12th US-Mexico Workshop on Optimization and Its Applications

January 9-13, 2023

Camino Real Zaashila
Huatulco, Oaxaca, Mexico

Contents

About

US-Mexico WOA

The 12th US-Mexico Workshop on Optimization and its Applications will be held from January 9–January 13, 2023 at Camino Real Zaashila, Huatulco, Oaxaca, Mexico. The workshop will also hold a special dinner celebrating Stephen Wright's (belated) 60th birthday.

Organizers

Jeff Linderoth, University of Wisconsin, Madison
Harvey D. Spangler Professor, Industrial & Systems Engineering

Jorge Nocedal, Northwestern University
Director, Center for Optimization and Statistical Learning
Walter P. Murphy Professor of Industrial Engineering and Management Sciences

Katya Scheinberg, Cornell University
Professor, Operations Research and Information Engineering

All local organization and registration was done by Eunae Jo, Northwestern University.

Sponsor

Northwestern | McCORMICK SCHOOL OF ENGINEERING
Center for Optimization and Statistical Learning

Participants

José Miguel Saavedra Aguilar	Centro de Investigación en Matemáticas
Ahmet Alacaoglu	University of Wisconsin-Madison
Mihai Anitescu	Argonne National Lab
Necdet Serhat Aybat	Penn State University
Albert Berahas	University of Michigan
Raghu Bollapragada	University of Texas at Austin
Johannes Brust	University of California, San Diego
Oscar Dalmau Cedeño	Centro de Investigación en Matemáticas
Mateo Diaz Diaz	Caltech
Lijun Ding	University of Wisconsin-Madison
Maryam Fazel	University of Washington
Michael Ferris	University of Wisconsin-Madison
Don Goldfarb	Columbia University
Oktay Günlük	Cornell University
Mert Gürbüzbalaban	Rutgers University
Simge Küçükyavuz	Northwestern University
Kangwook Lee	University of Wisconsin-Madison
Qin Li	University of Wisconsin-Madison
Haihao Lu	University of Chicago
Jeff Linderoth	University of Wisconsin-Madison
Jim Luedtke	University of Wisconsin-Madison
Matt Menickelly	Argonne National Lab
Miguel Oscar Almarales Milan	Centro de Investigación en Matemáticas
Hector Morales	Universidad Autónoma Metropolitana
David Morton	Northwestern University
Jorge Nocedal	Northwestern University
Courtney Paquette	McGill University
Ben Recht	University of California, Berkeley
Clément Royer	Université Paris Dauphine-PSL
Michael Saunders	Stanford University
Katya Scheinberg	Cornell University
Luis Nunes Vicente	Lehigh University
Andreas Wächter	Northwestern University
Stefan Wild	Berkeley Lab
Rebecca Willett	University of Chicago
Stephen Wright	University of Wisconsin-Madison

Timetable

Monday, January 9

9:15-9:35	**Welcoming Remarks**		
9:35–10:10	IS	**Albert Berahas** University of Michigan	Algorithms for Deterministically Constrained Stochastic Optimization
10:10–10:45	IS	**Raghu Bollapragada** University of Texas at Austin	Adaptive Sampling Sequential Quadratic Programming for Stochastic Constrained Optimization
10:45–11:05	**Coffee**		
11:05–11:40	IS	**Johannes Brust** University of California, San Diego	Weighted Trust-Region Methods with Hessian Estimates
11:40–12:15	IS	**Clément Royer** Université Paris Dauphine-PSL	Algorithms and Application for Special Classes of Nonlinear Least Squares Problems 2023
12:15–16:00	**Lunch and Break**		
16:00–16:35	IS	**Stephen Wright** University of Wisconsin-Madison	TBD
16:35–17:10	IS	**Ahmet Alacaoglu** University of Wisconsin-Madison	On the Complexity of a Practical Primal-Dual Coordinate Method
17:10–17:45	IS	**Necdet Serhat Aybat** Penn State University	SAPD+: An Accelerated Stochastic Method for Nonconvex-Concave Minimax Problems
17:45–18:05	**Coffee**		
18:05–18:40	IS	**Mert Gürbüzbalaban** Rutgers University	Robust and Risk-Averse Accelerated Gradient Methods
18:40–19:15	IS	**Kangwook Lee** University of Wisconsin-Madison	On a Bilevel Optimization Approach to Fair Classification

Tuesday, January 10

9:00–9:35	IS	**Mihai Anitescu** Argonne National Lab	Exponential Decay of Sensitivity in Dynamic Programming and other Graph-Indexed Optimization
9:35–10:10	IS	**Andreas Wächter** Northwestern University	A Smoothing-Based Decomposition Algorithm for Nonlinear Two-Stage Problems
10:10–10:45	IS	**Don Goldfarb** Columbia University	Efficient Second-Order Stochastic Methods for Deep Learning
10:45–11:05	**Coffee**		
11:05–11:40	IS	**Qin Li** University of Wisconsin-Madison	Interplay between differential equation and data
11:40-12:15	IS	**Rebecca Willett** University of Chicago	Optimization-inspired machine learning in the natural sciences
12:15–16:35	**Lunch and Break**		
16:35–17:10	IS	**Courtney Paquette** McGill University	Stochastic Algorithms in the Large: Batch Size Saturation, Stepsize Criticality, Generalization Performance, and Exact Dynamics
17:10–17:45	IS	**Mateo Diaz Diaz** Caltech	Clustering a mixture of Gaussians with unknown covariance
17:45–18:05	**Coffee**		
18:05–18:40	IS	**Lijun Ding** University of Wisconsin-Madison	Semidefinite programming: Conditioning and algorithm in its data science applications
18:40–19:15	IS	**Haihao Lu** University of Chicago	New Developments of First Order Methods for Linear Programming

Wednesday, January 11

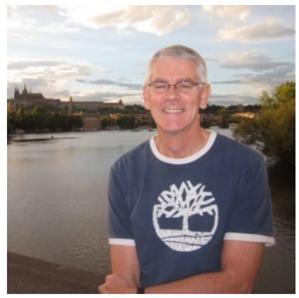
We will speak with the hotel and arrange for potential group outings for this day. A menu of potential outings will be presented.

19:00—	**Workshop Dinner**

The banquet dinner will honor Steve Wright's many contributions to mathematical optimization

---

Sometimes, there's a man, well, he's the man for his time and place. He fits right in there. And that's Steve Wright, in Optimization.



Thursday, January 12

9:00–9:35	IS	**Matt Menickelly** Argonne National Lab	Exploiting Structure in (Derivative-Free) Composite Nonsmooth Optimization
9:35–10:10	IS	**Jorge Nocedal** Northwestern University	Nonlinear Optimization in the Presence of Noise
10:10–10:45	IS	**Stefan Wild** Berkeley Lab	Randomized algorithms for large-scale derivative-free stochastic optimization
10:45–11:05	**Coffee**		
11:05–11:40	IS	**Luis Nunes Vicente** Lehigh University	A weak tail-bound probabilistic condition for function estimation in stochastic derivative-free optimization
11:40–12:15	IS	**Maryam Fazel** University of Washington	Multiplayer Performative Prediction: Learning in Decision-Dependent Games
12:15–16:00	**Lunch and Break**		
16:00–16:35	IS	**Jim Luedtke** University of Wisconsin-Madison	Improving Power Grid Resiliency with Bi-objective Stochastic Integer Optimization
16:35–17:10	IS	**Oktay Günlük** Cornell University	Incorporating Dantzig-Wolfe Decomposition Into Branch-and-Cut by Cutting Planes
17:10–17:45	IS	**Simge Küçükyavuz** Northwestern University	On Constrained Mixed-Integer DR-Submodular Minimization
17:45–18:05	**Coffee**		
18:05–18:40	IS	**Michael Saunders** Stanford University	Algorithm NCL to the rescue when LICQ fails
18:40–19:15	IS	**Ben Recht** University of California, Berkeley	This Isn’t ’Nam, There Are Rules: A History of Computational Game Play

Friday, January 13

9:00–9:35	IS	**Michael Ferris** University of Wisconsin-Madison	Optimization approaches to combat Gerrymandering
9:35–10:10	IS	**Hector Morales** Universidad Autónoma Metropolitana	A Variational Approach of Waveform Design In Tomographic Imaging Systems
10:10–10:45	IS	**Katya Scheinberg** Cornell University	Lower Bounds on Step Size Parameters in Adaptive Stochastic Methods
10:45–11:05	**Coffee**		
11:05–12:00	**Panel Discussion, Topic TBD**		
12:00	**Fin**		

List of Abstracts – Talks

Monday, January 9

Algorithms for Deterministically Constrained Stochastic Optimization

Albert Berahas IS

University of Michigan

In this talk, we propose and analyze stochastic optimization algorithms for deterministically constrained problems based on the sequential quadratic optimization (commonly known as SQP) methodology. We discuss the rationale behind our proposed techniques, convergence in expectation and complexity guarantees for our algorithms, and the results of preliminary numerical experiments that we have performed. This is joint work with Raghu Bollapragada, Frank E. Curtis, Michael O'Neill, Daniel P. Robinson, Jiahao Shi and Baoyu Zhou.

Adaptive Sampling Sequential Quadratic Programming for Stochastic Constrained Optimization

Raghu Bollapragada IS

University of Texas at Austin

This talk presents a methodology for using varying sample sizes in sequential quadratic programming (SQP) methods for solving equality-constrained stochastic optimization problems. We develop adaptive criteria for controlling the sample size, and the accuracy in the solutions of the SQP subproblems based on the variance estimates obtained as the optimization progresses. We establish global convergence results and sample complexity results for the proposed method and demonstrate the performance of the method on a subset of the CUTE problems and constrained classification tasks.

Weighted Trust-Region Methods with Hessian Estimates

Johannes Brust IS

University of California, San Diego

For unconstrained optimization problems in which only gradients of the objective function are available, we develop a trust-region method that updates a symmetric factorization of the Hessian approximation. Typically, in trust-region methods the main computational challenge is the solution of the trust-region subproblem to determine the next search direction. In order to overcome this challenge, the factored form of the matrix can be used to define trust-region subproblems in terms of weighted norms, which significantly reduce the computational costs of a subproblem. We develop two solvers based on a weighted l-2 and a weighted l-infinity norm. The methods are tested on a vast set of unconstrained optimization problems from CUTEst, on which they are competitive with state-of-the-art line-search implementations. (joint work with Philip E. Gill)

Algorithms and application for special classes of nonlinear least squares problems 2023

Clément Royer IS

Université Paris Dauphine-PSL

Nonlinear least squares are one of the most prevalent problems in nonlinear optimization, with applications ranging from data assimilation to machine learning. Those problems can also be found throughout Steve's career starting with the very topic of his PhD thesis.

In this talk, I will describe recent results on the complexity of algorithms for nonlinear least squares problems, with a focus on Gauss-Newton-type frameworks. I will cover the metrics of interest for such complexity results, and explain how those considerations extend to other formulations that arise in machine learning. I will then propose an algorithm with complexity guarantees for least-squares problems and beyond. Finally, I will present an application of this method to a special class of problems arising while training neural differential equations.

TBD

Stephen Wright IS

University of Wisconsin-Madison

TBD

On the Complexity of a Practical Primal-Dual Coordinate Method

Ahmet Alacaoglu IS

University of Wisconsin-Madison

We prove complexity bounds for the primal-dual algorithm with random extrapolation and coordinate descent (PURE-CD), which has been shown to obtain good practical performance for solving convex-concave min-max problems with bilinear coupling. Such problems arise in many machine learning contexts, including linear empirical risk minimization, matrix games, and image processing. Our results either match or improve the best-known complexities for dense and sparse (strongly)-convex-(strongly)-concave problems with bilinear coupling among first-order algorithms.

SAPD+: An Accelerated Stochastic Method for Nonconvex-Concave Minimax Problems

Necdet Serhat Aybat IS

Penn State University

We propose a new stochastic method SAPD+ for solving nonconvex-concave minimax problems of the form $\min \max L(x, y) = f(x) + \Phi(x, y) - g(y)$ where f, g are closed convex and $\Phi(x, y)$ is a smooth function that is weakly convex in x, (strongly) concave in y. For both strongly concave and merely concave settings, SAPD+ achieves the best known oracle complexities of $O(L\kappa/\epsilon^4)$ and $O(L^3/\epsilon^6)$, respectively, without assuming compactness of the problem domain, where κ is the condition number and L is the Lipschitz constant. We also propose SAPD+ with variance reduction, which enjoys the best known oracle complexity of $O(L\kappa^2/\epsilon^3)$ for weakly convex-strongly concave setting. We demonstrate the efficiency of SAPD+ on a distributionally robust learning problem with a weakly convex cost and also on a multi-class classification problem in deep learning.

Robust and Risk-Averse Accelerated Gradient Methods

Mert Gürbüzbalaban IS

Rutgers University

In the context of first-order algorithms subject to random gradient noise, we study the trade-offs between the convergence rate (which quantifies how fast the initial conditions are forgotten) and the "risk" of suboptimality, i.e., deviations from the expected suboptimality. We focus on a general class of momentum methods (GMM) that recover popular methods such as gradient descent (GD), accelerated gradient descent (AGD), and the heavy-ball (HB) method as special cases depending on the choice of GMM parameters. We use well-known risk measures "entropic risk" and "entropic value at risk" to quantify the risk of suboptimality. For strongly convex smooth minimization, we first obtain new convergence rate results for GMM with a unified theory that applies to both AGD and HB, improving some of the existing results for HB. We then provide explicit bounds on the entropic risk and entropic value at risk of suboptimality at a given iterate which also provides explicit bounds on the probability that the suboptimality exceeds a given threshold based on Chernoff's inequality. Our results unveil fundamental trade-offs between the convergence rate and the risk of suboptimality and result in Heisenberg-style uncertainty principles. We then plug the entropic risk and convergence rate estimates we obtained in a computationally tractable optimization framework and propose entropic risk-averse GMM (RA-GMM) and entropic risk-averse AGD (RA-AGD) methods that can select the GMM parameters to systematically trade-off the entropic value at risk with the convergence rate. We show that RA-AGD and RA-GMM improve performance on quadratic optimization and logistic regression problems compared to the standard choice of parameters. To our knowledge, our work is the first to resort to coherent measures to design the parameters of momentum methods systematically. We will then discuss how these ideas can be leveraged to develop efficient algorithms for non-convex optimization and min-max optimization problems, including those arising in distributionally robust stochastic optimization where we provide a provably convergent algorithm for non-smooth non-convex losses that are generalized differentiable in the sense of Norkin. This is joint work with Bugra Can, Yassine Laguel, Xuan Zhang, Serhat Aybat, Landi Zhu and Andrzej Ruszczyński.

On a bilevel optimization approach to fair classification

Kangwook Lee IS

University of Wisconsin-Madison

Bilevel optimization has been widely used in various machine learning areas such as meta-learning, hyperparameter optimization, and adversarial machine learning. In our recent work, we have proposed a bilevel optimization-based approach to fair machine learning. Our proposed framework, which we named FairBatch, keeps the standard training algorithm intact, viewing it as an inner optimizer. It then iteratively updates the weights of the training samples, where the update rules are dictated by a simple iterative solution to the underlying bilevel optimization problem. FairBatch's design is in stark contrast to the existing machine learning applications of bilevel optimization. While bilevel formulations are almost inherent in the existing applications, FairBatch's bilevel formulation looks rather artificial. Surprisingly, we show that this bilevel structure, introduced by design, allows us to tackle challenging problems, thanks to its compositional structure. We will share three recent applications of FairBatch: (1) fair training with corrupted data, (2) fair training with decentralized data, and (3) black-box fair training.

Tuesday, January 10

Exponential Decay of Sensitivity in Dynamic Programming and other Graph-Indexed Optimization

Mihai Anitescu IS

Argonne National Lab

We review recent results concerning the phenomenon of exponential decay of sensitivity in dynamic programming and its generalization to arbitrary graph-indexed structures. We describe its consequences for domain decomposition algorithms for such problems and for model predictive control. We discuss some of the remaining challenges.

A Smoothing-Based Decomposition Algorithm for Nonlinear Two-Stage Problems

Andreas Wächter IS

Northwestern University

We present a novel decomposition algorithm for nonlinear two-stage problems that utilizes properties of barrier functions to smooth the non-differentiabilty of the recourse functions. The advantages of this approach is that it can exploit the robustness and efficiency of existing nonlinear solvers and can easily be parallelized. Experimental results for large-scale optimal power flow problems are presented for networks that are decomposed into the transmission system and many distribution system.

Efficient Second-Order Stochastic Methods for Deep Learning

Don Goldfarb IS

Columbia University

We consider the training of Deep Neural Networks (DNNs), which due to the enormous number of parameters current DNNs have, using the Hessian matrix or a full approximation to it in a second-order method is prohibitive. Hence, we have proposed, and will describe in our talk, second-order quasi-Newton (QN), natural gradient (NG), and generalized Gauss-Newton (GGN) methods that use layer-wise block-diagonal approximations to these matrices that are competitive with and often outperform first-order methods. These methods include those that use layer-wise (i) Kronecker-factored BFGS and L-BFGS QN approximations, (ii) tensor normal covariance and (iii) mini-block Fisher matrix approximations, and (iv) Sherman-Morrison-Woodbury based variants of NG and GGN methods.

Interplay between differential equation and data

Qin Li IS

University of Wisconsin-Madison

To understand physical systems, we draw information from both differential equation (DE) modeling and data. DEs provide fundamental models and data fills in detailed descriptions, usually through the form of PDE-constrained optimization. The understanding of DEs also guides data collection, through the perspective of experimental design. In this talk, we showcase two examples in this front. Joint work with Ke Chen, Kit Newton, Steve Wright, Tan Bui and Leonardo Zepeda-Nunez.

Optimization-inspired machine learning in the natural sciences

Rebecca Willett

IS

University of Chicago

Machine learning has the potential to transform scientific research. This fundamental change cannot be realized through the straightforward application of existing off-the-shelf machine learning tools alone. Rather, we need novel methods for incorporating physical models and constraints into learning systems. In this talk, I will describe opportunities and emerging tools for addressing these challenges in the context of inverse problems, data assimilation, and simulator calibration – and how machine learning yields methods with high predictive skill and computational efficiency. These methods build upon insights from optimization and highlight how it can inspire and inform the design of machine learning systems in the natural sciences.

Stochastic Algorithms in the Large: Batch Size Saturation, Stepsize Criticality, Generalization Performance, and Exact Dynamics

Courtney Paquette

IS

McGill University

In this talk, I will present a framework for analyzing dynamics of stochastic optimization algorithms (e.g., stochastic gradient descent (SGD) and momentum (SGD+M)) when both the number of samples and dimensions are large. For the analysis, I will introduce a stochastic differential equation, called homogenized SGD. We show that homogenized SGD is the high-dimensional equivalent of SGD – for any quadratic statistic (e.g., population risk with quadratic loss), the statistic under the iterates of SGD converges to the statistic under homogenized SGD when the number of samples n and number of features d are polynomially related. By analyzing homogenized SGD, we provide exact non-asymptotic high-dimensional expressions for the training dynamics and generalization performance of SGD in terms of a solution of a Volterra integral equation. The analysis is formulated for data matrices and target vectors that satisfy a family of resolvent conditions, which can roughly be viewed as a weak form of delocalization of sample-side singular vectors of the data. By analyzing these limiting dynamics, we can provide insights into learning rate, momentum parameter, and batch size selection. For instance, we identify a stability measurement, the implicit conditioning ratio (ICR), which regulates the ability of SGD+M to accelerate the algorithm. When the batch size exceeds this ICR, SGD+M converges linearly at a rate of $O(1/\kappa)$, matching optimal full-batch momentum (in particular performing as well as a full-batch but with a fraction of the size). For batch sizes smaller than the ICR, in contrast, SGD+M has rates that scale like a multiple of the single batch SGD rate. We give explicit choices for the learning rate and momentum parameter in terms of the Hessian spectra that achieve this performance. Finally we show this model matches performances on real data sets.

Clustering a mixture of Gaussians with unknown covariance

Mateo Diaz Diaz IS

Caltech

Clustering is a fundamental data scientific task with broad applications. This talk investigates a simple clustering problem with data from a mixture of Gaussians that share a common but unknown and potentially ill-conditioned covariance matrix. We consider Gaussian mixtures with two equally-sized components and derive a Max-Cut integer program based on maximum likelihood estimation. We show its solutions achieve the optimal misclassification rate when the number of samples grows linearly in the dimension, up to a logarithmic factor. However, solving the Max-cut problem appears to be computationally intractable. To overcome this, we develop an efficient iterative algorithm that attains the optimal rate but requires a quadratic sample size. Although this sample complexity is worse than that of the Max-cut problem, we conjecture that no polynomial-time method can perform better. Furthermore, we present numerical and theoretical evidence that supports the existence of a statistical-computational gap.

Semidefinite programming: conditioning and algorithm in its data science applications

Lijun Ding IS

University of Wisconsin-Madison

Semidefinite programming (SDP) forms a class of convex optimization problems with remarkable modeling power. Apart from its classical applications in combinatorics and control, it also enjoys a range of applications in data science. This talk first discusses various concrete SDPs in data science and their conditioning. In particular, we show that even though Slater's constraint qualification condition may fail, these SDPs are still primal simple, i.e., they satisfy primal solution uniqueness and strict complementarity, which ensures the convergence of many existing algorithms. In the second part of the talk, based on the simplicity and computational structure shared by these problems, we design time- and space-efficient algorithms to solve these SDPs.

New Developments of First Order Methods for Linear Programming

Haihao Lu

IS

University of Chicago

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

Thursday, January 12

Exploiting Structure in (Derivative-Free) Composite Nonsmooth Optimization

Matt Menickelly

IS

Argonne National Lab

We present new methods for solving a broad class of bound-constrained nonsmooth composite minimization problems. These methods are specially designed for objectives that are some known non smooth mapping of outputs from a computationally expensive function. We provide rigorous convergence analysis and guarantees, and test the implementations extensively.

Nonlinear Optimization in the Presence of Noise

Jorge Nocedal IS

Northwestern University

We begin by presenting three case studies that illustrate the nature of noisy optimization problems arising in practice. They originate in atmospheric sciences, machine learning, and engineering design. We wish to understand the source of the noise (e.g. a lower fidelity model, sampling or reduced precision arithmetic), its properties, and how to estimate it. This sets the stage for the presentation of our goal of redesigning constrained and unconstrained nonlinear optimization methods to achieve noise tolerance.

Randomized algorithms for large-scale derivative-free stochastic optimization

Stefan Wild IS

~~Argonne National Lab~~ Berkeley Lab

We seek to solve high (at least for the derivative-free setting) dimensional problems by employing Johnson-Lindenstrauss transforms and thereby working in a trust-region framework in a reduced space. Joint work with Kwassi Joseph Dzahini.

A weak tail-bound probabilistic condition for function estimation in stochastic derivative-free optimization

Luis Nunes Vicente IS

Lehigh University

We use tail bounds to define a tailored probabilistic condition for function estimation that eases the theoretical analysis of stochastic derivative-free optimization methods. We focus on the unconstrained minimization of a potentially non-smooth function, whose values can only be estimated via stochastic observations, and give a simplified convergence proof for both a direct search and a basic trust-region scheme. We also study the trade-off between noise, algorithmic parameters, and number of samples needed per iteration. We prove in particular that under mild assumptions on the noise only $O(\Delta^{-2-\epsilon})$ samples are necessary at every iteration for any $\epsilon > 0$, rather than the $O(\Delta^{-4k})$ samples required in other works. This is joint work with Francesco Rinaldi and Damiano Zeffiro, University of Padova, Italy.

Multiplayer Performative Prediction: Learning in Decision-Dependent Games

Maryam Fazel

IS

University of Washington

When multiple learning algorithms interact with user populations, the population data reacts to competing decision makers' actions. In this talk we formulate a game-theoretic model to describe and study this phenomenon, called multi-player performative prediction. We show that for this model under mild assumptions, several algorithms, including repeated minimization and repeated stochastic gradient methods, converge to the so-called "performatively stable" points. Nash equilibria of the game are also of interest, but can be found efficiently only when the game is monotone; we present sufficient conditions for strong monotonicity of the game and use them to develop algorithms for finding Nash equilibria. Joint work with A. Narang, E. Faulkner, D. Drusvyatskiy, and L. Ratliff.

Improving Power Grid Resiliency with Bi-objective Stochastic Integer Optimization

Jim Luedtke

IS

University of Wisconsin-Madison

Designing a power grid that is both efficient on average and resilient to extreme weather events is a critical challenge. Traditional stochastic programming approaches to this are highly sensitive to sampling error due to the presence of low probability events with very high impacts. Stochastic programming also fails to account for system goals changing under extreme conditions. We present a bi-objective modeling approach that addresses these issues and illustrate it in the context of capacity planning in the electric grid.

Incorporating Dantzig-Wolfe Decomposition Into Branch-and-Cut by Cutting Planes

Oktay Günlük IS

Cornell University

Dantzig-Wolfe (DW) decomposition is a well-known technique in mixed integer programming (MIP) for decomposing and convexifying constraints to obtain potentially strong dual bounds. We investigate Fenchel cuts that can be derived using the DW decomposition algorithm and show that these cuts can provide the same dual bounds as DW decomposition. We show that these cuts, in essence, decompose the objective function cut one can simply write using the DW bound. Compared to the objective function cut, these Fenchel cuts lead to a formulation with lower dual degeneracy, and consequently a better computational performance under the standard branch-and-cut framework in the original space. We also discuss how to strengthen these cuts to improve the computational performance further. We test our approach on the Multiple Knapsack Assignment Problem and show that the proposed cuts are helpful in accelerating the solution time without the need to implement branch-and-price. (Joint work with Rui Chen and Andrea Lodi)

On Constrained Mixed-Integer DR-Submodular Minimization

Simge Küçükyavuz IS

Northwestern University

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

Algorithm NCL to the rescue when LICQ fails

Michael Saunders

IS

Stanford University

For general constrained optimization problems, LANCELOT is not troubled by LICQ because it solves a short sequence of bound-constrained subproblems. We call it a BCL method (bound-constrained augmented Lagrangian). Algorithm NCL solves an equivalent sequence of nonlinearly constrained subproblems that are suitable for interior methods such as IPOPT and KNITRO.

The AMPL implementation of NCL solved a specific (taxation policy) model with many nonlinear inequality constraints. The Julia implementation can solve the same model and more general problems from CUTEst.

This Isn't 'Nam, There Are Rules: A History of Computational Game Play

Ben Recht

IS

University of California, Berkeley

This talk will explore the history of using computers to play games. Beginning with ideas from Shannon, I will describe the optimization techniques developed to first play competitive Checkers and eventually to outmatch humans at Go. Along the way, I'll describe how these algorithmic ideas influenced our conceptions of strategies to master gameplay, war, and diplomacy.

Friday, January 13

Optimization approaches to combat Gerrymandering

Michael Ferris

IS

University of Wisconsin-Madison

Fair maps have been suggested to combat Gerrymandering using properties such as connectedness, compactness and population bounds as particular guiding principles. The political science literature points to this leading to unintentional Gerrymandering due to political biases in dense urban and sparse rural areas. We demonstrate this fact and propose new districting policies that provide maps that are robust to Gerrymandering practices. The approach uses optimization procedures to assign wards to districts, utilizes multi-representative districts with new assignment policies, and suggests maps and strategies from a model that is based on combatting adversarial users. The results will be demonstrated on a number of different case studies in the United States.

A Variational Approach of Waveform Design In Tomographic Imaging Systems

Hector Morales IS

Universidad Autónoma Metropolitana

In this talk, we analyze the problem of optimal waveform design for imaging systems through a dispersive medium. We use a scalar model for wave propagation, together with the single-scattering approximation, and we assume that measurements are polluted with thermal noise whose statistics are known. For image formation, we use a filtered back-projection algorithm in which the filter is determined by knowledge of the power spectral densities of the scene and noise. In this framework, we derive an optimal waveform by minimizing the mean-square error of the reconstructed image. We show the simulation results for imaging point scatterers embedded in a specific dispersive background. We show that for low signal-to-noise ratios, the optimal waveform resembles what is known as a precursor: a wave generated from propagating ultrawideband waveforms through the medium.

Lower Bounds on Step Size Parameters in Adaptive Stochastic Methods

Katya Scheinberg IS

Cornell University

Recently a variety of stochastic variants of adaptive methods have been developed and analyzed. These include stochastic step search, trust region and cubicly regularized Newton methods. Such methods adapt the step size parameter and use it to dictate the accuracy required or stochastic approximations. The step size parameters in these method can increase and decrease based on the perceived progress, but unlike the deterministic case they are not bounded away from zero. This, in principle, means that arbitrarily accurate estimates may be required by the methods. We show that it is possible to derive a lower bound on step size parameters in high probability for the methods in this general framework. This among other things, implies the first total sample complexity bound for these methods when the estimates are based on averaging stochastic samples.