DIMACS/TRIPODS/MOPTA

13-17 August '18

http://coral.ie.lehigh.edu/~mopta

Lehigh University
Bethlehem, PA, USA

Program and Abstracts

■ A very warm Welcome to DIMACS/TRIPODS/MOPTA

The Modeling and Optimization: Theory and Applications (MOPTA) conference is an annual event aiming to bring together a diverse group of people from both discrete and continuous optimization, working on both theoretical and applied aspects. The format consists of invited talks from distinguished speakers and selected contributed talks, spread over three days. This year the annual MOPTA conference is combined with an NSF-TRIPODS sponsored three day summer school for doctoral students and the NSF-DIMACS sponsored workshop on Optimization in Machine Learning which is a part of the DIMACS/Simons Collaboration on Bridging Continuous and Discrete Optimization. The summer school is held during the three days preceding the conference and is designed for doctoral students interested in improving their theoretical and practical skills related to optimization approaches in machine learning. The DIMACS sponsored workshop brings together invited lectures by top experts in the field as well as contributed poster presentations. The MOPTA part of the conference this year includes a variety on exciting new developments from different optimization areas and a special focus stream on optimization in energy. The conference brings together researchers from both theoretical and applied communities who do not usually have the chance to interact in the framework of a medium-scale event. The conference schedule and the mountain-top location are aimed to enhance interaction among the participants during breaks, catered lunches, a poster session with a cocktail reception and the conference banquet.

■ Directions

Conference Location

The **conference** will take place at:
Lehigh University
Iacocca Conference Center
111 Research Drive
Bethlehem, PA 18015

Conference Dinner Location

The **Cocktail Reception** is in the
University Center Grace Dining Room, 2nd Floor
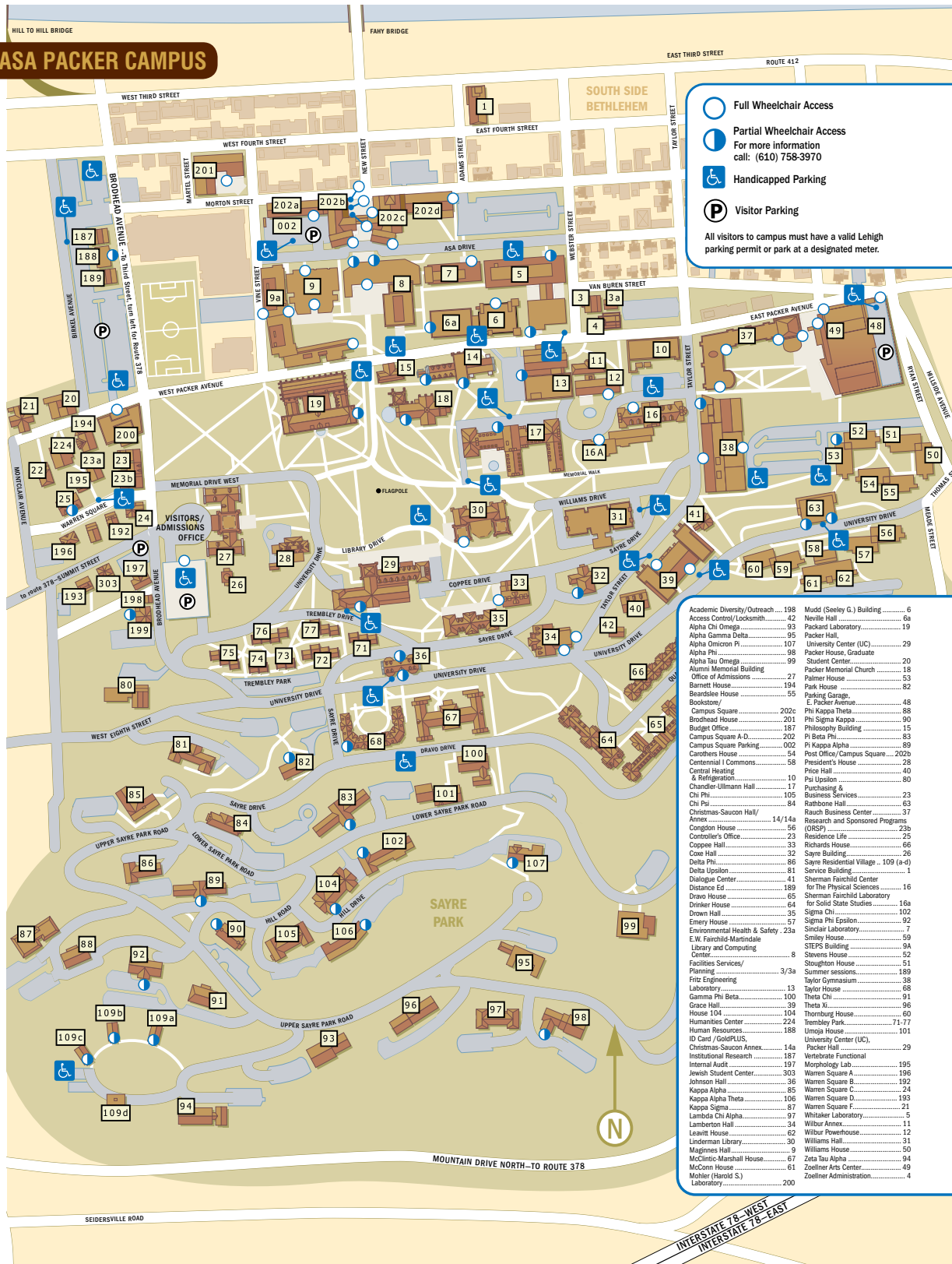29 Trembley Drive, Bethlehem, PA 18015

The **Banquet** is in the
University Center Faculty Lounge
ASA Packer Dining Room, 3rd Floor
29 Trembley Drive, Bethlehem, PA 18015

Graduate Student Social Location

Graduate students attending the Graduate Student Social – Please know that Packer House (see ASA Packer Campus map - Building 20), where the social is located, is within walking distance of Comfort Suites.
If you need to leave during the conference, please see the registration desk for information about taxis and car services.

■ Maps

ASA PACKER CAMPUS

HILL TO HILL BRIDGE

FAHY BRIDGE

EAST THIRD STREET

ROUTE 412

SOUTH SIDE BETHLEHEM

○ Full Wheelchair Access

◑ Partial Wheelchair Access
For more information
call: (610) 758-3970

♿ Handicapped Parking

Ⓟ Visitor Parking

All visitors to campus must have a valid Lehigh parking permit or park at a designated meter.

WEST THIRD STREET

WEST FOURTH STREET

EAST FOURTH STREET

NEW STREET

ADAMS STREET

WEBSTER STREET

TAYLOR STREET

MARTEL STREET

MORTON STREET

BRODHEAD AVENUE—to Third Street, turn left for Route 378

BIRKEL AVENUE

VINE STREET

ASA DRIVE

VAN BUREN STREET

EAST PACKER AVENUE

WEST PACKER AVENUE

MONTCLAIR AVENUE

to route 378—SUMMIT STREET

BRODHEAD AVENUE

MEMORIAL DRIVE WEST

MEMORIAL WALK

WILLIAMS DRIVE

SAYRE DRIVE

UNIVERSITY DRIVE

WARREN SQUARE

VISITORS/ ADMISSIONS OFFICE

● FLAGPOLE

LIBRARY DRIVE

COPPEE DRIVE

TREMBLEY DRIVE

TREMBLEY PARK

UNIVERSITY DRIVE

WEST EIGHTH STREET

SAYRE DRIVE

DRAVO DRIVE

LOWER SAYRE PARK ROAD

UPPER SAYRE PARK ROAD

HILL ROAD

HILL DRIVE

SAYRE PARK

MOUNTAIN DRIVE NORTH—TO ROUTE 378

SEIDERSVILLE ROAD

N

THOMAS STREET

HILLSIDE AVENUE

RYAN STREET

MEADE STREET

INTERSTATE 78—WEST
INTERSTATE 78—EAST

revised August 2013

LEHIGH UNIVERSITY

LEHIGH UNIVERSITY® CAMPUS MAPS

Alpha Tau Omega 99

Ben Franklin
TechVentures 116

Building D/Central
Heating 114

Building J/Mailing
and Printing 118

Gatehouse,
Mountaintop 119

Iacocca Hall 111

Imbt Laboratories 117

Jordan Hall 115

Sayre Field
Comfort Station 110

MOUNTAINTOP CAMPUS

LOOKOUT

LOWER SAYRE PARK ROAD

MOUNTAIN DRIVE NORTH

99

LOOKOUT DRIVE

SAYRE FIELD

110

UPPER SAYRE PARK ROAD

MOUNTAIN DRIVE SOUTH

MOUNTAIN DRIVE NORTH—TO ROUTE 378

ATLSS DRIVE

RESEARCH DRIVE

119

116

TO MURRAY H. GOODMAN CAMPUS

115

114

111

RESEARCH DRIVE

117

118

WILLIAMS STREET

APPLE ST.

N

Full Wheelchair Access

Partial Wheelchair Access
For more information
call: (610)758-3970

Parking

All visitors to campus must have valid Lehigh
parking permits or park at a designated meter.

MOUNTAIN DRIVE SOUTH

INTERSTATE 78—WEST

INTERSTATE 78—EAST

Child Care Center
(More House) 134

Communications
and Public Affairs.........125

Cundey Varsity House ...121

Diamond House.............143

Field Shop 127

Gibson House.................140

Murray H. Goodman
Stadium..........................150

Hartman House 137

Lewis Tennis Center......128

Manufacturers'
Resource Center............125

More House...................134

Mulvihill Golf Complex..149

Rauch Field House.......123

Severs House................131

Small Business
Development Center.....125

Squash Courts 122

Stabler Athletic &
Convocation Center......124

Stadium House.............150

Transportation
Services.........................126

Ulrich Complex 151

Baseball Field ....E

Cross Country
Start and Finish.................A

Kaufmann Fields
Soccer Practice C

Kaufmann Fields
Softball ....B

Outdoor Track ....F

South Fields......................D

Tennis Courts ....G

Whitehead Football
Practice Fields..................H

MURRAY H. GOODMAN CAMPUS

COLLEGE DRIVE

INTERSTATE 78—WEST

INTERSTATE 78—EAST

149

A

151

150

GOODMAN DRIVE

MOUNTAIN DRIVE SOUTH

H

121

122

F

D

123

124

C

G

E

B

127

128

125

131

134

126

DUH DRIVE

137

CREEK ROAD

N

FRIEDENSVILLE ROAD

140

143

BINGEN ROAD

Full Wheelchair Access

Partial Wheelchair Access
For more information
call: (610)758-3970

Parking

All visitors to campus must have valid Lehigh
parking permits or park at a designated meter.

revised August 2012

■ Plenary Talks

Monday 09:00–10:00. Suvrit Sra

Tractable nonconvex optimization via geometry

EECS, MIT

suvrit@mit.edu

Currently, in machine learning there is intense interest in nonconvex optimization. This interest is fueled by the rise of deep neural networks, and also by other more complex tasks in related areas. Although an understanding of why neural networks work so well remains elusive, there has been impressive progress in algorithms, software, and systems for nonconvex optimization. But in today's talk, I want to take a step back from algorithmic advances (fast nonconvex SGD, escaping saddle-point, etc.) — instead, I want to draw your attention to a new set of tools that expand our repertoire of tractable nonconvex optimization. In particular, I will present a rich subclass of nonconvex problems that can be solved to global optimality (or failing that, solved numerically more efficiently). The geometric concept that I'll discuss is geodesic convexity, which generalizes the usual vector-space (linear) notion of convexity to nonlinear spaces. I will outline how geometric thinking leads to improved models or insights for fundamental tasks in machine learning and statistics, including large-scale PCA, metric learning, and Gaussian mixture models. I will outline both theoretical and practical aspects, including iteration complexity theory, and conclude with some open problems.

Speaker Biography. Suvrit Sra is a faculty member in the EECS Department at MIT, and also a core faculty member of the Laboratory for Information and Decision Systems (LIDS) and the Institute for Data, Systems, and Society (IDSS) at Massachusetts Institute of Technology (MIT), as well as a member of the MIT-ML group. He obtained his PhD in Computer Science from the University of Texas at Austin. Before moving to MIT, he was a Senior Research Scientist at the Max Planck Institute for Intelligent Systems, Tübingen, Germany. He has held visiting faculty positions at UC Berkeley (EECS) and Carnegie Mellon University (Machine Learning Department) during 2013-2014. His research bridges a number of mathematical areas such as metric and differential geometry, matrix analysis, convex analysis, probability theory, and optimization with machine learning. He has been a co-chair for OPT2008-2016, NIPS workshops on "Optimization for Machine Learning," and has also co-edited a book with the same name (MIT Press, 2011).

Monday 13:30–14:30. Peter Bartlett

Representation, optimization and generalization properties of deep neural networks

UC Berkeley

peter@berkeley.edu
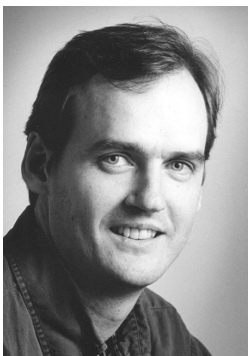
Deep neural networks have improved the state-of-the-art performance for prediction problems across an impressive range of application areas. This talk describes some recent results in three directions. First, we investigate the impact of depth on representational properties of deep residual networks, which compute near-identity maps at each layer, showing how their representational power improves with depth and that the functional optimization landscape has the desirable property that stationary points are optimal. Second, we study the implications for optimization in deep linear networks, showing how the success of a family of gradient descent algorithms that regularize towards the identity function depends on a positivity condition of the regression function. Third, we consider how the performance of deep networks on training data compares to their predictive accuracy, we demonstrate deviation bounds that scale with a certain "spectral complexity," and we compare the behavior of these bounds with the observed performance of these networks in practical problems.
[Joint work with Steve Evans, Dylan Foster, Dave Helmbold, Phil Long, and Matus Telgarsky.]

Speaker Biography. Peter Bartlett is a professor in the Computer Science Division and Department of Statistics and Associate Director of the Simons Institute for the Theory of Computing at the University of California at Berkeley. His research interests include machine learning and statistical learning theory. He is the co-author, with Martin Anthony, of the book Neural Network Learning: Theoretical Foundations. He has served as an associate editor of the journals Bernoulli, the Journal of Artificial Intelligence Research, the Journal of Machine Learning Research, the IEEE Transactions on Information Theory, Machine Learning, Mathematics of Operations Research, and Mathematics of Control Signals and Systems, and as program committee chair for COLT and NIPS. He was awarded the Malcolm McIntosh Prize for Physical Scientist of the Year in Australia in 2001, was chosen as an Institute of Mathematical Statistics Medallion Lecturer in 2008 and an IMS Fellow and Australian Laureate Fellow in 2011, and was elected to the Australian Academy of Science in 2015.

Tuesday 13:30–14:30. John Duchi

Better models in optimization

Stanford University

jduchi@stanford.edu

Many iterative methods for optimization use first- and second-order information to iteratively construct, then minimize, local models of the objective to be minimized. In this talk, I will discuss work my group has been doing on stochastic and non-stochastic optimization, in which we leverage alternative structure than standard first- or second-order information, which can often yield dramatic improvements in convergence and optimization accuracy. As particular applications, I will demonstrate the best-known empirical results (with strong theoretical guarantees) for solving phase retrieval problems, among others. Based on joint work with Hilal Asi and Feng Ruan.

Speaker Biography. John Duchi is an assistant professor of Statistics and Electrical Engineering and (by courtesy) Computer Science at Stanford University, with graduate degrees from UC Berkeley and undergraduate degrees from Stanford. His work focuses on large scale optimization problems arising out of statistical and machine learning problems, robustness and uncertain data problems, and information theoretic aspects of statistical learning. He has won a number of awards and fellowships, including best paper awards at the Neural Information Processing Systems conference, the International Conference on Machine Learning, an NSF CAREER award, a Sloan Fellowship in Mathematics, the Okawa Foundation Award, and the Association for Computing Machinery (ACM) Doctoral Dissertation Award (honorable mention).

Wednesday 09:00–10:00. Kilian Weinberger

Deep Learning with Dense Connectivity

Cornell University

kqw4@cornell.edu
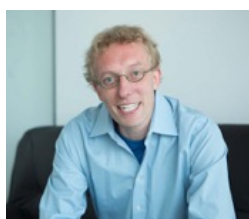
Although half a decade has passed since Frank Rosenblatt's original work on multi- layer perceptrons, modern artificial neural networks are still surprisingly similar to his original ideas.In this talk I will question some of their most fundamental design aspects. As networks have become much deeper than had been possible or had ever been imagined in the 1950s, it is no longer clear that the layer by layer connectivity pattern is a well-suited architectural choice. In the first part of the talk I will show that randomly removing layers during training can speed up the training process, make it more robust, and ultimately lead to better generalization. We refer to this process as learning with stochastic depth – as the effective depth of the networks varies for each minibatch. In the second part of the talk I will propose an alternative connectivity pattern, Dense Connectivity, which is inspired by the insights obtained from stochastic depth. Dense connectivity leads to substantial reductions in parameter sizes, faster convergence, and further improvement in generalization. Finally, I will investigate the question why deep neural networks are so well suited for natural images and provide evidence that they may linearize the underlying sub-manifold into a Euclidean feature space.

Speaker Biography. Kilian Weinberger is an Associate Professor in the Department of Computer Science at Cornell University. He received his Ph.D. from the University of Pennsylvania in Machine Learning under the supervision of Lawrence Saul and his undergraduate degree in Mathematics and Computing from the University of Oxford. During his career he has won several best paper awards at ICML (2004), CVPR (2004, 2017), AISTATS (2005) and KDD (2014, runner-up award). In 2011 he was awarded the Outstanding AAAI Senior Program Chair Award and in 2012 he received an NSF CAREER award. He was elected co-Program Chair for ICML 2016 and for AAAI 2018. In 2016 he was the recipient of the Daniel M Lazar '29 Excellence in Teaching Award. Kilian Weinberger's research focuses on Machine Learning and its applications. In particular, he focuses on learning under resource constraints, metric learning, Gaussian Processes, computer vision and deep learning. Before joining Cornell University, he was an Associate Professor at Washington University in St. Louis and before that he worked as a research scientist at Yahoo! Research in Santa Clara.

Wednesday 12:00–13:00. Steve Wright

Practical Conditional Gradient Algorithms

University of Wisconsin-Madison

swright@cs.wisc.edu
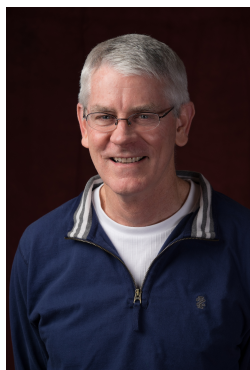
Interest in the conditional gradient algorithm, proposed by Frank and Wolfe in 1956, has revived in recent years because of its relevance to many data science paradigms. The basic algorithm for optimizing a smooth convex function over a closed convex compact set is appealing for its simplicity and its elementary convergence theory. However, the convergence is too slow for many applications. This talk describes enhancements of conditional gradient that improve its performance in several ways. A common feature of these enhancements is the maintenance of a "basis" of extreme points of the feasible set, being the solutions of the linear oracle from a subset of previous calls. We describe a "lazy" variant in which the linearized objective is minimized only over the convex hull of this basis on most iterations, a variant in which the objective is periodically reoptimized over the convex hull of this basis, and a sparsified variant in which this basis is sometimes reduced in size. (These techniques can be applied in tandem.) We show how convergence rates are affected by these techniques, and present computational results on a variety of applications.

This talk describes joint work with Sebastian Pokutta, Gábor Braun, Dan Tu, Nikhil Rao, and Parikshit Shah.

Speaker Biography. Stephen J. Wright holds the George B. Dantzig Professorship, the Sheldon Lubar Chair, and the Amar and Balinder Sohi Professorship of Computer Sciences at the University of Wisconsin-Madison. His research is in computational optimization and its applications to many areas of science and engineering, including data science. Prior to joining UW-Madison in 2001, Wright held positions at North Carolina State University (1986-90), Argonne National Laboratory (1990-2001), and the University of Chicago (2000-2001). He holds a Ph.D. in Mathematics from the University of Queensland, Australia. He has served as Chair of the Mathematical Optimization Society and as a Trustee of SIAM. He is a Fellow of SIAM. In 2014, he won the W.R.G. Baker Award from IEEE. Wright is the author / coauthor of widely used text and reference books in optimization including "Primal Dual Interior-Point Methods" and "Numerical Optimization". He has published widely on optimization theory, algorithms, software, and applications. Wright is current editor-in-chief of the SIAM Journal on Optimization and previously served as associate editor or editor-in-chief of Mathematical Programming (Series A), Mathematical Programming (Series B), SIAM Review, SIAM Journal on Scientific Computing, and several other journals and book series.

Thursday 09:00–10:00. Andrea Lodi

On big data, optimization and learning

Polytechnique Montreal, Canada

andrea.lodi@polymtl.ca

In this talk I review a couple of applications on Big Data that I personally like and I try to explain my point of view as a Mathematical Optimizer – especially concerned with discrete (integer) decisions – on the subject. I advocate a tight integration of Machine Learning and Mathematical Optimization (among others) to deal with the challenges of decision-making in Data Science. For such an integration I try to answer three questions: 1) what can optimization do for machine learning? 2) what can machine learning do for optimization? 3) which new applications can be solved by the combination of machine learning and optimization?

Speaker Biography. Andrea Lodi received the PhD in System Engineering from the University of Bologna in 2000 and he has been Herman Goldstine Fellow at the IBM Mathematical Sciences Department, NY in 2005-2006. He has been full professor of Operations Research at DEI, University of Bologna between 2007 and 2015. Since 2015 is Canada Excellence Research Chair in "Data Science for Real-time Decision Making" at the École Polytechnique de Montréal. His main research interests are in Mixed-Integer Linear and Nonlinear Programming and Data Science and his work has received several recognitions including the IBM and Google faculty awards. He is author of more than 100 publications in the top journals of the field of Mathematical Optimization and Data Science. He serves as Editor for several prestigious journals in the area. He has been network coordinator and principal investigator of two large EU projects/networks, and, since 2006, consultant of the IBM CPLEX research and development team. Finally, Andrea Lodi is the co-principal investigator of the project "Data Serving Canadians: Deep Learning and Optimization for the Knowledge Revolution", recently generously funded by the Canadian Federal Government under the Apogée Programme and scientific co-director of IVADO, the Montréal Institute for Data Valorization.

Thursday 12:00–13:00. Dan Bienstock

Variability-aware power operations

Columbia University

dano@columbia.edu

Power systems subject to uncertainty increasingly pose a challenge to safe and economical operation. A popular approach to dealing with uncertainty has been to rephrase standard problems so as to incorporate safety constraints, which are frequently stated, for example, as chance constraints, or modeled through scenario generation. In other words one obtains an optimization problem with constraints that guarantee high probability of safety. However, it has also been recognized that variability itself is a hazard, and that operating solutions that are nominally 'safe' may (undesirably) concentrate variability on inappropriate system components (for example, transformers and batteries). This point of view leads to the formulation of optimization problems that blend performance, safety and variability control, and we will discuss several examples. A challenge that is brought about by this approach is how to go about modeling variability. Ideally, we would refer to some type of real-time variance approach. Here we will discuss attributes of actual sensor data obtained through an industrial collaboration, in particular focusing on data features that do not readily fit theoretical models. If time permits we will outline a different application, dealing with counteracting data and physical attacks against transmission systems. Joint work with Apurv Shukla, Mauro Escobar and Michael Chertkov.

Speaker Biography.   Daniel Bienstock is Liu Family Professor at the IEOR Department at Columbia University, with joint appointments in the departments of Applied Math and Applied Physics, and Electrical Engineering. His current work focuses on polynomial optimization problems and on optimization and modeling of power systems. He is an Informs Fellow, was plenary speaker at the 2005 SIAM Optimization Conference and a semi-plenary speaker at the 2006 ISMP, received an IBM Faculty Partnership Award and an NSF Presidential Young Investigator Award, and is the author of two books.

Friday 09:00–10:00. David Morton

Optimizing Prioritized and Nested Solutions

Northwestern University

david.morton@northwestern.edu

A typical optimization model in operations research allocates limited resources among competing activities to derive an optimal portfolio of activities. In contrast, practitioners often form a rank-ordered list of activities, and select those with highest priority, at least when choosing an activity is a yes-no decision. Ranking schemes that score activities individually are known to be inferior. So, we describe a class of two-stage stochastic integer programs that accounts for structural and stochastic dependencies across activities and constructs an optimized priority list. We further discuss a class of optimization models, subject to a single "budget" constraint, that naturally leads to a family of optimal nested solutions at certain budget increments. Several applications both motivate the approach and illustrate results, ranging from a stochastic facility location model to a hierarchical graph clustering problem.

Speaker Biography.   Dave Morton is the David A. and Karen Richards Sachs Professor and Department Chair of Industrial Engineering and Management Sciences at Northwestern University. His research interests include stochastic and large-scale optimization with applications in security, public health, and energy systems. Prior to joining Northwestern, he was on the faculty at The University of Texas at Austin, worked as a Fulbright Research Scholar at Charles University in Prague, and was a National Research Council Postdoctoral Fellow in the Operations Research Department at the Naval Postgraduate School. He currently directs Northwestern's Center for Optimization and Statistical Learning http://osl.northwestern.edu.

Friday 12:00–13:00. Marija Ilic

New Energy Space Modeling and Implications on Complexity of Decision Making and Control in Electric Energy Systems

MIT

ilic@mit.edu

In this talk we present a recently-introduced multi-layered modeling framework for posing the problem of safe, robust and efficient design and control for rapidly changing electric energy systems. The proposed framework establishes dynamic relations between physical concepts such as stored energy, useful work, and wasted energy, on one hand; and modeling, simulation, and control of interactive modular complex dynamical systems, on the other. In particular, our recently introduced energy state-space modeling approach for electric energy systems is further interpreted using fundamental laws of physics in multi-physical systems, which are modeled as dynamically interacting modules. This approach is shown to be particularly well-suited for scalable optimization of large-scale complex systems. Instead of having to use simpler models, the proposed multi-layered modeling of system dynamics in energy space offers a promising basic method for modeling and controlling inter-dependencies across multi-physics subsystems for both ensuring feasible and near-optimal operation. It is illustrated how this approach can be used for understanding fundamental physical causes of inefficiencies created either at the component level or resulting from poor matching of their interactions. This talk is based on the recent paper by M. D. Ilic and R. Jaddivada entitled "Multi-layered interactive energy space modeling for near-optimal electrification of terrestrial, shipboard and aircraft systems", Annual Reviews in Control, available online May 2018. The paper provides theoretical foundations for Dynamic Monitoring and Decision Systems (DyMonDS) framework envisioned as the next-generation SCADA.

Speaker Biography.   Marija Ilic has retired as a Professor Emerita at Carnegie Mellon University. She is currently a Senior Staff in the Energy Systems Group 73 at the MIT Lincoln Laboratory, and a Visiting Professor at MIT Institute for Data, Systems and Society (IDSS). She is an IEEE Life Fellow. She was the first recipient of the NSF Presidential Young Investigator Award for Power Systems signed by late President Ronald Regan. In addition to her academic work, she is the founder of New Electricity Transmission Software Solutions, Inc. (NETSS, Inc.). She has co-authored several books on the subject of large-scale electric power systems, and has co-organized an annual multidisciplinary Electricity Industry conference series at Carnegie Mellon (http://www.ece.cmu.edu/ electriconf) with participants from academia, government, and industry. She was the founder and co-director of the Electric Energy Systems Group at Carnegie Mellon University (http://www.eesg.ece.cmu.edu).

■ AIMMS/MOPTA Optimization Modeling Competition 2018

The 10th AIMMS/MOPTA Optimization Modeling Competition is a result of cooperation between AIMMS and the organizers of the MOPTA conference. Teams of three graduate students participated and solved a stochastic supply chain problem.

The teams had to form a mathematical model of the problem, implement it in AIMMS, solve it, create a graphical user interface, and write a 15-page report on the project. We are happy that 13 teams from 9 countries registered to the competition. The panel of judges (Martin Takáč, Tamás Terlaky and Boris Defourny from Lehigh University and Deanne Zhang from AIMMS) selected the following three teams for the final:

Team "Opti Mice", Universidad de los Andes, Colombia
Nicolás Cabrera Malik, Daniel Cifuentes, Sebastian Cardona
advised by Camilo Gomez.

Team "Sparkles", NCSU, USA
Srinivasan Balan, Rahman Khorramfar, Rakesh Pandian Thangaraju
advised by Michael Kay

Team "ZIB", Zuse Institute Berlin, Germany
Mats Olthoff, Stanley Schade
advised by Boris Grimm.

The finalist teams will each give 30-minute presentations (20 minutes for the talk + 10 minutes for questions) on their work on Wednesday, August 15th, starting at 10:15am. The winning team will be announced at the conference banquet on Wednesday evening.

We thank all the teams for their participation. We believe that it has been a very positive experience for all parties involved in the process.

■ Instructions

For Speakers

- We ask all speakers to be familiar with the time and the location of their stream and talk, as specified in the conference booklet.

- Speakers should arrive at the location of their stream and talk 10 minutes prior to the scheduled start time of the session.

- Upon arrival you will be met by the chair of the session. Please introduce yourself and, if applicable, provide the chair with a copy of your presentation to upload onto the seminar room computer.

- Speakers should adhere to the allotted time slot. Anyone going over this time will be asked to stop by the chair.

- To aid you with the timing of your presentation, the chair will signal when you have 5 minutes and then 1 minute remaining for your presentation.

For Chairs

- Please arrive at the appropriate seminar room 10 minutes before the start of the stream you will be chairing. You should familiarize yourself with the equipment and ensure there are no obvious problems which would prevent the stream from running to schedule.

- In the event of a problem you should immediately seek the help of a local conference organizer.

- Delegates presenting in the stream should also be present in the seminar room 10 minutes before the start of the stream. You should introduce yourself to the speakers. You should discuss is everyone will use the same computer with uploaded presentations or switch computers.

- Your main role will be to ensure that the stream runs to time. Please allow time between talks for the comfort of the audience and to allow for movement between streams, when applicable.

- If a speaker fails to show for their talk, advise the audience to attend a talk in an alternative seminar room. Please, do not move the next talk forward.

- Before each speaker presents, you should introduce them and remind them that you will signal when, first 5 minutes, and then 1 minute, remain.

- Should a speaker go overtime, you must politely but firmly stop their presentation and move on to the next presentation.

- After each talk, thank the speaker, encourage applause, and open the floor to questions.

■ OPTE Special Issue: MOPTA 2018 on Energy and Optimization

Submission Due Date: October 30, 2018

Guest Editors:

Daniel Bienstock, Professor, Industrial Engineering and Operations Research/Applied Physics and Applied Mathematics, Columbia University. ⟨dano@ieor.columbia.edu⟩.

Luis F. Zuluaga, Associate Professor, Industrial and Systems Engineering, Lehigh University. ⟨luis.zuluaga@lehigh.edu⟩.

Aim

This call aims at publishing research work that showcases the interactions between Optimization and Engineering to address important problems arising in the Energy sector. The call encourages submissions from speakers at the upcoming MOPTA 2018, **but is open to submissions from any interested authors.** The call is **open to all types of papers, including theoretical, applied, and algorithmic articles, as well as articles that combine these features.**

Theme

The rapid and imperative changes in the Energy sector have created a wide range of challenges that are being addressed by a combination of Engineering and Optimization techniques. For example, consider the use of convex relaxations to obtain near-optimal alternating current optimal power flow (ACOPF) solutions, or for gas distribution network problems; the use of multi-level optimization to address commodity pricing problems, integration of renewable energy sources (RES) into electricity markets, or to immunize a commodity network against potential attacks; the use of uncertainty optimization techniques to address challenges related to the use of RES, immunize commodity networks against catastrophic events, and obtain robust equilibrium commodity prices that take into account uncertainty in commodity production or demand; the development of improved mathematical formulations and optimization solution methods, to address large-scale Energy related problems; among many other interesting research directions that are being explored towards the goal of designing commodity networks that are "smart", interconnected, efficient, reliable, resilient, and environmentally responsible. Beyond the study of individual commodity networks, nowadays, very interesting research addresses the challenging problem of studying the interaction between different commodity networks such as electricity, gas, and water networks, as well as the interaction of these networks with communication and transportation networks.

Submission Procedure

Please submit to the **Optimization and Engineering (OPTE)** journal at http://www.springer.com/mathematics/journal/11081 and select special issue "SI: MOPTA 2018". All submissions must be original and may not be under review by another publication. Interested authors should consult the journal's "Instructions for Authors", at http://www.springer.com/mathematics/journal/11081. All submitted papers will be reviewed on a peer review basis.

All inquiries should be directed to the attention of:

Luis F. Zuluaga, Guest Editor
Optimization and Engineering (OPTE) journal
⟨luis.zuluaga@lehigh.edu⟩

■ Conference Sponsors

■ DIMACS/TRIPODS/MOPTA 2018 Committee
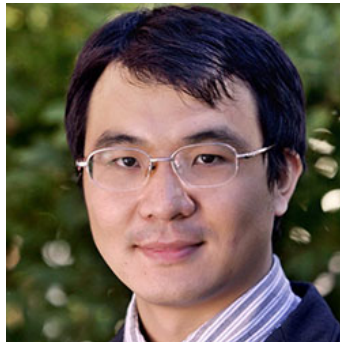
Katya Scheinberg, Lehigh
Conference Chair

Frank E. Curtis, Lehigh
SIAM Representative

Martin Takáč, Lehigh

Francesco Orabona, Boston U.

Han Liu, Northwestern U.

Satyen Kale, Google

Edo Liberty, Amazon

Stefanie Jegelka, MIT

Tamás Terlaky, Lehigh
MOPTA Founder

Yuri Dvorkin, NYU

Luis F. Zuluaga, Lehigh

Aberto Lamadrid, Lehigh

PROGRAM

DIMACS 2018	Monday, 13th August 2018	Tuesday, 14th August 2018	Wednesday, 15th August 2018
9:00 – 9:30	Sra	Singer	Weinberger
9:30 – 10:00		Jegelka	
10:00 – 10:15	Sridharan	Wright	Coffee Break
10:15 – 10:30			Robinson
10:30 – 10:45	Kale	Coffee Break	
10:45 – 11:00			Curtis
11:00 – 11:15	Coffee Break	Mobahi	
11:15 – 11:30			Haeffele
11:30 – 11:45	Abernethy	Toulis	
11:45 – 12:00			Coffee break
12:00 – 12:30	Hsu	Mokhtari	Wright
12:30 – 13:00	Lunch	Lunch	
13:00 – 13:30			Lunch
13:30 – 14:00	Bartlett	Duchi	
14:00 – 14:30			Cartis
14:30 – 15:00	Orabona	Lacoste-Julien	Schmidt
15:00 – 15:30	Andoni	Mahoney	Davis
15:30 – 16:00	Coffee Break	Coffee Break	Coffee Break
16:00 – 16:30	Ying	Ahmadi	Scheinberg
16:30 – 17:00	Belkin	Richtarik	Paquette
17:00 – 17:30	Langford	Poster Session and Cocktail Reception	Takac
17:30 – 18:00			
18:00 – 18:30			Cocktail Reception
18:30 – 19:00	Happy Hour		
19:00 – 20:00			Banquet
20:00 – 21:00			

MOPTA 2018	Program at a Glance	Wood Dining Room (WDR)	Governor's Suite	B131	B023
August 15, 2018 Wednesday					
08:00–09:00	Registration and Breakfast (at WDR)				
09:00 – 10:00	Plenary Talk	Kilian Weinberger			
10:00 – 10:15	Break (at WDR)				
10:15 – 11:45	Parallel Technical Sessions	Mopta Competition	DIMACS		
11:45 – 12:00	Coffee Break (at WDR)				
12:00 – 13:00	Plenary Talk	Steve Wright			
13:00 – 14:00	Lunch (at WDR)				
14:00 – 15:30	Parallel Technical Sessions	DIMACS	Novel Directions in Optimization	Metaheuristics and Their Applications	
15:30 – 16:00	Coffee Break (at WDR)				
16:00 – 17:30	Parallel Technical Sessions	DIMACS	Large Scale Mixed Integer Optimization	Modeling and Optimizing Energy Systems	
18:00 – 19:00	Cocktail Reception (Asa Packer Dining Room)				
19:00 – 21:00	Banquet (Asa Packer Dining Room)				
August 16, 2018 Thursday					
08:00 – 09:00	Registration and Breakfast (at WDR)				
09:00 – 10:00	Plenary Talk	Andrea Lodi			
10:00 – 10:15	Break (at WDR)				
10:15 – 11:45	Parallel Technical Sessions	New Directions in Optimization and Dynamical Systems	Stochastic Optimization	Iterative Methods	Applications in Healthcare
11:45 – 12:00	Coffee Break (at WDR)				
12:00 – 13:00	Plenary Talk	Dan Bienstock			
13:00 – 14:00	Lunch (at WDR)				
14:00 – 15:30	Parallel Technical Sessions	Optimization in Energy	Methods for Nonlinear Optimization	Recent Progress in Stochastic/Robust Optimization and Applications	Derivative Free and Black-Box Optimization
15:30 – 16:00	Coffee Break (at WDR)				
16:00 – 17:30	Parallel Technical Sessions	Learning and Energy	Nonlinear Optimization	Stochastic and Robust Optimization Algorithms and Applications	Reinforcement Learning for Supply Chain
19:00 – 21:00	Graduate Student Social - Packer House				
August 17, 2018 Friday					
08:00 – 09:00	Registration and Breakfast (at WDR)				
09:00 – 10:00	Plenary Talk	David Morton			
10:00 – 10:15	Break (at WDR)				
10:15 – 11:45	Parallel Technical Sessions	Resilience in Power Systems	Optimization and Learning	Conic Optimization and Integer Programming	Sparse Optimization
11:45 – 12:00	Coffee Break (at WDR)				
12:00 – 13:00	Plenary Talk	Marija Ilic			
13:00 – 14:00	Lunch (at WDR)				
14:00 – 15:30	Parallel Technical Sessions	Equilibrium and Complementarity Modeling in Energy Markets	Stochastic Gradient Decent and Conve Optimization	Optimization in Machine Learning	

08:00–08:45   **Registration and breakfast**

08:45–09:00   **Opening remarks**

09:00–10:00   **Plenary talk** (Wood Dining Room)   *Chair: Satyen Kale*

- **Tractable nonconvex optimization via geometry** *Suvrit Sra*

10:00–11:00   **DIMACS** (Wood Dining Room)   *Chair: Satyen Kale*

1. **Uniform Convergence of Gradients for Non-convex Learning and Optimization** *Karthik Sridharan*
2. **Logistic Regression: The Importance of Being Improper** *Satyen Kale*

11:00–11:30   **Coffee Break**

11:30–12:30   **DIMACS** (Wood Dining Room)   *Chair: Satyen Kale*

1. **Building Algorithms by Playing Games** *Jake Abernethy*
2. **Risk Bounds for Classification and Regression Models that Interpolate** *Daniel Hsu*

12:30–13:30   **Lunch**

13:30–14:30   **Plenary talk** (Wood Dining Room)   *Chair: Francesco Orabona*

- **Representation, optimization and generalization properties of deep neural networks** *Peter Bartlett*

14:30–15:30   **DIMACS** (Wood Dining Room)   *Chair: Francesco Orabona*

1. **Parameter-free Nonsmooth Convex Stochastic Optimization through Coin Betting** *Francesco Orabona*
2. **Data-dependent Hashing via Nonlinear Spectral Gaps** *Alexandr Andoni*

15:30–16:00   **Coffee Break**

16:00–17:30   **DIMACS** (Wood Dining Room)   *Chair: Francesco Orabona*

1. **Stochastic Optimization for AUC Maximization** *Yiming Ying*
2. **The Power of Interpolation: Machine Learning without Loss Functions and Regularization** *Mikhail Belkin*
3. **Contextual Reinforcement Learning** *John Langford*

18:30–20:00   **Happy Hour Social (Comfort Suites bar)**

08:00–09:00   **Registration and breakfast**

09:00–10:30   **DIMACS** (Wood Dining Room)   *Chair: Courtney Paquette*

1. **Maximizing Submodular Functions Exponentially Faster** *Yaron Singer*
2. **Robustness and Submodularity** *Stefanie Jegelka*
3. **Nonconvex Sparse Deconvolution: Geometry and Efficient Methods** *John Wright*

10:30–11:00   **Coffee Break**

11:00–12:30   **DIMACS** (Wood Dining Room)   *Chair: Courtney Paquette*

1. **Learning Over-Parameterized Models with Gradient Descent: An Average-Case Analysis over Quadratic Loss Functions** *Hossein Mobahi*
2. **Statistical Properties of Stochastic Gradient Descent** *Panos Toulis*
3. **Direct Runge-Kutta Discretization Achieves Acceleration** *Aryan Mokhtari*

12:30–13:30   **Lunch**

13:30–14:30   **Plenary talk** (Wood Dining Room)   *Chair: Martin Takac*

- **Better models in optimization** *John Duchi*

14:30–15:30   **DIMACS** (Wood Dining Room)   *Chair: Martin Takac*

1. **Frank-Wolfe Splitting via Augmented Lagrangian Method** *Simon Lacoste-Julien*
2. **Second Order Optimization and Non-convex Machine Learning** *Michael Mahoney*

15:30–16:00   **Coffee Break**

16:00–17:00   **DIMACS** (Wood Dining Room)   *Chair: Martin Takac*

1. **Optimization over Nonnegative Polynomials** *Amir Ali Ahmadi*
2. **Stochastic Quasi-gradient Methods: Variance Reduction via Jacobian Sketching** *Peter Richtarik*

17:00–19:00   **Posters** (Wood Dining Room)   *Chair: Katya Scheinberg*

1. **Feasible Level-set Methods for Optimization with Stochastic or Data-driven Constraints** *Qihang Lin*
2. **Bounding and Counting Linear Regions of Deep Neural Networks** *Thiago Serra*
3. **Level-set Methods for Finite-sum Constrained Convex Optimization** *Runchao Ma*
4. **A Machine Learning Approximation Algorithm for Fast Prediction of Solutions to Discrete Optimization Problems** *Sébastien Lachapelle*
5. **On the Convergence of Stochastic Gradient Descent with Adaptive Stepsizes** *Xiaoyu Li*
6. **Distributed First-order Algorithms with Gradient Tracking Converge to Second-order Stationary Solutions** *Amir Daneshmand*
7. **Estimation of Individualized Decision Rules Based on an Optimized Covariate-dependent Equivalent of Random Outcomes** *Zhengling Qi*
8. **A Stochastic Trust Region Algorithm Based on Careful Step Normalization** *Rui Shi*
9. **Revisiting the Foundations of Randomized Gossip Algorithms** *Nicolas Loizou*
10. **Underestimate Sequences via Quadratic Averaging** *Majid Jahani*
11. **A Unifying Scheme Of Primal-Dual Algorithms for Distributed Optimization** *Fatemeh Mansoori*
12. **Extrapolation of Finite Element Simulation with Graph Convolutional Lstm** *Yue Niu*
13. **Expected Risk and Auc Optimization without Dependence on the Data Set Size** *Minhan Li*
14. **Robust Learning of Trimmed Estimators via Manifold Sampling** *Matt Menickelly*
15. **Projective Splitting with Forward Steps: Asynchronous and Block-iterative Operator Splitting** *Patrick Johnstone*
16. **Inexact SARAH for Solving Stochastic Optimization Problems** *Lam Nguyen*
17. **Kernel Methods: Optimal Online Compressions, and Controlling Error Variance** *Alec Koppel*
18. **Towards Fast Computation of Certified Robustness for ReLU Networks** *Tsui-Wei (Lily) Weng*

08:00–09:00	**Registration and breakfast**	
09:00–10:00	**Plenary talk** (Wood Dining Room)	*Chair: Frank E. Curtis*

- **Deep Learning with Dense Connectivity** *Kilian Weinberger*

10:00–10:15	**Coffee Break**	
10:15–11:45	**Mopta Competition** (Wood Dining Room)	*Chair: Deanne Zhang*

1. **Team Opti Mice (Colombia)** *Daniel Cifuentes Daza*
2. **Team Sparkles (USA)** *Rakesh Pandian Thangaraju*
3. **Team ZIB (Germany)** *Mats Olthoff*

10:15–11:45	**DIMACS** (Governor's Suite)	*Chair: Frank E. Curtis*

1. **A Positive Outlook on Negative Curvature** *Daniel P. Robinson*
2. **How to Characterize Worst-case Performance of Algorithms for Nonconvex Optimization** *Frank Curtis*
3. **A Convex Lens for Non-convex Problems** *Benjamin Haeffele*

11:45–12:00	**Coffee Break**	
12:00–13:00	**Plenary talk** (Wood Dining Room)	*Chair: Katya Scheinberg*

- **Practical Conditional Gradient Algorithms** *Steve Wright*

13:00–14:00	**Lunch**	
14:00–15:30	**DIMACS** (Wood Dining Room)	*Chair: Katya Scheinberg*

1. **Dimensionality Reduction Techniques for Global Optimization** *Coralia Cartis*
2. **"Active-set Complexity" of Proximal Gradient: How Long Does It Take to Find the Sparsity Pattern?** *Mark Schmidt*
3. **Stochastic Methods for Non-smooth Non-convex Optimization** *Damek Davis*

14:00–15:30	**Novel Directions in Optimization** (Governor's Suite)	*Chair: Cemil Dibek*

1. **On the Complexity of Testing Attainment of the Optimal Value in Nonlinear Optimization** *Jeffrey Zhang*
2. **A Combinatorial Approach for Optimal Coloring of Perfect Graphs** *Cemil Dibek*
3. **Fast Fourier Linear Optimization** *Elahesadat Naghib*

14:00–15:30	**Metaheuristics and Their Applications** (B131)	*Chair: Yun Lu*

1. **How to Create a Elo-based Sports Ranking and Prediction Model** *Eric Landquist*
2. **A Population-based Metaheuristic Approach for Solving the Multi-demand Multidimensional Knapsack Problem** *Yun Lu*
3. **Analysis of Parameterless Metaheuristics as Applied to the MDMMKP** *Dylan Gaspar*

15:30–16:00	**Coffee Break**	
16:00–17:30	**DIMACS** (Wood Dining Room)	*Chair: Katya Scheinberg*

1. **New Framework For Convergence Analysis Of Stochastic Optimization Methods (Part 1)** *Katya Scheinberg*
2. **New Framework For Convergence Analysis Of Stochastic Optimization Methods (Part 2)** *Courtney Paquette*
3. **Do We Need 2nd Order Methods in Machine Learning?** *Martin Takac*

16:00–17:30	**Large Scale Mixed Integer Optimization** (Governor's Suite)	*Chair: Suresh Bolusani*

1. **Representabillity of Mixed-integer Bilevel Problem** *Sriram Sankaranarayanan*
2. **Red-Blue-Partitioned Network Optimization** *Matthew Johnson*
3. **Generalized Benders' Algorithm for Mixed Integer Bilevel Linear Optimization** *Suresh Bolusani*

16:00–17:30	**Modeling and Optimizing Energy Systems** (B131)	*Chair: Boris Defourny*

1. **Design of Time-delay Directed Dynamical Networks** *Shima Dezfulian*
2. **A Three-level Optimization Model for Fuel-supply Strategies of Natural Gas-fired Units** *Bining Zhao*
3. **On the Value of Dual-firing Power Generation Under Uncertain Gas Network Access** *Boris Defourny*

18:00–19:00	**Cocktail Reception (University Center, Asa Packer Campus)**	
19:00–21:00	**Banquet (University Center, Asa Packer Campus)**	

08:00–09:00   **Registration and breakfast**

09:00–10:00   **Plenary talk** (Wood Dining Room)   *Chair: Katya Scheinberg*

- **On big data, optimization and learning** *Andrea Lodi*

10:00–10:15   **Coffee Break**

10:15–11:45   **New Directions in Optimization and Dynamical Systems** (Wood Dining Room)   *Chair: Elahesadat Naghib*

1. **Sparsity Still Matters** *Robert Vanderbei*
2. **Optimization over Invariant Sets of Dynamical Systems** *Amir Ali Ahmadi*
3. **Motion Planning for Autonomous Vehicles Using MINLP** *Hande Benson*

10:15–11:45   **Stochastic Optimization** (Governor's Suite)   *Chair: Siqi Zhang*

1. **LNG Supply Chain Planning Under Demand and Supply Uncertainty : A Stochastic Programming Formulation and Solution Strategy.** *Armando Guarnaschelli*
2. **Distribution Systems Hardening against Natural Disasters** *Yushi Tan*
3. **Convergence Rate of Stochastic Mirror Descent for Non-smooth Non-convex Optimization** *Siqi Zhang*

10:15–11:45   **Iterative Methods** (B131)   *Chair: Mahmoud Rawashdeh*

1. **On Monotone Non-expansive Mapping and Their Approximation Fixed Point Results** *Buthinah Bin Dehaish*
2. **Tropical Optimization Problems: Recent Results and Applications Examples** *Nikolai Krivulin*
3. **Applying the Fractional Natural Decomposition Method to Solve Fractional Differential Equations in Multi-dimensional Space** *Mahmoud Rawashdeh*

10:15–11:45   **Applications in Healthcare** (B023)   *Chair: Secil Sozuer*

1. **Kinetic Parameter Identification Based on Spectroscopic Data - advancements Illustrated by Case Studies** *Christina Schenk*
2. **A Further Study on the Opioid Epidemic Dynamical Model with Random Perturbation** *Getachew Befekadu*
3. **Dynamic Appointment Scheduling Problem with Patient Preferences** *Secil Sozuer*

11:45–12:00   **Coffee Break**

12:00–13:00   **Plenary talk** (Wood Dining Room)   *Chair: Yuri Dvorkin*

- **Variability-aware power operations** *Dan Bienstock*

13:00–14:00   **Lunch**

14:00–15:30   **Optimization in Energy** (Wood Dining Room)   *Chair: Daniel Molzahn*

1. **Phase Transitions for Optimality Gaps in Optimal Power Flows** *Pascal Van Hentenryck*
2. **Statistical Learning for (Power System) Optimization: An Active Set Approach** *Line Roald*
3. **From Power System State Estimation to Low Rank Tensor Completion** *Cédric Josz*
4. **Optimal Power Flow with Robust Feasibility Guarantees** *Daniel Molzahn*

14:00–15:30   **Methods for Nonlinear Optimization** (Governor's Suite)   *Chair: Yuyang Rong*

1. **A Feasible Level-set Method for Optimization with Stochastic or Data-driven Constraints** *Qihang Lin*
2. **Level-set Methods for Finite-sum Constrained Convex Optimization** *Runchao Ma*
3. **An Inexact Penalty Sequential Linear Optimization Method for Constrained Nonlinear Optimization** *Yuyang Rong*

14:00–15:30   **Recent Progress in Stochastic/Robust Optimization and Applications** (B131)   *Chair: Guanglin Xu*

1. **A Copositive Approach for Multi-stage Robust Optimization Problems** *Grani A. Hanasusanto*
2. **Nurse Staffing under Uncertain Demand and Absenteeism** *Minseok Ryu*
3. **A Data-driven Distributionally Robust Optimization Approach for Appointment Scheduling With Random Service Durations and No-shows** *Guanglin Xu*

14:00–15:30   **Derivative Free and Black-Box Optimization** (B023)   *Chair: Albert Berahas*

1. **Efficient Optimal Design of Latent Energy Storage Systems** *Chunjian Pan*
2. **A New local Parallelization for Particle Swarm Optimization** *Abd AlRahman R. AlMomani; Ahmad Almomani*
3. **Derivative-free Optimization of Noisy Functions via Quasi-Newton Methods** *Albert Berahas*

15:30–16:00   **Coffee Break**

16:00–17:30   **Learning and Energy** (Wood Dining Room)   *Chair: Michael Chertkov*

1. **Data Recovery and Event Identification from Highly Quantized Measurements** *Meng Wang*

2. **Renewable Scenario Generation Using Adversarial Networks** *Baosen Zhang*

3. **Learning Power Flows with Support Vector Machines** *Ram Rajagopal*

4. **Learning and Control in Distribution Grids** *Michael Chertkov*

16:00–17:30   **Nonlinear Optimization** (Governor's Suite)   *Chair: Majid Jahani*

1. **Revisiting the Foundations of Randomized Gossip Algorithms** *Nicolas Loizou*

2. **Convergence Rates of Proximal Gradient Methods via the Convex Conjugate** *David Gutman*

3. **Efficient Distributed Hessian Free Algorithm for Large-scale Empirical Risk Minimization via Accumulating Sample Strategy** *Majid Jahani*

16:00–17:30   **Stochastic and Robust Optimization Algorithms and Applications** (B131)   *Chair: Fatma Kilinc-Karzan*

1. **SUNlayer: Stable Denoising with Generative Networks** *Soledad Villar*

2. **Stochastic ADMM Frameworks for Resolving Structured Stochastic Convex Programs** *Yue Xie*

3. **Exploiting Problem Structure in Optimization under Uncertainty via Online Convex Optimization** *Nam Ho-Nguyen*

16:00–17:30   **Reinforcement Learning for Supply Chain** (B023)   *Chair: Afshin OroojlooyJadid, MohamadReza Nazari*

1. **Concise Fuzzy Representation of Big Graphs: A Dimensionality Reduction Approach** *Faisal Abu-Khzam*

2. **RL for Inventory Optimization: Case on Beer Game** *Afshin Oroojlooy*

3. **Reinforcement Learning for Solving the Vehicle Routing Problem** *MohammadReza Nazari*

19:00–21:00   **Graduate Student Social - Packer House, 217 West Packer Avenue, Bethlehem**

08:00–09:00   **Registration and breakfast**

09:00–10:00   **Plenary talk** (Wood Dining Room)   *Chair: Tamas Terlaky*

- **Optimizing Prioritized and Nested Solutions** *David Morton*

10:00–10:15   **Coffee Break**

10:15–11:45   **Resilience in Power Systems** (Wood Dining Room)   *Chair: Russel Bent*

1. **Communication-constrained Expansion Planning for Resilient Distribution Systems** *Pascal Van Hentenryck*
2. **TBA** *J.P. Watson*
3. **Probabilistic N-k Failure-identification for Power Systems** *Harsha Nagarajan*
4. **Designing Resilient Distribution Systems under Natural Disasters** *Ruiwei Jiang*

10:15–11:45   **Optimization and Learning** (Governor's Suite)   *Chair: Zhiyuan Huang*

1. **A Trust-Region Method for Minimizing Regularized Non-convex Loss Functions** *Dimitri Papadimitriou*
2. **Bregman-divergence for Legendre Exponential Families** *Hyenkyun Woo*
3. **Learning-based Robust Optimization: Procedures and Statistical Guarantees** *Zhiyuan Huang*

10:15–11:45   **Conic Optimization and Integer Programming** (B131)   *Chair: Ali Mohammad Nezhad*

1. **Determine the Maximum Permissible Perturbation Set of SDP Problem With Unknown Perturbations** *Tingting Tang*
2. **Optimal Cutting Planes from the Group Relaxations** *Amitabh Basu*
3. **Covering Problem via Nonlinear Semidefinite Programming** *Walter Gomez*

10:15–11:45   **Sparse Optimization** (B023)   *Chair: Quoc Tran-Dinh*

1. **L0-regularized Sparsity for Probabilistic Mixture Models** *Dzung Phan*
2. **The CCP Selector: Scalable Algorithms for Sparse Ridge Regression from Chance-constrained Programming** *Weijun Xie*
3. **Accelerated Preconditioned Alternating Direction Methods of Multipliers with Non-ergodic Optimal Rates** *Quoc Tran-Dinh*

11:45–12:00   **Coffee Break**

12:00–13:00   **Plenary talk** (Wood Dining Room)   *Chair: Alberto Lamadrid*

- **New Energy Space Modeling and Implications on Complexity of Decision Making and Control in Electric Energy Systems** *Marija Ilic*

13:00–14:00   **Lunch**

14:00–15:30   **Equilibrium and Complementarity Modeling in Energy Markets** (Wood Dining Room)   *Chair: Jalal Kazempour*

1. **Efficiency and Welfare Distribution Effects From the Norwegian-Swedish Tradable Green Certificate Market** *Asgeir Tomasgard*
2. **A Column-and-constraint Decomposition Approach for Solving EPECs** *David Pozo*
3. **Market Integration of HVDC** *Spyros Chatzivasileiadis*
4. **Bi-level Network Planning with Generation-market Equilibria Subject to Transmission Costs Recovery** *Pengcheng Ding*

14:00–15:30   **Stochastic Gradient Decent and Conve Optimization** (Governor's Suite)   *Chair: Phuong Ha Nguyen*

1. **Reliable Machine Learning Using Unreliable Components: Error-runtime Trade-offs in Distributed SGD** *Sanghamitra Dutta*
2. **Complexity Bounds for Structured Convex Optimization** *Yuyuan Ouyang*
3. **Optimal Diminishing Stepsizes in SGD for Strongly Convex Objective Functions** *Phuong Ha Nguyen*

14:00–15:30   **Optimization in Machine Learning** (B131)   *Chair: Mohammad Pirhooshyaran*

1. **A Machine Learning Technique for Quadcopter State Estimation** *Arash Amini*
2. **Hyperparameter Tuning of Neural Networks(NNs) via Derivative Free Optimization(DFO)** *Mertcan Yetkin*
3. **Bidirectional LSTM Ensemble Structures for Multi-step Forecasting Of Ocean Wave Elevation** *Mohammad Pirhooshyaran*

ABSTRACTS

■ DIMACS

Room: **Wood Dining Room** (10:00 - 11:00)   Chair: _Satyen Kale_

1. Uniform Convergence of Gradients for Non-convex Learning and Optimization

Karthik Sridharan[1,*]

[1]Cornell University; *<sridharan@cs.cornell.edu>;

We introduce vector-valued Rademacher complexities as a user-friendly tool to bound the rate at which refined properties of the empirical risk such as gradients and Hessians converge to their population counterparts in non-convex settings. Our tools are simple, composable, and allow one to derive dimension-free uniform convergence bounds for gradients and hessians in a diverse range of non-convex learning problems under a robust set of assumptions. As an application of our techniques, we give a new analysis of batch gradient descent methods for non-convex generalized linear models and non-convex robust regression models, showing how to use any algorithm that finds approximate stationary points to obtain optimal sample complexity both in high (possibly infinite)- and low-dimensional regimes. This analysis applies under weaker distributional assumptions than in past works and applies even when multiple passes over the dataset are allowed.

Moving beyond smooth models we show - in contrast to the smooth case - even for simple models such as a single ReLU it is not possible to obtain dimension-independent convergence rates for gradients in the worst case. On the positive side, we show that it is still possible to obtain dimension-independent rates for this and other non-smooth models under a new type of distributional assumption.

2. Logistic Regression: The Importance of Being Improper

Satyen Kale[1,*]

[1]Google Research, New York; *<satyen@satyenkale.com>;

Learning linear predictors with the logistic loss - both in stochastic and online settings - is a fundamental task in machine learning and statistics, with direct connections to classification and boosting. Existing "fast rates" for this setting exhibit exponential dependence on the predictor norm, and Hazan et al. (2014) showed that this is unfortunately unimprovable. Starting with the simple observation that the logistic loss is 1-mixable, we design a new efficient improper learning algorithm for online logistic regression that circumvents the aforementioned lower bound with a regret bound exhibiting a doubly-exponential improvement in dependence on the predictor norm. This provides a positive resolution to a variant of the COLT 2012 open problem of McMahan and Streeter (2012) when improper learning is allowed. This improvement is obtained both in the online setting and, with some extra work, in the batch statistical setting with high probability. We also show that the improved dependence on predictor norm is near-optimal.

Leveraging this improved dependency on the predictor norm yields the following applications: (a) we give algorithms for online bandit multiclass learning with the logistic loss with an $\tilde{O}(\sqrt{n})$ relative mistake bound across essentially all parameter ranges, thus providing a solution to the COLT 2009 open problem of Abernethy and Rakhlin (2009), and (b) we give an adaptive algorithm for online multiclass boosting with optimal sample complexity, thus partially resolving an open problem of Beygelzimer et al. (2015) and Jung et al. (2017). Finally, we give information-theoretic bounds on the optimal rates for improper logistic regression with general function classes, thereby characterizing the extent to which our improvement for linear classes extends to other parameteric and even nonparametric settings.

■ DIMACS

Room: **Wood Dining Room** (11:30 - 12:30)   Chair: _Satyen Kale_

1. Building Algorithms by Playing Games

Jake Abernethy[1,*]

[1]Georgia Tech; *<prof@gatech.edu>;

A very popular trick for solving certain types of optimization problems is this: write your objective as the solution of a two-player zero-sum game, endow both players with an appropriate learning algorithm, watch how the opponents compete, and extract an (approximate) solution from the actions/decisions taken by the players throughout the process. This approach is very generic and provides a natural template to produce new and interesting algorithms. I will describe this framework and show how it applies in several scenarios, and describe recent work that draws a connection to the Frank-Wolfe algorithm and Nesterov's Accelerated Gradient Descent.

2. Risk Bounds for Classification and Regression Models that Interpolate

Daniel Hsu[1,*], _Misha Belkin_[2], _Partha Mitra_[3]

[1]Columbia University; *<djhsu@cs.columbia.edu>;[2]Ohio State University;[3]Cold Spring Harbor Laboratory;

Recent experiments with non-linear machine learning methods demonstrate the generalization ability of classification and regression models that interpolate noisy training data. It is difficult for existing generalization theory to explain these observations. On the other hand, there are classical examples of interpolating methods with non-trivial risk guarantees, the nearest neighbor rule being one of the most well-known and best-understood. I'll describe a few other such interpolating methods (old and new) with stronger risk guarantees compared to nearest neighbor in high dimensions.

■ DIMACS

Room: **Wood Dining Room** (14:30 - 15:30)   Chair: _F. Orabona_

1. Parameter-free Nonsmooth Convex Stochastic Optimization through Coin Betting

Francesco Orabona[1,*]

[1]Stony Brook University; *<francesco@orabona.com>;

Stochastic subgradient descent has become the method of choice for large-scale optimization of nonsmooth convex functions. However, in order to achieve the best theoretical and practical performance, it requires to tune its parameters: the stepsizes. These stepsizes are particularly critical in the unconstrained setting, where the distance between the initial point and the optimal solution can be arbitrary large. In this talk, I will show that stochastic optimization with Lipschitz convex losses can be reduced to a game of betting on a non-stochastic coin. Betting on a non-stochastic coin is a well-known problem that can be solved using tools from information theory. Moreover, optimal parameter-free coin betting algorithms are known, giving rise to novel parameter-free stochastic optimization algorithms. This approach is very general, i.e. it works for any norm, and it gives optimal rates in a number of settings, i.e. stochastic optimization in reproducing kernel Hilbert spaces, without any parameter/stepsize to tune. Empirical results will be shown as well.

2. Data-dependent Hashing via Nonlinear Spectral Gaps

Alexandr Andoni[1,], Assaf Naor, Aleksandar Nikolov, Ilya Razenshteyn, Erik Waingarten*

[1]Massachusetts Institute of Technology; [*]andoni@mit.edu;

We establish a generic reduction from nonlinear spectral gaps of metric spaces to space partitions, in the form of data-dependent Locality-Sensitive Hashing. This yields a new approach to the high-dimensional Approximate Near Neighbor Search problem (ANN). Using this reduction, we obtain a new ANN data structure under an arbitrary d-dimensional norm, where the query algorithm makes only a sublinear number of probes into the data structure.

Most importantly, the new data structure achieves a $O(\log d)$ approximation for an arbitrary norm. The only other such generic approach, via John's ellipsoid, would achieve square-root-d approximation only.

■ DIMACS

Room: **Wood Dining Room** (16:00 - 17:30)   Chair: *F. Orabona*

1. Stochastic Optimization for AUC Maximization

Yiming Ying[1,]*

[1]State University of New York, Albany; [*]yying@albany.edu;

Stochastic optimization algorithms such as stochastic gradient descent (SGD) update the model sequentially with cheap per-iteration costs, making them amenable for large-scale streaming data analysis. However, most of the existing studies focus on the classification accuracy which can not be directly applied to the important problems of maximizing the Area under the ROC curve (AUC) in imbalanced classification and bipartite ranking. In this talk, I will talk about our recent work on developing novel stochastic optimization algorithms for AUC maximization (aka bipartite ranking). Compared with the previous literature which requires high storage and per-iteration costs, our algorithms have both space and per-iteration costs of one datum and can achieve optimal convergence rates.

2. The Power of Interpolation: Machine Learning without Loss Functions and Regularization

Mikhail Belkin[1,]*

[1]Ohio State University; [*]mbelkin@cse.ohio-state.edu;

A striking feature of modern supervised machine learning is its consistent use of techniques that nearly interpolate the data. Deep networks often containing several orders of magnitude more parameters than data points, are trained to obtain near zero error on the training set. Yet, at odds with most theory, they show excellent test performance. It has become accepted wisdom that these properties are special to deep networks and require nonconvex analysis to understand.

In this talk I will show that classical (convex) kernel machines do, in fact, exhibit these unusual properties. Indeed, kernel machines explicitly constructed to interpolate the training data, show excellent test performance. Our empirical and theoretical results indicate that we are unlikely to make progress on understanding deep learning until we develop a fundamental understanding of classical "shallow" kernel classifiers in the "modern" near-interpolated setting.

Significantly, interpolating regimes lead to fast (exponential) convergence of SGD even with fixed step size, thus providing a cue toward explaining the efficiency of SGD in neural networks. Moreover, in the quadratic case we are able to derive explicit bounds on the step size and convergence rates in terms of the mini-batch size. These bounds are tight and are directly applicable to parameter selection, allowing us to construct very fast and accurate kernel machines, adaptive to both parallel computing resource (e.g., a GPU) and data. For example, training a kernel machine on full Imagenet dataset takes under an hour, on a single GPU. Smaller datasets, such as MNIST, take seconds. Finally, I will conclude by discussing the advantages of interpolation and arguing that these recent findings, as well as older observations on interpolation/overfitting in Adaboost and Random Forests suggest a need to revisit high-dimensional inference.

3. Contextual Reinforcement Learning

John Langford[1,]*

[1]Microsoft Research, New York; [*]jcl@microsoft.com;

The story of tabular reinforcement learning is nearly solved at a theoretical level, yet the algorithms coming from this process are typically useless in real-world settings. Real world settings often have a rich observation space, for example with an audio or video sensor. Treating these sensors as observations from a Markov Decision Process rapidly leads to statistical intractability. What's needed is a new model of the world. We've found a new model in a Contextual Decision Process which allows for a rich sensor space and yet still implies statistically tractable learning algorithms. This is the only model of reinforcement learning which allows generalization across any class of functions, exploration, and credit assignment.

■ DIMACS

Room: **Wood Dining Room** (09:00 - 10:30) Chair: *Courtney Paquette*

1. Maximizing Submodular Functions Exponentially Faster

Yaron Singer[1,], Eric Balkanski, Adam Breuer, Aviad Rubinstein*

[1]Harvard University; [*]yaron@seas.harvard.edu;

I'll describe a novel approach that yields algorithms whose parallel running time is exponentially faster than any previously known ones for submodular maximization. These algorithms reduce the number of parallel runtime from $\Omega(n)$ to $O(\log(n))$ while retaining optimal approximation guarantees. Time permitting I'll discuss parallelization for convex optimization. In contrast to the submodular case, we show information theoretic lower bounds indicating that parallelization cannot accelerate (non-smooth) convex optimization.

2. Robustness and Submodularity

Stefanie Jegelka[1,], Matthew Staib, Bryan Wilder*

[1]Massachusetts Institute of Technology;
[*]stefje@csail.mit.edu;

When critical decisions and predictions rely on observed data, robustness is an important consideration in learning and optimization. Robust formulations, however, can lead to more challenging, e.g., nonconvex, optimization problems. This talk will summarize some recent ideas at the intersection of robust optimization and submodular optimization. In particular, submodular optimization can help robust optimization, and vice versa: first, we show how ideas from discrete optimization lead to solving a nonconvex robust allocation or bidding problem; second, we develop algorithms for stochastic submodular optimization via robust submodular optimization. In both cases, the submodularity property offers the basis for a rich interplay of discrete and continuous optimization.

3. Nonconvex Sparse Deconvolution: Geometry and Efficient Methods

John Wright[1,], Yuqian Zhang, Yenson Lau, Han-Wen Kuo, Dar Gilboa, Sky Cheung, Abhay Pasupathy*

[1]Columbia University; [*]johnwright@ee.columbia.edu;

The problem of decomposing a given dataset as a superposition of basic motifs arises in a wide range of application areas, including neural spike sorting and the analysis of astrophysical and microscopy data. Motivated by these problems, we study a "short-and-sparse" deconvolution problem, in which the goal is to recover a short motif a from its convolution with a random spike train x. We formulate this problem as optimization over the sphere. We analyze the geometry of this (nonconvex) optimization problem, and argue that when the target spike train is sufficiently sparse, then on a region of the sphere, every local minimum is equivalent to the ground truth, up to symmetry (here a signed shift). This characterization obtains, e.g., for generic kernels of length k, when the sparsity rate of the spike train is proportional to $k^{-2/3}$ (i.e., roughly $k^{1/3}$ spikes in each length-k window). This geometric characterization implies that efficient methods obtain the ground truth under the same conditions.

Our analysis highlights the key roles of symmetry and negative curvature in the behavior of efficient methods – in particular, the role of a "dispersive" structure in promoting efficient convergence to global optimizers without the need to explicitly leverage second-order information. We sketch connections to broader families of "benign" nonconvex problems in data representation and imaging, in which efficient

methods obtain global optima independent of initialization. These problems include variants of sparse dictionary learning, tensor decomposition, and certain phase recovery problems.

■ DIMACS

Room: **Wood Dining Room** (11:00 - 12:30) Chair: *Courtney Paquette*

1. Learning Over-Parameterized Models with Gradient Descent: An Average-Case Analysis over Quadratic Loss Functions

Hossein Mobahi[1,], Misha Belkin[2], Partha Mitra[3]*

[1]Google; [*]hmobahi@csail.mit.edu;[2]The Ohio State University;[3]Cold Spring Harbor Laboratory;

With recent advancements in distributed computing infrastructure, the era of giga-dimensional optimization is now upon us. State-of-the-art models for high-dimensional problems are typically over-parameterized. Over-parameterization of a model often makes the associated loss function have small curvature in many directions. In such cases, the condition number of the Hessian matrix can be very large. Based on classic worst-case analysis of gradient descent, poor conditioning should have an adverse effect on the training time of a model. In practice, however, we observe precisely the opposite behavior, namely that training seems to proceed faster for such models. We suggest that a possible resolution may come from performing an average-case rather than a worst-case analysis. For concreteness and tractability, we focus the analysis on quadratic loss surfaces and establish an upper bound on the iteration complexity via average-case analysis. The result indicates that when most of the eigenvalues values of the Hessian are small, training becomes faster. This prediction is confirmed by our experiments on synthetic problems as well as preliminary experiments on deep networks applied to real data.

2. Statistical Properties of Stochastic Gradient Descent

Panos Toulis[1,]*

[1]University of Chicago; [*]panos.toulis@gmail.com;

Stochastic gradient descent (SGD) is remarkably multi-faceted: for machine learners it is a powerful optimization method, but for statisticians it is mainly a method for iterative estimation. While several important results are known for optimization properties of SGD, surprisingly little is known about its statistical properties. In this talk, I review recent results on doing statistics with SGD, which include analytic formulas for the asymptotic covariance matrix of SGD-based estimators and a numerically stable variant of SGD with implicit updates. Together these results open up the possibility of doing principled statistical analysis with SGD, including classical inference and hypothesis testing. Specifically about inference, I present current work showing that with appropriate selection of the learning rate the asymptotic covariance matrix of SGD is isotropic and parameter-free. As such, some SGD-based estimators can be easily transformed into pivotal quantities, which substantially simplifies inference. This is a unique and remarkable property of SGD, even compared to popular estimation methods favored by statisticians, such as maximum likelihood, highlighting the untapped potential of SGD for fast and principled estimation with large data sets.

3. Direct Runge-Kutta Discretization Achieves Acceleration

Aryan Mokhtari[1,*]

[1]Massachusetts Institute of Technology; [*]aryanm@mit.edu;

In this talk, we study gradient-based optimization methods obtained by directly discretizing a second-order ordinary differential equation (ODE) related to the continuous limit of Nesterov's accelerated gradient method. When the function is smooth enough, we show that acceleration can be achieved by a stable discretization of this ODE using standard Runge-Kutta integrators. Specifically, we prove that under Lipschitz-gradient, convexity and order-$(s+2)$ differentiability assumptions, the sequence of iterates generated by discretizing the proposed second-order ODE converges to the optimal solution at a rate of $O(N^{-2s/(s+1)})$, where s is the order of the Runge-Kutta numerical integrator. Furthermore, we introduce a new local flatness condition on the objective, under which rates even faster than $O(N^{-2})$ can be achieved with low-order integrators and only gradient information. Notably, this flatness condition is satisfied by several standard loss functions used in machine learning.

■ DIMACS

Room: **Wood Dining Room** (14:30 - 15:30)   Chair: *Martin Takac*

1. Frank-Wolfe Splitting via Augmented Lagrangian Method

Simon Lacoste-Julien[1,*]*, Gauthier Gidel, Fabian Pedregosa*

[1]University of Montreal; [*]slacoste@iro.umontreal.ca;

Minimizing a function over an intersection of convex sets is an important task in optimization that is often much more challenging than minimizing it over each individual constraint set. While traditional methods such as Frank-Wolfe (FW) or proximal gradient descent assume access to a linear or quadratic oracle on the intersection, splitting techniques take advantage of the structure of each sets, and only require access to the oracle on the individual constraints. In this work, we develop and analyze the Frank-Wolfe Augmented Lagrangian (FWAL) algorithm, a method for minimizing a smooth function over convex compact sets related by a "linear consistency" constraint that only requires access to a linear minimization oracle over the individual constraints. It is based on the Augmented Lagrangian Method (AL), also known as Method of Multipliers, but unlike most existing splitting methods, it only requires access to linear (instead of quadratic) minimization oracles. We use recent advances in the analysis of Frank-Wolfe and the alternating direction method of multipliers algorithms to prove a sublinear convergence rate for FWAL over general convex compact sets and a linear convergence rate over polytopes.

2. Second Order Optimization and Non-convex Machine Learning

Michael Mahoney[1,*]

[1]University of California, Berkeley; [*]mmahoney@stat.berkeley.edu;

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 to more robust second order 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, machine learning, and more traditional scientific computing applications, as well as non-convex variants of these basic methods. While this is of interest in many situations, it has received attention recently due to challenges associated with training so-called deep neural networks. We establish improved bounds for algorithms that incorporate sub-sampling as a way to improve computational efficiency, while maintaining the original convergence properties of these algorithms. These methods exploit recent results from Randomized Linear Algebra on approximate matrix multiplication. Within the context of second order optimization 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. Our results include extensions of the basic method to trust region and cubic regularization algorithms for non-convex optimization problems, interesting empirical observations on both convex and non-convex problems, as well as several non-obvious extensions.

■ DIMACS

Room: **Wood Dining Room** (16:00 - 17:00)   Chair: *Martin Takac*

1. Optimization over Nonnegative Polynomials

Amir Ali Ahmadi[1,*]*, Anirudha Majumdar*[2]*, Georgina Hall*[3]

[1]Princeton University;
[*]a_a_a@princeton.edu;[2]Princeton;[3]Princeton/INSEAD;

The problem of recognizing nonnegativity of a multivariate polynomial has a celebrated history, tracing back to Hilbert's 17th problem. In recent years, there has been much renewed interest in the topic because of a multitude of applications in applied and computational mathematics and the observation that one can optimize over an interesting subset of nonnegative polynomials using "sum of squares optimization".

In this talk, we first present two applications of nonnegative polynomials to problems that pop up in statistics and machine learning (shape-constrained regression and difference of convex programming). We then give an overview of our recent efforts to provide alternatives to sum of squares optimization that do not rely on semidefinite programming, but instead use linear programming, or second-order cone programming, or are altogether free of optimization. In particular, we present the first Positivstellensatz that certifies infeasibility of a set of polynomial inequalities simply by multiplying certain fixed polynomials together and checking nonnegativity of the coefficients of the resulting product. We also demonstrate the impact of our LP/SOCP-based algorithms on large-scale verification problems in control and robotics.

2. Stochastic Quasi-gradient Methods: Variance Reduction via Jacobian Sketching

Peter Richtarik[1,*]*, Robert M Gower, Francis Bach*
[1]The University of Edinburgh / KAUST;
[*]Peter.Richtarik@kaust.edu.sa;

We develop a new family of variance reduced stochastic gradient descent methods for minimizing the average of a very large number of smooth functions. Our method –JacSketch– is motivated by novel developments in randomized numerical linear algebra, and operates by maintaining a stochastic estimate of a Jacobian matrix composed of the gradients of individual functions. In each iteration, JacSketch efficiently updates the Jacobian matrix by first obtaining a random linear measurement of the true Jacobian through (cheap) sketching, and then projecting the previous estimate onto the solution space of a linear matrix equation whose solutions are consistent with the measurement. The Jacobian estimate is then used to compute a variance-reduced unbiased estimator of the gradient. Our strategy is analogous to the way quasi-Newton methods maintain an estimate of the Hessian, and

hence our method can be seen as a stochastic quasi-gradient method. We prove that for smooth and strongly convex functions, JacSketch converges linearly with a meaningful rate dictated by a single convergence theorem which applies to general sketches. We also provide a refined convergence theorem which applies to a smaller class of sketches. This enables us to obtain sharper complexity results for variants of JacSketch with importance sampling. By specializing our general approach to specific sketching strategies, JacSketch reduces to the stochastic average gradient (SAGA) method, and several of its existing and many new minibatch, reduced memory, and importance sampling variants. Our rate for SAGA with importance sampling is the current best-known rate for this method, resolving a conjecture by Schmidt et al (2015). The rates we obtain for minibatch SAGA are also superior to existing rates.

■ Posters

Room: **Wood Dining Room** (17:00 - 19:00) Chair: *Katya Scheinberg*

1. Feasible Level-set Methods for Optimization with Stochastic or Data-driven Constraints

Qihang Lin[1,∗], Selvaprabu (Selva) Nadarajah[2], Negar Soheili[2]

[1]University of Iowa; [∗]qihang-lin@uiowa.edu;[2]University of Illinois at Chicago;

We consider the constrained optimization where the objective function and the constraints are given as either finite sums or expectations. We propose a new feasible level-set method to solve this class of problems, which can produce a feasible solution path. To update a level parameter towards the optimality, our level-set method requires an oracle that generates upper and lower bounds as well as an affine-minorant of the level function. To construct the desired oracle, we reformulate the level function as the value of a saddle-point problem using the conjugate and perspective of constraints. Then a stochastic gradient method with a special Bregman divergence is proposed as the oracle for solving that saddle-point problem. The special divergence ensures the proximal mapping in each iteration can be solved in a closed form. The total complexity of both level-set methods using the proposed oracle are analyzed.

2. Bounding and Counting Linear Regions of Deep Neural Networks

Thiago Serra[1,∗], Christian Tjandraatmadja[2], Srikumar Ramalingam[3]

[1]Mitsubishi Electric Research Labs; [∗]tserra@gmail.com;[2]Carnegie Mellon University;[3]The University of Utah;

We investigate the complexity of deep neural networks (DNN) that represent piecewise linear (PWL) functions. In particular, we study the number of linear regions that a PWL function represented by a DNN can attain, both theoretically and empirically. We present (i) tighter upper and lower bounds for the maximum number of linear regions on rectifier networks, which are exact for inputs of dimension one; (ii) a first upper bound for multi-layer maxout networks; and (iii) a first method to perform exact enumeration or counting of the number of regions by modeling the DNN with a mixed-integer linear formulation. These bounds come from leveraging the dimension of the space defining each linear region. The results also indicate that a deep rectifier network can only have more linear regions than any shallow counterpart with same number of neurons if that number exceeds the dimension of the input.

3. Level-set Methods for Finite-sum Constrained Convex Optimization

Runchao Ma[1,∗], Qihang Lin[1], Tianbao Yang[1]

[1]University of Iowa; [∗]runchao-ma@uiowa.edu;

We consider the constrained optimization where the objective function and the constraints are defined as summation of finitely many loss functions. This model has applications in machine learning such as Neyman-Pearson classification. We consider two level-set methods to solve this class of problems, an existing inexact Newton method and a new feasible level-set method. To update the level parameter towards the optimality, both methods require an oracle that generates upper and lower bounds as well as an affine-minorant of the level function. To construct the desired oracle, we reformulate the level function as the value of a saddle-point problem using the conjugate and perspective of the loss functions. Then a stochastic variance-reduced gradient method with a special Bregman divergence is proposed as the oracle for solving that saddle-point problem. The special divergence ensures the proximal mapping in each iteration can be solved in a closed form. The total complexity of both level-set methods using the proposed oracle are analyzed.

4. A Machine Learning Approximation Algorithm for Fast Prediction of Solutions to Discrete Optimization Problems

Sébastien Lachapelle[1,∗], Eric Larsen[2], Sébastien Lachapelle[3], Yoshua Bengio[3], Emma Frejinger[2], Simon Lacoste-Julien[3], Andrea Lodi[4]

[1]Montreal Institute for Learning Algorithms (MILA); [∗]sebastien.lachapelle@umontreal.ca;[2]Department of Computer Science and Operations Research and CIRRELT, Université de Montréal;[3]Department of Computer Science and Operations Research and Montreal Institute for Learning Algorithms, Université de Montréal;[3]Department of Computer Science and Operations Research and Montreal Institute for Learning Algorithms, Université de Montréal;[4]Canada Excellence Research Chair, ÃL'cole Polytechnique de Montréal;

The paper presented provides a methodological contribution at the intersection of machine learning and operations research. Namely, we propose a methodology to predict descriptions of solutions to discrete stochastic optimization problems in very short computing time. We approximate the solutions based on supervised learning and the training dataset consists of a large number of deterministic problems that have been solved independently (and offline). Uncertainty regarding a subset of the inputs is addressed through sampling and aggregation methods.

Our motivating application concerns booking decisions of intermodal containers on double-stack trains. Under perfect information, this is the so-called load planning problem and it can be formulated by means of integer linear programming. However, the formulation cannot be used for the application at hand because of the restricted computational budget and unknown container weights.

The results show that standard deep learning algorithms allow to predict descriptions of solutions with high accuracy in very short time (milliseconds or less).

5. On the Convergence of Stochastic Gradient Descent with Adaptive Stepsizes

Xiaoyu Li[1,*]

[1]Stony Brook University; [*]xiaoyu.li@stonybrook.edu;

Stochastic gradient descent is the method of choice for large scale optimization of machine learning objective functions. Yet, its performance is greatly variable and heavily depends on the choice of the stepsizes. This has motivated a large body of research on adaptive stepsizes. However, there is currently a gap in our theoretical understanding of these methods, especially in the non-convex setting. In this work, we start closing this gap: we theoretically analyze the use of adaptive stepsizes, like the ones in AdaGrad, in the non-convex setting. We show sufficient conditions for almost sure convergence to a stationary point when the adaptive stepsizes are used, proving the first guarantee for AdaGrad in the non-convex setting. Moreover, we show explicit rates of convergence that automatically interpolates between $O(1/T)$ and $O(1/\sqrt{T})$ depending on the noise of the stochastic gradients, in both the convex and non-convex setting.

6. Distributed First-order Algorithms with Gradient Tracking Converge to Second-order Stationary Solutions

Amir Daneshmand[1,*], *Gesualdo Scutari*[1]

[1]Purdue University; [*]adaneshm@purdue.edu;

Several first-order optimization algorithms have been shown recently to converge to second-order stationary solutions of smooth nonconvex optimization problems, under mild conditions. There is no analogous guarantee for decentralized schemes solving nonconvex smooth multiagent problems over networks, modeled as directed static graphs. This work provides a positive answer to this open question. We prove that the family of decentralized algorithms employing distributed gradient tracking converges to second-order stationary solutions, under standard assumptions on the step-size.

7. Estimation of Individualized Decision Rules Based on an Optimized Covariate-dependent Equivalent of Random Outcomes

Zhengling Qi[1,*], *Ying Cui*[2], *Yufeng Liu*[1], *Jong-Shi Pang*[2]

[1]Department of Statistics and Operations Research, University of North Carolina; [*]qizl1027@live.unc.edu;[2]The Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California;

Recent exploration of optimal individualized decision rules (IDRs) for patients in precision medicine has attracted a lot of attention due to the heterogeneous responses of patients to different treatments. In the existing literature of precision medicine, an optimal IDR is defined as a decision function mapping from the patients' covariate space into the treatment space that maximizes the expected outcome of each individual. Motivated by the concept of Optimized Certainty Equivalent (OCE) introduced originally in the study of Ben-Tal and Teboulle (2007) that includes the popular conditional-value-of risk (CVaR) in the study of Rockafellar and Uryasev (2000), we propose a decision-rule based optimized covariates dependent equivalent (CDE) for individualized decision making problems. Our proposed IDR-CDE broadens the existing expected-mean outcome framework in precision medicine and enriches the previous concept of the OCE. Under a functional margin description of the decision rule modeled by an indicator function as in the literature of large-margin classifiers, we study the mathematical problem of estimating an optimal IDRs in two cases: in one case, an optimal solution can be obtained "explicitly" that involves the implicit evaluation of an OCE; the other case requires the numerical solution of an empirical minimization problem obtained by sampling the underlying distributions of the random variables involved. A major challenge of the latter optimization problem is that it involves a discontinuous objective function. We show that, under a mild condition at the population level of the model, the epigraphical formulation of this empirical optimization problem is a piecewise affine, thus difference-of-convex (dc), constrained dc, thus nonconvex, program. A simplified dc algorithm is employed to solve the resulting dc program whose convergence to a new kind of stationary solutions is established. Numerical experiments demonstrate that our overall approach outperforms existing methods in estimating optimal IDRs under heavy-tail distributions of the data. In addition to providing a risk-based approach for individualized medical treatments, which is new in the area of precision medicine, the main contributions of this work in general include: the broadening of the concept of the OCE, the epigraphical description of the empirical IDR-CDE minimization problem and its equivalent dc formulation, and the optimization of resulting piecewise affine constrained dc program.

8. A Stochastic Trust Region Algorithm Based on Careful Step Normalization

Rui Shi[1,*]

[1]Lehigh University; [*]rus415@lehigh.edu;

In this poster we present a new stochastic trust region method, deemed TRish, for solving stochastic and finite-sum minimization problems. We motivate our approach by illustrating how it can be derived from a trust region methodology. However, we also illustrate how a direct adaptation of a trust region methodology might fail to lead to general convergence guarantees. Hence, our approach involves a modified update scheme, which we prove possesses convergence guarantees that are similar to those for a traditional stochastic gradient (SG) method. We also present numerical results showing that TRish can outperform SG when solving convex and nonconvex machine learning test problems.

9. Revisiting the Foundations of Randomized Gossip Algorithms

Nicolas Loizou[1,*]

[1]University of Edinburgh; [*]n.loizou@sms.ed.ac.uk;

In this work we present a new framework for the analysis and design of randomized gossip algorithms for solving the average consensus problem. We present how randomized Kaczmarz-type methods - classical methods for solving linear systems - work as gossip algorithms when applied to a special system encoding the underlying network and explain their distributed nature. We reveal a hidden duality of randomized gossip algorithms, with the dual iterative process maintaining a set of numbers attached to the edges as opposed to nodes of the network. Subsequently, using the new framework we propose novel block and accelerated randomized gossip protocols.

10. Underestimate Sequences via Quadratic Averaging

Majid Jahani[1,*], *Chenxin Ma*[1], *Naga Venkata C. Gudapati*[1], *Rachael Tappenden*[2], *Martin Takac*[1]

[1]Lehigh University; [*]maj316@lehigh.edu;[2]University of Canterbury;

In this work we introduce the concept of an Underestimate Sequence (UES), which is a natural extension of Nesterov's estimate sequence. Our definition of a UES utilizes three sequences, one of which is a lower bound (or under-estimator) of the objective function. The question of how to construct an appropriate sequence of lower bounds is also addressed, and we present lower bounds for strongly convex smooth functions and for strongly convex composite functions, which

adhere to the UES framework. Further, we propose several first order methods for minimizing strongly convex functions in both the smooth and composite cases. The algorithms, based on efficiently updating lower bounds on the objective functions, have natural stopping conditions, which provides the user with a certificate of optimality. Convergence of all algorithms is guaranteed through the UES framework, and we show that all presented algorithms converge linearly, with the accelerated variants enjoying the optimal linear rate of convergence.

11. A Unifying Scheme Of Primal-Dual Algorithms for Distributed Optimization

Fatemeh Mansoori[1,*]

[1]PhD candidate;
[*]fatemehmansoori2019@u.northwestern.edu;

We study the problem of minimizing a sum of local convex objective functions over a network of processors/agents. Each agent in the network has access to a component of the objective function and can communicate with its neighbors. This problem can be formulated in a distributed framework by defining local copies of the decision variable for the agents. Each agent, then, works toward decreasing its local cost function, while keeping its variable equal to those of the neighbors. Many of the existing distributed algorithms with constant stepsize can only converge to a neighborhood of optimal solution. To circumvent this shortcoming, we propose to develop a class of distributed primal-dual algorithms based on augmented Lagrangian. The dual updates ensure exact convergence and the augmented quadratic term guarantees convergence. To improve convergence speed, we design algorithms with multiple primal updates per iteration. We can show that such algorithms converge to the optimal solution under appropriate constant stepsize choices. Simulation results confirm the superior of our algorithms to those with only one primal update at each iteration. The proposed class of algorithms and the convergence guarantees can be extended to the general form of linearly-constrained convex optimization problems.

12. Extrapolation of Finite Element Simulation with Graph Convolutional Lstm

Yue Niu[1,*]

[1]Lehigh University; [*]yun316@lehigh.edu;

Bioprosthetic heart valves are one of the best choices for heart valve replacement. However, the fact that their life span is generally limited to 15 years becomes a major concern of their clinical use. Several groups are studying the underlying mechanism of the degradation with different methods. Although numerical modeling has certain advantages over in vivo experiments, the computational cost for complete simulations is not yet acceptable. Here, we introduce our LSTM based neural network model to alleviate the resource consumption. It takes the simulation results of initial cycles and additional parameters that could easily be measured as guideline to predict the simulation result later. Due to the nature of finite element data structure, the graph convolution idea is also applied to smooth out the result and achieve smaller error. Results show that this model could keep track of the slow changes among the cycles. It is hoped that in this scheme, some slowly changing stages could be skipped, thus the simulation could be accelerated significantly.

13. Expected Risk and Auc Optimization without Dependence on the Data Set Size

Minhan Li[1,*]

[1]Lehigh University; [*]mil417@lehigh.edu;

We present a method for directly optimizing expected risk and expected AUC for linear classification proposed by Ghanbari and Scheinberg in 2017. This method constructs an optimization problem whose size is independent on the size of the data set. The empirical results show that it can produce competitive predictors in fraction of the time required by solving typical empirical risk minimization problem. However, for problems with large dimension this method is computationally costly. We propose further improvements to reduce the computational cost for large dimensional problems.

14. Robust Learning of Trimmed Estimators via Manifold Sampling

Matt Menickelly[1,*]

[1]Argonne National Laboratory; [*]mmenickelly@anl.gov;

We adapt a manifold sampling algorithm for nonsmooth, nonconvex formulations of learning while remaining robust to outliers in the training data. We demonstrate the approach on objectives based on trimmed loss. Empirical results show that the method has favorable scaling properties. Although savings in time come at the expense of not certifying global optimality, the algorithm consistently returns high-quality solutions on the trimmed linear regression and multi-class classification problems tested

15. Projective Splitting with Forward Steps: Asynchronous and Block-iterative Operator Splitting

Patrick Johnstone[1,*]

[1]Rutgers University; [*]patrick.r.johnstone@gmail.com;

This work is concerned with the classical problem of finding a zero of a sum of maximal monotone operators. For the projective splitting framework recently proposed by Combettes and Eckstein, we show how to replace the fundamental subproblem calculation using a backward step with one based on two forward steps. The resulting algorithms have the same kind of coordination procedure and can be implemented in the same block-iterative and potentially distributed and asynchronous manner, but may perform backward steps on some operators and forward steps on others. Prior algorithms in the projective splitting family have used only backward steps. Forward steps can be used for any Lipschitz-continuous operators provided the stepsize is bounded by the inverse of the Lipschitz constant. If the Lipschitz constant is unknown, a simple backtracking linesearch procedure may be used. Interestingly, this backtracking procedure also leads to a convergent algorithm even when the operator is only uniformly continuous, but not necessarily Lipschitz, provided it has full domain. For affine operators, the stepsize can be chosen adaptively without knowledge of the Lipschitz constant and without any additional forward steps. Convergence rates under various assumptions are also discussed along with preliminary experiments of several kinds of splitting algorithms on the lasso problem.

16. Inexact SARAH for Solving Stochastic Optimization Problems

Lam Nguyen[1,], Katya Scheinberg[1], Martin Takac[1]*

[1]Lehigh University; [*]lmn214@lehigh.edu;

We consider a general stochastic optimization problem which is often at the core of supervised learning, such as deep learning and linear classification. It is not able to apply SVRG and SARAH for this problem since these algorithms require an exact gradient information. We consider Inexact SARAH, which does not require to compute an exact gradient at each outer iteration; and analyze it in strongly convex, convex, and nonconvex cases. We also compare it with some existing stochastic algorithms.

17. Kernel Methods: Optimal Online Compressions, and Controlling Error Variance

Alec Koppel[1,]*

[1]U.S. Army Research Laboratory;
[*]akoppel@seas.upenn.edu;

In supervised learning, we learn a statistical model by minimizing some merit of fitness averaged over data. Doing so, however, ignores the model variance which quantifies the gap between the optimal within a hypothesized function class and the Bayes Risk. We propose to account for both the bias and variance by modifying training procedure to incorporate coherent risk which quantifies the uncertainty of a given decision. We develop the first online iterative solution to this problem when estimators belong to a reproducing kernel Hilbert space (RKHS), which we call Compositional Online Learning with Kernels (COLK). COLK addresses the fact that (i) minimizing risk functions requires handling objectives which are compositions of expected value problems by generalizing the two time-scale stochastic quasi-gradient method to RKHSs; and (ii) the RKHS-induced parameterization has complexity which is proportional to the iteration index which is mitigated through greedily constructed subspace projections. We establish almost sure convergence of COLK with attenuating step-sizes, and linear convergence in mean to a neighborhood with constant step-sizes, as well as the fact that its worst-case complexity is bounded. Experiments on data with heavy-tailed distributions illustrate that COLK exhibits robustness to outliers, consistent performance across training runs, and thus marks a step towards ameliorating the problem of over-fitting.

18. Towards Fast Computation of Certified Robustness for ReLU Networks

Tsui-Wei (Lily) Weng[1,]*

[1]MIT; [*]twweng@mit.edu;

Verifying the robustness property of a general Rectified Linear Unit (ReLU) network is an NP-complete problem. Although finding the exactminimum adversarial distortion is hard, giving a certified lower boundof the minimum distortion is possible. Current available methods of computing such a bound are either time-consuming or deliver low quality bounds that are too loose to be useful. In this work, we exploit the special structure of ReLU networks and provide two computationally efficient algorithms (Fast-Lin,Fast-Lip) that are able to certify non-trivial lower bounds of minimum adversarial distortions. Experiments show that (1) our methods deliver bounds close to (the gap is 2-3X) exact minimum distortions found by Reluplex in small networks while our algorithms are more than 10,000 times faster; (2) our methods deliver similar quality of bounds (the gap is within 35% and usually around 10%; sometimes our bounds are even better) for larger networks compared to the methods based on solving linear programming problems but our algorithms are 33-14,000 times faster; (3) our method is capable of solving large MNIST and CIFAR networks up to 7 layers with more than 10,000 neurons within tens of seconds on a single CPU core.

■ Mopta Competition

Room: **Wood Dining Room** (10:15 - 11:45)   Chair: *Deanne Zhang*

1. Team Opti Mice (Colombia)

Daniel Cifuentes Daza[1,*]

[1]Universidad de los Andes; [*];

2. Team Sparkles (USA)

Rakesh Pandian Thangaraju[1,*]

[1]NCSU; [*];

3. Team ZIB (Germany)

Mats Olthoff[1,*]

[1]Zuse Institute Berlin; [*];

■ DIMACS

Room: **Givernor's Suite** (10:15 - 11:45)   Chair: *Frank E. Curtis*

1. A Positive Outlook on Negative Curvature

Daniel P. Robinson[1,*]

[1]Johns Hopkins University; [*]daniel.p.robinson@jhu.edu;

The recent surge in interest in nonconvex models (e.g., in deep learning, subspace clustering, and dictionary learning) emphasizes a need for a fresh look at nonconvex optimization algorithms with provable convergence guarantees. A major factor in the design of such methods is the manner in which negative curvature is handled. In this talk, I present recent work that supports the following claims: (i) Commonly employed strategies for using negative curvature directions usually hurt algorithm performance; (ii) A new strategy based on upper-bounding models allows directions of negative curvature to be used while improving performance; and (iii) This strategy of using upper-bounding models is readily adapted for stochastic optimization, thus making it an attractive approach for large-scale "big data" problems. The talk also touches on worst-case complexity bounds and the pitfalls of attempting to associate such bounds with practical performance.

2. How to Characterize Worst-case Performance of Algorithms for Nonconvex Optimization

Frank Curtis[1,*]

[1]Lehigh University; [*]frank.e.curtis@lehigh.edu;

We posit that the manner in which worst-case complexity is analyzed for algorithms for solving nonconvex optimization problems leads to misleading characterizations of their performance. We propose a new strategy for analyzing complexity that attempts to more closely resemble actual practice. This is done by partitioning the search space into regions based on properties of the objective function and providing complexity bounds by region. (The partitioning is a theoretical tool only. The regions do not need to be known to an algorithm.) Our new characterization strategy offers new perspectives on the performance of well-known first- and second-order methods, and provides guidance for the design of new practical methods. For example, the strategy informs how the trust region radii should be chosen so that a second-order trust region method attains the same complexity as—and outperforms in practice—a cubic regularization approach.

3. A Convex Lens for Non-convex Problems

Benjamin Haeffele[1,*]

[1]Johns Hopkins University; [*]bhaeffele@jhu.edu;

A wide variety of non-convex problems can be characterized as the composition of a convex function with a convexity destroying transformation. Well known examples include many matrix/tensor factorization and neural network training formulations, where the loss is typically convex but convexity is destroyed by the matrix/tensor product or network mapping, respectively. This talk will describe a general framework that allows one to study a wide variety of non-convex optimization problems using tools from convex analysis. The analysis then provides sufficient conditions to guarantee when local minima are globally optimal as well as when no spurious local minima are present in the loss surface. Applications of the framework in matrix factorization, neural network training, separable-dictionary learning, and dropout regularization will be discussed.

■ DIMACS

Room: **Wood Dining Room** (14:00 - 15:30) Chair: *Katya Scheinberg*

1. Dimensionality Reduction Techniques for Global Optimization

Coralia Cartis[1,*], *Adilet Otemissov*[2]

[1]University of Oxford;
[*]Coralia.Cartis@maths.ox.ac.uk;[2]Turing Institute, London and Oxford University;

We show that the scalability challenges of Global Optimisation (GO) algorithms can be overcome for functions with low effective dimensionality, which are constant along certain linear subspaces. Such functions can often be found in applications, for example, in hyperparameter optimization for neural networks, heuristic algorithms for combinatorial optimization problems and complex engineering simulations. We propose the use of random subspace embeddings within a(ny) global minimisation algorithm, extending the approach in Wang et al (2013). We introduce two new frameworks, REGO (Random Embeddings for GO) and AREGO (Adaptive REGO), which transform the high-dimensional optimization problem into a low-dimensional one. In REGO, a new low-dimensional problem is formulated with bound constraints in the reduced space and solved with any GO solver. Using random matrix theory, we provide probabilistic bounds for the success of REGO, which indicate that this is dependent upon the dimension of the embedded subspace and the intrinsic dimension of the function, but independent of the ambient dimension. Numerical results show that high success rates can be achieved with only one embedding and that rates are independent of the ambient dimension of the problem. AREGO repeatedly solves a low-dimensional problem, each time with a different random subspace that is chosen using past information. Using results from conic integral geometry, we derive probabilistic bounds on the success of the reduced problem and show that AREGO is globally convergent with probability one for any Lipschitz function. In our numerical tests, we investigate the numerical efficiency of this adaptive approach, as well as its invariance to the ambient dimension of the problem. This work is joint with Adilet Otemissov (Turing Institute, London and Oxford University).

2. "Active-set Complexity" of Proximal Gradient: How Long Does It Take to Find the Sparsity Pattern?

Mark Schmidt[1,*]

[1]University of British Columbia; *schmidtm@cs.ubc.ca;

Proximal gradient methods have been found to be highly effective for solving minimization problems with non-negative constraints or ℓ_1-regularization. Under suitable nondegeneracy conditions, it is known that these algorithms identify the optimal sparsity pattern for these types of problems in a finite number of iterations. However, it is not known how many iterations this may take. We introduce the notion of the "active-set complexity", which in these cases is the number of iterations before an algorithm is guaranteed to have identified the final sparsity pattern. We further give a bound on the active-set complexity of proximal gradient methods in the common case of minimizing the sum of a strongly-convex smooth function and a separable convex non-smooth function.

3. Stochastic Methods for Non-smooth Nonconvex Optimization

Damek Davis[1,*]

[1]Cornell University; *dsd95@cornell.edu ;

We prove that the proximal stochastic subgradient method, applied to a weakly convex problem (i.e. difference of convex function and a quadratic), drives the gradient of the Moreau envelope to zero at the rate $O(k^{1/4})$. This class of problems captures a variety of nonsmooth nonconvex formulations, now widespread in data science. As a consequence, we obtain the long-sought convergence rate of the standard projected stochastic gradient method for minimizing a smooth nonconvex function on a closed convex set. In the talk, I will also highlight other stochastic methods for which we can establish similar guarantees.

■ Novel Directions in Optimization

Room: **Givernor's Suite** (14:00 - 15:30)   Chair: *Cemil Dibek*

1. On the Complexity of Testing Attainment of the Optimal Value in Nonlinear Optimization

Jeffrey Zhang[1,*], *Amir Ali Ahmadi*[1]

[1]Princeton University; *jeffz@princeton.edu;

We prove that unless P=NP, there exists no polynomial time (or even pseudo-polynomial time) algorithm that can test whether the optimal value of a nonlinear optimization problem where the objective and constraints are given by low-degree polynomials is attained. If the degrees of these polynomials are fixed, our results along with previously-known "Frank-Wolfe type" theorems imply that exactly one of two cases can occur: either the optimal value is attained on every instance, or it is strongly NP-hard to distinguish attainment from non-attainment. We also show that testing for some well-known sufficient conditions for attainment of the optimal value, such as coercivity of the objective function and closedness and boundedness of the feasible set, is strongly NP-hard. As a byproduct, our proofs imply that testing the Archimedean property of a quadratic module is strongly NP-hard, a property that is of independent interest to the convergence of the Lasserre hierarchy. Finally, we give semidefinite programming (SDP)-based sufficient conditions for attainment of the optimal value, in particular a new characterization of coercive polynomials that lends itself to an SDP hierarchy.

2. A Combinatorial Approach for Optimal Coloring of Perfect Graphs

Cemil Dibek[1,*]

[1]Princeton University; *cdibek@princeton.edu;

A graph G is perfect if for every induced subgraph H, the chromatic number of H equals the clique number of H. It is known that the chromatic number and a corresponding optimal coloring of perfect graphs can be computed in polynomial time using semidefinite programming due to the algorithm of Grotschel, Lovasz and Schrijver. However, this algorithm uses the ellipsoid method to solve semidefinite programs (SDP's), and hence it is not purely combinatorial. It is a well-known open question to construct a "combinatorial" polynomial-time algorithm that yields an optimal coloring of a perfect graph. The main goal of this work is to take a step forward in that direction. The method we use to approximately solve the Lovasz SDP is based on a meta-algorithm called Multiplicative Weights Update Algorithm. We give the starting ideas of how one can interpret the steps of the algorithm in the graph-theoretical language. This is a work in progress.

3. Fast Fourier Linear Optimization

Elahesadat Naghib[1,*], *Robert Vanderbei*[1]

[1]Princeton University; *enaghib@princeton.edu;

Many important problems such as Sphere Packing can be formulated as infinite dimensional Linear programing with constraints on the Fourier transform of the variables. We exploit the theoretical structure of these problems to find efficient ways for computing the solutions.

■ Metaheuristics and Their Applications

Room: **B131** (14:00 - 15:30)   Chair: *Yun Lu*

1. How to Create a Elo-based Sports Ranking and Prediction Model

Eric Landquist[1,*]

[1]Kutztown University; *elandqui@kutztown.edu;

There are numerous algorithmic strategies for objectively ranking sports teams and predicting the outcomes of games or competitions. In the 1950s, Arpad Elo created a rating system for professional chess players that has been adapted to various sports and other competitions. In this talk, we discuss frameworks for modifying Elo's original rating system for application to ranking sports teams. The objective of these frameworks is to maximize the accuracy of the probability of victory implied by the resulting model. One desirable feature of these models is that they can be used to predict the margin of victory of games. One framework applies the Jaya optimization algorithm and does not take the margin of victory into consideration, while a second framework does. We apply the latter framework to college basketball in order to predict the outcome of the 2018 Men's and Women's NCAA Tournament. Although this model correctly picked Villanova to win the Men's Tournament, it did not foresee UMBC defeating UVA in the first round nor Notre Dame winning the Women's Tournament.

2. A Population-based Metaheuristic Approach for Solving the Multi-demand Multidimensional Knapsack Problem

Yun Lu[1,]*

[1]Kutztown University; [*]lu@kutztown.edu;

The Multi-Demand Multidimensional Knapsack Problem (MDMKP) is a combinatorial optimization problem with real-world applications that is extremely difficult to solve due to conflicting constraints. In this study, we adapt a population-based (a collection of solutions) metaheuristic to efficiently generate near-optimal solutions to the MDMKP. This metaheuristic, called Jaya (victory in Sanskrit) was introduced in 2016 by Rao to solve continuous nonlinear engineering design problems. Since the MDMKP is a binary optimization problem (variables are bit strings, not continuous variables), we made modifications to the Jaya metaheuristic in order to effectively solve the MDMKP. For test purposes, we use 810 large MDMKP instances available to researchers on the web. In this talk, we will report empirical results we obtained from solving these 810 MDMKP instances using our new Jaya-based metaheuristic approach. Our results will be compared to the optimal (if known) or best known results for these problems.

3. Analysis of Parameterless Metaheuristics as Applied to the MDMMKP

Dylan Gaspar[1,]*

[1]Kutztown University; [*]dgasp528@live.kutztown.edu;

The Multiple Demand Multidimensional Multiple-choice Knapsack Problem (MDMMKP), defined by Lamine et al. (2012), is an NP-hard optimization problem with the objective being to maximize the value of objects placed in a knapsack under three categories of constraints. The constraints are multiple dimensional constraints, multiple demand constraints, and the multiple-choice constraints. The choice constraints allow for one and only one object within a given class to be selected. In this paper, we will analyze how well parameterless metaheuristics solve the MDMMKP. Specifically, how well both the Jaya and the Teaching Learning Based Optimization (TLBO) algorithms, along with some variations of these metaheuristics solve the MDMMKP will be analyzed. These metaheuristic solutions will be compared with CPLEX results to demonstrate that these metaheuristics are quite capable of finding high-quality results for a variety of large sized problem instances without the need for fine tuning of parameters other than population size and termination criteria.

■ DIMACS

Room: **Wood Dining Room** (16:00 - 17:30) Chair: *Katya Scheinberg*

1. New Framework For Convergence Analysis Of Stochastic Optimization Methods (Part 1)

Katya Scheinberg[1,]*

[1]Lehigh University; [*]katyas@lehigh.edu;

We will present a very general framework for unconstrained stochastic optimization which encompasses standard frameworks such as line search and trust region using random models. In particular this framework retains the desirable practical features such step acceptance criterion, trust region adjustment and ability to utilize of second order models. The framework is based on bounding the expected stopping time of a stochastic process, which satisfies certain assumptions. Then the convergence rates are derived for each method by ensuring that the stochastic processes generated by the method satisfies these assumptions. The methods include a version of a stochastic

trust-region method and a stochastic line-search methods and provide strong convergence analysis under weaker conditions than alternative approaches in the literature.

2. New Framework For Convergence Analysis Of Stochastic Optimization Methods (Part 2)

Courtney Paquette[1,]*

[1]Lehigh University/University of Waterloo;
[*]cop318@lehigh.edu;

We will present a very general framework for unconstrained stochastic optimization which encompasses standard frameworks such as line search and trust region using random models. In particular this framework retains the desirable practical features such step acceptance criterion, trust region adjustment and ability to utilize of second order models. The framework is based on bounding the expected stopping time of a stochastic process, which satisfies certain assumptions. Then the convergence rates are derived for each method by ensuring that the stochastic processes generated by the method satisfies these assumptions. The methods include a version of a stochastic trust-region method and a stochastic line-search methods and provide strong convergence analysis under weaker conditions than alternative approaches in the literature.

3. Do We Need 2nd Order Methods in Machine Learning?

Martin Takac[1,], Albert Berahas[1], Jie Liu[1], Rachael Tappenden[2]*

[1]Lehigh University; [*]Takac.MT@gmail.com;[2]University of Canterbury;

In this talk, we address the question if and when do we need 2nd order optimization methods for training deep neural networks and when are the SGD type algorithms sufficient.
We will further discuss some challenges when using stochastic and batch Quasi-Newton methods for training DNN. We will conclude the talk with preliminary numerical experiments.

■ Large Scale Mixed Integer Optimization

Room: **Givernor's Suite** (16:00 - 17:30)   Chair: *Suresh Bolusani*

1. Representabillity of Mixed-integer Bilevel Problem

Sriram Sankaranarayanan[1,], Amitabh Basu[1], Christopher Ryan[2]*

[1]Johns Hopkins University; [*]ssankar5@jhu.edu;[2]Booth School of Business, University of Chicago;

We give a precise description of the sets representable by mixed-integer bilevel problems in its full generality as well as with restrictions on the integrality constraints. We show that while a general mixed-integer bilevel problem could potentially represent sets that are not closed, their closure is a finite union of sets representable by mixed-integer problems. On restricting the integrality constraints to upper level only, we always get closed sets that are finite unions of sets representable by mixed-integer problems. A continuous bilevel problem represents a finite union of polyhedra. All these containments are tight that given a finite union of polyhedra, it can be represented by a continuous bilevel problem and given a finite union of mixed-integer representable sets, they can represented using a mixed-integer

bilevel problems. In this we also show an equivalence between bilevel problem, polyhedral reverse convex problem and a complementarity problem.

2. Red-Blue-Partitioned Network Optimization

Matthew Johnson[1,*]

[1]CUNY; [*]mpjohnson@gmail.com;

Arkin et al. (2017) recently introduced *partitioned pairs* network optimization problems: Given n pairs of points in a metric space, the task is to color one point from each pair red and the other blue, and then to compute two separate *network structures* or disjoint subgraphs of a specified sort, one on the graph induced by the red points and the other on the blue points. Three structures have been investigated by Arkin et al. (2017) —*perfect matchings*, *spanning trees*, and *traveling salesperson tours*—and the three objectives to optimize for when computing such pairs of structures: *min-sum*, *min-max*, and *bottleneck*. We provide improved approximation guarantees and/or strengthened hardness results for these nine NP-hard problem settings.

3. Generalized Benders' Algorithm for Mixed Integer Bilevel Linear Optimization

Suresh Bolusani[1,*], *Ted Ralphs*[1]

[1]Department of Industrial and Systems Engineering, Lehigh University; [*]bsuresh@lehigh.edu;

We propose a generalized Benders' approach to solving general mixed integer bilevel linear optimization problems in which continuous and integer variables are present in both first- and second-level problems. Our algorithm generalizes the previously proposed algorithm for stochastic optimization problems with mixed integer recourse. As with that algorithm, we make minimal assumptions, requiring only that first-level variables participating in the second-level problem be integer variables. As usual, the strategy is to project out the second-level variables in order to obtain a reformulation involving only first-level variables. This reformulation involves the so-called risk function, which is then iteratively approximated. The result is that the master problem is a single-level optimization problem involving only first-level variables, whereas the subproblem is the optimization problem resulting from fixing the first-level variables. We solve this latter problem by reformulating it as a mixed integer linear optimization problem. Computational results using an open-source implementation will be presented.

■ Modeling and Optimizing Energy Systems

Room: **B131** (16:00 - 17:30) Chair: *Boris Defourny*

1. Design of Time-delay Directed Dynamical Networks

Shima Dezfulian[1,*], *Yaser Ghaedsharaf*[1], *Nader Motee*[1]

[1]Lehigh University; [*]dezfulian@lehigh.edu;

We study design of a class of noisy time-delay directed dynamical networks in order to enhance their performance in reaching average consensus. The network consist of finitely many agents which could communicate with their neighbors through communication links. Agents and their communication links are represented by the nodes and edges of a directed weighted graph, respectively. The value of the non-negative edge weights show the communication strength. Moreover,

because of the time-delay, edge weights should satisfy dynamic stability criterion. Besides, to achieve average consensus, balancedness of the graph should be preserved. We exploit the H2-norm of the network as our performance measure, which is a measure of robustness to the external noise. Thus, we intend to improve the H2-norm by finding the optimal edge weights. We can cast this problem as a constrained optimization problem such that the objective function is the H2-norm and the constraints are the non-negativity of edge weights, networks stability condition, and preserving the balancedness of the graph. We assume fixed graph topology and constant time-delay. One of the main challenges is the cost of computing both the objective function and its gradient which involve solving the delay Lyapunov equation. To solve the delay Lyapunov equation, we make use of iterative methods and we suggest a preconditioner to improve the speed of convergence and numerical stability of iterative methods. Without an iterative method, the time complexity of solving the delay Lyapunov is $O(n^6)$, where n is the number of agents in the network. However, using the preconditioner, we solve the delay Lyapunov equation in a few iterations where time complexity of each iteration is $O(n^3)$.

2. A Three-level Optimization Model for Fuel-supply Strategies of Natural Gas-fired Units

Bining Zhao[1,*], *Rick S. Blum*[2], *Alberto Lamadrid*[2]

[1]ECE, Lehigh University; [*]biz218@lehigh.edu;[2]Lehigh University;

In this work, we provide a practical tool for a system operator to decide the type and amount of fuel- supply contract between gas-fired units and gas suppliers to minimize the potential impact of cyber- attacks on the communications between the operator and the gas suppliers. In this context, we develop a trilevel min-max-min optimization model to represent actions of a power system defender, an attacker, and an operator. The three agents share a common objective function, which is the expected operational cost and the cost of unserved energy. The strong duality theorem is used to combine the middle and the lower level problems into a single level one. Then, the original problem is transferred into a mixed-integer bilevel model, which is then solved by a Benders primal decomposition method (also called as C&CG decomposition algorithm).

3. On the Value of Dual-firing Power Generation Under Uncertain Gas Network Access

Boris Defourny[1,*], *Shu Tu*[1]

[1]Lehigh University; [*]defourny@lehigh.edu;

This work is concerned with the impact of gas network disruptions on dual-firing power generation. We formulate the question as a stochastic optimization problem for a risk-neutral operator, and study the sensitivity of the value of the dual-firing generating unit to gas network availability parameters.

■ New Directions in Optimization and Dynamical Systems

Room: **Wood Dining Room** (10:15 - 11:45) Chair: *Elahesadat Naghib*

1. Sparsity Still Matters

Robert Vanderbei[1,]*

[1]Princeton University; [*]rvdb@princeton.edu;

In this talk, I will focus on one particular issue that permeates much of the world of optimization, namely, the importance of understanding that "modeling matters" and, in particular, that exploiting sparsity in a problem's representation can be extremely beneficial.

2. Optimization over Invariant Sets of Dynamical Systems

Amir Ali Ahmadi[1,]*

[1]Princeton University; [*]a_a_a@princeton.edu;

We study the problem of optimizing over invariant sets of dynamical systems, which we refer to as "robust-to-dynamics optimization" (RDO). An RDO problem is an optimization problem specified by two pieces of input: (i) a mathematical program (an objective function $f : R^n \to R$ and a feasible set $\Omega \subseteq R^n$), and (ii) a dynamical system (a map $g : R^n \to R^n$). Its goal is to minimize f over the set $S \subseteq \Omega$ of initial conditions that forever remain in Ω under g. The focus of this talk is on the case where the mathematical program is a linear program and the dynamical system is either a known linear map, or an uncertain linear map that can change over time. In both cases, we study a converging sequence of polyhedral outer approximations and (lifted) spectrahedral inner approximations to S. Our inner approximations are optimized with respect to the objective function f and their semidefinite characterizationâĂŤwhich has a semidefinite constraint of fixed sizeâĂŤis obtained by applying polar duality to convex sets that are invariant under (multiple) linear maps. We characterize three barriers that can stop convergence of the outer approximations to S from being finite. We prove that once these barriers are removed, our inner and outer approximating procedures find an optimal solution and a certificate of optimality for the RDO problem in a finite number of steps. Moreover, in the case where the dynamics are linear, we show that this phenomenon occurs in a number of steps that can be computed in time polynomial in the bit size of the input data. Our analysis also leads to a polynomial-time algorithm for RDO instances where the spectral radius of the linear map is bounded above by any constant less than one. Joint work with Oktay Gunluk.

3. Motion Planning for Autonomous Vehicles Using MINLP

Hande Benson[1,]*

[1]Drexel University; [*]hvb22@drexel.edu;

We will present a mixed-integer nonlinear programming model, its centralized and decentralized solution, for motion planning in fleets of autonomous vehicles (land, water, and aerial) under communication constraints. We will show that a customized approach is necessary for real-time solutions.

■ Stochastic Optimization

Room: **Givernor's Suite** (10:15 - 11:45)   Chair: *Siqi Zhang*

1. LNG Supply Chain Planning Under Demand and Supply Uncertainty : A Stochastic Programming Formulation and Solution Strategy.

Armando Guarnaschelli[1,], Enrique Salomone[2]*

[1]Pontificia Universidad Catolica de Valparaiso; [*]armando.guarnaschelli@pucv.cl;[2]INGAR - CONICET - UTN;

In this work, a stochastic optimization model is introduced to create Annual Delivery Programs (ADPs) for Liquefied Natural Gas (LNG) supply chains. Every year, liquefaction plants create ADPs to efficiently supply their customers under long-medium and short-term contracts. This activity has taken great relevance since the SPOT market boom, in which both suppliers and clients may trade during the Gas-year. Spot sourcing and sales take place when supply chain partners look for more profits or need to stabilize variabilities in demand and raw gas supply. Moreover, the high cost of transportation and limited availability of specialized ships to carry it out, call for creating optimal delivery programs that maximize the value functions of both liquefaction plants and their customers. In this context, this proposal relies on stochastic programming to provide an optimized and reliable strategy for ADP creation. The stochastic programming model proves to be hard to solve. To overcome this weakness, a scenario reduction technique based on clustering is provided together with a specialized rolling horizon heuristic. Under this setting, an optimized ADP can be obtained in acceptable computing times. The bi-objective nature of this problem is tackled using goal programming on every step of the solution procedure. A discussion on the results of applying this approach to a real-world scenario is included.

2. Distribution Systems Hardening against Natural Disasters

Yushi Tan[1,]*

[1]University of Washington; [*]ystan@uw.edu;

Distribution systems are often crippled by catastrophic damage caused by natural disasters. Well-designed hardening can significantly speed-up the post-disaster restoration process. This performance is quantified by a resilience measure associated with the operability trajectory. The distribution system hardening problem can be formulated as a two-stage stochastic problem, where the inner operational problem addresses the proper scheduling of post-disaster repairs and the outer problem the judicious selection of components to harden. We propose a deterministic single crew approximation with two solution methods, an MILP formulation and a heuristic approach. We provide computational evidence based on several IEEE test feeders, which demonstrates that the heuristic approach provides near-optimal hardening solutions in a computationall efficient manner.

3. Convergence Rate of Stochastic Mirror Descent for Non-smooth Non-convex Optimization

Siqi Zhang[1,], Niao He[1]*

[1]University of Illinois at Urbana-Champaign; [*]siqiz4@illinois.edu;

We study the non-asymptotic stationary convergence behavior of Stochastic Mirror Descent (SMD) for a general class of nonconvex nonsmooth stochastic optimization problems, in which the objective can be decomposed into a relatively weakly convex function (possibly non-Lipschitz) and a simple non-smooth convex regularizer. We

prove that SMD, without the use of mini-batch, is guaranteed to converge to a stationary point in a convergence rate of $\mathcal{O}(1/\sqrt{t})$. The efficiency estimate matches with existing results for stochastic subgradient method, but is evaluated under a stronger stationarity measure. This appears to be the first non-asymptotic convergence analysis of SMD for solving relatively weakly convex problems.

■ Iterative Methods

Room: **B131** (10:15 - 11:45)   Chair: *Mahmoud Rawashdeh*

1. On Monotone Non-expansive Mapping and Their Approximation Fixed Point Results

Buthinah Bin Dehaish[1,*]

[1]kau; [*]bbindehaish@yahoo.com;

Suppose that C is a nonempty closed bounded and convex subset of a metric space X. Let T be a monotone nonexpansive mapping on C. During this talk we will present some existence fixed point result of this mapping. Furthermore, we will describe the behavior of its fixed point by using some constructive iteration.

2. Tropical Optimization Problems: Recent Results and Applications Examples

Nikolai Krivulin[1,*]

[1]Saint Petersburg State University; [*]nkk@math.spbu.ru;

We consider multidimensional optimization problems formulated in the tropical mathematics setting to minimize or maximize functions defined on vectors over idempotent semifields, subject to linear equality and inequality constraints. We start with a brief overview of known tropical optimization problems and solution approaches. Furthermore, some new problems are presented with nonlinear objective functions calculated using multiplicative conjugate transposition of vectors, including problems of Chebyshev approximation, problems of approximation in the Hilbert seminorm, and pseudo-quadratic problems. To solve these problems, we apply methods based on the reduction to the solution of parametrized inequalities, matrix sparsification, and other techniques. The methods offer direct solutions represented in a compact explicit vector form ready for further analysis and straightforward computation. We conclude with a short discussion of the application of the results obtained to practical problems in location analysis, project scheduling and decision making.

3. Applying the Fractional Natural Decomposition Method to Solve Fractional Differential Equations in Multi-dimensional Space

Mahmoud Rawashdeh[1,*]

[1]Jordan University of Science and Technology; [*]msalrawashdeh@just.edu.jo;

In recent years, interest in the fractional differential equations has been stimulated due to their wide applications in various fields of engineering and science. Various vital phenomena in electromagnetic, viscoelasticity, fluid mechanics, electrochemistry, biological population models, and signal processing are well described by fractional differential equations. Also, they are employed in social sciences such as food supplement, climate, finance, and economics. As a result, the importance of obtaining exact or approximate solutions of fractional linear and nonlinear differential equations in physics and applied mathematics is still a significant problem that needs new methods. We propose a new method called the inverse fractional natural transform method (IFNTM). We present theoretical results and apply them

to obtain approximate solutions of linear fractional ordinary differential equations (LFODEs) and partial differential equations (LFPDEs). The fractional derivatives are described in the Caputo sense. The algorithm described in this study is expected to be further employed to solve similar linear problems in fractional calculus.

■ Applications in Healthcare

Room: **B023** (10:15 - 11:45)   Chair: *Secil Sozuer*

1. Kinetic Parameter Identification Based on Spectroscopic Data - advancements Illustrated by Case Studies

Christina Schenk[1,*], *Lorenz T. Biegler*[1], *Lu Han*[2], *Jason Mustakis*[2]

[1]Carnegie Mellon University, Pittsburgh, PA, USA; [*]schenk@cmu.edu;[2]Pfizer Inc., Groton, CT, USA;

The development of drug manufacturing processes involves dealing with spectroscopic data. When dealing with spectroscopic data, the identification of parameters and variances still remains a challenging task. In many cases kinetic parameter identification from spectroscopic data has to be performed without knowing the absorbing species in advance, such that they have to be estimated as well. However, kinetic parameter estimation is important in order to lower production costs, i.e. save measurements and equipment. Furthermore, scaling up from laboratory to industrial level relies on accurate kinetic parameter values. That is why, we take a closer look at the development of optimization-based procedures in order to estimate the variances of the noise in the system variables and spectral measurements. Then, with the estimated variances we determine the concentration profiles and kinetic parameters simultaneously using adequate strategies. The work is based on the approach proposed by [1] using maximum likelihood principles for simultaneous estimation of reaction kinetics and curve resolution from process and spectral data. For this a new software environment was developed which is continuously enhanced. These investigations and advancements are presented within this talk and illustrated by several case studies of pharmaceutical processes.

References [1] W Chen, LT Biegler, and SG MuÃśoz. An approach for simultaneous estimation of reaction kinetics and curve resolution from process and spectral data. Journal of Chemometrics, 30:506âĂŞ522, 2016.

2. A Further Study on the Opioid Epidemic Dynamical Model with Random Perturbation

Getachew Befekadu[1,*], *Quanyan Zhu*[2]

[1]Department of Mechanical and Aerospace Engineering, University of Florida - REEF; [*]gbefekadu@ufl.edu;[2]Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University;

In this talk, we consider an opioid epidemic dynamical model with random perturbation that describes the interplay between regular prescription use, addictive use, and the process of rehabilitation from addiction and vice-versa. In particular, we provide two-sided bounds on the solution of the density function for the Fokker-Planck equation that corresponds to the opioid epidemic dynamical model, when the random perturbation enters only through the dynamics of the susceptible group in the compartmental model. Here, the proof for such two-sided bounds basically relies on the interpretation of the solution for the density function as the value function of a certain optimal stochastic control problem. Finally, as a possible interesting development in this direction, we also provide an estimate for the attainable exit

probability with which the solution for the randomly perturbed opioid epidemic dynamical model exits from a given bounded open domain during a certain time interval.

3. Dynamic Appointment Scheduling Problem with Patient Preferences

Secil Sozuer[1,], Miao Bai, PhD[2], Robert H. Storer, PhD[1]*

[1]Lehigh University; [*]ses515@lehigh.edu;[2]Mayo Clinic;

Healthcare providers are under growing pressure to improve efficiency due to an aging population and increasing expenditures. This research is designed to address a particular healthcare scheduling problem, dynamic and stochastic appointment scheduling with patient choice. In the first part of the study, we consider that the stochasticity is coming from only uncertain patient arrivals and uncertain service duration realizations. The aim is to find the optimal schedule start time for the patients in order to minimize the expected cost incurred from patient waiting time, server idle time, and server overtime. By conducting perturbation analysis for the gradient estimation, a Sample Average Approximation (SAA) and a Stochastic Approximation (SA) algorithm are proposed. Various sampling methods such as stratified sampling, weighted sampling, and pure random sampling are analyzed within SA algorithm. The structural properties of the sample path cost function and expected cost function are studied. Numerical experiments show the computational advantages of SAA and SA over the mathematical model. In the second part of the study, in addition to having uncertain patient arrivals and uncertain duration realizations, we also take patient choices into account. We consider an appointment system where the patients have preferences about the two appointment days. In order to simulate choice probabilities, we use discrete choice models. We investigate different scheduling policies and conduct empirical analysis by considering the trade-off between the expected cost and patient satisfaction.

■ Optimization in Energy

Room: **Wood Dining Room** (14:00 - 15:30) Chair: *Daniel Molzahn*

1. Phase Transitions for Optimality Gaps in Optimal Power Flows

Pascal Van Hentenryck[1,]*

[1]University of Michigan; [*]pvanhent@umich.edu;

This paper investigates phase transitions on the optimality gaps in Optimal Power Flow (OPF) problem on real-world power transmission systems operated in France. The experimental results study optimal power flow solutions for more than 6000 scenarios on the networks with various load profiles, voltage feasibility regions, generation capabilities and objective functions. The results show that bifurcations between primal solutions and the QC, SOCP, and SDP relaxation techniques frequently occur when approaching congestion points. Moreover, the results demonstrate the existence of multiple bifurcations for certain scenarios when load demands are increased uniformly. Preliminary analysis on these bifurcations were performed.

2. Statistical Learning for (Power System) Optimization: An Active Set Approach

Line Roald[1,], Sidhant Misra[2], Yee Sian Ng[3]*

[1]University of Wisconsin-Madison; [*]roald@wisc.edu;[2]Los Alamos National Laboratory;[3]MIT;

Many engineering applications such as power system optimization involve solving similar optimization problems over and over and over again, with only slightly varying input parameters. In this talk, we consider the problem of using information available through this repeated solution process to directly learn the optimal solution as a function of the input parameters, thus reducing the need of solving computationally expensive optimization problems in real time.

To overcome limitations of traditional machine learning methods, which often struggle to enforce feasibility constraints or to leverage the knowledge available in the mathematical model, we propose a learning framework based on identifying the relevant set of active constraints (the so-called active set). Using active sets as features enables efficient recovery of the optimal solution, inherently accounts for relevant safety constraints and provides more interpretable results. Further, the number of relevant active sets is finite and sometimes small, which make them simpler objects to learn. To identify the relevant active sets, we propose a streaming algorithm with rigorous probabilistic performance guarantees. The algorithm is demonstrated using the optimal power flow (OPF) problem with renewable energy as an example. We establish that the number of active sets is indeed typically small for this application, and discuss theoretical and practical implications for power systems operation.

3. From Power System State Estimation to Low Rank Tensor Completion

Cédric Josz[1,], Ouyang Yi[2]*

[1]University of California, Berkeley; [*]cedric.josz@gmail.com;[2]UC Berkeley;

A recent line of research in mathematical programming consists of proving that simple and practical algorithms can solve non-convex optimization problems. This has been of particular interest for machine learning applications, among others; a recent paper by Zhang et al. (https://www.ocf.berkeley.edu/ ryz/pdf/SE_2018_1.pdf)discusses the existence of spurious local minima in the context of power system state estimation - which can be viewed as a special case of matrix sensing. This paper finds that as the number of measurements increases, local search algorithms are less likely to get stuck at a spurious local minima. In particular, the Gauss-Newton algorithm for solving non-linear least squares generally finds the global minimizer. We consider replacing the least squares objective (i.e. L2 norm) with the L1 norm with the aim of better coping with outliers in the data. This leads us to develop new mathematical tools to handle non-convex nondifferentiable optimization problems. These apply to state estimation and more generally tensor completion.

4. Optimal Power Flow with Robust Feasibility Guarantees

Daniel Molzahn[1,], Line Roald[2]*

[1]Argonne National Laboratory; [*]dmolzahn@anl.gov;[2]Los Alamos National Laboratory;

Solutions to optimal power flow (OPF) problems provide minimum cost operating points that satisfy both engineering limits and the power flow equations corresponding to the network physics. To account for the forecast uncertainty and short-term fluctuations that are inherent to many renewable energy sources, this presentation describes a recently proposed iterative algorithm for OPF problems which provides operating points that are guaranteed to be robust to renewable fluctuations

within specified ranges. The algorithm is based on the observation that considering uncertainty leads to a tightening of the original, deterministic constraints in order to safely accommodate fluctuations due to uncertain generation. The main challenge in solving the robust AC OPF problem is to guarantee the existence of feasible solutions for all points within the uncertainty set. To overcome this challenge, the proposed algorithm (1) employs convex relaxations of the AC power flow equations to obtain a conservative estimate of the required tightenings and (2) uses a sufficient condition that ensures power flow solvability for all uncertainty realizations.

■ Methods for Nonlinear Optimization

Room: **Givernor's Suite** (14:00 - 15:30)   Chair: *Yuyang Rong*

1. A Feasible Level-set Method for Optimization with Stochastic or Data-driven Constraints

Qihang Lin[1,], Selvaprabu (Selva) Nadarajah[2], Negar Soheili[2]*

[1]University of Iowa; [*]`qihang-lin@uiowa.edu`;[2]University of Illinois at Chicago;

We consider the constrained optimization where the objective function and the constraints are given as either finite sums or expectations. We propose a new feasible level-set method to solve this class of problems, which can produce a feasible solution path. To update a level parameter towards the optimality, our level-set method requires an oracle that generates upper and lower bounds as well as an affine-minorant of the level function. To construct the desired oracle, we reformulate the level function as the value of a saddle-point problem using the conjugate and perspective of constraints. Then a stochastic gradient method with a special Bregman divergence is proposed as the oracle for solving that saddle-point problem. The special divergence ensures the proximal mapping in each iteration can be solved in a closed form. The total complexity of both level-set methods using the proposed oracle are analyzed.

2. Level-set Methods for Finite-sum Constrained Convex Optimization

Runchao Ma[1,], Qihang Lin[1], Tianbao Yang[1]*

[1]University of Iowa; [*]`runchao-ma@uiowa.edu`;

We consider the constrained optimization where the objective function and the constraints are defined as summation of finitely many loss functions. This model has applications in machine learning such as Neyman-Pearson classification. We consider two level-set methods to solve this class of problems, an existing inexact Newton method and a new feasible level-set method. To update the level parameter towards the optimality, both methods require an oracle that generates upper and lower bounds as well as an affine-minorant of the level function. To construct the desired oracle, we reformulate the level function as the value of a saddle-point problem using the conjugate and perspective of the loss functions. Then a stochastic variance-reduced gradient method with a special Bregman divergence is proposed as the oracle for solving that saddle-point problem. The special divergence ensures the proximal mapping in each iteration can be solved in a closed form. The total complexity of both level-set methods using the proposed oracle are analyzed.

3. An Inexact Penalty Sequential Linear Optimization Method for Constrained Nonlinear Optimization

Yuyang Rong[1,], Yuyang Rong[2], Hao Wang[2], Jiashan Wang[3], Hudie Zhou[4]*

[1]School of Information Science and Technology, ShanghaiTech University; [*]`rongyy@shanghaitech.edu.cn`;[2]Sch. of Info. Sci. and Tech., ShanghaiTech University;[3]Dept. of Mathematics, University of Washington;[4]Sch. of Phy. Sci. and Tech., ShanghaiTech University;

The primary focus of this paper is on designing inexact penalty sequential linear optimization (commonly known as SLP) methods for solving large-scale constrained nonlinear optimization problems. By controlling the inexactness of the linear optimization subproblem solution, we can significantly reduce the computational cost needed per each subproblem. A penalty parameter updating strategy during the subproblem solve enables the algorithm to automatically detect infeasibility. Global convergence for both feasible and infeasible cases are proved. Complexity analysis for the KKT residual is also derived under loose assumptions. Numerical experiments exhibit the ability of the proposed algorithm to find optimal solution through cheap computational cost.

■ Recent Progress in Stochastic/Robust Optimization and Applications

Room: **B131** (14:00 - 15:30)   Chair: *Guanglin Xu*

1. A Copositive Approach for Multi-stage Robust Optimization Problems

Grani A. Hanasusanto[1,], Guanglin Xu[2]*

[1]The University of Texas at Austin; [*]`grani.hanasusanto@utexas.edu`;[2]University of Minnesota;

In this talk, we study generic multi-stage linear robust optimization problems. We employ linear decision rules for the case when the objective coefficients, the recourse matrices, and the right-hand sides are uncertain, and utilize quadratic decision rules for the case when only the right-hand sides are uncertain. The emerging optimization problems are NP-hard but amenable to copositive programming reformulations that give rise to tight conservative approximations. We provide theoretical and numerical results to demonstrate the effectiveness of the copositive programming approach.

2. Nurse Staffing under Uncertain Demand and Absenteeism

Minseok Ryu[1,], Ruiwei Jiang[1]*

[1]University of Michigan; [*]`msryu@umich.edu`;

This paper describes a data-driven approach for nurse staffing decision under uncertain demand and absenteeism. We propose a distributionally robust nurse staffing (DRNS) model with both exogenous (stemming from demand uncertainty) and endogenous uncertainty (stemming from nurse absenteeism). We provide a separation approach to solve the DRNS model with general nurse pool structures. Also, we identify several classes of nurse pool structures that often arise in practice and show how the DRNS model in each of these structures can be reformulated as a monolithic mixed-integer linear program that facilitates off-the-shelf commercial software. Built upon the DRNS model, furthermore, we propose an optimal nurse pool design model,

which produces an optimal pool structure that minimizes the number of cross-training while achieving a target staffing cost.

3. A Data-driven Distributionally Robust Optimization Approach for Appointment Scheduling With Random Service Durations and No-shows

Guanglin Xu[1,]*

[1]University of Minnesota; [*]gxu@umn.edu;

We study a single-server appointment-scheduling problem where the number of appointees and the sequence of their arrivals are fixed, and the appointees have no-show behavior. The service durations and no-shows are stochastic, but the underlying true probability distribution is un- known. With a collection of historical observations, we develop an ambiguity set, which contains the true distribution with a pre-determined statistical confidence level. We then develop a distributioanlly robust optimization model, which minimizes the worst-case total expected cost of appointment waiting, server idleness, and server overtime, by optimizing the scheduled arrival times of the appointees. Under some mild conditions, we reformulate the resulting problem into a tractable convex program, and for the general case, we propose a copositive program- ming reformulation, which motivates a tight semidefinite-programming-based approximation. We validate the effectiveness of our approach on benchmark scheduling instances in the existing literature.

■ Derivative Free and Black-Box Optimization

Room: **B023** (14:00 - 15:30)   Chair: *Albert Berahas*

1. Efficient Optimal Design of Latent Energy Storage Systems

Chunjian Pan[1,]*

[1]Energy Research Center, Lehigh University; [*]chp313@lehigh.edu;

Due to the intermittent nature of renewable energy resources, latent thermal energy storage (LTES) systems can play an important role in matching supply with demand. Efficient approaches for the optimal design of LTES systems is of particular interest. LTES systems exhibit both nonlinear and transient behaviors. It is computationally expensive to simulate LTES systems, and thus also typically to optimize their design. LTES design optimizations are often based on parametric studies, which neglect the interaction between design variables in this complex system, resulting in unrealistic performance enhancements. In this work, in order to facilitate efficient mathematical optimizations of LTES systems, a combined data-driven and physical knowledge-based model is built. This modeling approach is validated for two LTES systems and integrated into an optimal design framework.

2. A New local Parallelization for Particle Swarm Optimization

Abd AlRahman R. AlMomani[1,] Ahmad Almomani[2,**]*

[1]Clarkson University [2]SUNY Geneseo; [*]almomaa@clarkson.edu [**]almomani@geneseo.edu;

Derivative-free methods are highly demanded in the last two decades for solving optimization problems. Derivative-Free Optimization (DFO) are applicable for these problems where the derivatives are not available or hard to compute. Particle Swarm Optimization (PSO)

considered one of the best DFO global solvers and had shown to be efficient, few parameters, and flexible optimization algorithm. But it suffers from slow convergence in the refined search stage (weak local searchability), even though PSO resists being trapped in a local optimum but sometimes does. In this talk, an Unsupervised Particle Swarm Optimization (UPSO) algorithm introduced that profoundly improve the reliability, cost, and robustness of PSO. Our approach introduces new position and velocity update strategy based on the weighted gravitational force between all swarm individuals. Nelder-Mead algorithm parallelized with UPSO and Numerical results proposed for most known standard benchmark problems.

3. Derivative-free Optimization of Noisy Functions via Quasi-Newton Methods

Albert Berahas[1,], Jorge Nocedal[1], Richard Byrd[2]*

[1]Northwestern University; [*]albertberahas@u.northwestern.edu;[2]UC Boulder;

We present a finite difference quasi-Newton method for the minimization of noisy functions. The method takes advantage of the scalability and power of BFGS updating, and employs an adaptive procedure for choosing the differencing interval at every iteration, based on an estimate of the noise. The noise estimation procedure is inexpensive but not always accurate, and to prevent failures the algorithm incorporates a recovery mechanism that takes appropriate action in the case when the line search procedure is unable to produce an acceptable point. A novel convergence analysis is presented that considers the effect of a (noisy) line search routine. We report results of numerical experiments comparing the method to a popular model based trust region method.

■ Learning and Energy

Room: **Wood Dining Room** (16:00 - 17:30) Chair: *Michael Chertkov*

1. Data Recovery and Event Identification from Highly Quantized Measurements

Meng Wang[1,], Ren Wang[1]*

[1]Rensselaer Polytechnic Institute; [*]wangm7@rpi.edu;

We propose to add noise and apply quantization to synchrophasor measurements to increase the data privacy and reduce the communication cost. This talk focuses on data recovery and event identification at the operators' side from the highly quantized measurements. Event identification can be achieved by solving a data clustering problem that aims to group measurements affected by the same event together. The recovery and clustering are achieved simultaneously by solving a nonconvex constrained maximum likelihood problem. A fast algorithm with the performance guarantee is proposed to solve the nonconvex problem. The proposed method is evaluated numerically on recorded synchrophasor datasets. With large amounts of measurements, even if the resolution is low, our developed clustering method performs comparably to those methods on high-resolution measurements. Thus, the operator can successfully extract information, while an eavesdropper with a limited number of measurement cannot extract useful information even using the same method as the operator.

2. Renewable Scenario Generation Using Adversarial Networks

Baosen Zhang[1,]*

[1]University of Washington; [*]zhangbao@uw.edu;

Scenario generation is an important step in the operation and planning of power systems. In this talk, we present a data-driven approach for scenario generation using the popular generative adversarial networks, where to deep neural networks are used in tandem. Compared with existing methods that are often hard to scale or sample from, our method is easy to train, robust, and captures both spatial and temporal patterns in renewable generation. In addition, we show that different conditional information can be embedded in the framework. Because of the feedforward nature of the neural networks, scenarios can be generated extremely efficiently.

3. Learning Power Flows with Support Vector Machines

Ram Rajagopal[1,*]

[1]Standford University; [*]ramr@stanford.edu;

Optimization and management of distribution networks traditionally requires representing the physical laws that relate the power state variables in the system. Typically this requires either prior knowledge or learning the network topology and parameters from time-series measurements collected across the system. Yet, in real-world systems there are significant uncertainties such as the presence of equipment with unknown control rules that prevent full representation of system physics. Instead, we propose utilizing machine learning to identify the relationships between measurements. We show that a specific choice of kernel support vector machines recovers the power flow equations correctly although in a different representation. The approach can correctly learn relationships even in the presence of significant unmodelled components. We conclude the talk demonstrating how machine learning can be utilized to perform a basic reactive power optimization task in the network utilizing an efficient algorithm.

4. Learning and Control in Distribution Grids

Michael Chertkov[1,*]

[1]Los Alamos National Laboratory; [*]chertkov@lanl.gov;

Distribution Networks provide the final tier in the transfer of electricity from generators to the end consumers. In recent years, smart controllable devices, residential generator/storage devices and distribution grid meters have expanded the availability of sensor data in distribution networks. We discuss learning and control problems in distribution grid using available time-series measurements. For a range of realistic operating conditions, machine learning algorithms based on the dynamics of power flows are proposed to learn the structure of the distribution network as well as to estimate the statistics of load profiles at missing nodes and/or line parameters. The learning methods are generalizable for estimation of in other network of interest like sensor networks and smart buildings. We also discuss a Markov Decision Process framework for network aware control of smart devices in distribution grids under uncertainty that co-optimizes user comfort and global objectives.

■ Nonlinear Optimization

Room: **Givernor's Suite** (16:00 - 17:30)   Chair: *Majid Jahani*

1. Revisiting the Foundations of Randomized Gossip Algorithms

Nicolas Loizou[1,*]

[1]University of Edinburgh; [*]n.loizou@sms.ed.ac.uk;

In this work we present a new framework for the analysis and design of randomized gossip algorithms for solving the average consensus problem. We present how randomized Kaczmarz-type methods - classical methods for solving linear systems - work as gossip algorithms when applied to a special system encoding the underlying network and explain their distributed nature. We reveal a hidden duality of randomized gossip algorithms, with the dual iterative process maintaining a set of numbers attached to the edges as opposed to nodes of the network. Subsequently, using the new framework we propose novel block and accelerated randomized gossip protocols.

2. Convergence Rates of Proximal Gradient Methods via the Convex Conjugate

David Gutman[1,*], *Javier Pena*[2]

[1]Department of Mathematics, Carnegie Mellon University; [*]dgutman@andrew.cmu.edu;[2]Tepper School of Business, Carnegie Mellon University;

In this talk we propose a novel proof of the $O(1/k)$ and $O(1/k^2)$ convergence rates of the proximal gradient and accelerated proximal gradient methods for composite convex minimization. The crux of the new proof is an upper bound constructed via the convex conjugate of the objective function.

3. Efficient Distributed Hessian Free Algorithm for Large-scale Empirical Risk Minimization via Accumulating Sample Strategy

Majid Jahani[1,*], *Xi He*[1], *Chenxin Ma*[1], *Aryan Mokhtari*[2], *Dheevatsa Mudigere*[3], *Alejandro Ribeiro*[4], *Martin Takac*[1]

[1]Lehigh University; [*]maj316@lehigh.edu;[2]Massachusetts Institute of Technology;[3]Intel Labs;[4]University of Pennsylvania;

In this paper, we propose a Distributed Accumulated Newton Conjugate gradiEnt (DANCE) method in which sample size is gradually increasing to quickly obtain a solution whose empirical loss is under satisfactory statistical accuracy. Our proposed method is multistage in which the solution of a stage serves as a warm start for the next stage which contains more samples (including the samples in the previous stage). The proposed multistage algorithm reduces the number of passes over data to achieve the statistical accuracy of the full training set. Moreover, our algorithm in nature is easy to be distributed and shares the strong scaling property indicating that acceleration is always expected by using more computing nodes. Various iteration complexity results regarding descent direction computation, communication efficiency and stopping criteria are analyzed under convex setting. Our numerical results illustrate that the proposed algorithm can outperform other comparable methods for training machine learning tasks including neural networks.

■ Stochastic and Robust Optimization Algorithms and Applications

Room: **B131** (16:00 - 17:30)   Chair: *Fatma Kilinc-Karzan*

1. SUNlayer: Stable Denoising with Generative Networks

Soledad Villar[1,], Dustin G. Mixon[2]*

[1]New York University; [*]soledad.villar@nyu.edu;[2]Ohio State;

It has been experimentally established that deep neural networks can be used to produce good generative models for real world data. It has also been established that such generative models can be exploited to solve classical inverse problems like compressed sensing and super resolution. In this work we focus on the classical signal processing problem of image denoising. We propose a theoretical setting that uses spherical harmonics to identify what mathematical properties of the activation functions will allow signal denoising with local methods.

2. Stochastic ADMM Frameworks for Resolving Structured Stochastic Convex Programs

Yue Xie[1,], Uday V. Shanbhag[2]*

[1]Pennsylvania State University; [*]YUX111@psu.edu;[2]PSU IME;

We consider the program: $\min\{E[f(x,\xi)] + E[g(y,\xi)]|Ax + By = b\}$, which finds application in regularized expected risk minimization and distributed regression. To exploit problem structure and allow for distributed computation, we propose a stochastic inexact ADMM (SI-ADMM) where subproblems are solved inexactly via stochastic approximation. A.s. convergence and geometric convergence rate of mean-squared error can be obtained for SI-ADMM. Meanwhile, the related program $\min\{E[f(x,\xi)] + g(y)|Ax + By = b\}$ where $g(y)$ is deterministic but non-smooth is also studied. We propose variable sample-size stochastic ADMM (VSS-ADMM) and its accelerated variant (AVSS-ADMM). It is shown that the gap of convergence rate is closed between these stochastic frameworks and their deterministic counterparts. Furthermore, VSS-ADMM may recover the canonical oracle complexity. Preliminary numerical experiments demonstrate some favorable properties of SI-ADMM and we observe that AVSS-ADMM is comparable to the state-of-art algorithm in addressing graph-guided fused Lasso.

3. Exploiting Problem Structure in Optimization under Uncertainty via Online Convex Optimization

Nam Ho-Nguyen[1,]*

[1]Carnegie Mellon University; [*]hnh@andrew.cmu.edu;

Online convex optimization (OCO) has seen much success for its ability to handle decision-making in dynamic, uncertain, and even adversarial environments. In this work, we exploit these favorable attributes of OCO to develop iterative frameworks for two paradigms in optimization under uncertainty. First, we consider the generic robust convex optimization (RCO) paradigm seeking solutions that are robust to worst-case noise e.g., data perturbations from a fixed uncertainty set. By reformulating this problem as a convex-nonconcave saddle point problem, we avoid relying on robust counterparts, and instead exploit OCO to solve it using only simple first-order updates. Second, we demonstrate how OCO can be utilized within the joint estimation-optimization (JEO) paradigm, where the data associated with the optimization problem are continuously updated via an estimation process during the optimization procedure. We illustrate our results for the JEO paradigm on the problem of dynamic estimation of non-parametric choice models from continuously collected observational data.

■ Reinforcement Learning for Supply Chain

Room: **B023** (16:00 - 17:30)   Chair: *A. OroojlooyJadid, M. Nazari*

1. Concise Fuzzy Representation of Big Graphs: A Dimensionality Reduction Approach

Faisal Abu-Khzam[1,]*

[1]Lebanese American University;
[*]faisal.abukhzam@lau.edu.lb;

The enormous amount of data to be represented using large graphs exceeds in some cases the resources of a conventional computer. Edges in particular can take up a considerable amount of memory as compared to the number of nodes. However, rigorous edge storage might not always be essential to be able to draw the needed conclusions. A similar problem takes records with many variables and attempts to extract the most discernible features. It is said that the "dimension" of this data is reduced. Following an approach with the same objective in mind, we map a graph representation to a k-dimensional space and answer queries of neighboring nodes by measuring Euclidean distances. The accuracy of our answers would decrease but would be compensated for by fuzzy logic which gives an idea about the likelihood of error. This method allows for reasonable representation in memory while maintaining a fair amount of useful information. Promising preliminary results are obtained and reported by testing the proposed approach on a number of Facebook graphs.

2. RL for Inventory Optimization: Case on Beer Game

Afshin Oroojlooy[1,], MohammadReza Nazari[1], Lawrence Snyder[1], Martin Takac[1]*

[1]Lehigh University; [*]oroojlooy@gmail.com;

The beer game is a widely used in-class game that is played in supply chain management classes to demonstrate the bullwhip effect. The game is a decentralized, multi-agent, cooperative problem that can be modeled as a serial supply chain network in which agents cooperatively attempt to minimize the total cost of the network even though each agent can only observe its own local information. Each agent chooses order quantities to replenish its stock. Under some conditions, a base-stock replenishment policy is known to be optimal. However, in a decentralized supply chain in which some agents (stages) may act irrationally (as they do in the beer game), there is no known optimal policy for an agent wishing to act optimally.

We propose a machine learning algorithm, based on deep Q-networks, to optimize the replenishment decisions at a given stage. When playing alongside agents who follow a base-stock policy, our algorithm obtains near-optimal order quantities. It performs much better than a base-stock policy when the other agents use a more realistic model of human ordering behavior. Unlike most other algorithms in the literature, our algorithm does not have any limits on the beer game parameter values. Like any deep learning algorithm, training the algorithm can be computationally intensive, but this can be performed ahead of time; the algorithm executes in real time when the game is played. Moreover, we propose a transfer learning approach so that the training performed for one agent and one set of cost coefficients can be adapted quickly for other agents and costs. Our algorithm can be extended to

other decentralized multi-agent cooperative games with partially observed information, which is a common type of situation in real-world supply chain problems.

3. Reinforcement Learning for Solving the Vehicle Routing Problem

MohammadReza Nazari[1,], Afshin Oroojlooy[1], Lawrence Snyder[1], Martin Takac[1]*

[1]Lehigh University; [*]mon314@lehigh.edu ;

We present an end-to-end framework for solving the Vehicle Routing Problem (VRP) using reinforcement learning. In this approach, we train a single model that finds near-optimal solutions for problem instances sampled from a given distribution, only by observing the reward signals and following feasibility rules. Our model represents a parameterized stochastic policy, and by applying a policy gradient algorithm to optimize its parameters, the trained model produces the solution as a sequence of consecutive actions in real time, without the need to re-train for every new problem instance. On capacitated VRP, our approach outperforms classical heuristics and Google's OR-Tools on medium-sized instances in solution quality with comparable computation time (after training). We demonstrate how our approach can handle problems with split delivery and explore the effect of such deliveries on the solution quality. Our proposed framework can be applied to other variants of the VRP such as the stochastic VRP, and has the potential to be applied more generally to combinatorial optimization problems.

■ Resilience in Power Systems

1. Communication-constrained Expansion Planning for Resilient Distribution Systems

Pascal Van Hentenryck[1,], Geunyeong Byeon[1], Russell Bent[2], Harsha Nagarajan[2]*

[1]University of Michigan; [*]pvanhent@umich.edu;[2]Los Alamos National Laboratory;

Distributed generation and remotely controlled switches have emerged as important technologies to improve the resiliency of distribution grids against extreme weather-related disturbances. Therefore it becomes important to study how best to place them on the grid in order to meet a resiliency criteria, while minimizing costs and capturing their dependencies on the associated communication systems that sustains their distributed operations. This paper introduces the Optimal Resilient Design Problem for Distribution and Communication Systems (ORDPDC) to address this need. The ORDPDC is formulated as a two-stage stochastic mixed-integer program that captures the physical laws of distribution systems, the communication connectivity of the smart grid components, and a set of scenarios which specifies which components are affected by potential disasters. The paper proposes an exact branch-and-price algorithm for the ORDPDC which features a strong lower bound and a variety of acceleration schemes to address degeneracy. The ORDPDC model and branch-and-price algorithm were evaluated on a variety of test cases with varying disaster intensities and network topologies. The results demonstrate the significant impact of the network topologies on the expansion plans and costs, as well as the computational benefits of the proposed approach.

2. TBA

J.P. Watson[1,]*

[1]Sandia National Laboratories; [*]jwatson@sandia.gov;

3. Probabilistic N-k Failure-identification for Power Systems

Harsha Nagarajan[1,], Kaarthik Sundar[1], Carleton Coffrin[1], Russell Bent[1]*

[1]Los Alamos National Laboratory; [*]harsha@lanl.gov;

This talk will deal with a probabilistic generalization of the N-k failure-identification problem in power transmission networks, where the probability of failure of each component in the network is known a priori and the goal of the problem is to find a set of k components that maximizes disruption to the system loads weighted by the probability of simultaneous failure of the k components. The resulting problem is formulated as a bilevel mixed-integer nonlinear program. Convex relaxations, linear approximations, and heuristics are developed to obtain feasible solutions that are close to the optimum. A general cutting-plane algorithm is proposed to solve the convex relaxation and linear approximations of the N-k problem. Extensive numerical results that corroborate the effectiveness of the proposed algorithms on small-, medium-, and large-scale test instances shall also be presented; the test instances include the IEEE 14-bus system, the IEEE single-area and three-area RTS96 systems, the IEEE 118-bus system, the WECC 240-bus test system, the 1354-bus PEGASE system, and the 2383-bus Polish winter-peak test system.

4. Designing Resilient Distribution Systems under Natural Disasters

Ruiwei Jiang[1,]*

[1]University of Michigan; [*]ruiwei@umich.edu;

Distribution system topology planning is a critical task for system operators in order to ensure the economic operation, resilience and reliability of the power supply. This paper proposes a novel distributionally robust defender-attacker-defender model for designing a distribution system to withstand the risk of disruptions imposed by natural disasters. The proposed model optimally configures the network topology and integrates distributed generation to effectively manage the loads. Moreover, the uncertainty of contingency occurrence due to natural disasters is taken into account. Using the moment information of asset failures based on component failure data, we construct an ambiguity set of plausible probability distributions of system contingencies, and minimize the load shedding with regard to the worst-case distribution within the ambiguity set. On the one hand, we consider the stochasticity of natural disasters and provide a less conservative configuration than that by a classical robust optimization approach. On the other hand, our approach considers distributional ambiguity and so is more reliable than stochastic programming. We recast the proposed model as a two-stage robust optimization formulation and solve it using the Column-and-Constraint Generation framework. We demonstrate the performance of the proposed approach in case studies.

■ Optimization and Learning

1. A Trust-Region Method for Minimizing Regularized Non-convex Loss Functions

Dimitri Papadimitriou[1,]*

[1]University of Antwerp;
[*]dimitri.papadimitriou@uantwerpen.be;

The training of deep neural networks is typically conducted via nonconvex optimization. Indeed, for nonlinear models, the nonlinear nature of the activation functions yields empirical loss functions that are nonconvex in the weight parameters. Even for linear models, i.e., when all activation functions are linear with respect to inputs and the output of the entire deep neural network is a chained product of weight matrices with the input vector, the (squared error) loss functions remain nonconvex. On the other hand, to circumvent the limits resulting from finding sharp minima (corresponding to weight parameters specified with high precision) of the empirical loss function, Hochreiter suggested in 1995 to find a large region in the weight parameter space with the property that each weight from that region can be given with low precision and lead to similar small error. In this paper, we propose to minimize the empirical loss (training error) together with weights precision (regularization error) by means of a Trust Region (TR)-based algorithm. When extended to nonconvex regularized objectives, this method contrasts to current techniques which either arbitrarily -sometimes strongly- convexify the empirical loss minimization problem or involve slowly converging Stochastic Gradient algorithms without guaranteeing the production of good predictors. TR methods instead provide i) better complexity bounds for convergence to first- and second-order critical points by means of rich set of iterative methods for TR subproblem solving, e.g., Steighaug-Toint and Generalized Lanczos Trust-Region (GLTR); and ii) fast escape from saddle points, e.g., by exploiting the Hessian information. In addition, they can be combined with approximation techniques (e.g., sub-sampling) that are effective in reducing computational cost associated to Hessian evaluation. The latter provides an essential property in solving

high-dimensional instances. Performance bounds of the TR-based algorithm are characterized against gradient descent together with numerical experiments for evaluation and comparison purposes.

2. Bregman-divergence for Legendre Exponential Families

Hyenkyun Woo[1],

[1]Korea University of Technology and Education;
*hyenkyun@gmail.com;

Bregman-divergence is a well-known generalized distance framework in various application area, such as machine learning, signal and image processing. In more specific, in case of Bregman-divergence associated with convex function of Legendre type, due to its dual structure of the Bregman divergence, it is easy to find global optimum in terms of second variable, irrespective of the convexity of the Bregman-divergence. In this talk, through dual structure of the Bregman- divergence associated with restricted convex function of Legendre type, we analyze the structure of the Legendre exponential families whose cumulant function is also conjugate convex function of Legendre type. Actually, the Legendre exponential families is an extended one of the regular exponential families including non-regular exponential families, such as the inverse Gaussian distribution. The main advantage of the Bregman-divergence associated with restricted convex function of Legendre type is that it offers systematic successive approximation tools to handle closed domain issues arising in non-regular exponential families and the statistical distributions having discrete random variables, such as Bernoulli distribution and Poisson distribution. As an application, we show how to use Bregman-divergence associated with restricted convex function of Legendre type for generalized center-based clustering problems.

3. Learning-based Robust Optimization: Procedures and Statistical Guarantees

Zhiyuan Huang[1],, Henry Lam[2], Jeff Hong[3]*

[1]University of Michigan; *zhyhuang@umich.edu;[2]Columbia University;[3]City University of Hong Kong;

We discuss a statistical framework to integrate data into robust optimization, based on prediction set learning and a simple data-splitting validation scheme that achieves finite-sample statistical guarantees on the feasibility of the underlying uncertain constraints. We demonstrate several features of the framework, including a dimension-free sample size requirement for the feasibility guarantees and the capability to self-improve existing solutions in terms of both optimality and feasibility.

■ Conic Optimization and Integer Programming

Room: **B131** (10:15 - 11:45)   Chair: *Ali Mohammad Nezhad*

1. Determine the Maximum Permissible Perturbation Set of SDP Problem With Unknown Perturbations

Tingting Tang[1],, Jonathan Hauenstein[2]*

[1]Department of Applied and Computational Mathematics and Statistics, University of Notre Dame; *ttang@nd.edu;[2]University of Notre Dame;

We study the property of the solution of semidefinite programs with multi-dimension perturbation variable using the Davidenko differential equations. Under the assumptions of strict complementary and non-degeneracy, we aim to find the *a priori* unknown maximal permissiable perturbation set where the semidefinite program after perturbation has unique optimum and solution set. A sweeping euler numerical method is developed to approximate the *a priori* unknown maximal perturbation set and solve the semidefinite program within this set. We prove local and global error bounds for this second-order sweeping Euler scheme and demonstrate results on several examples.

2. Optimal Cutting Planes from the Group Relaxations

Amitabh Basu[1],

[1]Johns Hopkins University; *Basu.amitabh@jhu.edu;

We study quantitative criteria for evaluating the strength of valid inequalities for Gomory and Johnson's finite and infinite group models and we describe the valid inequalities that are optimal for these criteria. We justify and focus on the criterion of maximizing the volume of the nonnegative orthant cut off by a valid inequality.
For the finite group model of prime order, we show that the unique maximizer is an automorphism of the *Gomory Mixed-Integer (GMI) cut* for a possibly *different* finite group problem of the same order.
We extend the notion of volume of a simplex to the infinite dimensional case. This is used to show that in the infinite group model, the GMI cut maximizes the volume of the nonnegative orthant cut off by an inequality.

3. Covering Problem via Nonlinear Semidefinite Programming

Walter Gomez[1],, Ernesto Birgin[2], Gabriel Haeser[2], Leonardo Mito[2], Daiana Viana[2]*

[1]Universidad de La Frontera;
*walter.gomez@ufrontera.cl;[2]University of Sao Paulo;

This talk is devoted to the problem of covering a given object with a fixed number of balls of minimal radius. A reformulation of the problem as nonlinear semidefinite programming is presented. Some theoretical convergence results to generalized optimality conditions are discussed. Numerical experiments showing the difficulties associated with this approach are finally presented.

■ Sparse Optimization

Room: **B023** (10:15 - 11:45)   Chair: *Quoc Tran-Dinh*

1. L0-regularized Sparsity for Probabilistic Mixture Models

Dzung Phan[1],

[1]IBM Research; *phandu@us.ibm.com;

This talk revisits a classical task of learning probabilistic mixture models. Our major goal is to sparsely learn the mixture weights to automatically determine the right number of clusters. The key idea is to use a novel Bernoulli prior on the mixture weights in a Bayesian learning framework, and formalize the task of determining the mixture weights as an L0-regularized optimization problem. By leveraging a specific mathematical structure, we derive a quadratic time algorithm for efficiently solving the non-convex L0-based problem. In experiments, we evaluate the performance of our proposed approach over existing sparse methods in terms of accuracy and scalability on synthetic and real data sets.

2. The CCP Selector: Scalable Algorithms for Sparse Ridge Regression from Chance-constrained Programming

Weijun Xie[1,*]*, Xinwei Deng*[2]

[1]Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg VA 24061; [*]wxie@vt.edu;[2]Department of Statistics, Virginia Tech, Blacksburg VA 24061;

Sparse regression and variable selection for large-scale data have been rapidly developed in the past decades. This work focuses on sparse ridge regression, which considers the exact L0 norm to pursue the sparsity. We pave out a theoretical foundation to understand why many existing approaches may not work well for this problem, in particular on large-scale datasets. Inspired by reformulating the problem as a chance-constrained program, we derive a novel mixed integer second order conic (MISOC) reformulation and prove that its continuous relaxation is equivalent to that of the convex integer formulation proposed in a recent work. Based upon these two formulations, we develop two new scalable algorithms, the greedy and randomized algorithms, for sparse ridge regression with desirable theoretical properties. The proposed algorithms are proved to yield near-optimal solutions under mild conditions. In the case of much larger dimensions, we propose to integrate the greedy algorithm with the randomized algorithm, which can greedily search the features from the nonzero subset identified by the continuous relaxation of the MISOC formulation. The merits of the proposed methods are elaborated through a set of numerical examples in comparison with several existing ones.

3. Accelerated Preconditioned Alternating Direction Methods of Multipliers with Non-ergodic Optimal Rates

Quoc Tran-Dinh[1,*]

[1]University of North Carolina at Chapel Hill; [*]quoctd@email.unc.edu;

We develop two new variants of alternating direction methods of multipliers (ADMM) and two parallel versions to solve a wide range class of constrained convex optimization problems. Our approach relies on a novel combination of the augmented Lagrangian framework, partial alternating/linearization scheme, Nesterov's acceleration technique, and a homotopy strategy. The proposed algorithms have the following new features compared to existing ADMM variants. Firstly, they have a Nesterov's acceleration step on the primal variables instead of the dual ones as in several existing ADMM variants. Secondly, they possess an optimal $O(1/k)$-convergence rate guarantees in a non-ergodic sense without any smoothness or strong convexity-type assumption, where k is the iteration counter. When one objective term is strongly convex, our algorithm achieves an optimal $O(1/k^2)$-non-ergodic rate. Thirdly, our methods have better per-iteration complexity than standard ADMM. Fourthly, we provide a set of conditions to derive update rules for algorithmic parameters and give a concrete update for these parameters as an example. Finally, when the objective function is separable, our methods can naturally be implemented in a parallel fashion. We also study some extensions of our methods and a connection to existing primal-dual methods. We verify our theoretical development via different numerical examples and compare our methods with some existing state-of-the-art algorithms.

■ Equilibrium and Complementarity Modeling in Energy Markets

Room: **Wood Dining Room** (14:00 - 15:30) Chair: *Jalal Kazempour*

1. Efficiency and Welfare Distribution Effects From the Norwegian-Swedish Tradable Green Certificate Market

Asgeir Tomasgard[1,*]*, Per Ivar Helgesen*[2]

[1]NTNU; [*]asgeir.tomasgard@ntnu.no;[2]NTNU and ENOVA;

We investigate economic impacts of the Norwegian-Swedish Green Certificate Market, which promotes electricity produced from renewable energy sources. We formulate a mixed complementarity, multi-region, partial equilibrium model, clearing both the electricity and green certificate markets under perfect competition. The model applies a linearized DC approximation of power flows in the transmission network imposing both Kirchhoff's current and voltage laws in a mixed complementarity formulation suitable for policy analysis.
The certificate scheme combines a subsidy to producers of renewable energy and a tax paid by consumers. The scheme increases the electricity supply and leads to welfare reallocation from old producers to new producers as well as consumers. Consumers pay for the support scheme through a tax, but still benefit in most scenarios due to price decreases in the wholesale markets. The deadweight loss and welfare transfers are reduced when new interconnectors are built or new demand enters the system. We show how geographical distribution effects and locations of new production are affected by the representation of the internal network between the model regions.

2. A Column-and-constraint Decomposition Approach for Solving EPECs

David Pozo[1,*]

[1]Skoltech; [*]d.pozo@skoltech.ru;

Equilibrium Problems with Equilibrium Constraints (EPECs) have been good mathematical models to represent hierarchical interactions among participants in several sectors such as deregulated power systems. However, they are hard to solve and present computational challenges because they are generally non-linear and non-convex and therefore a global solution for EPECs is seldom reached. There is neither a well-established theory on solving EPEC problems nor specific decomposition techniques because the bad properties inherit from the set of equilibrium/complementary constraints that forms the EPEC. Numerical methods for solving EPECs constitute an ongoing topic for many researchers. In this talk, we will present a column-and-row decomposition algorithm for solving EPECs where decisions/strategies are discrete. We prove that under mild assumptions the solution obtained is global. We illustrate the proposed methodology on the generation capacity expansion equilibrium problem.

3. Market Integration of HVDC

Spyros Chatzivasileiadis[1,*]*, Andrea Tosatto*[1]*, Tilman Weckesser*[1]*, Spyros Chatzivasileiadis*[1]

[1]Technical University of Denmark (DTU); [*]spchatz@elektro.dtu.dk;

Moving towards regional Supergrids, an increasing number of interconnections are formed by High Voltage Direct Current Lines (HVDC). Currently, HVDC line losses are not explicitly considered in market operations, resulting in additional costs for the TSOs. The introduction of tariffs to ensure budget balance for TSOs has been a central topic during the early stage of the implementation of the Internal Electricity Market of the EU, but HVDC lines have never been

considered. In this paper, we are introducing a methodology to determine a loss factor for HVDC lines to be integrated in the market clearing algorithm, and we study the impact of the introduction of HVDC grid tariffs on social welfare.

First, we focus on the identification of an appropriate loss factor. We propose and compare three different models: fixed, linear, and piece wise linear loss factors. Second, we investigate the appropriate cost allocation of HVDC losses among the different market participants and the design of an appropriate mechanism to explicitly account for the cost of losses. This is done by formulating the problem as a mathematical program with equilibrium constraints (MPEC), where lower-level problems are the profit maximization of each market participant and the upper-level problem is the social welfare maximization from a regulator's point of view.

4. Bi-level Network Planning with Generation-market Equilibria Subject to Transmission Costs Recovery

Pengcheng Ding[1,], Pengcheng Ding[1], Shisheng Cui[2], Uday Shanbhag[2], Ben Hobbs[1]*

[1]Johns Hopkins University;
[*]pangchingting@gmail.com;[2]Pennsylvania State University;

Transmission costs recovery, especially for the fixed operation and maintenance costs, is often neglected in transmission planning. We want to see how would a proactive transmission planner react to transmission charges for cost recovery. We have modelled two costs recovery schemes: one postage stamp type charge and one marginal MW-miles based charge adopted from UK's OFGEM. We built a bi-level program with the system operator on the upper level deciding transmission investments and incorporated the charging schemes into the lower level generation expansion and market clearing equilibrium. We illustrated the effects of the charges in a simple two-node system under both nodal and zonal markets. To understand the effects in a more realistic system, we have also done a case study in 5-node nodal network by formulating the planning problem as a Mixed Integer Mathematical Program with Equilibrium Constraints, which we solved combining a branching method and with binary relaxation. We have found the efficiency impacts of two costs recovery schemes varying significantly with different network and market structures.

Transmission network cost recovery, including both investment costs and fixed O&M expenses, is a very important issue in generation investment, since cost allocation and the specific means of transmission charging can have a large effect on where and what type of investment to make. However, this interaction of network cost recovery and generation decisions is usually neglected in transmission planning. Some transmission planning models are proactive, and anticipate where generation will be sited in response to alternative network configurations; however, those models usually assume efficient transmission pricing (pure locational marginal pricing).

In this presentation, we investigate how generation decisions would be affected by alternative network cost recovery methods, and how a proactive transmission planner can take those reactions into account. Multilevel transmission planning optimization models are used to explore these issues. We have modeled two potentially inefficient cost recovery schemes: one is a postage-stamp type charge and the other is a marginal MW-miles based charge similar that that used by UK's Office of Gas and Electricity Markets. We build a bi-level program. In the upper level, the system operator decides transmission investments; in the lower level, transmission charging schemes are incorporated into the lower level generation expansion and market clearing equilibrium.

We illustrated the effects of the charges in a simple two-node system under both nodal and zonal markets. To understand the effects in a more realistic system, we have also done a case study in 5-node nodal network by formulating the planning problem as a Mixed Integer Mathematical Program with Equilibrium Constraints, which we

solved combining a branching method and with binary relaxation. We have found that the efficiency impacts of the two cost recovery schemes (relative to the LMP ideal) vary depending on different network and market structures

■ Stochastic Gradient Decent and Conve Optimization

Room: **Givernor's Suite** (14:00 - 15:30) Chair: *Phuong Ha Nguyen*

1. Reliable Machine Learning Using Unreliable Components: Error-runtime Trade-offs in Distributed SGD

Sanghamitra Dutta[1,], Gauri Joshi[1], Soumyadip Ghosh[2], Parijat Dube[2], Priya Nagpurkar[2], Pulkit Grover[1]*

[1]Carnegie Mellon University;
[*]sanghamd@andrew.cmu.edu;[2]IBM TJ Watson Research Center;

Distributed Stochastic Gradient Descent (SGD) when run in a synchronous manner, suffers from delays in waiting for the slowest learners (stragglers). Asynchronous methods can alleviate stragglers, but cause gradient staleness that can adversely affect convergence. In this work we present a novel theoretical characterization of the speed-up offered by asynchronous methods by analyzing the trade-off between the error in the trained model and the actual training runtime (wall-clock time). The novelty in our work is that our runtime analysis considers random straggler delays, which helps us design and compare distributed SGD algorithms that strike a balance between stragglers and staleness. We also present a new convergence analysis of asynchronous SGD variants without bounded or exponential delay assumptions.

2. Complexity Bounds for Structured Convex Optimization

Yuyuan Ouyang[1,]*

[1]Clemson University; [*]yuyuano@clemson.edu;

We consider a class of convex optimization with certain structure. In particular, the structure of the problem may involve the smoothness of certain loss functions, saddle-point formulation, and linear equality constraints. In order to understand the efficiency of first-order methods, it is important to study the impact of the specific structure on the complexity of the problem. We will design worst-case instances that yield lower complexity bounds of any first-order method.

3. Optimal Diminishing Stepsizes in SGD for Strongly Convex Objective Functions

Phuong Ha Nguyen[1,], Lam M. Nguyen[2], Marten van Dijk[3]*

[1]University of Connecticut;
[*]phuong_ha.nguyen@uconn.edu;[2]Industrial and Systems Engineering,Lehigh University, USA;[3]Electrical and Computer Engineering, University of Connecticut, USA;

In this talk, we consider the Stochastic Gradient Algorithm (SGD) which performs the following iterations: $w_{t+1} = w_t - \eta_t \nabla f(w_t; \xi_t)$. Without *bounded gradient assumption*, we prove that the best possible *lower bound* and *upper bound* of the convergence rate for *any* possible stepsizes η_t are $\mathcal{O}(1/t)$ if we are given information about μ-strong convex, L-smooth properties, $N = 2\mathbb{E}[\|\nabla f(w_*; \xi)\|^2]$, and an oracle accessing $Y_t = \mathbb{E}[\|w_t - w_*\|^2]$ at the t-th iteration for updating w_{t+1},

where w_* is the optimal solution. This result implies that given only μ-strongly convex and L-smooth assumptions, the best *lower bound* and *upper bound* of the convergence rate for *any* stepsizes η_t for Stochastic Gradient Algorithm is $\mathcal{O}(1/t)$. We show this by constructing an example of an objective function for which the *lower bound* and *upper bound* of the smallest convergence rate that can be achieved by using a step size as a function of μ, L, N and Y_t are $\mathcal{O}(1/t)$. This result implies the optimality of the stepsize proposed in the study of Nguyen et al. (2018), i.e., $\eta_t = \frac{2}{\mu t + 4L}$ for $t \geq 0$. As an important conclusion, this talk shows that to achieve a better convergence, more information than μ, L, N and Y_t is needed.

■ Optimization in Machine Learning

Room: **B131** (14:00 - 15:30)   Chair: *Mohammad Pirhooshyaran*

1. A Machine Learning Technique for Quadcopter State Estimation

Arash Amini[1,], Yaser Ghaedsharaf[4], Nader Motee[1]*

[1]Lehigh University; [*]ara416@lehigh.edu;

Due to imperfections of sensors and calibration algorithms, those sensors fails to report the attitude of the vehicle with enough preciseness. Sensor fusion algorithms need tuning, which is time-consuming and those provide a low accuracy especially at small angles. Therefore we tried an alternative solution to improve our result and minimizing computation by using a neural network to predict better estimation of states of quadcopter based on previous onboard data during flight. The final result shows a significant improvement on the accuracy especially for small angles, by using a small neural network with two hidden layers and RELU Activision function to find a prediction for the attitude of the quadcopter.

2. Hyperparameter Tuning of Neural Networks(NNs) via Derivative Free Optimization(DFO)

Mertcan Yetkin[1,], Mohammad Pirhooshyaran[1], Katya Scheinberg[1]*

[1]Lehigh University; [*]mey316@lehigh.edu;

Derivative Free Optimization (DFO) has been shown to be a powerful tool for optimizing complex simulation and black-box functions. In this study, we investigate black-box functions arising in the optimization of the performance of the Neural Network (NN) structures. This performance is strongly dependent on parameters which define the architecture of the model. We aim to explore the effect of hyperparameter tuning through a model-based DFO which employs trust region framework to optimize the model. We extend the method to include discrete and bounded parameters of Convolutional and Recurrent Neural Networks (CNN and RNN). Then, we demonstrate performance improvements in terms of test accuracy on well-known datasets. Moreover, the effect of the parameter values of the testing accuracy is inspected for different number of training epochs to utilize the stochastic nature of our black-box function.

3. Bidirectional LSTM Ensemble Structures for Multi-step Forecasting Of Ocean Wave Elevation

Mohammad Pirhooshyaran[1,], Lawrence V. Snyder[2]*

[1]PhD; [*]mop216@lehigh.edu;[2]Lehigh University;

There exists an undeniable interest toward utilizing ocean energy as a renewable energy source. This research focuses on the Multiple step ahead prediction of ocean wave heights and power via innovative structure of Deep Neural Networks. An ensemble of Bidirectional Long short Term Memory (BLSTM) Networks is proposed to capture both long and short dependencies of the large historical data. Several optimization algorithms such as ADAM and RMSProb, SGD, etc., are considered to optimize the network. Furthermore, Derivative Free Optimization (DFO) framework is used to tune network parameters. The results indicate that the proposed model is more accurate in compare with conventional forecasting methods such as Support Vector Regression (SVR) or SARIMA.

Index of Authors