, we use the concept of the CVaR norm for geometric interpretation and to establish strong monotonicity, which guarantees Pareto efficient compromise solutions under mild assumptions about stakeholder priorities. We discuss a broad range of potential applications that involve complex decision-making processes. We demonstrate the developments by balancing stakeholder priorities on transportation, safety, water quality, and capital costs in a biowaste facility location case study. We also discuss extensions to energy system design and operation, such as combined cooling, heat and power systems, and solar power generators with energy storage.

Alexander W. Dowling
University of Wisconsin-Madison
Department of Chemical and Biological Engineering
adowling2@wisc.edu

CP7

Optimal Power Flow Based on Holomorphic Embedded Load Flow

Optimal Power Flow (OPF) is known as a difficult global optimization problem. Due to its importance in its own right and as a subproblem in other power systems optimization problems it has attracted a lot of attention eg by SDP relaxations. If solved by a standard NLP solver to local optimality the resulting solution may not just be suboptimal but also non-physical, in that it violates bus stability criteria. The Holomorphic Embedded Load Flow (HELM) method has been suggested recently by Antonio Trias as an alternative (and efficient) method of solving the related load flow problem. The claim (observed in practice, but not formally proven) is that HELM will never return unstable solutions. We present a novel method to wrap the HELM method in an SQP framework to arrive at a OPF solver that avoids unstable solutions. This does not avoid all local solutions, but a significant number depending on the problem in question. HELM-OPF is shown to be competetive with other local OPF solvers, but has a higher chance of finding the global solution.

Andreas Grothey
School of Mathematics
University of Edinburgh
A.Grothey@ed.ac.uk

Ian Wallace
University of Edinburgh
i.p.wallace@ed.ac.uk

CP7

Geometry and Topology Optimization of Sheet Metal Profiles by Using a Branch-and-Bound Framework

We present a new approach combining well established procedures from PDE-constrained and discrete optimization in order to find an optimal design of a multi-chambered profile. Given a starting profile design, a load case and corresponding design constraints (e.g., sheet thickness, chamber sizes), we want to find an optimal subdivision into a predefined number of chambers with optimal shape subject to structural stiffness. Within the presented optimization scheme a branch-and-bound tree is generated to model the topological decisions. Therefore, each level of the tree represents one additional chamber of the profile. Before the next chamber is added, the geometry of the current profile design is optimized by an SQP method. This is followed by an application of the SIMP approach to solve a relaxation of a topology optimization problem, preparing the further subdivision of the profile. Based on this relaxation, a best fitting feasible topology subject to manufacturability conditions is determined using a new mixed integer method employing shortest paths. To improve the running time, the finite element simulations for the geometry optimization and topology relaxation are performed with different levels of accuracy. Additionally, we present numerical experiments including different starting geometries, load scenarios and mesh sizes.

Benjamin M. Horn, Hendrik Lüthen, Marc Pfetsch
Research Group Optimization, Department of Mathematics
TU Darmstadt
bhorn@mathematik.tu-darmstadt.de,
luethen@mathematik.tu-darmstadt.de,
pfetsch@mathematik.tu-darmstadt.de

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

CP7

Classification of Mass Spectra Using a Multi-Objective Control Approach

Mass spectrometry is an analytical chemistry technique commonly employed for chemical identification. The resulting mass spectrum generated during mass spectrometry can be used to elucidate structural information about the chemical sample. Identification of an unknown chemical, given its mass spectrum and a curated library of mass spectra of known compounds, can usually be accomplished through a library search assuming that the instrument produces a sufficiently accurate spectrum and the spectrum of the unknown chemical is present in the library. Unfortunately, many applications see the synthesis of new chemical compounds occurring at a rate that exceeds the growth of curated libraries (e.g. illicit drug development). As such, identification must be supplemented with rigorous classification approaches to generate expectations of an unknown chemical samples identity. In this presentation, we discuss the unique challenges of mass spectral data analysis and introduce a multi-objective control-based mathematical framework for improved classification and identification of chemical compounds. Methods analysis and preliminary results will be discussed in detail.

Arun S. Moorthy
National Institute of Standards and Technology
arun.moorthy@nist.gov

Anthony Kearsley
Applied and Computational Mathematics Division
National Institute for Standards and Technology
anthony.kearsley@nist.gov

William E. Wallace
National Institute of Standards and Technology
william.wallace@nist.gov

CP7

Pricing and Clearing Combinatorial Markets with Singleton and Swap Orders

We present an algorithm that permits polynomial time market-clearing and -pricing for a certain class of combinatorial markets. The results are presented in the context of our main application: the futures opening auction problem. Futures contracts are an important tool to mitigate market risk and counterparty credit risk. The contracts can be traded with varying expiration dates and underlyings. Several exchanges also offer so-called futures contract combinations, which allow the traders for swapping two futures contracts with different expiration dates. In theory, the price is in both cases the difference of the two involved futures contracts. However, in particular in the opening auctions price inefficiencies often occur due to suboptimal clearing, leading to potential arbitrage opportunities. We present a minimum cost flow formulation of the futures opening auction problem that guarantees consistent prices. The core ideas are to model orders as arcs in a network, to enforce the equilibrium conditions with the help of two hierarchical objectives, and to combine these objectives into a single weighted objective while preserving the price information of dual optimal solutions. The resulting optimization problem can be solved in polynomial time and computational tests establish an empirical performance suitable for production environments.

Johannes C. Müller
FAU Erlangen-Nürnberg
JohannesMueller@fico.com

Sebastian Pokutta
Georgia Tech
sebastian.pokutta@isye.gatech.edu

Alexander Martin
FAU Erlangen-Nürnberg
alexander.martin@fau.de

Susanne Pape
FAU Erlangen-Nürnberg, Germany
Discrete Optimization
susanne.pape@math.uni-erlangen.de

Andrea Peter
University of Erlangen-Nuremberg
andrea.peter@math.uni-erlangen.de

Thomas Winter
Eurex Frankfurt AG
thomas.winter@eurexchange.com

CP8

Cybersecurity Using Interdiction

Recent cyber-attacks on private and public groups highlight the importance of a proper cybersecurity structure. Having well-structured cybersecurity decreases the vulnerability of the cyber physical system. We present an MDP to compute an optimal attack policy. The MDP has an exponential number of states, and is based on tracking the set of available attacks for each link in the network. We show that it is possible to compute values for each MDP state, and optimal attack policies using s-t reliability. We

also present a network interdiction model to find the optimal defense strategy for a cyber physical network. Unlike many cybersecurity models, our model considers the specifics of the network structure under attack. Software resulting from the model can assess the optimal defense strategy for a cyber physical system.

Murat Karatas, Nedialko Dimitrov
The University of Texas at Austin
mkaratas@utexas.edu, ned@austin.utexas.edu

CP8

Concavity and Bounds for the Stochastic Dynamic Programming Model on Social Networks

The problem of selecting an initial set of nodes in any network to create a cascade effect also known as the Influence Maximization (IM) problem has been extensively studied and has become an effective means of optimizing revenue for advertising companies via social networks. In this paper we investigate a Stochastic Dynamic Programming (SDP) approach to the IM problem which involves dividing the problem into stages or sub-problems, computing the solutions to these sub-problems and utilizing these solutions in future stages. The objective function in the SDP approach defines the value of an optimal solution recursively by expressing it in terms of optimal solutions for smaller problems and solutions are obtained in a bottom-up fashion. Our SDP algorithm not only identifies an optimal allocation but also its optimal value. In this paper, we investigate the Stochastic Dynamic Programming (SPD) approach to solving the Influence Maximization problem and demonstrate through our experimental results, the concavity of the objective function. It was found that the objective function was consistently concave, when tested on networks with varying node degrees. By defining and obtaining a secant line on our objective function, we provide a theoretical proof that the objective function for our SDP model is concave. Consequently, we obtain an upper bound for our optimal solution.

Trisha Lawrence
The University of Saskatchewan
trishalawren@gmail.com

CP8

Solution Attractor of Local Search System for the Traveling Salesman Problem

Although both the TSP and local search have huge literature, there is still a variety of open problems. The study of local search for TSP continues to be an interesting problem in combinatorial optimization and computational mathematics. We study local search for the TSP from the perspective of dynamical systems and treat a local search system as a discrete dynamical system. The attractor theory in dynamical systems provides the necessary and sufficient theoretical foundation to study the search behavior of local search system. In a local search system, all search trajectories converge into a small region of the solution space, called solution attractor. We will describe a procedure for constructing solution attractor of a local search system for TSP. This procedure can be used to build an attractor-based search system to solve the TSP and its variations. We will also present our empirical study on some important properties of the solution attractor, including convergence of local search trajectories, the size of the solution attractor, and quality of the tours in the solution attractor.

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

CP8

A Robust Supply Chain Network Equilibrium Model and its Analysis

Mathematical models associated with supply chain are one of the most important objects in management science, and hence many researchers have studied these models. Because supply chains involve many decision-makers (players) and they independently decide their behaviors, competitive situations often occur. To investigate competitive supply chain networks, Nagurney et al. (2002) proposed a supply chain network equilibrium (SCNE) model. On the other hand, recently, particular attention is paid to the risk management of supply chain. In this talk, to analyze competitive supply chain networks involving uncertainty, we propose a robust SCNE model by reconstructing the usual SCNE model. In the proposed model, each player cannot know exactly other players strategies, and they decide their strategy by assuming the worst case. We formulate the robust SCNE model as a variational inequality problem (VIP). The set associated with the VIP is constructed by second-order cone constraints. We show the existence and the uniqueness of the equilibrium under mild assumptions. Finally, some numerical results are given.

Yasushi Narushima
Yokohama National University
Department of Manegement System Science
narushima@ynu.ac.jp

Tatsuya Hirano
Yokohama National University
Faculty of International Social Science
hirano-tatsuya-fj@ynu.jp

CP8

Nonlinear Transient Optimization in Gas Networks

Non-constant operation scenarios and operation costs require non-stationary models for gas transport in pipelines. We present a transient optimization model that incorporates gas dynamics and technical network elements. Direct discretization leads to an NLP that can be solved for example with interior point methods. We discuss different ideas on exploitation of the problem structure and present results for simple networks and scenarios that show the potential to store highly volatile generated renewable energy in terms of pressure increase in gas networks.

Jan Thiedau
LU Hannover
Institute for Applied Mathematics
thiedau@ifam.uni-hannover.de

Marc C. Steinbach
Leibniz Universität Hannover
Institute for Applied Mathematics
mcs@ifam.uni-hannover.de

CP9

Geneva: Parametric Optimization in Distributed

and Parallel Environments

The Geneva library of parametric optimization algorithms covers execution on parallel devices ranging from GPGPUs and many-core systems over clusters to Grids and Clouds. Four optimization algorithms, including particle swarms and evolutionary algorithms have been implemented, and best solutions from one algorithm may be transferred to another. Parallelization is mostly transparent to the user, leaving little work for him to be done for execution on a wide range of devices, once his optimization problems have been defined. The entire library is available as Open Source, and targets particularly problems with very long running, computationally expensive optimization problems, often involving simulations. The definition of optimization problems may involve constraints between multiple parameters. Geneva was originally developed for and used in science, particularly particle physics (hence the name), but is today also used commercially. The presentation covers the architecture of Geneva and introduces use-cases. Note to organizers: This is meant as a presentation only, no paper-submission is intended.

Ruediger D. Berlich, Ariel Garcia, Sven Gabriel
Gemfony scientific
r.berlich@gemfony.eu, a.garcia@gemfony.eu, s.gabriel@gemfony.eu

CP9

Crash-Starting the Dual Simplex Method

This talk discusses how to find a good advanced basis start for the dual simplex method. In particular, it will discuss the applicability of techniques in the spirit of the primal simplex "idiot crash' implemented by Forrest in Clp. Via presolve, one-dimensional quadratic minimizations, crossover and further presolve, this unpublished technique can obtain a very good advanced basis for an LP of very much smaller dimension than the original problem. Once this reduced problem is solved using the primal simplex method, it yields a good (if not optimal) basis for the original problem, allowing it to be solved quickly using the primal simplex method. After giving an insight into the primal simplex "idiot crash' and discussing the LP problem characteristics for which the crash is effective, this talk will discuss similar techniques for crash-starting the dual simplex method.

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

CP9

Update on Expression Representations and Automatic Differentiation for Nonlinear AMPL Models

AMPL facilitates stating and solving nonlinear programming problems involving algebraically defined objectives and constraints. For solving such problems, the AMPL/solver interface library provides routines that compute objective functions, constraint residuals, and associated derivatives. Objectives and constraint bodies hitherto have been represented by "executable' expression graphs, in which each node points to its operands and to a function that computes the node's result. Nodes also store partial derivatives for use in computing gradients and Hessians by automatic differentiation. Storing these values makes the graphs nonreentrant. To enable several threads to evaluate the same expression at different points without having separate copies of the expression graphs, such details as variable values and partial derivatives must be stored in thread-specific arrays. I will describe and compare some expression-graph representations for use in computing function, gradient, and Hessian values, and for extracting some auxiliary problem information. This is an updated version of an earlier talk.

David M. Gay
AMPL Optimization, Inc.
dmg@acm.org

CP9

An Open-Source High Performance Dual Simplex Solver

This talk presents the techniques underlying hsol, an open-source high performance dual simplex solver, as well as observations on the effectiveness of its presolve, advanced basis crash and dual edge weight pricing strategies. Specifically, an insight will be given into three features. Firstly, the solver's two novel update procedures [described in an article awarded the prize for the best paper of 2015 in Computational Optimization and Applications]. Secondly, its exploitation of task and data parallelism via suboptimization and, thirdly, its advanced basis crash. Historically, it has not been considered worthwhile to start the dual simplex method from an advanced basis following presolve and crash, but this talk will present results demonstrating its effectiveness for important classes of LP problems.

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

Qi Huangfu
FICO
qi.huangfu@gmail.com

CP9

Model Enhancement via Annotations: Extension of EMP for Lagrangian Modifications

In this work we present a tool that performs Lagrangian modifications, like the ones occurring in the context of Quadratic Support (QS) functions [Aravkin et al., arXiv 1301:4566] used as penalties, or risk measures like CVaR [Rockafellar, Optimization of conditional value-at-risk, 2000]. It also enables the modeler beyond black and white constraints [Rockafellar, Lagrange Multipliers and Optimality, 1993] This tool is implemented within the EMP framework that is available in the GAMS modeling language. A particular attention has been paid to the ease of use: the modeler specify the data necessary to encode the transformation. The model is then augmented/transformed programmatically, solved and the solution is automatically reported. The user has the choice to up to 3 different schemes to solve the modified problem. An important feature of our approach is that it fully exploits the structure of the problem. We illustrate the approach with an electricity generation problem [Philpott et al., Equilibrium, uncertainty and risk in hydro-thermal electricity systems, 2016], and fitting problems. The value of having multiple formulations is highlighted by the computational results.

Olivier Huber
Wisconsin Institute for Discovery
University of Wisconsin-Madison

ohuber2@wisc.edu

Michael C. Ferris
University of Wisconsin
Department of Computer Science
ferris@cs.wisc.edu

CP10

A Second-Order Optimality Condition with First- and Second-Order Complementarity Associated to Global Convergence of Algorithms

We develop a new notion of second-order complementarity with respect to the tangent subspace associated to second-order necessary optimality conditions by the introduction of so-called tangent multipliers. We prove that around a local minimizer, a second-order stationarity residual can be driven to zero while controlling the growth of Lagrange multipliers and tangent multipliers, which gives a new strong second-order optimality condition without constraint qualifications. We prove that second-order variants of augmented Lagrangian and interior point methods satisfy our optimality condition. Finally, we present a companion minimal constraint qualification, weaker than the ones known for second-order methods, that ensures usual global convergence results to a classical second-order stationary point.

Gabriel Haeser
University of Sao Paulo
ghaeser@ime.usp.br

CP10

Some Convergence Properties of Regularization and Penalization Schemes for MPCCs

The mathematical programs with complementarity constraints (MPCCs) have been the subject of much recent research because of their theoretical and practical applications. From a nonlinear programming point of view, MPCCs are among the most highly degenerate problems, they violate at every feasible point the Mangasarian-Fromovitz constraint qualification which is a key ingredient for stability. As a consequence, most of the well developed theory for nonlinear programming cannot be applied directly to MPCCs. Another difficulty in dealing with MPCCs is their combinatorial nature due to the complementarity constraints, which implies that the optimality conditions for MPCCs are complex and not easy to verify. Therefore, developing an efficient algorithm for MPCCs is of a great interest. In this work, we propose an approach that combines a regularization scheme with a penalty method to solve the NLP reformulation of MPCCs. The relaxation scheme is to handle the lack of regularity of these problems and the penalty method is to penalize the equality nonlinear constraints and to update the relaxation parameter associated to the complementarity constraints. We investigate the convergence properties of this approach and we propose sufficient conditions to ensure that the cluster points of the sequence of stationary points generated by this approach are feasible for MPCCs.

Abdeslam Kadrani
Institut National de Statistique et d'Economie Appliquée
INSEA
kadrania@gmail.com

Jean-Pierre Dussault
Université de Sherbrooke
Jean-Pierre.Dussault@Usherbrooke.ca

Mounir Haddou, Tangi Migot
Institut National des Sciences Appliquées de Rennes
France
mounir.haddou@insa-rennes.fr, tangi.migot@insa-rennes.fr

Emilie Joannopoulos
Insa de Rennes / Université de Sherbrooke
emilie.joannopoulos@insa-rennes.fr

CP10

On Weak Conjugacy, Augmented Lagrangians and Duality in Nonconvex Optimization

The general duality theory for convex and nonconvex optimization problems, was comprehensively studied by Rockafellar and Wets. They used augmented Lagrangian functions in general forms and presented conditions for strong duality relations. The approach developed by Azimov and Kasimbeyli is based on a duality scheme, which uses the so-called weak conjugate functions. This conjugacy is constructed with respect to superlinear conic functions defined as an augmented norm term with a linear part added, instead of only linear functions used in convex analysis. A graph of such a function is a conical surface which serves as a supporting surface for a certain class of nonconvex sets. By using the mentioned class of superlinear functions, Azimov and Kasimbeyli introduced the concept of weak subdifferential – one of the natural generalizations of the classic subdifferential of the convex analysis and derived a collection of optimality conditions and duality relations for a wide class of nonconvex optimization problems. In this paper, we present the results obtained by both approaches. The duality results formulated in general forms, are similar in both approaches. We present duality results obtained for particular cases, that is for problems with equality and inequality constraints with illustrative examples.

Refail Kasimbeyli
Anadolu University
rkasimbeyli@anadolu.edu.tr

CP10

Mathematical Programms with Equilibrium Constraints: A Sequential Optimality Condition, New Constraint Qualifications and Algorithmic Consequences

Mathematical programs with equilibrium (or complementarity) constraints, MPECs for short, is a difficult class of constrained optimization problems. The feasible set has a very special structure and violates most of the standard constraint qualifications (CQs). Thus, the standard KKT conditionsare not necessary satisfied by minimizers and the convergence assumptions of many standard methods for solving constrained optimization problems are not fulfilled. This makes it necessary, both from a theoretical and numerical point of view, to consider suitable optimality conditions, tailored CQs, and specially designed algorithms for solving MPECs. In this paper, we present a new sequential optimality condition useful for the convergence analysis for several relaxation methods for solving MPECs. We also introduce a companion CQ for M-stationary that is weaker than the recently introduced MPEC relaxed constant positive linear dependence (MPEC-RCPLD). Relations between the old and new CQs as well as the algorith-

mic consequences will be discussed.

Alberto Ramos
Federal University of Parana
albertoramos@ufpr.br

CP10
Generalized Derivatives of Nonlinear Programs

Parametric sensitivities are obtained for NLPs exhibiting active set changes. By reformulating the KKT system as a nonsmooth equation system using any suitable nonlinear complementarity problem function, sensitivity information is furnished under conditions implied by, for example, a KKT point satisfying LICQ and strong second-order sufficiency. To accomplish this, a piecewise differentiable implicit function theorem that describes generalized derivative information is presented, which takes advantage of newly-found theory in the form of the lexicographic directional derivative. The sensitivity system found is a nonsmooth equation system that admits primal and dual sensitivities as its unique solution and recovers the classical results set forth by Fiacco and McCormick in the absence of active set changes. The results are computationally relevant, thanks to a recently developed nonsmooth vector forward mode of automatic differentiation; algorithms are detailed for calculating the sensitivity systems unique solution in a tractable way.

Peter G. Stechlinski
Process Systems Engineering Laboratory
Massachusetts Institute of Technology
pstechli@mit.edu

Kamil Khan
Argonne National Laboratory
Mathematics and Computer Science Division
kamil@alum.mit.edu

Kamil Khan, Paul I. Barton
Massachusetts Institute of Technology
Department of Chemical Engineering
kamil@alum.mit.edu, pib@mit.edu

CP10
On a Conjecture in Second-Order Optimality Conditions

In this work we deal with optimality conditions that can be verified by a nonlinear optimization algorithm, where only a single Lagrange multiplier is available. In particular, we deal with a conjecture formulated in [R. Andreani, J.M. Martnez, M.L. Schuverdt, “On second-order optimality conditions for nonlinear programming”, Optimization, 56:529–542, 2007], which states that whenever a local minimizer of a nonlinear optimization problem fulfills the Mangasarian-Fromovitz Constraint Qualification and the rank of the set of gradients of active constraints increases at most by one in a neighborhood of the minimizer, a second-order optimality condition that depends on one single Lagrange multiplier is satisfied. This conjecture generalizes previous results under a constant rank assumption or under a rank deficiency of at most one. In this work we prove the conjecture under the additional assumption that the jacobian matrix has a smooth singular value decomposition and we review previous attempts to solve it.

Daiana S. Viana
Avenida Prof. Almeida Prado1280, Butantã, São Paulo (SP)
CEP 05508-900
doriviana@hotmail.com

Roger Behling
Federal University of Santa Catarina
gerbehling@gmail.com

Gabriel Haeser
University of Sao Paulo
ghaeser@ime.usp.br

Alberto Ramos
Federal University of Parana
albertoramos@ufpr.br

CP11
Set-Valued Derivatives and Subdifferentials Based on Differences of Convex Sets

Arithmetic set operations for convex compact sets are an essential part of many set-valued numerical methods, e.g., in the calculation of reachable sets of control systems, in the interpolation of set-valued maps or in problems of set optimization. While the Minkowski sum is commonly accepted, there is a variety of convex difference notions, e.g., the geometric difference by Hadwiger/Pontryagin and the Demyanov difference. Alternatively, the set of convex compact subsets of R^n can also be embedded into vector spaces via pairs of convex sets, limits of differences of support functions and directed sets. Directed sets in R^n are described via a scalar continuous function and a lower-dimensional directed set, both parametrized by unit vectors. The embedding of a convex compact set into the space of directed sets suggests an obvious way for a realization in computational methods. Via directions attached to generalized supporting faces, a difference of directed sets being inverse to the embedded Minkowski sum can be introduced. From the visualized difference (either a convex or nonconvex subset of R^n), other convex set differences can be recovered. Based on this approach various notions of Lipschitz and differentiable set-valued maps and a new nonconvex subdifferential with exact calculus rules will be sketched. A visual comparison to the Dini, Michel-Penot, Clarke and Mordukhovich subdifferential is given offering a geometrical way to compute them.

Robert Baier
Department of Mathematics, University of Bayreuth, Germany
robert.baier@uni-bayreuth.de

CP11
Extending the Applicability of the Global Search Method DIRECT

Don Jones DIRECT algorithm for global optimization fills an interesting niche in the derivative-free-optimization taxonomy, with both significant advantages and significant limitations. We present a generalized version of this algorithm that allows mixed continuous/discrete search spaces in asynchronous parallel computing environments. The method can be effective for local, global, and multi-local optimization, for finding Pareto sets associated with multiple objectives, and for problems with hidden constraints.

Richard G. Carter

DNV GL
richard.carter@dnvgl.com

CP11

A Superlinearly Convergent Smoothing Newton Continuation Algorithm for Variational Inequalities Over Definable Sets

We use the concept of barrier-based smoothing approximations introduced by Chua and Li Li [C. B. Chua and Z. Li, *A barrier-based smoothing proximal point algorithm for NCPs over closed convex cones*, SIAM J. Optim., 23 (2013), pp. 745–769] to extend the smoothing Newton continuation algorithm of Hayashi at el [S. Hayashi, N. Yamashita, and M. Fukushima, *A combined smoothing and regularization method for monotone second-order cone complementarity problems*, SIAM J. Optim., 15 (2005), pp. 593–615] to variational inequalities over general closed convex sets X. We prove that when the underlying barrier has a gradient map that is definable in some o-minimal structure, the iterates generated converge superlinearly to a solution of the variational inequality. We further prove that if X is proper and definable in the o-minimal structure $\mathbf{R}_{\mathrm{an}}^{\mathbf{R}_{\mathrm{alg}}}$, then the gradient map of its universal barrier is definable in the o-minimal expansion $\mathbf{R}_{\mathrm{an,exp}}$. Finally, we consider the application of the algorithm to complementarity problems over epigraphs of matrix operator norm and nuclear norm, and present preliminary numerical results.

Chek Beng Chua
Nanyang Technological University
cbchua@ntu.edu.sg

Le Thi Khanh Hien
National University of Singaore
iseltkh@nus.edu.sg

CP11

Basis Function Selection in Approximate Dynamic Programming

Many complex problems arising in business, health care, and transportation can be modeled as sequential decision making problems under uncertainty, meaning that a decision maker has to make decisions periodically while some random events unfold over time. These problems can be conveniently modeled in the form of dynamic programs. However, for many practical problems, this formulation suffers from the curse of dimensionality, making exact calculation of the optimal policy intractable. In order to overcome this issue, approximate dynamic programming (ADP) methods have been developed to find an approximate optimal solution. A cornerstone of many ADP algorithms is defining a set of basis functions (or an approximation architecture) for approximating the cost-to-go function. Currently, the choice of basis functions requires prior expert knowledge about the problem, and is considered as more of an art than a science. We present, study, and apply a novel algorithm that automates generation of basis functions by efficiently selecting a subset of functions from a large pool of potential basis functions, and updating this set as more information becomes known about the problem. The benefit of our algorithm is twofold: first, it reduces the burden to come up with a well-informed set of basis functions that requires significant prior knowledge about the problem; second, since many potential candidates are considered for basis functions, it improves the quality of the approximation.

Alireza Sabouri
Haskayne School of Business
University of Calgary
alireza.sabouri@haskayne.ucalgary.ca

CP11

A Primal Majorized Newton-CG Augmented Lagrangian Method for Large-Scale Linearly Constrained Convex Programming

In this paper, we propose a primal majorized semismooth Newton-CG augmented Lagrangian method for large-scale linearly constrained convex programming problems, especially for some difficult problems which are nearly degenerate. The basic idea of this method is to apply the majorized semismooth Newton-CG augmented Lagrangian method to the primal convex problem. And we take two special nonlinear semidefinite programming problems as examples to illustrate the algorithm. Furthermore, we establish the global convergence and the iteration complexity of the algorithm. Numerical experiments demonstrate that our method works very well for the testing problems, especially for many ill-conditioned ones.

Chengjing Wang
Southwest Jiaotong University
renascencewang@hotmail.com

Peipei Tang
Zhejiang University City College
tangpp@zucc.edu.cn

CP11

Multi-Criteria Quadratic Programming Problems with Application to Portfolio Selection

Mathematical optimization is one of the most important tools for solving practical decision-making problems. It is widely applied in a variety of areas, including finance, network design and operation, supply chain management and engineering. It is worth noting that most problems in the real world are concerned with one or more criteria. However, there has not been much research into multi-objective quadratic programming problems with a norm regularization. In this paper, we investigate the tri-criteria quadratic programming problems with a cardinality constraint for portfolio management. The cardinality constrained investment strategy arises naturally due to the presence of various forms of market friction, such as transaction costs and management fees, or even due to the consideration of mental cost. In the theoretical framework, we study the optimality conditions for the proposed tri-criteria quadratic programming problem and show that the λ-space can be partitioned into some polyhedra and that on each polyhedron, the corresponding solution can be found by solving one scalarized quadratic programming problem. We will also discuss the properties of the Pareto optimal solution set of the problems. Acknowledgements: The work described in this paper was partially supported by grants from the Research Grants Council of the HKSAR, China (UGC/FDS14/P03/14 and UGC/FDS14/P02/15).

Carisa K. W. Yu
Department of Mathematics and Statistics
Hang Seng Management College

carisayu@hsmc.edu.hk

CP12

Density Estimation with Total Variation Penalized Maximum Likelihood Estimation

We introduce a new method for nonparametric density estimation from empirical data, which is conducive to scenario-based stochastic optimization methods. By penalizing a maximum likelihood approach with a total variation term, we avoid overfitting and the "dirac curse". Though the corresponding optimization problem is variational, we show that it reduces to a finite dimensional one. We investigate efficient methods for solving this convex problem, and then focus on the statistical properties of this density estimation method and its applications to stochastic programming.

Robert Bassett, James Sharpnack
University of California Davis
rbassett@math.ucdavis.edu, jsharpna@ucdavis.edu

CP12

Differential Evolution for Solving Maximum Clique Problem

The maximum clique problem (MCP) finds many important applications in different contexts. MCP is a classical NP-hard problem in combinatorial optimization, but it is still hard to get satisfactory results, especially when the number of vertices exceeds 200. In this talk, we apply Differential Evolution (DE) algorithm to MCP. DE is one of stochastic function minimizer and gets attention from many researches due to its effectiveness. To employ DE algorithm, we converted MPC problem into a standard quadratic optimization problem over a unit simplex. When DE algorithm generates the next generation by randomization steps, there is a possibility that the next candidate can get out the feasible region. However, the projection onto the feasible set can be done by the simple structure of the unit simplex. We can reduce the computation time of the projection by combining existing methods specialized for the unit simplex. Numerical results showed a favorable performance on most benchmark instances, some of which DE got better results than other existing algorithm. By choosing better parameters, we can further improve the results. DE can be considered as a very potential algorithm for solving MPC.

Hui Fang
Tokyo Institute of Technology
houkifh@gmail.com

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

CP12

High-Frequency Trading in Risk-Averse Portfolio Optimization with Higher-Order Risk Measures

This study examines the application of risk-averse optimization techniques to high-frequency trading (HFT) in portfolio management. First, I develop efficient clustering methods for scenario tree construction. Then, I construct a two-stage stochastic programming problem with higher-order conditional measures of risk, which is used to re-balance the portfolio on a rolling horizon basis, with transaction costs included. Finally, I present an extensive simulation study on both interday and high-frequency intraday real-world data of the methodology.

Sitki Gulten
Stockton Universtiy
sitki.gulten@stockton.edu

CP12

Improving Consistency in Intertemporal Decisions via Stochastic Optimization and Conditioning Coordination

Optimization models have been used since long to support operational planning in several areas and typically in different time frames: long-term strategic decisions, mid term tactical decisions and short term operational decisions. In general, longer term decisions impose constraints to the decision process in shorter horizons and one would like to compute efficient shorter term decisions. However, many times there are inconsistencies between these models, specially due to various sources of uncertainty, the dynamics of the process and different degrees of aggregation. We explore the question of how to improve consistency and address this problem using stochastic approaches and robust optimization. We also show how the implicit geometry and conditioning of the underlying problems might affect consistency, and we propose a coordination approach using condition measures instead of cost measures. We show some simulated results as well as applications to some real problems in the forest industry and also for a problem in the management of health systems.

Jorge R. Vera
Pontificia Universidad Católica de Chile
jvera@ing.puc.cl

Alfonso Lobos
University of California, Berkeley
alobos@berkeley.edu

Ana Batista
Pontificia Universidad Católica de Chile
abatista@uc.cl

CP13

Splitting Methods in Penalized Zero-Variance Discriminant Analysis

We propose improved methods for solving the sparse zero-variance discriminant analysis problem based on the alternating direction method of multipliers. Our approach avoids the expensive calculation of between and within-class scatter matrix typically used in linear discriminant analysis, and instead uses only the data matrix to compute discriminative directions. We present per-iteration complexity analyses and numerical simulations which establish improvement upon classical approaches for zero-variance discriminant analysis and high-dimensional LDA, as well as convergence results for our approach.

Brendan P. Ames
University of Alabama
Department of Mathematics
bpames@ua.edu

CP13

Fast Monte Carlo Algorithms for Tensors: Approx-

imating Tensor Multiplication and Decomposition

Motivated by applications in which the data may be formulated as a tensor, we consider methods for several common tensor factorization problems. These methods resolve many of the key algorithmic questions regarding the *randomized* tensor multiplication, low-rank decomposition, and nuclear norm minimization using a small number of lateral and/or horizontal slices of the data tensor. These methods make more efficient use of computational resources, such as the computation time, random access memory, and the number of passes over the data. Inspired by the recent development of randomized matrix factorization methods with rich theory and computational complexity, we propose the first polynomial time algorithms for the low-rank tensor(noisy) approximations that come with relative error guarantees. All theoretical and practical results are presented for third order tensors. However, they can be easily extended to the order-$p(p > 3)$ case. We conclude with some experimental results that show the promise of our approach for analyzing large-scale datasets.

Michailidis George, Davoud Ataee Tarzanagh
University of Florida
gmichail@ufl.edu, Tarzanagh@ufl.edu

CP13

Online Convolutional Dictionary Learning

Dictionary learning is a significant topic of signal processing. It is modeled by a nonconvex optimzation problem. Convolutional dictionary learning, a structured dictionary learning model which deals with convolutional signals, is powerful for inverse problems of images, such as denoising and super-resolution. However, traditional batch convolutional dictionary learning leads to huge memory costs, since all the training images are handled simultaneously and kept in memory. Thus, batch learning scheme can only adopts training set with limited size. Online learning deals with a single or a small number of training images in one iteration, and then accumulates the information step by step. Consequently, it can adopt large training set, even with infinitely size. There are existing works for online dictionary learning without convolution, but online convolutional dictionary learning is novel. The key issue is that the structures of sparse coding in standard signals and convolutional signals are different. Its not trivial to extend online scheme from standard dictionary learning to convolutional structures. We use down-sampling and mask-simulation techniques to solve this problem and get good results. Moreover, some techniques in optimization algorithms, such as coordinate descent, are used to accelerate our method.

Jialin Liu
Mathematics Department,
University of California, Los Angeles (UCLA)
danny19921123@gmail.com

Cristina Garcia-Cardona
CCS Division, Los Alamos National Laboratory
Los Alamos, NM
cristina.cgarcia@gmail.com

Brendt Wohlberg
Los Alamos National Laboratory
brendt@lanl.gov

Wotao Yin
University of California, Los Angeles
wotaoyin@math.ucla.edu

CP13

Sparsity-Constrained Gaussian Graphical Models

Gaussian graphical models (GGM) are graph matrices used to model conditional dependences between variables in a network. Our goal is to learn a sparse GGM by enforcing cardinality constraints for either interpretability of model or to avoid overfitting, which can be used in anomaly detection of sensor networks to detect faulty behaviors, avoid downtime of assets. We have developed an efficient algorithm incorporating projected gradients and an active-set method with Newton subproblems to solve the nonlinear optimization problem.

Dzung Phan
IBM T.J. Watson Research Center
phandu@us.ibm.com

Matt Menickelly
Lehigh University
mjm412@lehigh.edu

CP13

Numerical Identification of Sparse Chemical Reaction Networks

We are concerned with the numerical identification of sparse chemical reaction networks from noisy time series data of the species concentrations. After choosing suitable kinetics, the dynamics of the reaction network are usually modeled by a parameter-dependent system of ordinary differential equations. In classical approaches of system identification, after a discretization in time, the unknown parameters are estimated via least-squares approximations. It is well-known that such strategies yield inferior results in the case of a low network connectivity. In the last years, several recovery techniques have been proposed in the literature that promote sparse networks in the reconstructions. We report on the application of nonsmooth Tikhonov regularization with sparsity-promoting penalties to system identification problems. The optimality conditions are solved by globally convergent semismooth Newton techniques which were developed recently.

Thorsten Raasch
University of Mainz
Institute of Mathematics
raasch@uni-mainz.de

CP13

Proximal DC Algorithm for Sparse Optimization

Many applications such as in signal processing, machine learning and operations research seek sparse solutions by adopting the cardinality constraint or rank constraint. We formulate such problems as DC (Difference of two Convex functions) optimization problems and apply DC Algorithm (DCA) to them. While the DCA has been widely used for this type of problems, it often requires a large computation time to solve a sequence of convex subproblems. Our algorithm, which we call Proximal DC Algorithm (PDCA), overcomes this issue of the ordinary DCA by employing a special DC decomposition of the objective function. In PDCA, closed-form solutions can be obtained for the convex subproblems, leading to efficient performance in numerical experiments. We also discuss the theoretical as-

pects: PDCA can be viewed as a nonconvex variant of the proximal gradient methods (PGM), which provides an insight on the relation between PGM and DCAs.

Akiko Takeda
The Institute of Statistical Mathematics
atakeda@ism.ac.jp

Jun-ya Gotoh
Chuo University
jgoto@indsys.chuo-u.ac.jp

Katsuya Tono
The University of Tokyo
katsuya_tono@mist.i.u-tokyo.ac.jp

CP14

A Semidefinite Programming Approach for Energy Network Planning

This talk presents a mixed-integer quadratic formulation for transmission expansion planning with an AC network model. The model is solved by combining a semidefinite programming relaxation with valid inequalities in a branch-and-cut framework. Computational results are presented to show the performance of the proposed method.

Bissan Ghaddar
IBM Research
bghaddar@uwaterloo.ca

CP14

Acyclic Orientations and Nonlinear Flow Problems

We consider optimization problems for natural gas networks with complex topologies and many cycles like the German one. The existence of cycles implies that the flow direction of the gas is not known beforehand, which leads to weaker relaxations for the MINLP optimization problems arising in planning and operating gas networks. To improve these models, we study the underlying nonlinear flow problem from a combinatorial perspective. In fact, the flow in gas and water networks is driven by so-called potentials, which implies that the resulting flows are acyclic in the following sense: If each network arc is oriented in the direction of flow over this arc, then there is no directed cycle in the resulting network. We exploit the interplay of this acyclicity property with the network flow structure to construct stronger models and present some computational results.

Benjamin Hiller
Zuse Insitut Berlin
hiller@zib.de

CP14

Simultaneous Truss Topology - and Static Output Feedback Controller Design via Nonlinear Semidefinite Programming

We consider the problem of static output feedback controller design via the H_∞-control problem. The basic idea of the H_∞-control problem is to find an optimal controller that minimizes the H_∞-norm of the closed loop transfer function. With the help of the Bounded-Real Lemma the H_∞-control problem can be formulated as a nonlinear semidefinite programming problem. Moreover, we combine the H_∞-control problem with a topology optimization for truss structures under uncertain dynamic loads. The main advantage of our approach is that we can simultaneously optimize the topology of the mechanical system and the design of the static output feedback controller. We solve the resulting nonlinear semidefinite programming problem using a sequentially semidefinite programming approach. The considerations are complemented by numerical results for truss topology design under uncertain dynamic loads.

Anja Kuttich
TU Darmstadt
kuttich@mathematik.tu-darmstadt.de

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

CP14

Stochastic Planning of Environmentally Friendly Telecommunication Networks Using Column Generation

We present the network planning problem of an LTE cellular network with switchable base stations, ie. where the configuration of the antennae is dynamic, depending on demand. This is formulated as a two-stage stochastic program under demand uncertainty with integers in both stages. We develop its Dantzig-Wolfe reformulation and solve it using column generation. Preliminary results confirm the efficiency of the approach. Interestingly, the reformulated model often has a tight LP-gap.

Joe Naoum-Sawaya
IBM Research
jnaoumsa@uwaterloo.ca

CP14

Gradient Multicut: A Second Derivative Potts Model for Image Segmentation Using MILP

Unsupervised image segmentation and denoising are two fundamental tasks in image processing. Usually, graph based models such as graph cuts or multicuts are used for globally optimum segmentations and variational models are employed for denoising. Our approach addresses both problems at the same time. A (depth) image can be seen as a function $y : P \to \mathbb{R}$ giving the (depth) intensity of pixels located on a finite 2-dimensional grid P. The segmentation problem is tackled by fitting a piecewise linear function $f : P \to \mathbb{R}$ to y minimizing $\sum_{p\in P} |f(p) - y(p)|$ with an additional ℓ_0 regularization term to count the number of variations (with respect to segments) per row and column. Previous attempts are usually based on continuous or convex relaxations. We propose a novel MILP formulation of a second derivative Potts model, where binary variables are introduced to directly deal with the ℓ_0 norm. The model approximates the values of pixels in a segment by an affine plane and incorporates multicut constraints to enforce the connectedness of segments. As a by-product the image is denoised. Our approach could also be interpreted as nonparametric discontinuous piecewise linear fitting in 2D. To the best of our knowledge, it is the first mathematical programming model for finding globally optimum solutions. Numerical experiments demonstrate its performance.

Ruobing Shen
Heidelberg University
Institute of Computer Science

ruobing.shen@informatik.uni-heidelberg.de

Stéphane Canu
INSA de Rouen
Normandie Université
scanu@insa-rouen.fr

Gerhard Reinelt
University of Heidelberg
Institute of Computer Science
gerhard.reinelt@informatik.uni-heidelberg.de

CP14

Disjoint K-Path Min-Sum Approach Towards Robust Power Network Design

An electric power supplier needs to build a transmission line between two jurisdictions. Ideally, the power line configuration would optimize some utility, e.g., minimize the cost. Due to reliability concerns, the developer has to install not just 1, but 2 lines, separated by a certain distance from one to another, so that if one line fails, the electricity still gets delivered. Similar situations arise in other applications like communication networks, etc. We investigate the above problem, with emphasis on 2-path instances where the network topology representations are highly granular, resulting in large scale models. Specifically, we contrast the performance of a mixed-integer programming (MIP) model with that of a new graph-based model. On the MIP side, we observe that unlike the min-cost flow formulation of the shortest path, a similar MIP model of the disjoint 2-path min-sum problem suffers from loss of the total unimodularity. On the graph-theory side, we classify the exact problem as NP-complete. At the same time, we propose a polynomial approximation scheme, which appears to avoid local-search pitfalls encountered by other heuristics. Under two mild practical assumptions, the latter scheme yields an optimal solution to the problem, and appears to be extremely efficient computationally. We substantiate our finding with numerical experiments using real-life data from the Northeast region of Alberta, Canada, with network graphs nearing a hundred thousand nodes.

Yuriy Zinchenko
University of Calgary
Department of Mathematics and Statistics
yzinchen@ucalgary.ca

Haotian Song
Stern School of Business, New York University
, KMC 8-154, 44 W 4th St, New York, NY
hs2798@stern.nyu.edu

CP15

Copositive Optimization on Infinite Graphs

Combinatorial optimization problems can be formulated as conic convex optimization problems. Especially NP-hard problems can be written as copositive programs. In this case the hardness is moved into the copositivity constraint. We will show how to lift this theory to infinite dimension and we study operators in Hilbert spaces instead of matrices. For that purpose we develop a new theory of semidefinite and copositive optimization in infinite dimensional spaces. In this context we discuss some properties which are equivalent to the ones in finite dimension and we also point out differences. Understanding these properties is essential for developing solution approaches for problems of that kind.

Claudia Adams
Departement of Mathematics
Trier University
adamsc@uni-trier.de

Mirjam Dür
Department of Mathematics
University of Trier
duer@uni-trier.de

Leonhard Frerick
Departement of Mathematics
Trier University
frerick@uni-trier.de

CP15

The Cone of Positive Type Measures on a Compact Group

The cone of positive type measures on a compact group can be regarded as the natural dual of the cone of continuous positive type functions, and this duality can be seen as an infinite analog of the self-duality of the cone of positive semidefinite matrices. We present some details of this duality, and as an application we explain how the usual primal and dual semidefinite programming definitions of the Lovasz theta-function can be extended to accommodate infinite Cayley graphs over compact groups. From these extensions, one can deduce the Delsarte bounds for spherical codes, the Fourier analysis upper bounds of Vallentin and Oliveira for the measurable chromatic number of Euclidean space, and the upper bounds of DeCorte and Pikhurko for the Witsenhausen problem on orthogonal-free subsets of the unit sphere. Duality for these extensions will be investigated. Even though the known strong duality proofs for the theta-function fail in this setting, we can prove zero duality gap from Fourier analysis; attainment of the optimal value sometimes fails.

Evan Decorte
McGill University
pevdecorte@gmail.com

CP15

A Factorization Method for Completely Positive Matrices

A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is called completely positive if it can be factorized as $A = BB^T$ for some entrywise nonnegative matrix $B \in \mathbb{R}_+^{n\times k}$. These matrices play a role in optimization, since Burer (2009) showed in a seminal paper that any optimization problem with quadratic objective, linear constraints and binary variables can be equivalently formulated as a linear problem over the cone of completely positive matrices. Given a matrix completely positive matrix A, it is highly nontrivial to actually find a factorization $A = BB^T$ of the above form. Being able to compute such a factorization would help recover the solution of the underlying quadratic or combinatorial problem. In this talk, we present a factorization method for general completely positive matrices which is heuristic in nature but in our extensive numerical tests gave very fast results for almost all test instances.

Mirjam Duer
University of Trier
duer@uni-trier.de

Patrick Groetzner
Univeristy of Trier
groetzner@uni-trier.de

CP15

Convergence Rates of the Moment-Sum-of-Squares Hierarchy for Volume Approximation of Semialgebraic Sets

We derive bounds on the convergence rate of the moment-sum-of-squares hierarchy of semidefinite programs for approximating the volume of a compact basic semialgebraic set. We show that the rate is O(1/log log d) where d is the degree of the moments or polynomial sum-of-squares.

Didier Henrion
LAAS-CNRS, University of Toulouse, France
henrion@laas.fr

Milan Korda
University of California at Santa Barbara
USA
korda.m@gmail.com

CP15

A Bound on the Carathéodory Number

The Carathéodory number $\kappa(K)$ of a pointed closed convex cone K is the minimum integer k for which every element of K can be written as a sum of k elements belonging to extreme rays. The dimension of K is an upper bound of $\kappa(K)$ as a consequence of Carathéodory's Theorem. In this talk we provide a smaller upper bound $\ell(K) - 1$ of $\kappa(K)$, where $\ell(K)$ is the length of the longest chain of nonempty faces of K, thus tying the Carathéodory number with a key quantity that appears in the analysis of facial reduction algorithms. This bound is tight for several families of cones, which include symmetric cones, the so-called smooth cones, and rank one generated spectrahedral cones. We furnish a new proof of a result by Güler and Tunçel which states that the Carathéodory number of a symmetric cone is equal to its rank. We also extend Hildebrand's result for the Carathéodory number of rank one generated spectrahedral cones to Jordan algebra. We give a simple example showing that our bound fail to be sharp. Finally, we connect our discussion to the notion of cp-rank for completely positive matrices.

Masaru Ito
Tokyo Institute of Technology
ito.m@math.cst.nihon-u.ac.jp

Bruno Lourenco
Seikei University
lourenco@st.seikei.ac.jp

CP15

Conic Program Certificates from ADMM/Douglas-Rachford Splitting

Many practical optimization problems are formulated as conic programs, a current active area of research. One challenge with conic programs is that one often does not know, a priori, whether there is a solution to the conic program; it may be infeasible, unbounded or has a finite but unattainable optimal value. Solvers like *SeDuMi* and *SDPT3* provide certificates for strong infeasibility, but they struggle with weak infeasibility. To resolve this, we propose a method that finds (1) a solution when one exists (2) a certificate of strong/weak infeasibility when the problem is strongly/weakly infeasible (3) an unbounded direction when there is one and (4) a sufficient condition for identifying programs with finite but unattainable optimal values. The method is based on Douglas-Rachford Splitting (ADMM), and we establish the efficacy of it through theoretical analysis and numerical experiments.

Yanli Liu, Ernest Ryu
University of California, Los Angeles
Department of Mathematics
yanli@math.ucla.edu, eryu@math.ucla.edu

Wotao Yin
University of California, Los Angeles
wotaoyin@math.ucla.edu

CP16

Worst-Case Complexity Analysis of Convex Nonlinear Programming

We will review recent studies on the worst-case complexity analysis of derivative-based and derivative-free methods. Within convex smooth unconstrained setting, we present a unified framework for worst-case complexity analysis of both first and second-order methods when deriving the size of the gradient below some given threshold is desired. We then present two algorithms along with their complexity analysis for the minimization of nonsmooth convex constrained problems.

Rohollah Garmanjani
CMUC, Department of Mathematics, University of Coimbra
nima@mat.uc.pt

CP16

Alternating Projection on Convex Sets and Manifolds

Finding a feasible point for an optimization problem is equivalent to finding a point in the intersection of finitely many constraint sets. For convex sets, alternating projection is a common method to solve this problem. This has recently been extended to alternating projection on manifolds. Under certain conditions, there are results on the convergence speed in terms of the angle between the sets. In this talk I will give a short survey of these results and extend the given ideas to the case of alternating projection onto different types of sets.

Patrick Groetzner
Univeristy of Trier
groetzner@uni-trier.de

Mirjam Dür
Department of Mathematics
University of Trier
duer@uni-trier.de

CP16

A Two-Phase Algorithm for Large-Scale Qplogdet Optimization Problem

In this paper, we present a two-phase algorithm to solve a large-scale nonlinear semidefinite programming problem whose objective function is a sum of a convex quadratic

function and a log-determinant term (QPLogdet). In phase I, we adopt an inexact symmetric Gauss-Seidel (sGS) technique based alternating direction method of multipliers (ADMM)-type method to obtain a moderately accurate solution or generate a reasonably good initial point. In Phase II, we design an inexact accelerated block coordinate descent (ABCD) based augmented Lagrangian method (ALM) to obtain a highly accurate solution, where the inexact sGS techinque and the semismooth Newton-CG method are employed to solve the inner problem at each iteration. Numerical experiments demonstrate that our two-phase algorithm is efficient and robust.

Tang Peipei
Zhejiang University City College
tangpp@zucc.edu.cn

Wang Chengjing
Southwest Jiaotong University
renascencewang@hotmail.com

CP16

Large Gaps Between Gauss-Seidel Type Methods and Randomized Versions

A simple yet powerful idea for solving large-scale computational problems is to iteratively solve smaller subproblems and this idea appeared in, e.g., Gauss-Seidel (G-S), Kaczmarz, coordinate descent (CD), and ADMM. We prove for the first time that for all these methods, there are large gaps between the deterministic cyclic versions and the randomized versions. First we show an $O(n^2)$ gap in the convergence rate for CD/G-S/Kaczmarz methods. There are some examples showing cyclic CD (C-CD) can be much slower than randomized coordinate descent (R-CD), but there also exist many practical examples where C-CD is faster. We show that C-CD indeed can be $O(n^2)$ times slower than R-CD in the worst case, by establishing a lower bound that matches the upper bound. One difficulty is that the spectral radius of a non-symmetric iteration matrix does not necessarily constitute a lower bound for the convergence rate. Second we show a gap between divergence and convergence for ADMM. In particular, although cyclic multi-block ADMM was recently found to be possibly divergent, we show that RP-ADMM (randomly permuted ADMM) converges in expectation for solving linear systems (can be extended to quadratic objective). Our analysis reveals that random permutation has a symmetrization effect that nicely changes the spectrum of the iteration matrix. We believe RP-ADMM is potentially a competitive algorithm for large-scale linearly constrained problems.

Ruoyu Sun
UIUC
Facebook AI Research
sundirac@gmail.com

CP16

Fully Adaptive Admm for Deterministic and Stochastic Non-Smooth Optimization

Recently, alternating direction method of multipliers (ADMM) has received tremendous interests for solving numerous problems in machine learning, statistics and signal processing. A variety of algorithms of ADMM with fast convergence rates (e.g., linear rate) have been developed under smoothness, strong convexity and other regularity conditions. However, for general convex optimization problems (e.g., risk minimization with a non-smooth loss and a non-smooth regularizer), the best known iteration complexity of deterministic and stochastic ADMM is $O(1/\epsilon)$ and $O(1/\epsilon^2)$, respectively. In this paper, we propose fully adaptive ADMM for general convex optimization without smoothness and strong convexity assumptions but under a generic local error bound condition to enjoy faster convergence. In particular, we show that the proposed deterministic ADMM enjoys an improved iteration complexity of $\widetilde{O}(1/\epsilon^{1-\theta} log(1/\epsilon))$, where $\theta \in (0, 1]$ is a constant in the local error bound condition - a property of the objective function, and the proposed stochastic ADMM enjoys an iteration complexity of $\widetilde{O}(1/\epsilon^{2(1-\theta)})$, which is lower than $O(1/\epsilon^2)$ of standard stochastic ADMM. Without smoothness and strong convexity assumptions, we demonstrate that the proposed ADMMs can achieve linear convergence for minimizing a non-smooth loss with generalized lasso regularizers.

Yi Xu, Mingrui Liu, Qihang Lin, Tianbao Yang
University of Iowa
yi-xu@uiowa.edu, mingrui-liu@uiowa.edu, qihang-lin@uiowa.edu, tianbao-yang@uiowa.edu

CP17

An Algorithm with Infeasibility Detection for Linearly Constrained Optimization Problems

When solving a constrained optimization problem, there is a balance between moving toward a stationary point and satisfying the constraints. One method for minimizing violation of the constraints is to generate a sequence of iterates $\{x_n\}$ by applying Newton's method to the equality constraints, $c(x)$. That is, $x_{n+1} = x_n + d_n$ where

$$d_n = \arg\min \left\{ \|z - d_{n-1}\|^2 : \nabla c(d_{n-1})(z - d_{n-1}) + c(d_{n-1}) = 0,\ z \geq 0 \right\}.$$

However, even when the original constrained optimization problem is feasible, the Newton step may be infeasible. In this presentation, we discuss an algorithm with infeasibility detection for

$$\min \left\{ \|z\|^2 : Az = b,\ z \geq 0 \right\}$$

based on recent work by Burke, Curtis, Johnson, Robinson, Wächter, and Wang.

James D. Diffenderfer
University of Florida
jdiffen1@ufl.edu

William Hager
University of Florida
Department of Mathematics
hager@ufl.edu

CP17

Convex Relaxations with Second Order Cone Constraints for Nonconvex Quadratically Constrained Quadratic Programming

We present new convex relaxations for nonconvex quadratically constrained quadratic programming (QCQP) problems. Since the basic semidefinite programming relaxation is often too loose for general QCQP, recent research has focused on strengthening convex relaxations using valid linear or second order cone (SOC) inequalities. In this paper, we construct valid second order cone constraints for nonconvex QCQP and reduce the duality gap using these valid constraints. Specifically, we decompose and relax the nonconvex constraints to two SOC constraints and then linearize the products of the SOC constraints and linear constraints to achieve some new valid constraints. Moreover, we introduce and generalize two recent techniques for generating valid inequalities to further enhance our method. We demonstrate the efficiency of our results with numerical experiments.

Rujun Jiang
Rm810C, ERB, CUHK, Shatin, N.T., Hong Kong
rjjiang@se.cuhk.edu.hk

Duan Li
The Chinese University of Hong Kong
dli@se.cuhk.edu.hk

CP17

Differentiable McCormick Relaxations for Deterministic Global Optimization

Deterministic methods for nonconvex optimization require lower bounding information that is typically obtained using convex underestimators of the objective function and constraints. McCormick's relaxation scheme [Math. Program., 10 (1976): 147-175] is an established, efficient method for generating and computing appropriate convex underestimators of factorable functions, and was recently generalized and improved by Tsoukalas and Mitsos [J. Glob. Optim., 59 (2014): 633-662]. However, both of these schemes may yield nonsmooth underestimators, necessitating the use of nonsmooth optimization solvers to evaluate lower bounds. Such solvers exhibit poorer convergence properties and performance than their smooth counterparts. This presentation describes conditions under which the Tsoukalas-Mitsos scheme produces $\mathcal{C}^1$ relaxations, and shows how these conditions may be satisfied for any factorable function. The result is a $\mathcal{C}^1$ variant [J. Glob. Optim., in press] of McCormick's relaxations that retain the computational benefits of established schemes: the new relaxations may be evaluated automatically, cheaply, and accurately, apply even to nonsmooth functions, and converge rapidly to the relaxed function as the considered domain shrinks. Gradients may be evaluated using standard automatic differentiation techniques. A C++ implementation is discussed, along with extensions to implicit functions and dynamic systems.

Kamil Khan
Argonne National Laboratory
Mathematics and Computer Science Division
kamil@alum.mit.edu

Kamil Khan
Massachusetts Institute of Technology
Department of Chemical Engineering
kamil@alum.mit.edu

Harry Watson
MIT
hwatson@mit.edu

Paul I. Barton
Massachusetts Institute of Technology
Department of Chemical Engineering
pib@mit.edu

CP17

SQP Method for Constrained Vector Optimization Problem

In this paper a descent scheme is developed to solve nonlinear vector optimization problem with inequality constraints. At every iteration of the scheme, a decent direction is determined by solving a single objective quadratic subproblem using linear approximation of the objective functions as well as constraints and a quadratic restriction. A non-differentiable penalty function (l_∞) is used to restrict the constraint violations. It is proved that the descent sequence generated in this process converges to a KKT point under Mangaserian Fromovitz constraint qualifications. Global convergence of this scheme is justified with some mild assumptions. The paper contains numerical support for the theoretical developments.

Md Abu T. Ansary
Indian Institute of Technology Kharagpur
md.abutalha2009@gmail.com

Geetanjali Panda
Indian Institute of Technology, Kharagpur, India
geetanjali@maths.iitkgp.ernet.in

CP17

On the Fruitful Use of Damped Techniques Within Nonlinear Conjugate Gradient Methods

In this work we deal with the use of damped techniques within Nonlinear Conjugate Gradient (NCG) methods for large scale unconstrained optimization. Damped techniques were introduced by Powell for SQP Lagrangian BFGS methods and recently extended by Al-Baali to quasi-Newton methods [Al-Baali, Damped techniques for enforcing convergence of quasi-Newton methods, *Optim Method Softw*, 29, 919–936, 2014]. We consider their use for possibly improving the efficiency and the robustness of NCG methods in tackling difficult problems. In particular, we propose the use of damped techniques for a twofold aim: for modifying (unpreconditioned) NCG methods and for constructing preconditioners based on quasi-Newton updates [Caliciotti, Fasano, Roma, Novel preconditioners based on quasi-Newton updates for nonlinear conjugate gradient methods, to appear on *Optim Lett*], to be used in the preconditioned NCG schemes. In the first case, we obtain a novel NCG method, hence the necessity of ensuring its global convergence. In the second case, we propose a new general methodology to be used for defining new preconditioning strategies for NCG methods. In the latter case, the damped techniques enable to define efficient preconditioners which approximate the inverse of the Hessian matrix, while still preserving information provided by the secant equation or some of its modifications. An extensive numerical experience confirms the effectiveness of both the approaches proposed.

Massimo Roma
SAPIENZA - Università di Roma
roma@dis.uniroma1.it

Mehiddin Al-Baali
Sultan Qaboos University
albaali@squ.edu.om

Andrea Caliciotti

Dip. Ingegneria Informatica, Automatica e Gest.
SAPIENZA, Universita' di Roma, Rome, Italy
caliciotti@dis.uniroma1.it

Giovanni Fasano
Dip. Matematica Applicata
Universita' Ca' Foscari di Venezia
fasano@unive.it

CP17

Solving Two-Stage Adaptive Nonlinear Models via Dual Formulations

In this talk we derive a dualized formulation for two-stage adaptive nonlinear optimization models. In the literature, the original primal formulations are deemed intractable due to nonlinearity in the second stage decisions. In contrast to the primal formulation, our new dualized formulation is *linear* in the second stage decisions. Therefore, we can apply a wide array of methods to solve this model, including popular techniques such as affine policies. We also show how the dual formulation can be used to provide stronger lower bounds on the performance of these affine policies. Our results are demonstrated on a variety of examples where the nonlinearity in the second stage decisions appears naturally.

Frans de Ruiter, Dick Den Hertog
Tilburg University
f.j.c.t.deruiter@tilburguniversity.edu,
d.denhertog@tilburguniversity.edu

CP18

Efficient Combinatorial Optimization for Graph Partitioning Using Quantum Annealing

The recent availability of commercial D-Wave quantum annealers offers a new tool for solving NP-hard problems alongside existing approaches based on heuristics and exhaustive search (such as branch-and-bound). However, it is still unclear whether, and to what extent, the current state of quantum computing provides any advantage over existing classical approaches for a given problem class. In this study, we address this question by assessing the performance of the D-Wave 2X quantum computer on graph partitioning (GP), a well-studied and important NP-hard graph problem. We study several versions of the GP problem by providing formulations as quadratic unconstraint binary optimization (QUBO) problem, the type accepted by D-Wave, and by implementing those using three quantum solvers provided by D-Wave Systems (QSage, QBSolv, Sapi). We compare the performance of our quantum implementations to classical GP algorithms (METIS, simulated annealing, Gurobi) on a set of test graphs. We show that D-Wave can solve the GP problems to optimality in most cases, but so can the classical solvers, and in much shorter times. This is probably due to the restriction on graph sizes (between 6-20 vertices) which are compatible with current D-Wave hardware (with about 1000 qubits). We observe that the efficiency gap shrinks fast as the problem size increases, thus leading us to conjecture that for graphs with several hundred vertices a noticeable advantage of the quantum versions can be expected.

Georg Hahn
Imperial College London
g.hahn11@ic.ac.uk

Hristo Djidjev
Los Alamos National Laboratory
djidjev@lanl.gov

Guillaume Rizk
INRIA/Irisa France
guillaume.rizk@inria.fr

Guillaume Chapuis
Los Alamos National Laboratory
gchapuis@lanl.gov

CP18

Combinatorial Optimization in Survey Statistics

Many problems in statistics require the solution of large-scale combinatorial and discrete optimization problems. Oftentimes, heuristics are used to find a near-optimal solution, even though finding the global optimum is tractable in many cases. This observation is in particular addressed for the constrained nearest neighbor hot deck imputation problem. This problem arises in statistics where almost all population surveys suffer from missing responses. To create a complete and coherent data set, usually single imputation method are applied. The nearest neighbor hot deck imputation replaces the missing values with observed values from the closest donor with respect to a distance, e.g., the Gower distance. The constrained variant of the method additionally limits the number of donations per survey unit. This setup can be transformed to a weighted b-matching problem and solved efficiently via minimum cost flow algorithms.

Dennis Kreber, Ulf Friedrich, Jan Burgard
Trier University
kreberd@uni-trier.de, friedrich@uni-trier.de,
burgard@uni-trier.de

Sven De Vries
Universität Trier
devries@uni-trier.de

CP18

A New Approach for Solving Nonlinear Mixed Integer DC Programs Based on a Continuous Relaxation Without Integrality Gap and Smoothing Technique

In this talk, we consider a mixed integer DC program (MIDCP) whose objective function is a dc function, that is, a function that can be represented as the difference of two convex functions. Recently, T. Maehara, N. Marumo, and K. Mutota (2015) proposed a new continuous relaxation technique for solving discrete DC programs with only integral variables. Specifically, their approach does not give rise to a so-called integrality gap between the optimal values of the original discrete DC program and continuously relaxed problem. In this study, we firstly extend their result to the MIDCP case. Based on this, we present a new algorithm for solving the MIDCP. We further incorporate the smoothing technique into the proposed one to handle nonsmooth functions. Finally, we give some convergence results for the proposed algorithms.

Takayuki Okuno
Department of Applied Mathematics and Physics
Graduate School of Informatics Kyoto University
t_okuno@rs.tus.ac.jp

Yoshiko Ikebe
Department of Information and Computer Technology
Tokyo University of Science
yoshiko@rs.tus.ac.jp

CP18

PIPS-SBB: A Distributed-Memory Linear-Programming-Based Branch-and-Bound Solver for Stochastic Mixed-Integer Programs

Deterministic equivalent formulations of stochastic MIPs from applications such as unit commitment (UC) can exceed available memory on a single workstation. To overcome this issue, we have developed PIPS-SBB, a distributed-memory parallel stochastic mixed-integer programming (SMIP) solver based on the distributed-memory parallel stochastic linear programming solver PIPS-S. We will overview how PIPS-SBB uses the block-angular structure of constraints in SMIPs to illustrate the parallel data structures used in our branch-and-bound algorithm. We then discuss how to adapt common algorithms used within MIP solvers to preserve both the block-angular structure of SMIPs and our parallel data layout. These algorithms, such as cutting plane generators, presolve, and primal heuristics, are used to accelerate convergence of commercial and open-source MIP solvers, and are necessary to solve large problem instances quickly. We also discuss the effects of grouping natural blocks within SMIPs into larger, process-local blocks, which enables us to trade off between using more MPI processes and generating tighter cuts that also consume more memory. Finally, we report the performance of PIPS-SBB on test problem instances in the SMIP benchmarking test set SIPLIB and on some UC instances to show that our software is competitive with existing state-of-the-art MIP solver implementations.

Geoffrey M. Oxberry
Lawrence Livermore National Laboratory
oxberry1@llnl.gov

Lluis-Miquel Munguia
Georgia Institute of Technology
Stewart School of Industrial & Systems Engineering
lluis.munguia@gatech.edu

Cosmin G. Petra
Lawrence Livermore National
Center for Advanced Scientific Computing
petra1@llnl.gov

Pedro Sotorrio, Thomas Edmunds, Deepak Rajan
Lawrence Livermore National Laboratory
sotorrio1@llnl.gov, edmunds2@llnl.gov, rajan3@llnl.gov

CP18

Binary Variable Decompositions via ADMM in Mixed Integer Programs

We will consider a class of mixed integer programs where the problem is convex except for a binary constraint on the vector x. We will discuss heuristics for problems where the objective can be written as $f(g(x,y))$, for which we will replace the inner function $g(x,y)$ with its unconstrained version, while still being able to use ADMM to identify the binary variable such that $g(x,y)$ would be closest to its unconstrained version by only checking $2n$ values, rather than exponential 2^n, where n is the dimension of the binary variable. We will compare this heuristic both in terms of the performance and complexity to the case where only the binary vector x is replaced with its unconstrained counterpart, and then the unconstrained solution is rounded off to obtain the binary vector.

Michael Rotkowitz
Electrical and Computer Engineering
University of Maryland, College Park
mcrotk@umd.edu

Alborz Alavian
University of Maryland College Park
alavian@umd.edu

CP19

On Shape-Changing Trust-Region Subproblems

We propose a method for solving large-scale L-SR1 trust-region subproblems where the trust-region is defined by so-called *shape-changing* norms that depend on the eigen-decomposition of the approximation to the Hessian. These shape-changing norms simplify solving the subproblems. In particular, one of the norms results in an explicit expression of the solution, and in the other norm the solution can be characterized by a set of optimality conditions. Since the L-SR1 matrix is not guaranteed to remain positive definite, we also propose an analytic expression for the subproblem solution when the "hard-case" occurs.

Johannes J. Brust
UC Merced
jbrust@ucmerced.edu

Oleg Burdakov
Linkoping University
oleg.burdakov@liu.se

Jennifer Erway
Wake Forest University
erwayjb@wfu.edu

Roummel F. Marcia
University of California, Merced
rmarcia@ucmerced.edu

Ya-Xiang Yuan
Chinese Academy of Sciences
Inst of Comp Math & Sci/Eng, AMSS
yyx@lsec.cc.ac.cn

CP19

Scalable Adaptative Cubic Regularization Methods

Adaptative cubic regularization (ARC) methods for unconstrained optimization compute steps from linear systems with a shifted Hessian in the spirit of the modified Newton method. In the simplest case, the shift is a multiple of the identity, which is typically identified by trial and error. We propose a scalable implementation of ARC in which we solve a set of shifted systems concurrently by way of an appropriate Krylov solver

Jean-Pierre Dussault
Université de Sherbrooke
Jean-Pierre.Dussault@Usherbrooke.ca

Dominique Orban
GERAD and Dept. Mathematics and Industrial

Engineering
Ecole Polytechnique de Montreal
dominique.orban@gerad.ca

CP19

A Characterization of Symmetric Quasi-Newton Update Matrices for Quadratic Problems

In nonlinear optimization, a fundamental subproblem is to solve a linear system corresponding to an unconstrained quadric problem where the Hessian is symmetric positive definite. For quadratic problems, we give explicit conditions for the quasi-Newton update matrices which yield search directions parallel to those of the conjugate gradient method, if exact linesearch is used. The conditions are given for the quasi-Newton update matrices at a given iterate and provide a characterization of a family of symmetric rank-2 update matrices. We analyze the matrices and give necessary and sufficient conditions for preserving positive definiteness of the quasi-Newton matrix. We show that by utilizing the secant condition, Broyden's family is obtained, but in general it is not necessary for the final quasi-Newton matrix to coincide with the Hessian. We also discuss and evaluate possible extensions to nonquadratic problems.

David Ek
KTH Royal Institute of Technology
daviek@kth.se

Anders Forsgren
KTH Royal Institute of Technology
Dept. of Mathematics
andersf@kth.se

CP19

A Solver for Nonconvex Bound-Constrained Quadratic Optimization

We present a new algorithm for nonconvex bound-constrained quadratic optimization. In the strictly convex case, our method is equivalent to the state-of-the-art algorithm by Dostál and Schöberl [Comput. Optim. Appl., 30 (2005), pp. 23–43]. Unlike their method, however, we establish a convergence theory for our algorithm that holds even when the problems are nonconvex. This is achieved by carefully addressing the challenges associated with directions of negative curvature, in particular, those that may naturally arise when applying the conjugate gradient algorithm to an indefinite system of equations. Our presentation and analysis deal explicitly with both lower and upper bounds on the optimization variables, whereas the work by Dostál and Schöberl considers only strictly convex problems with lower bounds. To handle this generality, we introduce the reduced chopped gradient that is analogous to the reduced free gradient previously used. The reduced chopped gradient leads to a new condition that is used to determine when optimization over a given subspace should be terminated. This condition, although not equivalent, is motivated by a similar condition used by Dostál and Schöberl. Numerical results illustrate the superior performance of our method over commonly used solvers that employ gradient projection steps and subspace acceleration.

Hassan Mohy-Ud-Din
Yale University
Postdoctoral Associate
hassan.mohy-ud-din@yale.edu

CP19

Global Convergence of Memoryless Quasi-Newton Methods Based on Broyden Family for Unconstrained Optimization

Quasi-Newton methods are widely used for solving unconstrained optimization problems. For example, the BFGS and the DFP formulae are well-known as quasi-Newton update formulae. The Broyden family is a one parameter family of quasi-Newton update formulae which includes the BFGS and the DFP formulae. Although quasi-Newton methods are efficient for small or middle scale problems, it is difficult to apply quasi-Newton methods directly to large-scale unconstrained optimization problems, because they need the storage of memories for matrices. In order to overcome this difficulty, particular attention is paid to memoryless quasi-Newton methods. In this talk, we propose memoryless quasi-Newton methods based on the Broyden family that do not need any storage for matrices. We also deal with a scaling technique. The presented methods always satisfy the sufficient descent condition. Furthermore, we prove the global convergence property of the proposed methods for general objective functions under the assumption that the Wolfe conditions hold in line search. Finally, some numerical results are shown.

Shummin Nakayama
Tokyo University of Science,Graduate School of Science, Depa
1416702@ed.tus.ac.jp

Yasushi Narushima
Yokohama National University
Department of Manegement System Science
narushima@ynu.ac.jp

Hiroshi Yabe
Tokyo University of Science
Department of Mathematical Information Science
yabe@rs.kagu.tus.ac.jp

CP19

Necessary Optimality Conditions and Exact Penalization for Non-Lipschitz Nonlinear Programs

Including a non-Lipschitz term in the objective function has significantly enlarged the applicability of standard nonlinear programs. In particular it has recently been discovered that when the objective function belongs to a certain class of non-Lipschitz functions, local minimizers are often sparse. However when the objective function is not Lipschitz, standard constraint qualifications are no longer sufficient for Kaush-Kuch-Tucker (KKT) conditions to hold at a local minimizer, let alone ensuring an exact penalization. In this paper we extend quasi-normality and relaxed constant positive linear dependence condition to allow the non-Lipschitzness of the objective function and show that they are sufficient for KKT condition to be necessary for optimality. Moreover we also derive some exact penalty results.

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

Lei Guo

Shanghai Jiao Tong University
guolayne@sjtu.edu.cn

CP20

Solving Integer Programming Problems via Numerical Complex Integration

A novel analytic approach for the solution of integer programming problems with non-negative input is presented. The method combines generating function techniques with results from complex analysis in several variables and is primarily based on a multipath version of Cauchy's multivariate integral formula. The input data of the optimization problem is used as a parameter of an analytic function in the unit disc, which links the optimization problem to the evaluation of a complex path integral. This representation allows the formulation of an algorithm that relies on numerical quadrature. The theoretical description is supplemented with a discussion of challenges in practical implementations of the method. In particular, it is demonstrated how preprocessing with so-called path adaption methods can help to improve the condition number of the quadrature problem, whose efficient solution is essential for the algorithm. An especially promising variant of the path adaption idea solves a shortest path problem on a predefined grid inside the unit disc. This leads to a refined version of the algorithm with better numerical stability and overall performance.

Ulf Friedrich
Trier University
friedrich@uni-trier.de

CP20

Changing Graph Structure for Performing Fast, Approximate Inference in Graphical Models

Complexity of finding exact marginal distribution in probabilistic graphical models, known as marginal inference, is exponential in the treewidth of the underlying graph. We develop a method to perform approximate inference on discrete graphical models by modifying the graph to a new graph structure having low treewidth; on which performing exact inference is tractable. Performing exact inference on the new graph gives an approximate solution to the inference on original graph. We derive error bounds on the approximate inference solution as compared to the exact inference solution. We show that the problem of finding parameters of the new graph which gives the tightest error bounds can be formulated as a Linear Program (LP). The number of constraints in the LP grow exponentially with the number of nodes in the graph. To solve this issue, we develop a row generation algorithm to solve the LP. We also discuss heuristics for choosing the new graph.

Areesh Mittal
University of Texas at Austin
areeshmittal@utexas.edu

CP20

Intersection Cuts for Convex Mixed Integer Programs from Translated Cones

We develop a general framework for linear intersection cuts for convex integer programs with full-dimensional feasible regions by studying integer points of their translated tangent cones, generalizing the idea of Balas. For proper (i.e, full-dimensional, closed, convex, pointed) translated cones with fractional vertices, we show that under certain mild conditions all intersection cuts are indeed valid for the integer hull, and a large class of valid inequalities for the integer hull are intersection cuts, computable via polyhedral approximations. We also give necessary conditions for a class of valid inequalities to be tangent halfspaces of the integer hull of proper translated cones. We also show that valid inequalities for non-pointed regular translated cones can be derived as intersection cuts for associated proper translated cones under some mild assumptions.

Umakanta Pattanayak
IIT BOMBAY
umakanta@iitb.ac.in

Vishnu Narayanan
Assistant professor, Industrial Engineering and
Operations
Research, Indian Institute of Technology Bombay
vishnu@iitb.ac.in

CP20

An Approximation Algorithm for the Partial Covering 0-1 Integer Program

We deal with the partial covering 0-1 integer program. This problem is a generalization of some important problems such as the covering 0-1 integer program, the partial set multi-cover problem and the partial set cover problem. In this paper, we propose a $\max\{f, p+1\}$-approximation algorithm for the problem, where f is the largest number of non-zero coefficients in the constraints and p is the maximum number of constrains which are not satisfied.

Yotaro Takazawa, Shinji Mizuno, Tomonari Kitahara
Tokyo Institute of Technology
takazawa.y.ab@m.titech.ac.jp,
mizuno.s.ab@m.titech.ac.jp, kitahara.t.ab@m.titech.ac.jp

CP21

On the Minimization of the k-th Singular Value of a Matrix

We consider the problem of minimizing a particular singular value of a matrix variable, which is then subject to some convex constraints. Convex heuristics for this problem are discussed, including some counter-intuitive results regarding which is best, which then provide upper bounds on the value of the problem. The use of polynomial optimization formulations is considered, particularly for obtaining lower bounds on the value of the problem. We show that the main problem can also be formulated as an optimization problem with a bilinear matrix inequality (BMI), and discuss the use of this formulation.

Alborz Alavian
University of Maryland College Park
alavian@umd.edu

Michael Rotkowitz
Electrical and Computer Engineering
University of Maryland, College Park
mcrotk@umd.edu

CP21

A Stable Lagrangian-DNN Relaxation Algorithm

for Polynomial Optimization Problems

We discuss the Lagrangian doubly nonnegative (LDNN) relaxation method (Kim, Kojima, and Toh, 2016) for a class of polynomial optimization problems (POPs). The LDNN relaxation method first formulates a DNN relaxation of a POP and then apply the Lagrangian relaxation to the DNN problem. The resulting LDNN problem has a linear equality and two cone constraints. Since the dual of the LDNN problem has a single variable, it can be solved by the bisection method; the feasibility of a point can be evaluated by first-order methods (FOMs). In theory, when the parameter of the Lagrangian relaxation increases, the lower bounds obtained by LDNN improve. In practice, however, selecting the parameter is not obvious because the bisection method often suffers from numerical instability of FOMs when the parameter is large. In this work, we show that some of the constraints of DNN can be included in the cone constraints without increasing the computation cost at each iteration of FOMs. This allows us to avoid using Lagrangian relaxation for such constraints of DNN and make the bisection method stable. Numerical results are presented to demonstrate the stability of the proposed algorithm.

Naoki Ito
University of Tokyo
naoki_ito@mist.i.u-tokyo.ac.jp

Sunyoung Kim
Ewha W. University
Department of Mathematics
skim@ewha.ac.kr

Masakazu Kojima
Chuo University
kojimamasakazu@mac.com

Akiko Takeda
The Institute of Statistical Mathematics
atakeda@ism.ac.jp

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

CP21

Completely Positive and Positive Semidefinite Tensor Relaxations for Polynomial Optimization

Completely positive (CP) tensors, which correspond to a generalization of CP matrices, allow to reformulate or approximate general polynomial optimization problems (POPs) with a conic optimization problem over the cone of CP tensors. Similarly, completely positive semidefinite (CPSD) tensors, which correspond to a generalization of PSD matrices, can be used to approximate general POPs with a conic optimization problem over the cone of CPSD tensors. In this paper, we study CP and CPSD tensor relaxations for general POPs and compare them with the bounds obtained via a Lagrangian relaxation of the POP. This shows that existing results in this direction for quadratic POPs extend to general POPs. Also, we provide some tractable approximation strategies for CP and CPSD tensor relaxations. These approximation strategies show that, with a similar computational effort, bounds obtained from them for general POPs can be tighter than bounds for these problems obtained by reformulating the POP as a quadratic POP and subsequently using approximations for the quadratic problem using CP and PSD matrices. To illustrate our results, we present numerical results for these relaxation approaches on small scale fourth-order degree POPs.

Xiaolong Kuang
PhD Candidate at Lehigh University
kuangxiaolong0731@gmail.com

Luis Zuluaga
Lehigh University
luis.zuluaga@lehigh.edu

CP21

Combining SOS and Moment Relaxations with Branch and Bound to Extract Solutions to Global Polynomial Optimization Problems

We consider the problem of extracting approximate solutions to Global Polynomial Optimization (GPO) using a combination of branch and bound and Sum-of-Squares(SOS)/Moment relaxations. Specifically, we use hyperplanes to branch the feasible set of GPO problem. For each of the branches, SOS/Moment relaxations are used to find lower bounds to the optimal objective values of those GPO problems defined over the subdivided feasible sets. These bounds allow us to select branches that shrink the search domain of the GPO problem. For any given $\epsilon > 0$ and $\eta > 0$, there exist l and k sufficiently large s.t. after l branchings with SOS/Moment relaxations of degree k, the algorithm finds a point that is within ϵ-distance from a feasible and η-suboptimal point of the GPO problem. For a fixed degree of relaxations, the complexity of the algorithm is linear the number of branches and polynomial in the number of constraints. We illustrate the algorithm for a 6-variable, 6-constraint GPO problem for which the ideal generated by the equality constraints is not zero-dimensional - a case where the existing Moment-based approach for extracting the global minimizer fails.

Hesameddin Mohammadi, Matthew Peet
Arizona State University
hmoham17@asu.edu, mpeet@asu.edu

MS1

Finding Infimum Point with Respect to the Second Order Cone

We define the notion of infimum and supremum of a set of points with respect to the second order cone. These problems can be formulated as second order cone optimization and thus solvable by interior point methods in polynomial time. We present an extension of the simplex method to solve these problems. We outline both primal and dual versions of the simplex method. We also show some applications of infimum and supremum problems.

Marta Cavaleiro
Rutgers University
marta.cavaleiro@rutgers.edu

MS1

The Conic Optimizer in Mosek: A Status Report

We discuss recent developments in the conic optimizer in MOSEK. We present numerical results illustrating the improvements on challenging benchmark problems, and we discuss how many difficult problems can be solved more easily and reliably using model reformulations.

Joachim Dahl
MOSEK
joachim.dahl@mosek.com

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

MS1

Analysis of Newton's Method via Semidefinite Programming Performance Estimation Problems

We first consider the gradient (or steepest) descent method with exact line search applied to a strongly convex function with Lipschitz continuous gradient. We establish the exact worst-case rate of convergence of this scheme. We also give the tight worst-case complexity bound for a noisy variant of gradient descent method, where exact line-search is performed in a search direction that differs from negative gradient by at most a prescribed relative tolerance. Finally, we show that this analysis may be extended to analyse the worst-case performance of Newton's method for self-concordant functions. The proofs are computer-assisted, and rely on the resolution of semidefinite programming performance estimation problems as introduced in the paper [Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1-2):451-482, 2014].

Etienne De Klerk
Tilburg University
E.deKlerk@uvt.nl

Francois Glineur
Université Catholique de Louvain (UCL)
Center for Operations Research and Econometrics (CORE)
francois.glineur@uclouvain.be

Adrien Taylor
UCL, Belgium
adrien.taylor@uclouvain.be

MS1

Perturbation Analysis of Singular Semidefinite Programming via the Facial Reduction

In this talk, we study behavior of the optimal values and the solutions of semidefinite programs under perturbations on the coefficient matrices . If the feasible set of the primal problem has a nonempty interior, the optimal value changes continuously when only the constant terms (usually in the right hand sides) are perturbed. However the optimal value could change both continuously and discontinuously when we perturb the coefficient matrices of the variables in the constraints. We will use the Facial Reduction algorithm to find what kind of perturbations does not raise discontinuity. We note that, although the Facial Reduction algorithm is formulated with SDPs for general cases, it can be done without solving SDPs, if the given problem comes from matrix completion problems, polynomial optimization problems or H-infinity control problems. This is a joint work with H. Waki.

Yoshiyuki Sekiguchi
Tokyo University of Marine Science and Technology
yoshi-s@kaiyodai.ac.jp

MS2

Polynomial Norms

We establish when the d-th root of a multivariate degree-d polynomial produces a norm. In the quadratic case, it is well-known that this is the case if and only if the matrix associated with the quadratic form is positive definite. We present a generalization of this result to higher order polynomials and show that there is a hierarchy of semidefinite programs that can detect all polynomial norms. We also show that any norm can be approximated arbitrarily well with a polynomial norm.

Amir Ali Ahmadi
Princeton University
a_a_a@princeton.edu

Etienne De Klerk
Tilburg University
E.deKlerk@uvt.nl

Georgina Hall
Princeton University
gh4@princeton.edu

MS2

Convergence Rate of SoS Based Upper Bounds in Polynomial Optimization

We consider the problem of minimizing a polynomial f over a compact region K. As shown by Lasserre (2011), hierarchies of upper bounds converging to the minimum of f over K can be obtained by searching for a probability measure μ on K with a sum-of-squares density function h minimizing the expected value $\int_K h(x)f(x)\mu(dx)$. We will discuss several recent results about the convergence rate of these hierarchies. For a general compact set K satisfying a mild boundary condition (and selecting for μ the Lebesgue measure), one can show a convergence rate in $O(1/\sqrt{r})$, where $2r$ is the degree of h. When K is a convex body an improved convergence rate in $O(1/r)$ can be shown by exploiting a connection with simulated annealing based bounds. For simple regions like the hypercube, faster convergence rates can be shown by using other choices for the measure μ, as well a simpler hierarchy leading to bounds whose computation requires only elementary operations.

Monique Laurent
CWI, Amsterdam
and Tilburg University
M.Laurent@cwi.nl

Etienne De Klerk
Tilburg University
E.deKlerk@uvt.nl

Roxana Hess
LAAS CNRS Toulouse
rhess@laas.fr

Jean B. Lasserre
LAAS CNRS and Institute of Mathematics

lasserre@laas.fr

Zhao Sun
Symetrics, Utrecht
sunzhao1987@gmail.com

MS2

Symmetric Tensor Nuclear Norm

This talk discusses nuclear norms of symmetric tensors. As recently shown by Friedland and Lim, the nuclear norm of a symmetric tensor can be achieved at a symmetric decomposition. We show how to compute symmetric tensor nuclear norms, depending on the tensor order and the ground field. Lasserre relaxations are proposed for the computation. The theoretical properties of the relaxations are studied. For symmetric tensors, we can compute their nuclear norms, as well as the nuclear decomposition. The proposed methods can also be extended to nonsymmetric tensors.

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

MS2

Investigation of Crouzeix's Conjecture via Nonsmooth Optimization

Crouzeix's conjecture is among the most intriguing developments in matrix theory in recent years. Made in 2004 by Michel Crouzeix, it postulates that, for any polynomial p and any matrix A, $\|p(A)\| \leq 2\max(|p(z)| : z \in W(A))$, where the norm is the 2-norm and $W(A)$ is the field of values (numerical range) of A, that is the set of points attained by v^*Av for some vector v of unit length. Crouzeix proved in 2007 that the inequality above holds if 2 is replaced by 11.08, and very recently this was greatly improved by Palencia, replacing 2 by $1+\sqrt{2}$. Furthermore, it is known that the conjecture holds in a number of special cases, including $n = 2$. We use nonsmooth optimization to investigate the conjecture numerically by attempting to minimize the "Crouzeix ratio", defined as the quotient with numerator the right-hand side and denominator the left-hand side of the conjectured inequality. We present numerical results that lead to some theorems and further conjectures, including variational analysis of the Crouzeix ratio at conjectured global minimizers. All the computations strongly support the truth of Crouzeix's conjecture.

Michael L. Overton
New York University
Courant Instit. of Math. Sci.
overton@cs.nyu.edu

Anne Greenbaum
University of Washington
Mathematics Department
greenbau@math.washington.edu

Adrian Lewis
Cornell University
Oper. Res. and Indust. Eng.
aslewis@orie.cornell.edu

MS3

Optimization and Control in Free Boundary Fluid-Structure Interactions

We consider optimization and optimal control problems subject to free and moving boundary fluid-structure interactions. As the coupled fluid-structure state is the solution of a system of PDEs that are coupled through continuity relations defined on the free/moving interface, the investigation (existence of optimal controls, sensitivity equations, optimality conditions, etc.) is heavily dependent on the geometry of the problem, and falls into moving shape analysis framework.

Lorena Bociu
N. C. State University
Raleigh, NC
lvbociu@ncsu.edu

Lucas Castle
North Carolina State University
lcastle@ncsu.edu

MS3

Traffic Regulation via Controlled Speed Limit

We study an optimal control problem for traffic regulation via variable speed limit. The traffic flow dynamics is described with the Lighthill-Whitham-Richards (LWR) model with Newell-Daganzo flux function. We aim at minimizing the L^2 quadratic error to a desired outflow, given an inflow on a single road. We first provide existence of a minimizer and compute analytically the cost functional variations due to needle-like variation in the control policy. Then, we compare three strategies: instantaneous policy; random exploration of control space; steepest descent using numerical expression of gradient. We show that the gradient technique is able to achieve a cost within 10% of random exploration minimum with better computational performances.

Maria Laura Delle Monache
Inria
ml.dellemonache@inria.fr

Benedetto Piccoli
Rutgers University
piccoli@camden.rutgers.edu

Francesco Rossi
LSIS, Aix-Marseille Université
francesco.rossi@lsis.org

MS3

Sensitivity Analysis for Hyperbolic Systems of Conservation Laws

Sensitivity analysis (SA) concerns the quantification of changes in Partial Differential Equations (PDEs) solution due to perturbations in the model input. SA has many applications, among which uncertainty quantification, quick evaluation of close solutions, and optimization based on descent methods. Standard SA techniques for PDEs, such as the continuous sensitivity equation method, rely on the differentiation of the state variable. However, if the governing equations are hyperbolic PDEs, the state can exhibit discontinuities yielding Dirac delta functions in the sensitivity. The aim of this work is to modify the sensitivity equations to obtain a solution without delta functions. This is motivated by several reasons: firstly, a delta function cannot be seized numerically, leading to an incorrect solution for the sensitivity in the neighborhood of the state discontinuity; secondly, for some applications like optimization and evaluation of close solutions, the peaks appearing in the numerical solution of original sensitivity equations make sensitivities unusable. We propose a two-steps procedure: (i) definition of a correction term added to the right-hand side of the sensitivity equations; (ii) definition of a shock detector ensuring that the correction term is added only where needed. We show how this procedure can be applied to the Euler barotropic system with two different finite-volume formulations, based either on an exact Riemann solver or the approximate Roe solver.

Camilla Fiorini
University of Versailles
camilla.fiorini@uvsq.fr

Regis Duvigneau
INRIA Sophia Antipolis
regis.duvigneau@inria.fr

Christophe Chalons
t Université Versailles Saint-Quentin-en-Yvelines,
christophe.chalons@uvsq.fr

MS3

A Fokker-Planck Nash Game to Model Pedestrian Motion

Fokker-Planck-Kolmogorov(FPK) equations are PDEs which govern the dynamics of the probability density function (PDF) of continuous-time stochastic processes (e.g. Ito processes). In [1] a FPK-constrained control framework, where the drift was considered as control variable is developed and applied to crowd motion.

We consider herein the extension of the latter framework to the case where two crowds (or pedestrian teams) are competing through a Nash game. The players strategies are the drifts, which yield two uncoupled FPK equations for the corresponding PDFs. The interaction is done through cost functions : each player would prefer to avoid overcrowding (w.r.t. the other one, hence the coupling) additionally to have her own preferred trajectory and obstacle avoidance. In this particular setting, we prove the existence and uniqueness of the Nash equilibrium (NE). The NE is computed by means of a fixed point algorithm and adjoint-state method is used to compute the pseudo-gradients. We finally present some numerical experiments to illustrate which dynamics may arise from such equilibria.

[1] S. Roy and M. Annunziato and A. Borzì, A Fokker-Planck feedback control-constrained approach for modelling crowd motion, Journal of Computational and Theoretical Transport, 45(6), 2016, 442-458.

Alfio Borzi
Institute for Mathematics
University of Würzburg
alfio.borzi@uni-wuerzburg.de

Abderrahmane Habbal
UNIVERSITY COTE D'AZUR UCA, UNS, CNRS and Inria, FRANCE
habbal@unice.fr

Souvik Roy
Institute for Mathematics
University of Würzburg
1roysouvik@gmail.com

MS4

Low Rank Factor Analysis

Factor Analysis (FA) is a technique that is widely used to obtain a parsimonious representation of correlation structure among a set of variables in terms of a smaller number of common hidden factors. In this talk we revisit the classical rank-constrained FA problem and propose a flexible family of exact, smooth reformulations for this task. By coupling state-of-the-art techniques from nonlinear optimization with methods from global optimization, we show that our approach often finds certifiably optimal solutions and significantly outperforms existing publicly available methods across a variety of real and synthetic examples.

Dimitris Bertsimas
Massachusetts Insitute of Technology
dbertsim@mit.edu

Martin S. Copenhaver
Massachusetts Institute of Technology
mcopen@mit.edu

Rahul Mazumder
MIT
rahulmaz@mit.edu

MS4

Fitting Convex Sets to Data via Matrix Factorization

High-dimensional datasets arise prominently in a range of contemporary problem domains throughout science and technology. In many of these settings, the data are often constrained structurally so that they only have a few degrees of freedom relative to their ambient dimension. Methods such as manifold learning, dictionary learning, and others aim at computationally identifying such latent low-dimensional structure. We describe a new approach to inferring the low-dimensional structure underlying a dataset by fitting a convex set with favorable facial structure to the data. The convex set identifies a regularizer that is tractable to compute, and is useful for subsequent processing tasks. Our procedure is based on computing a structured matrix factorization and it includes several previous techniques as special cases. We illustrate the utility of our method with experimental demonstrations.

Yong Sheng Soh
California Institute of Technology
ysoh@caltech.edu

MS4

Interpreting Latent Variables in Factor Models via Convex Optimization

Latent or unobserved phenomena pose a significant difficulty in data analysis as they induce complicated and confounding dependencies among a collection of observed variables. Factor analysis is a prominent multivariate statistical modeling approach that addresses this challenge by identifying the effects of (a small number of) latent variables on a set of observed variables. However, the latent variables in a factor model are purely mathematical objects that are derived from the observed phenomena, and they do not have any interpretation associated to them. In this talk, we describe a systematic approach for attributing semantic information to the latent variables in a factor model. Our method is based on solving computationally tractable convex optimization problems, and it can be viewed as a generalization of the minimum-trace factor analysis procedure for fitting factor models via convex optimization. We analyze the theoretical consistency of our approach in a high-dimensional setting as well as its utility in practice via experimental demonstrations with real data.

Armeen Taeb
California Institute of Technology
California institute of technology
ataeb@caltech.edu

MS4

Generalized Permutohedra from Probabilistic Graphical Models

We consider the problem of learning a Bayesian network or directed acyclic graph (DAG) model. We propose the sparsest permutation algorithm, a nonparametric approach based on finding the ordering of the variables that yields the sparsest DAG. In order to compute the sparsest permutation, we introduce the DAG associahedron. This is a generalized permutohedron, the analog of the prominent graph associahedron for DAGs. We end by introducing a simplex-type algorithm on this convex polytope that provably converges to the sparsest permutation.

Caroline Uhler
Massachusetts Institute of Technology
cuhler@mit.edu

MS5

Linear Convergence of Gradient and Proximal-Gradient Methods Under the Polyak-Lojasiewicz Condition

In 1963, Polyak proposed a simple condition that is sufficient to show a global linear convergence rate for gradient descent. This condition is a special case of the Lojasiewicz inequality proposed in the same year, and it does not require strong convexity (or even convexity). In this work, we show that this much-older Polyak-Lojasiewicz (PL) inequality is actually weaker than the five main conditions that have been explored to show linear convergence rates without strong convexity over the last 25 years. We also use the PL inequality to give new analyses of coordinate descent and stochastic gradient for many non-strongly-convex (and some non-convex) functions. We further propose a generalization that applies to proximal-gradient methods for non-smooth optimization, leading to simple proofs of linear convergence for support vector machines and L1-regularized least squares without additional assumptions.

Hamed Karimi, Julie Nutini, Mark Schmidt
University of British Columbia
hamedkarim@gmail.com, jnutini@cs.ubc.ca, schmidtm@cs.ubc.ca

MS5

Stochastic Heavy Ball Method for Solving Linear Systems

Linear systems form the backbone of most numerical codes used in academia and industry. With the advent of theage of big data, practitioners are looking for ways to solve linear systems of unprecedented sizes. In this talk we present a stochastic variant of the Heavy Ball method for solving large scale linear systems. We prove linear convergence with an accelerated rate under no assumptions on the system beyond consistency.

Nicolas Loizou
The University of Edinburgh
s1461357@sms.ed.ac.uk

Peter Richtarik
University of Edinburgh
peter.richtarik@ed.ac.uk

MS5

Linear Convergence of the Randomized Feasible Descent Method Under the Weak Strong Convexity Assumption

In this paper we generalize the framework of the Feasible Descent Method (FDM) to a Randomized (R-FDM) and a Randomized Coordinate-wise Feasible Descent Method (RC-FDM) framework. We show that the famous SDCA algorithm for optimizing the SVM dual problem, or the stochastic coordinate descent method for the LASSO problem, fits into the framework of RC-FDM. We prove linear convergence for both R-FDM and RC-FDM under the weak strong convexity assumption. Moreover, we show that the duality gap converges linearly for RC-FDM, which implies that the duality gap also converges linearly for SDCA applied to the SVM dual problem.

Chenxin Ma
Lehigh University
chm514@lehigh.edu

Martin Takac
University of Edinburgh
takac.mt@gmail.com

MS5

Randomized Projection Methods for Convex Feasibility Problems

Finding a point in the intersection of a (finite or infinite) collection of convex sets, that is the convex feasibility problem, represents a modeling paradigm which was used for many decades for posing and solving various problems arising in engineering, computer science, applied mathematics, physics and other fields. In this paper, we first discuss different representations for a collection of convex sets and derive several stochastic reformulations of the convex feasibility problem. We then introduce a new general random projection algorithmic framework that at each iteration chooses randomly N sets from this collection, projects in parallel the present iterate on each chosen set and then computes the new update using an average of these projections with a certain step size. We derive convergence rate results for this general random projection method for both cases: sublinear convergence when we deal with general convex sets and linear convergence when the collection of sets satisfies a certain linear regularity condition. Moreover, our analysis allows large step sizes (that is step sizes larger than one) and we prove that the convergence rates depend directly on the number N of sets that are chosen randomly. We also show that some well-known projection methods, such as the alternating projection or the average projection schemes, result as special cases of our algorithmic framework.

Ion Necoara
University Politehnica Bucharest
ion.necoara@acse.pub.ro

Peter Richtarik
University of Edinburgh
peter.richtarik@ed.ac.uk

Andrei Patrascu
University Politehnica of Bucharest
andrei.patrascu@acse.pub.ro

MS6

Fractional Elliptic QVI: Theory and Numerics

This talk introduces an elliptic quasi-variational inequality (QVI) problem class with fractional diffusion of order $s \in (0, 1)$, studies existence and uniqueness of solutions and develops a solution algorithm. As the fractional diffusion prohibits the use of standard tools to study the QVI, instead we realize it as a Dirichlet-to-Neumann map for a nonuniformly elliptic equation posed on a semi-infinite cylinder. We first study existence and uniqueness of solution for this extended QVI and then transfer the results to the fractional QVI. This introduces a new paradigm in the field of fractional QVIs. We truncate the semi-infinite cylinder and show that the solution to the truncated problem converges to the solution to the extended problem, under fairly mild assumptions, as truncation parameter $\tau \to \infty$. Since the constraint set changes with the solution, we develop an argument using Mosco convergence. We state an algorithm to solve the truncated problem and show its convergence in function spaces. Finally, we conclude with several illustrative numerical examples.

Harbir Antil
George Mason University
Fairfax, VA
hantil@gmu.edu

Carlos N. Rautenberg
Humboldt-University of Berlin
carlos.rautenberg@math.hu-berlin.de

MS6

A Priori Error Estimates for Space-Time Finite Element Discretization of Parabolic Time-Optimal Control Problems

We consider the following time-optimal control problem (P) subject to a linear parabolic equation with control q and state u:

$$\text{Minimize } j(T, q) := T + \frac{\alpha}{2} \int_0^T \|q(t)\|^2_{L^2(\Omega)}\, dt$$

subject to

$$T > 0,\ q \in Q_{ad}$$
$$\partial_t u + Au = Bq \text{ in } (0,T) \times \Omega,\ u(0) = u_0 \text{in } \Omega,$$
$$\|u(T) - u_d\|_{L^2(\Omega)} \leq \delta,$$

where $u_d \in H^1_0(\Omega)$ is the desired state and $\delta > 0$ is given. Moreover, Q_{ad} denotes the set of admissible controls and $\alpha > 0$ the regularization parameter. Our aim is the numerical analysis for finite element discretizations of (P). It is worth mentioning that (P) is non-convex, possesses a nonlinear dependence on T and is subject to state constraints. We derive second order necessary as well as sufficient optimality conditions for (P) using a critical cone that leads to a minimal gap between necessity and sufficiency. For the numerical solution we consider a discontinuous Galerkin discretization scheme in time and a continuous Galerkin scheme in space. The main novelty are *a priori* discretization error estimates of optimal order for T and q under the hypothesis of second order sufficient optimality conditions.

Lucas Bonifacius
Technical University of Munich
Centre for Mathematical Sciences, M17
lucas.bonifacius@ma.tum.de

Boris Vexler
Technische Universitaet Muenchen
Centre for Mathematical Sciences, M1
vexler@ma.tum.de

Konstantin Pieper
Florida State University
Dept of Scientific Computing
kpieper@fsu.edu

MS6

Numerical Aspects and Error Estimates for Dirichlet Boundary Control Problems

In this talk we discuss convergence results for finite element discretized Dirichlet control problems in polygonal domains. We investigate unconstrained as well as control constrained problems. In both cases we discretize the state and the control by piecewise linear and continuous functions. The error estimates, which we obtain, mainly depend on the size of the interior angles but also on the presence of control constraints and the structure of the underlying mesh. For instance, considering non-convex domains, the convergence rates of the discrete optimal controls in the unconstrained case can even be worse than in the control constrained case.

Johannes Pfefferer
Technische Universität München
Zentrum Mathematik
pfefferer@ma.tum.de

Thomas Apel
Universität der Bundeswehr München
thomas.apel@unibw.de

Mariano Mateos
Unevesidad de Oviedo
33203 Gijon (Asturia), Spain
mmateos@uniovi.es

Arnd Rösch
University of Duisburg-Essen
arnd.roesch@uni-due.de

MS6

Sparse Control for Inverse Point Source Location with the Helmholtz Equation

We address the problem of recovering a superposition of localized sound sources, described by a limited number of frequencies, from point-wise measurements of the corresponding acoustic pressure. To this purpose, we investi-

gate a family of weighted, sparse, measure-valued control problems in combination with the Helmholtz equation on a bounded domain. The observation is given at a finite collection of points (representing M microphones) and solutions of the problem are finite sums of Dirac-delta functions (acoustic monopoles). To obtain well-posedness in a general setting, the weights are chosen unbounded near the observation points. Furthermore, optimality conditions and conditions for recovery in the small noise case are discussed, which motivates concrete choices of the weight. The numerical realization is based on an accelerated conditional gradient method and a finite element discretization. Numerical examples confirm that an appropriate choice of the weight significantly increases the number of cases where recovery is possible with the proposed problem formulation.

Konstantin Pieper
Florida State University
Dept of Scientific Computing
kpieper@fsu.edu

Bao Q. Tang, Philip Trautmann
Institute of Mathematics and Scientific Computing
University of Graz
quoc.tang@uni-graz.at, philip.trautmann@uni-graz.at

Daniel Walter
Technische Universität München
walter@ma.tum.de

MS7

Shape-Changing Limited-Memory SR1 Trust-Region Methods

In this talk, we discuss methods for solving the trust-region subproblem when a limited-memory symmetric rank-one (SR1) matrix is used in place of the true Hessian. We consider two different shape-changing norms to define the trust region. In both cases, the proposed methods take advantage of the two shape-changing norms to decompose the subproblem into two separate problems–both easy to solve. The proposed solver makes use of the structure of limited-memory SR1 matrices to find solutions that solve each subproblem to high accuracy even in the so-called "hard case". This is joint work with Johannes Brust, Oleg Burdakov, Roummel Marcia, and Ya-xiang Yuan.

Jennifer Erway
Wake Forest University
erwayjb@wfu.edu

Johannes J. Brust
UC Merced
jbrust@ucmerced.edu

Oleg Burdakov
Linkoping University
oleg.burdakov@liu.se

Roummel F. Marcia
University of California, Merced
rmarcia@ucmerced.edu

Yaxiang Yuan
Laboratory of Scientific and Engineering Computing
Chinese Academy of Science
yyx@lsec.cc.ac.cn

MS7

On Solving Symmetric Systems of Linear Equations Arising in Optimization

Solving systems of linear equations where the matrix is symmetric is an important part of many optimization methods intended for smooth optimization problems. It could be methods based on linear equations as such, for example active-set methods for linear and convex quadratic programming, or methods based on solving nonlinear equations by solving a sequence of linear equations, for example interior methods. It is not always so clear which equations to solve. Many methods are Newton-based, but it may be desirable to use quasi-Newton approximations of the Hessian. We discuss particular systems of symmetric linear equations arising in optimization. Some are genuinely indefinite, whereas others may be converted into positive definite matrices. A particular application is interior methods, where the indefinite augmented system may be identified by a doubly augmented system, which is positive definite.

Anders Forsgren
KTH Royal Institute of Technology
Dept. of Mathematics
andersf@kth.se

MS7

Using Nonlinear Optimization to Adjust Protein Similarity Score Matrices for Amino Acid Composition

The algorithm in BLAST (Basic Local Alignment Search Tool) and similar algorithms align protein sequences to assess whether the sequences are significantly similar. A positive score is awarded for each position for which there is an exact match at an aligned position, and a negative score is typically assigned each mismatch. These awards and penalties are known as similarity scores and depend on the specific pairs of amino acids aligned. Gaps within alignments are allowed but penalized with a negative score. The values for similarity scores and gap penalties may be chosento give on average good results. However, such generic scoring systems inevitably yield biologically undesirable results on sequences with skewed amino acid composition. We describe a nonlinear information-theoretic entropy minimization problem used within BLAST to adjust protein similarity scores for amino acid composition. We show that composition adjustment maintains the retrieval efficiency of BLAST searches, while improving the accuracy of BLAST statistics.

E. Michael Gertz
National Center for Biotechnology Information
National Institutes of Health
gertz@ncbi.nlm.nih.gov

MS7

Adding Primal Regularization to a Primal-Dual Augmented Lagrangian Approach

This talk focuses on the importance of adding explicit primal regularization to primal-dual augmented Lagrangian approaches. In particular we demonstrate how doing so provides mechanisms for implicitly bounding both primal and dual Lagrange multiplier estimates. This is of particular importance for primal-dual approaches that hope

to retain a one-to-one correspondence between the search direction computation and the corresponding primal-dual merit function. Without primal regularization, numerical difficulties exist when solving problems that are even moderately ill-conditioned or infeasible. The advantageous coupling of the search direction with the corresponding merit function is therefore typically abandoned as a solution is approached.

Joshua Griffin, Riadh Omheni
SAS Institute, Inc.
Joshua.Griffin@sas.com, riadh.omheni@sas.com

Wenwen Zhou
SAS Institute Inc.
Wenwen.Zhou@sas.com

MS8

A Distributed ADMM-Like Method for Resource Sharing Under Conic Constraints Over Time-Varying Networks

In this talk, decentralized methods for solving cooperative multi-agent optimization problems over a network of agents will be discussed, where only those agents connected by an edge can directly communicate with each other. The objective is to minimize the sum of agent-specific composite convex functions, i.e., each term in the sum is a private cost function belonging to an agent, subject to a conic constraint coupling all agents local decisions. An ADMM-like primal-dual algorithm is proposed to solve a saddle point formulation of the problem in a distributed fashion, where the consensus among agents is imposed on the agents' estimates of the dual price associated with the coupling constraint. We provide convergence rates both in suboptimality, infeasibility, and consensus violation in dual price estimates; examine the effect of underlying network topology on the convergence rate of the proposed decentralized algorithm; and show how to extend this method to handle communication networks with time-varying topology and directed networks. This optimization model abstracts a number of applications in machine learning, distributed control, and estimation using sensor networks.

Necdet S. Aybat, Erfan Yazdandoost Hamedani
Penn State University
nsa10@psu.edu, evy5047@psu.edu

MS8

Structured Nonconvex and Nonsmooth Optimization: Algorithms and Iteration Complexity Analysis

We consider in this paper some constrained nonconvex optimization models in block decision variables, with or without coupled affine constraints. In the case of no coupled constraints, we show a sublinear rate of convergence to an ϵ-stationary solution in the form of variational inequality for a generalized conditional gradient method, where the convergence rate is shown to be dependent on the Hölderian continuity of the gradient of the smooth part of the objective. For the model with coupled affine constraints, we introduce corresponding ϵ-stationarity conditions, and propose two proximal-type variants of the ADMM to solve such a model, assuming the proximal ADMM updates can be implemented for all the block variables except for the last block, for which either a gradient step or a majorization-minimization step is implemented. We show an iteration complexity bound of $O(1/\epsilon^2)$ to reach an ϵ-stationary solution for both algorithms. Moreover, we show that the same iteration complexity of a proximal BCD method follows immediately. Numerical results are provided to illustrate the efficacy of the proposed algorithms for tensor robust PCA.

Bo Jiang
School of Information Management and Engineering
Shanghai University of Finance and Economics
isyebojiang@163.com

MS8

Geometric Descent Method for Composite Convex Minimization

In the talk, we discuss how to extend the geometric descent method that is recently proposed by Bubeck, Lee and Singh, to solving composite convex minimization problems. Techniques that can potentially accelerate the algorithm are also discussed. Numerical results show that this method can be advantageous comparing with Nesterov's accelerated gradient method for some applications.

Shiqian Ma, Shixiang Chen
The Chinese University of Hong Kong
sqma@se.cuhk.edu.hk, sxchen@se.cuhk.edu.hk

MS8

Gradient Sampling Methods on Riemannian Manifolds and Algebraic Varieties

Gradient sampling (GS) is a conceptually simple method for optimization of locally Lipschitz functions. The idea is to approximate the subdifferential at a given point by a convex hull of randomly sampled nearby gradients. We consider generalizations of this approach to optimization problems posed on complete Riemannian manifolds or closed algebraic varieties admitting an a-regular stratification. The latter scenario is motivated by optimization problems on low-rank matrices or tensors. The main task is to prove that the minimum norm element in a convex hull of vectors, which are obtained by transporting gradients from nearby points to the current tangent space, is a descent direction with respect to a retraction, provided that the number of sampled gradients equals at least the dimension of the manifold/variety plus one. For varieties it is further necessary to discuss what should happen in the singular points, where one may be faced with a tangent cone that is not convex. Under reasonable assumptions, we are able to prove that with probability one the limit points of our proposed GS algorithms are Clarke stationary points relative to their strata. Numerical experiments illustrate that the approach can be useful on manifolds of not too high dimension.

Andre Uschmajew, Seyedehsomayeh Hosseini
University of Bonn
uschmajew@ins.uni-bonn.de, hosseini@ins.uni-bonn.de

MS9

Fast Algorithms for Nonconvex Robust Learning

We show how to transform any optimization problem that arises from fitting a machine learning model into one that (1) detects and removes contaminated data from the training set and (2) simultaneously fits the trimmed model on the remaining uncontaminated data. To solve the resulting nonconvex optimization problem, we introduce a fast stochastic proximal-gradient algorithm that incorporates prior knowledge through non-smooth regularization.

Aleksandr Aravkin
University of Washington
sasha.aravkin@gmail.com

Damek Davis
Cornell University
ORIE
dsd95@cornell.edu

MS9

Nonsmooth Optimization Using Taylor-Like Models: Error Bounds, Convergence, and Termination Criteria

Common algorithms rely on minimizing simple Taylor-like models of the objective function. The proximal gradient algorithm and methods of Gauss-Newton type for minimizing the composition of a convex function and a smooth map are typical examples. Our main result is an explicit relationship between the step-size of any such algorithm and the slope of the objective function at a nearby point. Consequently, we (1) show that the step-sizes can be reliably used to terminate the algorithm, (2) prove that as long as the step-sizes tend to zero, every limit point of the iterates is stationary, and (3) show that conditions, akin to classical quadratic growth, imply that the step-sizes linearly bound the distance of the iterates to the solution set. The latter so-called error bound property is typically used to establish linear (or faster) convergence guarantees. Analogous results hold when the step-size is replaced by the square root of the decrease in the model's value. All the results readily extend to when the models are minimized only inexactly.

Dmitriy Drusvyatskiy
University of Washington
University of Washington
ddrusv@uw.edu

MS9

Radial Subgradient Descent

We discuss a subgradient method for minimizing nonsmooth, non-Lipschitz convex optimization problems. We extend the work of Renegar by taking a different perspective, leading to an algorithm which is conceptually more natural, has notably improved convergence rate, and for which the analysis is surprisingly simple. Convergence occurs at a rate analogous to traditional methods that require Lipschitz continuity. At each iteration, the algorithm takes a subgradient step and then performs a line search to move radially towards (or away from) a known feasible point. Costly orthogonal projections typical of subgradient methods are entirely avoided.

Benjamin Grimmer
Cornell Unviersity
Department of Operations Research and Information
bdg79@cornell.edu

MS9

First-Order Methods for Singular Kalman smoothing with PLQ Penalties

Kalman smoothing is widely used in science and engineering to reconstruct hidden states from a series of measurements. Classic models assume process and measurement errors have non-degenerate (full-rank) variances. However, singular covariance models can capture integrated signals and colored noise in many applications. We use a constrained reformulation to allow singular covariances in both process and measurement errors. The resulting problem can be elegantly solved in the dual, where it reduces to inverting a single block tridiagonal positive definite system, just as in the classic case. We use this insight to develop efficient first-order methods for general singular smoothers, where process and measurement losses can be piecewise linear-quadratic functions (such as huber, and 1-norm), and simple regularizes on the state can be added as well.

Jonathan Jonker, Aleksandr Aravkin
University of Washington
jonkerjo@uw.edu, sasha.aravkin@gmail.com

MS10

A Joint Routing and Speed Optimization Problem

We study a variation of the classical vehicle routing problem with time-windows, where the overall speed used to traverse an arc is also a decision variable. These can be used to ensure that time windows are respected by effectively speeding up the vehicle along an arc. The tradeoff is that the objective function will also have terms which will take into account a (nonlinear) convex function of the speed, which can be used to model transportation models with energy savings and emission reduction considerations. It is worth noting that the speed optimization problem by itself is easy to solve. In other words, given a fixed route, there is an efficient algorithm to find the best set of speeds to use to traverse the route. On the other hand, if the routes are not fixed, the problem becomes the joint routing and speed optimization problem, which is extremely challenging to solve. Previous approaches involved a heuristic discretization of speeds and formulating the problem as a mixed-integer linear program, as well as formulating the exact problem as a mixed-integer nonlinear (convex) program. We propose a new set partitioning formulation for the problem, with a new pricing scheme which involves keeping track of the last time window that was hit. We test our formulation and algorithm on a variety of instances in the literature. Our algorithm is capable of solving instances of size much larger than previously reported.

Ricardo Fukasawa
University of Waterloo
rfukasaw@uwaterloo.ca

Qie He
University of Minnesota
qhe@umn.edu

Fernando Santos
Universidade Federal de Itajuba
fernandosantos@unifei.edu.br

Yongjia Song
Virginia Commonwealth University
ysong3@vcu.edu

MS10

Some Cut-Generating Functions for Second-Order

Conic Sets

We study cut generating functions for conic sets. Our first main result shows that if the conic set is bounded, then cut generating functions for integer linear programs can easily be adapted to give the integer hull of the conic integer program. Then we introduce a new class of cut generating functions which are non-decreasing with respect to second-order cone. We show that, under some minor technical conditions, these functions together with integer linear programming-based functions are sufficient to yield the integer hull of intersections of conic sections in $\mathbb{R}^2$.

Asteroide Santana
Georgia Institute of Technology
asteroide.santana@gatech.edu

Santanu S. Dey
Georgia Institute of Technology
School of Industrial and System Engineering
sdey30@isye.gatech.edu

MS10

Vehicle Routing Problems with Time Windows and Convex Node Costs

We consider a variant of the vehicle routing problems with time windows, where the objective includes the inconvenience cost modeled by a convex function on each node. We propose a novel set partitioning formulation for this mixed integer convex program using the underlying block structure of the solution, by considering all possible combinations of routes and block structures. In particular, given a route $r = (r_1, r_2, \ldots, r_n)$, we define a sequence of customers $B = \{r_b, r_{b+1}, \ldots, r_{b+n_b}\}(1 \leq b \leq b + n_b \leq n)$ as block, if there is no waiting between any of these customers. Based on the convexity of the node cost function, this block structure enables a recursive approach to calculate the optimal cost of a route, which enables separability in the labeling algorithm. This allows an efficient dominance checking between paths by only solving one dimensional convex optimization problems instead of any higher dimensional convex optimization problems. In our computational study, we show the effectiveness of our branch-and-price algorithm compared to state-of-the-art mixed integer convex programming solvers on a set of vehicle routing problems with time windows and quadratic inconvenience costs.

Yongjia Song
Virginia Commonwealth University
ysong3@vcu.edu

Qie He
University of Minnesota
qhe@umn.edu

MS10

A Polyhedral Approach to Optimal Control of Switched Linear Systems

Consider a discrete-time dynamic system $x_{k+1} = M_k x_k$ for $k = 0, 1, \ldots, T$. The matrix M_k can be chosen as one of the given two matrices A or B. The matrix selection problem is defined as follows: given an initial vector x_0, select matrices $M_1, \ldots, M_T$ such that a linear function over the states $x_1, \ldots, x_T$ is minimized. This problem is motivated by designing combination therapies for cancer treatment, and is closely related to the matrix mortality problem and switched linear systems studied by the control community. We present a poly-time dynamic programming algorithm for the case when A and B commute. We then formulate the general problem as a mixed-integer linear program, and derive several families of facet-defining inequalities for the resulting formulation.

Zeyang Wu
University of Minnesota
wuxx1164@umn.edu

MS11

Exponentially Accurate Temporal Decomposition for Long Horizon Dynamic Optimization Problem

We investigate a temporal decomposition approach to long-horizon dynamic optimization problems. The problems are discrete-time, linear, time-dependent and with bounds on the control variables. We prove that an overlapping domains temporal decomposition, while inexact, approaches the solution of the long-horizon dynamic optimization problem exponentially fast in the size of the overlap. The resulting subproblems share no solution information and thus can be computed with trivial parallelism. Our findings are demonstrated with a small, synthetic production cost model with real demand data.

Mihai Anitescu, Wanting Xu
University of Chicago
anitescu@uchicago.edu, wxu@uchicago.edu

MS11

Accelerating Newtons Method for Nonlinear Parabolic Control via Reduced-Order Modeling

We use projection based reduced order models (ROMs) to substantially decrease the computational cost of Newton's method for large-scale time-dependent optimal control problems. Previous approaches of using projection based ROMs in optimization solve sequences of computationally less expensive optimization problems, which are obtained by replacing the high fidelity discretization of the governing partial differential equation (PDE) by a ROM. This may not be feasible or efficient for several nonlinear PDEs because 1) ROMs may not be well-posed because the ROM solution does not satisfy properties obeyed by the solution of the high-fidelity PDE discretization, such as conservation properties, or non-negativity, 2) ROMs can become untrustworthy when the controls differ slightly from those at which the ROM was generated, or 3) validation that a ROM is still sufficiently accurate at a new control can be very expensive. We use high-fidelity PDE discretizations for objective function and gradient evaluation, and use state PDE and adjoint PDE solves to generate a ROM to compute approximate Hessian information. The ROM generation is computationally inexpensive, since it reuses information that is already computed for objective function and gradient evaluation, and the ROM-Hessian approximations are inexpensive to evaluate. The resulting method often enjoys convergence rates similar to those of Newton's method, but at the computational cost per iteration of the gradient method.

Matthias Heinkenschloss
Department of Computational and Applied Mathematics
Rice University
heinken@rice.edu

Caleb C. Magruder
Rice University, Computational and Applied Mathematics

cm47@rice.edu

MS11

New Multilevel Strategies in Optimal Feedback Control

The application of model-based optimization methods has become indispensable in science and engineering. Since all models contain systematic errors, the result of optimization can be very sensitive to such errors and even useless. Therefore, application of optimization to real-life processes demands taking into account uncertainties. One of the remedies is Nonlinear Model Predictive Control (NMPC), which is based on two steps: a simultaneous on-line estimation of the system state and parameters (Moving Horizon Estimation, MHE) and re-optimization of the optimal control for the current parameter and state values. The challenge is to solve these optimal control problems with high frequency in real time. During the last decade significant progress has been made in development of so-called multilevel real-time iterations to reduce the response time to a few milliseconds compared with standard methods. However, todays state-of-the-art is to perform MHE and NMPC separately. A next logical step is the development of a simultaneous MHE and NMPC in one step. This requires both the theoretical investigation of joint sensitivity, conditioning and stability analysis of coupled MHE and NMPC and an efficient generalization of the existing algorithmic approaches. In this talk we present combined multilevel real-time iterations based on coupling of the MHE and NMPC with inexact Newton methods which allows for further reduction of response times.

Ekaterina Kostina, Gregor Kriwet
Heidelberg University
ekaterina.kostina@iwr.uni-heidelberg.de,
gregor.kriwet@iwr.uni-heidelberg.de

Hans Georg Bock
IWR, University of Heidelberg
bock@iwr.uni-heidelberg.de

MS11

An Online Experimental Design Approach for Fast Detection of Extrasystoles

Networks described as graphs model many phenomena in the real world. Confronted with a spreading signal on the graph we are often interested in the (unknown) location of its origin. The available information are usually the arrival times of the signal at certain locations/nodes in the graph (measurement points). The shortest paths and linear regression are combined to estimate the source. This is a first step, but where should the measurements be placed to gain maximal information? What are requirements to have a unique estimate of the source? These questions will be addressed in our framework of linear regression on a graph. Based on these considerations an algorithm is proposed that places measurements iteratively on the nodes of the graph and gives a best guess for the source node in every iteration. Our test setting is derived from a heart disease (ventricular tachycardia), where a single source on one of the heart chambers initiates ill-timed heart beats. Treating the disease requires the exact source location before ablating it. After the heart chamber mesh is abstracted to a graph, our new method is applied. The talk concludes with well-performing simulation results.

Tobias Weber
Otto-von-Guericke Universität Magdeburg
tobias.weber@ovgu.de

Sebastian Sager
Otto-von-Guericke Universität Magdeburg, Germany
sager@ovgu.de

MS12

Sampling Stochastic Dynamic Programming for Mid-Term Hydropower Optimization

Rio Tinto Alcan (RTA) is a multinational aluminium producer with smelters in Quebec, Canada. RTA also owns and operates power houses on the Péribonka and Saguenay Rivers which supply energy to the smelters. The system, which is run by RTA's Quebec Power Operations Division, consists of 6 generating stations and 4 reservoirs, for an installed capacity of 2900 MW. One of the key decisions that have to be made for effective operation of this system is to determine the volume of water release per week at all generating stations. Stochastic Dynamic Programming (SDP) has been widely used for dealing with this kind of optimization problem. This optimization method can be applied on non-convex stochastic problems. However, the quality of the solution depends on several factors that must be carefully taken into account. For example, the temporal decomposition in SDP does not allow for modelling long time correlation in the inflows. When the system is composed of more than two reservoirs, the computation time becomes an issue and the way the cost-to-go function in SDP is calculated and interpolated as a huge impact. This presentation will show how we customized the SDP algorithm to get robust and operational software. In particular, we will show how to extend SDP to Sampling SDP for dealing with long-time correlation and how to solve efficiently the cost-to-go function in Sampling SDP.

Pascal Côté
Rio Tinto
pascal.cote@riotinto.com

James Merleau
IREQ
merleau.james@ireq.ca

Kenjy Demeester
Ecole Polytechnique Montreal
kenjy.dem@gmail.com

MS12

Optimal Control in Hydropower Management with a Simulation and Regression Approach to Dynamic Programming

Hydropower management presents complex control problems. Considering such problems as a series of real options, we develop a least-squares Monte Carlo, dynamic programming approach. The approach is applied to find optimal storage and production decisions associated to two of RioTinto's installations in Canada.

Michel Denault
HEC Montréal
michel.denault@hec.ca

MS12

**Direct Policy Search Gradient Method using Inflow

Scenarios

In the field of water reservoir operation design, stochastic dynamic programming (SDP) has been extensively used. This method imposes a time decomposition of the stochastic process to build the value functions. Thus, achieving a good modeling of the complex spatio-temporal correlation of the streamflow is tricky. The Direct Policy Search methods (DPS) methods optimize directly into the policy parameter within a given family of functions. The actual parametrization is evaluated by simulating over an ensemble of scenarios. Thus, no model for the random process is needed and the stochastic process is implicitly represented through simulation. So far, DPS is still slow to converge because it used with inefficient optimization methods such as evolutionary algorithms. We propose a method using a gradient estimation to perform the optimization step. The gradient is estimated using the Reinforce algorithm. This method does not require the derivative of the objective function but only the derivative of the policy function. The policy function used in our case is radial basis functions network. Sensibility of the DPS and improvement of the method are presented based on the real case of Kemano (British-Colombia, Canada). A comparison to SDP is performed and results suggest superiority of the DPS thanks to the better uncertainty modeling, with computation time of the same order of magnitude.

Quentin Desreumaux
University of Sherbrooke
quentin.desreumaux@usherbrooke.ca

Pascal Côté
Rio Tinto
pascal.cote@riotinto.com

Robert Leconte
University of Sherbrooke
robert.leconte@usherbrooke.ca

MS12

Numerical Assessment of an Adaptive Simplicial Approximation Method for Hydroelectric Planning

We present a novel approach for approximate stochastic dynamic programming (ASDP) over a continuous state space when the optimization phase has a near-convex structure. The approach entails a simplicial partitioning of the state space. Bounds on the true value function are used to refine the partition. The purpose of our method is to permit a user-chosen balance between computational burden and accuracy. Features of the proposed ASDP scheme are: (i) sample state grid is constructed on-line on the basis of optimality gap estimation; (ii) for revenue maximization, best piecewise linear concave fit to grid point evaluations; (iii) for each grid point, optimization using an easily computable generalized linear program; and (iv) value function converges to its concave envelope. The proposed method is suited for multi-echelon problems with stocks and flows embedded in a network structure, with possibly stochastic inflows and outflows, such as mid-term operations planning in an arborescent hydraulic energy production system. We empirically examine trade-offs between complexity and accuracy where complexity has two dimensions: number of reservoirs n and number of grid points κ. We present numerical tests to investigate (i) accuracy vs. state grid density, and (ii) accuracy vs. state space dimension. The approach is experimented on several configurations of hydroenergy systems.

Luckny Zéphyr
Cornell University
lz395@cornell.edu

Pascal Lang, Bernard F. Lamond
Université Laval
pascal.lang@fsa.ulaval.ca, bernard.lamond@fsa.ulaval.ca

Pascal Côté
Rio Tinto
pascal.cote@riotinto.com

MS13

Nonconvex Sparse Poisson Intensity Reconstruction for Time-Dependent Tomography

While convex optimization for low-light imaging has received some attention by the imaging community, nonconvex optimization techniques for photon-limited imaging are still in their nascent stages. In this talk, we propose a novel stage-based non-convex approach to recover high-resolution sparse signals from low-dimensional measurements corrupted by Poisson noise. We incorporate gradient-based information to construct a sequence of quadratic subproblems with a nonconvex ℓ_p-norm ($0 \leq p < 1$) penalty term to promote sparsity. We demonstrate the effectiveness of the proposed approach by solving two time-dependent medical imaging applications: bioluminescence tomography and fluorescence lifetime imaging.

Lasith Adhikari, Arnold D. Kim
Applied Mathematics Unit
University of California, Merced
ladhikari@ucmerced.edu, adkim@ucmerced.edu

Roummel F. Marcia
University of California, Merced
rmarcia@ucmerced.edu

MS13

Kernel-Based Reconstruction Algorithm for Anatomically-Aided Optical Tomography

Abstract not available at time of publication.

Reheman Baikejiang
University of California, Merced
rbaikejiang@ucmerced.edu

MS13

Extended Gauss-Newton-Type Methods for a Class of Matrix Optimization Problems

We develop a Gauss-Newton (GN) framework for solving a class of nonconvex optimization problems involving low-rank matrix variables. As opposed to standard GN method, our framework allows one to handle general smooth convex cost function via its surrogate. The main complexity-per-iteration consists of the inverse of two rank-size matrices and at most six small matrix multiplications to compute a closed form Gauss-Newton direction, and a backtracking linesearch. We show, under mild conditions, that the proposed algorithm globally and locally converges to a stationary point of the original nonconvex problem. We also show empirically that the Gauss-Newton algorithm achieves much higher accurate solutions compared to the well studied alternating direction method (ADM). Then, we specify our Gauss- Newton framework to handle the symmetric case and prove its convergence, where ADM is not applicable without lifting variables. Next, we incorporate our Gauss- Newton scheme into the alternating direction method of multipliers (ADMM) to design a GN-ADMM algorithm for solving the low-rank optimization problem. We prove, under a proper choice of the penalty parameter, that our GN-ADMM globally converges to a stationary point. Finally, we apply our algorithms to solve several problems in practice such as matrix completion, robust low-rank matrix recovery, and matrix recovery in quantum tomography. The experiments provide encouraging results to motivate the use of nonconvex optimization.

Quoc Tran-Dinh
Department of Statistics and Operations Research
UNC-Chapel Hill, USA
quoctd@email.unc.edu

Zheqi Zhang
Department of Statistics and Operations Research
UNC-Chapel Hill
zzhang1@live.unc.edu

MS14

Adaptive Sampling Strategies for Stochastic Optimization

Variance reduction techniques for stochastic gradient methods have recently received much attention. In this talk, we present an optimization method that achieves variance reduction by controlling sample sizes. It includes a test for detecting the transient behavior of the stochastic gradient (SG) method, and begins increasing sample sizes only after the initial phase of SG is complete. From this point forward, sample sizes are determined at every iteration by a new condition proposed in this talk. We present numerical experiments on logistic regression problems.

Raghu Bollapragada
Department of Industrial Engineering and Management Sciences
Northwestern University
vijayabollapragada2014@u.northwestern.edu

Richard H. Byrd
University of Colorado
richard@cs.colorado.edu

Jorge Nocedal
Department of Electrical and Computer Engineering
Northwestern University
j-nocedal@northwestern.edu

MS14

Exact and Inexact Subsampled Newton Methods for Optimization

We study the solution of stochastic optimization problems in which approximations to the gradient and Hessian are obtained through subsampling. We show how to coordinate the accuracy in the gradient and Hessian to yield a superlinear rate of convergence in expectation. We also consider inexact Newton methods and investigate what is the most effective linear solver in terms of computational complexity.

Jorge Nocedal
Department of Electrical and Computer Engineering
Northwestern University
j-nocedal@northwestern.edu

Raghu Bollapragada
Department of Industrial Engineering and Management Sciences
Northwestern University
vijayabollapragada2014@u.northwestern.edu

Richard H. Byrd
University of Colorado
richard@cs.colorado.edu

MS14

Random Projections for Faster Second Order Constrained Optimization Methods

In this talk we consider random projection methods for solving convex optimization problems with a convex constraint. First, we will describe a geometric relation between complexity and optimality of random projection methods and information theoretical lower bounds. Then, we will present a novel method, which iteratively refines the solutions subject to constraints and achieves statistical optimality. We then generalize our method to optimizing arbitrary convex functions of a large data matrix subject to constraints. The proposed method is a faster randomized version of the well-known Newton's Method. We show that the proposed method enables solving large scale optimization and statistical inference problems orders-of-magnitude faster than existing methods.

Mert Pilanci
Department of Mathematics
Stanford University
pilanci@stanford.edu

MS14

The Role of Over-Parametrization in Optimization with a Focus on Deep Learning

Optimization of a real valued non-convex function over a high dimensional space can be a challenge, however recent results in spin glass theory provide us with insight: Auffinger et. al., "Complexity of random smooth functions on the high-dimensional sphere", 2013 shows that the local minima of the Hamiltonian of a spherical spin system concentrate at a level that is slightly above its ground state. This concentration can empirically be observed even when there are finitely many spins. We demonstrate that the gradient descent finds those points whose energy is at the concentration level. We then move on to the experiments that observe a similar concentration when optimizing the loss functions of deep learning. In particular, we observe that the solutions found by both noisy and noise-less algorithms lead the system to similar levels for both the test and training loss values. We furthermore demonstrate that this kind of concentration is observed only when the system is large enough, in other words, when the system is over-parametrized.

Levent Sagun
Courant Institute of Mathematical Sciences
NYU

sagun@cims.nyu.edu

MS15

SVS-Free Algorithms for Low-Rank Matrix Recovery

We propose a non-convex algorithmic approach for recovering low-rank matrices from a small number of linear observations (samples). We demonstrate that our approach: (i) achieves optimal sample complexity; (ii) exhibits nearly-linear time complexity for constant-rank matrices with no extra spectral assumptions; and (iii) is stable to noise and model mismatch. To the best of our knowledge, this is the first approach that achieves all three of these properties. A distinction from existing methods is that our algorithm does not involve a full singular value decomposition (SVD), and instead uses a careful synthesis of techniques from randomized low-rank approximation schemes. Joint work with Piotr Indyk and Ludwig Schmidt.

Chinmay Hegde
Iowa State University
chinmay@iastate.edu

MS15

Nonconvex Proximal Primal Dual Methods

We propose a new decomposition approach named the proximal primal-dual algorithm (Prox-PDA) for smooth nonconvex linearly constrained optimization problems. The proposed approach is a primal-dual scheme, where the primal step minimizes certain approximation of the augmented Lagrangian of the problem, and the dual step performs an approximate dual ascent. The approximation used in the primal step is able to decompose the variable blocks, making it possible to obtain simple subproblems by leveraging the problem structures. Theoretically, we show that whenever the penalty parameter in the augmented Lagrangian is larger than a given threshold, the Prox-PDA converges to the set of stationary solutions, globally and in a sublinear manner (i.e., certain measure of stationarity decreases in the rate of $\mathcal{O}(\infty/\nabla)$, where r is the iteration counter). Interestingly, when applying a variant of the Prox-PDA to the problem of distributed nonconvex optimization (over a connected undirected graph), the resulting algorithm coincides with the popular EXTRA algorithm [Shi *et al* 2014], which is only known to work in *convex* cases. Our analysis implies that EXTRA and a few of its variants converge globally sublinearly to stationary solutions of certain nonconvex distributed optimization problem. There are many possible ways to generalize the Prox-PDA, and we present one such generalization and apply it to certain nonconvex distributed matrix factorization problem.

Mingyi Hong
Iowa State University
mingyi@iastate.edu

MS15

Stochastic Gradient Descent for Nonconvex Problems

Abstract not available at time of publication.

George Lan
Dept of ISE
University of Florida
glan@ise.ufl.edu

MS15

Distributed Stochastic Second-Order Method for Training Large Scale Deep Neural Networks

Training deep neural network is a high dimensional and a highly non-convex optimization problem. Stochastic gradient descent (SGD) algorithm and it's variations are the current state-of-the-art solvers for this task. However, due to non-convexity nature of the problem, it was observed that SGD slows down near saddle point. Recent empirical work claims that by detecting and escaping saddle point efficiently, it's more likely to improve training performance. With this objective, we revisit Hessian-free and Quasi-Newton methods for deep networks. We show that those methods obtain almost linear scaling and do not suffer from extensive communication when compared to distributed SGD methods.

Xi He
Lehigh University
xih314@lehigh.edu

Dheevatsa Mudigere
Parallel Computing Laboratory
Intel Corporation, Bangalore, India
dheevatsa.mudigere@intel.com

Mikhail Smelyanskiy
Intel Corporation
mikhail.smelyanskiy@intel.com

Martin Takac
University of Edinburgh
takac.mt@gmail.com

MS16

Characterization of the Robust Isolated Calmness for a Class of Conic Programming

This talk is devoted to studying the robust isolated calmness of the Karush-Kuhn-Tucker (KKT) solution mapping for a large class of interesting conic programming problems (including most commonly known ones arising from applications) at a locally optimal solution. Under the Robinson constraint qualification, we show that the KKT solution mapping is robustly isolated calm if and only if both the strict Robinson constraint qualification and the second order sufficient condition hold. This implies, among others, that at a locally optimal solution the second order sufficient condition is needed for the KKT solution mapping to have the Aubin property.

Chao Ding
National University of Singapore
dingchao@amss.ac.cn

Defeng Sun
Dept of Mathematics
National University of Singapore
matsundf@nus.edu.sg

Liwei Zhang
Dalian University of Technology

lwzhang@dlut.edu.cn

MS16

Primal-Dual Algorithm for Symmetric Programming Problems with Nonlinear Objective Functions

We describe a primal-dual algorithm for symmetric programming problems with nonlinear objective functions. A class of nonlinear objective functions is described for which the complexity estimates are (qualitatively) the same as for the case of linear objective functions. The class includes quantum entropy function arising in applications ranging from quantum physics to quantum information theory.

Leonid Faybusovich
Department of Mathematics
University of Notre Dame
leonid.faybusovich.1@nd.edu

MS16

Inner and Outer Approximations of the Semidefinite Cone Using SD Bases and their Applications to Some NP-Hard Problems

Some of the authors in previous paper introduced a semidefinite basis which is a basis of the space of symmetric matrices consisting of rank-1 semidefinite matrices. In this talk, we present polyhedral inner and outer approximations of the positive semidefinite cone and the completely positive cone by using semidefinite bases. The conical hull of a semidefinite basis gives an inner approximation of the positive semidefinite cone. If elements of a semidefinite basis are nonnegative matrices, it gives an inner approximation of the completely positive cone. Supporting hyperplanes at every point of a semidefinite basis configure an outer approximation of these cones. Preliminary numerical experiments suggest that linear optimization problems over our polyhedral cones give a good upper and lower bounds of standard quadratic programming and randomly generated doubly nonnegative programming.

Daigo Narushima
Tsukuba University
s1620483@sk.tsukuba.ac.jp

Akihiro Tanaka
University of Tsukuba
Faculty of Engineering, Information and Systems
tanaka.akihiro@sk.tsukuba.ac.jp

Akiko Yoshise
University of Tsukuba
Graduate School of Systems and Information Engineering
yoshise@sk.tsukuba.ac.jp

MS16

New Developments in Primal-Dual Interior-Point Algorithms for Convex Optimization

We will discuss some of the recent developments in primal-dual interior-point algorithms for convex optimization problems in conic form as well as in some other forms. For these approaches we will analyze the advantages and disadvantages of (i) utilizing only first-order information versus occasionally some second-order information and also (ii) enforcing, to various degrees, primal-dual symmetry of the underlying algorithms.

Levent Tuncel
University of Waterloo
Dept. of Combinatorics and Optimization
ltuncel@uwaterloo.ca

MS17

Tensor Eigenvalue Complementarity Problems

This talk discusses tensor eigenvalue complementarity problems. Basic properties of standard and complementarity tensor eigenvalues are given. We formulate tensor eigenvalue complementarity problems as constrained polynomial optimization. When one tensor is strictly copositive, the complementarity eigenvalues can be computed by solving polynomial optimization with normalization by strict copositivity. When no tensor is strictly copositive, we formulate the tensor eigenvalue complementarity problem equivalently as polynomial optimization by a randomization process. The complementarity eigenvalues can be computed sequentially. The formulated polynomial optimization can be solved by Lasserre's hierarchy of semidefinite relaxations. We show that it has finite convergence for general tensors.

Jinyan Fan
Department of Mathematics, Shanghai Jiao Tong University
Shanghai 200240, P.R. China
jyfan@sjtu.edu.cn

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

Anwa Zhou
Shanghai University, China
zhouanwa@shu.edu.cn

MS17

More NP-Hard Tensor Problems

We discuss exciting recent developments in the computational complexity of various tensor problems: (i) Shitov has shown that tensor rank over integers is undecidable and symmetric tensor rank is NP-hard, conjectured in earlier work of Hillar and Lim; (ii) Schaefer and Stefankovic have shown that determining tensor rank over any field F is ∃F-complete, providing an alternative proof for the NP-completeness of tensor rank over finite fields and a refinement of its NP-hardness over other fields; (iii) Friedland and Lim have shown that tensor nuclear norm is NP-hard for tensors of order 3 or higher over reals, and for tensors of order 4 or higher over complex (which remains true even if restricted to bi-Hermitian, bisymmetric, positive semidefinite, nonnegative-valued 4-tensors). Time permitting, we will discuss the complexity of more specific tensor problems in optimization such as deciding self-concordance of a convex function or membership problems in SOS and SOS-convex cones of quartic polynomials. We will suggest open problems on the complexity of two newly discovered notions — tensor nuclear rank and tensor network ranks.

Lek-Heng Lim
University of Chicago

lekheng@galton.uchicago.edu

MS17

Geometric Measure of Entanglement of Symmetric D-Qubits is Polynomial-Time Computable

We give a simple formula for finding the spectral norm of d-mode symmetric tensor in two variables over the complex or real numbers in terms of the complex or real roots of a corresponding polynomial in one complex variable. This result implies that the geometric measure of entanglement of symmetric d-qubits is polynomial-time computable. We discuss a generalization to d-mode symmetric tensor in more than two variables.

Shmuel Friedland
university of illinois at chicago
friedlan@uic.edu

Li Wang
University of Illinois at Chicago
liwang8@uic.edu

MS17

Real Eigenvalues of Nonsymmetric Tensors

This talk discusses the computation of real Z-eigenvalues of nonsymmetric tensors. A general nonsymmetric tensor has finitely many Z-eigenvalues, while there may be infinitely many ones for special tensors. We propose Lasserre type semidefinite relaxation methods for computing such eigenvalues. For every nonsymmetric tensor that has finitely many real Z-eigenvalues, we can compute all of them; each of them can be computed by solving a finite sequence of semidefinite relaxations.

Xinzhen Zhang
Department of Mathematics
Tianjin University
xzzhang@tju.edu.cn

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

MS18

High-Order Taylor Expansions for Compressible Flows

Sensitivity analysis for systems governed by partial differential equations is now commonly used by engineers to assess performance modification due to parameter changes. A typical illustration concerns shape optimization procedures based on the adjoint method, used in aeronautics to improve aerodynamic or structural performance of aircrafts. However, these approaches are usually limited to first-order derivatives and steady PDE systems, due to the complexity to extend the adjoint method to higher-order derivatives and the associated reverse time integration. Alternatively, this work investigates the use of the direct differentiation approach (continuous sensitivity equation method) to estimate high-order derivatives for unsteady flows. We show how this method can be efficiently implemented in existing solvers, in the perspective of providing a Taylor expansion of the PDE solution with respect to control parameters. Applications to optimization and uncertainty estimation are finally considered.

Regis Duvigneau
INRIA Sophia Antipolis
regis.duvigneau@inria.fr

MS18

A Comparison of Regularization Methods for Gaussian Processes

Gaussian Processes (GPs) are classical probabilistic models to represent the results of experiments on grids of points. They have numerous applications, in particular in nonlinear global optimization when the experiments (typically PDE simulations) are costly. GPs require the inversion of a covariance matrix. There are many situations, in particular optimization, when the density of experiments becomes higher in some regions of the search space, which makes the covariance matrix ill-conditionned, an issue which is handled in general through regularization techniques. Today, the need to better understand and improve regularization remains. The two most classical regularization methods are i) pseudoinverse (PI) and ii) adding a small positive constant to the main diagonal (which is called nugget regularization). This work provides new algebraic insights into PI and nugget regularizations. It is proven that pseudoinverse regularization averages the output values and makes the variance null at redundant points. On the opposite, nugget regularization lacks interpolation properties but preserves a non-zero variance at every point. However, these two regularization techniques become similar as the nugget value decreases. A new distribution-wise GP is then introduced which interpolates Gaussian distributions instead of data points and mitigates the drawbacks of pseudoinverse and nugget regularized GPs.

Rodolphe Le Riche
Ecole des Mines de St Etienne
leriche@emse.fr

Hossein Mohammadi, Nicolas Durrande, Eric Touboul, Xavier Bay
Ecole des Mines de St-Etienne
hossein.mohammadi@emse.fr, nicolas.durrande@emse.fr, touboul@emse.fr, bay@emse.fr

MS18

Bayesian Optimization Approaches to Compute Nash Equilibria

Game theory finds nowadays a broad range of applications in engineering and machine learning. However, in a derivative-free, expensive black-box context, very few algorithmic solutions are available to find game equilibria. Here, we propose a novel Gaussian-process based approach for solving games in this context. We follow a classical Bayesian optimization framework, with sequential sampling decisions based on acquisition functions. Two strategies are proposed, based either on the probability of achieving equilibrium or on the Stepwise Uncertainty Reduction paradigm. Practical and numerical aspects are discussed in order to enhance the scalability and reduce computation time. Our approach is evaluated on several synthetic game problems with varying number of players and decision space dimensions. We show that equilibria can be found reliably for a fraction of the cost (in terms of black-box evaluations) compared to classical, derivative-based algorithms.

Victor Picheny
MIAT Universite de Toulouse and INRA
victor.picheny@toulouse.inra.fr

Mickaël Binois
University of Chicago
mbinois@uchicago.edu

Abderrahmane Habbal
UNIVERSITY COTE D'AZUR UCA, UNS, CNRS and Inria, FRANCE
habbal@unice.fr

MS18

Surrogates and Classification Approaches for Efficient Global Optimization (EGO) with Inequality Constraints

In this work, we compare the use of Gaussian Process (GP) models for the constraints [Schonlau 1997] with a classification approach relying on a Least-Squares Support Vector Machine (LS-SVM) [Suykens and Vandewalle 1999]. We propose several adaptations of the classification approach in order to improve the efficiency of the EGO procedure, in particular an extension of the binary LS-SVM classifier to come-up with a probabilistic estimation of the feasible domain. The efficiencies of the GP models and classification methods are compared in term of computational complexities, distinguishing the construction of the GPmodels or LS-SVM classifier from the resolution of the optimization problem. The effect of the number of design parameters on the numerical costs is also investigated. The approaches are tested on the optimization of a complex non-linear Fluid-Structure Interaction system modeling a two dimensional flexible hydrofoil. Multi-design variables, defining the unloaded geometry of the foil and the characteristics of its elastic trailing edge, are used in the minimization of the foils drag, under constraints set to ensure minimal lift force and prevent cavitation at selected boat-speeds.

Matthieu Sacher
IRENAV
Ecole Navale
matthieu.sacher@ecole-navale.fr

Regis Duvigneau
INRIA Sophia Antipolis
regis.duvigneau@inria.fr

Olivier Le Maitre
Laboratoire d'Informatique pour la Mécanique et les Science
olm@limsi.fr

Mathieu Durand
K-EPSILON
mathieu@k-epsilon.com

Elisa Berrini
Université Côte d'Azur, INRIA, CNRS
elisa.berrini@inria.fr

Frederic Hauville, Jascque-Andre Astolfi
IRENAV
Ecole Navale
frederic.hauville@ecole-navale.fr, jacques-andre.astolfi@ecole-navale.fr

MS19

Ellipsoid Method vs Accelerated Gradient Descent

We propose a new method for unconstrained optimization of a smooth and strongly convex function, which attains the optimal rate of convergence of Nesterovs accelerated gradient descent. The new algorithm has a simple geometric interpretation, loosely inspired by the ellipsoid method. We provide some numerical evidence that the new method can be superior to Nesterovs accelerated gradient descent.

Sébastien Bubeck, Yin Tat Lee, Mohit Singh
Microsoft Research
sebubeck@microsoft.com, yintatlee@gmail.com, mohits@microsoft.com

MS19

A New Perspective on Boosting in Linear Regression via First Order Methods

Boosting is one of the most powerful and popular tools in machine learning/ statistics that is widely used in practice. They work extremely well in a variety of applications. However little is known about many of the statistical and computational properties of the algorithm, and in particular their interplay. We analyze boosting algorithms in linear regression from the perspective modern first-order methods in convex optimization. We show that classic boosting algorithms in linear regression, namely the incremental forward stagewise algorithm (FSe) and least squares boosting (LS-Boost-e), can be viewed as subgradient descent to minimize the maximum absolute correlation between features and residuals. We also propose a modification of FSe that yields an algorithm for the LASSO, and that computes the LASSO path. We derive novel comprehensive computational guarantees for all of these boosting algorithms, which provide, for the first time, a precise theoretical description of the amount of data-fidelity and regularization imparted by running a boosting algorithm with a pre-specified learning rate for a fixed but arbitrary number of iterations, for any dataset.

Rahul Mazumder
MIT
rahulmaz@mit.edu

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

Paul Grigas
UC Berkeley
pgrigas@berkeley.edu

MS19

An Optimal First-order Method Based on Optimal Quadratic Averaging

In a recent paper, Bubeck, Lee, and Singh introduced a new first order method for minimizing smooth strongly convex functions. Their geometric descent algorithm, largely inspired by the ellipsoid method, enjoys the optimal linear rate of convergence. We show that the same iterate sequence is generated by a scheme that in each iteration computes an optimal average of quadratic lower-models of the function. Indeed, the minimum of the averaged quadratic approaches the true minimum at an optimal rate. This intuitive viewpoint reveals clear connections to the original fast-gradient methods and cutting plane ideas, and leads to limited-memory extensions with improved performance.

Scott Roy
University of Washington
scottroy@uw.edu

Dmitriy Drusvyatskiy
University of Washington
University of Washington
ddrusv@uw.edu

Maryam Fazel
University of Washington
Electrical Engineering
mfazel@ee.washington.edu

MS19

A Unified Convergence Bound for Conjugate Gradient and Accelerated Gradient

We propose a single potential quantity that decreases at a constant rate on every iteration for both Nesterov's accelerated gradient method for smooth, strictly convex functions and linear conjugate gradient for convex quadratic functions. Analysis of this potential leads to the optimal convergence rate for both methods, namely, accuracy of ϵ is attained in time proportional to $\log(\epsilon)$ and the square root of L/ℓ, where L and ℓ are the parameters of strict convexity. The purpose of this analysis is to suggest new nonlinear extensions of conjugate gradient.

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

Sahar Karimi
University of Waterloo
sahar.karimi@gmail.com

MS20

Smoothing Technique for Nonsmooth Composite Minimization with Linear Operator

We introduce and analyze an algorithm for the minimization of convex functions that are the sum of differentiable terms and proximable terms composed with linear operators. The method builds upon the recently developed smoothed gap technique. In addition to a precise convergence speed result, valid even in the presence of linear equality constraints, it allows an explicit treatment of the gradient of differentiable functions and is equipped with a line search. We are also studying the consequences of restarting the algorithm at a given frequency. These features are not classical for primal-dual methods and allow us to solve difficult large scale convex optimization problems. We illustrate the performances of the algorithm on basis pursuit, TV-regularized least squares regression and L1 regression problems.

Volkan Cevher
EPFL
volkan.cevher@epfl.ch

Olivier Fercoq
Telecom-ParisTech, CNRS LTCI
Université Paris-Saclay,75013, Paris, France
olivier.fercoq@telecom-paristech.fr

Quang Nguyen
EPFL
quangnvmath@gmail.com

MS20

Let's Make Block Coordinate Descent Go Fast!

Block coordinate descent (BCD) methods are widely-used for large-scale numerical optimization. They are often ideal for large scale optimization problems because of their cheap iteration costs, low memory requirements, and amenability to parallelization. Three main algorithmic choices influence the performance of BCD methods; the block partitioning strategy, the block selection rule, and the block update rule. In this work we explore all three of these building blocks and propose variations for each that can lead to significantly faster BCD methods. We (i) propose new greedy block-selection strategies and provide a general convergence result that holds for general block partitioning and selection strategies; (ii) explore block update strategies that exploit problem structure or higher-order information to yield faster local convergence rates; (iii) consider optimal active manifold identification and superlinear or finite termination for problems with sparse solutions; and (iv) explore the use of message-passing to efficiently compute optimal block updates for problems with a certain sparsity structure. We support all of our findings with numerical results for the classic machine learning problems of logistic regression, multi-class logistic regression, support vector machines, and label propagation.

Julie Nutini, Issam Laradji, Mark Schmidt
University of British Columbia
jnutini@cs.ubc.ca, issamou(at)cs.ubc.ca, schmidtm@cs.ubc.ca

MS20

Efficiency of Coordinate Descent on Structured Problems

We present complexity bounds for a variant of the accelerated coordinate descent method and show how these estimates can be obtained using the technique of randomized estimate sequences. Our results show that coordinate decent methods often outperform the standard fast gradient methods on optimization problems with dense data. In many important situations, the computational expenses of the oracle and the method itself at each iteration of our scheme are perfectly balanced (both depend linearly on dimensions of the problem). As an application, we consider unconstrained convex quadratic optimization and problems arising from the application of the smoothing technique to nonsmooth problems. On some special problem instances, the provable acceleration factor with respect to the fast gradient method can reach the square root of the number of variables.

Sebastian Stich
Université catholique de Louvain
sebastian.stich@epfl.ch

Yurii Nesterov
CORE
Universite catholique de Louvain

yurii.nesterov@uclouvain.be

MS20

Flexible Coordinate Descent

I will present a novel randomized block coordinate descent method for the minimization of a convex composite objective function. The method uses (approximate) partial second-order (curvature) information, so that the algorithm performance is more robust when applied to highly nonseparable or ill conditioned problems. We call the method Flexible Coordinate Descent (FCD). At each iteration of FCD, a block of coordinates is sampled randomly, a quadratic model is formed about that block and the model is minimized approximately/ inexactly to determine the search direction. An inexpensive line search is then employed to ensure a monotonic decrease in the objective function and acceptance of large step sizes. I present preliminary numerical results to demonstrate the practical performance of the method.

Rachael Tappenden
Edinburgh
rachael.tappenden@canterbury.ac.nz

MS21

Ultrasound Tomography for Breast Cancer Detection

We present imaging techniques for breast cancer detection with ultrasound, which can be stated as a PDE-constrained optimization problem governed by the (visco-)acoustic wave equation. We discuss models and algorithmic strategies for the following three steps towards creating a tomographic image of the human breast tissues. First, the acquisition geometry of sending and receiving ultrasound transducers is determined using A-optimal experimental design with a randomized trace estimator. To reduce the computational costs, we create initial models from ray-based time-of-flight tomography and utilize spatial symmetries to extrapolate a 3D experimental design from a set of 2D configurations. Second, the emitting transducer elements acting as external pressure sources need to be calibrated, which amounts in solving a linear-quadratic control problem. In a third step, we discuss trust-region methods with inexactly computed derivatives to image the sound speed of the breast tissue. Here, the amount of measurement data and frequency bandwidths are gradually increased during the inversion subject to homogenization constraints. This enables us to increase the resolution without losing the long-wavelength components of the image, and to improve the computational efficiency. We show numerical examples using Salvus, an open-source package for full-waveform modeling and inversion based on PETSc and Eigen.

Christian Boehm, Naiara Korta Martiartu, Andreas Fichtner
ETH Zurich
christian.boehm@erdw.ethz.ch, naiara.korta@erdw.ethz.ch, andreas.fichtner@erdw.ethz.ch

MS21

On Optimal Experimental Design Problems for PDEs

Optimal experimental design (OED) tries to improve the accuracy of parameter estimation procedures under the influence of measurement errors. In this presentation we address OED problems in which the underlying state dynamics are given by an elliptic or parabolic partial differential equation. Improvements in the estimation accuracy are achieved by optimizing the experimental conditions, which may comprise initial conditions, volume or boundary control inputs, and sensor locations. We address formulations and algorithms for the solution of OED problems for PDEs, focusing on aspects related to the high dimension of the discretized problem.

Roland Herzog
Technische Universität Chemnitz
roland.herzog@mathematik.tu-chemnitz.de

MS21

Nonlinear Robust PDE Constrained Optimization Using a Second Order Approximation Technique and Model Order Reduction

We investigate a nonlinear constrained optimization problem with uncertain parameters. Using a robust worst-case formulation we obtain an optimization problem of bi-level structure. This type of problems are difficult to treat computationally and hence suitable solution methods have to be developed. We propose and investigate a strategy utilizing a quadratic approximation of the robust formulation as a surrogate model combined with an adaptive strategy to control the introduced error. The developed method is then applied to a shape optimization problem governed by a nonlinear partial differential equation with application in electric motor design. The goal is to optimize the volume and position of the permanent magnet in the rotor of the machine while maintaining a given performance level. These quantities are computed from the magnetic vector potentials given by the magnetostatic approximation of Maxwell's equation with transient movement of the rotor. Utilizing the robust optimization framework we account for uncertainties in material and production precision. The robustification of the optimization problem lead to high computational cost. Model order reduction is a promising choice when working in the context of partial differential equations. By generating reliable reduced order models with a posteriori error control the goal is to accelerate the computation. Numerical results are presented to validate the presented approach.

Oliver Lass, Stefan Ulbrich
Technische Universitaet Darmstadt
Fachbereich Mathematik
lass@mathematik.tu-darmstadt.de,
ulbrich@mathematik.tu-darmstadt.de

MS21

Approximation and Reduction of Optimal Control Problems in Infinite Dimension

This talk is concerned with the optimal control of nonlinear evolution equations in Hilbert spaces. The approach fits into the long tradition of seeking for slaving relationships between the small scales and the large ones but differ by the introduction of a new type of manifolds to do so, namely the finite-horizon parameterizing manifolds introduced by Chekroun and Liu (2015). A control in feedback form will be derived from the reduced HJB equation associated with the corresponding optimal control problem for the surrogate system. Rigorous error estimates will be also derived and application to a delay equation will be discussed. This is joint work with M. D. Chekroun (University of California, Los Angeles) and Axel Kröner (INRIA Saclay and CMAP, Ecole polytechnique, France).

Mickael Chekroun
University of California, Los Angeles
Department of Atmospheric and Oceanic Sciences
mchekroun@atmos.ucla.edu

Axel Kroener
INRIA Saclay and CMAP, Ecole Polytechnique
axel.kroener@inria.fr

Honghu Liu
Virginia Tech, Department of Mathematics
hhliu@vt.edu

MS22

Compact Representation of Quasi-Newton Matrices

Quasi-Newton methods for minimizing a continuously differentiable function generate a sequence of matrix approximations to the second derivative. Perhaps the most well-known class of quasi-Newton matrices is the Broyden family of updates. Two of the most widely used updates from this family of updates include the Broyden-Fletcher-Goldfarb-Shanno update and the symmetric rank-one update. Byrd et al. (1994) formulated a compact matrix representation of these updates. In this talk, we extend and generalize this result to (i) the Davidon-Fletcher-Powell update (ii) the restricted Broyden class of updates, and (iii) the full Broyden class of updates.

Omar DeGuchy
Applied Mathematics Unit
University of California, Merced
odeguchy@ucmerced.edu

Jennifer Erway
Wake Forest University
erwayjb@wfu.edu

Roummel F. Marcia
University of California, Merced
rmarcia@ucmerced.edu

MS22

An Enhanced Truncated-Newton Algorithm for Large-Scale Optimization

I present a truncated-Newton algorithm for solving unconstrained optimization problems. In many ways the method is similar to other truncated-Newton methods: apply linear CG to the Newton system, and terminate when it is deemed appropriate. The first contribution of this work is the inclusion of a termination condition that allows for a stronger local convergence theory that does not require any prior knowledge of the smallest eigenvalue of the Hessian matrix at a local minimizer. The second contribution is the use of a certain rank-1 modification to the Hessian that allows for a stronger global convergence theory that does not make any assumption about the boundedness of the Hessian matrices over the sequence of computed iterates. Numerical experience will be reported.

Daniel Robinson
Johns Hopkins University
daniel.p.robinson@gmail.com

MS22

Large-Scale Linear Algebra and its Role in Optimization

Numerical optimization has always needed reliable numerical linear algebra. Some worries go away when we use quad-precision floating-point (as is now possible with the GCC and Intel compilers). quad-MINOS has proved useful in systems biology for optimizing multiscale metabolic networks, and in computer experimental design for optimizing the IMSPE function with finite-difference gradients (confirming the existence of twin points). Recently quad-MINOS gave a surprising solution to the optimal control problem spring200. Quad versions of SNOPT and PDCO have been developed in f90 and C++ respectively. Changing a single line of source code leads to double-SNOPT or quad-SNOPT. The C++ PDCO can switch from double to quad at runtime.

Michael A. Saunders
Systems Optimization Laboratory, Stanford University
saunders@stanford.edu

MS22

Sequential Quadratic Programming Methods for Nonlinear Optimization

In this talk, we discuss the development of the optimization software package DNOPT for solving dense, small- to medium-scale nonlinear problems. The method is a sequential quadratic programming (SQP) method that utilizes either a quasi-Newton approximation of the Hessian of the Lagrangian or exact second derivative information in a rank-enforcing, inertia-controlling active-set method to solve the QP subproblems. Practical issues involving the formulation and implementation of the method will also be considered. Numerical results will be presented.

Elizabeth Wong, Philip E. Gill
University of California, San Diego
Department of Mathematics
elwong@ucsd.edu, pgill@ucsd.edu

MS23

L_p-Norm Regularization Algorithms for Optimization Over Permutation Matrices

Optimization problems over permutation matrices appear widely in facility layout, chip design, scheduling, pattern recognition, computer vision, graph matching, etc. Since this problem is NP-hard due to the combinatorial nature of permutation matrices, we relax the variable to be the more tractable doubly stochastic matrices and add an L_p-norm ($0 < p < 1$) regularization term to the objective function. The optimal solutions of the L_p-regularized problem are the same as the original problem if the regularization parameter is sufficiently large. A lower bound estimation of the nonzero entries of the stationary points and some connections between the local minimizers and the permutation matrices are further established. Then we propose an L_p regularization algorithm with local refinements. The algorithm approximately solves a sequence of L_p regularization subproblems by the projected gradient method using a nonmonotone line search with the Barzilai-Borwein step sizes. Its performance can be further improved if it is combined with certain local search methods, the cutting plane techniques as well as a new negative proximal point scheme. Extensive numerical results on QAPLIB show that our proposed algorithms can often find reasonably high quality solutions within a competitive amount of time.

Bo Jiang
School of Mathematical Sciences
Nanjing Normal University
jiangbo@njnu.edu.cn

Ya-Feng Liu
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
yafliu@lsec.cc.ac.cn

Zaiwen Wen
Peking University
Beijing International Center For Mathematical Research
wenzw@math.pku.edu.cn

MS23

Joint Power and Admission Control: A Sparse Optimization Perspective

In the first part of this talk, we shall consider the joint power and admission control problem for a wireless network consisting of multiple interfering links where the channel state information is assumed to be perfectly known. The goal is to support a maximum number of links at their specified signal to interference plus noise ratio (SINR) targets while using a minimum total transmission power. We first reformulate this NP-hard problem as a sparse ℓ_0-minimization problem and then approximate it by ℓ_1 convex approximation. We show that the ℓ_1 convex approximation can be equivalently reformulated as a linear program (without introducing any auxiliary variables) by exploiting the special structure of the problem. We use the solution of the above LP to iteratively remove strong interfering links (deflation). Numerical simulations show that the proposed approach compares favorably with the existing approaches in terms of the number of supported links, the total transmission power, and the execution time. In the second part of this talk, we shall discuss some interesting extensions of the first part. For instance, instead of doing convex (linear programming) approximations, we can do non-convex approximations to improve the approximation performance. If time is allowed, we shall also discuss how to extend the algorithm developed in the first part to the imperfect channel state information scenario.

Ya-Feng Liu
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
yafliu@lsec.cc.ac.cn

MS23

Randomized Block Proximal Damped Newton Method for Composite Self-Concordant Minimization

We consider the composite self-concordant (CSC) minimization problem, which minimizes the sum of a self-concordant function f and a (possibly nonsmooth) proper closed convex function g. The CSC minimization is the cornerstone of the path-following interior point methods for solving a broad class of convex optimization problems. It has also found numerous applications in machine learning. The proximal damped Newton (PDN) methods have been well studied in the literature for solving this problem and they enjoy a nice iteration complexity. Given that at each iteration these methods typically require evaluating or accessing the Hessian of f and also need to solve a proximal Newton subproblem, the cost per iteration can be prohibitively high for solving large-scale problems. To alleviate this difficulty, we propose a randomized block proximal damped Newton (RBPDN) method for solving the CSC minimization. Compared to the PDN methods, the computational cost per iteration of RBPDN is usually significantly lower. The computational experiment on a class of regularized logistic regression problems demonstrate that RBPDN is indeed promising in solving large-scale CSC minimization problems. The convergence of RBPDN is also analyzed

Zhaosong Lu
Department of Mathematics
Simon Fraser University
zhaosong@sfu.ca

MS23

A Distributed Semi-Asynchronous ADMM Type Algorithm for Network Traffic Engineering

In this work, we consider the traffic engineering problem in a large-scale hierarchical network arising in the next generation cloud-based wireless networks. We propose a distributed semi-asynchronous algorithm for this problem based on the so-called block successive upper bound minimization method of multipliers (BSUM-M). Theoretically, we show that the proposed algorithm converges to the global optimal solution under some assumptions on the degree of network asynchrony. We illustrate the effectiveness and efficiency of the proposed algorithm by comparing it with the state of the art commercial solvers in a networked environment.

Zhi-Quan Luo
Univerisity of Minnesota
Minneapolis, Minnesota
luozq@ece.umn.edu

MS24

An Accelerated Proximal Method for Minimizing Compositions of Convex Functions with Smooth Maps

I will describe an accelerated algorithm for minimizing compositions of finite-valued convex functions with smooth mappings. When applied to optimization problems having an additive composite form, the algorithm reduces to the method of Ghadimi and Lan. The method both realizes the best known complexity bound in optimality conditions for nonconvex problems with bounded domain, and achieves an accelerated rate under standard convexity assumptions. A natural convexity parameter of the composition quantifies the transition between the two modes of convergence. I end by analyzing the effect of inexact proximal subproblem solves on global performance – an essential feature of the composite setting – and compare the resulting scheme with algorithms based on smoothing.

Courtney Kempton
University of Washington
yumiko88@math.washington.edu

Dmitriy Drusvyatskiy
University of Washington
University of Washington

ddrusv@uw.edu

MS24

Nonasymptotic and Asymptotic Linear Convergence of an Almost Cyclic Shqp Dykstra's Algorithm for Polyhedral Problems

We show that an almost cyclic (or generalized Gauss-Seidel) Dykstra's algorithm which incorporates the SHQP (supporting halfspace- quadratic programming) strategy can achieve nonasymptotic and asymptotic linear convergence for polyhedral problems.

C.H. Jeffrey Pang
Mathematics, National University of Singapore
matpchj@nus.edu.sg

MS24

Whats the Shape of your Penalty?

Performance of machine learning approaches is strongly influenced by choice of misfit penalty, and correct settings of penalty parameters, such as the threshold of the Huber function. These parameter are typically chosen using expert knowledge, cross-validation, or black-box optimization, which are time consuming for large-scale applications. We present a data-driven approach that simultaneously solves inference problems and learns error structure and penalty parameters. We discuss theoretical properties of these joint problems, and present algorithms for their solution. We show numerical examples from the piecewise linear-quadratic (PLQ) family of penalties.

Zheng Peng, Aleksandr Aravkin
University of Washington
zhengp@uw.edu, sasha.aravkin@gmail.com

MS24

Optimization in Structure from Motion

Estimating the accurate poses of cameras and locations of 3D scene points from a collection of images obtained by the cameras is a classic problem in computer vision, referred to as "structure from motion" (SfM). There are two components in this pipeline: (a) estimating corresponding points between images, a.k.a., correspondence matching, (b) recovering 3D points and camera parameters from the corresponding 2D image points, a.k.a., bundle adjustment. Both these components have their unique optimization formulations, which become particularly challenging in the presence of noisy correspondences and as the scale of the dataset grows. We will touch upon the general optimization principles in these problems and discuss existing solutions along with their limitations. We will also point to exciting areas of opportunity for developing robust, large-scale solutions to these problems.

Karthik Ramamurthy
TBD
tbd2@tbd2.org

MS25

Pajarito: An Open-Source Mixed-Integer Conic Solver

Pajarito is an open-source solver written in Julia, integrated with the JuliaOpt ecosystem, and accessible through powerful algebraic modeling systems such as JuMP, Convex.jl, and CVXPY. Pajarito solves mixed-integer convex optimization problems (MICPs), which have some discrete variables but are convex when discreteness is relaxed. Unlike existing MICP solvers, which typically use nonlinear branch-and-bound and solve continuous relaxations at nodes, Pajarito performs iterative outer approximation on an extended formulation. Specifically, it constructs and refines a polyhedral relaxation in a higher-dimensional space, allowing it to converge more rapidly to the optimal MICP solution. A key contribution is the use of disciplined convex programming (DCP), an algebraic modeling concept designed to access powerful conic solvers via epigraph reformulations, to automate the construction of a conic extended formulation. Pajarito can solve MICPs with any mixture of second-order, exponential, and positive-semidefinite cones, and is easily extensible to other cones. We demonstrate that Pajarito is the fastest open-source mixed-integer nonlinear solver and illustrate some applications of mixed-integer semidefinite programming (MISDP).

Chris Coey
MIT
coey.chris@gmail.com

Miles Lubin
Operations Research Center
Massachusetts Institute of Technology
mlubin@mit.edu

Juan Pablo Vielma
Sloan School of Management
Massachusetts Institute of Technology
jvielma@mit.edu

MS25

SCIP-SDP: A Framework for Solving Mixed-Integer Semidefinite Programs

Mixed-integer semidefinite programs arise in many applications and several problem-specific solution approaches have been studied recently. In this talk, we present a generic branch-and-bound framework for solving such problems. We will discuss results regarding strong duality and the existence of relative interior points for SDPs appearing in a branch-and-bound tree and their implications for a practical implementation of an MISDP-solver. Furthermore, we will present techniques used to accelerate the solving process of SCIP-SDP. Finally, we will show numerical results examining the occurrence of the Slater condition in practical applications of mixed-integer semidefinite programming and demonstrating the applicability of the framework as well as the impact of the discussed techniques.

Tristan Gally
Technische Universität Darmstadt
gally@mathematik.tu-darmstadt.de

Marc E. Pfetsch, Stefan Ulbrich
Technische Universitaet Darmstadt
Fachbereich Mathematik
pfetsch@mathematik.tu-darmstadt.de,
ulbrich@mathematik.tu-darmstadt.de

MS25

Valid Quadratic Inequalities for Some Non-Convex Quadratic Sets

In recent years, the generalization of Balas disjunctive cuts for mixed integer linear optimization problems to mixed integer non-linear optimization problems has received significant attention. Among these studies, mixed integer second order cone optimization (MISOCO) is a special case. For MISOCO one has the disjunctive conic cuts approach. That generalization introduced the concept of disjunctive conic cuts (DCCs) and disjunctive cylindrical cuts (DCyCs). Specifically, it showed that under some mild assumptions the intersection of those DCCs and DCyCs with a closed convex set, given as the intersection of a second order cone and an affine set, is the convex hull of the intersection of the same set with a parallel linear disjunction. The key element in that analysis was the use of pencils of quadrics to find close forms for deriving the DCCs and DCyCs. In this work we use the same approach to show the existence of valid conic quadratic inequalities for hyperboloids and non-convex quadratic cones when the disjunction is defined by parallel hyperplanes. Also, for some of the cases, we show that for each of the branches of those sets, which are convex, the intersections with the DCCs or DCyCs still provides the convex hull of the intersection of the branches with a parallel linear disjunction.

Julio Goez
Norwegian School of Economics
Julio.Goez@nhh.no

Miguel F. Anjos
Mathematics and Industrial Engineering & GERAD
Polytechnique Montreal
anjos@stanfordalumni.org

MS25

Mixed-Integer Sum of Squares: Computation and Applications

Sum of squares optimization has seen a flurry of research and adoption in the control systems analysis over the past 15 years. Many of these problems involve discrete decisions, which can be naturally modeled using mixed-integer optimization. In this talk, we address computational aspects of mixed-integer sum of squares optimization, including both algorithmic and end-user modeling aspects. We apply these ideas to applications in optimal control and experimental design, with an emphasis on producing optimal (or provably sub-optimal) solutions for practical instances.

Joey Huchette
Operations Research Center
Massachusetts Institute of Technology
huchette@mit.edu

MS26

A Bundle Method for Transient Constrained Optimization Problems

The inclusion of dynamic constraints is the nominal objective of many optimization-based power system analyses. In current practice this is typically done off-line. We present an approach for the on-line inclusion of dynamic constraints in power grid optimization problems, and we apply it to transient security-constrained economic dispatch (TSCED). The approach is based on an encapsulation that allows for the separate simulation of transient system behavior, provided the simulation is equipped to return the sensitivity of constraint functions to the optimization problems. We define a bundle approach that incorporates the successively obtained sensitivity information to determine the solution of optimization problems with dynamic constraints. Numerical experiments are conducted on the IEEE 39-Bus System and show the benefits of the TSCED formulation.

Francois Gilbert, Shrirang Abhyankar, Hong Zhang
Argonne National Laboratory
francois.gilbert.fg@gmail.com, abhyshr@mcs.anl.gov, hongzhang@anl.gov

Mihai Anitescu
Argonne National Laboratory
Mathematics and Computer Science Division
anitescu@mcs.anl.gov

MS26

Towards Dual Control for Acute Myeloid Leukemia

Dual control is an extended feedback strategy compared to NMPC in which a sequence of optimal control problems with integrated OED are solved. For the parameter estimation OED is considered in the NMPC framework to obtain a control which provides well identifiable model parameters. The control strategy has two competitive tasks. Bringing the dynamic system into a desired state while simultaneously exciting the system such that the performed measurements provide minimal parameter uncertainty during estimation. Published dual control approaches focus on problem formulations with continuous measurement flow. We are interested in medical applications in which fast sampling rates are not necessary or available. In this setting the time points for taking measurements are of great importance and therefore should be considered during optimization. We present a new dual control algorithm in which the sampling decision is a further optimization variable due to a limited number of total measurements. The new algorithm is applied to a medical example. The mathematical model describes the dynamics of leukocytes of patients having acute myeloid leukemia. The model is used to investigate Leukopenia which is a clinically important side effect arising from the treatment with chemotherapy. The competitive tasks in this example are the reduction of intervals in which an infection due to Leukopenia is high and small excitations of the leukocytes' dynamics to identify the unknown parameters.

Felix Jost, Kristine Rinke
Otto-von-Guericke Universität Magdeburg
felix.jost@ovgu.de, kristine.rinke@ovgu.de

Sebastian Sager
Otto-von-Guericke Universität Magdeburg, Germany
sager@ovgu.de

Thuy T. T. Le
Otto-von-Guericke Universität Magdeburg
thuy.lethien@ovgu.de

MS26

Novel Computational Framework for Solving Complex Constrained Optimal Control Problems

A novel computational framework is described for solving complex constrained nonlinear optimal control problems. The framework has a wide variety of applications in mechanical and aerospace engineering. The basis of the framework is the new class of hp-adaptive Gaussian quadrature methods that transcribe the continuous optimal control problem to a finite-dimensional nonlinear optimization problem. The hp-adaptive methods have the feature that high accuracy can be obtained with a significantly smaller mesh when compared with traditional fixed-order methods while accurately capturing nonsmoothness or rapidly changing behavior. The hp-adaptive methods employed using advanced sparse nonlinear programming (NLP) solvers. The derivatives required by the NLP solvers are obtained using a new approach to algorithmic differentiation where efficient derivative source code is produced through a method that combines operator overloading with source transformation. The mathematical foundation of the framework is provided and examples are given that demonstrate the improvement over previously developed approaches.

Anil Rao
University of Florida
Mechanical and Aerospace Engineering
anilvrao@gmail.com

MS26

Architectures for Multi-Scale Predictive Control

We analyze the structure of the Euler-Lagrange conditions of a lifted long-horizon optimal control problem. The analysis reveals that the conditions can be solved by using block Gauss-Seidel schemes and we prove that such schemes can be implemented by solving sequences of short-horizon problems. The analysis also reveals that a receding-horizon control scheme is equivalent to performing a single Gauss-Seidel sweep. We also derive a strategy that uses adjoint information from a coarse long-horizon problem to correct the receding-horizon scheme and we observe that this strategy can be interpreted as a hierarchical control architecture in which a high-level controller transfers long-horizon information to a low-level, short-horizon controller. Our results bridge the gap between multigrid, hierarchical, and receding-horizon control.

Victor Zavala
University of Wisconsin-Madison
victor.zavala@wisc.edu

Yankai Cao
University of Wisconsin
cao27@wisc.edu

MS27

Model Formulations and Solution Approaches for the Hydropower Maintenance Scheduling Problem

Maintenance of generators is essential for the economic and reliable of operation of electrical power systems. As generators under maintenance are typically unavailable for electricity production, the generator maintenance scheduling problem (GMSP) should anticipate its impact on the system operation. Particularly in hydroelectric systems, the operation is affected by uncertain water inflows, nonlinearities in power production functions and interdependencies of the system variables in space and time. In this talk we present two MILP model formulations for the GMSP considering the specificities of the hydropower operation, and we discuss solution approaches for this problem.

Jesus A. Rodriguez Saras
Polytechnique Montreal
jesus.rodriguez@polymtl.ca

Miguel F. Anjos
Mathematics and Industrial Engineering & GERAD
Polytechnique Montreal
anjos@stanfordalumni.org

Guy Desaulniers
Mathematics and Industrial Engineering
Gerad & Polytechnique Montreal
guy.desaulniers@gerad.ca

Pascal Coté
Energie électrique
Rio Tinto Alcan & GERAD
pasca.cote@riotinto.com

MS27

Optimal Planning and Control of Large Scale Hydroelectric Systems: Information Flows and Modeling Tools

Operation of large-scale hydroelectric systems involves multiple decisions by different groups at different planning levels concerning when and how much water to release and power to generate at different generating units in a multi-reservoir system. These decisions are difficult because they must simultaneously consider technical, economic and reliability objectives as well as organizational and management boundaries. Several uncertainties must also be considered requiring the solution of stochastic problems in both supply and demand. The operation of such a complex system cannot be practically optimized by a single optimization model, no matter how sophisticated the model nor the computing power used as there are too many variables and unknowns and organizational levels to deal with. We present a Modeling and Information Flow Hierarchy we have developed by closely working with operations planning engineers at BC Hydro. The framework is divided into several computationally and operationally manageable levels; each provides information flows and feedbacks to the appropriate organizational and planning level, consisting of common measures such as resource prices, operational target levels, and constraints. This information is derived using stochastic and deterministic optimization and simulation algorithms, is passed from one level to the other to address a different aspect or planning horizon of the system ranging from long term operations planning to real-time operations.

Ziad Shawwash
UBC
shawwash@mail.ubc.ca

MS27

Stochastic Short-Term Hydropower Operations Planning

Short-term hydropower optimization is used on a daily basis to plan water releases, reservoir volumes and turbines in operation. This talk presents an optimization approach to solve the short-term hydropower problem with uncertain inflows. Scenario trees are built based on inflows forecasts and a multistage stochastic program is used to solve the problem. The scenario tree algorithm is based on the minimization of the nested distance and the multistage stochastic program is an innovative two-phase optimization process. Numerical results are presented for three hydropower plants located in the province of Quebec, Canada.

Sara Séguin

Université Laval
sara.seguin.1@ulaval.ca

Stein-Erik Fleten
Norwegian University of Science and Technology
stein-erik.fleten@iot.ntnu.no

Alois Pichler
NTNU and University of Vienna
aloisp@ntnu.no

Pascal Côté
Rio Tinto
pascal.cote@riotinto.com

Charles Audet
École Polytechnique de Montréal - GERAD
Charles.Audet@gerad.ca

MS27

Stochastic Co-Optimization of Transmission and Distribution-as-Microgrid Networks with Renewables and Energy Storage

The increase in renewable resource penetration in bulk power systems introduces challenges in operations, due to the inability to confidently predict resource availability. The result of inaccurate forecasting is a reduction in reliability, with volatility in prices, line flows and voltages on the power transmission network. However, the ability of microgrids to be operated in both interconnected and islanded modes shows potential to leverage these smaller systems to manage disturbances in the bulk power network, contributing to the integration of renewables in the power system. Therefore, more sophisticated tools are needed to operate the network efficiently. In this talk, we will present our in-progress work to co-optimize the interaction between the main network and microgrids, in the presence of high renewable penetration and distributed storage.

Luckny Zéphyr, Jialin Lui
Cornell University
lz395@cornell.edu, jl3455@cornell.edu

C. Lindsay Anderson
c lindsay anderson
landerson@cornell.edu

MS28

Incremental Algorithms for Additive Convex Cost Optimization and Learning

Motivated by machine learning problems over large data sets and distributed optimization over networks, we consider the problem of minimizing the sum of a large number of convex component functions. We start with an overview of this session by giving an introduction to incremental gradient methods and decentralized estimation techniques for solving such problems, which process component functions sequentially in time. We first consider deterministic cyclic incremental gradient methods (that process the component functions in a cycle, one at a time) and provide new convergence rate results under some assumptions. We then consider a randomized incremental gradient method, called the random reshuffling (RR) algorithm, which picks a uniformly random order/permutation and processes the component functions one at a time according to this order (i.e., samples functions without replacement in each cycle). We provide the first convergence rate guarantees for this method that outperform its popular with-replacement counterpart stochastic gradient descent (SGD). We finally consider incremental aggregated gradient methods, which compute a single component function gradient at each iteration while using outdated gradients of all component functions to approximate the global cost function gradient, and provide new linear rate results. This is joint work with Asu Ozdaglar and Pablo Parrilo.

Mert Gurbuzbalaban
Massachusetts Institute of Technology
mert.gurbuzbalaban@gmail.com

MS28

Distributed Learning: Parameter Estimation for the Exponential Family

In recent years, several distributed non-Bayesian algorithms have been proposed for the problem of distributed learning. Stability and explicit non-asymptotic convergence rates have be been shown for general estimation problems. Such convergence rates have been recently extended to continuous set of parameter or hypothesis. This allows the definition of an optimization based family of distributed algorithms for typical parameter estimation problems. Specifically we will describe how such non-Bayesian methods have "nice" close form explicit algorithmic representation in terms of the natural parameters for distributions in the exponential family.

Cesar A. Uribe
University of Illinois at Urbana-Champaign
cauribe2@illinois.edu

Angelia Nedich
Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign
angelia@illinois.edu

MS28

Global Convergence Rate of Proximal Incremental Aggregated Gradient Methods

We focus on the problem of minimizing the sum of smooth component functions (where the sum is strongly convex) and a non-smooth convex function, which arises in regularized empirical risk minimization in machine learning and distributed constrained optimization in wireless sensor networks and smart grids. We consider solving this problem using the proximal incremental aggregated gradient (PIAG) method, which at each iteration moves along an aggregated gradient (formed by incrementally updating gradients of component functions according to a deterministic order) and taking a proximal step with respect to the non-smooth function. While the convergence properties of this method with randomized orders (in updating gradients of component functions) have been investigated, this paper, to the best of our knowledge, is the first study that establishes the convergence rate properties of the PIAG method for any deterministic order. In particular, we show that the PIAG algorithm is globally convergent with a linear rate provided that the step size is sufficiently small. We explicitly identify the rate of convergence and the corresponding step size to achieve this convergence rate. Our results improve upon the best known condition number dependence of the convergence rate of the incremental aggregated gradient methods used for minimizing a sum of

smooth functions.

Nuri D. Vanli
MIT
denizcan@mit.edu

Mert Gurbuzbalaban
Massachusetts Institute of Technology
mert.gurbuzbalaban@gmail.com

Asu Ozdaglar
MIT
asuman@mit.edu

MS28

An Asynchronous Distributed Newton Method

We propose an asynchronous distributed Newton based method for minimizing a summation of convex objective functions. We show that the proposed method achieves at least linear rate of convergence and numerical studies shows great performance over existing methods.

Ermin Wei
Northwestern University
ermin.wei@northwestern.edu

Fatemeh Mansoori
northwestern university
mansoori.ftmh@gmail.com

MS29

A Numerical Investigation of Sketching and Subsampling

The concepts of subsampling and sketching have recently received much attention by the optimization and statistics communities. In this talk, we focus on their numerical performance, with the goal of providing new insights into their theoretical and computational properties. We pay particular attention to subsampled Newton and Newton-Sketch methods, as well as techniques for solving the Newton equations approximately. Our tests are performed on a collection of optimization problems arising in machine learning.

Albert Berahas
Department of Engineering Sciences and Applied Mathematics
Northwestern University
albertberahas@u.northwestern.com

Raghu Bollapragada
Department of Industrial Engineering and Management Sciences
Northwestern University
vijayabollapragada2014@u.northwestern.edu

Jorge Nocedal
Department of Electrical and Computer Engineering
Northwestern University
j-nocedal@northwestern.edu

MS29

Gradient Descent Efficiently Solves Cubic-Regularized Non-Convex Quadratic Problems

We consider the minimization of non-convex quadratic forms regularized by a cubic term, which exhibit multiple saddle points and poor local minima. Nonetheless, we prove that, under mild assumptions, gradient descent approximates the *global minimum* to within ε accuracy in $O(\varepsilon^{-1}\log(1/\varepsilon))$ steps for large ε and $O(\log(1/\varepsilon))$ steps for small ε (compared to a condition number we define), with at most logarithmic dependence on the problem dimension. We show how to use our results to develop a Hessian-free algorithm that finds second-order stationary points of general smooth non-convex functions. We complement our theoretical findings with experimental evidence corroborating our bounds, suggesting they are tight up to subpolynomial factors.

Yair Carmon
Statistics and Electrical Engineering
Stanford University
yairc@stanford.edu

John C. Duchi
Stanford University
Departments of Statistics and Electrical Engineering
jduchi@stanford.edu

MS29

Block BFGS Methods

We introduce a quasi-Newton method with block updates called Block BFGS. We show that this method, performed with inexact Armijo-Wolfe line searches, converges globally and superlinearly under the same convexity assumptions as BFGS. We also show that Block BFGS is globally convergent to a stationary point when applied to non-convex functions with bounded Hessian, and discuss other modifications for non-convex minimization. Numerical experiments comparing Block BFGS, BFGS and gradient descent are presented, as well as applications to machine learning.

Wenbo Gao, Donald Goldfarb
Columbia University
wg2279@columbia.edu, goldfarb@columbia.edu

MS29

Sub-Sampled Newton Methods and Large-Scale Machine Learning

A major challenge for large-scale machine learning, and one that will only increase in importance as we develop models that are more and more domain-informed, involves going beyond high-variance first-order optimization methods. Here, we consider the problem of minimizing the sum of a large number of functions over a convex constraint set, a problem that arises in many data analysis and machine learning applications, as well as many more traditional scientific applications. For this class of problems, we establish improved bounds for algorithms that incorporate sub-sampling as a way to improve computational efficiency, while maintaining their original convergence properties. These methods exploit recent results from Randomized Linear Algebra on approximate matrix multiplication. Within the context of second order methods, they provide quantitative convergence results for variants of Newton's methods, where the Hessian and/or the gradient is uniformly or non-uniformly sub-sampled, under much weaker assumptions than prior work.

Fred Roosta, Michael Mahoney
UC Berkeley

farbod@stat.berkeley.edu, mmahoney@stat.berkeley.edu

MS30

Matrix Completion, Saddlepoints, and Gradient Descent

Matrix completion is a fundamental machine learning problem with wide applications in collaborative filtering and recommender systems. Typically, matrix completion are solved by non-convex optimization procedures, which are empirically extremely successful. We prove that the symmetric matrix completion problem has no spurious local minima, meaning all local minima are also global. Thus the matrix completion objective has only saddlepoints an global minima. Next, we show that saddlepoints are easy to avoid for even Gradient Descent – arguably the simplest optimization procedure. We prove that with probability 1, randomly initialized Gradient Descent converges to a local minimizer. The same result holds for a large class of optimization algorithms including proximal point, mirror descent, and coordinate descent.

Jason Lee
University of Southern California
jasonlee@marshall.usc.edu

MS30

Regularized Hpe-Type Methods for Solving Monotone Inclusions with Improved Pointwise Iteration-Complexity Bounds

We discuss the iteration-complexity of new regularized hybrid proximal extragradient (HPE)-type methods for solving monotone inclusion problems (MIPs). The new (regularized HPE-type) methods essentially consist of instances of the standard HPE method applied toregularizations of the original MIP. It is shown that its pointwise iteration-complexity considerably improves the one of the HPE method while approaches (up to a logarithmic factor) the ergodic iteration-complexity of the latter method.

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

MS30

Stochastic Reformulations of Linear Systems and Fast Stochastic Iterative Method

We propose a new paradigm for solving linear systems. In our paradigm, the system is reformulated into a stochastic problem, and then solved with a randomized algorithm. Our reformulation can be equivalently seen as a stochastic optimization problem, stochastically preconditioned linear system, stochastic fixed point problem and as a probabilistic intersection problem. We propose and analyze basic and accelerated stochastic algorithms for solving the reformulated problem, with linear convergence rates.

Peter Richtarik
University of Edinburgh
peter.richtarik@ed.ac.uk

MS30

Serial and Parallel Coordinate Updates for Splitting Algorithms

Splitting algorithms are widely used in computational mathematics because they individually handle smaller and simpler parts of a problem that as a whole is more complicated. Many splitting algorithms can be abstracted by the fixed-point iteration $x^{k+1} = T(x^k)$, where T consists of a sequence of operations. This talk discusses the prevailing coordinate-friendly structure in T, with which one can update one, or a block of, components of x (while fixing the others) at a much lower cost than computing $T(x^k)$. Based on this structure, we develop coordinate-update splitting algorithms for problems that current coordinate descent algorithms do not apply, such as certain conic programs, portfolio optimization problems, as well as certain problems with nonsmooth nonseparable functions. Theoretically, we show that, when T has a fixed point and is nonexpansive, it coordinate-update algorithm converges to a fixed point (which is a solution to the original problem) under appropriate step sizes. This result applies to both cyclic, random, and (async-)parallel selections of coordinates.

Wotao Yin
University of California, Los Angeles
wotaoyin@math.ucla.edu

MS31

Stochastic Optimization for Optimal Transport

Abstract not available at time of publication.

Marco Cuturi
ENSAE / CREST
marcocuturicameto@gmail.com

MS31

Adversarial Embedding for Compositional Stochastic Optimization and Reinforcement Learning

While reinforcement learning received much attention recent years to address Markov decision problems, most existing reinforcement learning algorithms are either purely heuristic or lack of efficiency. A key challenge boils down to the difficulty with learning from conditional distributions with limited samples and minimizing objectives involving nested expectations. We propose a novel stochastic-approximation-based method that benefits from a fresh employ of Fenchel duality and kernel embedding techniques. To the best of our knowledge, this is the first algorithm that allows to take only one sample at a time from the conditional distribution and comes with provable theoretical guarantee. Finally, the proposed algorithm achieves the state-of-the-art empirical performances on several benchmark datasets comparing to the existing algorithms such as gradient-TD2 and residual gradient.

Niao He
University of Illinois at Urbana-Champaign
niaohe@illinois.edu

MS31

Accelerating Stochastic Gradient Descent

There is widespread sentiment that it is not possible to effectively utilize fast gradient methods (e.g. Nesterov's acceleration, conjugate gradient, BFGS) for the purposes of stochastic optimization due to their instability and error accumulation, a notion made precise in dAspremont

'08, Devolder, Glineur, & Nesterov 14]. This work strongly refutes this folklore by showing that acceleration, in a precise sense, is robust to statistical errors. In particular, this work provides an accelerated stochastic gradient descent algorithm that enjoys minimax optimal statistical estimation at a rate that is provably faster than stochastic gradient descent. The analysis involves introducing a novel technical lens of viewing accelerated stochastic gradient descent updates as a stochastic process, which may be of independent interest.

Sham Kakade
University of Washington
sham@cs.washington.edu

MS31

Restarting Accelerated Gradient Methods with a Rough Strong Convexity Estimate

We propose new restarting strategies for accelerated gradient and accelerated coordinate descent methods. Our main contribution is to show that the restarted method has a geometric rate of convergence for any restarting frequency, and so it allows us to take profit of restarting even when we do not know the strong convexity coefficient. The scheme can be combined with adaptive restarting, leading to the first provable convergence for adaptive restarting schemes with accelerated gradient methods. Finally, we illustrate the properties of the algorithm on a regularized logistic regression problem and on a Lasso problem.

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

Olivier Fercoq
Telecom-ParisTech, CNRS LTCI
Université Paris-Saclay,75013, Paris, France
olivier.fercoq@telecom-paristech.fr

MS32

Hybrid Optimal Control Problems for Partial Differential Equations

We provide optimality conditions and numerical results for hybrid optimal control problems governed by partial differential equations. These problems involve switching times to be optimized, at which the dynamics, the integral cost, and the bound constraints on the control can vary. The developed method to solve these problems is based on first and second-order optimality conditions. It is applied to different type of partial differential equations, such as semilinear parabolic equations, coupled systems or hyperbolic equations. Numerical illustrations are presented, whose qualitative aspects underlie the original motivations.

Sebastien Court
University of Graz
Institut for Mathematics and Scientific Computing
sebastien.court@uni-graz.at

Karl Kunisch
Karl-Franzens University Graz
Institute of Mathematics and Scientific Computing
karl.kunisch@uni-graz.at

Laurent Pfeiffer
University of Graz
laurent.pfeiffer@uni-graz.at

MS32

Optimal Control of PDEs in a Complex Space Setting; Application to the Schrodinger Equation

In this talk we discuss optimality conditions for abstract optimization problems over complex spaces. We then apply these results to optimal control problems with a semigroup structure. As an application we detail the case when the state equation is the Schrodinger one, with pointwise constraints on the bilinear control. We derive first and second order optimality conditions and address in particular the case that the control enters the state equation and cost function linearly.

M. Soledad Aronna
EMAp/FGV
Rio de Janeiro
soledad.aronna@fgv.br

Frédéric Bonnans
Inria-Saclay and CMAP, Ecole Polytechnique
frederic.bonnans@inria.fr

Axel Kroener
INRIA Saclay and CMAP, Ecole Polytechnique
axel.kroener@inria.fr

MS32

Binary Optimal Control of Semi Linear PDEs

We consider a specific combinatorial optimal control problem governed by a semilinear elliptic PDE. The control variable is given by a vector of integers. In addition, the optimization is subject to pointwise state constraints and possibly additional control constraints of arbitrary combinatorial structure. The objective is linear in the control variable. The particular structure of the optimal control problem under consideration allows to design an efficient outer approximation algorithm based on the decomposition of the optimal control problem into an integer linear programming (ILP) master problem and a subproblem for calculating linear cutting planes. These cutting planes rely on the pointwise concavity of the PDE solution operator in terms of the control variables, which we prove in the case of PDEs with a non-decreasing convex nonlinear part. The decomposition allows us to use standard solution techniques for ILPs as well as for PDEs. We further benefit from reoptimization strategies due to the iterative structure of the algorithm. Experimental results show the efficiency of our approach.

Christoph Buchheim, Christian Meyer
TU Dortmund
Faculty of Mathematics
christoph.buchheim@math.tu-dortmund.de,
cmeyer@mathematik.tu-dortmund.de

MS32

Optimal Control of a Regularized Phase Field Fracture Propagation Model

We consider an optimal control problem governed by a phase-field fracture model. One challenge of this model problem is a non-differentiable irreversibility condition on the fracture growth, which we relax using a penalization approach. We then discuss existence of a solution to the penalized fracture model, existence of at least one solution for the regularized optimal control problem, as well as first order optimality conditions.

Ira Neitzel
Universität Bonn
Institut für Numerische Simulation
neitzel@ins.uni-bonn.de

Thomas Wick
Ecole Polytechnique
thomas.wick@polytechnique.edu

Winnifried Wollner
TU Darmstadt
wollner@mathematik.tu-darmstadt.de

MS33

Parallel Algorithms for Robust PCA Using Marginalization

Robust PCA and its variants are popular models for separating data into a low-rank model while accounting for outliers. While there have been a number of algorithms proposed to solve the model, these algorithms do not perform well in a massively parallel environment such as the GPU due to high-communication costs of critical steps, e.g., SVD or QR computations. We show how marginalizing over one of the variables allows us to use Burer-Monteiro splitting techniques and apply many of the techniques from low-rank recovery literature. This approach works well on the GPU and the resulting algorithm is significantly faster than the prior state-of-the-art.

Stephen Becker
University of Colorado at Boulder
stephen.becker@colorado.edu

MS33

Gauge and Perspective Duality

Fenchel-Young duality is widely used in convex optimization and relies on the conjugacy operation for convex functions; however, alternative notions of duality relying on parallel operations exist as well. In particular, gauge and perspective duality are defined via the polarity operation on nonnegative functions. We present a perturbation argument for deriving gauge duality, thus placing it on equal footing with Fenchel-Young duality. This approach also yields explicit optimality conditions (analogous to KKT conditions), and a simple primal-from-dual recovery method based on existing algorithms. Numerical results confirm the usefulness of this approach in certain contexts (e.g. optimization over PLQ functions).

Kellie MacPhee, Aleksandr Aravkin
University of Washington
kmacphee@uw.edu, sasha.aravkin@gmail.com

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

Dmitriy Drusvyatskiy
University of Washington
University of Washington
ddrusv@uw.edu

Michael P. Friedlander
Dept of Computer Science
University of British Columbia
mpf@cs.ubc.ca

MS33

A Novel Methods for Saddle Point Problems: Primal-Dual and Variational Inequality Approaches

In this paper we propose a new method for monotone variational inequalities. We prove its convergence, establish its optimal ergodic rate of convergence, and under some additional assumptions obtain R-linear rate of convergence. We also introduce a line search which allows to use larger steps. Surprisingly, the same idea can be used to derive a new primal-dual method. Some preliminary numerical experiments are reported.

Yura Malitsky
University of Kyiv
y.malitsky@gmail.com

MS33

Sketchy Decisions: Convex Low-Rank Matrix Optimization with Optimal Storage

In this talk, we consider a fundamental class of convex matrix optimization problems with low-rank solutions. We show it is possible to solve these problem using far less memory than the natural size of the decision variable when the problem data has a concise representation. Our proposed method, SketchyCGM, is the first algorithm to offer provable convergence to an optimal point with an optimal memory footprint. SketchyCGM modifies a standard convex optimization method the conditional gradient method to work on a sketched version of the decision variable, and can recover the solution from this sketch. In contrast to recent work on non-convex methods for this problem class, SketchyCGM is a convex method; and our convergence guarantees do not rely on statistical assumptions.

Madeleine R. Udell
Stanford University
Stanford University
madeleine.udell@gmail.com

MS34

Multigrid Preconditioning for Space-Time Distributed Optimal Control of Parabolic Equations

This work is concerned with designing optimal order multigrid preconditioners for space-time distributed control of parabolic equations. The focus is on the reduced problem resulted from eliminating state and adjoint variables from the KKT system. Earlier numerical experiments have shown that our ability to design optimal order preconditioners depends strongly on the discretization of the parabolic equation, with several natural discretizations leading to suboptimal preconditioners. Using a continuous-in-space-discontinuous-in-time Galerkin discretization we obtain the desired optimality.

Andrei Draganescu
Department of Mathematics and Statistics
University of Maryland, Baltimore County

draga@umbc.edu

MS34

Flexible Decision Variables in Reservoir Operation Using Dimension Reduction Approach

The optimization of reservoir operation involves various competing objectives and it is essential to consider multiple objectives simultaneously. There are also various sources of uncertainty that should be considered as well. We consider the question of unquantifiable uncertainty in the context of providing a decision maker with the largest possible set of options which are feasible and potentially optimal. We model a range of options with random variables, and then attempt to maximize the variance in order to provide the largest flexibility possible. This becomes a multi-objective, dynamic PDE-constrained optimization problem for an optimal probability distribution of controls. We solve for this flexible control distribution by first identifying a suitable reduced basis via KL expansion, and then finding the set of Pareto optimal joint probability distributions for the KL random variables.

Nathan Gibson
Oregon State University
gibsonn@oregonstate.edu

Parnian Hosseini
Oregon State University
Civil Engineering
hosseinp@oregonstate.edu

Arturo Leon
University of Houston
Department of Civil and Environmental Engineering
aleon3@central.uh.edu

MS34

Recent Advances in Newton-Krylov Methods for NMPC

We present advances in Newton-Krylov (AKA continuation) methods for NMPC, pioneered by Prof. Ohtsuka. Our main results concern efficient preconditioning, leading to dramatically improved real-time Newton-Krylov MPC optimization. One idea is solving a forward recursion for the state and a backward recursion for the costate approximately, or reusing previously computed solutions, for the purpose of preconditioning. Another ingredient is discovery of sparse factorizations of high-quality preconditioners. Altogether, our fast and efficient preconditioning leads to optimal linear complexity in the number of gridpoints on the prediction horizon, for iterative solution of forward-difference Newton-Krylov NMPC. We suggest scenario/particle Newton-Krylov MPC in the case, where system dynamics or constraints discretely change on-line, for ensembles of predictions corresponding to various scenarios of anticipated changes and test it for minimum time problems. Simultaneous on-line computation of ensembles of controls, for several scenarios of changing in real time system dynamics/constraints, allows choosing and adapting the optimal destination. The Newton-Krylov MPC approach is extended to the case, where the state is implicitly constrained to a smooth manifold, and demonstrate its effectiveness for a problem on a hemisphere.

Andrew Knyazev
Mitsubishi Electric Research Laboratories (MERL)
knyazev@merl.com

Alexander Malyshev
Univ. of Bergen, Norway
alexander.malyshev@uib.no

MS34

Parameter Estimation for Nonlinear Hyperbolic PDE Using Transient Flow Observations

We introduce a method for parameter estimation for nonlinear hyperbolic partial differential equations (PDEs) on a metric graph that represent compressible fluid flow in pipelines subject to time-dependent boundary conditions and nodal controls. Flow dynamics are described on each edge by space- and time-dependent density and mass flux that evolve according to a system of coupled PDEs; momentum dissipation is modeled using the Darcy-Wiesbach friction approximation. Mass flow balance is maintained by applying the Kirchhoff-Neumann boundary conditions at network nodes, at which fluid may be injected or withdrawn from the network. Flow through the system may be actuated by augmenting density at the interface between nodes and adjoining pipes. We focus here on estimation of the Darcy-Wiesbach friction factor using partial observations of time-series for pressures and flows throughout the network. A rapid, scalable computational method for performing a non-linear least squares fit is developed.

Kaarthik Sundar
Los Alamos National Laboratory
kaarthik01sundar@gmail.com

MS35

Structural Properties of Affine Sparsity Constraints

We introduce a new constraint system for sparse variable selection in statistical learning. Such a system arises when there are logical conditions on the sparsity of certain unknown model parameters that need to be incorporated into their selection process. Formally, extending a cardinality constraint, an affine sparsity constraint (ASC) is defined by a linear inequality with two sets of variables: one set of continuous variables and the other set represented by their nonzero patterns. This paper aims to study an ASC system consisting of finitely many affine sparsity constraints. We investigate a number of fundamental structural properties of the solution set of such a non-standard system of inequalities, including its closedness and the description of its closure, continuous approximations and their set convergence, and characterizations of its tangent cones for use in optimization. Our study lays a solid mathematical foundation for solving optimization problems involving these affine sparsity constraints through their continuous approximations.

Jong-Shi Pang
University of Southern California
jongship@usc.edu

Hongbo Dong
Assistant Professor
Washington State University
hongbo.dong@wsu.edu

Miju Ahu
University of Southern California

xxx@xxx.xxx

MS35

Distributed Big-Data Optimization via Batch Gradient Tracking

We study distributed nonconvex bigdata optimization in multiagent networks. Our interest is on big-data problems wherein there is a large number of variables to optimize. If treated by means of standard distributed optimization algorithms, these large-scale problems may be intractable due to the prohibitive local computation and communication burden at each node. We propose a novel distributed solution method whereby at each iteration the agents update in an uncoordinated fashion only a block of the entire decision variables. To deal with the nonconvexity of the cost function, the novel scheme hinges on Successive Convex Approximation (SCA) techniques coupled with i) a tracking mechanism instrumental to estimate locally gradient averages; and ii) a novel block-wise consensus-based protocol to perform local block-averaging operations and gradient tacking. Asymptotic convergence to stationary solutions of the nonconvex problem is established. This is a joint work with (listed in alphabetic order): I. Notarnicola, G. Notarstefano, and Y. Sun.

Gesualdo Scutari
Purdue University
gscutari@purdue.edu

MS35

Stochastic Convex Optimization: Faster Local Growth Implies Faster Global Convergence

In this talk, I will present a new theory for first-order stochastic convex optimization, showing that the global iteration complexity is sufficiently quantified by a local growth rate of the objective function in a neighborhood of the optimal solutions. In particular, if the objective function $F(w)$ in the ϵ-sublevel set grows as fast as $\|w - w_*\|_2^{1/\theta}$, where w_* represents the closest optimal solution to w and $\theta \in (0, 1]$ quantifies the local growth rate, the iteration complexity of first-order stochastic optimization for achieving an ϵ-optimal solution can be $\widetilde{O}(1/\epsilon^{2(1-\theta)})$, which is *optimal at most* up to a logarithmic factor. This result is fundamentally better in contrast with the previous works that either assume a global growth condition in the entire domain or achieve a local faster convergence under the local faster growth condition. I will also present several stochastic algorithms for achieving the aforementioned iteration complexity and consider applications in machine learning.

Tianbao Yang
University of Iowa
tianbao-yang@uiowa.edu

MS35

Pathwise Coordinate Optimization for Nonconvex Sparse Learning: Algorithm and Theory

Nonconvex optimization naturally arises in many machine learning problems. Machine learning researchers exploit various nonconvex formulations to gain modeling flexibility, estimation robustness, adaptivity, and computational scalability. Although classical computational complexity theory has shown that solving nonconvex optimization is generally NP-hard in the worst case, practitioners have proposed numerous heuristic optimization algorithms, which achieve outstanding empirical performance in real-world applications. To bridge this gap between practice and theory, we propose a new generation of model-based optimization algorithms and theory, which incorporate the statistical thinking into modern optimization. Particularly, when designing practical computational algorithms, we take the underlying statistical models into consideration. Our novel algorithms exploit hidden geometric structures behind many nonconvex optimization problems, and can obtain global optima with the desired statistics properties in polynomial time with high probability.

Tuo Zhao
Johns Hopkins University
tourzhao@gatech.edu

MS36

Optimal Control of Problems with Hysteresis

We discuss optimality conditions for control problems for ordinary and parabolic equations which include nonsmooth nonlinearities of hysteresis type. These are obtained either by smooth approximation or by proving and utilizing results on generalized (weak) differentiability of hysteresis nonlinearities.

Martin Brokate
Technische Universität München
brokate@ma.tum.de

MS36

An Approach to the Differential Stability Analysis of Mosolov's Problem

The aim of this talk is to study the differentiability properties of the solution operator to the so-called Mosolov problem. This is a simplified version of the classical Bingham flow problem that takes the form of an elliptic variational inequality of the second kind and models the steady-state motion of a viscous-plastic medium in a cylindrical pipe. We derive differentiability results under suitable structural assumptions and show that the differential sensitivity analysis of Mosolov's problem is closely linked to the study of certain weighted Sobolev spaces. Some of the results on the fine properties of Sobolev functions needed for our investigation are also of independent interest.

Constantin Christof
Technische Universität Dortmund
constantin.christof@math.tu-dortmund.de

Christian Meyer
TU Dortmund
Faculty of Mathematics
cmeyer@mathematik.tu-dortmund.de

MS36

Optimal Control of Time Discretized Contact Problems

Our presentation deals with the optimal control of problems governed by time discretized, dynamic contact problems in linear viscoelasticity with respect to distributed controls. We discuss some of the subtleties introduced into the contact model by the linear non-penetration condition as well as a suitable formulation of the set of admissible states in the optimal control of the resulting variational inequality. A number of integrators for the time continuous form of dynamic contact problems allow for the analysis of corresponding time discretized problems. We employ a temporal finite element discretization, which corresponds to a contact implicit Newmark scheme. In the discretized setting, we obtain existence of a solution operator and, consequently, existence of minimizers under standard assumptions on the cost functional. We address different notions of capacity on the boundary of a domain and analyze differentiability properties of the solution operator. When the solution operator to the time discretized contact problem is directionally differentiable and the controls are dense, we derive a set of first order optimality conditions. Solutions to a modified adjoint equation can then be computed by an inverse time stepping scheme and allow for efficient computation of search directions for the fully discretized system in a simple, steepest-descent type optimization scheme.

Georg Müller, Anton Schiela
University of Bayreuth
georg.mueller@uni-bayreuth.de, anton.schiela@uni-bayreuth.de

MS36

Optimal Control of Thermoviscoplastic Systems

Elastoplastic deformations play a tremendous role in industrial forming. Many of these processes happen at non-isothermal conditions. Therefore, the optimization of such problems is of interest not only mathematically but also for applications. In this talk we will present the analysis of an optimal control problem governed by a thermo-visco(elasto)plastic model. We will point out the difficulties arising from the nonlinear coupling of the heat equation with the mechanical part of the model. Finally, we will discuss first numerical results.

Ailyn Stötzner, Roland Herzog
Technische Universität Chemnitz
ailyn.stoetzner@mathematik.tu-chemnitz.de,
roland.herzog@mathematik.tu-chemnitz.de

Christian Meyer
TU Dortmund
Faculty of Mathematics
cmeyer@mathematik.tu-dortmund.de

MS37

Sums of Squares and Matrix Completion

I will review recent results on semidefinite matrix completion and its relation with sums of squares. I will discuss low-rank completion, positive definite completion and applications to Gaussian graphical models.

Greg Blekherman
Georgia Institute of Technology
greg@math.gatech.edu

Rainer Sinn
Georgia Tech
sinn@math.gatech.edu

Mauricio Velasco
Universidad de los Andes
mvelasco@uniandes.edu.co

Mauricio Velasco
University of los Andes
mvelasco@uniandes.edu.co

MS37

Polynomial Optimization with Sum-of-Squares Interpolants

One of the most common approaches to polynomial optimization utilizes semidefinite programming (SDP) hierarchies, which arise from combining the algebraic theory of sum-of-squares polynomials with the observation that SOS polynomials are semidefinite representable. While theoretically satisfactory, the translation of problems involving sum-of-squares polynomials to SDPs can be impractical for a number of reasons. First, the SDP representation of sum-of-squares polynomials roughly squares the number of optimization variables, increasing the time and memory complexity of the solution algorithms by several orders of magnitude. The second problem is numerical: in the most common SDP formulation of the hierarchies, the dual variables are semidefinite matrices whose condition numbers grow exponentially with the degree of the polynomials involved; this is detrimental for afloating-point implementation. In the talk, we show that using a combination of polynomial interpolation and non-symmetric cone optimization we can optimize directly over sum-of-squares cones in a very efficient way, avoiding the aforementioned pitfalls of the semidefinite programming approach to sum-of-squares optimization.

David Papp
North Carolina State University
dpapp@ncsu.edu

Sercan Yildiz
University of North Carolina at Chapel Hill
syildiz@samsi.info

MS37

Polynomials as Sums of Few Squares

The classes of polynomials for which every nonnegative polynomial is a sum of squares is an important subject dating back to Hilbert. I will discuss the number of squares needed in these representations and give a tight bound in terms of the number of variables. This generalizes the celebrated result by Hilbert that every real nonnegative ternary quartic is a sum of three squares. Applying this theory to polynomials called biforms gives low-rank factorizations of positive semidefinite bivariate matrix polynomials. For polynomials in two variables, we can also count the number of representations as a sum of few squares.

Greg Blekherman
Georgia Institute of Technology
greg@math.gatech.edu

Daniel Plaumann
TU Dortmund
daniel.plaumann@math.tu-dortmund.de

Rainer Sinn
Georgia Tech
sinn@math.gatech.edu

Cynthia Vinzant
University of Michigan
Dept. of Mathematics

clvinzan@ncsu.edu

MS38

Deep Learning Meets Partial Differential Equations and Optimization

Abstract not available at time of publication

Eldad Haber
University of British Columbia, Vancouver, Canada
ehaber@eos.ubc.ca

MS38

Data-Driven Model Reduction via CUR-Factored Hankel Approximation

In this presentation, we consider obtaining reduced-order models for optimization and control from impulse-response data of dynamical systems. Traditional system identification methods for this task require the the singular value decomposition of the large, dense, block-structured Hankel matrix. We show that we can use certain other, faster matrix decompositions of the Hankel matrix. In particular, we use a fast algorithm to compute the CUR-decomposition of the Hankel matrix, and present an algorithm that integrates the decomposition into a subspace-based system identification method. We show a speedup factor of 25 compared to SVD-based approaches, while maintaining the same magnitude of error. Error bounds and stability proofs of the obtained reduced-order models are also provided. We demonstrate the approach on a numerical test problem.

Boris Kramer, Karen E. Willcox
Massachusetts Institute of Technology
bokramer@mit.edu, kwillcox@MIT.EDU

Alex A. Gorodetsky
Sandia National Laboratories
goroda@mit.edu

MS38

Adaptive Multiscale Finite Element Methods for PDE Parameter Estimation

PDE parameter-estimation problems arise frequently in many applications such as geophysical and medical imaging. The problem can be formulated as an optimization problem with constraints given by the underlying PDEs. We consider the so-called reduced formulation in which the PDEs are eliminated. The resulting unconstrained optimization problem is computationally expensive to solve because as the PDEs need to be solved numerous times for different parameters until an adequate reconstruction is found. We consider the case in which many measurements are available leading to a large number of PDE constraints. We present an optimization scheme that uses operator-induced interpolation to solve PDEs on a coarser mesh and thereby reduces the computational burden. The projection is computed using multiscale finite element methods (MS-FEM) and - in our formulation - is adapted in the course of optimization. Our optimization approach includes the derivatives of the adaptive multiscale basis. We demonstrate the potential of the method using examples from geophysical imaging.

Lars Ruthotto
Department of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

Samy Wu Fung
Emory University
400 Dowman Dr, 30322 Atlanta, GA, USA
samy.wu@emory.edu

MS39

Latest Results on the Binary Approximation Method for Polynomial Optimization Problems

We present experimental results with a cutting-plane algorithm for polynomial optimization problems which uses an approximate representation of continuous variables as weighted sums of binary variables. Rather than using the representation directly, we instead use it to separate over the convex hull of feasible solutions. This allows us to control the degree to which variables are approximately represented (as weighted sums of binary variables) as well as the precision to which the approximations are carried out, while also reaping the benefits of relying on a mature (linear) mixed-integer solver. Our experiments show some promising results. Joint work with Chen Chen and Gonzalo Munoz.

Daniel Bienstock
Columbia University IEOR and APAM Departments
IEOR Department
dano@columbia.edu

MS39

Characterizations of Mixed Integer Convex Quadratic Representable Sets

Representability results play a fundamental role in optimization since they provide geometric characterizations of the feasible sets that arise from optimization problems. In this talk we study the sets that appear in the feasibility version of mixed integer convex quadratic optimization problems. We provide a number of characterizations of the sets that can be obtained as the projection of such feasible regions in a higher dimensional space. Most notably, we provide a complete characterization in the case that the feasible region is bounded. This work is joint work with Jeffrey Poskin.

Alberto Del Pia
University of Wisconsin
delpia@wisc.edu

MS39

Relaxations and Convex Hull Results for the Intersection of Elementary Nonconvex Quadratics and General Regular Cones

We examine the structure of nonconvex sets obtained from the intersection of a regular cone K and an elementary nonconvex quadratic where the underlying matrix has rank two. Such nonconvex quadratics are fundamentally related to two-term disjunctions on the regular cone K, and thus the resulting sets can be used as relaxations for a number of problems. We develop a family of structured convex inequalities for these nonconvex sets. Under mild assumptions, these inequalities together characterize the closed convex hull of such a set in the original space, and if additional conditions are satisfied, a single inequality from this family is sufficient. For the case where K is the direct prod-

uct of positive semidefinite cone, second-order cones and a nonnegative orthant, we show that these inequalities can be represented in appropriate tractable forms in certain cases. For more general cases, and when K is the positive semidefinite cone, we present conic relaxations admitting some desirable properties.

Sercan Yildiz
University of North Carolina at Chapel Hill
syildiz@samsi.info

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

MS39

Embedding Formulations For Unions Of Convex Sets

There is often a strong trade-off between formulation strenght and size in mixed integer programming (MIP). When modelling unions of convex sets this trade-off can be resolved by adding auxiliary continuous variables. However, adding these variables can result in a deterioration of the computational effectiveness of a formulation. For this reason there has been considerable interest in constructing strong formulations that do not use auxiliary variables. In this work we introduce a technique to construct such formulations that is based on a geometric object known as the Cayley embedding. We show how this technique can be easily used to construct new formulations and to detect redundant non-linear inequalities in existing formulations.

Juan Pablo Vielma
Sloan School of Management
Massachusetts Institute of Technology
jvielma@mit.edu

MS40

The Adaptive Gradient Method for Optimization of Functions Observed with Deterministic Error

We consider unconstrained optimization problems where the objective function is observed with deterministic error and assumed to decay with increasing oracle effort. An important special case is the *integral optimization* problem where the objective function is an integral that can be evaluated with a computationally expensive numerical quadrature oracle or through a Quasi Monte Carlo oracle. We propose the *adaptive gradient method* (AGM), which evolves by repeatedly taking a step along the gradient direction, estimated after expending a carefully specified (strategic) amount of oracle effort. We present AGM's work complexity bounds when implemented on three basic classes of functions characterized by increasing levels of function structure. Our results quantify the effect of the "quality" of the oracle in use on the resulting convergence rate of AGM.

Fatemeh Hashemi
Virginia Tech
fatemeh@vt.edu

Raghu Pasupathy
Purdue University
pasupath@purdue.edu

Mike Taaffe
Virginia Tech, ISE
taaffe@vt.edu

MS40

Synchronous and Asynchronous Inexact Proximal Best-Response Schemes for Stochastic Nash Games: Rate Analysis and Complexity Statements

This work considers a stochastic Nash game, where each player solves a parametrized stochastic optimization problem. In deterministic regimes, best response schemes are convergent under a suitable spectral property on the proximal-response map. However, a direct application of this scheme to stochastic settings requires obtaining exact solutions to stochastic optimization problems at every step. Instead, we propose an inexact generalization of this scheme in which an inexact solution is computed via a fixed (but increasing) number of projected stochastic gradient steps. On the basis of this framework, we make several contributions: First, we propose a synchronous inexact best-response scheme where all players simultaneously update their decisions. Then we propose an asynchronous scheme for the setting when players are not required to make simultaneous decisions nor need they have have access to their rivals' latest information. Surprisingly, we show that the iterates of both regimes converge to the unique equilibrium in mean at a linear rate rather than a sub-linear rate. Furthermore, we establish the overall iteration complexity of the schemes in terms of projected stochastic gradient steps for computing an $\epsilon-$Nash equilibrium. The schemes are further extended to settings where players solve two-stage linear and quadratic stochastic Nash games. Finally, we carry out some numerical simulations to demonstrate the rate and complexity statements.

Jinlong Lei
Penn. State University
jxl800@psu.edu

Uday Shanbhag
Department of Industrial and Manufacturing Engineering
Pennsylvania State University
udaybag@psu.edu

Jong-Shi Pang, Suvrajeet Sen
University of Southern California
jongship@usc.edu, s.sen@usc.edu

MS40

Distributed Optimization Over Networks: New Rate Results

We consider the problem of distributed optimization over time-varying undirected graphs. We discuss a distributed algorithm (DIGing) for solving this problem based on a combination of an inexact gradient method and a gradient tracking technique. This algorithm deploys fixed step size but converges exactly to the global and consensual minimizer. Under strong convexity assumption, we prove that the algorithm converges at an R-linear (geometric) convergence rate as long as the step size is less than a specific bound; we give an explicit estimate of this rate over uniformly connected graph sequences and show it scales polynomially with the number of nodes. Numerical experiments demonstrate the efficacy of the introduced algorithm and validate our theoretical findings.

Angelia Nedich
Arizona State University

angelia.nedich@asu.edu

Alexander Olshevsky
University of Illinois at Urbana-Champaign
aolshev2@illinois.edu

Wei Shi
Boston University
wilburs@bu.edu

MS40

A Variable Sample-Size Stochastic Approximation Framework

Traditional stochastic approximation (SA) schemes employ a single gradient or a fixed batch of noisy gradients in computing a new iterate and requires a projection on a possibly complex convex set for every update. As a consequence, practical implementations may require the solution of a large number of projection problems, and may render this scheme impractical. We present a variable sample-size stochastic approximation (VSSA) scheme where the batch of noisy gradients may change in size across iterations and the scheme terminates when the budget is consumed. In this setting, we derive error bounds in strongly convex, convex differentiable, and convex nonsmooth regimes. In particular, we focus on quantifying the rate of convergence in terms of projection steps and in terms of simulation budget and comment on the optimality of the obtained rates. In addition, we present an accelerated VSSA scheme which displays the optimal accelerated rate of convergence in terms of projection steps if the sample size grows sufficiently fast. Preliminary numerics suggest that VSSA schemes outperform their traditional counterparts in terms of computational time, often by several orders of magnitude, with little drop in empirical performance.

Uday Shanbhag
Department of Industrial and Manufacturing Engineering
Pennsylvania State University
udaybag@psu.edu

Afrooz Jalilzadeh
Penn. State University
afrooz.jalilzadeh1992@gmail.com

MS41

An Augmented Lagrangian Filter Method for Real-Time Embedded Optimization

We present a filter line-search algorithm for non-convex continuous optimization that combines an augmented Lagrangian function and a constraint violation metric to accept and reject steps. The approach is motivated by real-time optimization applications that need to be executed on embedded computing platforms. The proposed method enables primal-dual regularization of the linear algebra system that in turn permits the use of solution strategies with lower computing overheads. We prove that the proposed the algorithm is globally convergent and we demonstrate the developments using a nonlinear real-time optimization application. Our numerical tests are performed on a standard desktop and on an embedded platform, and demonstrate that the proposed approach enables reductions in solution times of up to three orders of magnitude compared to an standard filter implementation.

Nai-Yuan Chiang
United Technologies Research Center
chiangn@utrc.utc.com

Rui Huang
United Technologies
huangr@utrc.utc.com

Victor Zavala
University of Wisconsin-Madison
victor.zavala@wisc.edu

MS41

A Parallelizable Augmented Lagrangian Method Applied to Large-Scale Non-Convex-Constrained Optimization Problems

We contribute improvements to a Lagrangian dual solution approach applied to large-scale optimization problems whose objective functions are convex, continuously differentiable and possibly nonlinear, while the non-relaxed constraint set is compact but not necessarily convex. Such problems arise, for example, in the split-variable deterministic reformulation of stochastic mixed-integer optimization problems. The dual solution approach needs to address the nonconvexity of the non-relaxed constraint set while being efficiently implementable in parallel. We adapt the augmented Lagrangian method framework to address the presence of nonconvexity in the non-relaxed constraint set and the need for efficient parallelization. The development of our approach is most naturally compared with the development of proximal bundle methods and especially with their use of serious step conditions. However, deviations from these developments allow for an improvement in efficiency with which parallelization can be utilized. Pivotal in our modification to the augmented Lagrangian method is the use of an integration of approaches based on the simplicial decomposition method (SDM) and the nonlinear block Gauss-Seidel (GS) method. Under mild conditions optimal dual convergence is proven, and we report computational results on test instances from the stochastic optimization literature. We demonstrate improvement in parallel speedup over a baseline parallel approach.

Brian C. Dandurand
Argonne National Laboratory
brian.dandurand@rmit.edu.au

Natashia Boland
Georgia Institute of Technology
natashia.boland@gmail.com

Jeffrey Christiansen
Mathematical Sciences,
School of Science, RMIT University
s3507717@student.rmit.edu.au

Andrew Eberhard
RMIT University
andy.eberhard@rmit.edu.au

Fabricio Oliveira
Mathematical Sciences,
School of Science, RMIT University
fabricio.oliveira@rmit.edu.au

MS41

An Efficient Dual Newton Strategy for Tree-Sparse

Quadratic Programs

Model Predictive Control (MPC) is an advanced control strategy that uses the model of the controlled plant to optimize its future behaviour, subject to constraints. However, when uncertainties are present in the employed model, the performance of the controller may deteriorate significantly. One common approach to address this problem is scenario-tree MPC, where uncertainties are handled via a finite number of realizations at each decision point. A drawback of this formulation is that the size of the underlying optimization problem grows exponentially in the horizon length and thus structure-exploiting solvers are essential to achieve acceptable computation times. In this talk, we propose to solve the tree-structured Quadratic Programs (QPs) arising in scenario-tree MPC with a dual Newton strategy, where the dynamic constraints are relaxed via Lagrange duality and the dual problem is solved with a semi-smooth Newton method. We exploit the underlying structure to factorize the dual Hessian in an efficient and parallelizable manner. We conclude the talk with a comparison of our method to other dual Newton strategy variants that differ in the choice of constraints that are dualized.

Dimitris Kouzoupis
University of Freiburg
dimitris.kouzoupis@imtek.uni-freiburg.de

Emil Klintberg
Chalmers University of Technology
kemil@chalmers.se

Gianluca Frison
Technical University of Denmark
giaf@dtu.dk

Sébastien Gros
Chalmers University of Technology
grosse@chalmers.se

Moritz Diehl
University of Freiburg
moritz.diehl@imtek.uni-freiburg.de

MS41

Scenario Decomposition for Stochastic 0-1 Optimization

We present a scenario decomposition algorithm for stochastic 0-1 programs. The algorithm recovers an optimal solution by iteratively exploring and cutting-off candidate solutions obtained from solving scenario subproblems. The scheme is applicable to quite general problem structures and can be implemented in a distributed framework. We provide a theoretical justification of the effectiveness of the proposed scheme.

Shabbir Ahmed
Georgia Institute of Technology, Industrial and
Systems Engineering
shabbir.ahmed@isye.gatech.edu

Deepak Rajan
Lawrence Livermore National Laboratory
rajan3@llnl.gov

MS42

Sketch and Project: A Tool for Designing Stochastic Quasi-Newton Methods and Stochastic Variance Reduced Gradient Methods

I will present a stochastic sketching and projecting technique, akin to the least change formulation of quasi-Newton methods, for designing new stochastic quasi-Newton methods. This includes a new and recently introduced stochastic block BFGS method. With the same sketching and projecting technique, I will present a unifying viewpoint of variance reduced stochastic gradient methods. Not only is this viewpoint expressive, but it also leads to simpler and stronger proofs of convergence.

Robert M. Gower, Peter Richtarik
University of Edinburgh
gower.robert@gmail.com, peter.richtarik@ed.ac.uk

Francis Bach
INRIA - École Normale Supérieure
francis.bach@inria.fr

MS42

Introduction to Riemannian BFGS Methods

Riemannian optimization is the task of finding an optimum of a real-valued function defined on a Riemannian manifold. Riemannian optimization has been a topic of much interest over the past few years due to many applications including computer vision, signal processing, and numerical linear algebra. So far, many effective and efficient optimization methods on Riemannian manifolds have been proposed and analyzed. In this presentation, we choose Riemannian BFGS methods as examples to briefly review the framework of Riemannian optimization and emphasize its differences from Euclidean optimization. A numerical experiment is used to show the performance of Riemannian BFGS methods.

Wen Huang
Rice University
wen.huang@rice.edu

MS42

GRANSO: An Open-Source Software Package Employing a BFGS-SQP Method for Constrained, Nonsmooth, Nonconvex Optimization

We present a new open-source optimization package implemented in Matlab intended to be an efficient solver for nonsmooth constrained optimization problems, without any special structure or assumptions imposed on the objective or constraint functions. Though nonsmooth optimization is a well-studied area, there is still little available software specifically designed to handle general nonsmooth constraint functions. In 2012, Curtis and Overton proposed a relevant gradient-sampling-based SQP method, proving convergence results under the assumption that the objective and constraint functions are locally Lipschitz, but if these functions are expensive to compute, its code can be quite slow due to the gradient sampling. In contrast, our new software, called GRANSO: GRadient-based Algorithm for Non-Smooth Optimization, is designed to use few function/gradient evaluations per iteration and can handle problems involving functions that are any or all of the following: smooth or nonsmooth, convex or nonconvex, and locally Lipschitz or non-locally Lipschitz. GRANSO only requires that gradients be supplied, even for nonsmooth problems. Although GRANSO's underlying BFGS-SQP method provides no convergence guarantees for nonsmooth or non-locally-Lipschitz problems, using a challenging test set of 200 nonsmooth, nonconvex, constrained optimization problems arising in controller design, we have observed GRANSO to be both reliable and significantly more efficient than available alternatives.

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

Tim Mitchell
Max Planck Institute for Dynamics of Complex Technical Syste
mitchell@mpi-magdeburg.mpg.de

Michael L. Overton
New York University
Courant Instit. of Math. Sci.
overton@cs.nyu.edu

MS42

Parallel quasi-Newton methods for PDE-constrained optimization

In this talk we will present a parallel quasi-Newton interior method for constrained optimization problems. The work is motivated by large-scale PDE-constrained topology optimization problems that occurs in optimal material design. Our parallel method exploits the fact that the interior-point linear system consists of thin dense blocks that border a low-rank update of a diagonal matrix. This structure stems from the use of reduced-space PDE-constrained optimization formulations and the use of secant Hessian approximations. The remainder of the computations (mat-vec, vec-vec operations) are parallelized by leveraging the domain decomposition present in the PDE solver used for finite element analysis and for the computation of the sensitivities. We will also report on the numerical performance when solving structural topology optimization problems in parallel and discuss the parallel efficiency of the proposed approach.

Cosmin G. Petra
Lawrence Livermore National
Center for Advanced Scientific Computing
petra1@llnl.gov

MS43

Semidefinite Approximations of Matrix Logarithm

We propose a new way to treat the exponential/relative entropy cone using symmetric cone solvers. Our approach is based on highly accurate rational (Padé) approximations of the logarithm function. The key to this approach is that our rational approximations, by construction, inherit the (operator) concavity of the logarithm. Importantly, our method extends to the matrix logarithm and other derived functions such as the matrix relative entropy, giving new semidefinite optimization-based tools for convex optimization involving these functions. We compare our method to the existing successive approximation scheme in CVX, and show that it can be much faster, especially for large problems.

Hamza Fawzi
University of Cambridge
h.fawzi@damtp.cam.ac.uk

James Saunderson, Pablo A. Parrilo
Massachusetts Institute of Technology
james.saunderson@monash.edu, parrilo@MIT.EDU

MS43

Optimal Resource Allocation to Control Epidemic Outbreaks in Networked Populations

We study the problem of controlling epidemic outbreaks in networked populations by distributing protection resources throughout the nodes of the network. We assume that two types of protection resources are available: (i) Preventive resources able to defend individuals in the population against the spreading of the disease (such as vaccines or disease-awareness campaigns), and (ii) corrective resources able to neutralize the spreading (such as antidotes). We assume that both preventive and corrective resources have an associated cost and study the problem of finding the cost-optimal distribution of resources throughout the networked population. We analyze these questions in the context of a viral outbreak and study the following two problems: (i) Given a fixed budget, find the optimal allocation of preventive and corrective resources in the network to achieve the highest level of disease containment, and (ii) when a budget is not specified, find the minimum budget required to eradicate the disease. We show that both resource allocation problems can be efficiently solved for a wide class of cost functions. We illustrate our approach by designing optimal protection strategies to contain an epidemic outbreak that propagates through the air transportation network.

Victor Preciado
University of Pennsylvania
preciado@seas.upenn.edu

MS43

Relative Entropy Relaxations for Signomial Optimization

Signomial programs (SPs) are optimization problems specified in terms of signomials, which are weighted sums of exponentials composed with linear functionals of a decision variable. SPs are non-convex optimization problems in general, and families of NP-hard problems can be reduced to SPs. We describe a hierarchy of convex relaxations to obtain successively tighter lower bounds of the optimal value of SPs. This sequence of lower bounds is computed by solving increasingly larger-sized relative entropy optimization problems, which are convex programs specified in terms of linear and relative entropy functions. We show how these ideas can be applied to a number of problems in various application domains (e.g. machine learning) where SPs appear in a very natural manner. We also discuss numerical aspects of our work, including a relative entropy optimization solver, and various experiments conducted using the same.

Parikshit Shah
University of Wisconsin/Philips Research
pshah@discovery.wisc.edu

Venkat Chandrasekaran

California Institute of Technology
venkatc@caltech.edu

MS43

A Positivstellensatz for Sums of Nonnegative Circuit Polynomials

Deciding nonnegativity of real polynomials is a fundamental problem in real algebraic geometry and polynomial optimization. Since this problem is NP-hard, one is interested in finding sufficient conditions (certificates) for nonnegativity, which are easier to check. Since the 19th century, the standard certificates for nonnegativity are sums of squares (SOS). In practice, SOS based semidefinite programming (SDP) is the standard method to solve polynomial optimization problems. In 2014, Iliman and I introduced an *entirely new* nonnegativity certificate based on *sums of nonnegative circuit polynomials (SONC)*, which are *independent* of sums of squares. We successfully applied SONCs to global nonnegativity problems. Recently, Dressler, Iliman, and I proved a Positivstellensatz for SONCs using results by Chandrasekaran and Shah. This Positivstellensatz provides a converging hierarchy of lower bounds for constrained polynomial optimization problems. These bounds can be computed efficiently via relative entropy programming. This result establishes SONCs as a promising alternative to SOS based SDP methods for hard polynomial optimization problems.

Timo de Wolff
Texas A&M University
dewolff@math.tamu.edu

Mareike Dressler
Goethe University, Frankfurt am Main
dressler@math.uni-frankfurt.de

Sadik Iliman
KPMG AG, Frankfurt am Main
sadik.iliman@gmx.net

MS44

Sparse Linear Programming for Optimal Adaptive Trial Design

Adaptive enrichment designs involve preplanned rules for modifying enrollment criteria based on accruing data in a randomized trial. We focus on designs where the overall population is partitioned into two predefined subpopulations, e.g., based on a biomarker or risk score measured at baseline. The goal is to learn which populations benefit from an experimental treatment. Two critical components of adaptive enrichment designs are the decision rule for modifying enrollment, and the multiple testing procedure. We provide a general method for simultaneously optimizing both of these components for two stage, adaptive enrichment designs. We minimize the expected sample size under constraints on power and the familywise Type I error rate. It is computationally infeasible to directly solve this optimization problem due to its nonconvexity. The key to our approach is a novel discrete representation of this optimization problem as a sparse linear program. We apply advanced optimization methods to solve this problem to high accuracy, revealing new, approximately optimal designs. This is a joint work with Michael Rosenblum and Han Liu

Ethan X. Fang
Pennsylvania State University
ethanfangxy@gmail.com

MS44

Large Scale Nonconvex ADMM

Abstract not available at time of publication.

Tom Luo
UMN
luozq@cuhk.edu.cn

MS44

Exploring the Second Order Sparsity in Large Scale Optimization

In this talk we shall demonstrate how the second order sparsity in important optimization problems such as the lasso models, semidefinite programming, and many others can be explored to induce efficient algorithms. The key idea is to incorporate the semismooth Newton methods into the augmented Lagrangian method framework in a way that the subproblems involved only need low to medium costs, e.g., for lasso problems with sparse solutions, the cost for solving the subproblems at each iteration is no more than that for most of the first order methods and can be much cheaper than some first order methods. Consequently, with the fast convergence rate in hand, usually asymptotically superlinear linear, we now reach at the stage of being able to solve many challenging large scale convex optimization problems efficiently and robustly.

Defeng Sun
Dept of Mathematics
National University of Singapore
matsundf@nus.edu.sg

MS44

Nonconvex Statisitcal Optimization: From Local Exploitation to Global Exploration

Taming nonconvexity requires a seamless integration of both global exploration and local exploitation of latent convexity. In this talk, we study a generic latent variable model that includes PCA, ICA, and dictionary learning as special cases. Using this statistical model as illustration, we propose two global exploration schemes for identifying the basin of attraction, namely tightening after relaxation and noise regularization, which allow further local exploitation via gradient-based methods. Interestingly, given a sufficiently good initialization, the local exploitation methods are often information-theoretic optimal. In sharp contrast, based on an oracle computational model, we prove that, one often has to pay a computational or statistical price in the global exploration stage to attain a good initialization. This is joint work with Zhuoran Yang, Junchi Li, and Han Liu.

Zhaoran Wang
Princeton Univ
zhaoranwang@gmail.com

MS45

Nested Min Cuts and Bipartite Queueing System

Abstract not available at time of publication.

Yichuan Ding, Yichuan Ding
Sauder School of Business

University of British Columbia
daniel.ding@sauder.ubc.ca, daniel.ding@sauder.ubc.ca

MS45

Combinatorial Algorithms for Clustering, Image Segmentation and Data Mining

Abstract not available at time of publication.

Dorit Hochbaum
Department of IE&OR
UC Berkeley
dhochbaum@berkeley.edu

MS45

Rerouting Flow in a Network

Abstract not available at time of publication.

Jannik Matuschke
TU Berlin
matuschke@math.tu-berlin.de

MS45

A Unified Framework for Determining the Structural Properties of Nested Min Cuts

Abstract not available at time of publication.

S. Thomas McCormick
Sauder School of Business, Vancouver, Canada
tom.mccormick@sauder.ubc.ca

MS46

Estimation and Computation of Log-Concave Density Estimators

We consider nonparametric maximum likelihood estimation of a log-concave density with further constraints imposed. These are useful for using likelihood ratio tests to perform inference, and in their own right. In particular we consider the mode-constrained log-concave nonparametric maximum likelihood estimator (MLE). One reason this MLE is useful is that it is simultaneously nonparametric, yet automatically selects optimal bandwidths without any user dependent tuning required, unlike e.g. kernel density estimators. We will discuss using the constrained MLE in modal regression, as well as its use in a likelihood ratio test for the mode. We will discuss the algorithms used for computation of the estimators.

Charles Doss
School of Statistics, University of Minnesota
313 Ford Hall, 224 Church St. SE, Minneapolis, MN
55455
cdoss@umn.edu

MS46

On the Accuracy of the Kiefer-Wolfowitz MLE for Gaussian Mixtures

Given real-valued observations $X_1, \ldots, X_n$, the Kiefer-Wolfowitz MLE for gaussian mixtures is defined as the (nonparametric) maximum likelihood estimator over the class of all gaussian mixtures (densities of the form $\int \phi(x-\theta)d\mu(\theta)$ where $\phi(x) := (2\pi)^{-1}\exp(-x^2/2)$ and μ is a probability measure on the real line). This estimator (contrary to most mixture estimates) can be computed via convex optimization. I shall present some results on the accuracy of this estimator when the data is generated i.i.d from a gaussian mixture. I will also briefly mention extensions to the multidimensional case.

Aditya Guntuboyina
Department of Statistics, University of California,
Berkeley
423 Evans Hall, Berkeley, CA 94720
aditya@stat.berkeley.edu

MS46

Consistency in Nonparametric Estimation Using Variational Analysis

This presentation discusses density estimation in situations when the sample (hard information) is supplemented by soft information about the random phenomenon. These situations arise broadly in operations research and management science where practical and computational reasons severely limit the sample size, but problem structure and past experiences could be brought in. We adopt a constrained maximum likelihood estimator that incorporates any, possibly random, soft information through an arbitrary collection of constraints. We illustrate the breadth of possibilities by discussing soft information about shape, support, continuity, smoothness, slope, location of modes, symmetry, density values, neighborhood of known density, moments, and distribution functions. The maximization takes place over spaces of extended real-valued semicontinuous functions and therefore allows us to consider essentially any conceivable density as well as convenient exponential transformations. The infinite dimensionality of the optimization problem is overcome by approximating splines tailored to these spaces.

Johannes O. Royset
Operations Research Department
Naval Postgraduate School
joroyset@nps.edu

MS46

On Multivariate Convex Regression

In the talk I will describe properties of the nonparametric least squares estimator (LSE) of a convex regression function when the predictor is multidimensional. We characterize the LSE and discuss the computation of the estimator. We study theoretical properties of the estimator: its consistency, global rate of convergence, and adaptation properties. I will also point to a few applications of convex regression.

Bodhisattva Sen
Department of Statistics, Columbia University
1255 Amsterdam Ave, Room # 1032 SSW, New York,
NY 10027
bodhi@stat.columbia.edu

MS47

Feasible Parameter Sets in Statinary Models of Gas Transportation

The concept of feasible parameter set established in parametric optimization long ago has regained attention in mathematical decision support in gas network operation. Here, parameters correspond to vectors of in- and mostly output of gas quantities to be sent through the network.

Feasible parameters correspond to valid nominations, and that is what gas companies are after since they must guarantee technical feasibility to the customers. Of similar, maybe even bigger, importance to the company is to be able to "prove' technical infeasibility if they turn down a nomination. Due to the uncertainty which is inherent to gas networks it is relevant to assign probabilities to subsets of nominations for being feasible. In the talk we are discussing feasible parameter sets for pipeline systems without meshes or with at most one. We conclude with an outlook to systems with a low number of mutually interwoven cycles.

Ralf Gollmer, Matthias Claus
University of Duisburg-Essen
ralf.gollmer@uni-due.de, matthias.claus@uni-due.de

MS47

Steady-State Models for Optimizing Gas Transportation: Analytical and Algebraic Perspectives

Steady-state models for gas transportation involving Kirchhoff's Laws obey interesting mathematical properties. From a more analytical point-of-view coercivity and monotonicity come to the fore. Algebraically considered, different types of polynomial descriptions are possible. This mobilizes application of ideas from, for instance, Gröbner bases theory and methods or from polynomial optimization in the presence of zero-dimensional ideals. In the talk, we elaborate on these topics and provide applications to feasibility analysis (nomination validation) of transportation orders in mildly meshed gas transportation networks with different densities of meshes. A specific feature of our models is in the presence of absolute values which has to do with flow directions. Therefore we present some rules to check for inconsistencies leading to elimination of irrelevant resolutions of absolute values.

Ruediger Schultz
University of Duisburg-Essen
ruediger.schultz@uni-due.de

MS47

On the Robustification of Uncertain Physical Parameters in Gas Networks

A class of stationary gas network problems is investigated. The challenge is to decide whether a set of demands can be satisfied by the network if not all characteristic data as, e.g. the roughness of the pipes are precisely known. To handle these uncertain physical parameters, a robust approach is suggested. It is shown that if, for a passive network, the pressure loss is modeled by the Weymouth equation, certificates for feasibility and infeasibility can be derived. The robustified problems can be rephrased as polynomial optimization problems. If active elements are involved such reformulations are not possible. If this case, the nonlinear pressure loss is approximated by a piecewise linearization scheme introducing auxiliary binary and continuous variables. As these variables approximate the pressure and flow along the pipes, they have to be adjustable with respect to the uncertain parameters. Thus, one is faced with the solution of a two-stage robust optimization problem with binary variables at second stage. In order to cope with this, a piecewise approximation of the second stage variables is suggested. It is demonstrated that in certain cases the underlying decomposition of the uncertainty set can be performed such that the second stage binary variables are eliminated. The practical relevance of the theoretical results will be demonstrated by a set of small to medium size numerical examples.

Michael Stingl
Friedrich-Alexander University of Erlangen-Nuremberg
stingl@math.fau.de

MS47

Control and Estimation Problems in Transient Gas Pipeline Flow

Renewed interest in the dynamic behavior of gas pipeline networks motivates a fundamental examination of modeling, simulation, optimal control, and estimation techniques for these systems. Mathematically, compressible fluid flow in a pipeline system can be represented by nonlinear hyperbolic partial differential equations (PDE) on a metric graph subject to time-dependent boundary conditions and nodal compatibility laws. In addition to mass flow balance at network nodes, where fluid may be injected or withdrawn from the network, flow can be actuated by compressors or regulators. The resulting highly nonlinear and ill-conditioned distributed system model defies traditional control-theoretic and computational treatment. After formulating a control system model and applying approximations, we can formulate constrained optimal control problems for flow scheduling, and state and/or parameter estimation problems, as large-scale nonlinear programs. Solutions using finer discretization converge to feasible optima of the continuous problem, which can be validated by simulation.

Anatoly Zlotnik
Los Alamos National Laboratory
azlotnik@lanl.gov

MS48

Robust Empirical Optimization is almost the same as Mean-Variance Optimization

We formulate a distributionally robust optimization problem where the empirical distribution plays the role of the nominal model, the decision maker optimizes against a worst-case alternative, and the deviation of the alternative distribution from the nominal is controlled by a ϕ-divergence penalty in the objective. Our main finding is that a large class of robust empirical optimization problems of this form are essentially equivalent to an in-sample mean-variance problem. Intuitively, controlling the variance reduces the sensitivity of a decision's expected reward to perturbations in the tail of the in-sample reward distribution. This in turn reduces the sensitivity of its out-of-sample performance to perturbations in the nominal model, which is precisely the notion of robustness. As a practical application of our main result, we introduce the notion of a robust mean-variance frontier, which can be used as a calibration tool for the robust model. To illustrate the usefulness of such frontiers we consider a real world portfolio optimization application. Our numerical experiments show that the first-order effect of robustness is variability reduction, followed by reduction in the mean reward.

Jun-ya Gotoh
Chuo University
jgoto@indsys.chuo-u.ac.jp

Michael Kim
University of Toronto

mikekim@mie.utoronto.ca

Andrew Lim
National University of Singapore
andrewlim@nus.edu.sg

MS48

Solving Stochastic Variational Inequalities by Progressive Hedging

Most of the research on stochastic variational inequalities has concentrated on models in which information about the uncertain future is revealed only once. Such models are inadequate to cover multistage stochastic programming, where information comes in stages that offer repeated opportunities for recourse decisions. That feature can be brought into stochastic variational inequalities by adapting them to a constraint of nonanticipativity. In that way not only stochastic programming but multistage multiagent games can be covered. A particular advantage of this approach is that it generates information price vectors which can be used to decompose the overall problem into a separate problem for each scenario. This fits with solution approaches like the progressive hedging algorithm. However, the ideas behind this algorithm offer the prospect of being applicable outside of a stochastic framework and under localized versions of convexity/monotonicity which often may be elicited from the problem structure in the manner of augmented Lagrangian methodology.

Terry Rockafellar
University of Washington
Department of Mathematics
rtr@math.washington.edu

MS48

Quasi-Monte Carlo and Sparse Grid Quadratures can be Efficient for Solving Two-Stage Stochastic Optimization Models

Sparse grid and randomized Quasi-Monte Carlo quadratures can be efficient for multivariate numerical integration have first order mixed partial Sobolev derivatives with respect to all variables. This remains true if a smoothing technique based on dimension reduction applies and the integrand in stochastic optimization models allows an approximate representation by lower order smooth ANOVA terms. We discuss the theoretical background of all methods and techniques, and illustrate the theory by a practical example originating from electric power industry.

Werner Roemisch
Humboldt University Berlin
Department of Mathematics
romisch@math.hu-berlin.de

Hernan Leovey
Humboldt-University Berlin, Institute of Mathematics
leovey@math.hu-berlin.de

MS48

Surrogate Models for Characterizing Tail Statistics of Systems Governed by PDEs

Performance of physical and engineering systems is often described by statistics of distribution tails such as quantile and probability of exceedance. However, these functions have poor mathematical properties for distributions described by sample statistics. Therefore, it is very difficult to optimize such quantities and build systems with the optimal characteristics. The recent trend is to use alternative characteristics, serving similar purposes: Conditional Value-at-Risk (CVaR) and Buffered Probability of Exceedance (bPOE). CVaR and bPOE can be optimized with convex and liner programming algorithms. We build surrogates for CVaR and bPOE of systems described by Partial Differential Equations (PDEs) with stochastic parameters. The surrogates are estimated with linear regression algorithms.

Stan Uryasev
University of Florida
uryasev@ufl.edu

MS49

Matrix Problems with No Spurious Local Minima

We show that there are no spurious local minima in the non-convex factorized parametrization of low-rank matrix recovery from incoherent linear measurements. With noisy measurements we show all local minima are very close to a global optimum. Together with a curvature bound at saddle points, this yields a polynomial time global convergence guarantee for stochastic gradient descent from random initialization.

Nathan Srebro
Toyota Technical Institute - Chicago
nati@ttic.edu

Srinadh Bhojanapalli
Toyota Technological Institute at Chicago
srinadh@ttic.edu

MS49

Fast Algorithms for Robust PCA via Gradient Descent

Abstract not available at time of publication.

Constantine Caramanis
The University of Texas at Austin
caramanis@mail.utexas.edu

MS49

Provably Robust Low-Rank Matrix Recovery via Median-Truncated Gradient Descent

Low-rank matrix recovery is of central interest to many applications such as collaborative filtering, sensor localization, and phase retrieval. Convex relaxation via nuclear norm minimization is a powerful method, but suffers from computational complexity when the problem dimension is high. Recently, it has been demonstrated that gradient descent works provably for low-rank matrix recovery if initialized using the spectral method, therefore achieving both computational and statistical efficacy. However, this approach fails in the presence of arbitrary, possibly adversarial, outliers in the measurements. In this talk, we will describe how to modify the gradient descent approach via a median-guided truncation strategy, and show this yields provably recovery guarantees for low-rank matrix recovery. While median has been well-known in the robust statistics literature, its utility in high-dimensional signal estimation is novel in this work. In particular, robust phase retrieval will be highlighted as a special case. This is based on joint work with Huishuai Zhang, Yuanxin Li, and Yingbin Liang.

Yuejie Chi
Ohio State University
chi.97@osu.edu

MS49

Low-Rank Matrix Completion (LRMC) Using Nuclear Norm (NN) with Facial Reduction (FR)

Minimization of the NN is often used as a surrogate, convex relaxation, for solving LRMC problems. The minimum NN problem can be solved as a trace minimization semidefinite program (SDP). The SDP and its dual are regular in the sense that they both satisfy strict feasibility. FR has been successful in regularizing many problems where strict feasibility fails, e.g., SDP relaxations of discrete optimization problems such as QAP, GP, as well as sensor network localization. Here we take advantage of the structure at optimality for the NN minimization and show that even though strict feasibility holds, the FR framework can be successfully applied to obtain a proper face that contains the optimal set. This can dramatically reduce the size of the final NN problem while guaranteeing a low-rank solution. We include numerical tests for both exact and noisy cases. (work with Shimeng Huang)

Henry Wolkowicz
University of Waterloo
Department of Combinatorics and Optimization
hwolkowicz@uwaterloo.ca

MS50

Thompson Sampling for Reinforcement Learning

We consider a variation of “Thompson Sampling” (TS), i.e., Bayesian posterior sampling heuristics for managing exploration-exploitation trade-off, to design an on-policy learner for a finite state, finite action, average reward reinforcement learning problem. We compare the average reward of TS to average reward of the optimal stationary policy, and demonstrate that TS achieves near-optimal worst case regret bound for weakly communicating MDPs.

Shipra Agrawal
Department of Industrial Engineering and Operations Research
Columbia University
sa3305@columbia.edu

MS50

Overview of Reinforcement Learning Algorithms

Reinforcement learning (RL) is an area of machine learning for creating agents that can learn long-term successful behavior through autonomous interaction with the environment. Modern RL is a topic studied in different communities, including computer science, optimal control, psychology, and neuroscience, among others. This talk gives an overview of RL Algorithms, introducing key challenges from the optimization perspective, and lay the common ground for other talks in this minisymposium.

Lihong Li
Microsoft Research
lihongli.cs@gmail.com

MS50

Gradient, Semi-Gradient and Pseudo-Gradient Reinforcement Learning

In this talk we I will present the establishment of a unified general framework for stochastic-gradient-based temporal-difference learning algorithms that use proximal gradient methods. The primal-dual saddle-point formulation is introduced, and state-of-the-art stochastic gradient solvers, such as mirror descent and extragradient are used to design several novel RL algorithms. The finite-sample analysis is given along with detailed empirical experiments to demonstrate the effectiveness of the proposed algorithms. Several important extensions, such as control learning, variance reduction, acceleration, and regularization will also be discussed in detail.

Bo Liu
Department of Computer Science and Software Engineering
Auburn University
boliu@auburn.edu

MS50

Algorithms for Safe Reinforcement Learning

As our ability to construct intelligent machines improves, so too must our ability to ensure that they are *safe*. In this talk I will review past efforts to ensure that machine learning algorithms, and particularly reinforcement learning algorithms, ensure different forms of safety. I will then discuss some ongoing research on this topic as well as future research directions.

Philip Thomas
School of Computer Science
Carnegie Mellon University
pthomascs@gmail.com

MS51

Symmetric Sums of Squares over K-Subset Hypercubes

We consider the problem of finding sum of squares (sos) expressions to establish the non-negativity of a symmetric polynomial over a discrete hypercube whose coordinates are indexed by k-element subsets of $[n]$. We develop a variant of the Gatermann-Parrilo symmetry-reduction method tailored to our setting that allows for several simplifications and a connection to Razborov's flag algebras. We show that every symmetric polynomial that has a sos expression of a fixed degree also has a succinct sos expression whose size depends only on the degree and not on the number of variables. This is joint work with James Saunderson, Mohit Singh and Rekha Thomas.

Annie Raymond
University of Washington
annieraymond@gmail.com

James Saunderson
Massachusetts Institute of Technology
james.saunderson@monash.edu

Mohit Singh
Microsoft Research

mohits@microsoft.com

Rekha Thomas
University of Washington
rrthomas@uw.edu

MS51

Lieb's Concavity Theorem, Matrix Geometric Means, and Semidefinite Optimization

A famous result of Lieb establishes that the map $(A, B) \mapsto \mathrm{tr}(K^*A^{1-t}KB^t)$ is jointly concave in the pair (A, B) of positive definite matrices, where K is a fixed matrix and $t \in [0, 1]$. This fact leads to concavity and convexity properties of many related functions that arise naturally, for instance, in quantum information. In this talk we discuss explicit descriptions of Lieb's function, for any rational $t \in [0, 1]$, in terms of the feasible regions of semidefinite optimization problems. Such descriptions allow us to use standard software for semidefinite programming to solve convex optimization problems involving Lieb's function, and a number of related functions.

James Saunderson
Massachusetts Institute of Technology
james.saunderson@monash.edu

Hamza Fawzi
University of Cambridge
h.fawzi@damtp.cam.ac.uk

MS51

Solving Generic Nonarchimedean Semidefinite Programs Using Stochastic Game Algorithms

One of the basic algorithmic questions associated with semidefinite programming is to decide whether a spectrahedon is empty or not. It is unknown whether this problem belongs to NP in the Turing machine model, and the state-of-the-art algorithms that solve this task exactly are based on cylindrical decomposition or the critical points method. We study the nonarchimedean analogue of this problem, replacing the field of real numbers by the field of Puiseux series. We introduce the notion of tropical spectrahedra and show that, under genericity conditions, these objects can be described explicitly by systems of polynomial inequalities in the tropical semiring. We demonstrate that these inequalities arise from Shapley operators associated with stochastic mean payoff games. As a consequence, we show that tropically generic semidefinite feasibility problems over Puiseux series can be solved efficiently using combinatorial algorithms designed for stochastic games. Finally, we discuss some consequences of this result for semidefinite programming over the reals.

Xavier Allamigeon
INRIA and CMAP, Ecole Polytechnique, CNRS
xavier.allamigeon@inria.fr

Stephane Gaubert
INRIA and CMAP, Ecole Polytechnique
stephane.gaubert@inria.fr

Mateusz Skomra
INRIA Saclay / Ecole Polytechnique
mateusz.skomra@polytechnique.edu

MS51

Do Sums of Squares Dream of Free Resolutions?

Let X be a variety defined by a homogeneous ideal I_X. We show that the number of steps for which the minimal free resolution of I_X is linear, is a computable lower bound for the next-to-minimal rank of the extreme rays of the cone dual to the sums of squares in X. As a consequence, we obtain: (1) A complete classification of totally real reduced schemes for which nonnegative quadratic forms are sums of squares generalizing Froberg's Theorem on PSD matrix completions. (2) New certificates of exactness for truncated moment problems on projective varieties.

Mauricio Velasco
Universidad de los Andes
mvelasco@uniandes.edu.co

Mauricio Velasco
University of los Andes
mvelasco@uniandes.edu.co

Greg Blekherman, Rainer Sinn
Georgia Institute of Technology
greg@math.gatech.edu, rsinn3@math.gatech.edu

MS52

New Results on the Simplex Method for Minimum Cost Flow Problems in Infinite Networks

We provide a simplex algorithm for a class of uncapacitated countably infinite network flow problems. Previous efforts to develop simplex algorithms for countably infinite network flow problems required explicit capacities on arcs with uniformity properties to facilitate duality arguments. In contrast, our method takes a "primal approach" by devising a simplex method that produces iterates that converge in optimal value without relying on duality arguments and facilitates the removal of explicit capacity bounds. Instead, our approach leverages a compactness assumption on the set of simplex iterates. This assumption holds in a variety of applied settings, including infinite-horizon production planning and nonstationary infinite-horizon dynamic programming.

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

Christopher T. Ryan
University of Chicago
chris.ryan@chicagobooth.edu

Robert L. Smith
Industrial and Operations Engineering
University of Michigan
rlsmith@umich.edu

MS52

Stability and Continuity in Robust Optimization by Mappings to Linear Semi-Infinite Optimization Problems

We consider the stability of robust optimization (RO) problems with respect to perturbations in their uncertainty

sets. In particular, we focus on robust linear optimization problems, including those with an infinite number of constraints, and consider uncertainty in both the cost function and constraints. We prove Lipschitz continuity of the optimal value and ϵ-approximate optimal solution set with respect to the Hausdorff distance between uncertainty sets. The Lipschitz constant can be calculated from the problem data. In addition, we prove closedness and upper semicontinuity for the optimal solution set with respect to the uncertainty set. In order to prove these results, we develop a novel transformation that maps RO problems to linear semi-infinite optimization (LSIO) problems in such a way that the distance between uncertainty sets of two RO problems correspond to a measure of distance between their equivalent LSIO problems. Using this isometry we leverage LSIO and Variational Analysis stability results to obtain stability results for RO problems.

Timothy C.Y. Chan, Philip Allen Mar
Department of Mechanical and Industrial Engineering
University of Toronto
tcychan@mie.utoronto.ca, philip.mar@mail.utoronto.ca

MS52

Strong Duality and Sensitivity Analysis in Semi-Infinite Linear Programming

Finite-dimensional linear programs satisfy strong duality (SD) and have the “dual pricing” (DP) property. The (DP) property ensures that, given a sufficiently small perturbation of the right-hand-side vector, there exists a dual solution that correctly “prices” the perturbation by computing the exact change in the optimal objective function value. These properties may fail in semi-infinite linear programming where the constraint vector space is infinite dimensional. Unlike the finite-dimensional case, in semi-infinite linear programs the constraint vector space is a modeling choice. We show that, for a sufficiently restricted vector space, both (SD) and (DP) always hold, at the cost of restricting the perturbations to that space. The main goal of the paper is to extend this restricted space to the largest possible constraint space where (SD) and (DP) hold. Once (SD) or (DP) fail for a given constraint space, then these conditions fail for all larger constraint spaces. We give sufficient conditions for when (SD) and (DP) hold in an extended constraint space. Our results require the use of linear functionals that are singular or purely finitely additive and thus not representable as finite support vectors. We use the extension of the Fourier-Motzkin elimination procedure to semi-infinite linear systems to understand these linear functionals.

Amitabh Basu
Johns Hopkins University
basu.amitabh@jhu.edu

Kipp Martin, Christopher T. Ryan
University of Chicago
kmartin@chicagobooth.edu, chris.ryan@chicagobooth.edu

MS52

Policy Iteration for Robust Countable-State Markov Decision Processes

Existing solution methods for robust Markov decision processes (MDPs) cannot be implemented “as-is” when the state-space is countable, since they entail infinite amounts of computation in each step. We present a policy iteration algorithm for robust countable-state MDPs with bounded costs, which performs finitely-implementable approximate variants of policy evaluation and policy improvement in each iteration. We prove that the value functions for the sequence of policies produced by this algorithm converge monotonically to the optimal value function. The policies themselves converge subsequentially to an optimal policy.

Saumya Sinha, Archis Ghate
University of Washington
saumya@uw.edu, archis@uw.edu

MS53

PETSc and PDE-Constrained Optimization

We describe TAO, Argonne's Toolkit for Advanced Optimization. TAO provides numerical optimization methods for high-performance computing. These methods are built upon PETSc, the Portable Extensible Toolkit for Advanced Optimization. I will discuss the overall design of the software describe some of the available methods, and show how to solve PDE constrained optimization problems using TAO. Time permitting, I will demonstrate the parallel availability of TAO.

Todd Munson
Argonne National Laboratory
Mathematics and Computer Science Division
tmunson@mcs.anl.gov

MS53

PDE-Constrained Shape Optimization with Fem-Tailored Discretization of Diffeomorphisms

PDE-constrained shape optimization problems are characterized by target functionals that depend both on the shape of a domain (the control) and on the solution of a boundary value problem formulated on that domain (the state). There are conflicting demands on the discretization of the domain. On the one hand, k^{th}-order finite element approximation of the state suggests $W^{k,\infty}$-smooth domain deformations. On the other hand, typical isoparametric finite element discretization only allows $W^{1,\infty}$ piecewise polynomial representations of the domain geometry. How can the control and state spaces be discretized to satisfy these conflicting demands? In this talk we develop a technique for cleanly decoupling a B-spline based *discretization of the control* and the *geometry used to solve the discretized state equation*. In particular, the discretized control satisfies the regularity requirement, which in turn guarantees that the standard finite element estimates hold. As a consequence, the state can be approximated efficiently with typical finite element software.

Alberto Paganini
seminar for applied mathematics
ETH, Zurich
paganini@maths.ox.ac.uk

Patrick E. Farrell
Mathematical Institute
University of Oxford
patrick.farrell@maths.ox.ac.uk

Florian Wechsung
University of Oxford

wechsung@maths.ox.ac.uk

MS53

Parallel-in-Time Methods for Large-Scale PDE-constrained Optimization

We study time-domain decomposition methods for optimality systems arising in optimal control problems with nonlinear transient partial differential equation (PDE) constraints. The methods are based on a concurrent solution of smaller optimality systems in each time subdomain. The coupling in time is recovered gradually at each optimization iteration through an inexact Krylov scheme combined with a matrix-free trust-region sequential quadratic programming (SQP) algorithm. The SQP algorithm controls the degree of inexactness based on feasibility and optimality criteria, and guarantees full time continuity at the optimal solution. We present numerical results involving optimal control of incompressible flows.

Denis Ridzal
Sandia National Laboratories
dridzal@sandia.gov

MS54

On Intersection of Two Mixing Sets with Applications to Joint Chance-Constrained Programs

We study the polyhedral structure of a generalization of a mixing set described by the intersection of two mixing sets with two shared continuous variables, where one continuous variable has a positive coefficient in one mixing set and a negative coefficient in the other. Our developments are motivated from a key substructure of linear joint chance-constrained programs (CCPs) with random right-hand sides from a finite probability space. The CCPs of interest immediately admit a mixed-integer programming reformulation. Nevertheless, such standard reformulations are difficult to solve at large-scale due to the weakness of their linear programming relaxations. In this talk, we provide a systemic polyhedral study of such joint CCPs by explicitly analyzing the system obtained from simultaneously considering two linear constraints inside the chance constraint. We propose a new class of valid inequalities in addition to the mixing inequalities and establish conditions under which these inequalities are facet defining. Moreover, under certain additional assumptions, we prove that these new valid inequalities along with the classical mixing inequalities are sufficient in terms of providing the closed convex hull description of our set. We complement our theoretical results with a computational study on the strength of the proposed inequalities.

Xiao Liu
The Ohio State University
liu.2738@osu.edu

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

Simge Kucukyavuz
Industrial and Systems Engineering
University of Washington
simge@uw.edu

MS54

Solving Standard Quadratic Programming By Cutting Planes

Standard quadratic programs are non-convex quadratic programs with the only constraint that variables must belong to a simplex. By a famous result of Motzkin and Straus, those problems are connected to the clique number of a graph. We propose cutting planes to obtain strong bounds: our cuts are derived in the context of Spatial Branch & Bound, where linearization variables represent products. Their validity is based on Motzkin-Straus result. We study the relation between these cuts and the ones obtained by the first RLT level. We present extensive computational results using the cuts in the context of the Spatial Branch & Bound implemented by the commercial solver CPLEX.

Andrea Lodi
École Polytechnique de Montréal
andrea.lodi@polymtl.ca

Pierre Bonami
IBM, Spain
pierre.bonami@es.ibm.com

Jonas Schweiger, Andrea Tramontani
IBM Italy
jonas@posteo.de, andrea.tramontani@it.ibm.com

MS54

Revisiting the Factorable Relaxation Scheme

In factorable programming, products of functions are relaxed by constructing outer-approximation of functions and McCormick envelopes for products. We construct tighter relaxations by exploiting the outer-approximation while relaxing function products. We show that convex hull of the product can be constructed in closed-form over cartesian product of some specific polytopes, where each polytope encapsulates information regarding outer-approximators and their bounds. The result generalizes to concave envelopes of supermodular functions linear in each argument. We provide preliminary computational experience with the new relaxations.

Mohit Tawarmalani
Krannert School of Management
mtawarma@purdue.edu

Taotao He
Purdue University
he135@purdue.edu

MS54

Optimization Methodologies for Truss Topology Design Problems

We consider various mathematical models for truss topology design problems with the objective to minimize the weight, where the non-convex Euler bucking constraints are also considered. We propose solution approaches to either alleviate the non-convexity of the problem, or to obtain a tight lower bound of the optimal objective value. Using a discretization approach, the non-convex optimization model is transformed to a mixed integer linear optimization model. We provide numerical experiments to demonstrate the performance of the discretization approach and the Mc-Cormick relaxation.

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

Mohammad Shahabsafa, Ali Mohammad Nezhad
Dept. Industrial and Systems Engineering
Lehigh University
mos313@lehigh.edu, alm413@lehigh.edu

Luis Zuluaga
Lehigh University
luis.zuluaga@lehigh.edu

Sicheng He, John T. Hwang, Joaquim R. R. A Martins
University of Michigan
hschsc@umich.edu, hwangjt@umich.edu, jrram@umich.edu

MS55

Restarted Sub-Gradient Method for Non-Smooth Optimization under Error Bound Conditions

In this paper, we show that, when applied to a class of convex optimization problems that satisfies some error bound conditions, a restarted sub-gradient (RSG) method can find an optimal solution with a much lower iteration complexity than the standard sub-gradient method. This class of problems include polyhedral optimization, and the optimization whose objective functions have local quadratic growth, or admit a local Kurdyka-Lojasiewicz inequality. Moreover, we show that a homotopy smoothing method can further reduce the complexity of RSG when the problem satisfies an additional structured non-smooth property

Qihang Lin, Tianbao Yang, Yi Xu
University of Iowa
qihang-lin@uiowa.edu, tianbao-yang@uiowa.edu, yi-xu@uiowa.edu

Yan Yan
University of Technology Sydney
yan.yan-3@student.uts.edu.au

MS55

Further Properties of the Forward-Backward Envelope with Applications to Difference-of-Convex Programming

We further study the forward-backward envelope first introduced by Patrinos and Bemporad for problems whose objective is the sum of a convex function and a twice continuously differentiable possibly nonconvex function with Lipschitz continuous gradient. We derive sufficient conditions for the corresponding forward-backward envelope to be a level-bounded and Kurdyka- Lojasiewicz function with an exponent of $\frac{1}{2}$. We also demonstrate how to minimize some difference-of-convex regularized least squares problems by minimizing a suitably constructed forward-backward envelope. Finally, numerical results are shown.

Tianxiang Liu
the Hong Kong Polytechnic University
tiskyliu@polyu.edu.hk

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

MS55

An Inertial Forward-Douglas-Rachford Splitting Method and Applications

We propose an inertial Forward-Douglas-Rachford splitting method for solving monotone inclusions involving three operator. The weak and strong convergence of the iteration are proved under suitable condition. Some applications to minimization problems, system of monotone inclusions are demonstrated. Numerical results are provided.

Bang Cong Vu, Volkan Cevher, Alp Yurtsever
EPFL
bang.vu@epfl.ch, volkan.cevher@epfl.ch, alp.yurtsever@epfl.ch

MS55

The Common-Directions Method for Regularized Empirical Risk Minimization

State-of-the-art first-order optimization methods are able to achieve optimal linear convergence rate, but cannot obtain quadratic convergence. In this work, we propose a first-order method for regularized empirical risk minimization that exploits the problem structure to efficiently combine multiple update directions to obtain both optimal global linear convergence rate and local quadratic convergence. Experimental results show that our method outperforms state-of-the-art first- and second-order optimization methods in terms of the number of data accesses, while is competitive in training time.

Po-Wei Wang
Carnegie Mellon University
poweiw@cs.cmu.edu

Ching-Pei Lee
University of Wisconsin-Madison
ching-pei@cs.wisc.edu

Chih-Jen Lin
National Taiwan University
cjlin@csie.ntu.edu.tw

MS56

Nonlinear Model Predictive Control on a Heterogeneous Computing Platform

Model Predictive Control (MPC) is a computationally demanding control technique that allows explicit performance optimization and systematic constraint handling. Although traditionally MPC algorithms have been implemented on general-purpose CPU-based machines, recent developments in reconfigurable computing resulted in high-speed and resource-efficient hardware implementations of MPC. We present an implementation of a nonlinear MPC on a heterogeneous embedded processer that combines advantages of conventional CPUs and reconfigurable platforms, namely Field-Programmable Gate Arrays (FPGAs). We also present a method for scheduling sparse matrix-vector multiplications in a Lanczos kernel to enable significant improvements in terms of the computation time vs resource usage when computing a search direction. A case study is used to demonstrate that a heterogeneous implementation can accelerate a pure software realization by a factor of 10, while achieving a 10x memory saving compared to existing approaches.

Bulat Khusainov
Imperial College
b.khusainov@imperial.ac.uk

Eric C. Kerrigan
Imperial College London
e.kerrigan@imperial.ac.uk

Andrea Suardi
Imperial College London
Electrical and Electronic Engineering
a.suardi@imperial.ac.uk

George Constantinides
Imperial College London
g.constantinides@imperial.ac.uk

MS56

Parallel Temporal Decomposition for Long-Term Production Cost Model in Power Grid

Abstract not available at time of publication.

Kibaek Kim
Mathematics and Computer Science Division
Argonne National Laboratory
kimk@anl.gov

MS56

Algorithms for Large-Scale Mixed Integer Programming

Abstract not available at time of publication.

Ted Ralphs
Department of Industrial and Systems Engineering
Lehigh University
ted@lehigh.edu

MS56

Comparison between ug[PIPS-SBB, MPI] and Parallel PIPS-SBB on Large Scale Distributed Memory Computing Environment

In this talk, we present and compare two frameworks for distributed-memory branch-and-bound (B&B) tree search. Both of these are implemented as extensions of PIPS-SBB, which is already a parallel distributed-memory stochastic MIP solver based on the distributed memory stochastic LP solver PIPS-S. As result, both frameworks include two levels of distributed-memory parallelism, for solving LP relaxations and for B&B tree search. The first framework relies on UG, part of the SCIP project, and implements an external coarse parallelization of PIPS-SBB. The second framework, Parallel PIPS-SBB implements a novel internal fine-grained parallelization of PIPS-SBB. We present computational results that evaluate the effectiveness of both frameworks.

Yuji Shinano
Zuse Institute Berlin
Takustr. 7 14195 Berlin
shinano@zib.de

Lluis Miquel Munguia
Georgia Institute of Technology
lluis.munguia@gatech.edu

Geoffrey M. Oxberry, Deepak Rajan
Lawrence Livermore National Laboratory
oxberry1@llnl.gov, rajan3@llnl.gov

MS57

Cubic Regularization For Symmetric Rank-1 and Nonlinear Conjugate Gradient Methods

Regularization techniques have been used to help existing algorithms solve "difficult" nonlinear optimization problems. Over the last decade, regularization has been proposed to remedy issues with equality constraints and equilibrium constraints, bound Lagrange multipliers, and identify infeasible problems. In this talk, we will focus on the application of cubic regularization in the context of the symmetric rank-one and conjugate gradient methods for nonlinear programming.

Hande Y. Benson
Drexel University
Department of Decision Sciences
benson@drexel.edu

David Shanno
Rutgers University (Retired)
shannod@comcast.net

MS57

Quadratic Regularization with Cubic Descent for Unconstrained Optimization

Cubic-regularization and trust-region methods with worst-case first-order complexity $O(\varepsilon^{-3/2})$ and worst-case second-order complexity $O(\varepsilon^{-3})$ have been developed in the last few years. In this paper it is proved that the same complexities are achieved by means of a quadratic-regularization method with a cubic sufficient-descent condition instead of the more usual predicted-reduction based descent. Asymptotic convergence and order of convergence results are also presented. Finally, some numerical experiments comparing the new algorithm with a well-established quadratic regularization method are shown.

Ernesto G. Birgin
Department of Computer Science - IME
University of São Paulo
egbirgin@ime.usp.br

Jose Mario Martinez
IMECC-UNICAMP
martinez@ime.unicamp.br

MS57

Primal-Dual Interior Methods for Nonlinear Optimization

Regularization and stabilization are vital tools for resolving the numerical and theoretical difficulties associated with ill-posed or degenerate optimization problems. Broadly

speaking, regularization involves perturbing the underlying linear equations so that they are always nonsingular. Stabilized methods are designed to provide a sequence of iterates with fast local convergence even when the gradients of the constraints satisfied at a solution are linearly dependent. We discuss the crucial role of regularization and stabilization in the formulation and analysis of modern interior methods for nonlinear optimization. In particular, we establish the close relationship between regularization and stabilization, and propose a new interior method based on formulating an associated "simpler" optimization subproblem defined in terms of both primal and dual variables.

Philip E. Gill
University of California, San Diego
Department of Mathematics
pgill@ucsd.edu

Vyacheslav Kungurtsev
Electical Engineering Department, KU Leuven
vkungurt@esat.kuleuven.be

Daniel Robinson
Johns Hopkins University
daniel.p.robinson@gmail.com

MS57

Implementation of an Active Set Algorithm for Nonlinear Optimization with Polyhedral Constraints

An active set algorithm is developed for solving an optimization problem whose feasible set is a polyhedron. The gradient projection algorithm is used to identify active constraints, while the conjugate gradient algorithm is used to accelerate convergence. The algorithm exploits a recently developed scheme for projection onto a polyhedron, and makes special adjustments when a loss of local convexity is detected.

James D. Diffenderfer
University of Florida
jdiffen1@ufl.edu

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

MS58

Topology and Geometry of Deep Rectified Network Optimization Landscapes

The loss surface of deep neural networks has recently attracted interest in the optimization and machine learning communities as a prime example of non-convex problem. Some insights were recently gained using spin glass models, but at the expense of strongly simplifying the nonlinear nature of the model. In this work, we do not make any such assumption and study conditions that prevent the existence of bad local minima. We first take a topological approach by studying the connectedness of the level sets. We formalize two important facts: (i) the landscape of deep linear networks has a radically different topology from that of deep half-rectified ones, and (ii) that the landscape for non-linear case is fundamentally controlled by the interplay between the smoothness of the data distribution and model over-parametrization. Our main result is that half-rectified single layer networks are asymptotically connected, and we provide explicit bounds that reveal the aforementioned interplay. The conditioning of GD is the next challenge we address. We study this question through the geometry of the level sets, and we introduce an algorithm to efficiently estimate the regularity of such sets. Our empirical results show that these level sets remain connected throughout all the learning phase, suggesting a nearly convex behavior, but they become exponentially more curvy as the energy level decays, in accordance to what is observed in practice with very low curvature attractors.

Joan Bruna
New York University
Courant Institute
bruna@cims.nyu.edu

Daniel Freeman
UC Berkley
daniel.freeman@berkeley.edu

MS58

Advances in Nonconvex Optimization for Deep Learning

In this talk I will discuss new developments in non-convex optimization for deep neural networks. I will review our work along with new developments showing that neural nets do not significantly suffer from local minima unlike previously thought. In fact, the loss surface of neural nets exhibit a particular structure with saddle points leading the way to near-optimal local minima

Yann Dauphin
Facebook AI Research
yann@dauphin.io

MS58

On the Rate of Convergence for Online Principal Component Estimation and Tensor Decomposition

Principal component analysis (PCA) and tensor component analysis has been a prominent tool for high-dimensional data analysis. Online algorithms that estimate the component by processing streaming data are of tremendous practical and theoretical interests. In this talk, we cast these methods into stochastic nonconvex optimization problems, and we analyze the online algorithms as a stochastic approximation iteration. This talk is divided into two parts. In the first half, we first attempt to understand the dynamics of stochastic gradient method via ordinary and stochastic differential equation approximations. In the second half, we prove for the first time a nearly optimal convergence rate result for both online PCA algorithm and online tensor decomposition. We show that the finite-sample error closely matches the corresponding results of minimax information lower bound.

Chris Junchi Li
Princeton University

junchil@princeton.edu

MS58

When Are Nonconvex Optimization Problems Not Scary?

General nonconvex optimization problems are NP-hard. In applied disciplines, however, nonconvex problems abound, and heuristic algorithms are often surprisingly effective. The ability of nonconvex heuristics to find high-quality solutions for practical problems remains largely mysterious.
In this talk, I will describe a family of nonconvex problems which can be solved efficiently. This family has the characteristic structure that (1) every local minimizer is also global, and (2) the objective function has a negative directional curvature around all saddle points (ridable saddle). Natural nonconvex formulations for a number of important problems in signal processing and machine learning lie in this family, including the eigenvector problem, complete dictionary learning (CDL), generalized phase retrieval (GPR), orthogonal tensor decomposition, and various synchronization problems. This benign geometric structure allows a number of optimization methods to efficiently find a global minimizer, without special initializations. This geometric analysis has helped to produce new computational guarantees for several practical problems, such as CDL and GPR.
To complete and enrich the framework is an ongoing research effort. I will highlight challenges from both theoretical and algorithmic sides. Joint work with Qing Qu and John Wright. See my PhD thesis
http://sunju.org/pub/docs/thesis.pdf
for details.

Ju Sun
Stanford University
sunjunus@gmail.com

Qing Qu
Columbia University
Electrical Engineering
qq2105@columbia.edu

John Wright
Department of Electrical Engineering
Columbia University
johnwright@ee.columbia.edu

MS59

Newton-Type Methods Under Inexact Information

Second-order methods have been studied much less than the first-order ones due to their computational burden at each iteration. In this talk, we present Newton-type methods with cubic regularization for solving (strongly) convex optimization problems using noisy information of the objective function. We provide rates of convergence of these methods and show that they perform better in practice than their exact counterparts and optimal first-order methods.

Saeed Ghadimi
Princeton University
sghadimi@princeton.edu

Tong Zhang
Rutgers
tzhang@stat.rutgers.edu

Han Liu
Princeton University
hanliu@princeton.edu

MS59

On the Efficient Computation of a Generalized Jacobian of the Projector Over the Birkhoff Polytope

In this talk, we derive an explicit formula, as well as an efficient procedure, for constructing a generalized Jacobian for the projector of a given square matrix onto the Birkhoff polytope, i.e., the set of doubly stochastic matrices. To guarantee the high efficiency of our procedure, a semismooth Newton method for solving the dual of the projection problem is proposed and efficiently implemented. Extensive numerical experiments are presented to demonstrate the merits and effectiveness of our method by comparing its performance against other powerful solvers such as the commercial software Gurobi and the academic code PPROJ [Hager and Zhang, 2016]. In particular, our algorithm is able to solve the projection problem with over one billion variables and nonnegative constraints to a very high accuracy in less than 15 minutes on a modest desktop computer. In order to further demonstrate the importance of our procedure, we also propose a highly efficient augmented Lagrangian method (ALM) for solving a class of convex quadratic programming (QP) problems constrained by the Birkhoff polytope. The resulted ALM is demonstrated to be much more efficient than Gurobi in solving a collection of QP problems arising from the relaxation of quadratic assignment problems.

Xudong Li
NUS
matlixu@nus.edu.sg

Defeng Sun
Dept of Mathematics
National University of Singapore
matsundf@nus.edu.sg

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

MS59

Hyperparameter Tuning of Machine Learning Algorithms Using Stochastic Derivative Free Optimization

Abstract not available at time of publication.

Katya Scheinberg
T.J. Watson Research Center, IBM, P.O.Box 218
Yorktown Heights, NY 10598
katyas@watson.ibm.com

MS59

Inexact Proximal Stochastic Gradient Method for Convex Composite Optimization

We study an inexact proximal stochastic gradient (IPSG) method for convex composite optimization, whose objective function is a summation of an average of a large number of smooth convex functions and a convex, but possibly nonsmooth, function. The variance reduction technique is incorporated in the method to reduce the stochastic gradient variance. The main feature of this IPSG algorithm is to allow solving the proximal subproblems inexactly while still keeping the global convergence with desirable complexity bounds. Di?erent accuracy criteria are proposed for solving the subproblem, under which the global convergence and the component gradient complexity bounds are derived for the both cases when the objective function is strongly convex or generally convex. Preliminary numerical experiment shows the overall e?ciency of the IPSG algorithm.

Xiao Wang
University of Chinese Academy of Sciences
wangxiao@ucas.ac.cn

Shuxiong Wang
Chinese Academy of Sciences, China
wsx@lsec.cc.ac.cn

Hongchao Zhang
Department of Mathematics and CCT
Louisiana State University
hozhang@math.lsu.edu

MS60

MILP Relaxations for Instationary Gas Network Optimization Problems

We present an new MINLP model for instationary gas network optimization problems using time-expanded graphs and discretization of time. Furthermore we describe a method solving the obtained MINLP based on constructing MILP relaxations of the underlying MINLP by Piecewise Linear Functions. Both the theoretical background and numerical results are treated.

Robert Burlacu
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Mathematik, EDOM
robert.burlacu@fau.de

MS60

Global Optimization of Ideal Multi-component Distillation Columns

Many separation tasks in chemical engineering are based on distillation. The determination of an optimal design for such separation processes often requires solving a mixed-integer non-convex optimization problem to global optimality, and is, hence, very challenging, in general. From a practical point of view, solving those problems within a reasonable time (if possible) is of particular interest. This makes it necessary to develop problem-specific optimization techniques that exploit the underlying physical and mathematical properties of the considered processes and model formulations. In this talk, a reformulation for ideal multi-component distillation column processes is considered. The reformulation is based on a suitable linear transformation of the concentration variables that yield monotonic concentration profiles. The monotonic behavoir especially allows to design a problem-specific bound-tightening strategy that can be used to solve the underlying mixed-integer non-convex optimization problems. Numerical examples are presented, demonstrating the usefulness of the reformulation and the bound-tightenting strategy.

Dennis Michaels
Technische Universität Dortmund
Fakultät für Mathematik
dennis.michaels@tu-dortmund.de

Nick Mertens
Technische Universität Dortmund Fakultät für Mathematik
nich.mertens@math.tu-dortmund.de

Achim Kienle
Otto-von-Guericke-Universität Magdeburg
Institut für Automatisierungstechnik
achim.kienle@ovgu.de

Christian Kunde
Unaffilated
christian.kunde@ovgu.de

MS60

Complexity of MINLP from Gas Network Optimization

We discuss the complexity of various gas network optimization problems. The main result is that validation of nomination in presence of compressors or valves is weakly NP-complete for series-parallel graphs and strongly NP-complete on general graphs. This is exactly what is predicted by the classical results on dynamic programming algorithms for these problems which are widely applied in practice. We also show how the results can be used to preprocess these problems to make them easier solvable using standard solvers. These results also generalize to potential driven flow problems, e.g., in water networks.

Lars Schewe
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Mathematik, EDOM
lars.schewe@math.uni-erlangen.de

Martin Gro
TU Berlin
gross@math.tu-berlin.de

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

Martin Skutella
TU Berlin
skutella@math.tu-berlin.de

Martin Schmidt
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Mathematik
mar.schmidt@fau.de

MS60

Approximate Convex Decomposition in Gas Network Optimization

In some optimization problems, the feasible region of the constraint set can be described as the union of convex polytopes and thus forms itself a nonconvex polytope. Known approaches to treat such feasible regions involve Disjunctive Programming or the use of binary indicator variables for each of the original convex polytopes. We apply a decomposition method known as Approximate Convex Decomposition to such sets in 2 or 3 dimensions in order to obtain relaxations that can be hierarchically refined when required. Thereby, each hierarchy level yields a decomposition that reduces some measure of concavity until a specified tolerance threshold is attained. We will illustrate our approach with the example of optimizing the operational details of gas compressor stations.

Tom Walther
Zuse Institute Berlin
walther@zib.de

Benjamin Hiller
Zuse Insitut Berlin
hiller@zib.de

MS61

The Case for Competition: Efficient Computation of Equilibria

Equilibrium problems are widespread in the economics and physical literatures, and often involve interactions between competing players or physical systems. These include examples of generalized Nash Equilibria and Multiple Optimization Problems with Equilibrium Constraints. While Extended Mathematical Programming provides a convenient mechanism to describe these interactions, and to process them into models for solution by existing complementarity and optimization methods, they are often limited by an assumption of exclusive control or knowledge. In many cases, shared constraints, linked multipliers, and shared implicit variables are necessary to model behavior within the equilibrium. We describe a framework that allows each of these features within competitive equilibrium problems and show how to exploit the structure in efficient computational schemes. Several large scale applications to electricity planning will be demonstrated that show the power of the framework and the speed of solution.

Michael C. Ferris
University of Wisconsin
Department of Computer Science
ferris@cs.wisc.edu

Youngdae Kim
Computer Science Department
University of Wisconsin Madison
youngdae@cs.wisc.edu

MS61

A Full-Newton Step Interior Point Method for Sufficient LCP

An improved Infeasible Full Newton-step Interior-Point method for $P_*(\kappa)$ Linear Complementarity Problem (LCP) is considered. For most methods of this type each iteration consists of one feasibility step and a few centering steps while in this version of the algorithm each iteration consists of only one feasibility step. This improvement has been achieved by a much tighter estimate of the proximity measure after the feasibility step. Furthermore, although the analysis of the algorithm depends on the constant κ, the algorithm itself does not; thus, the method is suitable for the class of sufficient LCPs. However, the best iteration bound known for these types of methods is still achieved.

Goran Lesaja
Georgia Southern University
Department of Mathematical Sciences
goran@georgiasouthern.edu

Florian A. Potra
University of Maryland Baltimore County
Department of Mathematics and Statistics
potra@math.umbc.edu

MS61

Inexact Interior Point Methods for Complementarity Problems

We consider Primal-dual Inexact Interior point methods for linear complementarity problems. At each iteration, computing the step requires the solution of a linear system, and the computational cost can be reduced by using iterative methods where the accuracy of the solution is related to the quality of the Interior Point iterate. We focus on symmetric indefinite KKT formulations for this system and on the accuracy of the inexact steps for different stopping criteria proposed in the literature. Our aim is to analyze the effect of inexactness in the late stage of the Interior Point method when the linear systems become severely ill-conditioned.

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

Valeria Simoncini
Universita' di Bologna
valeria.simoncini@unibo.it

MS61

Weighted Complementarity Problems and Applications

The weighted complementarity problem (wCP) is a new paradigm in applied mathematics that provides a unifying framework for analyzing and solving a variety of equilibrium problems in economics, multibody dynamics, atmospheric chemistry and other areas in science and technology. It represents a far reaching generalization of the notion of a complementarity problem (CP). Since many of the very powerful CP solvers developed over the past two decades can be extended to wCP, formulating an equilibrium problem as a wCP opens the possibility of devising highly efficient algorithms for its numerical solution. For example, Fisher's competitive market equilibrium model can be formulated as a wCP, while the Arrow-Debreu competitive market equilibrium problem (due to Nobel prize laureates Kenneth Joseph Arrow and Gerard Debreu) can be formulated as a self-dual wCP. The talk will show how equilibrium problems can be formulated as wCP and will present an interior point method that can efficiently solve such problems. Several properties of wCP will be analyzed.

Florian A. Potra
University of Maryland Baltimore County
Department of Mathematics and Statistics
potra@math.umbc.edu

MS62

Euclidean Distance Matrix Optimization Model for Ordinal Embedding

When the coordinates of a set of points are known, the pairwise Euclidean distances among the points can be easily computed. Conversely, if the Euclidean distance matrix is given, a set of coordinates for those points can be computed through the well known classical Multi-Dimensional Scaling (MDS). In this paper, we consider the case where some of the distances are far from being accurate (containing large noises or even missing). In such a situation, the order of the known distances (i.e., some distances are larger than others) is valuable information that often yields far more accurate construction of the points than just using the magnitude of the known distances. The methods making use of the order information is collectively known as non-metric MDS. A challenging computational issue among all existing nonmetric MDS methods is that there are often a large number of ordinal constraints. In this paper, we cast this problem as a matrix optimization problem with ordinal constraints. We then adapt an existing smoothing Newton method to our matrix problem. Extensive numerical results demonstrate the efficiency of the algorithm, which can potentially handle a very large number of ordinal constraints.

Qingna Li
School of Mathematics and Statistics
Beijing Institute of Technology
qnl@bit.edu.cn

Houduo Qi
University of Southampton
h.qi@soton.ac.uk

MS62

Sparse Recovery via Partial Regularization

In the context of sparse recovery, it is known that most of existing regularizers such as L1 suffer from some bias incurred by some leading entries (in magnitude) of the associated vector. We propose a class of models with partial regularizers to neutralize this bias for recovering sparse vectors. We show that every local minimizer of these models is sufficiently sparse or the magnitude of all its nonzero entries is above a uniform constant depending only on the data of the linear system. Moreover, for a class of partial regularizers, any global minimizer of these models is a sparsest solution to the linear system. We also establish some sufficient conditions for local or global recovery of the sparsest solution to the linear system, among which one of the conditions is weaker than the best known restricted isometry property (RIP) condition for sparse recovery by L1. In addition, a first-order feasible augmented Lagrangian (FAL) method is proposed for solving these models, in which each subproblem is solved by a nonmonotone proximal gradient (NPG) method. The global convergence of this method is also established. Numerical results on compressed sensing and sparse logistic regression demonstrate that the proposed models substantially outperform the widely used ones in the literature in terms of solution quality.

Zhaosong Lu
Department of Mathematics
Simon Fraser University
zhaosong@sfu.ca

Xiaorui Li
Simon Fraser University
xla97@sfu.ca

MS62

The Sparsest Solutions to Z-Tensor Complementarity Problems

Finding the sparsest solutions to a tensor complementarity problem is generally NP-hard due to the non-convexity and non-continuity of the involved ell_0 norm. In this talk, a special type of tensor complementarity problems with Z-tensors has been considered. Under some mild conditions, we show that to pursuit the sparsest solutions is equivalent to solving polynomial programming with a linear objective function. The involved conditions guarantee the desired exact relaxation and also allow to achieve a global optimal solution to the relaxed nonconvex polynomial programming problem. Particularly, in comparison to existing exact relaxation conditions, such as RIP-type ones, our proposed conditions are easy to verify.

Ziyan Luo
State key laboratory of rail traffic control and safety, Be
starkeynature@hotmail.com

MS62

A Generalized Alternating Direction Method of Multipliers with Semi-Proximal Terms for Convex Composite Conic Programming

In this talk, we propose a generalized alternating direction method of multipliers (ADMM) with semi-proximal terms for solving a class of convex composite conic optimization problems, of which some are high-dimensional, to moderate accuracy. Our primary motivation is that this method, together with properly chosen semi-proximal terms, such as those generated by the recent advance of symmetric Gauss-Seidel technique, is applicable to tackling these problems. Moreover, the proposed method, which relaxes both the primal and the dual variables in a natural way with one relaxation factor in the interval $(0, 2)$, has the potential of enhancing the performance of the classic ADMM. Extensive numerical experiments on various doubly non-negative semidefinite programming problems, with or without inequality constraints, are conducted. The corresponding results showed that all these multi-block problems can be successively solved, and the advantage of using the relaxation step was apparently observed.

Yunhai Xiao
College of Mathematics and Statistics, Henan University
yhxiao@henu.edu.cn

Liang Chen
Hunan University
chl@hnu.edu.cn

Donghui Li
South China Normal University
dhli@scnu.edu.cn

MS63

Risky Robust Optimization in Machine Learning and Support Vector Machines

Robust Optimization has proven to be an intuitive and fruitful framework for optimization under uncertainty. Typically, Robust Optimization approaches uncertainty from a pessimistic, or worst-case, viewpoint and can be overly conservative. We diverge from this tradition, and show that an optimistic, or best-case, viewpoint of uncertainty should not be ignored. We show that this viewpoint can yield Robust Optimization problems with advantageous properties, particularly w.r.t. sparsity of the resulting decision vector. We focus our analysis on Robust Support Vector Machines, and show that this viewpoint helps to unify convex and non-convex formulations, helping to also explain the success of non-convex regularization for inducing sparsity in hyperplane selection.

Matthew Norton
University of Florida
mdnorton@ufl.edu

MS63

The Application of Stochastic Optimization in Chemical Engineering Process

In this talk, we discuss how to model a few important chemical engineering processes using multi-stage stochastic programming. We then present stochastic first-order algorithms and establish their rate of convergence for solving these types of problems. Some preliminary numerical results will also be presented to demonstrate the effectiveness of these algorithms.

Zhaohui Tong, Hanxi Bao
University of Florida
ztong@ufl.edu, baohanxi@ufl.edu

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

MS63

Statistical Inversion and Risk Averse Control for Additive Manufacturing Processes

We demonstrate an optimization process in which statistical inversion methods determine dynamical-model coefficients with corresponding statistical characterizations, which are in turn mapped to a risk-averse optimization formulation to solve for a control strategy. Appropriate risk measures are selected to achieve an a priori determined performance strategy. This approach is demonstrated on a direct write additive manufacturing process in which the goal is to infer model parameters and control several stages of the process to achieve material property targets. Direct write consists of slurry preparation, printing and sintering to develop ceramic components and therefore the challenge is to not only solve the coupled inversion-control problem but to solve this coupled problem for multiple stages. Furthermore, uncertainties need to be propagated throughout these stages. PDE-constrained optimization methods serve as the foundation for this work with finite element discretizations, adjoint-based sensitivities, trust-region methods, and Newton-Krylov solvers. Our final AM produced parts must achieve tight tolerances for a range of different material properties. Accordingly, a range of risk measures are considered with a specific emphasis on reliability. A numerical example demonstrates that the use of risk measures as part of the objective function results in optimal solutions.

Bart G. Van Bloemen Waanders
Sandia National Laboratories
bartv@sandia.gov

MS64

Nonconvex Low-Rank Estimation: Robustness and Linear Convergence

Many problems in statistics and machine learning involve fitting a low-rank matrix to noisy data. The resulting rank-constrained optimization problem is often solved using semidefinite relaxation, which enjoys strong statistical guarantees but does not scale well to large-scale datasets. Recently, nonconvex approaches — such as gradient descent over the low-rank space — have emerged as a computationally efficient alternative for low-rank estimation problems. We develop a unified framework characterizing the convergence behaviors of this nonconvex approach and the statistical errors of the resulting fixed points. Our results provide insights on why it is expected to work in general, and provide linear convergence rates. When specialized to specific problems, our general theory yields concrete guarantees for matrix completion, robust PCA, community detection, sparse PCA, matrix sensing, 1-bit matrix completion, and others. For these problems nonconvex methods enjoy near-linear running time, have statistical performance matching (sometimes beating) the best known results achieved by SDP relaxation, and are robust to gross corruption in data. Moreover, our framework applies to projected gradient descent for constrained formulations, and does not require unrealistic assumptions such as sample splitting.

Yudong Chen
University of California, Berkeley
Department of Electrical Engineering and Computer Sciences
yudong.chen@berkeley.edu

MS64

Faster Projection-Free Convex Optimization with Structured Matrices

Many important problems in machine learning and signal processing can be directly formulated as minimization of a non-convex function over a certain set of matrices. Fortunately, many of these problems admit natural and well-studied convex relaxations with strong theoretical guarantees. Prime examples include the problems of Low-rank Matrix Completion and Multi-class Classification. On the downside, algorithms that are applicable for solving these convex relaxations either converge fast, but require prohibitive SVD computations on each iteration (i.e., orthogonal projection), or rely only on cheap single eigenvector computations, but converge with an inferior rate. In this talk I will present a new variant of the classical Frank-Wolfe method that relies only on cheap eigenvector computations. I will show that the new method enjoys (roughly) a quadratic improvement in convergence rate over the standard method, under an additional strong-convexity assumption.

Dan Garber
Technion
dangar@tx.technion.ac.il

MS64

Dropping Convexity for Faster Semidefinite Optimization

Abstract not available at time of publication.

Sujay Sanghavi
Electrical and Computer Enigneering
University of Texas at Austin
sanghavi@mail.utexas.edu

MS64

Nonconvex Optimization and Nonlinear Models for

Matrix Completion

Most recent results in matrix completion assume that the matrix under consideration is low-rank. If the matrix is $p \times n$, this model is equivalent to assuming we observe n points in a p-dimensional ambient space that lie along a low-dimensional hyperplane. In real-world settings, however, observations are more accurately modeled as lying along a low-dimensional nonlinear submanifold. In this talk, I will describe two approaches to matrix completion in the presence of such nonlinearities. The first approach models the submanifold as a monotonic transformation of a hyperplane – that is, each element of the matrix is a monotonic function of the corresponding entry of an underlying low-rank matrix. I will describe a novel matrix completion method that alternates between low-rank matrix estimation and monotonic function estimation to estimate the missing matrix elements. The second approach uses a piecewise linear approximation to the submanifold by modeling the n points as lying in a union of low-rank subspaces. I will describe a novel method for subspace clustering with missing data and matrix completion which is based on group-sparsity and alternating minimization.

Rebecca Willett
University of Wisconsin-Madison
willett@discovery.wisc.edu

MS65

Stochastic Variance Reduction Methods for Policy Evaluation

Policy evaluation is a crucial step in many reinforcement-learning procedures, which estimates a value function that predicts states' long-term value under a given policy. In this work, we focus on policy evaluation with linear function approximation over a fixed dataset. We first transform the empirical policy evaluation problem into a (quadratic) convex-concave saddle point problem, and then present a primal-dual batch gradient method, as well as two stochastic variance reduction methods for solving the problem. These algorithms scale linearly in both sample size and feature dimension. Moreover, they achieve linear convergence even when the saddle-point problem has only strong concavity in the dual variables but no strong convexity in the primal variables. Numerical experiments on benchmark problems demonstrate the effectiveness of our methods.

Simon Du
Machine Learning Department, Carnegie Mellon University
ssdu@cs.cmu.edu

Jianshu Chen, Lihong Li, Lin Xiao
Microsoft Research
jianshuc@microsoft.com, lihongli.cs@gmail.com, lin.xiao@microsoft.com

Dengyong Zhou
Microsoft Research, Redmond
dengyong.zhou@microsoft.com

MS65

Primal-Dual Reinforcement Learning

We consider the online estimation of the optimal policy estimation of Markov decision processes. We propose a Stochastic Primal-Dual (SPD) method which exploits the inherent minimax duality of Bellman equations. SPD updates a few coordinates of the value and policy estimates as state transitions are sampled. We show that the SPD has superior space and computational complexity and it finds with high probability an ϵ-optimal policy using near optimal samples and iterations.

Yichen Chen
Department of Computer Science
Princeton University
yichenc@princeton.edu

Mengdi Wang
Department of Operations Research and Financial Engineering
Princeton University
mengdiw@princeton.edu

MS65

A Generalized Reduced Linear Program for Markov Decision Processes

The efficient computation of near-optimal policies in Markov Decision Processes (MDPs) is of major interest in various scientific and engineering applications. One approach of major interest is to combine linear function approximation with linear programming, leading to what is known as "Approximate Linear Programming" (ALP). While ALP allows for a compact representation of value functions, the number of constraints in the standard ALP formulation is still intractable. One way to overcome this is to reduce the number of constraints. We provide a new analysis of a generalized version of the resulting reduced ALP that complements previous results. In particular, as opposed to previous results, the new analysis allows us to derive a policy error bound that is applicable regardless of how the constraints are selected and shows a graceful degradation that connects features, the optimal value function and the constraints selected in an easy-to-interpret fashion, suggesting specific ways of reducing the constraints, while avoiding the knowledge of the stationary distribution of the optimal policy.

Chandrashekar Lakshmi Narayanan
Department of Computing Science
University of Alberta
chandrurec5@gmail.com

Shalabh Bhatnagar
Department of Computer Science and Automation
Indian Institute of Science
shalabh.bhatnagar2012@gmail.com

Csaba Szepesvari
Department of Computing Science
University of Alberta
szepesva@ualberta.ca

MS65

Robust Optimization for Data-Limited Reinforcement Learning

One of the goals of robust reinforcement learning is to compute good policies from limited data with high confidence. We develop and analyze new practical model-based approaches for limited-data reinforcement learning based on robust Markov decision processes (RMDP). Standard RMDP methods are based on formulations with rectangular uncertainty sets. We show that rectangular models are not appropriate in common data-limited settings and propose new non-rectangular formulations. Most of these formulations are, unfortunately, NP-hard. We develop approximate approximate algorithms and empirically demonstrate their advantages over regular methods in several domains, including energy arbitrage in smart grids.

Marek Petrik
Department of Computer Science
University of New Hampshire
mpetrik@cs.unh.edu

MS66

Smoothing for Improved Worst-Case Competitive Ratio in Online Optimization

In Online Optimization, the data in an optimization problem is revealed over time, and at each step a decision variable needs to be set without knowing the future data. This setup covers online resource allocation, from classical inventory problems to the 'Adwords' problem popular in online advertising. We discuss several algorithms for a class of online convex optimization problems. Our focus is on the algorithms' competitive ratio, i.e., the ratio of the objective achieved by the algorithm to that of the optimal offline sequence of decisions, for worst-case input data. We discuss bounds on this ratio for a primal-dual greedy algorithm, show how smoothing the objective can improve this bound, and for separable functions, how to seek the optimal smoothing by solving a convex optimization problem. This approach allows us to design effective smoothing, customized for a given cost function.

Maryam Fazel
University of Washington
Electrical Engineering
mfazel@ee.washington.edu

MS66

Automating the Analysis and Design of Large-Scale Optimization Algorithms

First-order iterative algorithms such as gradient descent, fast/accelerated methods, and operator-splitting methods such as ADMM can be viewed as discrete-time dynamical systems. We will show that if the function being optimized is strongly convex, for example, computing the worst-case performance of a particular algorithm is equivalent to solving a robust control problem. This amounts to establishing feasibility of a small semidefinite program whose size is independent of the dimension of the function's domain. Our unified approach allows for the efficient and automatic evaluation of worst-case performance bounds for a wide variety of popular algorithms. The bounds derived in this manner either match or improve upon the best known bounds from the literature. Finally, our framework can be used to search for algorithms that meet desired performance specifications, thus establishing a new and principled methodology for designing new algorithms.

Laurent Lessard
University of Wisconsin-Madison
laurent.lessard@wisc.edu

MS66

Randomization Vs. Acceleration

Abstract not available at time of publication.

Pablo A. Parrilo
Massachusetts Institute of Technology
parrilo@MIT.EDU

MS66

Low-Recurrence Width Polynomial Families in Linear Algebra and Optimization

Abstract not available at time of publication.

Christopher Re
Department of Computer Science
Stanford University
chrismre@cs.stanford.edu

MS67

Application of SOCP in Power Efficient Networks

In this paper, a mixed-integer second-order cone programming (MISOCP) model based on robust optimization is proposed to achieve power saving in the networks. Most of the researches in the literature to minimize the power in the networks are based on exact traffic demands, but in practice, estimating traffic demands exactly is a difficult task, because traffic demands fluctuate in general. In our work, taking account of some fluctuations in the estimated traffic demands, we apply the robust optimization technique to obtain an MISOCP. The objective is to minimize the power consumption in the networks by deactivating some unnecessary links. Numerical experiments are presented to compare our model with other models in the literature.

Bimal C. Das, Eiji Oki, Masakazu Muramatsu
The University of Electro-Communications
bcdasdiu@gmail.com, eiji.oki@uec.ac.jp, masakazumuramatsu@uec.ac.jp

MS67

A Lifted-Polyhedral-Programming Approach for Optimal Contribution Problems

An important issue in tree breeding is to improve tree performance genetically in seed orchard without losing genetic diversity. The optimal contribution selection is classified into two types; unequal and equal problems. While the proportion of genes in unequal problems are not necessarily same, the proportion in equal problem should be equal to 0 or $\frac{1}{N}$ for a certain integer N. [Pong-Wong et.al, 2007] introduced semidefinite programming (SDP) to obtain an accurate value, but it demands long computation time. [Yamashita et al., 2015] then proposed a second-order cone programming (SOCP) approach to reduce computation time. Since the methods are employed to solve the unequal contribution selection, we should develop new methods for the equal contribution selection problem. Extending the notion of the SOCP approach, the equal problem is much harder than the unequal problem. Extending the notion of the SOCP approach, the equal contribution selection can be described with mixed-integer SOCP (MISOCP) problems which involve both second-order cone constraints and integer constraints. In order to solve the equal selection, we develop a relaxation approach using Lifted Polyhedral Programming proposed by Ben-Tal and Nemirovski. Numerical results show that this approach is effective compared with existing methods.

Sena Safarina
Tokyo Institute of Technology
sena4@is.titech.ac.jp

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

MS67

Constrained Penalized Spline Estimation by Second-Order Cone Programming

We consider regression by B-splines with a penalty on high-order finite differences of the coefficients of adjacent B-splines. The penalty prevents overfitting. The underlying function is assumed to be nonnegative. The model is casted as a second-order cone programming problem, which can be solved efficiently by modern optimization techniques. The method is implemented in MATLAB.

Yu Xia
Lakehead University
yxia@lakeheadu.ca

Farid Alizadeh
RUTCOR and School of Business,
Rutgers University
alizadeh@rutcor.rutgers.edu

MS67

A Steep-Ascend Method for MI-SOCP Arising from Tree Breeding

An optimal contribution problem arising from tree breeding is an optimization problem to determine the contribution of candidate genotypes for seed orchards. A diversity constraint to improve a long-term performance of the seed orchards is a second-order cone constraint, and each contribution should be an integer as an equally deployment. Therefore, the optimal contribution problem can be described as a mixed-integer SOCP problem. In this talk, we discuss LP, SOCP and SDP relaxations of the optimal contribution problem. In particular, we build an efficient SOCP relaxation by exploiting the sparsity and the structure of the diversity constraint. We also consider the tightness of the SDP relaxation. Furthermore, we propose a steep-ascent method to generate a feasible solution from the three relaxation. We move the second-order constraint to the objective function as a penalty term and implement a heuristic method to search a feasible solution with a high long-term performance. Numerical results including data of Scot Pine show that the steep-ascent method combined with the SOCP relaxation obtained a favorable solution in a short time.

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

Tim Mullin
The Swedish Forestry Research Institute (Skogforsk)
tim.mullin@skogforsk.se

Sena Safarina
Tokyo Institute of Technology
sena4@is.titech.ac.jp

MS68

Conic Infimum and Supremum and Some Applications

Finding the minimum or the maximum of a set of numbers is probably the most widely used basic operation in optimization algorithms. We study the extension of these operations to partial orders induced by proper cones. Specifically, we define the notion of infimum and supremum of a set of points in Euclidean space with respect to the partial order induced by a proper cone. We examine conditions under which these concepts are well-defined and present infimum and supremum induced by concrete examples such as nonnegative polynomial, second order, semidefinite cone and polyhedral cones. Some applications of conic infimum and supremum, particularly in generalization of network flow problems will be discussed.

Farid Alizadeh
RUTCOR and School of Business,
Rutgers University
alizadeh@rutcor.rutgers.edu

MS68

On Positive Duality Gaps in Semidefinite Programming

In semidefinite programming (SDP) one often encounters positive duality gaps, i.e., the situation when the primal and dual optimal values differ. We present a systematic study of positive duality gaps, and generate a library of SDPs with positive duality gaps. Our instances turn out to be extremely challenging for both commercial and research solvers.

Gabor Pataki
University of North Carolina, at Chapel Hill
gabor@unc.edu

MS68

Dimension Reduction for SDP

We propose a new method for simplifying semidefinite programs (SDP) inspired by symmetry reduction. Specifically, we show if an orthogonal projection satisfies certain invariance conditions, restricting to its range yields an equivalent primal-dual pair over a lower-dimensional symmetric cone—namely, the cone-of-squares of a Jordan subalgebra of symmetric matrices. We then give a simple algorithm for minimizing the rank of this projection and hence the dimension of this cone. Through the theory of Jordan algebras, the proposed method easily extends to linear programming, second-order cone programming, and, more generally, symmetric cone optimization

Frank Permenter, Pablo A. Parrilo
Massachusetts Institute of Technology
fperment@mit.edu, parrilo@MIT.EDU

MS68

On Facial Structure of Convex Sets

Whilst faces of a polytope form a well structured lattice, in which faces of each possible dimension are present, the latter property is not true for general compact convex sets. We address the question of which dimensional patterns are possible for the faces of general closed convex sets. We show that for any finite sequence of positive integers there exist compact convex sets which only have extreme points and faces with dimensions from this prescribed sequence.

We also discuss another approach to dimensionality, considering the dimension of the union of all faces of the same dimension. We show that the questions arising from this approach are highly nontrivial and give examples of convex sets for which the sets of extreme points have fractal dimension.

Vera Roshchina
Federation University Australia
vera.roshchina@gmail.com

MS69

Exploiting Active Subspaces in Nonconvex Optimization of Physics-Based Models

Given a function of several variables, the active subspace is the span of a set of directions in the function's domain along which variable perturbations change the function's value the most, on average. I will discuss how to exploit an nonconvex objective function's active subspace to accelerate minimization.

Paul Constantine
Colorado School of Mines
paul.constantine@mines.edu

MS69

Deep Learning Meets Differential Equations

Abstract not available at time of publication.

Eldad Haber
Department of Mathematics
The University of British Columbia
haber@math.ubc.ca

MS69

Fast Solvers for Optimization Problems Constrained by PDEs with Uncertain Inputs

Optimization problems constrained by deterministic stationary partial differential equations (PDEs) are computationally challenging. This is even more so if the constraints are deterministic unsteady PDEs since one would then need to solve a system of PDEs coupled globally in time and space, and time-stepping methods quickly reach their limitations due to the enormous demand for storage. Yet, more challenging than the afore-mentioned are problems constrained by unsteady PDEs involving (countably many) parametric or uncertain inputs. This class of problems often leads to prohibitively high dimensional saddle-point system with tensor product structure, especially when discretized with the stochastic Galerkin finite element method (SGFEM). Moreover, a typical model for an optimal control problem with stochastic inputs (SOCP) will usually be used for the quantification of the statistics of the system response – a task that could in turn result in additional enormous computational expense. In this talk, we consider two prototypical model SOCPs and discretize them with SGFEM. We derive and analyze tensor-based, low-rank iterative solvers for the resulting stochastic Galerkin systems. The developed solvers are quite efficient in the reduction of temporal and storage requirements of the high-dimensional linear systems. Finally, we illustrate the effectiveness of our solvers with numerical experiments.

Akwum Onwunta
Max Planck Institute, Magdeburg, Germany
onwunta@mpi-magdeburg.mpg.de

Peter Benner
Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
benner@mpi-magdeburg.mpg.de

Sergey Dolgov
University of Bath
Department of Mathematical Sciences
s.dolgov@bath.ac.uk

Martin Stoll
Max Planck Institute, Magdeburg
stollm@mpi-magdeburg.mpg.de

MS69

Semidefinite Programming Relaxations and Multigrid Methods in Optimal Control

We discuss recent developments for solving Semidefinite Programming relaxations of optimal control problems using multigrid techniques. Multigrid methods are known to lead to efficient algorithms for optimal control problems. However, the SDP relaxation of the model does not maintain the geometric information present in the original model. Despite this issue, we discuss theoretical and algorithmic developments that allow us to take advantage of the inherent structure of optimal control problems.

Panos Parpas
Imperial College London
panos.parpas@imperial.ac.uk

MS70

Stochastic Variance Reduction Methods for Saddle-Point Problems

We consider convex-concave saddle-point problems where the objective functions may be split in many components, and extend recent stochastic variance reduction methods to provide the first large-scale linearly convergent algorithms for this class of problems which is common in machine learning. While the algorithmic extension is straightforward, it comes with challenges and opportunities: (a) the convex minimization analysis does not apply and we use the notion of monotone operators to prove convergence, showing in particular that the same algorithm applies to a larger class of problems, such as variational inequalities, (b) the split does need to be done with convex-concave terms.

Francis Bach
INRIA - École Normale Supérieure
francis.bach@inria.fr

Balamurugan Palaniappan
SIERRA Project Team, INRIA, Paris, France.
balamurugan.palaniappan@inria.fr

MS70

On a Scaled ϵ-Subgradient Method with Adaptive Stepsize Rule

A wide range of image reconstruction problems [Bertero et al, 2009] requires that the functional to be minimized has a non differentiable term. A general formulation for such problems is

$$\min_{x \in R^n} f(x) + \Phi(x)$$

where $f, \Phi : R^n \to R \cup \{\infty\}$ are convex, proper, lsc functions and dom$\Phi \subset$ dom f. In this work we generalize the Forward-Backward (FB) scheme, considering in the forward step the ϵ–subgradient of f and in the backward step a variable metric. The proposed scheme can be written as

$$x^{k+1} = \mathrm{prox}_{\alpha_k \Phi, D_k^{-1}} \left(x^k - \epsilon_k D_k u^k \right)$$

where D_k is a symmetric pos def matrix with bounded eigenvalues, $u_k \in \partial_{\epsilon_k} f(u^k)$ for $\epsilon_k \geq 0$ and α_k is chosen with an adaptive rule (another option is to set it via a priori selected divergent, square summable sequence). Beside the theoretical generalization, the proposed method can be numerically efficient with suitable choices of D_k and α_k. We describe a concrete example [Bonettini et al, 2016] arising in image deblurring from Poisson data, where $f(x) = KL(Hx + b; g) + \beta TV(x)$, with KL the generalized Kullback–Leibler functional, $g \in R^n$ the detected image, corrupted by Poisson noise, $H \in R^{n^2}$ the blurring operator, $b \in R^+$ a background term; TV is the Total Variation functional, $\beta > 0$ the regularization parameter and Φ the indicator function of the nonnegative orthant.

Silvia Bonettini
Dipartimento di Matematica e Informatica
Universit{à} di Ferrara
silvia.bonettini@unife.it

Alessandro Benfenati
Université Paris-Est Marne-la-Vallée
Laboratoire d'Informatique Gaspard Monge (LIGM)
alessandro.benfenati@esiee.fr

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

MS70

Duality Based Iterative Regularization Techniques for Inverse Problems

Many applied problems in science and engineering can be modelled as noisy inverse problems. The high dimensional size of such problems requires the development of ever more efficient numerical procedures to solve them. In this talk, we will focus on iterative regularization techniques, where the number of iterations controls both the computational complexity of the method, and its regularization properties. Key in our approach will be the use of duality techniques. Numerical results showing state of the art performance will be also discussed.

Guillaume Garrigos
Istituto Italiano di Tecnologia
guillaume.garrigos@iit.it

Lorenzo Rosasco
Instituto Italiano di Technologia and
Massachusetts Institute of Technology
lrosasco@mit.edu

Silvia Villa
Dipartimento di Matematica
Politecnico di Milano
silvia.villa@polimi.it

MS70

Stochastic Forward-Douglas-Rachford Splitting for Monotone Inclusions

We propose a stochastic Forward-Douglas-Rachford Splitting framework for finding a zero point of the sum of three maximally monotone operators in real separable Hilbert space, where one of the operators is cocoercive. We characterize the rate of convergence in expectation in the case of strongly monotone operators. We provide guidance on step-size sequences that achieve this rate, even if the strong convexity parameter is unknown. Finally, we present numerical examples from convex optimization with multiple constraints and/or regularizers that support the effectiveness of our method.

Volkan Cevher, Bang Cong Vu, Alp Yurtsever
EPFL
volkan.cevher@epfl.ch, bang.vu@epfl.ch,
alp.yurtsever@epfl.ch

MS71

Estimating Clarke Subgradients of Non-Regular Integrands by Smoothing

Abstract not available at time of publication.

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

MS71

A New Class of Matrix Support Functionals with Applications

Abstract not available at time of publication.

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

MS71

Convergence of a Scholtes-Type Relaxation Method for Optimization Problems with Cardinality Constraints

Optimization problems with cardinality constraints have applications for example in portfolio optimization. However, due to the discrete-valued cardinality constraint, they are not easy to solve. We provide a continuous reformulation of cardinality constraints and discuss the convergence of a Scholtes-type relaxation method for the resulting nonlinear programs with orthogonality constraints. Furthermore, we show preliminary numerical results for portfolio optimization problems with different risk measures.

Alexandra Schwartz
TU Darmstadt
Graduate School CE
schwartz@gsc.tu-darmstadt.de

Michal Cervinka, Martin Branda
Czech Academy of Sciences, Charles University Prague
cervinka@utia.cas.cz, branda@karlin.mff.cuni.cz

Max Bucher
TU Darmstadt

bucher@gsc.tu-darmstadt.de

MS72

A Gaussian Process Trust-Region Method for Derivative-Free Nonlinear Constrained Stochastic Optimization

We present the algorithm SNOWPAC for derivative-free constrained stochastic optimization. The algorithm builds on a model-based approach for deterministic nonlinear constrained DFO that introduces an "inner boundary path" to locally convexify the feasible domain and ensure feasible trial steps. We extend this deterministic method via a generalized trust region approach that accounts for noisy evaluations of the objective and constraints. To reduce the impact of noise, we fit consistent Gaussian process models to past evaluations. Our approach incorporates a wide variety of probabilistic risk or deviation measures in both the objective and the constraints. We demonstrate the efficiency of the algorithm using several numerical benchmarks and comparisons.

Youssef M. Marzouk, Florian Augustin
Massachusetts Institute of Technology
ymarz@mit.edu, fmaugust@mit.edu

MS72

DFO/STORM Approaches to Machine Learning Settings

In previous work, we had proposed an algorithmic framework STORM (STochastic Optimization using Random Models), an adaptation of trust-region methods, to solve unconstrained stochastic optimization problems and almost sure convergence was proven under fairly general conditions on the stochasticity. In particular, we assumed that on a given iteration, fully-linear models of an unknown ground truth function could only be constructed with some fixed probability (bounded away from 1), and likewise estimates of a current iterate and trial iterate could only be computed with error in $\mathcal{O}(\delta_k^2)$ with some fixed probability (bounded away from 1), where δ_k denotes a trust region radius parameter. In this talk, we will discuss how this can be applied to various settings in machine learning. In particular, we will consider a trust-region algorithm that uses batch stochastic gradients (and Hessians) for model-building and inexact zeroth-order function evaluations of common loss functions for classification tasks. We will also consider an algorithm within the STORM framework that uses only inexact zeroth-order information for the minimization of a 01-loss function.

Matt Menickelly
Lehigh University
mjm412@lehigh.edu

MS72

Some Convergence Rate Results for Derivative-Based and Derivative-Free Stochastic Optimization

We consider unconstrained optimization problems where only "stochastic" estimates of the objective function are observable as replicates from a Monte Carlo oracle. We present ASTRO and ASTRO-DF — derivative-based and derivative-free implementations of a class of trust-region algorithms, where a stochastic local interpolation model is constructed, optimized, and updated iteratively. Function estimation and model construction within ASTRO and ASTRO-DF are adaptive in the sense that the extent of Monte Carlo sampling is determined by continuously monitoring and balancing metrics of sampling error (or variance) and structural error (or model bias). ASTRO and ASTRO-DF converge with probability one; more interestingly, the asymptotic relationship between sampling and the trust region radii is characterizable in each case, shedding light on algorithm efficiency.

Raghu Pasupathy
ISE Department
Virginia Tech
pasupath@vt.edu

Sara Shashaani
University of Michigan
sara.shashaani@gmail.com

MS72

Direct Search based on Probabilistic Feasible Descent for Bound and Linearly Constrained Problems

Direct-search methods form a popular class of derivative-free optimization algorithms, that explore the objective function along suitably chosen sets of directions. Under the presence of constraints, such sets must conform to the geometry of the feasible set and a number of deterministic ways of generating them have been proposed in the literature. However, recent new developments in the unconstrained setting have shown that generating the polling directions randomly, ensuring a form of probabilistic descent, can lead to significant gains in efficiency, and still deliver global convergence at appropriated rates with reassuring probabilities. In this talk, we describe a direct-search scheme for solving linearly constrained optimization problems, based on randomly generating directions that guarantee probabilistic feasible descent, in a generalization from the unconstrained case. By applying martingale-type arguments to assess the quality of the directions used throughout the algorithms, we are able to prove global convergence with probability one as well as convergence rates with overwhelming probability. Such rates or worst complexity bounds indicate, as in the unconstrained case, possible interesting gains over the deterministic counterparts. Numerical experiments comparing our approaches to existing deterministic direct-search solvers seem to also support such prospects.

Serge Gratton
ENSEEIHT, Toulouse, France
serge.gratton@enseeiht.fr

Clément W. Royer
Wisconsin Institute for Discovery
University of Wisconsin-Madison
croyer2@wisc.edu

Luis N. Vicente, Zaikun Zhang
University of Coimbra
lnv@mat.uc.pt, zaikun.zhang@polyu.edu.hk

MS73

Stochastic Nonlinear Programming for Wind Turbine Control

We present a stochastic programming formulation to optimize wind turbine pitch and torque controllers to maximize extracted power from uncertain wind profiles and while satisfying a probabilistic constraint on the maximum long-term mechanical stress experienced by the turbine. The probabilistic constraint matches stress profiles to a Gumbel distribution to predict long-term failure probabilities.

Yankai Cao
University of Wisconsin
cao27@wisc.edu

Fernando D'Amato
General Electric Company
damato@ge.com

Victor Zavala
University of Wisconsin-Madison
victor.zavala@wisc.edu

MS73

Logical Benders Decomposition for Quadratic Programs with Complementarity Constraints and Binary Variables

We study a logical Benders decomposition approach to solving binary-constrained quadratic programs with linear complementarity constraints. It is based on a satisfiability master problem to which feasibility cuts are added that are computed from primal and dual subproblems for chosen complementarity pieces and binary variables. Interpreting the logical Benders decomposition approach from a branch-and-bound point of view, we propose several new methods for strengthening the feasibility cuts and guiding the master problem solution. Their efficiency is assessed by numerical experiments.

Francisco Jara-Moroni
Northwestern University
franciscojaramoroni2013@u.northwestern.edu

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

Jong-Shi Pang
University of Southern California
jongship@usc.edu

John E. Mitchell
Rensselaer Polytechnic Institute
110, 8th Street, Troy, NY, 12180
mitchj@rpi.edu

MS73

New Solution Approaches for the Maximum Reliability Stochastic Network Interdiction Problem

We investigate new methods to solve the maximum-reliability stochastic network interdiction problem (SNIP). To begin, we introduce a new extensive formulation that is significantly more compact than the previously introduced extensive form. We then propose a new path-based formulation, which can be solved by delayed constraint generation of paths and cuts derived from submodularity. We apply results from a recent paper by Ahmed and Atamturk for a structured submodular set to obtain stronger valid inequalities for this formulation. These cuts are then embedded within a branch-and-cut (BC) algorithm. Computational results demonstrate that the proposed compact formulation and new BC algorithm are significantly more efficient than existing methods on a set of test instances from the literature.

Eli Towle
University of Wisconsin-Madison
etowle@wisc.edu

James Luedtke
University of Wisconsin-Madison
Department of Industrial and Systems Engineering
jrluedt1@wisc.edu

MS73

Solving chance-constrained problems using a kernel VaR estimator

We present a reformulation of nonlinear optimization problems with chance constraints that replaces a probabilistic constraint by a nonparametric value-at-risk estimator. The estimator smooths the historical VaR via a kernel, resulting in a better approximation of the quantile and reducing the nonconvexity of the feasible region. An optimization algorithm is proposed that permits the efficient treatment of joint chance constraints. Theoretical and empirical convergence properties are presented.

Andreas Waechter, Alejandra Pena-Ordieres
Northwestern University
IEMS Department
waechter@iems.northwestern.edu, alejandra.pena@u.northwestern.edu

MS74

Stochastic Derivative Free Optimization for Hyperparameter Tuning in Machine Learning Problems

The performance of many machine learning algorithms depends on model hyperparameters and thus on the method which sets these hyperparameters. In particular, hyperparameter optimization can be applied to tune model related hyperparameters of convex machine learning problems (kernel SVM, logistic regression), deep learning (DNN, DBN), and optimizing nonsmooth nonconvex loss functions, such as AUC. This class of optimization problems can be viewed as a blackbox optimization problem, which can be considered as a (possibly smooth) stochastic optimization problem. In this work, we utilize the Trust Region based Derivative Free Optimization (DFO-TR) as a hyperparameter optimization method. We present computational comparison of DFO-TR method with the state-of-the-art methods such as some model-based Bayesian optimization algorithms as well as the traditional random search approach.

Hiva Ghanbari, Katya Scheinberg
Lehigh University
coralia.cartis@maths.ox.ac.uk, kas410@lehigh.edu

MS74

A Levenberg-Marquardt Method for Large-Scale Noisy Nonlinear Least Squares Problems

Nonlinear least squares problems arise in many applications. We are going to consider the case in which we don't have at disposal the exact value of the function, of the gradient and of the Jacobian matrix, but just approximations are known. We propose a Levenberg-Marquardt method that is suitable to deal with the noise in the function, gradient and Jacobian and guarantees global convergence to a solution of the unperturbed problem as the noise level tends to zero. Our procedure starts with a given noise level and we propose a mechanism to measure if the noise in the function values is too high to achieve a good reduction in the unperturbed objective function. In that case we assume to be able to reduce the noise. At each iteration an inexact Levenberg-Marquardt step is taken, and this makes the method suitable also to large-scale problems. Numerical results will be provided.

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

Serge Gratton
ENSEEIHT, Toulouse, France
serge.gratton@enseeiht.fr

Elisa Riccietti
University of Florence
Italy
elisa.riccietti@unifi.it

MS74

Beyond SGD: Faster Stochastic Methods for Non-convex Optimization

We study nonconvex finite-sum problems (Empirical Risk Minimization) and analyze stochastic variance reduced (VR) methods for solving them efficiently. Variance reduction methods have recently surged into prominence for convex optimization, especially due to their edge over stochastic gradient (SGD); but their theoretical analysis has almost exclusively assumed convexity. In contrast, we prove non-asymptotic rates of convergence (to stationary points) of SVRG (and SAGA) for nonconvex optimization, and show that these methods converge provably faster than SGD and even batch gradient descent. We also analyze a subclass of nonconvex problems on which SVRG attains linear convergence to the global optimum. Our analysis extends, albeit is non-obvious ways to handle nonsmooth, constrained, and parallel (mini-batch) variants — time permitting, we will comment on these and point out very curious open problems relating to these methods.

Suvrit Sra
Massachusetts Institute of Technology
suvrit@mit.edu

Sashank Reddi
Carnegie Mellon University
sjakkamr@cs.cmu.edu

MS74

Exact Worst-Case Performance of First-order Methods: Recent Developments

We introduce the performance estimation approach. This methodology aims at automatically analyzing the convergence properties of first-order algorithms for solving (composite convex) optimization problems. In particular, it allows obtaining tight guarantees for fixed-step first-order methods involving a variety of different oracles - namely explicit, projected, proximal, conditional and inexact (sub)gradient steps - and a variety of convergence measures. During the presentation, we will present the methodology and illustrate how it can be used for developing new algorithms. On the way, we will present a toolbox allowing to easily use the approach for studying simple first-order methods.

Adrien Taylor
UCL, Belgium
adrien.taylor@uclouvain.be

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

Francois Glineur
Université Catholique de Louvain (UCL)
Center for Operations Research and Econometrics (CORE)
francois.glineur@uclouvain.be

MS75

A SMART Stochastic Algorithm for Nonconvex Optimization with Applications to Robust Machine Learning

In this paper, we show how to transform any optimization problem that arises from fitting a machine learning model into one that (1) detects and removes contaminated data from the training set while (2) simultaneously fitting the trimmed model on the uncontaminated data that remains. To solve the resulting nonconvex optimization problem, we introduce a fast stochastic proximal-gradient algorithm that incorporates prior knowledge through nonsmooth regularization. For datasets of size n, our approach requires $O(n^{2/3}/\varepsilon)$ gradient evaluations to reach ε-accuracy and, when a certain error bound holds, the complexity improves to $O(\kappa n^{2/3}\log(1/\varepsilon))$. These rates are $n^{1/3}$ times better than those achieved by typical, full gradient methods.

Damek Davis
Cornell University
ORIE
dsd95@cornell.edu

MS75

Improving the Optimized Gradient Method for Large-Scale Convex Optimization

We recently proposed the optimized gradient method (OGM) [Math. Prog., 151:83-107, Sep. 2016] that is as efficient as Nesterov's fast gradient method (FGM), and satisfies a worst-case cost-function convergence bound that is twice smaller than that of FGM, and that is optimal for large-dimensional smooth convex problems. Here, we present two approaches to improving and extending the OGM. First, we adapt an adaptive restarting scheme originally developed for FGM to heuristically improve the convergence rate of OGM, particularly when the iterates enter a locally well-conditioned region. Second, we propose a new algorithm called OGM-OG (optimized over gradient) by optimizing the step coefficients with respect to the rate of gradient norm decrease, whereas we optimized the original OGM with respect to the cost function decrease.

Donghwan Kim, Jeffrey Fessler
University of Michigan

kimdongh@umich.edu, fessler@umich.edu

MS75

Accelerated Primal-Dual Method for Affinely Constrained Problems

Motivated by big data applications, first-order methods have been extremely popular in recent years. However, naive gradient methods generally converge slowly. Hence, much efforts have been made to accelerate various first-order methods. This talk presents two accelerated methods towards solving structured linearly constrained composite convex programming. The first method is the accelerated linearized augmented Lagrangian method (LALM). Assuming merely weak convexity, we show that LALM owns $O(1/k^2)$ if its parameters are adapted. The second method is the accelerated linearized alternating direction method of multipliers (LADMM). In addition to the composite convexity, it further assumes two-block structure. Different from classic ADMM, our method allows linearization to the objective and also augmented term to make the update simple. Assuming strong convexity on one block variable, we show that LADMM also enjoys $O(1/k^2)$ convergence with adaptive parameters. This result is a significant improvement over that in [Goldstein et. al, SIIMS'14], which requires strong convexity on both block variables and no linearization to the objective or augmented term. Numerical experiments will be shown to demonstrate the validness of acceleration and superior performance of the proposed methods over existing ones.

Yangyang Xu
University of Alabama
yangyang@ima.umn.edu

MS75

Primal-Dual Algorithms for the Sum of Three Operators

In this talk, I will introduce existing primal-dual algorithms for minimizing f(x) + g(x) + h(Lx), where f is convex and Lipschitz differentiable, g and h are convex (possibly nondifferentiable), and L is a linear operator. Then, I will propose a new primal-dual algorithm, which is a generalization of two primal-dual algorithms for solving the sum of two operators. In addition, the new algorithm recovers many operator splitting methods involving two and three operators.

Ming Yan
Michigan State University
Department of CMSE; Department of Mathematics
yanm@math.msu.edu

MS76

A Copositive Approach for Two-Stage Adjustable Robust Optimization with Uncertain Right-Hand Sides

We study two-stage adjustable robust linear programming in which the right-hand sides are uncertain and belong to a convex, compact uncertainty set. This problem is NP-hard, and the affine policy is a popular, tractable approximation. We prove that under standard and simple conditions, the two-stage problem can be reformulated as a copositive optimization problem, which in turn leads to a class of tractable, semidefinite-based approximations that are at least as strong as the affine policy. We investigate several examples from the literature demonstrating that our tractable approximations significantly improve the affine policy. In particular, our approach solves exactly in polynomial time a class of instances of increasing size for which the affine policy admits an arbitrarily large gap.

Guanglin Xu
Department of Management Sciences
The University of Iowa
guanglin-xu@uiowa.edu

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

MS76

Approximation Guarantees for Monomial Convexification in Polynomial Optimization

Consider minimizing a polynomial function over a compact convex set. Convexifying each monomial separately yields a lower bound on the global optimum. For any monomial whose domain is a subset of the $[0, 1]^n$ box, we give upper bounds on the approximation error produced by the closure convex hull of this monomial. The error measure is the maximum absolute deviation between the actual and approximatedproduct value. Our bounds are functions of the monomial degree. Special structures of the domain for which our bounds are tight are also analyzed. For a multilinear monomial over the $[1, r]^n$ box, where r is a positive real, we give refined error bounds. As a step towards addressing mixed-sign variable domains, we provide error bounds for a multilinear monomial over the $[-1, 1]^n$ box. All the above anaylses together imply an upper bound on the additive error in the global optimum due to monomial convexifications of a polynomial.

Akshay Gupte
Department of Mathematical Sciences
Clemson University
agupte@clemson.edu

Warren Adams, Yibo Xu
Clemson University
wadams@clemson.edu, yibox@clemson.edu

MS76

The Multilinear Polytope for Gamma-acyclic Hypergraphs

We consider the Multilinear polytope defined as the convex hull of the set of binary points satisfying a collection of multilinear equations. Such sets are of fundamental importance in many types of mixed-integer nonlinear optimization problems, such as binary polynomial optimization. Utilizing an equivalent hypergraph representation, we study the facial structure of the Multilinear polytope in conjunction with the acyclicity degree of the underlying hypergraph. We provide explicit characterizations of the Multilinear polytopes corresponding to Berge-acylic and gamma-acyclic hypergraphs. As the Multilinear polytope for gamma-acyclic hypergraphs may contain exponentially many facets in general, we present a highly efficient polynomial algorithm to solve the separation problem. Our results imply polynomial solvability of the corresponding classes of binary polynomial optimization problems and provide new types of cutting planes for a variety of mixed-integer non-

linear optimization problems.

Aida Khajavirad
IBM TJ Watson research center
aida@cmu.edu

Alberto Del Pia
University of Wisconsin
delpia@wisc.edu

MS76

Low-Complexity Relaxations and Convex Hulls of Disjunctions on the Positive Semidefinite Cone and Other Regular Cones

In this talk we consider linear two-term disjunctions on a regular cone $\mathbf{K}$. The resulting disjunctive sets provide fundamental non-convex relaxations for mixed-integer conic programs. We develop a family of structured convex inequalities that are valid for a disjunctive set of this form. Under mild assumptions on the choice of disjunction, these inequalities collectively describe the closed convex hull of the disjunctive set in the space of the original variables, and if additional conditions are satisfied, a single inequality from this family is adequate for a complete closed convex hull characterization. In the cases where the cone $\mathbf{K}$ is the positive semidefinite cone or a direct product of second-order cones and nonnegative rays, we show that these inequalities admit equivalent conic representations for certain choices of disjunction. For more general disjunctions, we present low-complexity convex relaxations with favorable tightness properties in the space of the original variables. Along the way, we also establish connections between two-term disjunctions and non-convex sets defined by rank-two quadratics.

Sercan Yildiz
University of North Carolina at Chapel Hill
syildiz@samsi.info

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

MS77

Global Multi-Objective Optimization and an Application to Robust Optimization

For globally solving multi-objective optimization problems, so far parameter dependent scalarizations have to be solved iteratively by a global procedure. We present here a direct approach based on a branch and bound scheme and on convex underestimators. A new discarding test is presented which uses outer approximations of convex multi-objective optimization problems. We apply these techniques also for computing a covering of the solutions of a robust multiobjective optimization problem. There, decision uncertainty is taken into account by considering to each variable all possible realizations and the correspondent objective function values. By choosing a robust approach this leads to a special set optimization problem.

Gabriele Eichfelder
Institute of Mathematics
Technische Universität Ilmenau
Gabriele.Eichfelder@tu-ilmenau.de

Julia Niebling
Technische Universität Ilmenau
julia.niebling@tu-ilmenau.de

MS77

Multicriteria Tradeoff Analysis for Optimization under Uncertainty

We present a theoretical analysis and propose new methodological approaches to compute improved tradeoff solutions for optimization problems under uncertainty. Our results use several concepts from multiobjective programming including scalarizations and extended notions of robustness and proper efficiency. Their potential impact will be demonstrated on applications of financial portfolio selection and price-aware electric vehicle charging.

Alexander Engau
University of Colorado, Denver, USA.
aengau@alumni.clemson.edu

MS77

Vector Optimization Methods for Solving Matrix Completion Problems

Motivated by applications in machine learning, e.g. the well known Netflix Challenge, we are interested in the following problem: Given an incomplete $m \times n$-data matrix M, which entries are only known for indices $(i, j) \in \Omega$, we are looking for a completion $X \in \mathbb{R}^{m \times n}$ of M. As it is often the case in real-world problems, we also demand that X has a special structure. For example, let us assume X has to be a sparse and low-rank matrix. Therefore, we can formulate the following matrix optimization problem

$$\min_X \sum_{(i,j)\in\Omega} [X_{ij} - M_{ij}]^2 + \lambda \|X\|_* + \mu \|X\|_1.$$

Here, $\|\cdot\|_*$ denotes the nuclear norm which promotes low-rank solutions, and $\|\cdot\|_1$ denotes the ℓ_1-norm which promotes the sparsity of X. The quality of a solution for this problem, with respect to the underlying application, highly depends on a good choice of the regularization parameters λ and μ. In this talk, we will discuss the parameter choice from the viewpoint of vector optimization.

Christopher Schneider
Friedrich-Schiller-University of Jena
christopher.schneider@uni-jena.de

MS77

Indifference Pricing via Convex Vector Optimization

When the preferences are incomplete, they can be represented by vector valued utility functionals. In this case, the utility maximization problem is a convex vector optimization problem. We allow utility functions to be multivariate and we define the utility buy and sell prices as set valued functions of the claim. The set-valued buy and sell prices recover the complete preference case and they satisfy some monotonicity and convexity properties as expected. It is possible to approximate these set valued prices by solving convex vector optimization problems.

Firdevs Ulus
Bilkent University
firdevs@bilkent.edu.tr

Birgit Rudloff

Vienna University of Economics and Business
brudloff@wu.ac.at

MS78

MM Algorithms for Mixture Modeling and Robust, Structured Regression

We present a computational algorithm for performing mixture modeling and robust, structured regression using the L2E method, a minimum distance estimator. Previous implementations for the method were limited to fitting only a handful of parameters. We introduce an iterative majorization framework for extending the L2E method proposed in Scott (2001, 2009) to handling high-dimensional models. We demonstrate our algorithm on simulated and real data examples.

Eric Chi
Department of Statistics
North Carolina State University
eric_chi@ncsu.edu

MS78

An Overview of MM Algorithms

This talk will survey the history, theory, and applications of the MM principle, a framework for constructing monotone optimization algorithms in high-dimensional models. The MM principle transfers optimization from the objective function to a surrogate function and simpli es matters by: (a) separating the variables of a problem, (b) avoiding large matrix inversions, (c) linearizing a problem, (d) restoring symmetry, (e) dealing with equality and inequality constraints gracefully, and (f) turning a nondi erentiable problem into a smooth problem. The art in devising an MM algorithm lies in choosing a tractable surrogate function that hugs the objective function as tightly possible. The EM principle from statistics is a special case of the MM principle. Modern mathematical themes such as sparsity and parallelization mesh well with the MM principle.

Kenneth Lange
UCLA
klange@ucla.edu

MS78

The MM Principle for Split Feasibility Problems

We present a majorization-minimization algorithm for the non-linear split feasibility problem. The classical multi-set split feasibility problem seeks a point in the intersection of finitely many closed convex domain constraints, whose image under a linear mapping also lies in the intersection of finitely many closed convex range constraints. Split feasibility generalizes important inverse problems including convex feasibility, linear complementarity, and regression with constraint sets. When a feasible point does not exist, methods that proceed by minimizing a proximity function can be used to obtain optimal approximate solutions to the problem. Our work extends the proximity function approach for the linear case to allow for non-linear mappings. Our algorithm is amenable to quasi-Newton acceleration, and comes complete with convergence guarantees under mild assumptions, and applies to proximity functions in terms of arbitrary Bregman divergences. We explore several examples including regression for constrained generalized linear models and rank-restricted matrix regression, and consider a case study in optimization for intensity-modulated radiation therapy.

Jason Xu
UCLA
jasonxu90@gmail.com

MS78

MM Algorithms For Variance Components Models

Variance components estimation and mixed model analysis are central themes in statistics with applications in numerous scientific disciplines. Despite the best efforts of generations of statisticians and numerical analysts, maximum likelihood estimation and restricted maximum likelihood estimation of variance component models remain numerically challenging. In this talk, we present a novel iterative algorithm for variance components estimation based on the minorization-maximization (MM) principle. MM algorithm is trivial to implement and competitive on large data problems. The algorithm readily extends to more complicated problems such as linear mixed models, multivariate response models possibly with missing data, maximum a posteriori estimation, and penalized estimation. We demonstrate, both numerically and theoretically, that it converges faster than the classical EM algorithm when the number of variance components is greater than two.

Hua Zhou
UCLA
huazhou@ucla.edu

MS79

Conic Programming Reformulations of Two-Stage Distributionally Robust Linear Programs over Wasserstein Balls

Adaptive robust optimization problems are usually solved *approximately* by restricting the adaptive decisions to simple parametric decision rules. However, the corresponding approximation error can be substantial. In this paper we show that two-stage robust and distributionally robust linear programs can often be reformulated *exactly* as conic programs that scale polynomially with the problem dimensions. Specifically, when the ambiguity set constitutes a 2-Wasserstein ball centered at a discrete distribution, then the distributionally robust linear program is equivalent to a copositive program (if the problem has complete recourse) or can be approximated arbitrarily closely by a sequence of copositive programs (if the problem has sufficiently expensive recourse). These results directly extend to the classical robust setting and motivate strong tractable approximations of two-stage problems based on semidefinite approximations of the copositive cone. We also demonstrate that the two-stage distributionally robust optimization problem is equivalent to a tractable linear program when the ambiguity set constitutes a 1-Wasserstein ball centered at a discrete distribution and there are no support constraints.

Grani A. Hanasusanto
Operations Research and Industrial Engineering
The University of Texas at Austin
grani.hanasusanto@utexas.edu

Daniel Kuhn
École Polytechnique Fédérale de Lausanne

daniel.kuhn@epfl.ch

MS79

Appointment Scheduling under Schedule-Dependent Patient No-Show Behavior

This paper studies an appointment scheduling problem under schedule- dependent patient no-show behavior. The problem is motivated by our studies of independent datasets from countries in two continents which identify a significant time-of-day effect on patient show-up probabilities. We deploy a distributionally robust model, which minimizes the worst case total expected cost of patient waiting and service provider's idle and overtime, by optimizing the scheduled arrival times of patients. We show that this model under schedule- dependent patient show-up behavior can be reformulated as a copositive program and then be approximated by semidefinite programs. These formulations are obtained by a new technique that uses a completely positive program to equivalently represent a linear program with uncertainties present in both the objective function and the right-hand side of the constraint sets. To tackle the case when patient no-shows are endogenous on the schedule, we construct a set of dual prices to guide the search for a good schedule and use the technique iteratively to obtain a near optimal solution. Our computational studies reveal a significant reduction in total expected cost by taking into account the time-of-day variation in patient show-up probabilities as opposed to ignoring it.

Chung-Piaw Teo
Department of Decision Sciences
NUS Business School
bizteocp@nus.edu.sg

Qingxia Kong
Erasmus University Rotterdam
q.kong@rsm.nl

Shan Li
City University of New York
shan.li@baruch.cuny.edu

Nan Liu
Columbia University
nl2320@cumc.columbia.edu

Zhenzhen Yan
National University of Singapore
a0109727@u.nus.edu

MS79

A Data-Driven Distributionally Robust Bound on the Expected Optimal Value of Uncertain Mixed 0-1 Linear Programming

We study the expected optimal value of a mixed 0-1 programming problem with uncertain objective coefficients following a joint distribution. We assume that the true distribution is not known exactly, but a set of independent samples can be observed. Using the Wasserstein metric, we construct an ambiguity set centered at the empirical distribution from the observed samples and containing all distributions that could have generated the observed samples with a high statistical confidence. The problem of interest is to investigate the bound on the expected optimal value over the Wasserstein ambiguity set. Under standard assumptions, we reformulate the problem into a copositive programming problem, which naturally leads to a tractable semidefinite-based approximation. We compare our approach with a moment-based approach from the literature for two applications. The numerical results illustrate the effectiveness of our approach.

Guanglin Xu
Department of Management Sciences
The University of Iowa
guanglin-xu@uiowa.edu

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

MS79

Distribution-Free Robust Optimization with Shape Constraints

One of the major critiques of the distributional free robust bounds is the discrete nature of its optimum distribution. Increasing failure rate (IFR) and log-concave are two widely accepted and used distribution classes in various research domains. Even though the corresponding math programming problems involving such distributions are highly non-convex, we establish an optimal solution structure via analyzing a relaxed problem. Furthermore, we develop numerical toolbox for optimally solving the distributional free moment problems with IFR and log-concave distribution assumptions.

Jiang Bo
Shanghai University of Finance and Economics
isyebojiang@163.com

Xi Chen
New York Unviersity
xchen3@stern.nyu.edu

Simai He
Shanghai University of Finance and Economics
simaihe@mail.shufe.edu.cn

Christopher T. Ryan
University of Chicago
chris.ryan@chicagobooth.edu

Teng Zhang
Stanford University
tengz@stanford.edu

MS80

Guarantees for Subspace Learning by Incremental Gradient Descent on the Grassmannian

It has been observed in a variety of contexts that gradient descent methods have great success in solving low-rank matrix factorization problems, despite the relevant problem formulation being non-convex. The Singular Value Decomposition (SVD) is the solution to a non-convex optimization problem, and there are several highly successful linear-algebraic algorithms for solving it. Unfortunately, these algorithms cannot easily be extended to problems with regularizers or missing data. In this talk I will focus on the problem where we seek the d-dimensional subspace spanned by a streaming data matrix. We apply the natural first order incremental gradient descent method, constraining the gradient method to the Grassmannian. I will discuss these algorithms and theoretical results in the standard l2 norm subspace learning problem with noisy or undersampled data. Additionally I will present algorithms and empirical results for the robust PCA and sparse PCA problems. (Joint work with Dejiao Zhang.)

Laura Balzano
Electrical Engineering Dept
University of Wisconsin
girasole@umich.edu

MS80

Nonconvex Statisitcal Optimization: Global Exploration Versus Local Exploitation

Taming nonconvexity requires a seamless integration of both global exploration and local exploitation of latent convexity. In this talk, we study a generic latent variable model that includes PCA, ICA, and dictionary learning as special cases. Using this statistical model as illustration, we propose two global exploration schemes for identifying the basin of attraction, namely tightening after relaxation and noise regularization, which allow further local exploitation via gradient-based methods. Interestingly, given a sufficiently good initialization, the local exploitation methods are often information-theoretic optimal. In sharp contrast, based on an oracle computational model we prove that, one must pay a computational or statistical price in the global exploration stage to attain a good initialization (Based on joint work with Zhaoran Wang, Zhuoran Yang, and Junchi Li).

Han Liu
Princeton University
hanliu@princeton.edu

MS80

Theory for Local Optima in Nonconvex Regression Problems

We present results for high-dimensional linear regression using robust M-estimators with a regularization term. We show that when the derivative of the loss function is bounded, our estimators are robust with respect to heavy-tailed noise distributions and outliers in the response variables, with the usual order of k log p/n rates for high-dimensional statistical estimation. Our results continue a line of recent work concerning local optima of nonconvex M-estimators with possibly nonconvex penalties, where we adapt the theory to settings where the loss function only satisfies a form of restricted strong convexity within a local neighborhood. We also discuss second-order results concerning the asymptotic normality of our estimators, and provide a two-step M-estimation algorithm for obtaining statistically efficient solutions within the local region.

Po-Ling Loh
University of Wisconsin
loh@ece.wisc.edu

MS81

The Langevin MCMC: Theory and Methods

The complexity and sheer size of modern data sets, to whichever increasingly demanding questions are posed, give rise to major challenges. Traditional simulation methods often scale poorly with data size and model complexity, and thus fail for the most complex of modern problems. We are considering the problem of sampling from a log-concave distribution. Many problems in machine learning fail into this framework, such has linear ill-posed inverse problems with sparsity inducing priors, or large scale Bayesian binary regression.. The purpose of this lecture is to explain how we can use ideas which have proven very useful in machine learning community to solve large scale optimization problems to design efficient sampling algorithms. Most of the efficient algorithms know so far may be seen as variants of the gradient descent algorithms, most often coupled with partial updates (coordinates descent algorithms). This of course suggests to study methods derived from Euler discretization of the Langevin diffusion. Partial updates may in this context as Gibbs steps This algorithm may be generalized in the non-smooth case by regularizing the objective function. The Moreau-Yosida inf-convolution algorithm is an appropriate candidate in such case. We will prove convergence results for these algorithms with explicit convergence bounds both in Wasserstein distance and in total variation. Numerical illustrations will be presented to illustrate our results.

Eric Moulines
Télécom ParisTech
eric.moulines@polytechnique.edu

Alain Durmus
Telecom ParisTech
alain.durmus@telecom-paristech.fr

Nicolas Brosse
Ecole Polytechnique
nicolas.brosse@polytechnique.edu

MS81

Converging on the Ultimate Optimization Algorithm for Machine Learning

In this talk I'll survey recent advances in algorithms for minimizing the cost functions arising from convex machine learning models. I'll talk about recent work on stochastic methods incorporating acceleration, Newton updates, mini-batches, and non-uniform sampling for minimizing finite sums of convex functions.

Mark Schmidt
University of British Columbia
schmidtm@cs.ubc.ca

MS81

Fast Convergence of Newton-Type Methods on High-Dimensional Problems

We study the convergence rate of Newton-type methods on high-dimensional problems. The high-dimensional nature of the problem precludes the usual global strong convexity and smoothness that underlie the classical analysis of such methods. We find that restricted version of these conditions which typically arise in the study of the statistical properties of the solutions are also enough to ensure good computational properties of Newton-type methods. We explore the algorithmic consequences in distributed and online settings.

Yuekai Sun
University of Michigan

yuekai@umich.edu

MS82

Relative Entropy Optimization and Dynamical Systems

We discuss relative entropy convex programs, which represent a generalization of geometric programs and second-order cone programs. We describe an application using these ideas to hitting-time estimation in certain dynamical systems, which in turn leads to an approach for checking set membership.

Venkat Chandrasekaran
California Institute of Technology
venkatc@caltech.edu

Parikshit Shah
University of Wisconsin/Philips Research
pshah@discovery.wisc.edu

MS82

Gradient Descent Learns Linear Systems

We prove that gradient descent efficiently converges to the global optimizer of the maximum likelihood objective of an unknown linear time-invariant dynamical system from a sequence of noisy observations generated by the system. Even though the objective function is non-convex, we provide polynomial running time and sample complexity bounds under strong but natural assumptions. Linear systems identification has been studied for many decades, yet, to the best of our knowledge, these are the first polynomial guarantees for the problem we consider.

Moritz Hardt
Google, Inc.
mrtz@google.com

Tengyu Ma
Princeton University
tengyu@cs.princeton.edu

Benjamin Recht
University of California, Berkeley
brecht@berkeley.edu

MS82

Efficient Methods for Online Node Classification in a Network

We consider the problem of sequentially classifying nodes in a network. A natural formulation leads to an NP hard combinatorial objective. We discuss a relaxation technique for solving a certain dynamic program that leads to a linear-time algorithm with a nearly-optimal mistake bound.

Alexander Rakhlin
University of Pennsylvania
rakhlin@wharton.upenn.edu

MS82

Randomized Approximation of Feed-Forward Network Using Compositional Kernels

We describe and analyze a simple random feature scheme (RFS) from prescribed compositional kernels. The compositional kernels we use are inspired by the structure of neural networks and convolutional kernels. The resulting scheme is simple and yields sparse and efficiently computable features. Each random feature can be represented as an algebraic expression over a *small* number of (random) paths in a composition tree. Thus, compositional random features can be stored very compactly. The discrete nature of the generation process enables de-duplication of repeated features, further compacting the representation and increasing the diversity of the embeddings. Our approach rather complements and can be combined with previous random feature schemes.

Amit Daniely, Roy Frostig, Vineet Gupta
Google Brain
amitdaniely@google.com, frostig@google.com,
vineet@google.com

Yoram Singer
Princeton University
ysinger@cs.princeton.edu

MS83

On the Construction of Exact Augmented Lagrangian Functions for Nonlinear Semidefinite Optimization

Nonlinear semidefinite programming (NSDP) problems extend the well-known nonlinear programming and the linear semidefinite programming problems, and have many applications, for example, in feedback control and structural optimization. The research associated to NSDP is, however, relatively scarce and few methods have been developed. In this work, we propose a general exact augmented Lagrangian function for NSDP problems. By exact, we mean that the unconstrained minimization of the constructed function gives a solution of the original problem, when the penalty parameter is large enough. Differently from the traditional exact penalty function, the exact augmented Lagrangian is defined on the product space of the problem's variables and of the multipliers. Here, we will give a general formula for such a function and, based on that, we will construct another one, which is continuously differentiable, and with guaranteed exactness property.

Ellen H. Fukuda
State University of Campinas
ellen@i.kyoto-u.ac.jp

Bruno Lourenco
Seikei University
lourenco@st.seikei.ac.jp

MS83

Sequential Injective Algorithm for Weakly Univalent Vector Equation and Its Application to Mixed Second-Order Cone Complementarity Problem

It is known that the conic complementarity problems and the variational inequality problems are reformulated equivalently as vector equations by using the natural residual or Fischer-Burmeister function. Moreover, under some mild assumptions, those vector equations possess the weak univalence property. In this study, we first provide a sequential injective algorithm for a weakly univalent vector equation. We note that the algorithm can be cast as a *prototype* for many kinds of algorithm such as the smoothing Newton method, regularized smoothing Newton method, semi-smooth Newton method, etc. Then, we apply the prototype algorithm and the convergence analysis to the regularized smoothing Newton algorithm for mixed nonlinear second-order cone complementarity problems. We prove the global convergence property under the Cartesian P_0 assumption, which is strictly weaker than the monotonicity assumption.

Shunsuke Hayashi
Graduate School of Information Sciences, Tohoku University
s_hayashi@plan.civil.tohoku.ac.jp

MS83

An Extension of Chubanov's Algorithm to Symmetric Cones

We present a new algorithm for homogeneous feasibility problems over symmetric cones and discuss its complexity. We will also present the special case where the cone is the positive semidefinite matrices. Our work is an extension of Chubanov's algorithm for linear programs and, analogously, it consists of a basic procedure and a step where the possible solutions are confined to an intersection of the cone and a half-space. It also has some flavor of the ellipsoid method and of interior point methods. Some of the distinguishing features are that progress is measured through volumes and we make use of the so-called "spectral norms". We will also briefly compare our approach to a recent work by Peña and Soheili.

Bruno Lourenco
Seikei University
lourenco@st.seikei.ac.jp

Tomonari Kitahara
Tokyo Institute of Technology
kitahara.t.ab@m.titech.ac.jp

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

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

MS83

Sub-Homogeneous Optimization Problems and its Applications

We consider an optimization problem with sub-homogeneous functions in its objective and constraint functions. Since the absolute value function is sub-homogeneous, our problem is a generalization of the absolute value optimization (AVO) proposed by Mangasarian in 2007. The problem includes the 0-1 integer programming problem, the linear programming with linear complementarity constraints, and Lasso regularized regression. We give a new dual formulation of our problem that has a closed form. It is an extension of the AVO dual problem given by Mangasarian. We show that the new dual formulation is essentially the same as the Lagrangean dual, which does not have a closed form in general. To the best of our knowledge, this property is new even for AVO.

Shota Yamanaka, Nobuo Yamashita
Kyoto University
shota@amp.i.kyoto-u.ac.jp, nobuo@i.kyoto-u.ac.jp

MS84

Power Cones in Second-Order Cone Form and Dual Recovery

Although rational power cones (or equivalently, geometric mean cones) can be represented using second-order cones, the transformation is cumbersome. We prove presolve and postsolve steps to automate this transformation based on a surprisingly size-economic reformulation procedure. Primal and dual values can thus be computed for the original power cone formulation using second-order cone solvers. Computational results with MOSEK are given.

Henrik Friberg
MOSEK Optimization
henrik.alsing.friberg@mosek.com

MS84

Partial Polyhedrality and Facial Reduction

In general conic programming, there may be a positive duality gap or the optimal value may not be attained when there's no interior feasible solution. Facial Reduction Algorithms (FRAs) give a remedy for this situation by giving the minimal cone of the feasible region or by finding that the problem is infeasible. However, we usually do not use FRAs for linear programming (LP), because the well-known strong duality theorem holds for LP; no such nasty cases even if no interior feasible points exist. In this sense, polyhedral cones are 'harmless.' So, our claim is that polyhedral cones should be treated separately and properly even when they appear in general cones. We say that a cone is partially polyhedral when it is written as a direct product of a polyhedral cone and another general cone. We propose a new concept of distance to polyhedrality, and construct a new FRA, FRA-poly, to exploit the partial polyhedrality. As a result, the number of iterations is drastically reduced when the distance to the polyhedrality is small; in particular, in the case of the doubly nonnegative cone, FRA-poly gives a worst case bound $O(n)$ whereas the classical FRA needs $O(n^2)$. Of possible independent interest, we prove a variant of Gordan-Stiemke's Theorem and a proper separation theorem that takes into account partial polyhedrality.

Masakazu Muramatsu
University of Electro-Communications
muramatu@cs.uec.ac.jp

Bruno Lourenco
Seikei University
lourenco@st.seikei.ac.jp

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

MS84

Low-Order Complexity Results for SOS/SDP Methods in Real Algebra

In this talk, I will present some new algorithmic approaches to polynomial optimization and related problems that are based on Semidefinite optimization and Sums of squares and which have low orders of complexity. These results give resolution to some long standing unsolved problems in the area.

Motakuri Ramana
United Airlines
mvrmath@zoho.com

MS84

On An Algorithm for Conic Linear Program

Recently, Chubanov proposed an interesting new polynomial-time algorithm for linear programs. In this talk, we extend his algorithm to second-order cone programming. The algorithm deals with a homogeneous feasibility problem and, given a tolerance value epsilon, finds a primal interior-feasible solution or a dual nonzero feasible-solution, or concludes that there is no epsilon-interior feasible solution to the primal system in O(n log 1/epsilon) iterations of the basic procedure, by generating a series of the direct product of shrinking obliquely truncated second-order cones and intervals. Roughly, the basic procedure corresponds to one iteration of interior-point algorithm. The algorithm has similarity to the ellipsoid method in the sense that it generates a series of shrinking convex bodies of the same type, but it also has some flavor of the interior-point method in that it utilizes the automorphism group of the cone.

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

Tomonari Kitahara
Tokyo Institute of Technology
kitahara.t.ab@m.titech.ac.jp

MS85

A Multigrid Algorithm for SDP Relaxations of Sparse Polynomial Problems Arising in PDE Optimization

We propose a multigrid approach for the global optimization of a class of polynomial optimization problems with sparse support. The problems we consider arise from the discretization of infinite dimensional optimization problems. In many of these applications, the level of discretization can be used to obtain a hierarchy of optimization models that captures the underlying infinite dimensional problem at different degrees of fidelity. The main difficulty of applying multigrid methods to SPDs is that the geometric information between grids is lost when the original problem is approximated via an SDP relaxation. We show how a multigrid approach can be applied by developing projection operators to relate the primal and dual variables of the SDP relaxation between lower and higher levels in the hierarchy of discretizations. We develop sufficient conditions for the operators to be useful in applications. Our conditions are easy to verify in practice, and we discuss how they can be used with infeasible interior point methods. Our preliminary results highlight two promising advantages of following a multigrid approach in contrast with a pure interior point method: the percentage of problems that can be solved to a high accuracy is higher, and the time necessary to find a solution can be reduced, especially for large scale problems.

Juan Campos Salazar
Imperial College London
juan.campos-salazar12@imperial.ac.uk

MS85

Numerical and Theoretical Aspects of Shape Optimization for Fluid-Structure Interaction

In this talk we consider shape optimization for unsteady fluid-structure interaction problems that couple Navier-Stokes equations with non-linear elasticity equations. We focus on the monolithic approach in the Lagrangian framework. It is obtained by applying an ALE transformation to the fluid equations that are usually formulated on the time dependent physical domain. Shape optimization by the method of mappings approach requires another transformation which maps the ALE reference domain to a reference domain for shape optimization. This yields an optimal control setting and therefore can be used to drive an optimization algorithm with adjoint based gradient computation. Numerical results for our implementation, which builds on FEniCS, dolfin-adjoint, and IPOPT, are presented. In addition, some theoretical aspects on the differentiability of the state variables of fluid-structure interaction models in optimal control settings are discussed.

Johannes Haubner
Technical University of Munich
johannes.haubner@ma.tum.de

Michael Ulbrich
Technische Universitaet Muenchen
Chair of Mathematical Optimization
mulbrich@ma.tum.de

MS85

Primal-Dual Interior-Point Multigrid Method for Topology Optimization

An interior point method for the structural topology optimization is proposed. The linear systems arising in the method are solved by the conjugate gradient method preconditioned by geometric multigrid. The resulting method is then compared with the so-called optimality condition method, an established technique in topology optimization. This method is also equipped with the multigrid preconditioned conjugate gradient algorithm. We conclude that, for large scale problems, the interior point method with an inexact iterative linear solver is superior to any other variant studied in the paper.

Michal Kocvara
School of Mathematics
University of Birmingham
m.kocvara@bham.ac.uk

MS86

New Active Set Frank-Wolfe Variants for Minimization over the Simplex and the ℓ_1 Ball

In this talk, we focus on active-set variants of the Frank-Wolfe algorithm for minimizing a function over the simplex and the ℓ_1 ball. We first describe an active-set estimate (i.e. an estimate of the indices of the zero variables in the optimal solution) for the considered problem that tries to quickly identify as many active variables as possible at a given point, while guaranteeing that approximate optimality conditions are satisfied. We then embed our estimate into different Frank-Wolfe variants and analyze their convergence properties. Finally, we report some numerical results showing the effectiveness of the new approaches.

Andrea Cristofari
DIIAG Sapienza University of Rome
cristofari@dis.uniroma1.it

Marianna De Santis
Università di Roma La Sapienza
Department of Computer, Control, and Management Engineering
mdesantis@dis.uniroma1.it

Stefano Lucidi
University of Rome
lucidi@dis.uniroma1.it

Francesco Rinaldi
University of Padova
rinaldi@math.unipd.it

MS86

Convex Relaxation Without Lifting: Solving Large Semidefinite Programs without all the Pain

Many non-convex problems can be converted into convex problems by "lifting", or mapping the problem into a higher dimensional space. Very often, this results in convex, but intractable semidefinite programs that are too big for convex solvers. In this talk, we discuss new methods for semidefinite and rank-constrained problems that don't require lifting to higher dimensions. First, we discuss biconvex relaxations of non-convex problems that have most of the benefits of convex liftings, but without the added dimensionality. Then, we discuss PhaseMax - a new convex relaxation for rank constrained problems that provably recovers global solutions without added dimensionality.

Tom Goldstein
Department of Computer Science
University of Maryland
tomg@cs.umd.edu

MS86

Application of Recent First-Order Methods to DIC Microscopy

One of the most successful techniques to observe biological processes in living cells without staining is differential interference contrast (DIC) microscopy. DIC images have a typical three dimensional appearance and show high contrast, however they are not suited for topographical and morphological interpretation, for which it is essential to recover the changes in phase of light, i.e. the phase function. The inverse problem of estimating the specimen's phase function from a set of DIC color images acquired at different angles is nonlinear and ill-posed, and can be reformulated as the following nonconvex minimization problem

$$\min_{\phi \in R^N} J(\phi) \equiv J_0(\phi) + \mu J_{TV}(\phi)$$

where J_0 is a least squares term measuring the distance between the observed images and the predicted images, and J_{TV} is the total variation regularizer. In this talk we propose an efficient proximal-gradient method to address the problem of phase estimation in DIC microscopy. The method exploits a stepsize rule based on a Lanczos-like process applied to the Hessian matrix of J_0, and is equipped with an Armijo-like linesearch in order to guarantee convergence. When J_{TV} is chosen as a smoothed version of the total variation, the method reduces to a gradient descent method. We show that the proposed tool is able to obtain in both the smooth and nonsmooth case accurate reconstructions with a reduced computational demand.

Simone Rebegoldi
Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università di Modena e Reggio Emilia
simone.rebegoldi@unife.it

Lola Bautista
Escuela de Ingenieria de Sistemas y Escuela de Fisica
Universidad Industrial de Santander
bautista@i3s.unice.fr

Laure Blanc-Féraud
INRIA
Sophia Antipolis, France
laure.blanc_feraud@sophia.inria.fr

Marco Prato
Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Universit{a} di Modena e Reggio Emilia
marco.prato@unimore.it

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

Arturo Plata
Escuela de Ingenieria de Sistemas y Escuela de Fisica
Universidad Industrial de Santander
aplatag@yahoo.com

MS86

Fast Algorithms for Non-Convex and Non-Smooth Euler's Elastica Regularization Problems

In this talk, I introduce two efficient optimization algorithms for solving Euler's elastica regularized optimization problems. The minimizing functional is non-smooth, non-convex, and involves high-order derivatives, so that traditional gradient descent based methods converge very slowly. Recent alternating minimization methods show fast convergence when a good choice of parameters is used. The objective of this talk is to introduce efficient algorithms which have simple structures with fewer parameters. These methods are based on operator splitting and alternating direction method of multipliers, and subproblems can be solved efficiently by Fourier transforms and shrinkage operators. I briefly present the analytical properties of each algorithm, as well as numerical experiments, including comparison with some existing state-of-the-art algorithms to show the efficiency and the effectiveness of the proposed methods.

Maryam Yashtini
School of Mathematics
Georgia Institute of Technology

myashtini3@math.gatech.edu

MS87

Fast Proximal L1-Banach Descent Methods

Abstract not available at time of publication.

Marwa El Halabi
EPFL
gosia.baltaian@epfl.ch

MS87

Local Behaviour of Proximal Splitting Methods: Identification, Linear and Finite Convergence

Abstract not available at time of publication.

Jalal Fadili
Normandie Univ, ENSICAEN
Jalal.Fadili@greyc.ensicaen.fr

MS87

Almost Nonexpansive Operators

Abstract not available at time of publication.

Matthew K. Tam
Institute for Numerical and Applied Mathematics
University of Goettingen
m.tam@math.uni-goettingen.de

MS88

A Trust Region Method for Solving Derivative-Free Problems with Binary and Continuous Variables

Trust region methods are used to solve various black-box optimization problems, especially when no derivative information is available. In this talk, we will consider an extension of trust region methods for mixed-integer nonlinear programming (MINLP). There are both theoretical and computational innovations to handle the binary variables, including restricting the quadratic model, solving mixed integer quadratic problems and handling well-poisedness. Whereas, of necessity, we address globality with respect to the binary variables, we are content to obtain good local minima for the continuous variables, at least in part because our typical context involves expensive simulations. We report computational results on analytic and real-life problems and compare the results with existing methods.

Andrew R. Conn
IBM T J Watson Research Ctr
arconn@us.ibm.com

Claudia D'Ambrosio
CNRS, LIX, France
dambrosio@lix.polytechnique.fr

Leo Liberti
Ecole Polytechnique
liberti@lix.polytechnique.fr

Delphine Sinoquet
IFP Energies nouvelles
delphine.sinoquet@ifpen.fr

MS88

Order-Based Error for Managing Ensembles of Surrogates in Derivative-Free Optimization

We investigate surrogate-assisted strategies for derivative-free optimization using the mesh adaptive direct search (MADS) blackbox optimization algorithm. In particular, we build an ensemble of surrogate models to be used within the search step of MADS, and examine different methods for selecting the best model for a given problem at hand. To do so, we introduce an order-based error tailored to surrogate-based search. We report computational experiments for analytical benchmark problems and engineering design applications. Results demonstrate that different metrics may result in different model choices and that the use of order-based metrics improves performance.

Sebastien Le Digabel
Ecole Polytechnique de Montreal
Sebastien.Le.Digabel@gerad.ca

Bastien Talgorn
GERAD
bastientalgorn@yahoo.fr

Charles Audet
École Polytechnique de Montréal - GERAD
Charles.Audet@gerad.ca

Michael Kokkolaras
McGill University
michael.kokkolaras@mcgill.ca

MS88

A New Derivative-Free Model-Based Method for Unconstrained Nonsmooth Optimization

We consider the unconstrained optimization of a nonsmooth function, where first-order information is unavailable or impractical to obtain. We propose a model based method that explicitly takes into account the nonsmoothness of the objective function. A theoretical analysis concerning the global convergence of the approach is carried out.Furthermore, some results related to a preliminary numerical experience are reported.

Giampaolo Liuzzi
CNR - IASI
liuzzi@iasi.cnr.it

Stefano Lucidi
University of Rome
lucidi@dis.uniroma1.it

Francesco Rinaldi
University of Padova
rinaldi@math.unipd.it

Luis Nunes Vicente
University of Coimbra
lnv@mat.uc.pt

MS88

A New Derivative-Free Linesearch Method for

Solving Integer Programming Problems

In this talk, we define a new derivative-free method for integer programming problems with both bound constraints on the variables and general nonlinear constraints. The approach combines a nonmonotone linesearch with a specific penalty approach for handling the nonlinear constraints. The use of both suitable randomly generated search directions and specific stepsizes in the linesearch guarantee that all the points are generated in the integer lattice. We analyze the theoretical properties of the method and show extensive numerical experiments on both bound constrained and nonlinearly constrained problems.

Giampaolo Liuzzi
CNR - IASI
liuzzi@iasi.cnr.it

Stefano Lucidi
University of Rome
lucidi@dis.uniroma1.it

Francesco Rinaldi
University of Padova
rinaldi@math.unipd.it

MS89

R-Linear Convergence of Limited Memory Steepest Descent

The limited memory steepest descent method (LMSD) proposed by Fletcher is an extension of the Barzilai-Borwein "two-point step size" strategy for steepest descent methods for solving unconstrained optimization problems. It is known that the Barzilai-Borwein strategy yields a method with an R-linear rate of convergence when it is employed to minimize a strongly convex quadratic. Our work extends this analysis for LMSD, also for strongly convex quadratics. In particular, it is shown that the method is R-linearly convergent for any choice of the history length parameter. The results of numerical experiments are provided to illustrate behaviors of the method that are revealed through the theoretical analysis.

Frank E. Curtis, Wei Guo
Industrial and Systems Engineering
Lehigh University
frank.e.curtis@gmail.com, weg411@lehigh.edu

MS89

Linear Algebra Issues in Nonlinear Optimization

Abstract not available at time of publication.

Dominique Orban
GERAD and Dept. Mathematics and Industrial Engineering
Ecole Polytechnique de Montreal
dominique.orban@gerad.ca

MS89

Worst-Case Iteration Complexity of a Trust Region Algorithm for Equality Constrained Optimization

We present an optimization method for minimizing a twice-continuously differentiable nonconvex function subject to twice-continuously differentiable equality constraints. The method consists of a new trust funnel scheme combined with the recently proposed TRACE algorithm that ensures convergence to an ϵ-neighborhood of the feasible region in $\mathcal{O}(\epsilon^{-3/2})$ number of iterations. The algorithm then continues computing iterates using the TRACE algorithm that aim at reducing the objective function while maintaining near feasibility. The resulting scheme has the same complexity as for the SHORT-STEP ARC algorithm. The subproblems that we employ during both phases take the objective and constraint functions into account. This is in contrast to the feasibility phase used in SHORT-STEP ARC, which temporarily ignores the objective function. The empirical performance of such an approach is illustrated through numerical results.

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

Daniel Robinson
Johns Hopkins University
daniel.p.robinson@gmail.com

Mohammadreza Samadi
Lehigh University
mos213@lehigh.edu

MS89

A Space Transformation Framework for Nonlinear Optimization

We present a space transformation framework for nonlinear optimization. Instead of tackling the problem in the original space, each iteration of this framework seeks for a trial step by modeling and approximately solving the optimization problem in another space. We establish the global convergence and worst case iteration complexity of the framework. Then we show that the framework can be specialized to a parallel space decomposition framework for nonlinear optimization, which can be regarded as an extension of the domain decomposition method for PDEs. A feature of the decomposition framework is that it incorporates the restricted additive Schwarz methodology into the synchronization phase of the method. We will illustrate how this decomposition framework can be applied to design parallel algorithms for optimization problems with or without derivatives.

Zaikun Zhang, Luis Nunes Vicente
University of Coimbra
zaikun.zhang@polyu.edu.hk, lnv@mat.uc.pt

Serge Gratton
ENSEEIHT, Toulouse, France
serge.gratton@enseeiht.fr

MS90

Universal Regularization Methods: Varying the Power, the Smoothness and the Accuracy

Adaptive cubic regularization methods have recently emerged as a credible alternative to linesearch and trust-region for smooth nonconvex optimization, with optimal complexity amongst second-order methods. Here we consider a general/new class of adaptive regularization methods, that use first- or higher-order local Taylor models of the objective regularized by a(ny) power of the step size and applied to convexly-constrained optimization problems. We investigate the worst-case complexity/global rate of convergence of these algorithms, in the presence of varying (unknown) smoothness of the objective. We find that the methods automatically adapt their complexity to the degree of smoothness of the objective, while also taking advantage of the power of the regularization step to satisfy increasingly better bounds with the order of the models. The bounds vary continuously and robustly, with respect to the regularization power, the model accuracy and the degree of smoothness.

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

Nick Gould
Numerical Analysis Group
Rutherford Appleton Laboratory
nick.gould@stfc.ac.uk

Philippe L. Toint
University of Namur, Belgium
philippe.toint@fundp.ac.be

MS90

A Line-Search Approach Deriving by Adaptive Regularization Framework Using cubics, with a Worst-Case Iteration Complexity of O(e-3/2)

We consider solving unconstrained optimization problems by means of two popular globalization techniques: trust-region (TR) algorithms and adaptive regularized framework using cubics (ARC). Both techniques require the solution of a so-called "subproblem' in which a trial step is computed by solving an optimization problem involving an approximation of the objective function, called "the model". The latter is supposed to be adequate in a neighborhood of the current iterate. In this work, we address an important practical question related with the choice of the norm for defining the neighborhood. Given a symmetric positive definite matrix M that satisfies a specific secant equation, we propose the use of the M-scaled norm – defined by $\|x\|_M = \sqrt{x^T Mx}$ for all $x \in \mathbb{R}^n$ – in both TR and ARC techniques. We show that the use of this norm induces remarkable relations between the trial step of both methods that can be used to obtain efficient practical algorithms. Using the M-scaled norm we propose a possible way to derive line-search algorithms enjoying the same worst case complexity as the ARC algorithms. The good potential of the proposed algorithms is illustrated on a set of numerical experiments.

Youssef Diouane
Institut Supérieur de lAéronautique et de lEspace (ISAE)
Toulouse, France
youssef.diouane@isae.fr

El houcine Bergou
INRA, Jouy-en-Josas, France
elhoucine.bergou@inra.fr

Serge Gratton
ENSEEIHT, Toulouse, France
serge.gratton@enseeiht.fr

MS90

Minimal Constraints Qualification that Ensure Convergence to KKT Points

The Karush-Kuhn-Tucker condition (KKT) is arguably the most important property related to optimality in nonlinear optimization and hence many optimization algorithms were developed that search for KKT points. Naturally, such algorithms can only generate sequences that conform to approximate versions of KKT. Therefore some type of continuity argument is needed to extend this approximate version of KKT to the exact version at the limit points. This talk will recall a hierarchy of approximate KKT conditions associated to different optimization methods and introduce the respective minimal conditions that must hold at the limit points in order to ensure that they are actually KKT. Such conditions are special constraint qualifications that we call *strict*. In this sense we will be able to fully characterize the minimal constraint qualifications needed to ensure the global convergence of different algorithms to KKT points. Also we will draw a complete picture of the relationship between these approximate, or sequential, optimality conditions, their respective minimal constraint qualifications and other classical constraint qualifications that appear in the literature.

Roberto Andreani
DMA-IMECC-UNICAMP
andreani@ime.unicamp.br

José Mario Martínez
University of Campinas
martinez@ime.unicamp.br

Alberto Ramos
Federal University of Parana
albertoramos@ufpr.br

Paulo J. S. Silva
University of Campinas
pjssilva@ime.unicamp.br

MS90

Monotone Properties of the Barzilai-Borwein Method

The Barzilai-Borwein method is one of the simpliest methods for unconstrained optimization, which uses the steepest descent direction with special stepsizes based on the curvature informaiton in the direction of the previous step. It is well-konwn that the BB method is much more efficient than the steepest descent method, particularly for large-scale problems. However, normally the BB method is not a monotone method. In this talk, we explore certain monotone properties of the BB method when the objective function is a convex quadratic function.

Yaxiang Yuan
Laboratory of Scientific and Engineering Computing
Chinese Academy of Science
yyx@lsec.cc.ac.cn

MS91

A Distributed Optimization Method for Large-Scale Sparse Support Vector Machines with an Application to Healthcare Problems

As we are moving into the era of big data, large-scale solutions that are computationally efficient are necessary. In this context, the distributed setting becomes very relevant. We are specifically motivated by healthcare problems, where (medical research) participants could use their mobile devices distributing storage and most importantly computational load, to develop predictive models. We present an iterative splitting distributed algorithm to find an ℓ_1 regularized Support Vector Machines (SVM) separation hyperplane with improved guarantees on the convergence rate. We compare our method to other approaches that solve the sparse SVM problem, show results on a synthetic dataset and discuss the pros and cons of each one. Lastly, we apply our framework in a real healthcare problem, where based on the de-identified Electronic Health Records of patients, we aim to predict heart-related hospitalizations in a target year. The induced sparsity in the algorithm helps us identify the important medical factors that lead to hospitalization, accounting for model interpretability, which is crucial in the medical domain.

Theodora Brisimi, Alex Olshevsky, Ioannis Paschalidis, Wei Shi
Boston University
tbrisimi@bu.edu, alexols@bu.edu, yannisp@bu.edu, wilburs@bu.edu

MS91

Decentralized Stochastic and Online Optimization

We present a decentralized primal-dual gradient method for optimizing a class of finite-sum convex optimization problem whose objective function is given by the summation of m local objective functions which are distributed over a network of m agents. In our method, each agent alternatively updates its primal and dual estimates by computing the primal and dual proximal operator, and by communicating these estimates with other agents in the network. We show that this method converges to an ϵ-optimal solution with no more than $1/\epsilon$ communication rounds when the problem is nonsmooth and stochastic.

Soomin Lee
Georgia Tech
soomin.lee@isye.gatech.edu

MS91

A Comprehensive Linear Speedup Analysis for Asynchronous Stochastic Parallel Optimization from Zeroth-Order to First-Order

Asynchronous parallel optimization received substantial successes and extensive attention recently. One of core theoretical questions is how much speedup (or benefit) the asynchronous parallelization can bring to us. This paper provides a comprehensive and generic analysis to study the speedup property for a broad range of asynchronous parallel stochastic algorithms from the zeroth order to the first order methods. Our result recovers or improves existing analysis on special cases, provides more insights for understanding the asynchronous parallel behaviors, and suggests a novel asynchronous parallel zeroth order method for the first time. Our experiments provide novel applications of the proposed asynchronous parallel zeroth order method on hyper parameter tuning and model blending problems.

Ji Liu
University of Rochester
jliu@cs.rochester.edu

MS91

Decentralized Consensus Optimization on Networks with Delayed and Stochastic Gradients

Decentralized consensus optimization has extensive applications in many emerging big data, machine learning, and sensor network problems. In decentralized computing, nodes in a network privately hold parts of the objective function and need to collaboratively solve for the consensual optimal solution of the total objective, while they can only communicate with their immediate neighbors during updates. In real-world networks, it is often difficult and sometimes impossible to synchronize these nodes, and as a result they have to use stale (and stochastic) gradient information which may steer their iterates away from the optimal solution. In this talk, we focus on a decentralized consensus algorithm by taking the delays of gradients into consideration. We show that, as long as the random delays are bounded in expectation and a proper diminishing step size policy is employed, the iterates generated by this algorithm still converge to a consensual optimal solution. Convergence rates of both objective and consensus are derived. Numerical results on some synthetic optimization problems and on real seismic tomography will also be presented.

Xiaojing Ye, Benjamin Sirb
Department of Mathematics & Statistics
Georgia State University
xye@gsu.edu, bsirb1@student.gsu.edu

MS92

Adaptive Convex Relaxations for Mixed-Integer Optimization of Gas Pipeline Dynamics

We present an adaptive partitioning method in combination with optimality-based bound tightening approaches to strengthen standard convex relaxations for dynamic pipeline flow optimization problems, which are mathematically formulated as nonconvex mixed-integer nonlinear programs (MINLP). Transient equations for pipeline flow are approximated on a coarse space-time grid and then outer-approximated with adaptively tightened convex domains. Computation time and solution accuracy for objective functions that maximize compressor efficiency or economic welfare, respectively, on meshed gas networks are evaluated and compared to outcomes obtained using primal-dual interior point methods. The approach is extended to include discrete variables that determine whether gas compressors are active in each optimization time period. Such mixed-integer problems are solved to near global-optimality using the proposed method. Efficiency and scalability of the technique makes it promising for extension to the general class of large-scale problems that involves dynamic (differential equation) constraints as well as discrete decision variables.

Harsha Nagarajan, Anatoly Zlotnik
Los Alamos National Laboratory
harsha@lanl.gov, azlotnik@lanl.gov

Fei Wu, Ramteen Sioshansi
Ohio State University

wu.1557@buckeyemail.osu.edu, sioshansi.1@osu.edu

MS92

Penalty Alternating Direction Methods for Mixed-Integer Nonlinar Optimization

We present a penalty algorithm that is based on an alternating direction method for the penalty subproblems. This work is motivated by a preceding paper ("Solving power-constrained gas transportation problems using an MIP-based alternating direction method", Comput. Chem. Eng. 2015(82)), in which a variant of the method is used to solve large-scale non-convex mixed-integer nonlinear feasibility problems from steady-state gas transport. In this talk, the extensions of the method are discussed and we give a sketch of the convergence theory for the new algorithm. The practical strength of the proposed method is demonstrated by a computational study, in which we apply the method to large-scale real-world problems from steady-state gas transport including, e.g., pooling effects. Moreover, we discuss the capabilities of the method as a primal heuristic for general mixed-integer nonlinear optimization problems and show its strength on the MINLPLib2 test set.

Martin Schmidt
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Mathematik
mar.schmidt@fau.de

Bjoern Geissler, Antonio Morsi
Friedrich-Alexander-Universität Erlangen Nürnberg,
Department Mathematik, EDOM
bjoern.geissler@math.uni-erlangen.de,
antonio.morsi@math.uni-erlangen.de

Lars Schewe
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Mathematik, EDOM
lars.schewe@math.uni-erlangen.de

MS92

MIP-Based Instantaneous Control of Mixed-Integer PDE-Constrained Gas Transport Problems

We study the transient optimization of gas transport networks including both discrete controls due to switching of controllable elements and nonlinear fluid dynamics that are described by the system of partial differential Euler equations. This combination leads to mixed-integer optimization problems subject to nonlinear hyperbolic partial differential equations on a graph. We propose an instantaneous control approach in which suitable Euler discretizations yield systems of ordinary differential equations on a graph. We show that this networked system of ordinary differential equations is well-posed and that affine-linear solutions of these systems can be derived analytically. As a consequence, we obtain finite dimensional mixed-integer linear optimization problems for every time step that can be solved to global optimality using general-purpose solvers. We illustrate our approach in practice by presenting numerical results of realistic gas transport networks.

Mathias Sirvent
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Mathematik
mathias.sirvent@fau.de

Martin Gugat
Friedrich-Alexander-University Erlangen-Nuremberg
(FAU)
Germany
martin.gugat@fau.de

Günter Leugering
University Erlangen-Nuremberg
Institute of Applied Math.
leugering@am.uni-erlangen.de

Alexander Martin
University of Erlangen-Nürnberg
alexander.martin@math.uni-erlangen.de

Martin Schmidt
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Mathematik
mar.schmidt@fau.de

David Wintergerst
Friedrich-Alexander-Universität Erlangen-Nürnberg
wintergerst@math.fau.de

MS92

Reformulating Discrete Aspects in Operative Planning of Gas and Water Networks

Operative planning in gas and water networks usually involves discrete decisions and other discrete model aspects with constraints that couple the resulting integer variables over time or space. While certain aspects are genuinely discrete, other aspects admit reformulations using continuous variables and constraints that can be helpful in reducing the combinatorial complexity of solving the overall optimization problem. We discuss several examples of such reformulations that have proved effective in apllications.

Marc C. Steinbach
Leibniz Universität Hannover
Institute for Applied Mathematics
mcs@ifam.uni-hannover.de

MS93

Relationships Between Constrained and Unconstrained Multi-Objective Optimization and Application in Location Theory

In this talk we investigate relationships between constrained and unconstrained multi-objective optimization problems. We mainly focus on generalized convex multi-objective optimization problems, i.e., the objective function is a componentwise generalized convex (e.g., quasi-convex or semi-strictly quasi-convex) function and the feasible domain is a convex set. Beside the field of location theory the assumptions of generalized convexity are found in several branches of Economics. We derive a characterization of the set of efficient solutions of a constrained multi-objective optimization problem using characterizations of the sets of efficient solutions of unconstrained multi-objective optimization problems. We demonstrate the usefulness of the results by applying it on constrained multi-objective location problems. Using our new results we show that special classes of constrained multi-objective location problems (e.g., point-objective location problems, Weber location problems and center location problems) can be completely solved with the help of algorithms for the unconstrained case. At the end of the talk, we present some information about the current development of the

MATLAB-based software tool "Facility Location Optimizer".

Christian Guenther
Martin-Luther-Universität Halle-Wittenberg
heodor-Lieser-Strae 5, 06120 Halle, Germany
christian.guenther@mathematik.uni-halle.de

Christiane Tammer
Faculty of Sciences III, Institute of Mathematics
University of Halle-Wittenberg
christiane.tammer@mathematik.uni-halle.de

MS93

An Implicit Filtering Algorithm for Derivative Free Multiobjective Optimization

We consider multiobjective optimization problems where the variables are subject to bound constraints. We assume that the objective functions are continuously differentiable but the derivatives are difficult to compute or not available. The method we present is a multiobjective variant of the implicit filtering algorithm, which can be viewed as a line search method based on a difference-approximated steepest descent direction, coupled with coordinate search. At each iteration, coordinate search with a fixed step size is performed as long as new nondominated points are encountered. Then, the computed objective function values are employed to approximate the gradients of the objective functions, a linear programming problem is solved to approximate the multiobjective descent direction, and a line search is performed along the computed direction. In case of failure of the line-search, the step size is reduced and a new iteration started. Global convergence to Pareto points is stated. Computational experiments are presented and discussed.

Guido Cocchi
Dip. Ingegneria dell'Informazione
Università degli Studi di Firenze
guido.cocchi@unifi.it

Giampaolo Liuzzi
Istituto di Analisi dei Sistemi ed Informatica
CNR, Rome, Italy
giampaolo.liuzzi@iasi.cnr.it

Alessandra Papini
Dipartimento di Energetica 'S. Stecco'
Univerity of Florence
alessandra.papini@unifi.it

Marco Sciandrone
Università degli Studi di Firenze
marco.sciandrone@unifi.it

MS93

An Iterative Approach for Multiobjective Optimization Problems with Heterogeneous Functions

In many applications of multiobjective optimization the objective functions are heterogeneous, for example regarding the evaluation time. But methods for computing efficient points of such problems do not distinguish between the objectives so far, but treat them all equally. In this presentation an idea to take into account the heterogeneity of the objective functions is introduced. For this purpose an iterative approach is considered, which is based on the trust region concept for scalar optimization problems.

Jana Thomann
TU Ilmenau
jana.thomann@tu-ilmenau.de

MS94

New Analytical and Stochastic Results on Understanding Deep Convolutional Neural Network

Abstract not available at time of publication.

Fang Han
University of Washington
fanghan@uw.edu

MS94

Probabilistic Image Models and Extensions of the Perona-Malik Filter

The Perona-Malik model has been very successful at restoring images from noisy input. In this talk, we show how the Perona-Malik model can be reinterpreted and extended using the language of Gaussian scale mixtures. Specifically, we show how the expectation-maximization EM algorithm applied to Gaussian scale mixtures leads to the lagged-diffusivity algorithm for computing stationary points of the Perona-Malik diffusion equations. Moreover, we show how mean field approximations to these Gaussian scale mixtures lead to a modification of the lagged-diffusivity algorithm that better captures the uncertainties in the restoration. Since this modification can be hard to compute in practice we propose relaxations to the mean field objective to make the algorithm computationally feasible. Our numerical experiments show that this modified lagged-diffusivity algorithm often performs better at restoring textured areas and fuzzy edges than the unmodified algorithm. As a second application of the Gaussian scale mixture framework, we show how an efficient sampling procedure can be obtained for the probabilistic model, making the computation of the conditional mean and other expectations algorithmically feasible. Again, the resulting algorithm has a strong resemblance to the lagged-diffusivity algorithm.

Dirk Lorenz
TU Braunschweig
d.lorenz@tu-braunschweig.de

Lars Mescheder
MPI Tübingen
lars.mescheder@tuebingen.mpg.de

MS94

A Family of Stochastic Surrogate Optimization Algorithms

We present a family of optimization techniques for solving large-scale structured optimization problems that typically occur in machine learning, where the objective function is either a large finite sum, or an expectation. Our approaches are based on incrementally updating a model (surrogate) of the objective. When using upper-bounds, we obtain incremental and stochastic majorization-minimization algorithms, which are shown to asymptotically provide stationary points for non-convex problems, and we recover classical convergence rates of first-order methods in the convex case. When the objective is strongly convex, we

also show that using upper bounds is not necessarily appropriate. By using instead strong convexity lower bounds, we obtain much faster convergence, both in theory and in practice. Besides, our algorithms for finite sums can be accelerated in the sense of Nesterov, providing significant improvements when the problem is ill-conditioned.

Julien Mairal
Inria Grenoble
julien.mairal@m4x.org

MS94

Expectation-Maximization Algorithms for Partially Observed Continuous-Time Markov Processes

Continuous-time Markov processes enjoy numerous applications in sciences and engineering. In many settings, multiple independent realizations of these processes are observed at unevenly spaced time points. In such scenarios, estimation of state transition rates is a nontrivial optimization problem, because these rates can be functions of many parameters related to the covariates attached to each realization of the process (e.g., each realization may correspond to a patient disease trajectory with patient-specific covariates). I will discuss expectation-maximization (EM) algorithms to solve such optimization problems in the context of finite state space models, birth-death processes, and multitype branching processes, with applications to analyses of electronic medical records and intrahost evolution of bacteria.

Vladimir Minin
University of Washington
vminin@uw.edu

MS95

Global Optimization with Orthogonality Constraints via Stochastic Diffusion on Manifolds

Orthogonality constrained problems have wide applications in many problems including p-harmonic flow, 1-bit compressive sensing, eigenvalues problems in electronic structure calculation and many others. One of the main challenges of orthogonality constraints are the non-convexity which makes these problems hard to achieve their global optimizers. In this talk, I will discuss our recent work of an global optimization method for orthogonality constrained problems. Our method is based on an idea of iterative perturbing local minimizers via a stochastic differential equation on the Stiefel manifold. We theoretically analyze the global convergence guarantee of the proposed method, and our numerical tests also validate its effectiveness.

Rongjie Lai
Rensselaer Polytechnic Institute
rongjielai@gmail.com

Zaiwen Wen, honglin yuan, xiaoyi gu
peking university
wenzw@pku,edu.cn, yhlmath@pku.edu.cn, 1300017646@pku.edu.cn

MS95

Reducing the Cost of the Fock Exchange Operator

The Fock exchange operator plays a central role in modern quantum chemistry and materials science. I will describe the recently developed adaptive compression strategy for reducing the cost of the Fock exchange operator from an optimization perspective, and discuss the progress for real materials calculations using planewaves as well as quantum chemistry basis sets.

Lin Lin
University of California, Berkeley
Lawrence Berkeley National Laboratory
linlin@math.berkeley.edu

MS95

Electronic Structure Calculation using Semidefinite Programs

Electronic structure calculation can be modelled by a variational approach where the two-body reduced density matrix (RDM) is the unknown variable. It can lead to an extremely large-scale semidefinite programming (SDP) problem. This talk will present a practically efficient second-order type method.

Zaiwen Wen
Peking University
Beijing International Center For Mathematical Research
wenzw@math.pku.edu.cn

MS95

A Newton-Krylov Method for Solving Coupled Cluster Equations

The coupled cluster approach to solving the quantum many body problem is considered one of the most accurate approaches in quantum chemistry. However, the coupled cluster equations are a system of nonlinear equations with many variables (tens of thousands to millions). These equations are currently solved by either an inexact Newton method or a quasi-Newton method called Direct Inversion in Iterative Subspace (DIIS). These methods experience difficulty in convergence in some cases. In this talk, we will describe using a Newton-Krylov method to overcome the convergence difficulty. The effectiveness of the Newton-Krylov method depends largely on the construction of an effective preconditioner for solving the Newton correction equation iteratively.

Chao Yang
Lawrence Berkeley National Lab
cyang@lbl.gov

Jiri Brabec
Czeck Academy of Sciences
jiri.brabec@jh-inst.cas.cz

Karol Kolwalski
Pacific Northwest National Laboratory
karol.kowalski@pnnl.gov

MS96

New Results for Sparse Methods for Logistic Regression and Related Classification Problems

We propose and analyze methods in logistic regression that induce sparse solutions. We introduce a novel condition number associated with the data for logistic regression that measures the degree of non-separability of the data (how close the data is to being separable), and that naturally enters the analysis and the computational properties of methods. We take advantage of very recent new results in first-order methods for convex optimization due to Bach and others, to present new computational guarantees for the performance of methods for logistic regression. In the high-dimensional regime in particular, these guarantees precisely characterize how the methods impart both data-fidelity and implicit regularization, for any dataset.

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

Paul Grigas
UC Berkeley
pgrigas@berkeley.edu

Rahul Mazumder
MIT
rahulmaz@mit.edu

MS96

Fast Algorithms for Large Scale Generalized Distance Weighted Discrimination

High dimension low sample size statistical analysis is important in a wide range of applications. In such situations, the highly appealing discrimination method, support vector machine, can be improved to alleviate data piling at the margin. This leads naturally to the development of distance weighted discrimination (DWD), which can be modeled as a second-order cone programming problem and solved by interior-point methods when the scale (in sample size and feature dimension) of the data is moderate. Here, we design a scalable and robust algorithm for solving large scale generalized DWD problems. Numerical experiments on real data sets from the UCI repository demonstrate that our algorithm is highly efficient in solving large scale problems, and sometimes can even be more efficient than the popular LIBSVM for solving the corresponding SVM problems.

Kim-Chuan Toh
National University of Singapore
Department of Mathematics
mattohkc@nus.edu.sg

Defeng Sun
Dept of Mathematics
National University of Singapore
matsundf@nus.edu.sg

Xin-Yee Lam
National University of Singapore
a0074534@u.nus.edu

J.S. Marron
University of North Carolina
marron@email.unc.edu

MS96

Stochastic First-Order Methods in Data Analysis and Reinforcement Learning

Stochastic first-order methods provide a basic algorithmic tool for online learning and data analysis In this talk, we survey several innovative applications including risk-averse optimization, online principal component analysis, and reinforcement learning. We will show that rate of convergence analysis of the stochastic optimization algorithms provide sample complexity analysis for these online learning applications.

Mengdi Wang
Princeton University
mengdiw@princeton.edu

MS96

Random Permutations Fix a Worst Case for Cyclic Coordinate Descent

Variants of the coordinate descent approach for minimizing a nonlinear function are distinguished in part by the order in which coordinates are considered for relaxation. Three common orderings are cyclic (CCD), in which we cycle through the components of x in order; randomized (RCD), in which the component to update is selected randomly and independently at each iteration; and random-permutations cyclic (RPCD), which differs from CCD only in that a random permutation is applied to the variables at the start of each cycle. Known convergence guarantees are weaker for CCD and RPCD than for RCD, though in most practical cases, computational performance is similar among all these variants. There is a certain family of quadratic functions for which CCD is significantly slower than for RCD; a recent paper of Sun and Ye has explored the poor behavior of CCD on this family. The RPCD approach performs well on this family, and this paper explains this good behavior with a tight analysis. We discuss generalizations of the analysis to quadratics in which the Hessian has a single dominant eigenvalue whose components are not too diverse.

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

Ching-Pei Lee
University of Wisconsin-Madison
ching-pei@cs.wisc.edu

MS97

Communication Avoiding Primal and Dual Block Coordinate Descent Methods

Primal and dual block coordinate descent are iterative methods for solving regularized and unregularized optimization problems. Distributed-memory parallel implementations of these methods have become popular in analyzing large machine learning datasets. However, existing implementations communicate at every iteration which, on modern data center and supercomputing architectures, often dominates the cost of floating-point computation. We evaluate the performance of new communication-avoiding variants of the primal and dual block coordinate descent for ridge regularized least squares. We test our communication-avoiding algorithms on dense datasets (both synthetic and real) to illustrate the speedups attainable on a Cray XC30 supercomputer. We evaluate the strong scaling performance on up to 12288 cores to illustrate the speedups attainable. Our results show that the communication-avoiding algorithms achieve speedups of 1.2x – 4.8x on real problems obtained from LIBSVM.

Aditya Devarakonda, Kimon Fountoulakis, James Demmel, Michael Mahoney

UC Berkeley
aditya@eecs.berkeley.edu, kfount@berkeley.edu, demmel@eecs.berkeley.edu, mmahoney@stat.berkeley.edu

MS97

AIDE: Fast and Communication Efficient Distributed Optimization

In this paper, we present two new communication-efficient methods for distributed minimization of an average of functions. The first algorithm is an inexact variant of the DANE algorithm that allows any local algorithm to return an approximate solution to a local subproblem. We show that such a strategy does not affect the theoretical guarantees of DANE significantly. In fact, our approach can be viewed as a robustification strategy since the method is substantially better behaved than DANE on data partition arising in practice. It is well known that DANE algorithm does not match the communication complexity lower bounds. To bridge this gap, we propose an accelerated variant of the first method, called AIDE, that not only matches the communication lower bounds but can also be implemented using a purely first-order oracle. Our empirical results show that AIDE is superior to other communication efficient algorithms in settings that naturally arise in machine learning applications.

Sashank Reddi
Carnegie Mellon University
sjakkamr@cs.cmu.edu

Jakub Konecny
The University of Edinburgh
kubo.konecny@gmail.com

Peter Richtarik
University of Edinburgh
peter.richtarik@ed.ac.uk

Barnabas Poczos
Carnegie Mellon University
bapoczos@cs.cmu.edu

Alex Smola
Yahoo! Research, Santa Clara, California
alex@smola.org

MS97

Speeding-Up Large-Scale Machine Learning Using Graphs and Codes

In this talk I will present three simple combinatorial ideas to speed up parallel and distributed learning algorithms. We will start off with serial equivalence in asynchronous parallel optimization, its significance, and how we can guarantee it using a recent phase transition result on graphs. We continue on the issue of stragglers (i.e., slow nodes) in distributed systems, where we will use erasure codes to robustify gradient based algorithms against delays. In our third example, we will reduce the high communication cost of data-shuffling in distributed learning, using the seemingly unrelated notion of coded caching. For all the above, we will see real world experiments that indicate how these simple ideas can significantly improve the speed and accuracy of large-scale learning.

Dimitris Papailiopoulos
UC Berkeley
dimitris@papail.io

MS97

A General Framework for Communication-Efficient Distributed Optimization

The scale of modern datasets necessitates the development of efficient distributed optimization methods for machine learning. We present a general-purpose framework for the distributed environment, CoCoA, that has an efficient communication scheme and is applicable to a wide variety of problems in machine learning and signal processing. We extend the framework to cover general non-strongly convex regularizers, including L1-regularized problems like lasso, sparse logistic regression, and elastic net regularization, and show how earlier work can be derived as a special case. We provide convergence guarantees for the class of convex regularized loss minimization objectives, leveraging a novel approach in handling non-strongly convex regularizers and non-smooth loss functions. The resulting framework has markedly improved performance over state-of-the-art methods, as we illustrate with an extensive set of experiments on real distributed datasets.

Virginia Smith
UC Berkeley
vsmith@berkeley.edu

MS98

Perturbed Iterates Analysis for Stochastic Optimization

We introduce and analyze stochastic optimization methods where the input to each update is perturbed by bounded noise. We show that this framework forms the basis of a unified approach to analyze asynchronous implementations of stochastic optimization algorithms, Using our perturbed iterate framework, we provide new analyses of the HogWild algorithm and asynchronous stochastic coordinate descent, that are simpler than earlier analyses, remove many assumptions of previous models, and in some cases yield improved upper bounds on the convergence rates. We proceed to apply our framework to develop and analyze KroMagnon: a novel, parallel, sparse stochastic variance-reduced gradient (SVRG) algorithm. We demonstrate experimentally on a 16-core machine that the sparse and parallel version of SVRG is in some cases more than four orders of magnitude faster than the standard SVRG algorithm.

Horia Mania, Xinghao Pan
University of California, Berkeley
hmania@berkeley.edu, xinghao@berkeley.edu

Dimitris Papailiopoulos
University of Wisconsin-Madison
dimitris@ece.wisc.edu

Benjamin Recht, Kannan Ramchandran
University of California, Berkeley
brecht@berkeley.edu, kannanr@eecs.berkeley.edu

Michael Jordan
University of California at Berkeley
jordan@eecs.berkeley.edu

MS98

Robustness Mechanisms for Statistical Generaliza-

tion in Large-Capacity Models

Recent successes in neural networks have demonstrated that models with an excessive numbers parameters can achieve tremendous improvements in pattern recognition. Moreover, empirical evidence demonstrates that such performance is achievable without any obvious forms of regularization or capacity control. These findings reveal that traditional learning theory fails to explain why large neural networks generalize. In this talk, I will discuss possible mechanisms to explain generalization in such large models, appealing to insights from linear predictors. I will close by proposing some possible paths towards a framework of generalization that explains contemporary experimental findings in large-scale machine learning.

Benjamin Recht
University of California, Berkeley
brecht@berkeley.edu

MS98

Iterative Optimization Methods: From Robustness to Regularization

Abstract not available at time of publication.

Lorenzo Rosasco
Instituto Italiano di Technologia and
Massachusetts Institute of Technology
lrosasco@mit.edu

MS98

Implicit Regularization Through Optimization

Abstract not available at time of publication.

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

MS99

Acceleration of Randomized Block Coordinate Descent Methods via Identification Function

We present a block coordinate descent method, based on the paper [Peter Richtárik, Martin Takáč, Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function, Math. Program. (2014)], applied to the minimization problem

$$\begin{aligned} \min_x \quad & f(x) + \psi(x) \\ \text{s.t.} \quad & l \leq x \leq u \end{aligned} \tag{1}$$

where f is a smooth function and ψ is a convex, possibly nonsmoth, function with a block coordinate separable structure.
Using the concept of identification function presented in the paper [Francisco Facchinei, Andreas Fischer, Christian Kanzow, On the accurate identification of active constraints, SIAM Journal on Optimization (1998)], we propose an identification function for (1), capable of finding the active constraints of the problem. In particular, for $\psi(x) = \lambda\|x\|_1,\ \lambda > 0$, we, additionally, obtain information about the null coordinates of stationary point of (1).
We construct a nonuniform randomized method that exploits the information of the identification function, increasing the probability of selecting the block coordinates that are active, according to this strategy.
Our numerical experimentation produced promising results, including tests of a parallel implementation of the algorithm.

Ronaldo Lopes, Paulo J. S. Silva, Sandra A. Santos
University of Campinas
ronaldolps3@gmail.com, pjssilva@ime.unicamp.br, sandra@ime.unicamp.br

MS99

A First-Order Primal-Dual Algorithm with Bregman Distance Functions

In this talk we analyze the ergodic convergence rate for a first-order primal-dual algorithm for non-smooth convex-cave saddle-point problems of the form

$$\min_x \max_y \langle Kx, y\rangle + f(x) + g(x) - h^*(y),$$

where K is a linear operator, f, h^* are convex functions with simple-to-compute proximal mappings and g is a convex function with Lipschitz continuous gradient. The presented algorithm alternates proximal gradient steps in the primal and dual variables and hence it is well suited for large scale problem as they arise in numerous imaging and machine learning applications. We present the algorithm in its most general form using proximal-like Bregman distance functions, which can have advantages for example in case of simplex constraints. It turns out that the algorithm has an ergodic $O(1/k)$ convergence rate in the most general case but it can also be accelerated (via dynamic choices of the step sizes) to achieve the optimal $O(1/k^2)$ convergence rate for problems strongly convex in the primal or dual variable. Furthermore, if the problem is strongly convex in both the primal and dual variables, it also yields an optimal linear convergence rate. We show applications of the algorithm to important problems in imaging.

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology
pock@icg.tugraz.at

Antonin Chambolle
Centre de Mathématiques Appliquées
École Polytechnique, France
antonin.chambolle@cmap.polytechnique.fr

MS99

A Two-Phase Gradient Method for Singly Linearly Constrained Quadratic Programming Problems with Lower and Upper Bounds

We propose an algorithmic framework for the solution of Quadratic Programming (QP) problems with a single linear equality constraint and bounds on the variables. Inspired by the GPCG algorithm [J.J. Moré and G. Toraldo, SIAM J. Optim., 1, 1991] for bound-constrained convex QP problems, our algorithm alternates between two phases until convergence: an identification phase, which performs spectral Gradient Projection steps until either a candidate active set is identified or no reasonable progress is made, and a minimization phase, which reduces the objective function in the working set defined by the identification phase, by applying either the Conjugate Gradient method or spectral Gradient Projection methods recently proposed for unconstrained QP (see, e.g., [R. De Asmundis, D. di Serafino, W.W. Hager, G. Toraldo and H. Zhang, Comput. Optim. Appl. 59, 2014]). Convergence results as well as numerical experiments are presented, showing the effectiveness of the proposed approach.

Daniela di Serafino
Second University of Naples
Department of Mathematics and Physics
daniela.diserafino@unicampania.it

Gerardo Toraldo
University of Naples "Federico II"
Department of Agricultural Engineering
toraldo@unina.it

Marco Viola
Sapienza Università di Roma
Dip. di Ingegneria Informatica, Automatica e Gestionale
marco.viola@uniroma1.it

MS100

Pareto Fronts with Gaussian Process Conditional Simulations

In the case of expensive or time consuming black-box simulators, efficient surrogate-based methods to perform multi-objective optimization have been proposed. Many of them rely on Gaussian Process (GP) models, that are exploited to drive the optimization to the set of optimal compromise solutions in the design and objective spaces, i.e., the Pareto set and Pareto front, respectively. Yet, when it comes to predicting the actual location of these sets, say to select promising designs interactively, the common practice of taking the predictive mean of the models may prove to be insufficient, as it does not take into account their uncertainties. To this end, we propose the use of GP conditional simulations to get realizations of Pareto sets and Pareto fronts. From the latter, based on concepts from random closed set theory, i.e., the Vorobev expectation and deviation, we obtain a continuous approximation of the Pareto front and a measure of the remaining variability of the set of non-dominated points. Then, to benefit from the structure of the Pareto set, generally a low-dimensional manifold in the design space, we train a Gaussian Process Latent Variable Model (GP-LVM) from the conditional Pareto set simulations. The combined GP-LVM and GP models enable selecting points on the estimated Pareto front interactively, and more accurately. Finally, we show how this methodology naturally adapts to accommodate noise in observations.

Mickaël Binois
University of Chicago
mbinois@uchicago.edu

Victor Picheny
MIAT Universite de Toulouse and INRA
victor.picheny@toulouse.inra.fr

MS100

Derivative-Free Methods for Hyperparameter Optimization of Neural Networks

We discuss the application of derivative-free optimization methodologies in the context of machine learning. More specifically, our main application will be determining values of the hyperparameters of deep neural networks. We will discuss different formulations for the problem, including different ways to handle categorical variables. We will present a computational evaluation to determine which formulations work better for the problem at hand, showing along the way that derivative-free methodologies can perform better than the popular random search approach, and are competitive with state-of-the-art algorithms for hyperparameter optimization algorithms.

Giacomo Nannicini
Carnegie Mellon University
Tepper School of Business
nannicini@us.ibm.com

MS100

A Hybrid Framework for Constrained Derivative-Free Optimization Combining Accelerated Random Search with a Trust Region Approach

The Accelerated Random Search (ARS) algorithm by Appel, Labarre, and Radulovic (2004) has been proven to converge to the global minimum faster than the classic Pure Random Search algorithm. This talk presents a hybrid method that uses a derivative-free trust-region algorithm in conjunction with an extension of the Accelerated Random Search (ARS) algorithm to constrained optimization. Subsets of sample points generated by ARS are passed to the trust-region algorithm, which is then used to search for local minima. In the numerical experiments, Constrained ARS is combined with CONORBIT (Regis and Wild 2016), which is a trust-region method that uses interpolating Radial Basis Function (RBF) models, and the resulting hybrid algorithm is compared with alternative methods on a series of test problems for constrained optimization.

Rommel G. Regis
Saint Joseph's University
rregis@sju.edu

MS100

Bayesian Optimization under Mixed Constraints with a Slack-Variable Augmented Lagrangian

Bayesian optimization methods sequentially update a posterior distribution over the objective function. To address blackbox-constrained optimization problems, we employ augmented Lagrangian-based merit functions in order to define a sampling criterion. We show how the form of the merit function is critical for sampling efficiency. Numerical results show that the resulting method may be especially powerful when there are several disjoint local minimizers.

Stefan Wild
Argonne National Laboratory
wild@mcs.anl.gov

MS101

Scaling Up Gauss-Newton Methods for Expensive Least-Squares Problems

Nonlinear least squares problems are used for a range of important scientific applications such as data fitting, data assimilation for weather forecasting and climate modelling. A typical feature of such problems is that they involve a huge amount of data, making them very high dimensional and extremely expensive to solve. Similar big data problems also arise in machine learning where the optimization is often performed using randomized block-coordinate gradient descent methods. Inspired by such approaches, we present a randomized block-coordinate adaptation for nonlinear least squares problems. We provide global convergence guarantees for our approach and encouraging numerical results.

Jaroslav Fowkes
Emirates Data Science Laboratory
University of Oxford
jaroslav.fowkes@maths.ox.ac.uk

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

MS101

New Approaches for Global Optimization Methods

We present some dimensionality reduction techniques for global optimization algorithms, in order to increase their scalability. Inspired by ideas in machine learning, and extending the approach of random projections in Zhang et al (2016), we present some new algorithmic approaches for global optimisation with theoretical guarantees of good behaviour and encouraging numerical results.

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

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

MS101

Estimation Performance and Convergence Rate of the Generalized Power Method for Phase Synchronization

The problem of phase synchronization consists of estimating a collection of phases based on noisy measurements of relative phases, and has numerous applications. Motivated by its order-optimal estimation error in the sense of Cramer-Rao bound, two approaches have been recently developed to compute the maximum likelihood estimator (MLE). One is the semidefinite relaxation (SDR) and the other is to directly solve the non-convex quadratic optimization formulation of the MLE by the generalized power method (GPM). In this paper, we focus on the latter approach. Our contribution is twofold. First, we bound the rate at which the estimation error decreases and use this bound to show that all iterates—not just the MLE—achieve an estimation error of the same order as the Cramer-Rao bound. This implies that one can terminate the GPM at any iteration and still obtain an estimator with theoretically guaranteed estimation error. Second, we show that under the same assumption on the noise power as that for the SDR method, the GPM will converge to the MLE at a linear rate with high probability. This answers a question raised by Boumal and shows that GPM is competitive in terms of both theoretical guarantees and numerical efficiency with the SDR method. As a by-product, we give an alternative proof of a result of Boumal, which asserts that every second-order critical point of the aforementioned non-convex quadratic optimization is globally optimal in a certain noise regime.

Man Chung Yue, Huikang Liu
Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
mcyue214@gmail.com, hkliu@se.cuhk.edu.hk

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

MS102

Convergence of Asynchronous Algorithms with Unbounded Delay

Asynchronous algorithms have the potential to vastly speed up parallel computations by eliminating the idle time in which computing nodes wait for the updates of other nodes. Their advantages may be even greater at scale, where synchronization may be the dominant computational cost. However proving convergence of asynchronous algorithms can be complicated. In this talk we prove the convergence of a very general asynchronous-parallel algorithm under unbounded delays, which can be either stochastic or deterministic. This is done using novel Lyapunov function techniques, which can be generalized to other convergence analyses.

Robert R. Hannah
University of California, Los Angeles
RobertHannah89@gmail.com

Wotao Yin
University of California, Los Angeles
Department of Mathematics
wotao.yin@gmail.com

MS102

Accelerating Optimization under Uncertainty via Online Convex Optimization

We consider two paradigms that account for uncertainty in optimization models: robust optimization (RO) and joint estimation-optimization (JEO). We examine recent developments on efficient and scalable iterative first-order methods for these problems, and show that these iterative methods can be viewed through the lens of online convex optimization (OCO). Our applications of interest present further flexibility in OCO via three simple modifications to standard OCO assumptions: we introduce two new concepts of weighted regret and online saddle point problems and study the possibility of making lookahead (anticipatory) decisions. We demonstrate how these new flexibilities are instrumental in exploiting structural properties of functions, and provide improved convergence rates for iterative RO and JEO methods.

Nam Ho-Nguyen
Carnegie Mellon University
hnh@andrew.cmu.edu

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

MS102

Faster Convergence Rates for Subgradient Methods under an Error Bound Condition

The purpose of this work is to provide new analyses of several subgradient methods for convex functions satisfying an error bound condition. Under mild assumptions this error bound condition is equivalent to the well-known Kurdukya-Łojasiewicz (KL) inequality and is weaker than uniform convexity. To this end, there are two main contributions. First, new nonergodic, nonasymptotic, convergence rates are provided for the classical subgradient method under the common scenario of decaying but nonsummable stepsizes. Second, for a constant but sufficiently small stepsize, it is shown that the subgradient method achieves linear convergence up to a certain tolerance level that depends on the stepsize and other problem parameters. While this was previously known for smooth convex functions satisfying the Polyak-Łojasiewicz inequality, we extend it to the more general KL inequality and do not require smoothness. When the KL exponent is less than or equal to 1/2 we show that for a particular choice of decaying stepsize the subgradient method achieves the same convergence rate as a recently proposed restarted method. We also derive new convergence rates for stochastic problems.

Patrick Johnstone, Pierre Moulin
University of Illinois at UrbanaChampaign
prjohns2@illinois.edu, moulin@ifp.uiuc.edu

MS102

QuickeNing: A Generic Quasi-Newton Algorithm for Faster Gradient-Based Optimization

We propose an approach to accelerate gradient-based optimization algorithms by giving them the ability to exploit curvature information using quasi-Newton update rules. The proposed scheme, called QuickeNing, is generic and can be applied to a large class of first-order methods such as incremental and block-coordinate algorithms; it is also compatible with composite objectives, meaning that it has the ability to provide exactly sparse solutions when the objective involves a sparsity-inducing regularization. QuickeNing relies on limited-memory BFGS rules, making it appropriate for solving high-dimensional optimization problems; with no line-search, it is also simple to use and to implement. Besides, it enjoys a worst-case linear convergence rate for strongly convex problems. We present experimental results where QuickeNing gives significant improvements over competing methods for solving large-scale high-dimensional machine learning problems.

Hongzhou Lin
Microsoft-Inria
hongzhou.lin@inria.fr

Julien Mairal
Inria Grenoble
julien.mairal@m4x.org

Zaid Harcahoui
University of Washington
zaid.harchaoui@gmail.com

MS103

Necessary Optimality Conditions for Some Nonconvex Facility Location Problems

The problem of locating a new facility with simultaneous consideration of existing attraction and repulsion points is a problem with great practical relevance e.g. in the fields of economy, city planning or industrial design. Unfortunately, the consideration of negative weights makes this problem in general a nonconvex one, so none of the established algorithms for location problems are able to solve it. We will therefore present a new approach to derive necessary optimality conditions for such problems using the nonconvex subdifferentials by Ioffe and Kruger/Mordukhovich. While there are many strong theoretical results on these subdifferentials, it is rarely possible to explicitly calculate them or use them for applications. After giving a brief review on definition, properties and calculus of the mentioned subdifferentials we will precisely calculate the corresponding subdifferentials for certain distance functions. By taking advantage of the special structure of the problems we can then derive necessary optimality conditions for some instances of scalar semi-obnoxious facility location problems. Furthermore, we will present some new scalarization results and use them to establish necessary optimality conditions for multicriteria semi-obnoxious facility location problems. At the end of the talk, we will give an outlook on open questions and possible future developments.

Marcus Hillmann
Martin-Luther-Universität Halle-Wittenberg
Theodor-Lieser-Strae 5, 06120 Halle, Germany
marcus.hillmann@mathematik.uni-halle.de

MS103

Nonsmooth Optimization Framework for Elastography Inverse Problem of Tumor Identification

This talk will focus on a nonsmooth formulation of the inverse problem of identifying a variable parameter in mixed variational problems. We will talk about various optimization formulations for this inverse problem including a recently introduced convex modified output-least-squares, a nonconvex output-least-squares, among others. We will present some new stability estimates for the inverse problem under data perturbation using the OLS and the MOLS formulations. Using techniques from multi-objective optimization, we will shed some light onto a proper selection of the optimal regularization parameter. We will present applications of our theoretical results to the elastography inverse problem of tumor identification. Numerical examples will be given.

Baasansuren Jadamba
Rochester Institute of Technology
bxjsma@rit.edu

MS103

Stable Identification in Ill-Posed Variational Problems

In this talk, we will discuss a non-regular PDE-constrained optimization problem with scalar as well as vector objectives. By employing a constructible representation of the constraint cone, we will devise a new family of dilating cones which is closely related to the Henig dilating cone, and use it to introduce a family of conically regularized problems. We will give new stability estimates for the regularized problems in terms of the regularization parameter. We will apply our results to L1-constrained least-squares problems. This is a joint work with Prof. Miguel Sama.

Akhtar Khan
Rochester Institute of Technology
aaksma@rit.edu

MS103

Pontryagin Maximum Principle for Optimal Con-

trol on Banach Manifolds

We prove the Pontryagin Maximum Principle for optimal control problems in which the state evolves on a Banach manifold, introducing and illustrating the method of Lagrangian charts for analysis of geometric control problems. In addition we demonstrate connections between a type of controllability, exact penalization of constraints, and normality of extremals.

Robert Kipka
Kent University,
North Canton, OH 44720, USA
rkipka@kent.edu

Yuri Ledyaev
Western Michigan University
yuri.ledyaev@wmich.edu

MS104

Numerical Experiments on Dynamically Choosing Cutting Models and Scalings in Bundle Methods

Bundle methods are frequently employed for minimizing the sum of convex functions. Large scale problems of this type arise e.g. in Lagrangian relaxation of resource constraints that couple a multitude of objects involved in some scheduling task or in scenario based stochastic programming applications. If a few objects or scenarios turn out to be critical in this process, it may be worth to provide more detailed cutting models for those while the others are subsumed in a common model. Which ones are critical is typically not known in advance and may well change throughout the optimization process. We report on experiments within the callable library ConicBundle for dynamically identifying critical parts and adapting the bundle choices and scaling information accordingly.

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

MS104

On Solving the Quadratic Shortest Path Problem

The goal of the quadratic shortest path problem (QSPP) is to find a path in a digraph such that the sum of weights of arcs plus the sum of interaction costs over all pairs of arcs on the path is minimized. An instance of the quadratic shortest path problem (QSPP) is said to be linearizable if there exists an instance of the shortest path problem (SPP) on the same graph such that the cost of each s-t path in both problems is the same. Although the linearization problem for quadratic assignment problem is well studied, checking the linearizability of the QSPP is still an open question. In this talk, we present an algorithm that examines whether a QSPP instance on the directed grid graph is linearizable. This algorithm has complexity $\mathcal{O}(p^3q^2 + p^2q^3)$ where $p, q \geq 2$ are the size of the directed grid graph. We will also consider other special cases of the QSPP.

Hao Hu
Tilburg University
h.hu@tilburguniversity.edu

MS104

Computational Study of Some Valid Inequalities for k-Way Graph Partitioning

We consider the maximum k-cut problem that consists in partitioning the vertex set of a graph into k subsets such that the sum of the weights of edges joining vertices in different subsets is maximized. We focus on identifying effective classes of inequalities to tighten the semidefinite programming relaxation. We carry out an experimental study of four classes of inequalities from the literature: clique, general clique, wheel and bicycle wheel. We considered multiple combinations of these classes and tested them on both dense and sparse instances for different values of k. Our computational results suggest that the bicycle wheel and wheel are the strongest inequalities for $k = 3$, and that for greater k, the wheel inequalities are the strongest by far. Furthermore, we observe an improvement in the performance for all choices of k when both bicycle wheel and wheel are used, at the cost of substantially increased CPU time on average when compared with using only one of them.

Vilmar Rodrigues de Sousa, Miguel F. Anjos
Mathematics and Industrial Engineering & GERAD
Polytechnique Montreal
vilmar.rodrigues-de-sousa@polymtl.ca,
anjos@stanfordalumni.org

Sebastien Le Digabel
Ecole Polytechnique de Montreal
Sebastien.Le.Digabel@gerad.ca

MS104

Graph Bisection Revisited

The graph bisection problem (GBP) is the problem of partitioning the vertex set of a graph into two sets of given sizes such that the total weight of edges joining these sets is minimized. In this talk we present a semidefinite programming relaxation for the GBP with a matrix variable of order n, where n is the order of the graph. Our relaxation is equivalent to the currently strongest semidefinite programming relaxation by Wolkowicz and Zhao from 1999, with matrices of order $2n$. To further improve our relaxation, we impose valid inequalities. Our strengthened SDP relaxation outperforms all previously considered SDP bounds for the GBP, including those tailored for highly symmetric graphs.

Renata Sotirov
Tilburg University
r.sotirov@uvt.nl

MS105

From Clifford Algebras and Rigidity Theory to Large Completely Positive Semidefinite Rank

Is the set of quantum correlations closed? This is an important question in quantum information theory and we use it as a motivation for this talk. We approach this question by looking at a related object: the cone of completely positive semidefinite matrices. An $n \times n$ matrix A is called completely positive semidefinite (cpsd) if there exist positive semidefinite $d \times d$ matrices $X_1, \dots, X_n$ such that $A = (\mathrm{Tr}(X_iX_j))$. The smallest such d is called the cpsd-rank of A. If there exists an upper bound on this rank in terms of n then the cone of completely positive semidefinite matrices is closed. Since the set of quantum correlations forms an affine slice of the cpsd cone this would settle our motivating question. We show that if such an upper bound exists then it has to be at least exponential in n. For this we use the seminal work of Tsirelson, connecting bipartite correlations to quantum correlations, and the above mentioned connection between quantum correlations and cpsd matrices. Bipartite correlations arise as a projection of the more commonly known correlation matrices. We proceed to show the optimality of our construction by connecting recent results from rigidity theory to the work of Tsirelson.

Sander Gribling, David de Laat
Centrum Wiskund & Informatica, The Netherlands
s.j.gribling@cwi.nl, mail@daviddelaat.nl

Monique Laurent
CWI, Amsterdam
and Tilburg University
M.Laurent@cwi.nl

MS105

Quantum and Non-Signalling Graph Isomorphisms

We introduce a two-player nonlocal game, called the (G, H)-*isomorphism game*, where classical players can win with certainty if and only if the graphs G and H are isomorphic. We then define the notions of quantum and non-signalling isomorphism, by considering perfect quantum and non-signalling strategies for the (G, H)-isomorphism game, respectively. First, we prove that non-signalling isomorphism coincides with the well-studied notion of fractional isomorphism. Second, we show that quantum isomorphism is equivalent to the feasibility of two polynomial systems in non-commuting variables, obtained by relaxing the standard integer programming formulations for graph isomorphism to Hermitian variables. On the basis of this correspondence, we show there exist graphs that are quantum isomorphic but not isomorphic.

David Roberson
Nanyang Technological University
davideroberson@gmail.com

MS105

Some Problems Concerned with Graph Decompositions

We consider various computational questions arising from the decomposition of (combinatorial) Laplacian matrices of graphs as convex combinations of Kronecker products. The questions have been originally proposed in the context of quantum information theory, in connection with a special case of the separability problem for density matrices. While the general problem is known to be NP-hard, the restricted one is feasible in a number of cases, but its complexity has not been yet determined.

Simone Severini
Department of Computer Science
University College London
simoseve@gmail.com

MS105

Quantum Correlations: Conic Formulations, Dimension Bounds and Matrix Factorizations

In this talk we focus on the sets of two-party correlations generated from a Bell scenario involving two spatially separated systems with respect to various physical models. We show that the sets of quantum and classical correlations can be expressed as projections of affine sections of the completely positive semidefinite (cpsd) cone and the completely positive cone, respectively. Furthermore, the smallest size of a quantum representation for a quantum correlation is equal to the smallest cpsd-rank over all its cpsd completions. This correspondence allows to set up a dictionary between properties of quantum correlations and features the cpsd cone and additionally, between dimensionality questions concerning quantum correlations and algebraic properties of the cpsd-rank. My goal in this talk is to give a brief summary of the most important results obtained in both directions.

Antonios Varvitsiotis
Nanyang Technological University, Singapore
Centre for Quantum Technologies, Singapore
avarvits@gmail.com

MS106

Parallel Subspace Correction Method for Strongly Convex Problem

We investigate a parallel subspace correction (PSC) framework for strongly convex optimization based domain decomposition method. At each iteration, the algorithm solve subproblems simultaneously to derive directions on subspaces and take steps sizes along the directions to find a new point. We establish the linear convergence rate of our algorithm under some mild condition about inexact solution of the subproblem, and design a new step size strategy which can both guarantee the sufficient reduction of the objective function and the uniform upper bound of the step sizes. The numerical comparison about with other existing PSC methods shows the potential of our algorithm. Moreover, we compare our method with the BB method for solving large scale strongly problem from PDE which shows that our method is very efficient and powerful.

Qian Dong
Academy of Mathematics and Systems Science, Chinese Academy
dongqian@nssc.ac.cn

Xin Liu
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
liuxin@lsec.cc.ac.cn

Zaiwen Wen
Peking University
Beijing International Center For Mathematical Research
wendouble@gmail.com

Ya-Xiang Yuan
Chinese Academy of Sciences
Inst of Comp Math & Sci/Eng, AMSS
yyx@lsec.cc.ac.cn

MS106

Distributed Algorithms for Orthogonal Constrained Optimization Problem

To construct distributed approach for solving orthogonal constrained optimization problems is usually regarded as impossible mission, due to the low scalability of orthogonalization procedure. In this talk, we propose a Jacobi type column-wise block coordinate descent method for solving a class of orthogonal constrained optimization problems, and establish the global iterate convergence to stationary point of our proposed approach. A distributed algorithm is consequently implemented. Numerical experiments illustrate that the new algorithm has brilliant performance and high scalability in solving discretized Kohn-Sham total energy minimization problems.

Xin Liu, Bin Gao
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
liuxin@lsec.cc.ac.cn, gaobin@lsec.cc.ac.cn

Ya-Xiang Yuan
Chinese Academy of Sciences
Inst of Comp Math & Sci/Eng, AMSS
yyx@lsec.cc.ac.cn

MS106

New Gradient Methods Improving Yuan's Step-Size

We propose a new framework to generate stepsizes for gradient methods by improving Yuan's stepsize. By adopting different criterions, we propose several new gradient methods. For 2-dimensional convex quadratic problems, we prove that the new methods have either finite terminations or converges R-superlinearly. For general n-dimensional convex quadratic problems, we prove that the new methods converges R-linearly. Numerical experiments show that the new methods enjoys lower complexity and outperforms the compared gradient methods.

Cong Sun
Beijing University of Posts and Telecommunications
suncong86@bupt.edu.cn

Jinpeng Liu
Beihang University
610147729@qq.com

MS106

Total Least Squares with Tikhonov Regularization: Hidden Convexity and Efficient Global Algorithms

The Tikhonov regularized total least squares (TRTLS) is a non-convex optimization problem dealing with an overestimated system of linear equations. A standard approach for (TRTLS) is to reformulate it as a problem of finding a zero point of some decreasing concave nonsmooth univariate function such that the classical bisection search and Dinkelbach's method can be applied. In this paper, by exploring the hidden convexity of (TRTLS), we reformulate it as a new problem of finding a zero point of a strictly decreasing, smooth and concave univariate function. This allows us to apply the classical Newtons method to the reformulated problem, which converges globally to the unique root with an asymptotic quadratic convergence rate. Moreover, at every iteration of the Newtons method, no optimization subproblem such as the extended trust-region subproblem is needed to evaluate the new univariate function value as it has an explicit expression. Promising numerical results based on the new algorithm are reported.

Meijia Yang
Beihang University
15201644541@163.com

Yong Xia
School of Mathematics and System Sciences
Beihang University
yxia@buaa.edu.cn

Jiulin Wang
Beihang University
buaa1109wjl@163.com

Jiming Peng
University of Houston
jopeng@uh.edu

MS107

Taylor Approximation for PDE-Constrained Optimal Control Problems under High-Dimensional Uncertainty

In this talk, we present an efficient method based on Taylor approximation for PDE-constrained optimal control problems under high-dimensional uncertainty. The computational complexity of the method does not depend on the nominal but only on the intrinsic dimension of the uncertain parameter, thus the curse of dimensionality is broken for intrinsically low-dimensional problems. Further correction for the Taylor approximation is proposed, which leads to an unbiased evaluation of the statistical moments in the objective function. We apply our method for a turbulence model with infinite-dimensional random viscosity.

Peng Chen
UT Austin
peng@ices.utexas.edu

Omar Ghattas
The University of Texas at Austin
omar@ices.utexas.edu

Umberto Villa
University of Texas at Austin
uvilla@ices.utexas.edu

MS107

Robust Gradient-Based Shape Optimization for CFD Applications

With advances in numerical techniques and increase of available computational power, Computational Fluid Mechanics has evolved from an analysis tool towards a design tool. Despite the widespread use of optimization algorithms to design engineering parts, these are mostly employed on deterministic problems where the operating and manufacturing conditions are well known. In reality however, operational and manufacturing conditions are seldom deterministic, e.g. wind direction and manufacturing tolerances. For such problems, a deterministic optimization approach may result in an optimal that is very sensitive to the uncertainties, i.e., lacks robustness, thus leading to a lower overall performance. In a robust optimization approach, uncertainties are accounted by introducing the variance of the objective as a second objective. This approach however, suffers from the curse of dimensionality. In this context we developed a gradient based optimization technique using adjoint methods combined with Polynomial Chaos Expansions (PCE) to efficiently calculate the gradients (mean objective and variance). These gradients are then combined, using weighted averaging, to find the best solution for each problem. The proposed approach is applied to shape optimization of internal and external flows. Different weighting factors allowed to control the level of robustness of each design. Overall a more robust design led to a worse mean performance. For this cases a Paretto front is obtained.

João Duarte Carrilho Miranda
Vrije Universiteit Brussel
joao.miranda@vub.ac.be

Francesco Contino
Vrije Universiteit Brussel
Brussels University (VUB)
fcontino@vub.ac.be

Chris Lacor
Vrije Universiteit Brussel
chris.lacor@vub.ac.be

MS107

Robustness Measures for Robust Design with Multiple Objectives

Engineering design often involves the optimization of different competing objectives. Here, the aim is to find a set of solutions that fulfill the concept of Pareto optimality. A further significant step to realistic multi-objective designs is to take into account uncertainties affecting the system for finding robust optimal solutions. In a previous work, multi-objective optimization and robust design were combined using the epsilon-constraint approach together with the non-intrusive polynomial chaos approach in the context of aerodynamic shape optimization. The computational effort for solving the sequence of resulting single-objective optimization problems was reduced by using a one-shot approach suited for constrained optimization problems as well as dimension-adaptivesparse grids. The aim is to extend the existing approach and investigate on robustness measures in the context of multi-objective optimization. Apart from the use of common stochastic quantities like mean and variance, we propose the use of a signed distance function as an additional constraint for optimization to account for average losses in the multi-objective space. The use of more sophisticated robustness measures can make the search for a global optimum more difficult and more computationally expensive. As a result, a global search is conducted on a surrogate model that is built using a Gaussian process model. The model is refined for optimization using an expected improvement method.

Lisa Kusch
TU Kaiserslautern
lisa.kusch@scicomp.uni-kl.de

Lionel Mathelin
LIMSI - CNRS, France
mathelin@limsi.fr

Nicolas R. Gauger
TU Kaiserslautern
nicolas.gauger@scicomp.uni-kl.de

MS107

Optimal Sensor Placement Subject to Efficiency and Robustness Constraints

We consider the problem of optimally placing sensor networks for parabolic stochastically perturbed processes under constraints involving the robustness of measurements and efficiency criteria. The mathematical formulation arises from applications in engineering and science that include location of thermostats in large buildings and industrial processes. We formulate an optimization problem where the Riccati equation is a constraint, the design variables are sensor locations, and the objective functional involves the trace of the solution to the Riccati equation. In addition, we consider constraints associated with robustness and efficiency: The robustness criterion is formulated in terms of perturbations to the differential operator of the underlying process and the efficiency one involves performance of prescribed sensor locations. We address well-posedness of the problem, and provide numerical tests.

Carlos N. Rautenberg
Humboldt-University of Berlin
carlos.rautenberg@math.hu-berlin.de

MS108

Dual Extrapolation for Solving Variational Inequalities with a Linear Optimization Oracle

We present a novel method for solving variational inequalities that relies only on the ability to solve simple (e.g., linear optimization) subproblems over the region of interest. The main idea of the method is to apply a generalization of the Frank-Wolfe method to solve the subproblems within the "dual extrapolation" method for variational inequalities due to Nesterov. By combining analyses of the dual extrapolation method and the Frank-Wolfe method, we show that our method is able to achieve a solution with accuracy ϵ using at most: (i) $O(1/\epsilon^4)$ simple (e.g., linear optimization) subproblem oracle calls in the case of monotone operators with bounded variation, and (ii) $O(1/\epsilon^2)$ simple (e.g., linear optimization) subproblem oracle calls in the case of Lipschitz continuous operators. We also discuss extensions and related results that apply to non-smooth convex optimization.

Paul Grigas
UC Berkeley
pgrigas@berkeley.edu

Yurii Nesterov
CORE
Universite catholique de Louvain
yurii.nesterov@uclouvain.be

MS108

Relatively-Smooth and Relatively-Continuous Convex Optimization by First-Order Methods, and Applications

The usual approach to developing and analyzing first-order methods for smooth or non-smooth convex optimization assumes that the gradient or the function is uniformly smooth with some Lipschitz constant L or M, respectively. However, in many settings the convex function $f(\cdot)$ satisfies neither of these two assumptions– for example, $f(x) := -\ln\det(HXH^T)$ in D-optimal design, or $f(w) := \frac{1}{n}\sum_{i=1}^{n}\max\{0, 1 - y_i x_i^T w\} + \frac{\lambda}{2}\|w\|_2^2$ in support vector machines. Herein we develop and present a different way to define smoothness, strong convexity and "uniform" continuity that is based on a user-specified "reference function" $h(\cdot)$ that should be computationally tractable; and we show that most convex functions are smooth (or continuous) with respect to a corresponding relatively-simple reference function $h(\cdot)$. We extend two standard algorithms – the prox gradient scheme and the dual averages scheme – to our new smooth setting, and we extend the subgradient scheme to our new non-smooth setting, with associated computational guarantees. Last of all, we apply our new approach to develop a new first-order method for the D-optimal design problem, with associated computational complexity analysis. This is a joint work with Robert Freund (MIT) and Yurii Nesterov (UCL).

Haihao Lu
MIT
Operations Research Center
haihao@mit.edu

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

Yurii Nesterov
CORE
Universite catholique de Louvain
yurii.nesterov@uclouvain.be

MS108

Detecting Communities by Voting Model

In this talk, we analyze a voting model based on random preferences of participants. Each voter can choose a party with certain probability, which depends on the divergence between his personal preferences and the position of the party. Our model represents a rare example of a community detection (or, clustering) model with unique equilibrium solution. This is achieved by allowing the flexible positions for the parties (centers of clusters). We propose an efficient algorithm for finding the equilibrium state with linear rate of convergence. It can be interpreted an alternating minimization scheme. Again, this is a rare example when such a scheme has provable global rate of convergence.

Yurii Nesterov
CORE
Universite catholique de Louvain
yurii.nesterov@uclouvain.be

MS108

Sample Complexity Analysis for Low-Order Optimization Algorithms

In this talk we discuss sample complexities for various zeroth-order algorithms for (stochastic) convex optimization. Essentially, the analysis is meant to understand how many times one will need to call the (stochastic) zeroth-order oracle in order to get an ϵ optimal solution in expectation. Such algorithms and analysis are important for various Bayesian optimization models and applications arising from revenue management. We shall also extend our discussions to include certain non-convex models.

Shuzhong Zhang
Department of Industrial & Systems Engineering
University of Minnesota
zhangs@umn.edu

MS109

Prioritizing Hepatitis C Treatment Decisions in U.S. Prisons

Hepatitis C virus (HCV) prevalence in prison systems is ten times higher than the general population, and hence prison systems offer a unique opportunity to control the HCV epidemic. New treatment drugs are effective, but providing treatment to all inmates is prohibitively expensive, which precludes universal HCV treatment in prison systems. As such, current practice recommends prioritizing treatment in prisons; however, there is controversy about how prioritization decision should be made. We propose a restless bandit modeling framework to support hepatitis C treatment prioritization decisions in U.S. prisons. We first prove indexability and derive several structural properties of the well-known Whittle's index, based on which, we derive a closed-form expression for Whittles index for patients with advanced liver disease. From the interpretation of this closed-form expression, we anticipate that the performance of the Whittle's index would degrade as the treatment capacity increases; and we further propose a capacity-adjusted closed-form index policy. We parameterize and validate our model using real-world data from Georgia state prison system and published studies. We test the performance of our proposed policy using a detailed, clinically-realistic simulation model and show that our proposed policy can significantly improve the overall effectiveness of the hepatitis C treatment programs in prisons compared several benchmark policies, including the Whittle's index policy.

Turgay Ayer
Georgia Institute of Technology
ayer@isye.gatech.edu

Can Zhang, Anthony Bonifonte
Georgia Tech
czhang2012@gatech.edu, abonifon@gatech.edu

Anne Spaulding
Emory University
aspauld@emory.edu

Jagpreet Chhatwal
Harvard Medical School
jagchhatwal@mgh.harvard.edu

MS109

Optimized Treatment Schedules for Chronic Myeloid Leukemia

Over the past decade, several targeted therapies have been developed to treat Chronic Myeloid Leukemia (CML). Despite an initial response to therapy, drug resistance remains a problem for some CML patients. Recent studies have shown that resistance mutations that preexist treatment can be detected in a substantial number of patients, and that this may be associated with eventual treatment failure. One proposed method to extend treatment efficacy is to use a combination of multiple targeted therapies. However, the design of such combination therapies (timing, sequence, etc.) remains an open challenge. In this work we mathematically model the dynamics of CML response to combination therapy and analyze the impact of combination treatment schedules on treatment efficacy in patients with preexisting resistance. We then propose an optimization problem to find the best schedule of multiple therapies based on the CML evolution according to our ordinary differential equation model. The resulting optimization problem contains ordinary differential equation constraints and integer variables. By exploiting the problem structure, we are able to find near optimal solutions within an hour. We determine optimal combination strategies that maximize time until treatment failure, using parameters estimated from clinical data in the literature. This is joint work with Jasmine Foo, Kevin Leder, Junfeng Zhu from the University of Minnesota and David Dingli from the Mayo Clinic.

Qie He
University of Minnesota
qhe@umn.edu

MS109

Response-Guided Dosing

Within the broad field of personalized medicine, there has been a recent surge of clinical interest in the idea of response-guided dosing. Roughly speaking, the goal is to devise strategies that administer the right dose to the right patient at the right time. We will present stochastic models that attempt to formalize such optimal dosing problems. Theoretical results about the structure of optimal dosing strategies and associated solution methods rooted in convex optimization, stochastic dynamic programming, robust optimization, and Bayesian learning will be described. Computational results on rheumatoid arthritis will be discussed.

Jakob Kotas
University of Portland
jkotas@uw.edu

Archis Ghate
University of Washington
archis@uw.edu

MS110

Mesh Adaptive Direct Search Algorithms for Multifidelity Optimization Problems

Multifidelity optimization occurs in engineering design when multiple engineering simulation codes are available at different levels of fidelity. An example of this might be in aerodynamic optimization of the shape of a wing, in which the computation of aerodynamic quantities, such as lift and drag, can be computed using a full Navier-Stokes solver, or an Euler solver, or a linearized potential code. High fidelity simulations are more accurate, but also more computationally expensive. The goal of this work is the design of algorithms that optimize with respect to the high-fidelity simulation, but exploit the use of lower fidelity codes as much as possible. Two new surrogate-based mesh adaptive direct search (MADS) algorithms will be presented, in which interpolating surrogates are constructed and updated from previously evaluated iterates of the algorithm to speed convergence. The first algorithm employs a recursive Search step that optimizes a surrogate function constructed from the next lower fidelity level simulation augmented with an interpolating surrogate that accounts for the difference between adjacent levels of fidelity. The second approach is an augmentation of the optimization problem, in which the fidelity level is incorporated as a variable in the problem, and a relaxable constraint is added to force the solution to be at the highest level of fidelity. Some preliminary numerical results are presented.

Mark Abramson
Brigham Young University
abramson@mathematics.byu.edu

MS110

Hybrid Derivative-Free Methods for Composite Nonsmooth Optimization

In this talk, the combination of global solvers and derivative-free trust-region algorithms is considered for minimizing a composite function $\Phi(x) = f(x) + h(c(x))$, where f and c are smooth and h is convex but may be non-smooth. Global convergence results and worst-case complexity bounds are discussed. The performance of the algorithms is illustrated considering the parameter estimation of biochemical models.

Geovani N. Grapiglia
Federal University of Parana
Brazil
grapiglia@ufpr.br

MS110

Manifold Sampling for Piecewise Linear Nonconvex Optimization

We develop a manifold sampling algorithm for the unconstrained minimization of a nonsmooth composite function $h \circ F$ when h is a nonsmooth, piecewise linear function and F is smooth but expensive to evaluate. The trust-region algorithm classifies points as in the domain of h as belonging to different manifolds, and uses this knowledge when computing search directions. Since h is known, classifying objective manifolds using only the function values F_i is simple. We prove that all cluster points of the sequence of algorithm iterates are Clarke stationary; this holds though points evaluated by the algorithm are not assumed to be differentiable and when only approximate derivatives of F are available. Numerical results show that manifold sampling using zero-order information is competitive with gradient sampling algorithms that are given exact gradient values.

Jeffrey Larson
Argonne National Laboratory
jmlarson@anl.gov

Kamil Khan
Massachusetts Institute of Technology
Department of Chemical Engineering
kamil@alum.mit.edu

Kamil Khan
Argonne National Laboratory
Mathematics and Computer Science Division
kamil@alum.mit.edu

Stefan Wild
Argonne National Laboratory
wild@mcs.anl.gov

MS110

A Derivative-Free Vu-Algorithm for Convex Finite-

Max Functions

The VU-algorithm is a superlinearly convergent method for minimizing nonsmooth, convex functions. At each iteration, the algorithm separates the space into two orthogonal subspaces, called the V-space and the U-space, such that the nonsmoothness of the objective function is due to its projection onto the V-space only, and on the U-space the projection is smooth. This structure allows for a quasi-Newton step parallel to the U-space, then parallel to the V-space a proximal-point step is taken. The point of this work is to establish a derivative-free variant of the VU-algorithm for convex, finite-max objective functions. It is shown that approximate subgradients are sufficient to get convergence to within acceptable tolerance in this setting.

Chayne Planiden
University of British Columbia Okanagan
chayne.planiden@ubc.ca

MS111

Control and Estimation of Traffic Flow on Networks

This talk presents a new mixed integer programming formulation of traffic density estimation and control problems in highway networks modeled by the Lighthill Whitham Richards Partial Differential Equation. We first present an equivalent formulation of the problem using an Hamilton-Jacobi equation. Then, using a semi-analytic formula, we show that the model constraints resulting from the Hamilton-Jacobi equation result in linear constraints, albeit with unknown integers. We then pose the problem of estimating the density of traffic, given incomplete and inaccurate traffic data, as a Mixed Integer Program. We also show that the same framework can be used for boundary flow control on single highway links and networks. We then present a numerical implementation of both estimation and control methods on realistic examples involving experimental measurement data.

Simoni Michele
UT Austin
m.simoni@utexas.edu

Edward Canepa
King Abdullah University of Science and Technology
Department of Electrical Engineering
edward.canepa@kaust.edu.sa

Christian Claudel
University of Texas at Austin
christian.claudel@utexas.edu

MS111

Total Variation Diminishing Runge-Kutta Methods for the Optimal Control of Conservation Laws: Stability and Order-Conditions

Optimal control problems subject to conservation laws are among challenging optimization problems due to difficulties that arise when shocks are present in the solution. Such optimization problems have application, for instance, in gas networks. Beside theoretical difficulties at the continuous level, discretization of such optimal control problems should be done with care in order to guarantee convergence. In this talk, we present stability results of the total variation diminishing (TVD) Runge-Kutta (RK) methods for such optimal control problems. In particular we show that enforcing strong stability preserving (SSP) for both forward and adjoint problem results to a first order time-discretization. However, requiring SSP only for forward problem is sufficient to obtain a stable discrete adjoint. We also present order-conditions for the TVD-RK methods in the optimal control context.

Soheil Hajian
Humboldt University of Berlin
soheil.hajian@hu-berlin.de

Michael Hintermueller
Weierstrass Institute for Applied Analysis and Stochastics
hintermueller@wias-berlin.de

MS111

Optimal Control of Hyperbolic Balance Laws with State Constraints

This talk deals with the treatment of pointwise state constraints in the context of optimal control of hyperbolic nonlinear scalar balance laws. We study an optimal control problem governed by balance laws with an off/on switching device that allows to interrupt the flux, where the switching times between on- and off-modes are the control variables. Such a problem arises, for example, when considering the optimal control of the traffic density on a one-way street with a traffic light in the middle. The appearance of state constraints presents a special challenge, since solutions of nonlinear hyperbolic balance laws develop discontinuities after finite time, which prohibits the use of standard methods. In this talk, we will build upon the recently developed sensitivity- and adjoint calculus by Pfaff and Ulbrich to derive necessary optimality conditions. In addition, we will use Moreau-Yosida regularization for the algorithmic treatment of the pointwise state constraints. Hereby, we will prove convergence of the optimal controls and weak convergence of the corresponding Lagrange multiplier estimates of the regularized problems.

Johann Schmitt
Technische Universitaet Darmstadt
jschmitt@mathematik.tu-darmstadt.de

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

MS111

Chance Constrained Optimization on Gas Networks Governed by the Isothermal Euler Equations

We study optimization problems on gas networks governed by the isothermal Euler equations with nonconstant compressibility factor under uncertainty. The deterministic problem consists of choosing a pressure control such that prescribed pressure bounds are fulfilled on the whole network. However, the demands of customers, i.e., the boundary mass flows, are typically unknown. Statistical information for the demands is available and allows the construction of a distribution function (e.g. normal) around a mainly temperature dependent mean value. We consider a chance constrained (or probabilistic constrained) optimization problem, i.e., we replace our original constraints by one constraint stating that the probability to be feasible should exceed a prescribed threshold.

David Wintergerst

Friedrich-Alexander-Universität Erlangen-Nürnberg
wintergerst@math.fau.de

MS112

Rapid, Robust, and Reliable Blind Deconvolution via Nonconvex Optimization

We study the question of reconstructing two signals f and g from their convolution $y = f * g$. This problem, known as *blind deconvolution*, pervades many areas of science and technology, including astronomy, medical imaging, optics, and wireless communications. A key challenge of this intricate non-convex optimization problem is that it might exhibit many local minima. We present an efficient numerical algorithm that is guaranteed to recover the exact solution, when the number of measurements is (up to log-factors) slightly larger than the information-theoretical minimum, and under reasonable conditions on f and g. The proposed regularized gradient descent algorithm converges at a geometric rate and is provably robust in the presence of noise. To the best of our knowledge, our algorithm is the first blind deconvolution algorithm that is numerically efficient, robust against noise, and comes with rigorous recovery guarantees under certain subspace conditions. Moreover, numerical experiments do not only provide empirical evidence of our theory, but they also demonstrate that our method yields excellent performance even in situations beyond our theoretical framework.

Xiaodong Li
Department of Statistics
University of California Davis
xdgli@ucdavis.edu

Shuyang Ling
Department of Mathematics
University of California at Davis
syling@math.ucdavis.edu

Thomas Strohmer
University of California,Davis
Department of Mathematics
strohmer@math.ucdavis.edu

Ke Wei
Department of Mathematics
University of California at Davis
kewei@math.ucdavis.edu

MS112

Breaking Sample Complexity Barriers via Nonconvex Optimization?

In the past decade there has been significant progress in understanding when convex relaxations are effective for finding low complexity models from a near minimal number of data samples (e.g. sparse/low rank recovery from a few linear measurements). Despite such advances convex optimization techniques are often prohibitive in practice due to computational/memory constraints. Furthermore, in some cases convex programs are suboptimal in terms of sample complexity and provably require significantly more data samples than what is required to uniquely specify the low complexity model of interest. In fact for many such problems certain sample complexity barriers have emerged so that there are no known computationally tractable algorithm that can beat the sample complexity achieved by such convex relaxations. Motivated by a problem in imaging, In this talk I will discuss my recent results towards breaking such barriers via natural nonconvex optimization techniques.

Mahdi Soltanolkotabi
University of Southern California
soltanol@usc.edu

MS112

Guarantees of Riemannian Optimization for Low Rank Matrix Reconstruction

We establish theoretical recovery guarantees of a family of Riemannian optimization algorithms for low rank matrix reconstruction, which is about recovering an $m \times n$ rank r matrix from $p < mn$ number of linear measurements. The algorithms are first interpreted as the iterative hard thresholding algorithms with subspace projections. Then, based on this connection, we prove that with a proper initial guess the Riemannian gradient descent method and a restarted variant of the Riemannian conjugate gradient method are guaranteed to converge to the measured rank r matrix provided the number of measurements is proportional to nr^2 up to a log factor. Empirical evaluation shows that the algorithms are able to recover a low rank matrix from nearly the minimum number of measurements necessary.

Ke Wei
Department of Mathematics
University of California at Davis
kewei@math.ucdavis.edu

Jian-Feng Cai
Department of Mathematics
Hong Kong University of Science and Technology
jfcai@ust.hk

Tony Chan
HKUST
tonyfchan@ust.hk

Shingyu Leung
Hong Kong University of Science and Technology
masyleung@ust.hk

MS112

Nonconvex Recovery of Low-Complexity Models

General nonconvex optimization problems are NP-hard. In applied disciplines, however, nonconvex problems abound, and heuristic algorithms are often surprisingly effective. The ability of nonconvex heuristics to find high-quality solutions for practical problems remains largely mysterious. In this talk, I will describe a family of nonconvex problems which can be solved efficiently. This family has the characteristic structure that (1) every local minimizer is also global, and (2) the objective function has a negative directional curvature around all saddle points (ridable saddle). Natural (nonconvex) formulations for a number of important problems in signal processing and machine learning lie in this family, including the eigenvector problem, complete dictionary learning (CDL), generalized phase retrieval (GPR), orthogonal tensor decomposition, and various synchronization and deconvolution problems. This benign geometric structure allows a number of optimization methods to efficiently find a global minimizer, without special initializations. This geometric approach to solving nonconvex problems has led to new types of computational guarantees for a number of practical problems, including dictionary learning, generalized phase retrieval, and sparse blind deconvolution. I will describe several open challenges in terms of both theory and algorithms, and give applications of these ideas to computer vision and the analysis of microscopy data.

John Wright
Department of Electrical Engineering
Columbia University
johnwright@ee.columbia.edu

Ju Sun
Stanford University
sunjunus@gmail.com

Qing Qu, Yuqian Zhang
Columbia University
Electrical Engineering
qq2105@columbia.edu, yz2409@columbia.edu

MS113

Two-Stage Linear Decision Rules for Multi-Stage Stochastic Programming

Multi-stage stochastic linear programs (MSLPs) are notoriously hard to solve in general. An approximation of MSLP can be obtained by applying linear decision rules (LDRs) on the recourse decisions, i.e., by restricting the set of feasible policies into affine functions of uncertain parameters. This reduces MSLP into a static problem which provides an upper bound on the optimal objective value. Similarly, it is possible to obtain a lower bound by applying LDRs on the dual decisions. Under certain (somehow restrictive) assumptions, the restricted primal and dual problems are both tractable linear programs. In this work, we introduce two-stage LDR whose application reduces MSLP (or its dual) into a two-stage stochastic linear program which can be solved efficiently by means of decomposition. In addition to potentially yielding better policies and bounds, this approach requires many fewer assumptions than are required to obtain an explicit reformulation when using the standard static LDR approach. As an illustrative example we apply our approach to a capacity expansion model, and find that the two-stage LDR policy has expected cost between 20% and 34% lower than the static LDR policy, and in the dual yields lower bounds that are between 0.1% and 3.3% better.

Merve Bodur
University of Toronto
Department of Mechanical and Industrial Engineering
bodur@mie.utoronto.ca

James Luedtke
University of Wisconsin-Madison
Department of Industrial and Systems Engineering
jrluedt1@wisc.edu

MS113

Progressive Hedging Like Methods for Computing Lagrangian Dual Bounds in Stochastic Mixed-Integer Programming

A new primal-dual decomposition method is presented, based on an integration of the Progressive Hedging (PH) and the Frank-Wolfe (FW) methods, referred to as FW-PH, for computing high-quality Lagrangian bounds of the stochastic mixed-integer programming problem (SMIP). The deterministic equivalent (DE) form of the SMIP typically lacks convexity due to the assumed integrality restrictions, and since PH is a specialization of the alternating direction method of multipliers (ADMM), the application of PH to the SMIP is not theoretically supported due to this lack of convexity. Thus, PH applied to the SMIP is understood as a heuristic approach without convergence guarantees, where either cycling or suboptimal convergence is possible. Although Lagrangian bounds may be computed after each PH iteration, these bounds often show limited improvement with increasing number of iterations, and the amount of improvement is highly sensitive to the value of the PH penalty parameter. Motivated by these observations, we modify the PH method so that the generated sequence of Lagrangian bounds is guaranteed to converge to the optimal Lagrangian bound for any positive-valued PH penalty parameter. The new integrated method is shown to be both theoretically supported due to an integration of established theory for ADMM and the FW method, and practically implementable. Numerical experiments demonstrate the improvement of FW-PH over the PH method for computing high-quality Lagrangian bounds.

Andrew C. Eberhard
RMIT University, Discipline of Mathematical Sciences
School
of Science
andy.eb@rmit.edu.au

Natashia Boland
University of Newcastle
natashia.boland@newcastle.edu.au

Brian C. Dandurand
Argonne National Laboratory
brian.dandurand@rmit.edu.au

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

Fabricio Oliveira, Jeffrey Christiansen
Mathematical Sciences,
School of Science, RMIT University
fabricio.oliveira@rmit.edu.au,
s3507717@student.rmit.edu.au

MS113

A Sequential Sampling Algorithm for Stochastic Multistage Programs

Stochastic Dual Dynamic Programming (SDDP) has become a prominent approach to tackle multistage stochastic programs. The convergence argument for this algorithm rely on existence of finitely many Benders cuts that can be generated. This is possible only when the probability distributions associated with underlying stochastic processes defining the optimization problem are known, and often represented as a scenario tree. However, in many applications one may not have a priori description of scenarios, and their probabilities. For such cases, the traditional Benders cuts in SDDP are no longer available, and one has to resort to sequential sampling approach where empirical estimates of the minorants can be calculated. We will discuss convergence of one such sequential sampling algorithm which we refer to as the Stochastic Dynamic Linear Programming. This algorithm is a dynamic extension of the regularized two-stage stochastic decomposition for stage-wise independent multistage stochastic linear programs. It turns out that the use of regularization becomes the key to convergence of such algorithms. We will also present results from our computational experiments conducted on a short-term distributed storage control problem. These results show that our distribution-free approach provides prescriptive solutions and values which are statistically indistinguishable from those obtained from SDDP, while improving computational times significantly.

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

MS113

Nested Decomposition of Multistage Stochastic Integer Programs with Binary State Variables

We propose a valid nested decomposition algorithm for multistage stochastic integer programming problems when the state variables are binary. We prove finite convergence of the algorithm as long as the cuts satisfy some sufficient conditions. We discuss the use of well known Benders and integer optimality cuts within this algorithm, and introduce new cuts derived from a reformulation of the problem where local copies of state variables are introduced. We propose a stochastic variant of this algorithm and prove its finite convergence with probability one. Numerical experiment on a large-scale generation expansion planning problem will be presented.

Shabbir Ahmed
Georgia Institute of Technology, Industrial and Systems Engineering
shabbir.ahmed@isye.gatech.edu

Andy Sun
Georgia Tech
College of Engineering
andy.sun@isye.gatech.edu

MS114

Mixed-Integer Linear Programming for a PDE-Constrained Dynamic Network Flow

To compute a numerical solution of PDEs it is common practice to find a suitable finite dimensional approximation of the differential operators and solve the obtained finite dimensional system, which is linear if the PDE is linear. In case the PDE is under an outer influence (control), a natural task is to find an optimal control with respect to a certain objective.
In this talk general methods and concepts are given for linear PDEs with a finite set of control variables, which can be either binary or continuous. The presence of the binary variables restricts the feasible controls to a set of hyperplanes of the control space and makes conventional methods inapplicable. We show that under the assumption that all additional constraints of the system are linear, these problems can be modeled as a mixed-integer linear program (MILP), and numerically solved by a linear programming (LP) based branch-and-cut method. If the discretized PDE is directly used as constraints in the MILP, the problem size scales with the resolution of the discretization in both space and time. A reformulation of the model allows to solve the PDE in a preprocessing step, which not only helps to significantly reduce the number of constraints and computation time, but also makes it possible to use adaptive finite element methods. It is further shown that the scaling of state-constraints with finer discretization can be reduced significantly by enforcing them in a lazy-constraint-callback.

Fabian Gnegel
Helmut-Schmidt Universitaet
Hamburg, Germany
gnegelf@hsu-hh.de

Armin Fügenschuh
Helmut Schmidt University Hamburg
fuegenschuh@hsu-hh.de

Marcus Stiemer
Helmut-Schmidt-University of
Federal Armed Forces Hamburg
m.stiemer@hsu-hh.de

Michael Dudzinski
Helmut Schmidt University of the Federal Armed Forces Hamburg
dudzinski@hsu-hh.de

MS114

Mixed-Integer PDE-Constrained Optimization for Gas Networks

We discuss a PDE-constrained optimal control problem that arises in the optimal control of natural gas transport networks subject to nonlinear friction. Following a traditional first-discretize-then-optimize approach, we start from a MINLP formulation of the problem and focus on issues that arise when applying decoupling and rounding methods to it. While rounding methods such as Sum-Up-Rounding, which solve the MINLP approximately by solving an NLP and an MILP, are a powerful tool that is used to great effect in the field of optimal control, they generally require solutions to the relaxed NLP to resemble solutions to the original MINLP. We discuss several aspects of our problem formulation that render it inaccessible to classical decoupling methods and present reformulations, approximations and solution techniques that work around these issues and open up alternate sources of information that may be exploited by approximate MINLP solvers.

Mirko Hahn
Argonne National Laboratory
hahnm@anl.gov

MS114

Gradient Descent Methods for Mixed-Integer PDE-Constrained Optimization

Motivated by challenging tasks such as minimal cost operation of gas, water or traffic networks we consider mixed-integer optimization problems subject to evolution type PDE constraints. The integer restrictions model discrete control components, for example valves, gates or signal signs in the mentioned applications. We consider relaxation and re-parameterization techniques in order to solve such problems in an appropriate sense locally using gradient descent methods. The descent strategies are chosen such that gradients of certain sub-problems can be obtained numerically very efficiently by solutions of suitable adjoint PDEs. The performance of these methods will be demonstrated on numerical examples.

Falk M. Hante
Friedrich-Alexander-Universität Erlangen-Nürnberg
falk.hante@fau.de

MS114

Introduction to Mixed-Integer PDE Constrained Optimization

We introduce mixed-integer PDE-constrained optimization problems, discuss some theoretical and computational challenges of this new class of problems, and propose a classification of this problem class. We present a test-problem library for mixed-integer PDE-constrained optimization written in AMPL, and present preliminary numerical results using existing mixed-integer solvers. We emphasize problem formulations that ensure convex relaxations, and tight relaxations that result in smaller search trees. We also discuss transformations that simplify the problem formulations. Our goal is to highlight the impact of modeling choices at the interface between integer programming and PDE constrained optimization.

Sven Leyffer
Argonne National Laboratory
leyffer@mcs.anl.gov

MS115

Compositions of Convex Functions and Fully Linear Models

Model-based optimization methods use an approximation of the objective function is to guide the optimization algorithm. Proving convergence of such methods often applies an assumption that the approximations form fully linear models – an assumption that requires the true objective function to be smooth. However, some recent methods have loosened this assumption and instead worked with functions that are compositions of smooth functions with simple convex functions (the max-function or the ℓ-1 norm). In this talk, we examine the error bounds resulting from the composition of a convex lower semi-continuous function with a smooth vector-valued function when it is possible to provide fully linear models for each component of the vector-valued function.

Warren Hare
University of British Columbia, Okanagan Campus
warren.hare@ubc.ca

MS115

From Least Squares Solutions of Linear Inequality Systems to Convex Least Squares Problems

In our contributions, we consider least squares solutions (if any) in problems of three different types: for a finite number of linear inequality systems, for infinitely many linear systems, and for general convex systems. Our objective focuses on existence results (for properly defined least squares solutions), characterizations of such solutions, and in some cases algorithms for computing them.

Jean-Baptiste Hiriart-Urruty
University of Toulouse
France
jbhu@math.univ-toulouse.fr

MS115

VU Decomposition and Partial Smoothness for Sublinear Functions

VU decomposition and partial smoothness are related notions for generic nonsmooth functions. Sublinear functions are widely used in optimization algorithms as they themselves contain important structures. We discuss how to characterize several concepts related to partial smoothness for this type of functions. In particular, for a closed sublinear function h we introduce a relaxed decomposition depending on certain V_ε and U_ε subspaces that converge continuously to the V- and U-counterparts as ε tends to zero. To the new $V_\varepsilon U_\varepsilon$ decomposition corresponds a relaxed smooth manifold that contains the manifold where h is partly smooth.

Shuai Liu
IMPA, Brazil
shuailiu@impa.br

Claudia A. Sagastizabal
IMPA
Visiting Researcher
sagastiz@impa.br

Mikhail V. Solodov
Inst de Math Pura e Applicada
solodov@impa.br

MS115

A Unified Approach to Error Bounds for Structured Convex Optimization Problems

Error bounds, which refer to inequalities that bound the distance of vectors in a test set to a given set by a residual function, have proven to be extremely useful in analyzing the convergence rates of a host of iterative methods for solving optimization problems. In this paper, we present a new framework for establishing error bounds for a class of structured convex optimization problems, in which the objective function is the sum of a smooth convex function and a general closed proper convex function. Such a class encapsulates not only fairly general constrained minimization problems but also various regularized loss minimization formulations in machine learning, signal processing, and statistics. Using our framework, we show that a number of existing error bound results can be recovered in a unified and transparent manner. To further demonstrate the power of our framework, we apply it to a class of nuclear-norm regularized loss minimization problems and establish a new error bound for this class under a strict complementarity-type regularity condition. We then complement this result by constructing an example to show that the said error bound could fail to hold without the regularity condition. We believe that our approach will find further applications in the study of error bounds for structured convex optimization problems.

Zirui Zhou
Simon Fraser University
ziruiz@sfu.ca

MS116

Relaxations of the Complementarity-Based Formu-

lations of the L0 Problem

In many high-dimensional statistical learning problems there is a reason to believe that the solution of interest is sparse. To induce sparsity, the objective function of such learning problems includes ℓ_0-norm of the decision variable that penalizes based on the number of nonzero elements. For computational tractability, the nonconvex ℓ_0-norm is usually replaced with its convex ℓ_1-norm surrogate. From the learning perspective, however, recent studies have shown that the local solutions to nonconvex problems could be superior to the global solutions of convex problems. One way to model the ℓ_0-norm is by introducing nonconvex complementarity constraints. In this talk, we present our findings on convex relaxations of such nonconvex problems.

Sam Davanloo, Ben Chaiken
Ohio State University
davanloo-tajbakhsh.1@osu.edu, chaiken.2@osu.edu

MS116

Rank Minimization using a Complementarity Approach

The problem of minimizing the rank of a matrix subject to constraints can be cast equivalently as a semidefinite program with complementarity constraints (SDCMPCC). The formulation requires two positive semidefinite matrices to be complementary. We investigate calmness of locally optimal solutions to the SDCMPCC formulation and hence show that any locally optimal solution is a KKT point, under an assumption on the original set of constraints. We develop a penalty formulation of the problem. We present calmness results for locally optimal solutions to the penalty formulation. We also develop a proximal alternating linearized minimization (PALM) scheme for the penalty formulation, and investigate the incorporation of a momentum term into the algorithm. Computational results are presented.

John E. Mitchell
Rensselaer Polytechnic Institute
110, 8th Street, Troy, NY, 12180
mitchj@rpi.edu

Xin Shen
Monsanto
St Louis MO
xin.shen@monsanto.com

MS116

Algorithms for Dictionary Learning and Sparse Coding Problems

This talk discusses numerical schemes for sparse coding and its extension, convolutionary sparse coding, arising from machine learning, high dimensional statistics, and big data analysis. The sparse coding problems give rise to a nonconvex, nonsmooth, and large-scale optimization problem. By using the recent techniques for sparse optimization, we develop first order necessary and second order sufficient optimality conditions for a local minimizer. Based on these conditions, a Newton scheme is proposed to achieve the local 2nd order convergence, where convex program is solved at each step. Aubin's property is established and used to prove the 2nd order convergence. Finally, statistical properties of the proposed algorithms will be discussed.

Jinglai Shen
University of Maryland Baltimore County
shenj@umbc.edu

Xiao Wang
Dept. of Statistics, Purdue University
wangxiao@purdue.edu

MS116

On the Global Resolution ℓ_0-Norm Minimization Problems via ADMM Schemes

We consider the minimization of an objective function containing an ℓ_0-norm term which implicitly emphasizes the need for a sparse minimizer. Such a problem is widely considered in image processing and statistical learning. However, because of the nonconvex and discontinuous nature of this norm, many use the ℓ_1-norm or other variants as substitutes, and far less is known about directly solving the ℓ_0 problem. In this work, we consider an approach for the direct resolution of this problem through a reformulation as an equivalent continuous mathematical program wtih complementarity constraints (MPCC). We propose the solution of this problem by considering an ADMM scheme which requires the solution of two subproblems in each iteration. Of these, the first is convex while the second is nonconvex. However, we observe that this nonconvex subproblem possesses a hidden convexity property and its solution can be provided in closed form. Based on this observation, both subproblems may be solved efficiently. Numerical experiments indicate that the scheme produces near global optimal solutions and the effort scales modestly with problem size.

Yue Xie
Pennsylvania State University
yux111@psu.edu

Uday Shanbhag
Department of Industrial and Manufacturing Engineering
Pennsylvania State University
udaybag@psu.edu

MS117

Solving a Quadratic Problem on a Bounded Integer Domain Using a Quantum Annealer

Quantum annealers solve Quadratic Unconstrained Binary Optimization (QUBO) problems via transformation to Ising models. Current implementations of quantum devices, however, have limited numbers of qubits and are furthermore prone to various sources of noise. In practice, this restricts the usage of the quantum devices to a limited number of qubits and a limited range of applicable ferromagnetic biases and couplings. One way to solve unconstrained quadratic integer problems using quantum annealers is encoding integer variables as binary ones. In this talk, we discuss several integer encodings and propose one that is more robust to the limitations of the currently available quantum devices.

Sahar Karimi
University of Waterloo

sahar.karimi@gmail.com

MS117

Optimization: Quantum vs Classical

Can quantum computers meet the tantalizing promise of solving complex calculations – such as optimization problems or database queries – faster than classical computers based on transistor technologies? Although IBM recently opened up their five-qubit programmable quantum computer to the public to tinker with, the holy grail of a useful large-scale programmable universal quantum computer is decades away. While working mid-scale programmable special-purpose quantum optimization machines exist, a conclusive detection of quantum speedup remains controversial despite promising results by Google Inc. In this talk I outline how we can predict the typical difficulty of optimization problems without solving them with the goal of "tickling" any quantumness out of these machines while, at the same time, aiding in the search for the "killer" application domain where quantum optimization might excel. Finally, an overview of different sequential, non-tailored, as well as specialized tailored classical state-of-the-art algorithms is given. Current quantum annealing technologies must outperform these to claim the crown in the race for quantum speedup.

Helmut G. Katzgraber
Texas A&M University
helmut@katzgraber.org

MS117

Building a Performance Model for Quantum Annealing

A primary objective in optimization research is to build performance models for heuristic algorithms, which describe the relationships between input properties, algorithm parameters, and performance. D-Wave's annealing based quantum computers fall within the adiabatic quantum model of computation. Since theoretical results in this model are sparse, we use empirical methods to understand and characterize performance. I will describe some challenges that arise when developing performance models for these novel computing systems, and survey what is known so far.

Catherine McGeoch
D-Wave Systems Inc.
cmcgeoch@dwavesys.com

MS117

Leveraging Quantum Computing for Optimizing Data Analysis

Abstract not available at time of publication.

Immanuel Trummer
Cornell University
itrummer@cornell.edu

MS118

Estimating Large Covariance Matrices Using Proximity Operators

The covariance matrix is one of the most fundamental objects in statistics, and yet its reliable estimation is notoriously difficult in high dimensional settings. The sample covariance matrix is known to perform poorly as an estimator when the number of variables is large relative to the sample size. Indeed, the key to tractable estimation of a high-dimensional covariance matrix lies in exploiting any knowledge of the covariance structure. Convex regularization penalties are a natural way to incorporate such problem-specific structural knowledge. In this talk, we discuss several recently introduced estimators of the covariance matrix which are formulated as proximity operators of certain convex penalties.

Jacob Bien
Department of Biological Statistics
Cornell University
jbien@cornell.edu

MS118

Proximal Methods for Penalized Concomitant M-Estimators

In statistics, the task of simultaneously estimating a regression vector and an additional model parameter is often referred to as concomitant estimation. Huber introduced a generic method for formulating "maximum likelihood-type" estimators (or M-estimators) with a concomitant parameter from a convex criterion. Recent work in penalized regression has extended this framework to the high-dimensional setting. In this contribution, we provide a generic proximal algorithmic approach to solving convex optimization problems associated with this class of penalized concomitant M-estimators. The class of estimators includes the scaled Lasso and the TREX as special cases. We present novel proximity operators arising from different concomitant estimators and show their applicability in standard proximal algorithm schemes. We illustrate both the algorithmic performance of the optimization routines and the statistical performance of the different estimators on selected synthetic and real-world examples.

Christian L. Mueller
Simons Center for Data Analysis
Simons Foundation
cmueller@simonsfoundation.org

Patrick L. Combettes
Department of Mathematics
North Carolina State University
plc@math.ncsu.edu

MS118

A New Perspective on Stochastic 0th and 1st Order Proximal Methods

Abstract not available at time of publication.

Vianney Perchet
CMLA Research Center for Applied Maths
Ecole normale supérieure de Cachan
vianney.perchet@gmail.com

MS118

Hierarchical Convex Optimization and Proximal Splitting for High-Dimensional Statistics

The proximal splitting algorithms are powerful convex optimization tools to exploit multiple prior knowledge for improvements of variety of high dimensional estimation tasks in data sciences and engineerings. These algorithms can iteratively approximate an unspecial vector among possibly infinitely many minimizers of a superposition of multiple nonsmooth convex functions. With elegant translations of the solution set, i.e., the set of all minimizers, into the fixed point sets of nonexpansive mappings, the hybrid steepest descent method allows a hierarchical optimization, i.e., further strategic selection of a most desirable vector among the solution set, by minimizing an additional convex function over the solution set. In this talk, we introduce fixed point theoretic interpretations of the proximal splitting algorithms and their enhancements by the hybrid steepest descent method with applications to recent advanced statistical estimation problems. The applications include certain enhancements of the state-of-the-art concomitant estimators, e.g., the Sqrt-Lasso estimator, the TREX estimator, and the generalized TREX estimator. This talk is based on a joint work with Dr.Yamagishi of the Tokyo Institute of Technology.

Isao Yamada
Dept. Communications and Computer Engineering
Tokyo Institute of Technology
isao@ict.e.titech.ac.jp

MS119

Sketchy Decisions: Convex Optimization with Optimal Storage

This work proposes a novel paradigm for designing optimization algorithms for large-scale convex problems with structured solutions. The two key ideas are (i) to maintain only a sketch of the decision variable and (ii) to use algorithmic templates that are compatible with these sketches. A natural target for this approach is the class of convex matrix optimization problems with low-rank solutions. This talk develops and analyzes a new sketch-based algorithm for a model problem from this class. Numerical evidence establishes that the algorithm scales to huge instances while retaining the advantages of convex optimization. joint work with Alp Yurtsever, Quoc Tran Dinh, Madeliene Udell, and Joel Tropp

Volkan Cevher, Alp Yurtsever, Quoc Tran Dinh
EPFL
volkan.cevher@epfl.ch, alp.yurtsever@epfl.ch,
quoc.trandinh@epfl.ch

Madeleine R. Udell
Stanford University
Stanford University
madeleine.udell@gmail.com

Joel Tropp
Applied and Computational Mathematics
Caltech
jtropp@acm.caltech.edu

MS119

Recent Advances in Frank-Wolfe Optimization

We will review the venerable Frank-Wolfe algorithm and its variants in the context of modern applications in machine learning and signal processing. We will summarize some of its latest algorithmic and theoretical improvements, including a global linear convergence analysis of variants of Frank-Wolfe that seeded further developments presented in this symposium. We will also mention current open questions, including the application of Frank-Wolfe to saddle point problems.

Simon Lacoste-Julien
University of Montreal
slacoste@iro.umontreal.ca

MS119

Decomposition-Invariant Linearly Convergent Conditional Gradient Algorithm for Structured Polytopes

Recently, several works have shown that natural modifications of the classical conditional gradient method (aka Frank-Wolfe algorithm) for constrained convex optimization, provably converge with a linear rate when the feasible set is a polytope, and the objective is smooth and strongly-convex. However, all of these results suffer from significant shortcomings: 1) large memory requirement (and running time) due to storing an explicit convex decomposition of the current iterate, and 2) the worst case convergence rate depends unfavorably on the dimension. In this work we present a new conditional gradient variant and a corresponding analysis that improves on both of the above shortcomings. In particular, both memory and computation overheads are only linear in the dimension, and in addition, in case the optimal solution is sparse, the new convergence rate replaces a factor which is at least linear in the dimension in previous work, with a linear dependence on the number of non-zeros in the optimal solution. At the heart of our method, and corresponding analysis, is a novel way to compute decomposition-invariant away-steps. While our theoretical guarantees do not apply to any polytope, they apply to several important structured polytopes that capture central concepts such as paths in graphs, perfect matchings in bipartite graphs, marginal distributions that arise in structured prediction tasks, and more.

Dan Garber
TTI Chicago
dgarber@ttic.edu

Ofer Meshi
Google
meshi@google.com

MS119

Polytope Conditioning and Linear Convergence of the Frank-Wolfe Algorithm

It is known that the gradient descent algorithm converges linearly when applied to a strongly convex function with Lipschitz gradient. In this case the rate of convergence is determined by condition number of the function. In a similar vein, it has been shown that a variant of the Frank-Wolfe algorithm with away steps converges linearly when applied a strongly convex function over a polytope. In a nice extension of the unconstrained case, the rate of convergence is determined by the product of the condition number of the function and a certain condition number of the polytope. We shed new light into the latter type of polytope conditioning. In particular, we show that previous and seemingly different approaches to define a suitable condition measure for the polytope are essentially equivalent to each other. Perhaps more interesting, they can all be unified via a parameter of the polytope that formalizes a key premise linked to the linear convergence of the algorithm. We also give new insight into the linear convergence property. For a convex quadratic objective, we show that the rate of convergence is determined by the condition number of a suitably scaled polytope.

Javier Pena, Daniel Rodriguez
Carnegie Mellon University
jfp@andrew.cmu.edu, drod@cmu.edu

MS120

Quantum Bilinear Optimization

We study optimization programs given by a bilinear form over non-commutative variables subject to linear inequalities. Problems of this form include the entangled value of two-prover games, entanglement-assisted coding for classical channels and quantum-proof randomness extractors. We introduce an asymptotically converging hierarchy of efficiently computable semidefinite programming (SDP) relaxations for this quantum optimization. This allows us to give upper bounds on the quantum advantage for all of these problems. Compared to previous work of Pironio, Navascues and Acin, our hierarchy has additional constraints. By means of examples, we illustrate the importance of these new constraints both in practice and for analytical properties. Moreover, this allows us to give a hierarchy of SDP outer approximations for the completely positive semidefinite cone introduced by Laurent and Piovesan.

Mario Berta
Institute for Quantum Information and Matter
California Institute of Technology, USA
berta@caltech.edu

MS120

Quantum Channels, Spectrahedra and a Tracial Hahn-Banach Theorem

This talk will discuss matrix convex sets and their tracial analogs which we call contractively tracial convex sets. In both contexts completely positive (cp) maps play a central role: unital cp maps in the case of matrix convex sets and trace preserving cp (CPTP) maps in the case of contractively tracial convex sets. CPTP maps, also known as quantum channels, are fundamental objects in quantum information theory. Free convexity is intimately connected with Linear Matrix Inequalities (LMIs) $L(x) = A_0 + A_1x_1 + ... + A_gx_g \succeq 0$ and their matrix convex solution sets $\{X : L(X)$ is positive semidefinite$\}$, called free spectrahedra. The Effros-Winkler Hahn-Banach Separation Theorem for matrix convex sets states that matrix convex sets are solution sets of LMIs with operator coefficients. Motivated in part by cp interpolation problems, we will develop the foundations of convex analysis and duality in the tracial setting, including tracial analogs of the Effros-Winkler Theorem. This is based on joint work with J. William Helton and Scott McCullough.

Igor Klep
Department of Mathematics
The University of Auckland
igor.klep@auckland.ac.nz

MS120

Convex Separation from Convex Optimization for Large-Scale Problems

We present a scheme, based on Gilbert's algorithm for quadratic minimization [SIAM J. Contrl., vol. 4, pp. 61-80, 1966], to prove separation between a point and an arbitrary convex set S?Rn via calls to an oracle able to perform linear optimizations over S. Compared to other methods, our scheme has almost negligible memory requirements and the number of calls to the optimization oracle does not depend on the dimensionality n of the underlying space. We study the speed of convergence of the scheme under different promises on the shape of the set S and/or the location of the point, validating the accuracy of our theoretical bounds with numerical examples. Finally, we present some applications of the scheme in quantum information theory. There we find that our algorithm out-performs existing linear programming methods for certain large scale problems, allowing us to certify nonlocality in bipartite scenarios with upto 42 measurement settings. We apply the algorithm to upper bound the visibility of two-qubit Werner states, hence improving known lower bounds on Grothendieck's constant KG(3). Similarly, we compute new upper bounds on the visibility of GHZ states and on the steerability limit of Werner states for a fixed number of measurement settings.

Miguel Navascues
Institute of Quantum Optics and Quantum Information
Austian Academy of Sciences
miguel.navascues@oeaw.ac.at

MS120

Using Noncommutative Polynomial Optimization for Matrix Factorization Ranks

In this talk I will explain how tracial noncommutative polynomial optimization can be used to compute lower bounds on the cpsd-rank (completely positive semidefinite rank) of a matrix. Here the cpsd-rank is a quantum inspired generalization of the cp-rank, where we consider Gram factorizations by positive semidefinite matrices instead of nonnegative vectors. Inspired by this approach we have also derived new techniques to lower bound the cp-rank, and I will compare this to recent results by Fawzi and Parrilo. Finally, I will discuss how this relates to the dimension required to realize quantum correlations.

Sander Gribling, David de Laat
Centrum Wiskund & Informatica, The Netherlands
s.j.gribling@cwi.nl, mail@daviddelaat.nl

Monique Laurent
CWI, Amsterdam
and Tilburg University
M.Laurent@cwi.nl

MS121

A Dual Active-Set Algorithm for Regularized Monotonic Regression

Monotonic (isotonic) Regression (MR) is a powerful tool used for solving a wide range of important applied problems. One of its features, which poses a limitation on its use in some areas, is that it produces a piecewise constant fitted response. For smoothing the fitted response, we introduce a regularization term in the MR formulated as a least distance problem with monotonicity constraints. The resulting Smoothed Monotonic Regression (SMR) is a convex quadratic optimization problem. We focus on the SMR, where the set of observations is completely (linearly) ordered. Our Smoothed Pool-Adjacent-Violators (SPAV) algorithm is designed for solving the SMR. It belongs to the class of dual active-set algorithms. We proved its finite convergence to the optimal solution in, at most, n iterations, where n is the problem size. One of its advantages is that the active set is progressively enlarging by including one or, typically, more constraints per iteration. This resulted in solving large-scale SMR test problems in a few iterations, whereas the size of that problems was prohibitively too large for the conventional quadratic optimization solvers. Although the complexity of the SPAV algorithm is $O(n^2)$, its running time was growing in our computational experiments almost linearly with n.

Oleg Burdakov, Oleg Sysoev
Linkoping University
oleg.burdakov@liu.se, oleg.sysoev@liu.se

MS121

Parameter Estimation and Variable Selection for Big Systems of High-Dimensional Linear Ordinary Differential Equations

In this talk, we will introduce several new methods, which are based on a new matrix framework, to solve an inverse problem in a big system of high-dimensional linear ordinary differential equations. In the procedure to reconstruct the system, both parameter estimation and variable selection problems are discuessed. Finally, several applications are introduced to show the efficiency of our algorithms for real-world challenges.

Leqin WU
Jinan University
twuleqin@jnu.edu.cn

Ya-Xiang Yuan
Chinese Academy of Sciences
Inst of Comp Math & Sci/Eng, AMSS
yyx@lsec.cc.ac.cn

Xing Qiu
University of Rochester
xing_qiu@urmc.rochester.edu

Hulin Wu
Department of Biostatistics
University of Texas Health Science Center at Houston
hulin.wu@uth.tmc.edu

MS121

A TV-SCAD Approach for Image Deblurring with Impulsive Noise

We consider image deblurring problem in the presence of impulsive noise. It is well known that total variation (TV) regularization with L1-norm data fitting (TVL1) works well only when the level of impulsive noise corruption is low. For high level impulsive noise, the TVL1 model works poorly mainly because the corrupted data is over fitted. In this paper, we propose to use TV regularization with nonconvex SCAD data fitting for image deblurring with high level impulsive noise. By decomposing the SCAD function as the difference of two convex functions, the TV-SCAD model is reformulated as a DC programming. We then adopt a classical DC algorithm to solve the TV-SCAD model via solving a sequence of TVL1-equivalent problems. Theoretically, it is guaranteed that any limit point of the sequence of points generated by the proposed approach is a critical point of the nonconvex optimization problem. Numerically, we present extensive numerical results to demonstrate the superiority of the proposed approach.

Junfeng Yang
Nanjing University, China
jfyang@nju.edu.cn

MS121

Completely Positive Binary Tensors

A symmetric tensor is completely positive (CP) if it is a sum of tensor powers of nonnegative vectors. This paper characterizes completely positive binary tensors. We show that a binary tensor is completely positive if and only if it satisfies two linear matrix inequalities. This result can be used to determine whether a binary tensor is completely positive or not. When it is, we give an algorithm for computing its cp-rank and the decomposition. When the order is odd, we show that the cp-rank decomposition is unique. When the order is even, we completely characterize when the cp-rank decomposition is unique. We also discuss how to compute the nearest cp-approximation when a binary tensor is not completely positive.

Jinyan Fan
Department of Mathematics, Shanghai Jiao Tong University
Shanghai 200240, P.R. China
jyfan@sjtu.edu.cn

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

Anwa Zhou
Shanghai University, China
zhouanwa@shu.edu.cn

MS122

Adaptive Finite Element Solvers for MPECs in Function Space

Abstract not available at time of publication.

Michael Hintermueller
Humboldt-University of Berlin
hint@math.hu-berlin.de

MS122

PDE-Constrained Optimization Using Epi-Regularized Risk Measures

Many science and engineering applications necessitate the optimization of PDEs with uncertain inputs such as coefficients, boundary conditions and initial conditions. In this talk, I formulate these problems as risk-averse optimization problems in Banach space. Such problems are often nonsmooth and require an enormous number of samples to accurately quantify the uncertainty in the governing PDE. To circumvent these issues, I present a general smoothing technique for risk measures based on epigraphical calculus. I show that the resulting smoothed risk measures are differentiable and epi-converge to the original nonsmooth risk measure. Moreover, I prove consistency of this approximation for both minimizers and stationary points. I conclude with numerical examples demonstrating these results.

Drew P. Kouri
Optimization and Uncertainty Quantification
Sandia National Laboratories

dpkouri@sandia.gov

MS122

Optimal Control of Quasilinear Parabolic Equations with State Constraints

We present a modern abstract framework for optimal control problems governed by quasilinear parabolic evolution equations in divergence form. The optimal control problem is presumed to also include state constraints. To obtain continuity of the state, we build upon maximal parabolic regularity techniques combined with recent elliptic regularity results for negative Sobolev spaces with integrability order greater than the space dimension. This way, we are able to handle rather involved forcing terms in the quasilinear evolution equation, including such which arise as solution operators to a number of subordinated equations; in this sense, we also treat systems of PDEs at once. For the optimal control problem, the quasilinear structure gives rise to further complications, in particular the lack of suitable a priori bounds for sequences of states. Moreover, we cannot exclude the phenomenon of blow-up of solutions. We thus restrict the optimal control problem to the set of controls admitting global-in-time solutions and show that doing so, under standard assumptions, gives the same optimality conditions as one would have obtained if the problem of blow-up did not arise. The downside of this ansatz is that the proof of existence of globally optimal controls becomes harder, for which we propose to use the objective functional to add favorable properties to an infimal sequence.

Hannes Meinlschmidt
TU Darmstadt
meinlschmidt@mathematik.tu-darmstadt.de

MS122

Risk-Averse PDE-Constrained Optimization and Nash Equilibrium Problems

Uncertainty is ubiquitous in engineering and the natural sciences. Thus, whenever we model real-world problems it is important that this uncertainty is incorporated. For many applications, this leads to partial differential equations (PDEs) with uncertain inputs. Upon transitioning from modeling and simulation to optimization, we typically arrive at a stochastic optimization problem with distributed parameters or decision variables. In order to handle the randomness of the objective functional and control-to-state mapping, we make use of risk measures. This leads to non-smooth, stochastic PDE-constrained optimization problems. After establishing reasonable conditions that allow us prove the existence of a solution and derive first-order optimality conditions, we discuss smoothing techniques (including their asymptotic behavior). In addition, we demonstrate how these ideas can be extended to (multi-objective) PDE-constrained generalized Nash equilibrium problems and so-called MOPECs (multiple optimization problems with equilibrium constraints). We then demonstrate the effects of various risk measures on the overall cost and design of the decisions via numerical examples for each of the problem classes mentioned above.

Thomas M. Surowiec
Department of Mathematics and Computer Science
Philipps University of Marburg
thomas.surowiec@gmail.com

MS123

A Universal Catalyst for First-Order Optimization

Abstract not available at time of publication.

Zaid Harcahoui
University of Washington
zaid.harchaoui@gmail.com

MS123

A Novel, Simple Interpretation of Nesterovs Accelerated Method as a Combination of Gradient and Mirror Descent

Abstract not available at time of publication.

Lorenzo Orecchia
Boston University
orecchia@bu.edu

MS123

A Variational Perspective on Accelerated Methods in Optimization

Abstract not available at time of publication.

Ashia Wilson
Statistics Dept.
U.C. Berkeley
ashia@berkeley.edu

MS123

Regularized Nonlinear Acceleration

We describe a convergence acceleration technique for generic optimization problems. Our scheme computes estimates of the optimum from a nonlinear average of the iterates produced by any optimization method. The weights in this average are computed via a simple linear system, whose solution can be updated online. This acceleration scheme runs in parallel to the base algorithm, providing improved estimates of the solution on the fly, while the original optimization method is running. Numerical experiments are detailed on classical classification problems.

Alexandre d'Aspremont
CNRS - ENS, Paris
aspremon@di.ens.fr

MS124

Optimization of Spatiotemporal Fractionation in Photon Radiotherapy with Optimality Bounds

Radiotherapy treatments are typically fractionated, which means that radiation dose is delivered over several days or weeks rather than in a single treatment session. In current clinical practice, the total radiation dose is split evenly and the patient is treated in the same manner on every day of the treatment. It has recently been demonstrated that treatments can be improved by altering the radiation dose distribution from day to day in a practice called nonuniform fractionation. In contrast to the uniformly fractionated treatments that can be computed with convex optimization algorithms, nonuniformly fractionated treatments are based on a biological dose-response model called biologically effective dose (BED), which leads to a nonconvex quadratic formulation. In the current work, we solve nonuniformly fractionated treatment plan models to local optimality and use a semidefinite programming relaxation to prove the near-optimality of the computed plans. Using clinical liver cases, we demonstrate that nonuniform fractionation can substantially reduce the biological effect of radiation in the healthy liver tissue while maintaining treatment effectiveness. In addition, we find that the computed locally optimal treatment plans are close to realizing the maximum potential benefit of nonuniform fractionation. This is joint work with David Papp and Jan Unkelbach.

Melissa Gaddy
Department of Mathematics
North Carolina State University
mrgaddy@ncsu.edu

David Papp
North Carolina State University
dpapp@ncsu.edu

Jan Unkelbach
Department of Radiation Oncology
Massachusetts General Hospital and Harvard Medical School
junkelbach@partners.org

MS124

Multi-Modality Optimal Radiation Therapy

Each radiation type, e.g. X-rays, protons, neutrons, currently used for treating cancers has its unique radiobiological power (cell-kills for a given physical dose) and physical dose distribution (relative dose between tumor and normal tissue). We present a multi-modality optimal radiation therapy (MMORT) approach, which balances a trade-off between different radiation types by optimizing the number of fractions and a fractional dose for each modality. We maximize the biological effect, as defined by a linear-quadratic cell survival model, on the tumor while constraining the normal tissue damage to an acceptable level. We first present a spatiotemporally separated MMORT model to gain qualitative insights into the problem. This formulation leads to a QCQP, which is solved using KKT conditions. Second, we present a spatiotemporally integrated formulation. Optimization variables in this case are the radiation intensities of the beamlets, which then determine a spatial dose distribution, and the number of fractions for each modality. Various OAR constraint types prevalent in practice, i.e., mean dose, maximum dose, and dose-volume constraints, are incorporated. Numerical simulations using dual modalities with multiple OAR constraints will be used to provide structural insights into the optimal solutions and quantify the potential benefits of MMORT.

Minsun Kim
Department of Radiation Oncology
University of Washington School of Medicine
mk688@uw.edu

Archis Ghate, Sevnaz Nourollahi
University of Washington
archis@uw.edu, sevnaz@uw.edu

MS124

Biologically-Informed Radiotherapy Planning using a Global Optimization Approach

Radiotherapy is one of the main modalities for cancer treatment. The goal of radiotherapy is to deliver sufficient radiation dose to the tumor region to eradicate the disease while sparing the surrounding healthy tissue to the largest extent possible. To achieve this goal, radiotherapy plans for individual cancer patients are designed to deliver the desired spatial dose distribution to the patient. The radiotherapy plan will be then used on a daily basis to deliver a daily fraction of the prescribed radiation dose over the course of the treatment. However, there is biological evidence suggesting that additional therapeutic gain may be achieved if we allow for temporal variation in the radiotherapy plan. This research aims at developing a spatiotemporal radiotherapy planning approach to evaluate the potential benefit of varying radiotherapy plans and thus the spatial dose distribution over the treatment course. The spatiotemporal planning problem is modeled as a non-convex quadratically-constrained quadratic programming problem, which is solved using global optimization techniques. The proposed approach is applied to a clinical cancer case in order to test the computational performance of the solution method and to quantify the potential therapeutic benefit of varying the radiotherapy plan over the course of the treatment.

Ehsan Salari
Wichita State University
ehsan.salari@wichita.edu

Ali Adibi
Department of Industrial, Systems, and Manufacturing Eng.
Wichita State University
axadibi1@wichita.edu

MS124

Adaptive SBRT Planning for Interfraction Motion

Conventional treatment is delivered in 30 fractions (treatments) over approximately one month. However, SBRT delivers up to 5 fractions in up to two weeks; interfraction motion (changes in patient geometry between fractions) can result in overdosing organs at risk (OAR) near the tumor. Current adaptive strategies may reduce dose to OARs, but at the cost of under-dosing the tumor without compensation in future fractions. We investigate adaptive planning strategies that compensate for fractions with these so-called "unfavorable" geometries that under-dose the tumor by boosting (increasing) dose in fractions with "favorable" geometry (while still satisfying OAR dose limits) via a stochastic programming approach.

Victor Wu
University of Michigan
vwwu@umich.edu

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

Randall Ten Haken
University of Michigan

rtk@umich.edu

Kristy Brock
MD Anderson
kkbrock@mdanderson.org

Martha Matuszak
University of Michigan
marthamm@med.umich.edu

MS125

Asynchronous Parallel Operator Splitting Methods for Convex Stochastic Programming

We describe an decomposition method for convex stochastic programs that resembles progressive hedging but only needs to process a subset of the scenarios at each iteration, can vary the penalty parameter by both iteration and scenario, and can operate asynchronously. We discuss how to implement the algorithm and how it is derived from more general block-iterative and projective splitting frameworks. We describe some applications and computational results.

Jonathan Eckstein
Rutgers University
jeckstei@business.rutgers.edu

Jean-Paul Watson
Sandia National Laboratories
Discrete Math and Complex Systems
jwatson@sandia.gov

MS125

Object-Parallel Augmented Lagrangian Solution of Continuous Stochastic Programs

We describe an "object-parallel' C++ approach to implementing iterative optimization methods, with the goal of coding the algorithms in a readable, MATLAB-like style, but with efficient parallel implementation of underlying low-level linear algebra operations. Recent advances in augmented Lagrangian algorithms and mathematical modeling environments, combined with our "object-parallel' software development techniques suggest efficiently scalable solver implementations for stochastic programming problems, without employing decomposition methods.

Gyorgy Matyasfalvi, Jonathan Eckstein
Rutgers University
matyasfalvi@gmail.com, jeckstei@business.rutgers.edu

MS125

Scenario-Based Decomposition for Parallel Solution of the Contingency-Constrained ACOPF

We present a nonlinear stochastic programming formulation for a large-scale contingency-constrained optimal power flow problem. Using a rectangular IV formulation to model AC power flow in the transmission network, we construct a nonlinear, multi-scenario optimization formulation where each scenario considers nominal operation followed by a failure an individual transmission element. Given the number of potential failures in the network, these problems are very large; yet need to be solved rapidly. In this paper, we demonstrate that this multi-scenario problem can be solved quickly using a parallel decomposition approach based on progressive hedging and nonlinear interior-point methods. Parallel and serial timing results are shown using test cases from Matpower, a MATLAB-based framework for power flow analysis.

Jean-Paul Watson
Sandia National Laboratories
Discrete Math and Complex Systems
jwatson@sandia.gov

Carl Laird, Anya Castillo
Sandia National Laboratories
cdlaird@sandia.gov, arcasti@sandia.gov

MS125

Highly Parallelized, Exact Progressive Hedging for Stochastic MIPs

Progressive hedging, though an effective heuristic for stochastic mixed integer programs (MIPs), is not guaranteed convergence in this case. Here, we describe BBPH, a branch and bound algorithm that uses PH within each node that, given enough time, will always converge to the optimal solution. In addition to providing a theoretically convergent wrapper for PH applied to MIPs, computational results demonstrate that for some difficult problem instances branch and bound can find improved solutions.

David Woodruff
University of California, Davis
dlwoodruff@ucdavis.edu

MS126

Finding Planted Graphs with Few Eigenvalues using the Schur-Horn Relaxation

Extracting structured planted subgraphs inside large graphs is a fundamental question that arises in a range of domains. We describe a tractable approach based on convex optimization to recover certain families of graphs embedded in larger graphs containing spurious edges. Our method relies on tractable semidefinite descriptions of the spectrum of a matrix, and we give conditions on the eigenstructure of a planted graph in relation to the noise level under which our algorithm succeeds.

Utkan Onur Candogan
California Institute Of Technology
utkan@caltech.edu

Venkat Chandrasekaran
California Institute of Technology
venkatc@caltech.edu

MS126

Graph Structure in Polynomial Systems: Chordal Networks

We introduce a novel representation of structured polynomial ideals, which we refer to as *chordal networks*. The sparsity structure of a polynomial system is often described by a graph that captures the interactions among the variables. Chordal networks provide a computationally convenient decomposition of a polynomial ideal into simpler (triangular) polynomial sets, while preserving its underlying graphical structure. We show that many interesting families of polynomial ideals admit compact chordal network representations (of size linear in the number of variables), even though the number of components could be exponentially large. Chordal networks can be computed for arbitrary polynomial systems using a refinement of the chordal elimination algorithm from (Cifuentes2016). Furthermore, they can be effectively used to obtain several properties of the variety, such as its dimension, cardinality and equidimensional components, as well as an efficient probabilistic test for radical ideal membership. We apply our methods to examples from algebraic statistics and vector addition systems; for these instances, algorithms based on chordal networks outperform existing techniques by orders of magnitude.

Diego Cifuentes, Pablo A. Parrilo
Massachusetts Institute of Technology
diegcif@mit.edu, parrilo@MIT.EDU

MS126

Local Graph Profiles: Algorithms and Large-Scale Applications

We study the problem of counting the number of subgraphs incident on each vertex and each edge in a large graph. These are called the vertex-local and edge-local graph profile, respectively, and describe the graphs connectivity structure. We present a state-of-the-art algorithm for local profile counting of 5-subgraphs which can be parallelized in a shared-memory architecture. Additionally, graph profile calculation is applied to graph theoretic optimization problems such as clustering and classification.

Ethan R. Elenberg
The University of Texas at Austin
elenberg@utexas.edu

Alex Dimakis
University of Texas at Austin
dimakis@austin.utexas.edu

MS126

What You Ask Is What You Get: Query Design and Robust Algorithms for Crowdsourced Clustering

We consider the problem of clustering unlabeled data using a crowd of non-expert workers. For example, clustering a database of images of birds into different species with the collective help of workers who are not experts in bird classification. While they may not be able to label an image directly, they can compare the images and judge whether they are similar. As the workers are not experts, the answers obtained are noisy. Therefore it is crucial to (1) design queries that can reduce the noise levels in the responses, and (2) design algorithms to reliably cluster the noisy data. We consider two query models, edge query model (comparing two images per query), and triangle query model (comparing three images per query). Under natural modeling assumptions, we show that triangle queries provide more reliable answers than edge queries. For both the query models, we analyze a robust convex clustering algorithm that attempts to decompose a partially observed adjacency matrix into low-rank and sparse components. In this setting, we obtain sufficient conditions for the exact recovery of the clusters. We also provide experiments on real datasets as well as numerical simulations that validate our theoretical results. In particular, (1) triangle queries significantly outperform edge queries and (2) using the convex algorithms to denoise the data substantially reduces the number of errors in clustering.

Ramya Korlakai Vinayak
California Institute of Technology
ramya@caltech.edu

MS127

Semidefinite Programs with a Dash of Smoothness: Why and When the Low-Rank Approach Works

Semidefinite programs (SDP's) can be solved in polynomial time by interior point methods, but scalability can be an issue. To address this shortcoming, over a decade ago, Burer and Monteiro proposed to solve SDP's with few equality constraints via rank-restricted, non-convex surrogates. Remarkably, for some applications, local optimization methods seem to converge to global optima of these non-convex surrogates reliably. Although some early theory supports this empirical success, a complete explanation of it remains an open question. In this presentation, we show that the Burer-Monteiro formulation of SDP's in a certain class almost never has any spurious local optima, that is: the non-convexity of the low-rank formulation is benign. This class of SDP's covers applications such as max-cut, community detection in the stochastic block model, robust PCA, phase retrieval and synchronization of rotations. The crucial assumption we make is that the low-rank problem lives on a manifold, so that theory and algorithms from optimization on manifolds can be used. Optimization on manifolds is about minimizing a cost function over a smooth manifold, such as spheres, low-rank matrices, orthonormal frames, rotations, etc. We will present the basic framework as well as parts of the more general convergence theory, including recent complexity results. (Toolbox: http://www.manopt.org) Select parts are joint work with P.-A. Absil, A. Bandeira, C. Cartis and V. Voroninski.

Nicolas Boumal
Mathematics Department
Princeton University
nboumal@princeton.edu

MS127

Fast and Provable Algorithms for Low-Rank Hankel Matrix Completion

In this talk, we consider the problem of low-rank Hankel matrix completion, with applications to the reconstruction of spectrally sparse signals from a random subset of n regular time domain samples. We introduce an iterative hard thresholding (IHT) algorithm and a fast iterative hard thresholding (FIHT) algorithm for efficient low rank Hankel matrix completion to reconstruct spectrally sparse signals. Theoretical recovery guarantees have been established for FIHT, showing that $O(r^2 \log^2(n))$ number of samples are sufficient for exact recovery with high probability. Empirical performance comparisons establish significant computational advantages for IHT and FIHT. In particular, numerical simulations on 3D arrays demonstrate the capability of FIHT on handling large and high-dimensional real data.

Jian-Feng Cai
Department of Mathematics
Hong Kong University of Science and Technology
jfcai@ust.hk

Tianming Wang

Department of Mathematics
University of Iowa
tianming-wang@uiowa.edu

Ke Wei
Department of Mathematics
University of California at Davis
kewei@math.ucdavis.edu

MS127

The Projected Power Method: An Efficient Nonconvex Approach for Joint Alignment from Pairwise Differences

Various applications involve assigning discrete label values to a collection of objects based on some noisy data. Due to the discrete—and hence nonconvex—structure of the problem, computing the maximum likelihood estimates (MLE) becomes intractable at first sight. This paper makes progress towards efficient computation of the MLE by focusing on a concrete joint alignment problem—that is, the problem of recovering n discrete variables $\{x_i | i = 1, ..., n\}$ given noisy observations of their modulo differences $\{x_i - x_j \bmod m\}$. We propose a novel low-complexity procedure, which operates in a lifted space by representing distinct label values in orthogonal directions, and which attempts to optimize quadratic functions over hyper cubes. Starting with a first guess computed via a special method, the algorithm successively refines the iterates via projected power iterations. We prove that the proposed projected power method makes no error—and hence converges to the MLE—in a suitable regime. Numerical experiments have been carried out on both synthetic and real data to demonstrate the practicality of our algorithm. We expect this algorithmic framework to be effective for a broad range of discrete assignment problems.

Yuxin Chen
Department of Statistics
Stanford University
yuxin.chen@princeton.edu

Emmanuel Candes
Stanford University
Departments of Mathematics and of Statistics
candes@stanford.edu

MS127

Gradient Descent for Rectangular Matrix Completion

From recommender systems to healthcare analytics low-rank recovery from partial observations is prevalent in modern data analysis. There has been significant progress over the last decade in providing rigorous guarantees for low-rank recovery problems based on convex relaxation techniques. However, the computational complexity of these algorithms render them impractical for large-scale applications. Recent advances in nonconvex optimization explain the surprising effectiveness of simple first-order algorithms for many low rank matrix recovery problems, especially for positive semidefinite matrices. The common theme of these algorithms is to work directly with the low rank factors of the semidefinite variable. In this talk, I will discuss how similar ideas can be applied to rectangular matrix completion. We provide rigorous convergence guarantee to show such simple algorithms are effective and can overcome the scalability limits faced by popular convex relaxation approach.

Qinqing Zheng
Department of Compute Science
University of Chicago
qinqing@cs.uchicago.edu

John Lafferty
Department of Statistics
University of Chicago
lafferty@galton.uchicago.edu

MS128

Local Complexities and Adaptive Algorithms in Stochastic Convex Optimization

We extend the traditional worst-case, minimax analysis of stochastic convex optimization by introducing a localized form of minimax complexity for individual functions. Our main result gives function-specific lower and upper bounds on the number of stochastic subgradient evaluations needed to optimize either the function or its "hardest local alternative" to a given numerical precision. The bounds are expressed in terms of a localized and computational analogue of the modulus of continuity that is central to statistical minimax analysis. We show how the computational modulus of continuity can be explicitly calculated in concrete cases, and relates to the curvature of the function at the optimum. We also prove a superefficiency result that demonstrates it is a meaningful benchmark, acting as a computational analogue of the Fisher information in statistical estimation. We also discuss adaptive algorithms that optimally achieve these rates, showing that fast convergence is often possible.

John C. Duchi
Stanford University
Departments of Statistics and Electrical Engineering
jduchi@stanford.edu

Sabyasachi Chatterjee
University of Chicago
sabyasachi@uchicago.edu

John Lafferty
Department of Statistics
University of Chicago
lafferty@galton.uchicago.edu

Yuancheng Zhu
University of Pennsylvania
yuancheng.zhu@gmail.com

MS128

Efficient Second Order Online Learning by Sketching

We propose Sketched Online Newton (SON), an online second order learning algorithm that enjoys substantially improved regret guarantees for ill-conditioned data. SON is an enhanced version of the Online Newton Step, which, via sketching techniques enjoys a running time linear in the dimension and sketch size. We further develop sparse forms of the sketching methods (such as Oja's rule), making the computation linear in the sparsity of features. Together, the algorithm eliminates all computational obstacles in previous second order online learning approaches.

Alekh Agarwal

Microsoft Research New York
alekha@microsoft.com

Haipeng Luo
Microsoft Research
haipeng@microsoft.com

MS128

Oracle Complexity of Second-Order Methods

Second-order iterative methods, which utilize Hessians as well as gradients, are an integral part of the mathematical optimization toolbox, and have seen a resurgence of interest recently in the context of various large-scale optimization problems. In this talk, I'll describe some recent results on the oracle complexity of such methods, and their (possibly unexpected) limitations in improving on gradient-based methods in terms of worst-case guarantees.

Ohad Shamir, Yossi Arjevani, Ron Shiff
Weizmann Institute of Science
ohad.shamir@weizmann.ac.il,
yossi.arjevani@weizmann.ac.il, ron.shiff@weizmann.ac.il

MS128

Tight Complexity Bounds for Optimizing Composite Objectives

We provide tight upper and lower bounds on the complexity of minimizing the average of m convex functions using gradient and prox oracles of the component functions. We show a significant gap between the complexity of deterministic vs randomized optimization. For smooth functions, we show that accelerated gradient descent and an accelerated variant of SVRG are optimal in the deterministic and randomized settings respectively, and that a gradient oracle is sufficient for the optimal rate. For non-smooth functions, having access to prox oracles reduces the complexity and we present optimal methods based on smoothing that improve over methods using just gradient accesses.

Blake Woodworth
Toyota Technological Institute Chicago
blake@ttic.edu

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

MS129

Switching Controls for Reaction-Diffusion Equation

We present first steps towards an efficient numerical method for optimization of advection-diffusion processes described by a partial differential equation, and subject to non-smoothness arising from phase transitions and from switching controls. Our approach is based on a decomposition approach to mixed-integer optimal control with combinatorial constraints, and on a simultaneous optimization framework, in which a discretized PDE enters the optimization problem as a nonlinear constraint. We discuss the application to a Rankine cycle process, widely established in process engineering for recovery of energy from an exhaust heat source, that has recently found novel use as a promising hybridization concept for heavy duty trucks.

Christian Kirches
Institute for Mathematical Optimization
Braunschweig University of Technology
c.kirches@tu-bs.de

Paul Manns
Technische Universität Braunschweig
pfmanns@gmail.com

MS129

On the (Non)Convexity of Different Mixed-Integer Optimal Control Formulations

There is a huge variety of different approaches to solve mixed-integer optimal control problems. The differences exist on several levels and can usually be combined to obtain the final numerical approach: 1) which method is used to solve the optimal control problem in the first place, e.g., a direct or indirect method, dynamic programming, or moment based approaches; 2) which modeling techniques are used to represent logical implications, e.g., bigM, disjunctive programming, or vanishing constraints; 3) which method is used to discretize the differential equations, e.g., single shooting, multiple shooting, collocation and 4) can variable substitutions or reformulations be applied, e.g., a switching time optimization. We discuss the aspect of (non)convexity of some of these approaches and give insight on different of the aforementioned levels, e.g., multiple shooting vs single shooting in the context of global optimal control and an "increase' in nonconvexity due to the enhanced time transformation technique.

Sebastian Sager
Otto-von-Guericke Universität Magdeburg, Germany
sager@ovgu.de

MS129

New Decomposition Strategies for Mixed-Integer Optimal Control

Mixed-Integer Optimal Control Problems (MIOCPs) are optimization problems that combine the difficulties of combinatorial decisions with underlying dynamical systems. Usually, these combinatorial decisions are characterized as switching decisions between the systems different operations modes. A common example is the optimal gear switching of a truck.
We consider a first discretize, then optimize approach resulting in a mixed-integer nonlinear program (MINLP) combined with the outer convexification relaxation method, where the binary variables only enter affinely in the MIOCP. Sager [S. Sager. Numerical methods for mixed-integer optimal control problems. Toenning, Luebeck, Marburg: Der andere Verlag; 2005] proposed to decompose the MINLP into a NLP and a MILP. In this study we focus on new theoretical and numerical insights regarding this decomposition process. We mainly investigate the associated MILP, also described as combinatorial integral approximation (CIA) or control approximation problem in integral sense.
Commonly used approaches in this context are the sum up rounding strategy and branch and bound [S. Sager, M. Jung, C. Kirches. Combinatorial integral approximation. Mathematical Methods of Operations Research 73.3 (2011): 363-380.], which we both apply to a altered version of the CIA.

Clemens Zelle
Otto-von-Guericke Universitaet
Magdeburg, Germany

clemens.zeile@ovgu.de

Sebastian Sager
Otto-von-Guericke Universität Magdeburg, Germany
sager@ovgu.de

MS130

Decomposable Nonsmooth Optimization: Methods and Applications

In this talk, we consider nonsmooth optimization problems with decomposable structures. Such decomposable structures include special structures such as the difference of convex decompositions and smooth compositions of nonsmooth convex functions. We will discuss versions of the bundle method for solving such problems. Preliminary numerical results will be presented and comparison of these methods with other methods of nonsmooth optimization will be reported. We will also consider some applications of decomposable nonsmooth optimization in machine learning and regression analysis.

Adil Baghirov
CIAO, School of ITMS,
The University of Ballarat
a.bagirov@federation.edu.au

MS130

A Chain Rule for Convexly Generated Spectral Max Functions

Eigenvalue optimization problems arise in the control of continuous and discrete time dynamical systems. The spectral abscissa (the largest real part of an eigenvalue of a matrix) and the spectral radius (the largest eigenvalue in modulus) are examples of functions of eigenvalues, or spectral functions, connected to such problems. More specifically, the abscissa and radius are spectral max functions—the maximum of a real-valued function over the spectrum of a matrix. In 2001, Burke and Overton characterized the regular subdifferential of the spectral abscissa and showed that the spectral abscissa is subdifferentially regular in the sense of Clarke when all active eigenvalues are nonderogatory. In this talk, we describe the new techniques used to obtain these results for the a more general class of convexly generated spectral max functions, and demonstrate their application to the spectral radius.

Julia Eaton
University of Washington
Tacoma
jreaton@uw.edu

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

MS130

Epsilon Subdifferential Computation in Computational Convex Analysis

Computational convex analysis focuses on the computation of convex operators that routinely appear in convex analysis e.g. Legendre-Fenchel conjugate, Moreau envelope, etc. After recalling what computation the (open source) CCA numerical library allows, we will switch the focus from convex transforms like the Legendre-Fenchel transform to convex operators like the subdifferential. We will present results on efficiently computing the epsilon subdifferential of a convex univariate function.

Yves Lucet
UBC
yves.lucet@ubc.ca

Warren Hare
University of British Columbia, Okanagan Campus
warren.hare@ubc.ca

Anuj Bajaj
Wayne State University
anuj.bajaj91@gmail.com

MS130

The Nesterov Smoothing Technique and Minimizing Differences of Convex Functions with Applications to Facility Location and Clustering

In this talk, we propose new optimization algorithms to solve a number of problems in multifacility location and hierarchical clustering. Our methods are based on the Nesterov smoothing technique and algorithms for minimizing differences of convex functions. The algorithms obtained improve existing methods for solving these problems.

Mau Nam Nguyen
Fariborz Maseeh Department of Mathematics and Statistics
Portland State University
mnn3@pdx.edu

Wondi Geremew
Stockton University
wondi.geremew@stockton.edu

Sam Reynolds, Tuyen Tran
Portland State University
ser6@pdx.edu, tuyen2@pdx.edu

MS131

Network-Flow Polytopes Satisfy the Hirsch Conjecture

We solve a problem in the combinatorics of polyhedra motivated by the network simplex method. We show that the Hirsch conjecture holds for the diameter of the graphs of all network-flow polytopes, in particular the diameter of a network-flow polytope for a network with n nodes and m arcs is never more than $m + n - 1$. A key step to prove this is to show the same result for classical transportation polytopes.

Steffen Borgwardt
University of Colorado Denver
Department of Mathematical and Statistical Sciences
steffen.borgwardt@ucdenver.edu

MS131

Superlinear Diameters for Combinatorial Abstractions of Polyhedra

The asymptotic growth of the diameters of polyhedra and their combinatorial abstractions is relevant to the efficiency of the simplex method for linear optimization. In this talk, we outline the construction of a combinatorial abstraction of polyhedra whose diameter is superlinear, which gives evidence against the Linear Hirsch Conjecture. This is joint work with Tristram C. Bogart.

Edward Kim
University of Wisconsin-La Crosse
ekim@uwlax.edu

MS131

On the Circuit Diameter Conjecture

A key concept in optimization is the combinatorial diameter of a polyhedron. From the point of view of optimization, we would like to relate it to the number of facets f and dimension d of the polyhedron. In the seminal paper of Klee and Walkup, the Hirsch conjecture, that the bound is $f-d$, was shown to be equivalent to several seemingly simpler statements, and was disproved for unbounded polyhedra through the construction of a particular 4-dimensional polyhedron with 8 facets. The Hirsch bound for polytopes was only recently narrowly exceed by Santos. We consider analogous questions for a variant of the combinatorial diameter called the circuit diameter. In this variant, paths are built from the circuit directions of the polyhedron, and can travel through the interior. We are able to recover the equivalence results that hold in the combinatorial case. Further, we show that validity of the circuit analogue of the non-revisiting conjecture for polytopes would imply a linear bound on the circuit diameter of all unbounded polyhedra. Finally, we prove a circuit version of the 4-step conjecture. Our methods require adapting the notion of simplicity to work with circuits. We show that it suffices to consider such circuit simple polyhedra for studying circuit analogues of the Hirsch conjecture, and use this to prove the equivalences of the different variants. This is joint work with S. Borgwardt and T. Yusun.

Steffen Borgwardt
University of Colorado Denver
Department of Mathematical and Statistical Sciences
steffen.borgwardt@ucdenver.edu

Tamon Stephen, Timothy Yusun
Simon Fraser University
tamon@sfu.ca, tyusun@sfu.ca

MS131

A New Model for Realization Spaces of Polytopes

The theory of psd lifts of polytopes offers a new model for realization spaces of polytopes via the slack ideal associated to the polytope. In this talk I will discuss this model and various features of it.

Rekha Thomas
University of Washington
rrthomas@uw.edu

Joao Gouveia
Universidade de Coimbra
jgouveia@mat.uc.pt

Antonio Macchio
University of Coimbra
macchia.antonello@gmail.com

Amy Wiebe
University of Washington
awiebe@uw.edu

MS132

Classical Approximation Algorithms for Quantum Constraint Satisfaction Problems

The study of approximation algorithms for Boolean satisfiability problems such as MAX-k-SAT is a well-established line of research. In the quantum setting, there is a physically motivated generalization of MAX-k-SAT known as the k-Local Hamiltonian problem (k-LH), which is of interest for two reasons: From a complexity theoretic perspective, k-LH is complete for the quantum analogue of NP, and from a physics perspective, k-LH asks one to estimate the energy of a quantum system when cooled to very low temperatures. For the latter reason in particular, the condensed matter physics community has devoted decades to developing heuristic algorithms for k-LH and related problems. However, recent years have witnessed the development of the first classical approximation algorithms for k-LH. This talk will overview and discuss some of these existing results, as well as (time permitting) preview recent work in progress on generalizing the Goemans-Williamson algorithm for MAX-CUT to approximate physically motivated special cases of k-LH (joint work with Yi-Kai Liu (NIST, USA)).

Sevag Gharibian
Virginia Commonwealth University
sgharibian@vcu.edu

MS132

Parameter Tuning for Optimization in Quantum Annealing Processors

D-Wave quantum computing systems employ quantum annealing, a computing paradigm related to simulated annealing but different in several important ways. In this talk I will provide an overview of the D-Wave computing framework from the perspective of a classical optimization researcher: How can I make this thing solve a constraint satisfaction problem? I will then give a gentle introduction to practicalities. How can we minimize the effect of noise and error when solving a problem? Finally, I will introduce a more advanced perspective: What is a quantum Boltzmann distribution, and what can it do for you? This talk is intended for a general optimization audience.

Andrew King
D-Wave Systems
aking@dwavesys.com

MS132

Analyzing Quantum Cryptographic Protocols using Semidefinite Programming

Die-rolling is the cryptographic task where two mistrustful, remote parties wish to generate a random D-sided die-roll over a communication channel. Optimal quantum protocols for this task have been given by Aharon and Silman (New Journal of Physics, 2010) but are based on optimal weak coin-flipping protocols which are currently very complicated and not very well understood. In this talk, I will first present very simple classical protocols for die-rolling which have decent (and sometimes optimal) security which is in stark contrast to many other two-party cryptographic tasks. I will also present quantum protocols and discuss how to analyze their security using semidefinite programming. By modifying optimal solutions to these semidefinite programs, I will show how to design quantum protocols which have security provably better than the given classical protocols.

Jamie Sikora
Centre for Quantum Technologies
National University of Singapore
jamiesikora@gmail.com

MS132

Quantum Speed-Ups for Semidefinite Programming

We give a quantum algorithm for solving semidefinite programs (SDPs). It has worst case running time $n^{\frac{1}{2}} m^{\frac{1}{2}} s \ \mathrm{poly}(\log(n), \log(m), R, 1/\delta)$, with n and s the dimension and sparsity of the input matrices, respectively, m the number of constraints, δ the accuracy of the solution, and R an upper bound on the trace of the optimal solution. This gives a square-root unconditional speed-up over any classical method for solving SDPs both in n and m. We prove the algorithm cannot be substantially improved giving a $\Omega(n^{\frac{1}{2}} + m^{\frac{1}{2}})$ quantum lower bound for solving semidefinite programs with constant s, R and δ. In some instances, the algorithm offers even exponential speedups.

Krysta Svore
Microsoft Research
ksvore@microsoft.com

MS133

Perspective-Based Proximal Data Processing

We show that many problems in data processing involve convex perspective functions. We analyze pertinent properties of such functions and show how to handle them in proximal algorithms. Applications to high-dimensional statistics are presented.

Patrick L. Combettes
Department of Mathematics
North Carolina State University
plc@math.ncsu.edu

MS133

Efficient Bayesian Computation by Proximal Markov Chain Monte Carlo: When Langevin meets Moreau

Modern imaging methods rely strongly on Bayesian inference techniques to solve challenging problems. Currently, the predominant Bayesian computation approach is convex optimisation, which scales efficiently to high dimensions and delivers accurate point estimation results. However, to perform more advanced analyses it is necessary to use more computationally intensive techniques such as Markov chain Monte Carlo (MCMC) methods. This paper presents a new and highly efficient MCMC methodology to perform Bayesian computation for high dimensional models that are log-concave and non-smooth, a class of models that is key in imaging sciences. The method is based on a regularised unadjusted Langevin algorithm that uses Moreau-Yoshida envelopes and proximal operators to construct Markov chains with favourable convergence properties. In addition to scaling efficiently to high dimensions, the method is straightforward to apply to models that are currently solved by proximal optimisation. We provide a detailed theoretical analysis of the method, including asymptotic and non-asymptotic convergence results with easily verifiable conditions, and explicit bounds on the convergence rates. The methodology is demonstrated with experiments related to image deconvolution and tomographic reconstruction where we conduct challenging Bayesian analyses related to uncertainty quantification and model selection without ground truth available. Joint work with Dr. Alain Durmus and Prof. Eric Moulines.

Marcelo Pereyra
School of Mathematics
University of Bristol
m.pereyra@hw.ac.uk

MS133

Estimating High Dimensional Generalized Additive Models via Proximal Methods

In this talk, we discuss a general framework for sparse additive regression. We allow the class of each additive component to be quite general (characterized by seminorm smoothness this includes monotonicity of derivatives, sobolev/holder smoothness). We show that by minimizing a simple convex problem, we can estimate these functions at the minimax rate (over functions in that smooth additive class). In addition we show that the penalized regression problem can be efficiently solved using a proximal gradient descent algorithm: Each prox-step decouples into p-univariate penalized regression problems; each of these univariate penalized problems can in turn be written as a simple update of the solution to a univariate non-parametric regression problem (for which we often have efficient algorithms). For some important classes (eg. total-variation smoothness), this means the additive model can be fit in roughly the same order of computation as the LASSO (n*p operations per iteration), and in practice can scale easily to any dataset which fits in memory. In addition, because the prox-step splits over the features, our algorithm is simply to parallelize.

Noah Simon
Department of Biostatistics
University of Washington
nrsimon@u.washington.edu

MS133

Alternating Projections, ADMM, and Parallel Coordinate Descent

Abstract not available at time of publication.

Ryan Tibshirani
Department of Statistics and Machine Learning
Department
Carnegie Mellon University
ryantibs@cmu.edu

MS134

A Unified Optimization View on Generalized Matching Pursuit and Frank-Wolfe

Two of the most fundamental prototypes of greedy optimization are the matching pursuit and Frank- Wolfe algorithms. In this paper we take a unified view on both classes of methods, leading to the first explicit convergence rates of matching pursuit methods in an optimization sense, for general sets of atoms. We derive sublinear (1/t) convergence for both classes on general smooth objectives, and linear convergence on strongly convex objectives, as well as a clear correspondence of algorithm variants. Our presented algorithms and rates are affine invariant, and do not need any incoherence or sparsity assumptions.

Francesco Locatello
ETH Zürich
francesco.locatello@gmail.com

Rajiv Khanna
UT Austin
rajivak@utexas.edu

Michael Tschannen
ETH Zürich
michaelt@nari.ee.ethz.ch

Martin Jaggi
Berkeley
martin.jaggi@epfl.ch

MS134
Three Variations on a Frank-Wolfe Theme

In this presentation, we discuss two frameworks where an adaptation of the Frank-Wolfe linearization strategy for solving monotone variational inequalities achieves fast (at least linear) convergence. This analysis yields insight into the determination of efficient FW directions.

Patrice Marcotte
DIRO, CRT
Université de Montréal
marcotte@iro.umontreal.ca

MS134
Forward-Backward Methods for Atomic Norm Constrained Optimization

In many signal processing and machine learning applications, the aim is to reconstruct a signal that has a simple representation with respect to a certain basis or frame. Fundamental elements of the basis known as atoms allow us to define atomic norms that can be used to formulate convex regularizations for the reconstruction problem. Efficient algorithms are available to solve these formulations in certain special cases, but an approach that works well for general atomic norms, both in terms of speed and reconstruction accuracy, remains to be found. I will describe an algorithm that produces solutions with succinct atomic representations for reconstruction problems. The algorithm (called CoGEnT) combines the Conditional Gradient method with additional enhancement steps to reduce the basis size. Experimental evidence suggests that such a method is superior to standard greedy approaches on several problems of interest, while also retaining the convergence guarantees of the standard Conditional Gradient methods.

Nikhil Rao
Dept. of ECE
University of Wisconsin
nikhilrao86@gmail.com

MS134
Conditional Gradient Sliding for Convex Optimization

In this talk, we present a new conditional gradient type method for convex optimization by calling a linear optimization (LO) oracle to minimize a series of linear functions over the feasible set. Different from the classic conditional gradient method, the conditional gradient sliding (CGS) algorithm developed herein can skip the computation of gradients from time to time, and as a result, can achieve the optimal complexity bounds in terms of not only the number of calls to the LO oracle, but also the number of gradient evaluations. More specifically, we show that the CGS method requires $\mathcal{O}(1/\sqrt{\epsilon})$ and $\mathcal{O}(\log(1/\epsilon))$ gradient evaluations, respectively, for solving smooth and strongly convex problems, while still maintaining the optimal $\mathcal{O}(1/\epsilon)$ bound on the number of calls to LO oracle. We also develop variants of the CGS method which can achieve the optimal complexity bounds for solving stochastic optimization problems and an important class of saddle point optimization problems. To the best of our knowledge, this is the first time that these types of projection-free optimal first-order methods have been developed in the literature. Some preliminary numerical results have also been provided to demonstrate the advantages of the CGS method.

Guanghui Lan
Georgia Tech
tba

Yi Zhou
University of Florida
sonrisa_zhou@hotmail.com

PP1
An Optimal Investment Strategy with Maximal Risk Aversion

In this work we consider the problem of an insurance company where the wealth of the insurer is described by a Cramer-Lundberg process. The insurer has the possibility to invest in an incomplete market that includes a risky asset with stochastic volatility and a bank account. The risky asset is correlated to an economic factor modeled as a diffusion process. Under the exponential utility function a Hamilton-Jacobi-Bellman equation is solved to obtain the optimal investment strategy maximizing the expected utility function for insurers. We suggest a mixed Finite Difference Monte-Carlo method to solve numerically the (HJB) equation. In order to show the connection between the Insurers decision and the correlation factor several simulations are presented in the case of Scott model.

Mohamed Badaoui
National Polytechnic Institute, IPN
Escuela Superior de Ingeniería Mecánica y Eléctrica, ESIME-Z
mbadaoui@ipn.mx

PP1
Compact Representation of the Full Broyden Class of Quasi-Newton Updates

Quasi-newton methods are an effective tool used to find the local maxima or minima of a function when second order information is unavailable or too expensive. Compact representations of these approximations allow us to efficiently compute their eigenvalues thereby giving us the ability to perform sensitivity analysis on the resultant matrices. In this talk we present the compact representation for matrices belonging to the full Broyden class of quasi-Newton updates allowing for different members of the Broyden class to be used at each iteration. Furthermore, we demonstrate how to compute the inverse compact representation of these matrices and demonstrate through numerical experiments how they can be used to efficiently solve linear systems involving Broyden class matrices to high accuracy.

Omar DeGuchy
Applied Mathematics Unit
University of California, Merced
odeguchy@ucmerced.edu

Jennifer Erway
Department of Mathematics
Wake Forest University
erwayjb@wfu.com

Roummel F. Marcia
University of California, Merced
rmarcia@ucmerced.edu

PP1

Working Breakdown in a Repairable Queue with Fixed Size Bulk Service and Variable Breakdown and Variable Repairable Rates

This paper considers the repairable queuing system with service in a fixed batch size with one operating unit. It is assumed that the operating units may breakdown with different rates while in operation, either in idle state or in busy state, respectively, according to the Poisson manner. When the system is down, it under-goes two processes, first a substitute server is serving with slower rate with same fixed batch size rule, we call this state as 'working breakdown state' and secondly, a repairman will repair the defective server, where the repair time follows an exponential distribution with variable repairable rate depending on the substitute server's operational state, whether it is idle or busy, respectively. The customers are arriving to the system according to the Poisson manner depending on the state of the server (idle, busy, idle working breakdown and busy working breakdown state). A customer on arrival either decide to join the system with some probability or balk immediately. We use probability generating function technique to derive the explicit expressions for the steady state probabilities of the system length, the steady state availability of the server and various performance measures of the model.

Gopal K. Gupta, A. Banerjee
Dept. Of Mathematical Science
IIT (BHU) , Varanasi, India
gopalgupta.bhu90@gmail.com, anuradha.mat@iitbhu.ac.in

PP1

Second Order Riemannian Methods for Low-Rank Tensor Completion

The goal of tensor completion is to fill in missing entries of a partially known tensor (possibly inculding some noise) under a low-rank constraint. This may be formulated as a least-squares problem. The set of tensors of a given multilinear rank is known to admit a Riemannian manifold structure, thus methods of Riemannian optimization are applicable. In our work, we derive the Riemannian Hessian of an objective function on the low-rank tensor manifolds and discuss the convergence properties of second order methods for the tensor completion problem, both theoretically and numerically. We compare our approach to Riemannian tensor completion methods from recent literature. Our examples include the recovery of multidimensional images, approximation of multivariate functions and recovery of partially missing data from survey statistics.

Gennadij Heidel
Trier University
heidel@uni-trier.de

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

PP1

Modeling and Optimization of New Feeding Systems in the Pig Industry: Linear and Bilinear Problems

Feed represents 70% of the production cost in the pig industry,so, in the current economic context, it is important to reduce it. This study is mainly focus on modeling the feeding system that satisfy all animal requirements. Currently, farmers use a three phase feeding system. For each phase a single feed is used and the energy density of each one is fixed. In that case,the associated mathematical model is linear. In a previous study, we found a method that reduces feeding cost by 4.0%. The idea behind this feeding system is to take two feeds and to blend them daily in order to satisfy the requirements. The mathematical model associated to it is a bilinear model. Recently we propose a new feeding system that mix both methods (feeding system using phases and feeding system using two feeds). The concept is the following:for each phase, one is a feeding system using two feeds (optimized at the same time as the total cost minimization and not necessarily complete) which will be blended together to satisfy the daily animal requirements. For example, for three phases, one will use four feeds (A, B, C, and D):A and B will be used in the first phase, B and C in the second phase and finally C and D in the last phase. Compared to the traditional feeding system, this leads to a feed cost reduction of 5.2%. All these models are described in this poster as well as the numerical results. We also included tricriterion model and results that minimize feeding cost, and excretion.

Emilie Joannopoulos
Insa de Rennes / Université de Sherbrooke
emilie.joannopoulos@insa-rennes.fr

Francois Dubeau, Jean-Pierre Dussault
Université de Sherbrooke
francois.dubeau@usherbrooke.ca,
Jean-Pierre.Dussault@Usherbrooke.ca

Mounir Haddou
Insa de Rennes
mhaddou@insa-rennes.fr

Candido Pomar
Agriculture and Agri-Food Canada
pomarc@em.agr.ca

PP1

A Weighted Random Forest Algorithm to Improve

Classification Performance

Random forest (RF) is a machine learning method for classification problems. It applies bootstrap and random attribution selection to construct a number of trees for prediction. In original RF, each tree has equal weights during aggregation. In reality, some trees always perform better than others. So weighted random forest (wRF) with heavier weight for better performance tree may potentially improve the overall predictive performance. The cardiotocography dataset from https://archive.ics.uci.edu/ml/datasets/Cardiotocography was downloaded to compare the predictive performance of original RF and wRF. Half of the data was used as the training data to construct the RF model. The other half of the data was the testing dataset to test the RF model and computer the predictive accuracy for each tree. The predictive accuracy was then used to calculate the aggregation weight for each tree. The weight was calculated according the logistic regression curve with predictive accuracy as the independent variable. With the weights calculated, the testing dataset was again classified with the RF model, but this time the aggregation was done with weighted vote from each tree. The predictive accuracy rates for original RF and wRF are 94.08% and 96.15% respectively. The results show improved predictive performance.

Weihui Li
Wentworth Institute of Technology
liw1@wit.edu

Xiaxin Mei
wentworth institute of technology
meix1@wit.edu

PP1

Conic Program Certificates from ADMM/Douglas-Rachford Splitting

Many practical optimization problems are formulated as conic programs, a current active area of research. One challenge with conic programs is that one often does not know, a priori, whether there is a solution to the conic program; it may be infeasible, unbounded or has a finite but unattainable optimal value. Solvers like *SeDuMi* and *SDPT3* provide certificates for strong infeasibility, but they struggle with weak infeasibility. To resolve this, we propose a method that finds (1) a solution when one exists (2) a certificate of strong/weak infeasibility when the problem is strongly/weakly infeasible (3) an unbounded direction when there is one and (4) a sufficient condition for identifying programs with finite but unattainable optimal values. The method is based on Douglas-Rachford Splitting (ADMM), and we establish the efficacy of it through theoretical analysis and numerical experiments.

Yanli Liu, Ernest Ryu
University of California, Los Angeles
Department of Mathematics
yanli@math.ucla.edu, eryu@math.ucla.edu

Wotao Yin
University of California, Los Angeles
wotaoyin@math.ucla.edu

PP1

Hybrid Method in Solving Large-Scale Alternating-Current Optimal Power Flows

Many steady-state problems in power systems, including rectangular power-voltage formulations of optimal power flows in the alternating-current model (ACOPF), can be cast as non-convex polynomial optimization problems (POP). For a POP, one can derive strong convex relaxations, or rather hierarchies of ever stronger, but ever larger relaxations. Readily available second-order methods for solving convex relaxations on real-world large-scale power systems have been popular due to the fast quadratic local convergence. However, they usually suffer from three main issues, namely high computational costs, prohibitive memory requirements and convergence to local optima. Randomized first-order methods (RMs) have been successful in machine learning, due to their much lower per-iteration computational and memory requirements, convergence to the optima of convex problems with high probability, and acceleration with the parallel/distributed variants. We hence study means of switching from RMs for solving a convex relaxation to Newtons method working on the original non-convex problem, which allows for convergence under the same conditions as in solvers for the convex relaxation, but with an improved rate of convergence. We also propose a backtracking framework which allows the practical search for globally optimal solutions. We illustrate our approach on the ACOPF using Polish instances and demonstrate the benefits of employing the hybrid schema.

Jie Liu
Lehigh University
jild13@lehigh.edu

Alan Liddell
University of Notre Dame
Dept. Appl. Comp. Math. Stat.
aliddel1@nd.edu

Jakub Marecek
IBM Research Ireland
jakub.marecek@ie.ibm.com

Martin Takac
University of Edinburgh
takac.mt@gmail.com

PP1

Stochastic Optimization Framework for a Fresh Fruit Supply Chain

We present a stochastic optimization framework for the dragon fruit supply chain in Vietnam. The dragon fruit supply chain faces several uncertainties in yields, price, and demand, being highly sensitive to weather conditions and global uncertainty factors. The risk factors in the dragon fruit supply chain also depend on species for example, the red varieties, while more profitable than the white varieties, also have higher export risk because they are subject to global prices and adverse geopolitical conditions. The stochastic planning model framework is developed based on the two-stage stochastic approach suggested from Darby-Dowman et al. (2000). In the first stage, we determine optimal planting plans for dragon fruit under second-stage uncertainties. The second stage optimization problem looks at optimizing harvest and distribution through standard trader channels. The timing of sowing and harvesting decisions is taken into account, because the dragon fruit plant has a six year life-cycle with maximum average yield in the fourth year. Another modelling consideration in this industry is the prevalence of forward-buying contracts with locked-in prices. Some preliminary computational results will be discussed.

Tri-Dung Nguyen
Dalhousie University
Industrial Engineering Department, Faculty of Engineering
dn755440@dal.ca

Uday Venkatadri
Dalhousie University
Industrial Engineering Department, Faculty of Engineering Sp
uday.venkatadri@dal.ca

Tri Nguyen-Quang
Dalhousie University
Biofluids and Biosystems Modeling Lab. (BBML)
tri.nguyen-quang@dal.ca

Claver Diallo
Dalhousie University
Industrial Engineering Department, Faculty of Engineering Sp
claver.diallo@dal.ca

PP1

Hybrid Hu-Storey Method for Large-Scale Nonlinear Monotone Systems

Nonlinear monotone systems arise in various practical situations and applications, so many iterative methods for solving these systems have been developed. We propose a hybrid method for solving large-scale monotone systems, which is based on derivative-free conjugate gradient approach and hyperplane projection technique. The conjugate gradient approach is efficient for large-scale systems due to low memory, while projection method is suitable for monotone equations because it enables simply globalization. The derivative-free, function-value-based line search is combined with Hu-Storey search direction and projection procedure, in order to construct a globally convergent method. Numerical experiments indicate the robustness of proposed method.

Sanja Rapajic, Papp Zoltan
University of Novi Sad, Faculty of Sciences,
Department of Mathematics and Informatics
sanja@dmi.uns.ac.rs, pappzoki@gmail.com

PP1

Bacterial Community Analysis via Convex Optimization

Metagenomics is the study of communities of bacteria through all their sampled DNA. One of the main problems of interest is identifying the presence and relative abundance of organisms in a given environmental sample. Previous approaches, namely the algorithm Quikr and subsequent iterations, have cast this problem as an ℓ_2 regularized ℓ_1 regression ($\min_x ||x||_1 + \lambda||Ax - y||_2^2$) which chooses the most parsimonious mixture of columns of the training database (encoded in the matrix A) that fit the data (encoded by k-mer counts in the vector y). However, the matrix A can have similar if not indistinguishable columns. One way to capture the uncertainty about which columns of A to choose is to instead use the Ordered weighted ℓ_1-norm (OWL-norm) instead of the ℓ_1-norm and recast the problem as ℓ_2 regularized OWL-norm regression. We reduce this problem to a Non-negative least square optimization with linear inequality constraints and investigate an active set algorithm to solve this problem efficiently.

David Koslicki, Hooman Zabeti
Oregon State University
david.koslicki@math.oregonstate.edu, zabetih@oregonstate.edu

PP1

PDE Constrained Optimal Control of Hydropower Systems

Hydropower is the most widely used renewable source of energy, however the operations present many challenges. We formulate the reservoir operation problem as a continuous in time optimal control problem to determine the reservoir outflow which maximizes the revenue produced, yet also meets long term planning targets. The problem is constrained by the PDE dynamics within the channel. Additionally we have system state bound and ramping constraints. We solve for the optimal control using the adjoint approach and show numerical results for a test problem.

John Zaheer, Nathan Gibson
Oregon State University
zaheerj@oregonstate.edu, gibsonn@oregonstate.edu