MOPTA 2023

Modeling and Optimization: Theory and Applications

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

16-18 August '23

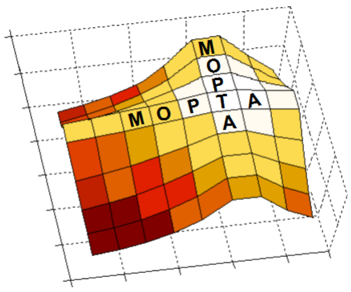
Lehigh University
Bethlehem, PA, USA

Program and Abstracts

■ Welcome to MOPTA 2023!

The Modeling and Optimization: Theory and Applications (MOPTA) conference is an annual event hosted by the Department of Industrial and Systems Engineering at Lehigh University. MOPTA aims at bringing 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. The goal is to present a diverse set of exciting new developments from different optimization areas while at the same time providing a setting that will allow increased interaction among the participants. We aim to bring together researchers from both the theoretical and applied communities who do not usually have the chance to interact in the framework of a medium-scale event.

Tamás Terlaky
Lehigh University



■ Conference Details

Conference Statistics

- 3 day conference;
- 7 plenary speakers;
- 32 sessions, 96 talks;
- Over 135 registrations;
- Optimization and modeling competition, 13 teams from around the world;
- Poster session and competition, 8 participants.

Conference Location

The **conference** will take place at:
Lehigh University
Rauch Business Center
621 Taylor Street
Bethlehem, PA 18015
US

Conference Dinner Location

The **Cocktail Reception** and the **Banquet** will take place at:
Lamberton Hall, 690 Taylor St, Bethlehem, PA 18015

Student Social Location

Student Social is a social event for undergraduate/graduate students and postdocs. Pizza and soda will be served. Student Social will take place at Packer House, which is located at 217 W. Packer Avenue on Lehigh University's Asa Packer Campus (10 minute walk from the conference venue).

■ Lehigh University's Asa Packer Campus Map

230 W. Packer House 22
Alpha Epsilon Pi 102
Access Control/Locksmith 42
Alpha Gamma Delta 95
Alpha Omicron Pi 107
Alpha Phi 98
Alumni Memorial Building
  Office of Admissions 27
  Office of Financial Aid 27
Alumni Parking Pavilion 26a
Barnett House 194
Beardslee House 55
Bookstore/Farrington Square 202c
Brinker Lofts 1
Brodhead House 201
Budget Office 198
Business Services 300
Business Innovation Building 206
Carothers House 54
Centennial I Commons 58
Central Heating & Refrigeration 10
Chandler-Ullmann Hall 17
Chi Phi 105
Chi Psi 84
Christmas-Saucon Hall/Annex 14/a
Clayton University Center 29
Communications, Government and
  Public Affairs Office 331
Congdon House 56
Controller's Office 300
Coppee Hall 31
Coxe Hall 32
Cultural Greek Council House 85
Delta Chi 86
Delta Upsilon 81
Dialogue Center 41
Dravo House 65
Drinker House 64
Drown Hall 35
Emery House 57
Environmental Health & Safety 23a
E.W. Fairchild-Martindale Library
  and Computing Center 8
Facilities Services/Planning 40
Farrington Square A-D 202
Farrington Square Parking 002
Fritz Engineering Laboratory 13
Gamma Phi Beta 100
Grace Hall 39
Health, Science, Technology
  Building 205
Cultural Greek Council House 85
House 87 87
House 88 88
House 89 89
House 90 90
House 93 93
House 99 99
Housing Services 63
Human Resources 300
Humanities Center 224
ID Card/GoldPLUS
  Access Control/Locksmith 42
Jewish Student Center 21
Johnson Hall 36
Human Resources 300
Kappa Alpha Theta 106
Kappa Delta 104
Lamberton Hall 34
Leavitt House 62
Linderman Library 30
Maginnes Hall 9
McClintic-Marshall House 67
McConn House 61
Mohler (Harold S.) Laboratory 200
Mudd (Seeley G.) Building 6
Neville Hall 6a
Packard Laboratory 19
Packer House, Graduate
  Student Center 20
Packer Memorial Church 18
Palmer House 53
Parking Services Office 199
Phi Delta Theta 97
Philosophy Building 15
Pi Beta Phi 83
Psychology 23
Police Department 47
President's House 28
Price Hall 40
Psi Upsilon 80
Purchasing 300
Rathbone Hall 63
Rauch Business Center 37
Research and Sponsored
Programs 23b
Richards House 66
Risk Management/Internal Audit . 197
Sayre Building 26
Sayre Park Village .... 109 (a-d)
Sherman Fairchild Center
  for The Physical Sciences 16
Sherman Fairchild Laboratory
  for Solid State Studies 16A
Sigma Phi Delta 96
Sigma Phi Epsilon 92
Sinclair Laboratory 7
Singleton, Hitch & Maida
Houses .... 69a, 69b, & 69c
Smiley House 59
South Side Commons 204
STEPS Building 9a
Stevens House 52
Stoughton House 51
Sustainability Office 23
Taylor Gymnasium 38
Taylor House 68
The Lehigh Store and Mail
Center 202c
Theta Chi 91
Thornburg House 60
Trembley Park 71-77
Umoja House 101
Warren Square A 196
Warren Square B 192
Warren Square C 24
Warren Square D 193
Warren Square E 25
Whitaker Laboratory 5
Wilbur Annex 11
Wilbur Powerhouse 12
Williams Hall 33
Williams House 50
Zeta Tau Alpha 94
Zoellner Arts Center 49
Zoellner Parking Garage 48

LEHIGH UNIVERSITY

■ Plenary Talks

Wednesday 9:00–10:00 am. Omar Ghattas

Geometric Neural Surrogates for Optimization Problems Governed by PDEs

The University of Texas at Austin

omar@oden.utexas.edu

Optimization problems governed by PDEs with high- or infinite-dimensional decision/parameter spaces abound across numerous scientific and engineering disciplines. When the PDEs are characterized by random coefficient fields, these optimization problems are often prohibitive to solve. Examples include Bayesian inverse problems (BIPs), Bayesian optimal experimental design (BOED), and optimal control under uncertainty (OCU). Efficient evaluation of the maps from parameters or decision variables to objective functions, which involves solution of the forward PDEs, is key to making such problems tractable. Surrogate approximations of these maps can greatly accelerate solution of these optimization problems, provided an accurate surrogate can be trained with modest numbers of PDE solves. Unfortunately, constructing such surrogates presents significant challenges when the parameter/decision dimension is high and the forward model is expensive.

Deep neural networks (DNNs) have emerged as leading contenders for overcoming these challenges. We demonstrate that black box application of DNNs for problems with infinite dimensional parameter fields leads to poor results when training data are limited due to the expense of the model. Instead, by constructing a network architecture that captures the geometry of the map – in particular its smoothness, anisotropy, and intrinsic low-dimensionality – as revealed through adjoint-PDE-based Gauss-Newton Hessians, one can construct a dimension-independent reduced basis DNN surrogate with superior generalization properties using just limited training data. We employ this reduced basis DNN surrogate to make tractable the solution of PDE-constrained BIPs, BOED, and OCU, with applications to inverse wave scattering, inverse hyperelasticity, inverse earthquake subduction, and optimal flow control.

This work is joint with Thorsten Becker, Lianghao Cao, Peng Chen, Dingcheng Luo, J. Tinsley Oden, Thomas O'Leary-Roseberry, Simone Puel, Umberto Villa, and Keyi Wu.

Speaker Biography. Dr. Omar Ghattas is Professor of Mechanical Engineering at The University of Texas at Austin and holds the Fletcher Stuckey Pratt Chair in Engineering. He is also the Director of the OPTIMUS (OPTimization, Inverse problems, Machine learning, and Uncertainty for complex Systems) Center in the Oden Institute for Computational Engineering and Sciences. He is a member of the faculty in the Computational Science, Engineering, and Mathematics (CSEM) interdisciplinary PhD program in the Oden Institute, and holds courtesy appointments in Geological Sciences, Computer Science, and Biomedical Engineering. Before moving to UT Austin in 2005, he spent 16 years on the faculty of Carnegie Mellon University. He holds BSE (civil and environmental engineering) and MS and PhD (computational mechanics) degrees from Duke University. With collaborators, he received the ACM Gordon Bell Prize in 2003 (for Special Achievement) and again in 2015 (for Scalability), and was a finalist for the 2008, 2010, and 2012 Bell Prizes. He received the 2019 SIAM Computational Science & Engineering Best Paper Prize, and the 2019 SIAM Geosciences Career Prize. He is a Fellow of the Society for Industrial and Applied Mathematics (SIAM) and serves on the National Academies Committee on Applied and Theoretical Statistics. He is director of the M2dt Center, a DOE ASCR-funded multi-institutional collaboration developing the mathematical foundations for digital twins, and serves as Co-PI for TACC's Frontera HPC system.

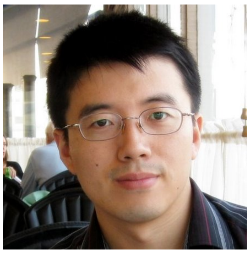
Ghattas's research focuses on advanced mathematical, computational, and statistical theory and algorithms for large-scale inverse and optimization problems governed by models of complex engineered and natural systems. He and his group are developing algorithms to overcome the challenges of Bayesian inverse problems, Bayesian optimal experimental design, and stochastic optimal control & design for large-scale complex systems. To do so, they develop structure-exploiting methods for dimension reduction, surrogates, and neural network approximation, along with high performance computing algorithms. All of these components are integrated and coupled together to form frameworks for digital twins. Driving applications include those in geophysics and climate science (ice sheet dynamics, ice-ocean interaction, seismology, subsurface flows, poroelasticity, tsunamis), advanced materials and manufacturing processes (metamaterials, nanomaterials, additive manufacturing), and gravitational wave inference.

Wednesday 12:00–1:00 pm. Wotao Yin

Using Graph Neural Networks to Solve (Mixed Integer) Linear Programs

DAMO Academy, Alibaba Group US

wotao.yin@alibaba-inc.com



This talk will delve into the fascinating interactions between GNNs (graph neural networks) and mathematical optimization. We have recently found that, by defining linear programs on specific graphs, we can use GNNs to determine their feasibility and solve them to arbitrary precision. To extend this surprising result to mixed-integer linear programs, we analyze the limitations of GNNs and show that, by breaking the symmetry of foldable instances, GNNs can determine their feasibility and solve them to arbitrary precision.
These findings not only deepen our understanding of the expressive ability of GNNs, but also open up new avenues for the application of these deep learning models in solving optimization problems.

Speaker Biography. Dr. Yin is a director of the Alibaba DAMO Academy, Decision Intelligence Lab. His lab introduced the MindOpt optimization solver and used machine learning and operations research technologies to improve the computing efficiency of Alibaba Cloud and other businesses. Prior to joining DAMO Academy, he was a professor in the Department of Mathematics at UCLA, where he made many contributions to the theory and applications of distributed computing, optimization algorithms, machine learning, and image processing. He has received the NSF CAREER Award, the Sloan Research Award, the Morningstar Medal in Applied Mathematics, the DAMO Award, and the Egon Balas Award. Since 2018, Clarivate Analyese has been listed him as one of the world's 1% highly cited scholars.

Thursday 8:30–9:30 pm. Laura Albert

Advancing Critical Infrastructure Protection through Optimization

University of Wisconsin-Madison

laura@engr.wisc.edu

Optimization techniques such as linear and integer programming have been widely applied to problems in critical infrastructure protection. This research has led to advances in optimization theory and application, leading to new algorithmic techniques, models, and policy insights. Advancing optimization through societally relevant applications has been a central theme in my academic research career, spanning applications from security to public safety and election resilience. Using stories from my research, I will offer insight into identifying problems worthy of study, overcoming modeling challenges, creating data-driven modeling frameworks, and influencing policy. As an avid science blogger, I will also discuss opportunities for communicating advances in science and engineering research to the public.

Speaker Biography. Laura Albert, Ph.D., is a Professor and the David Gustafson Department Chair of Industrial & Systems Engineering at the University of Wisconsin-Madison. She is also the President of the Institute for Operations Research and the Management Sciences (INFORMS). Professor Albert's research interests are in the field of operations research, with a particular focus on applications in the public sector. She has been awarded many honors for her research, including the American Association for the Advancement of Science (AAAS) Fellow Award, Institute of Industrial and Systems Engineers (IISE) Fellow Award, the INFORMS Impact Prize, a National Science Foundation CAREER award, and a Fulbright Award. She is the author of the blog "Punk Rock Operations Research."

Thursday 11:30 am–12:30 pm. Xiaodi Wu

Hamiltonian-oriented Quantum Algorithm Design and Programming

University of Maryland

xwu@cs.umd.edu

This is an exciting time for quantum computing where early-stage quantum computers become available at your fingertips through clouds. The conventional design of quantum algorithms is centered around the abstraction of quantum circuits and relies on a digital mindset for application design and implementation. While serving as an elegant mathematical interface, circuit-based digital abstraction usually fails to capture the native programmability of quantum devices, and incurs large overheads, which significantly restricts its near-term feasibility where the computing resource is the major limitation. We propose the so-called Hamiltonian-oriented quantum algorithm design which is directly built on the abstraction of continuous-time Hamiltonian evolution. We illustrate that the Hamiltonian-oriented design not only allows more efficient implementation of quantum algorithms but also inspires novel quantum algorithms, especially in optimization and scientific computing, which are hard to perceive in the circuit model. Moreover, we also introduce a programming language called SIMUQ (SIMUlation language for Quantum) that enables easy implementation of Hamiltonian-based quantum applications for domain experts on heterogeneous quantum devices.

Speaker Biography. Xiaodi Wu is an Associate Professor in the Department of Computer Science and a Fellow of the Joint Center for Quantum Information and Computer Science (QuICS) at the University of Maryland, College Park. He is also an Amazon Visiting Academic. Before that, he was an assistant professor at the University of Oregon, a postdoc associate at MIT, and a Simons Research Fellow at UC Berkeley. He received his Ph.D. in computer science from the University of Michigan, Ann Arbor in 2013, and his B.S. degree in mathematics and physics from Tsinghua University in 2008. His research aims to bridge the gap between the theoretical foundation of quantum computing and the limitation of realistic quantum machines. He is a recipient of the Sloan Research Fellowship, NSF CAREER, and AFOSR YIP awards.

Thursday 3:15–4:15 pm. Fatma Kılınç-Karzan

New Perspectives on Deriving Compact Extended Formulations for Optimization Problems with Indicator Variables

Carnegie Mellon University

fkilinc@andrew.cmu.edu

Applications in statistics and data sciences require modeling an inherit sparsity as well as structural relationships among variables. While such considerations are crucial for creating reliable models, the resulting optimization problems are typically nonconvex and often NP-Hard. As a result, there is a growing interest in structured relaxation techniques that provide strong convex relaxations to such problems. Sparsity and structural relations are usually modeled by introducing indicator variables associated with the original continuous variables and enforcing complementarity restrictions between them. We consider a broad class of optimization problems, with convex objective functions of the continuous variables and arbitrary constraints on their associated indicators. By studying the recessive and rounding directions associated with the epigraph forms of these problems and a connection with projection matrices, we establish that the given complementarity constraints can be reduced to fewer and much simpler to handle relations in a lifted space. Consequently, our insights integrated with the generalized disjunctive programming approaches that utilize perspective functions provide an effective way to derive compact, ideal, and conic-representable extended formulations. The size of our extended formulations depends on the "rank" of the function and the presence of sign restrictions. Moreover, our techniques highlight that the complexity of the convex hull characterizations of these conic mixed-binary sets with complementarity restrictions hinges solely on the complexity of the convex hull characterization of a set defined by only the indicator variables. This then paves the way to benefit from the extensive research on convex hull descriptions of binary sets and related sophisticated optimization software. In addition, our results unify and generalize previous results established for special cases, e.g., perspective reformulation for separable functions, non-separable rank-one functions, or low-rank quadratic functions optimized over domains with combinatorial constraints on indicator variables and possibly sign constraints on continuous variables. On the computational side, we illustrate the effectiveness of our approach on sparse structured logistic regression problems.

Speaker Biography.
Fatma Kılınç-Karzan is an Associate Professor of Operations Research at Tepper School of Business, Carnegie Mellon University. She completed her PhD in Industrial and Systems Engineering from Georgia Institute of Technology in 2011. Her research interests are on foundations of convex optimization and structured nonconvex optimization. Her research is supported by NSF, ONR and AFOSR and has received the 2015 INFORMS Optimization Society Prize for Young Researchers, the 2014 INFORMS JFIG Best Paper Award, and an NSF CAREER Award in 2015. She is an associate editor for journals including Mathematical Programming, SIAM J on Optimization, Mathematics of Operations Research, Operations Research, and INFORMS Journal on Computing, and serves on the editorial board of MOS-SIAM book series on Optimization. She is also an elected member of Mathematical Optimization Society Council and has served as an elected member in the Board of Directors of INFORMS Computing Society (2021–2023).

Friday 8:30–9:30 am. Stefanie Jegelka

Machine Learning for Discrete Optimization:
Graph Neural Networks, Generalization under Shifts, and Loss Functions

Massachusetts Institute of Technology

stefje@csail.mit.edu

Graph Neural Networks (GNNs) have become a popular tool for learning algorithmic tasks, in particular related to combinatorial optimization. In this talk, we will focus on the "algorithmic reasoning" task of learning a full algorithm. Instead of competing on empirical benchmarks, we will aim to get a better understanding of the model's behavior and generalization properties, i.e., the performance on hold-out data, which is an important question in learning-supported optimization too. We will try to understand in particular out-of-distribution generalization in widely used message passing GNNs, with an eye on applications in learning for optimization: what may be an appropriate metric for measuring shift in the data? Under what conditions will a GNN generalize to larger graphs? In the last part of the talk, we will take a brief look at objective (loss) functions for learning with discrete objects, beyond GNNs.

This talk is based on joint work with Ching-Yao Chuang, Keyulu Xu, Joshua Robinson, Nikos Karalias, Jingling Li, Mozhi Zhang, Simon S. Du, Ken-ichi Kawarabayashi and Andreas Loukas.

Speaker Biography. Stefanie Jegelka is an Associate Professor in the Department of EECS at MIT. Before joining MIT, she was a postdoctoral researcher at UC Berkeley, and obtained her PhD from ETH Zurich and the Max Planck Institute for Intelligent Systems. Stefanie has received a Sloan Research Fellowship, an NSF CAREER Award, a DARPA Young Faculty Award, the German Pattern Recognition Award, a Best Paper Award at ICML and an invitation as sectional lecturer at the International Congress of Mathematicians. She has co-organized multiple workshops on (discrete) optimization in machine learning and graph representation learning, and has served as an Action Editor at JMLR and a program chair of the International Conference on Machine Learning (ICML) 2022. Her research interests span the theory and practice of algorithmic machine learning, in particular, learning problems that involve combinatorial, algebraic or geometric structure.

Friday 2:00–3:30 pm. Julie Ivy

Learning from Data for Decision Making in Health and Humanitarian Logistics:
Insights and Challenges in a World with Increasing AI

North Carolina State University

jsivy@ncsu.edu

Decision making to satisfy the basic human needs of health, food, and education is complex. We present an overview of two illustrative studies using data to inform decision making in health care delivery associated with sepsis and hunger relief.

In the first study, we integrate electronic health record (EHR) data with clinical expertise to develop a continuous-time Markov decision process model of the natural history of sepsis. We use this model to better understand the stochastic nature of patients' health trajectories and determine the optimal treatment policy to minimize mortality and morbidity. Specifically, the optimal health states for first anti-infective and first fluid are identified. We formulate this as a stopping problem in which the patient leaves the system when he or she receives the first treatment (intervention) and receives a lump sum reward. Our objective is to find the optimal first intervention for health states to minimize expected mortality and morbidity. We explore the effect of the complex trade-offs associated with the intervention costs and patient disposition costs which are subjective and difficult to estimate. Our model captures the natural progression along sepsis trajectory using a clinically defined treatment delayed population. The model translates observations of patient health as defined by vitals and laboratory results recorded during hospitalization in the EHR to capture the complex evolution of sepsis within a patient population. This framework provides key insights into sepsis patients' stochastic trajectories and informs clinical decision making associated with caring for these patients as their health dynamically evolves.

In the second study, in collaboration with our food bank partner in North Carolina, we develop a single-period, weighted multi-criteria optimization model that provides the decision-maker the flexibility to capture their preferences over the three criteria of equity, effectiveness, and efficiency, and explore the resulting trade-offs. Food banks are challenged with juggling multiple criteria such as equity, effectiveness, and efficiency when making distribution decisions. Models that assume predetermined weights on multiple criteria may produce inaccurate results as the preference of food banks over these criteria may vary over time, and as a function of supply and demand. We introduce a novel algorithm to elicit the inherent preference of a food bank by analyzing its actions within a single-period. The algorithm does not require direct interaction with the decision-maker. The non-interactive nature of this algorithm is especially significant for humanitarian organizations such as food banks which lack the resources to interact with modelers on a regular basis. We explore the implications of different decision-maker preferences for the criteria on distribution policies.

Speaker Biography. Julie Simmons Ivy is a professor in the Edward P. Fitts Department of Industrial and Systems Engineering and Fitts Faculty Fellow in Health Systems Engineering. She previously spent several years on the faculty of the Stephen M. Ross School of Business at the University of Michigan. She received her B.S. and Ph.D. in Industrial and Operations Engineering at the University of Michigan. She also received her M.S. in Industrial and Systems Engineering with a focus on Operations Research at Georgia Tech. She is a President of the Health Systems Engineering Alliance (HSEA) Board of Directors. She is an active member of the Institute of Operations Research and Management Science (INFORMS), Dr. Ivy served as the 2007 Chair (President) of the INFORMS Health Applications Society and the 2012 – 13 President for the INFORMS Minority Issues Forum. Her research interests are mathematical modeling of stochastic dynamic systems with an emphasis on statistics and decision analysis as applied to health care, public health, and humanitarian logistics. This research has made an impact on how researchers and practitioners address complex societal issues, such as health disparities, public health preparedness, hunger relief, student performance, and personalized medical decision-making and has been funded by CDC, NSF, Clinton Health Access Initiative, and the UNC Cancer Center.

Dr. Ivy's primary research interests are in the mathematical modeling of stochastic dynamic systems with emphasis on statistics and decision analysis as applied to health care, manufacturing, and service environments. The focus of her research is decision making under conditions of uncertainty with the objective of improving the decision quality. Dr. Ivy's research program seeks to develop novel concepts of maintenance and monitoring policies and associated scientific theories, and apply them specifically to two important application domains: industrial and medical decision making. She has extensive background in stochastic modeling, in particular the application of partially observable Markov decision processes (POMDPs) and Markov decision processes (MDPs). Dr. Ivy's medical decision making research relates to studying the cost-effectiveness of mammography screening, dynamic breast cancer screening policy development, false positive prediction as a function of breast cancer screening policy, the impact of comorbidity on breast cancer patient outcomes modeling birth delivery choice as a function of long term consequences such as pelvic floor dysfunction, patient-centered pharmaceutical inventory management, and public health preparedness. In addition to her research in medical decision making, Dr. Ivy also works in the area of humanitarian logistics particularly as it relates to hunger relief and equitable food distribution. Her research has been funded by the NSF and the Centers for Disease Control.

■ AIMMS/MOPTA Optimization Modeling Competition 2023

The 15th 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 challenging problem regarding the planning of electric vehicle charging stations. 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. For more information about the competition and the full problem description, please see `https://coral.ise.lehigh.edu/ mopta/competition`.

Problem Description

Electric vehicles (EVs) are becoming increasingly popular due to their environmental and economic benefits. They offer a viable alternative to petrol and diesel vehicles, and generate fewer total greenhouse gas emissions than fossil fuel vehicles. They are becoming more affordable, efficient, and powerful, with advances in battery technology and improved charging infrastructure. EVs can now travel longer distances, faster, and with greater efficiency than ever before. The Biden administration has recently introduced an electric vehicle incentive that provides consumers with a tax credit when they purchase or lease a new electric vehicle. This incentive is part of the administration's plan to reduce emissions and combat climate change. It is anticipated that this incentive will encourage more people to switch to electric vehicles, helping to reduce emissions and save money on fuel costs.

While most electric vehicle charging occurs at home, there is an increasing trend of commercial electric vehicle charging stations being built across the country. These charging stations provide a safe and convenient way for electric vehicle owners to charge their vehicles. They are typically located near public places such as shopping malls, hotels, and other popular locations. The construction of these charging stations can help electric vehicle owners maximize their vehicle's capabilities and decrease range anxiety for potential buyers. Hence, it can make electric vehicles more accessible and encourage more people to make the switch to electric vehicles.

Finalists

We are happy that 13 teams from 7 countries around the world registered for the competition. The panel of judges (Xiu Yang and Tommaso Giovannelli from Lehigh University and Gabriela Servidone from AIMMS) selected the following three teams as finalists:

Team "Clean Air Avengers", University of Bern, Switzerland
Nicklas Klein
advised by Norbert Trautmann.

Team "OptiCoffee", Universidad de Los Andes, Colombia
Ariel Rojas and Juan Betancourt
advised by Daniel Yamín

Team "Vultures", Vrije Universiteit Amsterdam, Netherlands
Fergus Hathorn and Iker Olarra Maldonado
advised by Alessandro Zocca.

Each finalist team will give a 30-minute presentation (25 minutes for the talk + 5 minutes for questions) on their work during the 15th AIMMS-MOPTA Optimization Modeling Competition session on Thursday, August 17, 2023, starting at 9:45am in the Perella Auditorium (Rauch 184). The winning team will be announced at the conference banquet on Thursday evening (Prizes: 1st place $1200 + full scholarship, not including travel, for the AIMMS Campus 2024 event; 2nd place $600; 3rd place $300). In addition, the 10 best teams that used AIMMS as the software platform for their submissions will be awarded $100 in prize money.

We thank all the teams for contributing to the competition. This was another successful and positive experience for all participants and MOPTA organizers. Thank you to **AIMMS** for sponsoring the competition!

■ Poster Session and Competition

The poster session and competition will be held on Wednesday, August 16, 2023, from 5:15-6:15pm in the Rauch atrium. A poster presenter needs to bring a hard copy of the poster, which should not exceed 36" x 48" in size. All posters presented in the poster session will enter the poster competition automatically. Up to three posters will be selected as finalists of the competition by a panel of judges composed of Albert S. Berahas from University of Michigan (chair), Sadan Kulturel-Konak from Pennsylvania State University, Berks, and Javier Peña from Carnegie Mellon University. The winner will be announced at the conference banquet. The list of posters and their presenters is below.

- **A new multiobjective heuristic for creating political redistricting plans while maximizing similarity to a previously-used plan**
  Brendan Ruskey, Lehigh University
- **Dual dynamic programming for stochastic programs over an infinite horizon**
  Caleb Ju, Georgia Tech
- **Provable Convergence of Tensor Decomposition-Based Neural Network Training**
  Chenyang Li, New Jersey Institute of Technology
- **Stochastic Methods for Multi-Level Multi-Objective Optimization**
  Griffin Kent, Lehigh University
- **Moving anchor accelerated algorithms for smooth minimax problems**
  James K. Alcala, University of California, Riverside
- **High Probability Sample Complexity Bounds for Adaptive Optimization Methods with Stochastic Oracles**
  Miaolan Xie, Cornell University
- **Multisecant Quasi-Newton methods**
  Mokhwa Lee, Stony Brook University
- **Higher Order Markov Random Fields and Binary Polynomial Optimization**
  Yakun Wang, Lehigh University

■ Instructions for Talks

For Speakers

- We ask all speakers to be familiar with the time and location of their sessions.
- Speakers should arrive at the location of their sessions 10 minutes prior to the scheduled start time.
- Upon arrival you will be met by the chair of the session. Please introduce yourself and provide the chair with a copy of your presentation for uploading onto the computer in the session room.
- Talks in a session with three speakers are 25 minutes long plus 5 minutes for questions and answers. Talks in a session with four speakers are 20 minutes long plus 2–3 minutes for questions and answers. Anyone going over this time will be asked to stop by the chair.

For Chairs

- Please arrive at the appropriate location 10 minutes before the start of the session you will be chairing.
- In case you encounter any issues during the conference, please do not hesitate to seek the help of our conference staff (in each conference room, there will be IT support available to assist with loading presentations).
- Speakers presenting in the session should also arrive at the session room 10 minutes before the start of the session. You should introduce yourself to the speakers. They will provide you with electronic copies of their presentations to be loaded onto the computer in the session room.
- Your main role will be to ensure that the session runs to time. Each speaker in a session with three talks has 25 minutes for presentation followed by 5 minutes of questions and answers. Each speaker in a session with four talks has 20 minutes for presentation followed by 2–3 minutes of questions and answers.
- If a speaker fails to show up 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, your role will involve introducing them to the audience and reminding everyone that all questions and interruptions are to be reserved for the scheduled 5-minute question and answer section following each presentation.
- In the event that a speaker exceeds their allocated time, your role is to politely but firmly stop their presentation and proceed to the question and answer section of the time slot.
- After each talk, thank the speaker, encourage applause, and open the floor to questions.

■ Gold Sponsor

sas

■ Sponsors

AIMMS

GUROBI OPTIMIZATION

INDUSTRIAL AND SYSTEMS ENGINEERING ISE LEHIGH UNIVERSITY

LEHIGH UNIVERSITY

■ Cooperating Societies

siam In cooperation with Society for Industrial and Applied Mathematics

informs

■ Exhibitors

Springer

■ MOPTA 2023 Committee

Tommaso Giovannelli
MOPTA Conference Chair

Xiu Yang
MOPTA Competition Chair

Tamás Terlaky

Frank E. Curtis
SIAM Representative

Lawrence V. Snyder
Co-director of IDISC

Luis Nunes Vicente
ISE Department Chair

Aida Khajavirad

Alberto Lamadrid

Akwum Onwunta

Ted K. Ralphs

Daniel P. Robinson

Karmel S. Shehadeh

Luis F. Zuluaga

Albert S. Berahas

PROGRAM

MOPTA 2023 Program Overview

Wed, August 16, 2023

Time	Event
8:00-8:40am	Registration and Breakfast (Outside of RB 184)
8:40-9:00am	Opening Session (RB 184)
9:00-10:00am	**Plenary Talk #1** Omar Ghattas (RB 184)
10:00-10:15am	Coffee Break
10:15-11:45am	**Parallel Sessions Wed.1** (RB 85, 91, 241, 271)
11:45am-12:00pm	Coffee Break
12:00-1:00pm	**Plenary Talk #2** Wotao Yin (RB 184)
1:00-2:00pm	Lunch Break (RB 291-293)
2:00-3:30pm	**Parallel Sessions Wed.2** (RB 85, 91, 241, 271)
3:30-3:45pm	Coffee Break
3:45-5:15pm	**Parallel Sessions Wed.3** (RB 85, 91, 241, 271)
5:15-6:15pm	Poster Session/Reception (Rauch Atrium)
6:15-6:30pm	Break
6:30-8:30pm	Student Social (Packer House)

Thu, August 17, 2023

Time	Event
8:00-8:30am	Registration and Breakfast (Outside of RB 184)
8:30-9:30am	**Plenary Talk #3** Laura Albert (RB 184)
9:30-9:45am	Coffee Break
9:45-11:15am	**AIMMS-MOPTA Competition** (RB 184) **Parallel Sessions Thu.1** (RB 85, 91, 241, 271)
11:15-11:30am	Coffee Break
11:30am-12:30pm	**Plenary Talk #4** Xiaodi Wu (RB 184)
12:30-1:30pm	Lunch Break (RB 291-293)
1:30-3:00pm	**Parallel Sessions Thu.2** (RB 85, 91, 241, 271)
3:00-3:15pm	Coffee Break
3:15-4:15pm	**Plenary Talk #5** Fatma Kılınç-Karzan (RB 184)
4:15-4:30pm	Coffee Break
4:30-6:00pm	**Parallel Sessions Thu.3** (RB 85, 91, 241, 271)
6:00-6:15pm	Break
6:15-7:15pm	Cocktail Reception (Lamberton Hall)
7:15-9:00pm	Conference Banquet (Lamberton Hall)

Fri, August 18, 2023

Time	Event
8:00-8:30am	Registration and Breakfast (Outside of RB 184)
8:30-9:30am	**Plenary Talk #6** Stefanie Jegelka (RB 184)
9:30-9:45am	Coffee Break
9:45-11:15am	**Parallel Sessions Fri.1** (RB 85, 91, 241, 271)
11:15-11:30am	Coffee Break
11:30am-1:00pm	**Parallel Sessions Fri.2** (RB 85, 91, 241, 271)
1:00-2:00pm	Lunch Break (RB 291-293)
2:00-3:30pm	**Plenary Talk #7** **Schantz ISE-COH-HSE** Julie Ivy (RB 184)
3:30-3:40pm	Closing Session (RB 184)

8:00–8:40am **Registration**

8:40–9:00am **Opening Session**

9:00–10:00am **Plenary talk** (RB 184) *Chair: Xiu Yang*

- **Geometric Neural Surrogates for Optimization Problems Governed by PDEs** *Omar Ghattas*

10:00–10:15am **Coffee Break**

10:15–11:45am **Advances in Large-Scale Nonlinear Optimization I** (RB 85) *Chair: Albert S. Berahas*

1. **Stochastic Algorithms for Continuous Constrained Optimization** *Frank E. Curtis*
2. **Adaptive SQP Methods for Nonlinear Equality Constrained Stochastic Optimization** *Michael O'Neill*
3. **Inexact Bundle Method for Two-Stage Stochastic Programming** *Baoyu Zhou*

10:15–11:45am **Conic Optimization: Theory and Applications** (RB 91) *Chair: Pouya Sampourmahani*

1. **A Faster Interior-Point Method for Sum-of-Squares Optimization** *Shunhua Jiang*
2. **Higher-Order Newton Methods** *Abraar Chaudhry*
3. **A Higher-order Interior Point Method For Semidefinite Optimization Problems Failing Strict Complementarity Condition** *Pouya Sampourmahani*

10:15–11:45am **Bilevel Stochastic Methods for Optimization and Learning** (RB 241) *Chair: Tommaso Giovannelli*

1. **Bilevel optimization with a multi-objective lower-level problem: Risk-neutral and risk-averse formulations** *Griffin Kent*
2. **Solving Some Stochastic Bilevel Optimization Problems is Nearly as Easy as Solving Single-level Problems** *Tianyi Chen*
3. **Inexact bilevel stochastic gradient methods for constrained and unconstrained lower-level problems** *Tommaso Giovannelli*

10:15–11:45am **Robust and Stochastic Optimization for Dynamic Decision Making Under Uncertainty** (RB 271) *Chair: Caleb Ju*

1. **Data-driven Multistage Distributionally Robust Optimization with Nested Distance** *Rui Gao*
2. **Analytical Stochastic Dynamic Programming for Vehicle-to-Grid Fleet Service Provision** *Ningkun Zheng*
3. **First-order Policy Optimization for Robust Markov Decision Process** *Yan Li*

11:45am–12:00pm **Coffee Break**

12:00–1:00pm **Plenary talk** (RB 184) *Chair: Frank E. Curtis*

- **Using Graph Neural Networks to Solve (Mixed Integer) Linear Programs** *Wotao Yin*

1:00–2:00pm **Lunch**

2:00–3:30pm **Advances in Large-Scale Nonlinear Optimization II** (RB 85) *Chair: Albert S. Berahas*

1. **High Probability Complexity Bounds for Adaptive Optimization Methods with Stochastic Oracles** *Miaolan Xie*
2. **A Stochastic-Gradient-based Interior-Point Algorithm for Solving Smooth Bound-Constrained Optimization Problems** *Qi Wang*
3. **Exploiting Negative Curvature in Conjunction with Adaptive Sampling: Theoretical Results and a Practical Algorithm** *Wanping Dong*

2:00–3:30pm **Quantum Computing and Optimization I** (RB 91) *Chairs: Mohammadhossein Mohammadisiahroudi and Zeguan Wu*

1. **BQPhy: NextGen Quantum-based Engineering Software** *Rut Lineswala*
2. **University Quantum Annealing: Algorithms, Hardware, Applications** *Ananth Tenneti*
3. **Potential Energy Supremacy In Quantum Economy** *Hansheng Jiang*

2:00–3:30pm **Optimization for Data Science** (RB 241) *Chairs: Mateo Díaz and Javier Peña*

1. **Learning Dynamics of Overparametrized Linear Networks** *Rene Vidal*
2. **Provably Faster Gradient Descent via Long Steps** *Benjamin David Grimmer*
3. **Any-dimensional equivariant neural networks** *Mateo Díaz*

2:00–3:30pm **Advanced Decomposition and Hybrid Strategies for Optimization under Uncertainty** (RB 271) *Chair: Joshua Pulsipher*

1. **A Lagrangian dual method for two-stage robust optimization with binary uncertainties** *Anirudh Subramanyam*
2. **Optimization Under Uncertainty with Bayesian Hybrid Models** *Kyla Jones*
3. **Worst-Case Optimization Under Uncertainty With Hybrid Models for Robust Process Systems** *Matthew D. Stuber*

3:30–3:45pm **Coffee Break**

3:45–5:15pm **Special Topics in Derivative-Free Optimization** (RB 85) *Chair: Yunsoo Ha*

1. **Consistency And Complexity Of Adaptive Sampling Based Trust-region Optimization** *Yunsoo Ha*
2. **Stochastic optimization in random subspaces: trust-region framework and subspace selection strategies** *Kwassi Joseph Dzahini*
3. **Sequential stochastic blackbox optimization with zeroth order gradient estimators** *Romain Couderc*

3:45–5:15pm **Tensor Optimization** (RB 91) *Chair: Mengqi Hu*

1. **Koopman Optimization** *Mengqi Hu*
2. **Federated Gradient Matching Pursuit** *Jing Qin*
3. **A bilevel optimization approach to inverse mean-field games** *Jiajia Yu*

3:45–5:15pm **First-Order Methods for Convex Optimization** (RB 241) *Chairs: Javier Peña and Mateo Díaz*

1. **Away-Step Frank-Wolfe Method for Minimizing Logarithmically-Homogeneous Barriers** *Renbo Zhao*
2. **On the Geometry and Refined Rate of Primal-Dual Hybrid Gradient for Linear Programming** *Jinwen Yang*
3. **Affine invariant convergence rates of Frank-Wolfe algorithms on polytopes** *Javier Peña*

3:45–5:15pm **Optimization and Applications I** (RB 271) *Chair: Abdullah Konak*

1. **Optimal and equilibrium solutions to the backup server problem** *Vasco F. Alves*
2. **An Efficient Algorithm to the Integrated Shift and Task Scheduling Problem** *G S R MURTHY*
3. **Data Envelopment Analysis to Develop Composite Resilience Indexes for Sustainable Future** *Abdullah Konak*

5:15–6:15pm **Poster Session and Competition** (Rauch Atrium)

- **A new multiobjective heuristic for creating political redistricting plans while maximizing similarity to a previously-used plan** *Brendan Ruskey*
- **Dual dynamic programming for stochastic programs over an infinite horizon** *Caleb Ju*
- **Provable Convergence of Tensor Decomposition-Based Neural Network Training** *Chenyang Li*
- **Stochastic Methods for Multi-Level Multi-Objective Optimization** *Griffin Kent*
- **Moving anchor accelerated algorithms for smooth minimax problems** *James K. Alcala*
- **High Probability Sample Complexity Bounds for Adaptive Optimization Methods with Stochastic Oracles** *Miaolan Xie*
- **Multisecant Quasi-Newton methods** *Mokhwa Lee*
- **Higher Order Markov Random Fields and Binary Polynomial Optimization** *Yakun Wang*

6:15–6:30pm **Break**

6:30–8:30pm **Student Social** (Packer House)

08:00–08:30am **Registration**

08:30–09:30am **Plenary talk** (RB 184) *Chair: Karmel S. Shehadeh*

- **Advancing Critical Infrastructure Protection through Optimization** *Laura Albert*

9:30–9:45am **Coffee Break**

9:45–11:15am **AIMMS-MOPTA Optimization Modeling Competition** (RB 184) *Chair: Xiu Yang*

9:45–11:15am **Recent Advances in Frank-Wolfe-type Methods** (RB 85) *Chair: Xin Jiang*

1. **Projection-free Methods for Solving Saddle Point Problems** *Erfan Yazdandoost Hamedani*
2. **Multiplicative Gradient Method: When and Why It Works** *Renbo Zhao*
3. **A globally convergent difference-of-convex algorithmic framework and application to information theory** *Xin Jiang*

9:45–11:15am **Applications of Optimization to Dynamical Systems** (RB 91) *Chair: Abraar Chaudhry*

1. **Meta-Learning Operators to Optimality from Multi-Task Non-IID Data** *Nikolai Matni*
2. **Automated reachability analysis of neural network dynamical systems** *Mahyar Fazlyab*
3. **Optimal distributed state estimation using convex optimization** *Juncal Arbelaiz*

9:45–11:15am **Learning for Optimization: Performance and Robustness** (RB 241) *Chair: Bartolomeo Stellato*

1. **End-to-End Learning to Warm-Start Fixed Point Optimizers** *Rajiv Sambharya*
2. **Learning for Robust Optimization** *Irina Wang*
3. **Learning Rationality Parameters in Potential Games** *Stefan Clarke*
4. **On the System Efficiency and Backdoor Robustness of Federated Learning** *Hongyi Wang*

9:45–11:15am **Recent Advances in Stochastic Optimization** (RB 271) *Chair: Miaolan Xie*

1. **Balancing Communication and Computation in Gradient Tracking Algorithms for Decentralized Optimization** *Albert S. Berahas*
2. **Stochastic Optimization for Variational Quantum Algorithms** *Matt Menickelly*
3. **Stochastic approximation with decision-dependent distributions: asymptotic normality and optimality** *Mateo Díaz*

11:15–11:30am **Coffee Break**

11:30am–12:30pm **Plenary talk** (RB 184) *Chair: Tamás Terlaky*

- **Hamiltonian-oriented Quantum Algorithm Design and Programming** *Xiaodi Wu*

12:30–1:30pm **Lunch**

1:30–3:00pm **Nonlinear and Stochastic Optimization Algorithms I** (RB 85) *Organizer: Yutong Dai, Chair: Daniel P. Robinson*

1. **A Riemannian ADMM** *Shiqian Ma*
2. **Some Primal-Dual Theory for Subgradient Methods for Strongly Convex Optimization** *Benjamin David Grimmer*
3. **Inexact Proximal-Methods for Constrained Optimization** *Daniel P. Robinson*

1:30–3:00pm **Quantum Computing and Optimization II** (RB 91) *Chair: Mohammadhossein Mohammadisiahroudi*

1. **Graph decomposition techniques for solving CO problems with VQAs** *Rebekah Herrman*
2. **Quantum Hamiltonian Descent** *Jiaqi Leng*
3. **BEINIT: Avoiding Barren Plateaus in Variational Quantum Algorithms** *Ankit Kulshrestha*

1:30–3:00pm **Optimization Algorithms for Adversarial Training** (RB 241) *Chair: Jiancong Xiao*

1. **Uniformly Stable Algorithms for Adversarial Training** *Jiancong Xiao*
2. **Better Representations via Adversarial Training in Pre-Training: A Theoretical Perspective** *Yue Xing*
3. **Secure Distributed Optimization Under Gradient Attacks** *Shuhua Yu*

1:30–3:00pm **Political Redistricting and Multi-Objective Optimization** (RB 271) *Chair: Brendan Ruskey*

1. **Optimization models for fair political redistricting** *Rahul Swamy*
2. **The Tradeoff Between District Compactness and Minority Representation in US Redistricting** *Brendan Ruskey*
3. **Menu Engineering at Assisted Living Facilities** *Sadan Kulturel-Konak*

3:00–3:15pm **Coffee Break**

3:15–4:15pm **Plenary talk** (RB 184) *Chair: Tommaso Giovannelli*

- **New Perspectives on Deriving Compact Extended Formulations for Optimization Problems with Indicator Variables**
  Fatma Kılınç-Karzan

4:15–4:30pm **Coffee Break**

4:30–6:00pm **Nonlinear and Stochastic Optimization Algorithms II** (RB 85) *Organizer: Yutong Dai, Chair: Xiaoyi Qu*

1. **Developments in Quasi-Newton Methods for Numerical Optimization** *Jeb Runnoe*
2. **A Stochastic GDA Method With Backtracking For Solving Nonconvex Strongly Concave Minimax Problems** *Qiushui Xu*
3. **A Sequential Proximal Algorithm for Minimizing a Composite Objective Function with Equality Constraints** *Xiaoyi Qu*

4:30–6:00pm **Quantum Computing and Optimization III** (RB 91) *Chair: Mohammadhossein Mohammadisiahroudi*

1. **New perspectives on quantum interior point methods** *Brandon Augustino*
2. **A Quantum Framework for Topological Data Analysis** *Bernardo Ameneyro*
3. **An Inexact Feasible Quantum Interior Point Method for Linear and Quadratic Optimization** *Zeguan Wu*

4:30–6:00pm **Machine Learning and Discrete Optimization** (RB 241) *Chair: Mohammad Hesam Shaelaie*

1. **Power System Informed Reinforcement Learning for Unit Commitment** *Robert Ferrando*
2. **Scarf's algorithm and stable matchings** *Chengyue He*
3. **Toward a unified framework for branch & bound and reinforcement learning** *Mohammad Hesam Shaelaie*

4:30–6:00pm **Fairness in Optimization** (RB 271) *Chair: Man Yiu Tsang*

1. **A Unified Framework for Analyzing and Optimizing a Class of Convex Inequity Measures** *Man Yiu Tsang*
2. **Bridging Two Fairness Perspectives: Group Parity Metrics and Social Welfare Functions** *Violet (Xinying) Chen*
3. **Equity-promoting Integer Programming Approaches For Medical Resident Rotation Scheduling** *Shutian Li*

6:00–6:15pm **Break**

6:15–7:15pm **Cocktail Reception**

7:15–9:00pm **Conference Banquet**

8:00–8:30am **Registration**

8:30–9:30am **Plenary talk** (RB 184) *Chair: Ted K. Ralphs*

- **Machine Learning for Discrete Optimization: Graph Neural Networks, Generalization under Shifts, and Loss Functions** *Stefanie Jegelka*

9:30–9:45am **Coffee Break**

9:45–11:15am **Bilevel Optimization** (RB 85) *Chair: Federico Battista*

1. **Generating Improving Solutions/Directions for Mixed Integer Bilevel Linear Problems** *Federico Battista*
2. **Robust solutions to mixed integer bilevel (two-stage) linear optimization with bounded rationality** *Yu Xie*
3. **On the stability of a bilevel optimisation program with a convex lower level problem** *Emanuele Pizzari*

9:45–11:15am **Advances in Mixed-Integer Programming and Max-Cut Problems** (RB 91) *Chair: Nimita Shinde*

1. **Warm Starting Series of Mixed Integer Linear Programs with Fixed Dimensions via Disjunctive Cuts** *Sean Kelley*
2. **An SU(2)-symmetric Semidefinite Programming Hierarchy for Quantum MaxCut** *Cunlu Zhou*
3. **Memory-efficient approximation algorithm for Max-Cut, Max-k-Cut and Correlation Clustering** *Nimita Shinde*

9:45–11:15am **Advances in Nonconvex Optimization** (RB 241) *Chair: Wei Liu*

1. **Gaussian smoothing gradient descent for high-dimensional nonconvex optimization** *Andrew Starnes*
2. **Multisecant Quasi-Newton methods** *Mokhwa Lee*
3. **Lower Complexity Bound of First-order Methods for Affinely Constrained Nonconvex Nonsmooth Problems** *Wei Liu*

9:45–11:15am **Optimization and Applications II** (RB 271) *Chair: Tuyen Tran*

1. **Real-Time Personalized Order Holding** *Mohammad Reza Aminian*
2. **Evaluation of Manual Assembly Line Balancing Problem in Flexible Manufacturing and Uncertain Operator Performance** *Liam J. Cahalane*
3. **Solving multifacility Location Problems Based on Mixed Integer Programming** *Tuyen Tran*

11:15–11:30am **Coffee Break**

11:30am–1:00pm **Advances in Constrained Composite Optimization** (RB 85) *Chair: Dimitri Papadimitriou*

1. **A single-loop Lagrangian method for nonconvex optimization with nonlinear constraints** *Dimitri Papadimitriou*
2. **A primal-dual splitting method for nonlinear composite problems** *Bang Vu*
3. **Inexact proximal augmented Lagrangian method for compositional problems with convex constraints** *Yangyang Xu*

11:30am–1:00pm **Healthcare Optimization** (RB 91) *Chair: Mohammadhossein Mohammadisiahroudi*

1. **Treatment Planning Optimization for Proton Therapy** *Mohammadhossein Mohammadisiahroudi*
2. **Supervised Inverse Optimization** *Felix Parker*
3. **BiomedGPT: A Unified and Generalist Biomedical Generative Pre-trained Transformer for Vision, Language, and Multimodal Tasks** *Jun Yu*

11:30am–1:00pm **Advances in Continuous Optimization for Learning Problems** (RB 241) *Chair: César A. Uribe*

1. **Applications of DC Programming to bilevel hierarchical clustering problem** *Tuyen Tran*
2. **On the Performance of Gradient Tracking with Local Updates** *Edward Duc Hien Nguyen*
3. **On First-Order Meta-Reinforcement Learning with Moreau Envelopes** *César A. Uribe*

11:30am–1:00pm **Optimization and Applications III** (RB 271) *Chair: Billy Jin*

1. **Probabilistic Submodularity of Maximizing Anticoordination in Finite Network Games** *Soham Das*
2. **Stochastic Programming Inspired Modeling Techniques for Shaping Dynamic Trajectories** *Joshua Pulsipher*
3. **Optimal Tradeoffs for Two-Stage Bipartite Matching with Advice** *Billy Jin*

1:00–2:00pm **Lunch**

2:00–3:30pm **Plenary talk/Schantz ISE-COH-HSE** (RB 184) *Chairs: Luis Nunes Vicente and Ana I. Alexandrescu*

- **Learning from Data for Decision Making in Health and Humanitarian Logistics: Insights and Challenges in a World with Increasing AI** *Julie Ivy*

3:30–3:40pm **Closing Session**

ABSTRACTS

■ Advances in Large-Scale Nonlinear Optimization I

Room: **RB 85** (10:15–11:45am) Chair: *Albert S. Berahas*

1. Stochastic Algorithms for Continuous Constrained Optimization

Frank E. Curtis[1,*]

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

I will discuss the interesting features that drive the convergence guarantees for a set of adaptive stochastic algorithms that my collaborators and I have proposed for solving nonlinearly constrained optimization problems. These algorithms are of the sequential quadratic optimization and interior-point varieties, and they operate in the fully stochastic regime in which we prove convergence-in-expectation and almost-sure-convergence guarantees.

2. Adaptive SQP Methods for Nonlinear Equality Constrained Stochastic Optimization

Michael O'Neill[1,*]

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

We develop a Sequential Quadratic Programming algorithm forminimizing a stochastic objective function subject to deterministic equalityconstraints. The method uses an adaptive stepsize rule that is not dependenton a pre-specified sequence and a complexity result is derived. In addition,new procedures are developed for estimating the merit parameter, which aremore robust to the stochasticity than those proposed in prior work. Numericalvalidation of the proposed algorithms is also presented.

3. Inexact Bundle Method for Two-Stage Stochastic Programming

Baoyu Zhou[1,*], *Haihao Lu*[1], *John R. Birge*[1]

[1]University of Chicago; *zbaoyu@umich.edu;

We propose an inexact bundle method for solving two-stage stochastic programming problems that arise from important application areas including revenue management and power system. We consider the setting that it is intractable to compute true objective function and gradient information, and instead only estimates of objective function and gradient values are available. Under common assumptions, the algorithm generates a sequence of iterates converging to a neighborhood of optimality, where the radius of neighborhood depends on the level of inexactness from stochastic objective function estimates. The numerical results demonstrate empirical performance of our proposed algorithm.

■ Conic Optimization: Theory and Applications

Room: **RB 91** (10:15–11:45am) Chair: *Pouya Sampourmahani*

1. A Faster Interior-Point Method for Sum-of-Squares Optimization

Shunhua Jiang[1,*], *Bento Natura*[2], *Omri Weinstein*[3]

[1]Columbia University; *sj3005@columbia.edu; [2]ICERM and Georgia Tech ; [3]The Hebrew University and Columbia University;

Sum-of-squares (SOS) optimization is a conic optimization problem that optimizes a linear objective while restricting the polynomials to be sum of squares. It is a central tool in polynomial optimization and capture convex programming in the Lasserre hierarchy. In this talk I will present a recent result in ICALP '22 that shows a polynomially faster interior-point method for sum-of-squares optimization. The centerpiece of our algorithm is a dynamic data structure for maintaining the inverse of the Hessian of the SOS barrier function under the polynomial interpolant basis.

2. Higher-Order Newton Methods

Abraar Chaudhry[1,*], *Amir Ali Ahmadi*[1], *Jeff Zhang*[2]

[1]Princeton University; *azc@princeton.edu; [2]Carnegie Mellon University;

We present generalizations of Newton's method that incorporate derivatives of arbitrarily high order, but maintain a polynomial dependence on dimension in their cost per iteration. At each step, our algorithms use semidefinite programming to construct and minimize a convex approximation to the Taylor expansion of the function we wish to minimize. We analyze the convergence rates of our higher-order Newton methods and compare their basins of attraction around local minima to those of the classical Newton method.

3. A Higher-order Interior Point Method For Semidefinite Optimization Problems Failing Strict Complementarity Condition

Pouya Sampourmahani[1,*], *Ali Mohammad-Nezhad*[2], *Tamás Terlaky*[1]

[1]Lehigh University; *pos220@lehigh.edu; [2]University of North Carolina at Chapel Hill;

Semidefinite optimization (SDO) problems are known to be solved efficiently using interior point methods (IPMs), however superlinear convergence is achieved only under the strict complementarity condition. Failing strict complementarity leads to losing analyticity of the central path at optimality. Based on a semi-algebraic description of the central path, we reparametrize the central path and then propose an IPM that exploits higher-order derivatives of the parametrized central path. We discuss the local convergence of our higher-order IPM and then present the numerical results.

■ Bilevel Stochastic Methods for Optimization and Learning

Room: **RB 241** (10:15–11:45am) Chair: *Tommaso Giovannelli*

1. Bilevel optimization with a multi-objective lower-level problem: Risk-neutral and risk-averse formulations

Griffin Kent[1,*], *Tommaso Giovannelli*[1], *Luis Nunes Vicente*[1]

[1]Lehigh University; *gdk220@lehigh.edu;

In this work, we propose different formulations and gradient-based algorithms for deterministic and stochastic bilevel problems with conflicting objectives in the lower level. Such problems have received little attention in the deterministic case and have never been studied from a stochastic approximation viewpoint despite the recent advances in stochastic methods for single-level, bilevel, and multi-objective optimization.

To solve bilevel problems with a multi-objective lower level, different approaches can be considered depending on the interpretation of the lower-level optimality. An optimistic formulation that was previously introduced for the deterministic case consists of minimizing the upper-level function over all non-dominated lower-level solutions. We develop new risk-neutral and risk-averse formulations, address their main computational challenges, and develop the corresponding deterministic and stochastic gradient-based algorithms.

2. Solving Some Stochastic Bilevel Optimization Problems is Nearly as Easy as Solving Single-level Problems

Tianyi Chen[1,*]

[1] Rensselaer Polytechnic Institute; *chent18@rpi.edu;

Stochastic bilevel optimization, including stochastic compositional and min-max optimization as special cases, is gaining popularity in many machine learning applications. While the three problems share a nested structure, existing works often treat them separately, thus developing problem-specific algorithms and analyses. Among various exciting developments, simple SGD-type updates (potentially on multiple variables) are still prevalent in solving this class of nested problems. Still, they are believed to have a slower convergence rate than single-level problems. This talk unifies several (projected) SGD-type updates for stochastic bilevel problems into a single SGD approach that we term the Alternating Implicit Stochastic gradient dEscenT (AISET) method. By leveraging the smoothness of the problem, this talk reveals that for a growing class of stochastic bilevel optimization problems, AISET matches SGD for solving single-level nonconvex optimization problems in terms of sample complexity.

3. Inexact bilevel stochastic gradient methods for constrained and unconstrained lower-level problems

Tommaso Giovannelli[1,*], *Griffin Kent*[1], *Luis Nunes Vicente*[1]

[1] Lehigh University; *tog220@lehigh.edu;

Two-level stochastic optimization formulations have become instrumental in a number of machine learning contexts such as continual learning, neural architecture search, adversarial learning, and hyperparameter tuning. Practical stochastic bilevel optimization problems become challenging in optimization or learning scenarios where the number of variables is high or there are constraints.

In this work, we introduce a bilevel stochastic gradient method for bilevel problems with lower level constraints. We also present a comprehensive convergence theory that covers all inexact calculations of the adjoint gradient (also called hypergradient) and addresses both the lower-level unconstrained and constrained cases. To promote the use of bilevel optimization in large-scale learning, we introduce a practical bilevel stochastic gradient method (BSG-1) that does not require second-order derivatives and, in the lower-level unconstrained case, dismisses any system solves and matrix-vector products.

■ Robust and Stochastic Optimization for Dynamic Decision Making Under Uncertainty

Room: **RB 271** (10:15–11:45am) Chair: *Caleb Ju*

1. Data-driven Multistage Distributionally Robust Optimization with Nested Distance

Rui Gao[1,*]

[1] University of Texas at Austin;
*rui.gao@mccombs.utexas.edu;

We study multistage distributionally robust optimization in which the uncertainty set consists of stochastic processes that are close to a scenario tree in the nested distance. Compared to other choices such as Wasserstein distance between stochastic processes, the nested distance accounts for information evolution, making it hedge against a plausible family of data processes. Due to the non-convexity of the nested distance uncertainty set, the resulting minimax problem is notoriously difficult to solve. In spite of this challenge, in this paper, we develop an equivalent robust dynamic programming reformulation. This reformulation has two important implications: (1) Modeling-wise, it unveils that the considered single minimax multistage-static formulation based on the nested distance for stochastic processes is equivalent to a nested minimax multistage-dynamic formulation based on one-period nested Wasserstein distance, thus both of which admit a time-consistent robust optimal policy. (2) Computation-wise, we identify conditions under which the robust Bellman recursion can be interpreted as norm-regularized sample average approximation and solved via tractable convex programs. We develop a stochastic dual dynamic programming algorithm and apply it to dynamic portfolio selection. Numerical experiments demonstrate the superior out-of-sample performance of our robust approach.

2. Analytical Stochastic Dynamic Programming for Vehicle-to-Grid Fleet Service Provision

Ningkun Zheng[1,*], *Joshua Jaworski*[1], *Yousuf Baker*[1], *Xiaoxiang Liu*[1], *Bolun Xu*[1]

[1] Columbia University; *nz2343@columbia.edu;

The surging adoption of electric vehicles (EV) calls for accurate and efficient approaches to coordinate with the power grid operation. Hence, it becomes imperative to integrate accurate battery models and electricity price uncertainties while ensuring the computational tractability of control algorithms. To this end, we propose an analytical method employing Stochastic Dynamic Programming for EV fleet vehicle-to-grid (V2G) charging control. Using the proposed algorithm, we investigate the economic benefit of V2G using real-time price data from New York State and a real-world charging network dataset. Our result shows the importance of using more accurate nonlinear battery models in V2G controllers and evaluating the cost of price uncertainties over V2G. Furthermore, the proposed algorithm can be effectively amalgamated with a learning-based value function prediction model to better manage price uncertainty.

3. First-order Policy Optimization for Robust Markov Decision Process

Yan Li[1,*]

[1] Georgia Institute of Technology; *yli939@gatech.edu;

We consider the problem of solving robust Markov decision process (MDP), which involves a set of discounted, finite state, finite action space MDPs with uncertain transition kernels. The goal of planning is to find a robust policy that optimizes the worst-case values against the transition uncertainties, and thus encompasses the standard MDP planning as a special case. For (s, a)-rectangular uncertainty sets, we develop a policy-based first-order method, namely the robust policy mirror descent (RPMD), and establish an $O(log(1/\epsilon))$ iteration complexity for finding an epsilon-optimal policy. We further develop a stochastic variant of the robust policy mirror descent method, named SRPMD, when the first-order information is only available through online interactions with the nominal environment. For general Bregman divergences, we establish an $O(1/\epsilon^2)$ sample complexity. The prior convergence of RPMD is applicable to any Bregman divergence, provided the policy space has a bounded radius measured by the divergence when centering at the initial policy. These aforementioned results appear to be new for policy-based first-order methods applied to the robust MDP problem.

■ Advances in Large-Scale Nonlinear Optimization II

Room: **RB 85** (2:00–3:30pm) Chair: *Albert S. Berahas*

1. High Probability Complexity Bounds for Adaptive Optimization Methods with Stochastic Oracles

Miaolan Xie[1,], Billy Jin[1], Katya Scheinberg[1]*

[1]Cornell University; [*]mx229@cornell.edu;

We consider a simple adaptive optimization framework for continuous optimization where the (explicit or implicit) step size in each iteration is adaptively adjusted by the estimated progress of the algorithm instead of requiring manual tuning or using a pre-specified sequence of step sizes. The framework accommodates a stochastic setting where function value, gradient (and possibly Hessian) estimates are available only through noisy probabilistic oracles (can be biased and possibly arbitrarily bad with some constant probability). This framework is very general and encompasses stochastic variants of line search, quasi-Newton, cubic regularized Newton methods for unconstrained problems, and stochastic SQP methods for constrained problems. The probabilistic oracles capture multiple standard settings including expected loss minimization in machine learning, zeroth-order (derivative-free), and low-precision optimization. Under reasonable conditions on the oracles, we derive high probability bounds on the sample (and iteration) complexity of the algorithms.

2. A Stochastic-Gradient-based Interior-Point Algorithm for Solving Smooth Bound-Constrained Optimization Problems

Qi Wang[1,], Frank E. Curtis[1], Vyacheslav Kungurtsev[2], Daniel P. Robinson[1]*

[1]Lehigh University; [*]qiw420@lehigh.edu; [2]Czech Technical University ;

A stochastic-gradient-based interior-point algorithm for minimizing a continuously differentiable objective function (that may be nonconvex) subject to bound constraints is presented, analyzed, and demonstrated through experimental results. The algorithm is unique from other interior-point methods for solving smooth (nonconvex) optimization problems since the search directions are computed using stochastic gradient estimates. It is also unique in its use of inner neighborhoods of the feasible region—defined by a positive and vanishing neighborhood-parameter sequence—in which the iterates are forced to remain. It is shown that with a careful balance between the barrier, step-size, and neighborhood sequences, the proposed algorithm satisfies convergence guarantees in both deterministic and stochastic settings. The results of numerical experiments show that in both settings the algorithm can outperform a projected-(stochastic)-gradient method.

3. Exploiting Negative Curvature in Conjunction with Adaptive Sampling: Theoretical Results and a Practical Algorithm

Wanping Dong[1,], Albert S. Berahas[1], Raghu Bollapragada[2]*

[1]University of Michigan; [*]wanpingd@umich.edu; [2]University of Texas at Austin;

With the emergence of deep learning in recent years, nonconvex optimization has gained more interest and focus, but its nonconvex properties bring us great challenges in algorithm design. Our goal is to develop algorithms that have second-order convergence and affordable complexity so that they can be used for large-scale problems. Therefore, we'll exploit negative curvature with adaptive sampling strategy. First, we provide the convergence result for the deterministic setting where the gradients and Hessians are inexact. In the second part, I want to show our result in the stochastic setting where the gradients are constructed by random samples. Except for the theory part, I'll also talk about our practical algorithm which is the Newton-CG method with negative curvature detection and adaptive sampling. Our experiment results on different nonconvex problems are shown after that.

■ Quantum Computing and Optimization I

Room: **RB 91** (2:00–3:30pm) Chair: *M. Mohammadisiahroudi, Z. Wu*

1. BQPhy: NextGen Quantum-based Engineering Software

Rut Lineswala[1,]*

[1]BosonQ Psi; [*]rutlineswala@bosonqpsi.com;

Quantum Computation has the potential to provide significant computational advantages over traditional computing, and harnessing its power can have lasting effects on industries. BosonQ Psi aims to do just that by delivering Quantum solutions targeted to assist complex Engineering industries in product design and optimization.

With their flagship product, BQPhy, enterprises can readily perform Design Optimization and Engineering Simulations using Quantum-Integrated methods to provide a great advantage. Design Optimization helps engineers and designers systematically explore and refine design alternatives to achieve optimal solutions based on predefined objectives and constraints. By leveraging novel Quantum algorithms, BQPhy Design Optimization enables users to streamline the design process to drive innovation, enhance product performance, and reduce costs, ultimately poised to shape the future of product design across various industries. Engineering simulation products play a crucial role in designing and analyzing complex systems across different engineering disciplines, challenged with huge costs, time, and physical limitations. BQPhy Engineering Simulations employ the novel HQCFE framework in its engineering simulation modules to perform complex system-level Multiphysics problems to minimize the need for physical prototypes, reduce physical testing costs, speed up the design process, and enhance product reliability and efficiency.

Along with BQPhy, BosonQ Psi has also developed Airline Route Optimization (ARO) application. ARO is critical for maximizing operational efficiency and profitability in the aviation industry. Leveraging advanced Quantum algorithms and considering factors such as flight schedules, aircraft performance, and operational constraints, BQP ARO aims to help airlines make informed decisions to optimize their route planning and resource allocation.

2. Quantum Annealing: Algorithms, Hardware, Applications

Ananth Tenneti[1,*], *Sridhar Tayur*[1]

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

Quantum annealing has emerged as a promising approach to solving difficult combinatorial optimization algorithms. A quantum annealer is a computational device which solves Quadratic Unconstrained Binary Optimization (QUBO) problems framed as an Ising Hamiltonian. In this talk, we report on progress made at Quantum Technologies Group at Carnegie Mellon University (in collaboration with several other institutions).

We first describe a hybrid quantum-classical optimization algorithm, the Graver Augmented Multi-seed algorithm (GAMA), for solving binary constrained, non-linear optimization problems. The GAMA algorithm utilizes augmentation along graver basis elements starting from multiple initial feasible solutions to find an optimal solution. We show its performance on Binary Classification Problem (X-ray images for pneumonia), the Cardinality Boolean Quadratic Problems (CBQP), Quadratic Semi-Assignment Problems (QSAP) and Quadratic Assignment Problems (QAP). We also discuss the performance of annealers based on two photonic ising machines, the temporal multiplexed ising machine (TMIM) and the spatial photonic ising machine (SPIM) on the Max-cut and NPP problems, comparing them to D-Wave and Gurobi.

3. Potential Energy Supremacy In Quantum Economy

Hansheng Jiang[1,*]

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

Energy cost is crucial in the modern computing industry, like large-language models. Governments prefer low energy consumption for computing providers' market development. This paper explores the potential energy advantages of quantum computing over classical counterparts without considering computational complexity. By employing game-theoretical models with energy constraints for quantum and classical companies, we demonstrate that quantum firms may achieve higher profits within given energy limits and be more energy-efficient in the Nash equilibrium, regardless of computational efficiency assumptions for quantum and classical algorithms. Consequently, quantum computing could offer a more sustainable approach to the computing industry.

■ Optimization for Data Science

Room: **RB 241** (2:00–3:30pm) Chair: *Mateo Díaz and Javier Peña*

1. Learning Dynamics of Overparametrized Linear Networks

Rene Vidal[1,*], *Hancheng Min*[2], *Salma Tarmoun*[2], *Ziqing Xu*[2], *Enrique Mallada*[2]

[1]University of Pennsylvania;
[*]vidalr@seas.upenn.edu;[2]Johns Hopkins University;

Deep networks have led to significant improvements in the performance of AI systems. However, the mathematical reasons for this success remain elusive. For example, contrary to the common belief that overparameterization may hurt generalization and optimization, recent work suggests that overparameterization may bias the optimization algorithm towards solutions that generalize well — a phenomenon known as implicit regularization or implicit bias — and may also accelerate convergence — a phenomenon known as implicit acceleration. This lecture will study both phenomena through the lens of dynamical systems. In particular, we will provide a detailed analysis of the dynamics of gradient flow and gradient descent for overparametrized linear models showing that convergence to equilibrium depends on the imbalance between input and output weights (which is fixed at initialization) and the margin of the initial solution. The talk will also provide an analysis of the implicit bias, showing that large hidden layer width, together with (properly scaled) random initialization, constrains the network parameters to converge to a solution which is close to the min-norm solution.

2. Provably Faster Gradient Descent via Long Steps

Benjamin David Grimmer[1,*]

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

This work establishes provably faster convergence rates for gradient descent via a computer-generated analysis technique. Our theory allows nonconstant stepsize policies with frequent long steps potentially violating descent by analyzing the overall effect of many iterations at once rather than the typical one-iteration inductions used in most first-order method analyses. We show long steps, which may increase the objective value in the short term, lead to provably faster convergence in the long term. A conjecture towards proving a faster $O(1/T \log T)$ rate for gradient descent is also motivated along with simple numerical validation.

3. Any-dimensional equivariant neural networks

Mateo Díaz[1,*], *Eitan Levin*[2]

[1]Johns Hopkins University; [*]mateodd@jhu.edu; [2]Caltech;

Traditional supervised learning aims to learn an unknown mapping by fitting a function to a set of input-output pairs with a fixed dimension. The fitted function is then defined on inputs of the same dimension. However, in many settings, the unknown mapping takes inputs in any dimension; examples include graph parameters defined on graphs of any size and physics quantities defined on an arbitrary number of particles. We leverage a newly-discovered phenomenon in algebraic topology, called representation stability, to define equivariant neural networks that can be trained with data in a fixed dimension and then extended to accept inputs in any dimension. Our approach is user-friendly, requiring only the network architecture and the groups for equivariance, and can be combined with any training procedure. We provide a simple open-source implementation of our methods and offer preliminary numerical experiments.

■ Advanced Decomposition and Hybrid Strategies for Optimization under Uncertainty

Room: **RB 271** (2:00–3:30pm) Chair: *Joshua Pulsipher*

1. A Lagrangian dual method for two-stage robust optimization with binary uncertainties

Anirudh Subramanyam[1,*]

[1]The Pennsylvania State University; [*]subramanyam@psu.edu;

This paper presents a new exact method to calculate worst-case parameter realizations in two-stage robust optimization problems with categorical or binary-valued uncertain data. Traditional exact algorithms for these problems, notably Benders decomposition and column-and-constraint generation, compute worst-case parameter realizations by solving mixed-integer bilinear optimization subproblems. However, their numerical solution can be computationally expensive not only due to their resulting large size after reformulating the bilinear terms, but also because decision-independent bounds on their variables are typically unknown. We propose an alternative Lagrangian dual method that circumvents these difficulties and is readily integrated in either algorithm. We specialize the method to problems where the binary parameters switch on or off constraints as these are commonly encountered in applications, and discuss extensions to problems that lack relatively complete recourse and to those with integer recourse. Numerical experiments provide evidence of significant computational improvements over existing methods.

2. Optimization Under Uncertainty with Bayesian Hybrid Models

Kyla Jones[1,*], *Alexander Dowling*[1]

[1]University of Notre Dame; [*]kjones29@nd.edu;

In chemical engineering, physicochemical models utilize experimental data to make system predictions where data collection is cumbersome. The standard practice is to construct these models from detailed system knowledge, such as reaction kinetics, thermodynamics, and transport phenomena. However, computational tractability constraints demand simplifications and approximations that induce uncertainty in the model form. Furthermore, when the underlying phenomena are not known a priori, inadequate (i.e., incorrect) models can result in predictions that systematically differ from reality. Thus, uncertainty quantification (UQ) is essential for providing reliable computer model predictions. What is not known is how to synergize physicochemical knowledge with the sound statistical inference of process data to make engineering decisions robust to real-world uncertainty.
This work builds upon the seminal Kennedy and O'Hagan framework for Bayesian calibration of computer models for optimization under uncertainty. Our objective is to quantify the uncertainty of the system by adding a nonparametric correction term to account for systematic bias in optimization problems. We model the correction term as a Gaussian process and demonstrate probabilistic approaches to parameter estimation, including maximum likelihood estimation and Bayesian approaches. Our numerical results demonstrate the effectiveness of these techniques.

3. Worst-Case Optimization Under Uncertainty With Hybrid Models for Robust Process Systems

Matthew D. Stuber[1,*], *Chenyu Wang*[1], *Matthew E. Wilhelm*[1], *Robert X. Gottlieb*[1], *Pengfei Xu*[1], *Dimitri Alston*[1]

[1]University of Connecticut ; [*]matthew.stuber@uconn.edu;

Pessimistic "worst-case" model-based design is necessary to ensure that uncertainty (epistemic and aleatoric) is rigorously accounted for prior to the construction and deployment of complex performance/safety-critical process systems. Deterministic global optimization (DGO) is required to solve such problems as systems models are most often nonlinear. However, since such problems are NP-hard, high complexity models pose a barrier to current state-of-the-art DGO methods and software. Machine learning (ML) models have the potential to significantly reduce this complexity, but by themselves, they lack interpretability, extrapolability, and any guaranteed satisfaction of the physical laws. In turn, hybrid models that combine first-principles mechanistic models with ML data-driven models overcome these limitations and have the potential to improve model prediction accuracy and simultaneously reduce overall model complexity over pure first-principles models. Furthermore, hybrid modeling architectures can also leverage special features of ML models to overcome numerical challenges often associated with complex first-principles models (e.g., function domain violations).
In this talk, we will discuss some of our recent advances in solving worst-case design formulations as semi-infinite programs (SIPs) with hybrid models embedded. Namely, we will formalize this approach for general hybrid modeling architectures that involve complex first-principles models and a variety of ML modeling architectures and approaches that include both regression and classification. Within this context, we will also discuss our state-of-the-art, readily extensible DGO software package "EAGO" and showcase its capabilities and performance on relevant case studies.

■ Special Topics in Derivative-Free Optimization

Room: **RB 85** (3:45–5:15pm) Chair: *Yunsoo Ha*

1. Consistency And Complexity Of Adaptive Sampling Based Trust-region Optimization

Yunsoo Ha[1,*], *Sara Shashaani*[1]

[1]North Carolina State University; [*]yha3@ncsu.edu;

We present complexity results of a class of stochastic nonconvex optimization problems with trust-region methods and adaptive sampling schemes devised for zeroth order (ASTRO-DF) and first order (ASTRO) stochastic oracles. We present two metrics for measuring efficiency; the classical iteration complexity and work or sampling complexity. The latter is more appropriate for adaptive sampling procedures where the total number of queries to the stochastic oracle varies per iteration.

2. Stochastic optimization in random subspaces: trust-region framework and subspace selection strategies

Kwassi Joseph Dzahini[1,*], *Stefan M. Wild*[2]

[1]Argonne National Lab; [*]kdzahini@anl.gov; [2]Lawrence Berkeley National Laboratory;

This work proposes a framework for large-scale stochastic derivative-free optimization (DFO) by introducing STARS, a trust-region method which achieves scalability using random models in random low-dimensional affine subspaces. STARS significantly reduces per-iteration costs in terms of function evaluations, thus yielding strong performance on large-scale stochastic DFO problems. The user-determined dimension of these subspaces can be chosen independently of the dimension of the problem via Johnson–Lindenstrauss transforms such as hashing-like matrices or those obtained using Haar-distributed orthogonal random matrices. For convergence purposes, both a particular quality of the subspace and the accuracies of random function estimates and models are required to hold with sufficiently high, but fixed, probabilities. Convergence and expected complexity results of STARS are obtained using martingale theory.

3. Sequential stochastic blackbox optimization with zeroth order gradient estimators

Romain Couderc[1,*], *Charles Audet*[2], *Jean Bigeon*[3], *Michael Kokkolaras*[4]

[1]GSCOP, Grenoble-Alpes University, GERAD, Ecole Polytechnique de Montreal; [*]romain.couderc@polymtl.ca; [2]GERAD, Ecole Polytechnique de Montreal; [3]Nantes University, École Centrale Nantes; [4]GERAD, McGill University;

This work considers stochastic optimization problems in which the objective function values can only be computed by a blackbox corrupted by some random noise following an unknown distribution. The proposed method is based on sequential stochastic optimization (SSO): the original problem is decomposed into a sequence of subproblems. Each of these subproblems is solved using a zeroth order version of a sign stochastic gradient descent with momentum algorithm (ZO-Signum) and with an increasingly fine precision. This decomposition allows a good exploration of the space while maintaining the efficiency of the algorithm once it gets close to the solution. Under Lipschitz continuity assumption on the blackbox, a convergence rate in expectation is derived for the ZO-signum algorithm. Moreover, if the blackbox is smooth and convex or locally convex around its minima, a convergence rate to an epsilon-optimal point of the problem may be obtained for the SSO algorithm. Numerical experiments are conducted to compare the SSO algorithm with other state-of-the-art algorithms and to demonstrate its competitiveness.

■ Tensor Optimization

Room: **RB 91** (3:45–5:15pm) Chair: *Mengqi Hu*

1. Koopman Optimization

Mengqi Hu[1,*], *Bian Li*[2], *Yifei Lou*[3], *Xiu Yang*[2]

[1]University of North Carolina at Chapel Hill; [*]humengqi@unc.edu; [2]Lehigh University; [3]University of Texas at Dallas;

In this paper, we propose to solve an optimization problem by reformulating it into a dynamical system, followed by an adaptive spectral Koopman (ASK) method. Specifically, our solution is built upon the Koopman operator that can be used to approximate the evolution of an ordinary differential equation (ODE) by a finite number of eigenfunctions and eigenvalues.
We demonstrate its applicability and accuracy on a broad spectrum of classical optimization testing functions, evidenced by consistently smaller gradient norms. The method maintains competitive computational efficiency, and works particularly well when the corresponding dynamical system can be accurately approximated by its local dynamics.
The method's performance combined with the promising theoretical guarantee indicates ASK's vast potential in solving various optimization problems.
This exploration paves the way for future endeavors in optimization theory and practice.

2. Federated Gradient Matching Pursuit

Jing Qin[1,*], *Halyun Jeong*[2], *Deanna Needell*[2]

[1]University of Kentucky; [*]jing.qin@uky.edu; [2]University of California, Los Angeles;

Due to the rapid development of communication technologies and growth of decentralized data on clients, collaborative machine learning has become the main interest while providing privacy-preserving frameworks. In particular, federated learning provides such a solution to learn a shared model while keeping training data at local clients. On the other hand, in a wide range of machine learning and signal processing applications, the desired solution naturally has a certain structure that can be framed as sparsity with respect to a certain dictionary. This problem can be formulated as an optimization problem with sparsity constraints, and solving it efficiently has been one important topic in the traditional centralized setting. In this talk, we introduce a federated gradient matching pursuit (FedGradMP) algorithm to solve the sparsity constrained minimization problem in the federated learning setting.

3. A bilevel optimization approach to inverse mean-field games

Jiajia Yu[1,*]

[1]Duke University; [*]yuj12@rpi.edu;

Mean-field games study the Nash Equilibrium in a non-cooperative game with infinitely many agents. It has wide applications in various fields, such as economics, engineering, pandemic control, traffic flow models, and social dynamics. Recently, mean-field game/control problems have been extended into machine learning and reinforcement learning. In the game, each agent aims to minimize a combination of dynamic cost, interaction cost, and terminal cost by controlling its own state trajectory. In most existing works, it is required to know the cost functions in order to solve mean-field games. However, in practice, the cost functions are not always easy to obtain. On the contrary, the state distribution, the strategy of agents and sometimes the value function at the Nash Equilibrium can be observed.
In this talk, I will talk about a bilevel optimization formulation for inverse mean-field games and study the numerical methods for solving the bilevel problem. With the bilevel formulation, we preserve the convexity of the objective function and the linearity of the constraint in the forward problem. This formulation permits us to solve the problem with a gradient-based optimization algorithm that has a nice convergence guarantee. I will present our numerical results for two classes of examples, inverse mean-field games with unknown obstacles and unknown metrics. If time permits, I will talk about the numerical stability of these two inverse problems and the local unique identifiability result for the unknown obstacle problems.

■ First-Order Methods for Convex Optimization

Room: **RB 241** (3:45–5:15pm) Chair: *Javier Peña and Mateo Díaz*

1. Away-Step Frank-Wolfe Method for Minimizing Logarithmically-Homogeneous Barriers

Renbo Zhao[1,*]

[1]MIT; [*]renboz@mit.edu; [1]University of Iowa;

We present and analyze a new away-step Frank-Wolfe method for the convex optimization problem $\min_{x\in\mathcal{X}} f(\mathsf{A}x) + c(x)$, where f is a θ-logarithmically-homogeneous self-concordant barrier, A is a linear operator, $c(\cdot)$ is a linear function and $\mathcal{X}$ is a nonempty polytope. We establish the global linear convergence rate of our Frank-Wolfe method in terms of both the objective gap and the Frank-Wolfe gap. This, in particular, settles the question raised in Ahipasaoglu, Sun and Todd (2008) on the global linear convergence of the away-step Frank-Wolfe method specialized to the D-optimal design problem. Experimental results on D-optimal design and inference of multi-dimensional Hawkes processes are in excellent agreement with our theory, and demonstrate the superior numerical performance of our methods over other first-order methods.

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

Jinwen Yang[1,*], *Haihao Lu*[1]

[1]University of Chicago; [*]jinweny@uchicago.edu;

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

3. Affine invariant convergence rates of Frank-Wolfe algorithms on polytopes

Javier Peña[1,*]

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

This talk will discuss affine invariant convergence properties of several variants of the Frank-Wolfe algorithm over a polytope. We will consider the classical Frank-Wolfe algorithm as well as the following variants: away steps, blended pairwise, and decomposition-invariant pairwise. We will discuss results for two types of stepsizes: exact line-search and open-loop steps. Our discussion will highlight the key role played by a suitable condition measure of the polytope, namely a kind of affine invariant "facial distance".

■ Optimization and Applications I

Room: **RB 271** (3:45–5:15pm) Chair: *Abdullah Konak*

1. Optimal and equilibrium solutions to the backup server problem

Dr. Vasco F Alves[1,*], *Dr. Jiesen Wang*[2], *Prof. Refael Hassin*[2]

[1]University of Birmingham ; *v.alves@bham.ac.uk; [2]Tel Aviv University;

Backup servers are often used in service systems to lower operating costs and increase reliability. Backup servers may have a lower service quality or a slower service rate. We analyze a system with one primary and one backup server and compute the system and individually optimal server-selection strategies. We also compare the optimal strategy with the better of two simple strategies: using only the primary server or never waiting for the primary server when the backup server is idle. We discuss when such a simple strategy is a good approximation. Finally, we compare the individually optimal strategy to the socially-optimal.

2. An Efficient Algorithm to the Integrated Shift and Task Scheduling Problem

G S R MURTHY[1,*]

[1]Indian Statistical Institute, India; *murthygsr@gmail.com;

We present a new solution methodology for the integrated task and staff scheduling problem that has a wide range of applications. This problem becomes computationally very complex when one considers flexible shifts with statutory regulations and tasks with precedence relationships and flexible start times. Existing methods to find optimal solutions to these problems use integer programming formulations and iterative procedures. These methods are computationally very expensive both in terms of running time and coding efforts. Further, the test instances cited in the literature consider limited flexibility. Against this backdrop, our methodology is simple and considers all possible shift pattern, flexible start times for the tasks and precedence relationships. The problem is solved in two stages, each stage involving solving an integer linear programming problem. Our method produces an optimal solution if the objective function depends only on the shift patterns and their positioning, and near optimal solutions if the objective is to minimize the number of workers. For the latter case, we provide a lower bound on the number of workers. With this, we establish the efficacy of our solutions through test cases. For large size problem instances, we developed a new technique, the split technique, which can handle very large size problems with solution time increasing only in a linear fashion. The efficacy of our method is established through several test instances where we compare the solution times with the existing solutions. Reduction in solution times varies from 75 to 99 percent.

3. Data Envelopment Analysis to Develop Composite Resilience Indexes for Sustainable Future

Abdullah Konak[1,*], *Michael Gregory Jacobson*[1], *Adelaide Cassia Nardocci*[2], *Nazmiye Ozkan*[3], *Tanaya Sarmah*[3], *Karina Sass*[2], *Marcos Roberto Benso*[2], *Elisabeth Shrimpton*[3], *Marina Batalini De Macedo*[4], *E Mario Mendiondo*[2], *Pedro Gustavo*[5], *Alina Rodriguez*[1], *Greicelene J. da Silva*[2]

[1]The Pennsylvania State University, Berks; *auk3@psu.edu;
[2]University of Sao Paulo, Brazil; [3]Cranfield University, UK;
[4]UNIFEI, Brazil; [5]marcosbenso@usp.br;

Research on environmental risk modeling relies on numerous indicators to quantify the magnitude and frequency of extreme climate events, their ecological, economic, and social impact, and coping mechanisms that can reduce or avoid their adverse effects. Index-based approaches greatly simplify the process of quantifying, comparing, and monitoring risks associated with natural hazards because a large set of indicators can be summarized into a few key performance indicators. Data fusion techniques are frequently used along with expert opinions to create key performance indicators. This paper presents a data envelopment analysis (DEA) approach for creating a composite index to evaluate policy decisions for reducing drought's negative impact on people's livelihoods and comparing drought resilience of different systems. DEA is a linear programming methodology that quantifies the relative efficiency of multiple decision-making units (DMUs) with multiple inputs and outputs. The inputs and outputs are decoupled in a typical DEA model, leading to linear models. In the proposed approach, the relations between the input and outputs are partially known. Therefore, new mathematical models are proposed to find the efficient frontier for the cases where inputs and outputs are loosely coupled. The resulting models are challenging to solve yet better represent real-life drought resilience scenarios. The paper presents applications of these new DEA models and computational techniques to solve them efficiently.

■ Poster Session and Competition

Room: **Rauch Atrium** (5:15–6:15pm) Chair: *Albert S. Berahas*

1. A new multiobjective heuristic for creating political redistricting plans while maximizing similarity to a previously-used plan

Brendan Ruskey[1,*], *Lawrence V. Snyder*[1]

[1]Lehigh University; *bjr221@lehigh.edu;

In this project, we introduce a multiobjective genetic algorithm (MOGA) for generating political redistricting plans. Unlike all existing MOGAs for redistricting, and most other heuristic algorithms for redistricting, our MOGA produces plans that maximize similarity to an existing plan. Our focus on similarity with a given plan addresses the real-life phenomenon that new redistricting plans are typically created with the previous plan in mind. As part of our algorithm, in order to promote plans that resemble some given original plan, we utilize a new initialization method for setting the initial population of the MOGA. In addition to similarity with the given plan, our algorithm considers the classic objectives of population deviation and compactness, and returns a set of redistricting plans that are nondominated in these three objectives. Our numerical experiments demonstrate the effectiveness of our MOGA in generating redistricting plans that balance the objectives of similarity, compactness, and population deviation.

2. Dual dynamic programming for stochastic programs over an infinite horizon

Caleb Ju[1,*], *Guanghui Lan*[1]

[1]Georgia Tech; *cju33@gatech.edu;

We consider a dual dynamic programming algorithm for solving stochastic programs over an infinite horizon. We show non-asymptotic convergence results when using an explorative strategy, and we then enhance this result by reducing the dependence of the effective planning horizon from quadratic to linear. This improvement is achieved by combining the forward and backward phases from dual dynamic programming into a single iteration. We then apply our algorithms to a class of problems called hierarchical stationary stochastic programs, where the cost function is a stochastic multi-stage program. The hierarchical program can model problems with a hierarchy of decision-making, e.g., how long-term decisions influence day-to-day operations. We show that when the subproblems are solved inexactly via a dynamic stochastic approximation-type method, the resulting hierarchical dual dynamic programming can find approximately optimal solutions in finite time. Preliminary numerical results show the practical benefits of using the explorative strategy for solving the Brazilian hydro-thermal planning problem and economic dispatch, as well as the potential to exploit parallel computing.

3. Provable Convergence of Tensor Decomposition-Based Neural Network Training

Chenyang Li[1,*], *Bo Shen*[1]

[1]New Jersey Institute of Technology; *cl237@njit.edu;

Tensorized neural network, coupled with advanced tensor decomposition such as tensor train decomposition, is a commonly used model compression technique and finds broad applications where the memory of computing devices is limited. However, training tensorized neural network is challenging. Existing training strategies often lead to severe performance deterioration. In this paper, we revisit the training procedure from a novel perspective of nonconvex optimization by designing an appropriate objective function. Then we propose TenBCD, a tensor block coordinate descent algorithm to solve it. One advantage of our algorithm is that a closed-form iteration scheme can be derived in some scenarios, which reduces the computational complexity considerably and is gradient-free. With the Kurdyka-Lojasiewicz property of our objective function, we show that our algorithm converges to a critical point at the rate of O(1/K), where K denotes the number of iterations. Lastly, extensive experiments demonstrate the efficiency and superior performance of our training framework.

4. Stochastic Methods for Multi-Level Multi-Objective Optimization

Griffin Kent[1,*], *Tommaso Giovannelli*[1], *Luis Nunes Vicente*[1]

[1]Lehigh University; *gdk220@lehigh.edu;

We introduce a bilevel stochastic gradient method for bilevel problems with lower level constraints. We also present a comprehensive convergence theory that covers all inexact calculations of the adjoint gradient (also called hypergradient) and addresses both the lower-level unconstrained and constrained cases. To promote the use of bilevel optimization in large-scale learning, we introduce a practical bilevel stochastic gradient method (BSG-1) that does not require second-order derivatives and, in the lower-level unconstrained case, dismisses any system solves and matrix-vector products.
To solve bilevel problems with a multi-objective lower level, different approaches can be considered depending on the interpretation of the lower-level optimality. An optimistic formulation that was previously introduced for the deterministic case consists of minimizing the upper-level function over all non-dominated lower-level solutions. We develop new risk-neutral and risk-averse formulations, address their main computational challenges, and develop the corresponding deterministic and stochastic gradient-based algorithms.

5. Moving anchor accelerated algorithms for smooth minimax problems

James K. Alcala[1,*], *Yat Tin Chow*[1]

[1]University of California, Riverside; *jalca014@ucr.edu;

Our work introduces a moving anchor technique to extragradient algorithms for smooth structured minimax problems. First, our moving anchor technique is introduced into the original algorithmic anchoring framework known as EAG. We match the optimal order of convergence in terms of worst-case complexity on the squared gradient, $O(1/k^2)$. As many problems of practical interest are nonconvex-nonconcave, the recently developed FEG class of algorithms brings order-optimal methods developed within EAG to the nonconvex-nonconcave problem settings. We introduce the moving anchor methods to the FEG class of algorithms and again obtain order-optimal complexity results. In both problem settings, a variety of numerical examples demonstrate the efficacy of our algorithms. A proximal-point version of our algorithms is developed. Future directions are also discussed.

6. High Probability Sample Complexity Bounds for Adaptive Optimization Methods with Stochastic Oracles

Miaolan Xie[1,*]

[1]Cornell University; *mx229@cornell.edu;

We consider a simple adaptive optimization framework for continuous optimization where the (explicit or implicit) step size in each iteration is adaptively adjusted by the estimated progress of the algorithm instead of requiring manual tuning or using a pre-specified sequence of step sizes. The framework accommodates a stochastic setting where function value, gradient (and possibly Hessian) estimates are available only through noisy probabilistic oracles (can be biased and possibly arbitrarily bad with some constant probability). This framework is very general and encompasses stochastic variants of line search, quasi-Newton, cubic regularized Newton methods for unconstrained problems, and stochastic SQP methods for constrained problems. The probabilistic oracles capture multiple standard settings including expected loss minimization in machine learning, zeroth-order (derivative-free) and low-precision optimization. Under reasonable conditions on the oracles, we derive high probability bounds on the sample complexity of the algorithms.

7. Multisecant Quasi-Newton methods

Mokhwa Lee[1,*]

[1]Stony Brook University; [*]mokhwa.lee@stonybrook.edu;

When dealing with a large-scale optimization problem, classical second-order methods, such as Newton's method, are no longer practical because it requires iteratively solving a large-scale linear system of order n. For this reason, Quasi-Newton(QN) methods, like BFGS or Broyden's method, are introduced because they are more efficient than Newton's method. This project focuses on multi-secant extensions of the BFGS method, to improve its Hessian approximation properties. Unfortunately, doing so sacrifices the matrix estimate's positive semidefiniteness, and steps are no longer assured to be descent directions. Therefore, we apply a perturbation strategy to construct an almost-secant positive-definite Hessian estimate matrix. This strategy has a low computational cost, involving only rank-2 updates with variable and gradient successive differences. We also explore several ways of improving this method, accepting and rejecting older updates according to several non-degeneracy metrics. Future goals include extending these techniques to limited memory versions.

8. Higher Order Markov Random Fields and Binary Polynomial Optimization

Yakun Wang[1,*]

[1]Lehigh University; [*]yaw220@lehigh.edu;

Backup servers are often used in service systems to lower operating costs and increase reliability. Backup servers may have a lower service quality or a slower service rate. We analyze a system with one primary and one backup server and compute the system and individually optimal server-selection strategies. We also compare the optimal strategy with the better of two simple strategies: using only the primary server or never waiting for the primary server when the backup server is idle. We discuss when such a simple strategy is a good approximation. Finally, we compare the individually optimal strategy to the socially-optimal.

■ AIMMS-MOPTA Optimization Modeling Competition

Room: **RB 184** (9:45 – 11:15am) Chair: *Xiu Yang*

■ Recent Advances in Frank-Wolfe-type Methods

Room: **RB 85** (9:45 – 11:15am) Chair: *Xin Jiang*

1. Projection-free Methods for Solving Saddle Point Problems

Erfan Yazdandoost Hamedani[1,*], *Morteza Boroun*[1], *Afrooz Jalilzadeh*[1]

[1]The University of Arizona; *erfany@arizona.edu;

This talk addresses constrained saddle point optimization problems with nonconvex-concave and smooth objective functions, which arise in many machine learning problems. Existing methods rely on computationally expensive projection onto constraint sets. To fill this gap, we propose efficient single-loop projection-free methods using first-order information. We introduce a primal-dual conditional gradient method utilizing linear minimization oracles, achieving an ε-stationary solution in O(ε^{-6}) iterations when the maximization problem's constraint set is strongly convex. Additionally, we present a one-sided projection-free method with regularized projected gradient ascent, achieving an ε-stationary solution in O(ε^{-4}) iterations when the projection is easy to compute.

2. Multiplicative Gradient Method: When and Why It Works

Renbo Zhao[1,*]

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

The multiplicative gradient (MG) method was originally developed for the positron emission tomography problem, where in each iteration one simply multiplies the current iterate with the current gradient to produce the next iterate. Its convergence rate was recently discovered by Zhao (2022). Despite this special case, the understanding of when and why MG works in general is lacking. In this talk, we unravel this mystery by identifying a general class of convex problems that involve optimizing a log-homogeneous function over a slice of a symmetric cone, and propose a generalization of MG. We show that the generalized MG method has $O(ln(n)/t)$ convergence rate on this class of problems, where n is the rank of the symmetric cone. Our theory not only subsumes the existing results as special cases, but also helps to identify new applications where the MG method significantly outperforms the state-of-the-art.

3. A globally convergent difference-of-convex algorithmic framework and application to information theory

Xin Jiang[1,*], *Chaorui Yao* [2]

[1]Lehigh University; *xjiang@lehigh.edu; [2]University of California, Los Angeles;

The difference-of-convex algorithm (DCA) is a conceptually simple method for the minimization of (possibly) nonconvex functions that are expressed as the difference of two convex functions. At each iteration, DCA constructs a global overestimator of the objective and solves the resulting convex subproblem. Despite its conceptual simplicity, the theoretical understanding and algorithmic framework of DCA needs further investigation. In this paper, global convergence of DCA at a linear rate is established under an extended Polyak-Lojasiewicz condition. The proposed condition holds for a class of DC programs with a bounded, closed, and convex constraint set, for which global convergence of DCA cannot be covered by existing analyses. Moreover, the DCProx computational framework is proposed, in which the DCA subproblems are solved by a primal–dual proximal algorithm with Bregman distances.
With a suitable choice of Bregman distances, DCProx has simple update rules with cheap per-iteration complexity. As an application, DCA is applied to several fundamental problems in network information theory, for which existing solvers cannot compute the global optimum.
For these problems, our analysis proves the global convergence of DCA, and more importantly, DCProx solves the DCA subproblems efficiently.

■ Applications of Optimization to Dynamical Systems

Room: **RB 91** (9:45 – 11:15am) Chair: *Abraar Chaudhry*

1. Meta-Learning Operators to Optimality from Multi-Task Non-IID Data

Nikolai Matni[1,*], *Thomas Zhang*[2], *Leonardo F. Toso*[2], *James Anderson*[2]

[1]University of Pennsylvania; *nmatni@seas.upenn.edu;
[2]Columbia University;

A powerful concept behind much of the recent progress in machine learning is the extraction of common features across data from heterogeneous sources or tasks. Intuitively, using all of one's data to learn a common representation function benefits both computational effort and statistical generalization by leaving a smaller number of parameters to fine-tune on a given task. Toward theoretically grounding these merits, we propose a general setting of recovering linear operators M from noisy vector measurements y = Mx + w, where the covariates x may be both non-i.i.d. and non-isotropic. We demonstrate that existing isotropy-agnostic meta-learning approaches incur biases on the representation update, which causes the scaling of the noise terms to lose favorable dependence on the number of source tasks. This in turn can cause the sample complexity of representation learning to be bottlenecked by the single-task data size. We introduce an adaptation, De-bias & Feature-Whiten (DFW), of the popular alternating minimization-descent (AMD) scheme proposed in Collins et al., 2021, and establish linear convergence to the optimal representation with noise level scaling down with the total source data size. This leads to generalization bounds on the same order as an oracle empirical risk minimizer. We verify the importance of DFW on various numerical simulations. In particular, we show that vanilla alternating-minimization descent fails catastrophically even for iid, but mildly non-isotropic data. Our analysis unifies and generalizes prior work, and provides a flexible framework for a wider range of applications, such as in controls and dynamical systems.

2. Automated reachability analysis of neural network dynamical systems

Mahyar Fazlyab[1,*]

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

Going beyond machine learning tasks, neural networks also arise in a variety of control and robotics problems, where they function as feedback control policies, motion planners, perception modules/observers, or models of dynamical systems. However, adopting these approaches in safety-critical domains (such as robots working alongside humans) has been hampered due to a lack of stability and safety guarantees, which can be primarily attributed to the large-scale and compositional structure of neural networks. These challenges only exacerbate when neural networks are integrated into feedback loops, in which time evolution adds another axis of complexity. In this talk, we present a novel framework based on non-(convex) optimization and robust control that can provide stability, safety, and robustness certificates for NN-driven systems.

3. Optimal distributed state estimation using convex optimization

Juncal Arbelaiz[1,*]

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

The celebrated Kalman filter has been widely applied in diverse fields to optimally estimate the state of dynamical systems. Despite this success, some challenges remain regarding its implementation in spatially distributed and interconnected systems: Kalman filters are optimal under the implicit assumption that all-to-all (i.e., centralized) communications are available within the system, which requires a centralized fusion center; however, centralized communications are often infeasible in spatially distributed systems. Such a challenge motivates the design of distributed Kalman filters, in which communications are limited.
In this talk, I will focus on the design of optimal distributed Kalman filters for spatially invariant systems with linear spatiotemporal dynamics described by PDEs. This is a particular instance of infinite-dimensional filtering. First, I will briefly review important structural properties of the Kalman filter in this problem set-up. Then, I will utilize these properties to formulate the distributed Kalman filtering problem as a convex functional optimization. I will discuss the sensitivity of the suboptimality gap of centralized and decentralized filter architectures to different system parameters and highlight the usefulness of dimensional analysis for this task.

■ Learning for Optimization: Performance and Robustness

Room: **RB 241** (9:45 – 11:15am) Chair: *Bartolomeo Stellato*

1. End-to-End Learning to Warm-Start Fixed Point Optimizers

Rajiv Sambharya[1,*], *Vinit Ranjan*[1], *Georgina Hall*[2], *Brandon Amos*[3], *Bartolomeo Stellato*[1]

[1]Princeton University; *rajivs@princeton.edu; [2]INSEAD; [3]Meta AI;

We introduce a machine learning framework to warm-start fixed-point algorithms for parametric convex optimization. Our architecture consists of a neural network mapping problem parameters to warm-starts, followed by a predefined number of fixed-point iterations, with the loss function defined as the final fixed-point residual. In this way, the neural network predicts warm-starts with the end-to-end goal of minimizing the downstream loss. We provide PAC-Bayes generalization bounds on unseen data for common classes of fixed-point operators, including contractive and averaged operators. We apply our framework to popular algorithms including ADMM and Douglas-Rachford splitting to solve problems in control, statistics, and energy. Our approach significantly reduces the number of iterations required to solve these problems through learned warm-starts.

2. Learning for Robust Optimization

Irina Wang[1,*], *Cole Becker*[1], *Bart Van Parys*[2]

[1]Princeton University; *iywang@princeton.edu; [2]MIT Sloan School of Management;

We propose a data-driven technique to automatically learn the uncertainty sets in robust optimization. Our method reshapes the uncertainty sets by minimizing the expected performance across a family of problems while guaranteeing constraint satisfaction. We learn the uncertainty sets using a novel stochastic augmented Lagrangian method that relies on differentiating the solutions of the robust optimization problems with respect to the parameters of the uncertainty set. We show sublinear convergence to stationary points under mild assumptions, and finite-sample probabilistic guarantees of constraint satisfaction using empirical process theory. Our approach is very flexible and can learn a wide variety of uncertainty sets while preserving tractability. Numerical experiments show that our method outperforms traditional approaches in robust and distributionally robust optimization in terms of out of sample performance and constraint satisfaction guarantees. We implemented our method in the open-source package LROPT.

3. Learning Rationality Parameters in Potential Games

Stefan Clarke[1,*]

[1]Princeton University; *sc8647@princeton.edu;

We propose a stochastic first-order algorithm to learn the rationality parameters of simultaneous and non-cooperative potential games, i.e, the parameters of the agents' optimization problems. Our technique combines an active-set step that enforces that the agents play at a Nash equilibrium and an implicit-differentiation step to update the estimates of the rationality parameters. We detail the convergence properties of our algorithm and perform numerical experiments on Cournot and congestion games, showing that our algorithm effectively finds high-quality solutions (in terms of out-of-sample loss) and scales to large datasets.

4. On the System Efficiency and Backdoor Robustness of Federated Learning

Hongyi Wang[1,*]

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

Federated learning (FL) offers a privacy-preserving collaborative learning solution, though it grapples with high communication costs and susceptibility to attacks. In this talk, Dr. Wang will initially present FedMA, which matches and averages hidden elements bearing similar signatures. FedMA surpasses widely-used FL algorithms, all while reducing communication overhead. Subsequently, Dr. Wang will explore FL's robustness, emphasizing vulnerabilities to backdoor attacks and asserting that robustness to backdoors entails resilience to adversarial examples with guarantees. He will introduce edge-case backdoors, which trigger misclassifications on atypical inputs, potentially jeopardizing fairness across diverse ML tasks.

■ Recent Advances in Stochastic Optimization

Room: **RB 271** (9:45 – 11:15am) Chair: *Miaolan Xie*

1. Balancing Communication and Computation in Gradient Tracking Algorithms for Decentralized Optimization

Albert S. Berahas[1,*], *Raghu Bollapragada*[2], *Shagun Gupta*[2]

[1]University of Michigan; [*]albertberahas@gmail.com; [2]UT Austin;

In this talk, we present a framework that unifies gradient tracking methods for decentralized optimization. We establish unified theoretical convergence results for the algorithmic framework with any composition of communication and computation steps, and quantify the improvements achieved as a result of this flexibility. The framework recovers the results of popular gradient tracking methods as special cases, and allows for a direct comparison of these methods. Finally, we illustrate the performance of the proposed methods on quadratic functions and binary classification problems.

2. Stochastic Optimization for Variational Quantum Algorithms

Matt Menickelly[1,*], *Stefan M. Wild* [2], *Miaolan Xie*[3]

[1]Argonne National Lab; [*]mmenickelly@anl.gov; [2]Lawrence Berkeley National Laboratory; [3]Cornell University;

Variational quantum algorithms (VQAs), which have risen to prominence in the noisy intermediate-scale quantum setting, require the implementation of a stochastic optimizer on classical hardware. To date, most research in this setting has employed algorithms based on the stochastic gradient (SG) iteration as the stochastic classical optimizer. However, VQAs pose interesting challenges in that 1) there are practical hardware reasons for why the computation of derivatives for VQA cost functions remains prohibitive on near-term hardware, and 2) many standard assumptions in the analysis of SG methods simply do not hold in the VQA setting. In this presentation, I will discuss my research on various fronts in this problem, discussing both advances in SG methods (the development of quasi-Newton methods in the absence of common random numbers), as well as an empirical comparison of derivative-free and derivative-based methods on this class of problems.

3. Stochastic approximation with decision-dependent distributions: asymptotic normality and optimality

Mateo Díaz[1,*], *Joshua Cutler*[2], *Dmitriy Drusvyatskiy*[2]

[1]Johns Hopkins University; [*]mateodd@jhu.edu; [2]University of Washington;

We analyze a stochastic approximation algorithm for decision-dependent problems, wherein the data distribution used by the algorithm evolves along the iterate sequence. The primary examples of such problems appear in performative prediction and its multiplayer extensions. We show that under mild assumptions, the deviation between the average iterate of the algorithm and the solution is asymptotically normal, with a covariance that nicely decouples the effects of the gradient noise and the distributional shift. Moreover, building on the work of Hájek and Le Cam, we show that the asymptotic performance of the algorithm is locally minimax optimal.

■ Nonlinear and Stochastic Optimization Algorithms I

Room: **RB 85** (1:30–3:00 pm) Chair: *Daniel P. Robinson*

1. A Riemannian ADMM

Shiqian Ma[1,*], *Jiaxiang Li*[2], *Tejes Srivastava*[2]

[1]Rice University; [*]sqma@rice.edu; [2]UC Davis;

We consider a class of Riemannian optimization problems where the objective is the sum of a smooth function and a nonsmooth function, considered in the ambient space. This class of problems finds important applications in machine learning and statistics such as the sparse principal component analysis, sparse spectral clustering, and orthogonal dictionary learning. We propose a Riemannian alternating direction method of multipliers (ADMM) to solve this class of problems. Our algorithm adopts easily computable steps in each iteration. The iteration complexity of the proposed algorithm for obtaining an epsilon-stationary point is analyzed under mild assumptions. To the best of our knowledge, this is the first Riemannian ADMM with provable convergence guarantee for solving Riemannian optimization problem with nonsmooth objective. Numerical experiments are conducted to demonstrate the advantage of the proposed method.

2. Some Primal-Dual Theory for Subgradient Methods for Strongly Convex Optimization

Benjamin David Grimmer[1,*], *Danlin Li*[1]

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

We consider (stochastic) subgradient methods for strongly convex but potentially nonsmooth non-Lipschitz optimization. We provide new equivalent dual descriptions (in the style of dual averaging) for the classic subgradient method, the proximal subgradient method, and the switching subgradient method. These equivalences enable optimal convergence guarantees in terms of both their classic primal gap and a not previously analyzed dual gap for strongly convex optimization. Consequently, our theory provides these classic methods with simple, optimal stopping criteria and optimality certificates at no added computational cost. Our results apply under nearly any stepsize selection and for a range of non-Lipschitz ill-conditioned problems where the early iterations of the subgradient method may diverge exponentially quickly (a phenomenon which, to the best of our knowledge, no prior works address). Even in the presence of such undesirable behaviors, our theory still ensures and bounds eventual convergence.

3. Inexact Proximal-Methods for Constrained Optimization

Daniel P. Robinson[1,*], *Xiaoyi Qu*[1], *Yutong Dai*[1], *Frank E. Curtis*[1]

[1]Lehigh University; [*]daniel.p.robinson@lehigh.edu;

We discuss several recent advances in proximal-methods for sparse optimization. We first present a new subproblem solver that allows for inexact solutions of the proximal-gradient subproblem to be computed. We then present inexact stopping conditions that allow us to derive complexity and support identification results. We also introduce a proximal method for solving deterministic optimization problems, as well as to minimize stochastic objective functions subject to deterministic constraints.

■ Quantum Computing and Optimization II

Room: **RB 91** (1:30–3:00 pm) Chair: *M. Mohammadisiahroudi*

1. Graph decomposition techniques for solving CO problems with VQAs

Rebekah Herrman[1,*], *Moises Ponce*[1], *Phillip Lotshaw*[2], *Sarah Powers*[2], *George Siopsis*[1], *Travis Humble*[2], *James Ostrowski*[1]

[1]The University of Tennessee Knoxville Laboratory; *rherrma2@utk.edu; [2]Oak Ridge National Laboratory;

The quantum approximate optimization algorithm (QAOA) has the potential to approximately solve complex combinatorial optimization problems in polynomial time. However, current noisy quantum devices cannot solve large problems due to hardware constraints. In this presentation, we discuss a recent algorithm that decomposes the QAOA input problem graph into a smaller problem and solves Max-Cut using QAOA on the reduced graph. On average, the reduced problems require only approximately 1/10 of the number of vertices than the original MaxCut instances. Furthermore, the average approximation ratio of the original MaxCut problems is 0.75, while the approximation ratios of the decomposed graphs are on average of 0.96.

2. Quantum Hamiltonian Descent

Jiaqi Leng[1,*], *Ethan Hickman*[1], *Joseph Li*[1], *Xiaodi Wu*[1]

[1]University of Maryland; *jiaqil@umd.edu;

Gradient descent is a fundamental algorithm in the theory and practice of continuous optimization. Identifying its counterpart in quantum computing is not only theoretically interesting but also could lead to practical quantum algorithms. A conventional approach to quantum speedups in optimization relies on the quantum acceleration of intermediate steps of corresponding classical algorithms while keeping the quality of their solutions. We propose Quantum Hamiltonian Descent (QHD) as a truly quantum counterpart of classical gradient methods, derived from the path integral of classical dynamical systems corresponding to the continuous-time limit of classical gradient descent algorithms. We establish QHD's convergence to the global optimum in both convex and non-convex settings. More importantly, QHD is described as a time-dependent Hamiltonian evolution that can be efficiently simulated on both digital and analog quantum computers. By embedding QHD into the so-called Quantum Ising Machine (including D-Wave and others), we empirically observe that the D-Wave-implemented QHD outperforms a selection of state-of-the-art gradient-based classical solvers and the standard quantum adiabatic algorithm, based on the time-to-solution metric, on non-convex quadratic programming instances up to 75 dimensions.

3. BEINIT: Avoiding Barren Plateaus in Variational Quantum Algorithms

Ankit Kulshrestha[1,*]

[1]University of Delaware; *akulshr@udel.edu;

Barren plateaus are a notorious problem in the optimization of variational quantum algorithms and pose a critical obstacle in the quest for more efficient quantum machine learning algorithms. Many potential reasons for barren plateaus have been identified but few solutions have been proposed to avoid them in practice. Existing solutions are mainly focused on the initialization of unitary gate parameters without taking into account the changes induced by input data. In this paper, we propose an alternative strategy which initializes the parameters of a unitary gate by drawing from a beta distribution. The hyperparameters of the beta distribution are estimated from the data. To further prevent barren plateau during training we add a novel perturbation at every gradient descent step. Taking these ideas together, we empirically show that our proposed framework significantly reduces the possibility of a complex quantum neural network getting stuck in a barren plateau

■ Optimization Algorithms for Adversarial Training

Room: **RB 241** (1:30–3:00 pm) Chair: *Jiancong Xiao*

1. Uniformly Stable Algorithms for Adversarial Training

Jiancong Xiao[1,*]

[1]University of Pennsylvania; *jiancongxiao14@gmail.com;
In adversarial machine learning, neural networks suffer from a significant issue known as robust overfitting, where the robust test accuracy decreases over epochs (Rice at al., 2020). Recent research conducted by Xing et al. (2021) and Xiao et al. (2022) has focused on studying the uniform stability of adversarial training. Their investigations revealed that SGD-based adversarial training fails to exhibit uniform stability, and the derived stability bounds align with the observed phenomenon of robust overfitting in experiments. This motivates us to develop uniform stable algorithms specifically tailored for adversarial training. Recognizing that the non-uniform stability comes from the non-smoothness of the adversarial loss, we employ a Moreau envelope function to reframe the original problem as a min-min problem, separating the non-strong convexity and non-smoothness of the adversarial loss. In light of this, we introduce ME-$\mathcal{A}$, an approach that alternates between solving the inner and outer minimization problems. We provide a proof demonstrating that ME-$\mathcal{A}$ achieves uniform stability without incurring any additional computational overhead. Beyond its application in adversarial training, this represents a fundamental result in uniform stability analysis, as ME-$\mathcal{A}$ is the first algorithm to exhibit uniform stability for (weakly) convex, non-smooth problems. Finally, we show the efficacy of ME-$\mathcal{A}$ in mitigating the issue of robust overfitting in practical scenarios.

2. Better Representations via Adversarial Training in Pre-Training: A Theoretical Perspective

Yue Xing[1,*], *Xiaofeng Lin*[2], *Qifan Song*[3], *Yi Xu*[4], *Belinda Zeng*[4], *Guang Cheng*[2,4]

[1]Michigan State University; *xingyue1@msu.edu; [2]University of California, Los Angeles; [3]Purdue University ; [4]Amazon Search-M5;
Pre-training is known to generate universal representations for downstream tasks in large scale deep learning such as large language models. Existing literature, e.g., Kim et al (2020), empirically observe that the downstream tasks can inherit the adversarial robustness of the pre-trained model. We provide theoretical justification for this robustness inheritance phenomenon. Our theoretical results reveal that feature purification plays an important role in connecting the adversarial robustness of the pre-trained model and the downstream tasks in two-layer neural networks. Specifically, we show that (i) with adversarial training, each hidden node tends to pick only one (or a few) feature; (ii) without adversarial training, the hidden nodes can be vulnerable to attacks. This observation is valid for both supervised pre-training and contrastive learning. With purified nodes, it turns out that clean training is enough to achieve adversarial robustness in downstream tasks.

3. Secure Distributed Optimization Under Gradient Attacks

Shuhua Yu[1,*]

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

We study secure distributed optimization against arbitrary gradient attacks in multi-agent networks. In distributed optimization, there is no central server to coordinate local updates, and each agent can only communicate with its neighbors on a predefined network. We consider the scenario where out of n networked agents, a fixed but unknown fraction ρ of the agents are under arbitrary gradient attacks in that their stochastic gradient oracles return arbitrary information to derail the optimization process, and the goal is to minimize the sum of local objective functions on unattacked agents. We propose a distributed stochastic gradient method that combines local variance reduction and clipping (CLIP-VRG). We show that, in a connected network, when the unattacked local objective functions are convex and smooth, share a common minimizer, and their sum is strongly convex, CLIP-VRG leads to almost sure convergence of the iterates to the exact sum cost minimizer at all agents. We quantify a tight upper bound on the fraction ρ of attacked agents in terms of problem parameters such as the condition number of the associated sum cost that guarantee exact convergence of CLIP-VRG, and characterize its asymptotic convergence rate. Finally, we empirically demonstrate the effectiveness of the proposed method under gradient attacks on both synthetic and real-world image classification datasets.

■ Political Redistricting and Multi-Objective Optimization

Room: **RB 271** (1:30–3:00 pm) Chair: *Brendan Ruskey*

1. Optimization models for fair political redistricting

Rahul Swamy[1,*], *Douglas King*[1], *Sheldon Jacobson*[1]

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

Political districting in the United States is a decennial process of redrawing the boundaries of congressional and state legislative districts. Fairness in political districting has been an important topic of subjective debate, with district plans having consequences for multiple stakeholders such as the voters, political parties, candidates and independent commissions. Even though districting as an optimization problem has been well-studied, existing models primarily rely on non-political fairness measures such as the compactness of districts. This talk provides optimization formulations for fairness-based redistricting for fundamental notions of fairness such as proportionality, symmetry and competitiveness. Our algorithmic framework generates approximate-Pareto optimal solutions that illustrate the trade-off between compactness and each of the partisan fairness metrics.

2. The Tradeoff Between District Compactness and Minority Representation in US Redistricting

Brendan Ruskey[1,*], *Lawrence V. Snyder*[1]

[1]Lehigh University; *bjr221@lehigh.edu;

Every ten years in the United States, lawmakers sort the citizens of each state into groups of roughly equal population in what is called a "redistricting plan". These districts are expected to be geometrically compact, and should not disadvantage members of any minority racial group. In many redistricting plans, so-called "opportunity districts" are drawn to promote minority representation in Congress. However, in practice, prioritizing compactness typically comes at the cost of having less minority representation via opportunity districts. In this project, we present a metric for evaluating the compactness of minority opportunity districts. Specifically, we solve a mixed-integer program (MIP) to find optimally-compact opportunity districts, whose compactness we can compare against the opportunity districts of a proposed plan. We also explore how, when compactness is treated as a constraint, increasing the required level of compactness diminishes the level of minority representation that a redistricting plan can achieve, as well as the effect that US Census non-response among minority citizens has on district compactness and minority representation.

3. Menu Engineering at Assisted Living Facilities

Sadan Kulturel-Konak[1,*], *Abdullah Konak*[1], *Nagesh Gavirneni*[2]

[1]The Pennsylvania State University, Berks; *sxk70@psu.edu;
[2]Cornell University;

This project focuses on designing healthy menus for assisted living facilities under multiple, often conflicting, objectives and complex nutritional constraints. The proposed system considers various factors in the construction of menus, including the United States Department of Agriculture (USDA) healthy eating guidelines (HEG), chefs' choices and experiences, diversity of menu items, cost, and residents' preferences with different dietary requirements/diet patterns. A modeling framework was developed based on the input collected by interviewing facility managers, chefs, and dietitians.
We address this complex multi-objective optimization problem using Mixed Integer Linear Programming (MILP). We propose a pseudo diet type concept using clustering and principal component analysis to represent multiple distinct eating patterns in the same diet type. The modeling framework utilizes various approaches to incorporate different stakeholders' perspectives in the decision processes, including a satisficing approach to represent the menu diversification goals of chefs and goal programming to prioritize nutritionists' points of view while ensuring the autonomy of patrons. We develop an iterative greedy search algorithm to solve this large-scale problem. The search algorithm first finds an initial feasible solution using the branch and bound in CPLEX within an allowed CPU. Then it improves the feasible solution by solving the model only for subsets of decision variables. The efficacy of the proposed approach is demonstrated using two use cases of a setting containing over 300 menu items and three diet profiles. The final menus are evaluated based on the following performance metrics: total preference score, deviations from the HEG, and total cost.

■ Nonlinear and Stochastic Optimization Algorithms II

Room: **RB 85** (4:30–6:00 pm) Chair: *Xiaoyi Qu*

1. Recent Developments in Quasi-Newton Methods for Numerical Optimization

Jeb Runnoe[1,*], *Philip E. Gill*[1], *Ziyan Zhu*[1]

[1]University of California, San Diego; *jrunnoe@ucsd.edu;

Quasi-Newton methods form the basis of many effective methods for unconstrained and constrained optimization. Quasi-Newton methods require only the first-derivatives of the problem to be provided and update an estimate of the Hessian matrix of second derivatives to reflect new approximate curvature information found during each iteration. In the years following the publication of the Davidon-Fletcher-Powell (DFP) method in 1963 the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update emerged as the best update formula for use in unconstrained minimization. More recently, a number of quasi-Newton methods have been proposed that are intended to improve on the efficiency and reliability of the BFGS method. Unfortunately, there is no known analytical means of determining the relative performance of these methods on a general nonlinear function, and there is no accepted standard set of test problems that may be used to verify that results reported in the literature are comparable. In this talk we will discuss ongoing work to provide a thorough derivation, implementation, and numerical comparison of these methods in a systematic and consistent way. We will look in detail at several modifications, discuss their relative benefits, and review relevant numerical results.

2. A Stochastic GDA Method With Backtracking For Solving Nonconvex Strongly Concave Minimax Problems

Qiushui Xu[1,*], *Xuan Zhang*[1], *Necdet Serhat Aybat*[1], *Mert Gurbuzbalaban*[2]

[1]The Pennsylvania State University; *qjx5019@psu.edu;
[2]Rutgers University;

We propose a stochastic Gradient Descent Ascent method with backtracking (SGDA-B) to solve nonconvex-strongly-concave (NSC) minimax problems $\min_x \max_y \sum_{i=1}^N g_i(x_i) + f(x,y) - h(y)$, where h and g_i for $i = 1, \ldots, N$ are closed convex, f is L-smooth and mu-strongly concave in y. SGDA-B does not require to know L, and using random block-coordinate updates it can compute an ε-stationary point with probability p within $O(\kappa^2 L\varepsilon^{-4} log(1/p))$ stochastic oracle calls, where $\kappa = L/mu$. To our knowledge, SGDA-B is the first GDA-type method with backtracking to solve NSC minimax problems. We provide numerical results for SGDA-B on a distributionally robust learning problem.

3. A Sequential Proximal Algorithm for Minimizing a Composite Objective Function with Equality Constraints

Xiaoyi Qu[1,*], *Yutong Dai*[1], *Daniel P. Robinson*[1]

[1]Lehigh University; *xiq322@lehigh.edu;

We consider composite objective function subject to equality constraints and present an algorithm that borrows ideas from sequential quadratic programming and proximal method. While state-of-the-art line search SQP method cannot cope with nonsmooth objective function and proximal method cannot handle complex constraints, their combined usage is shown to successfully offset their respective shortcomings. In particular, due to the presence of nonsmooth terms, we solve a trust region subproblem and a QP subproblem sequentially to obtain the search direction. Under reasonable assumptions, subsequential convergence from remote starting point (either feasible or infeasible) is proved for the proposed algorithm. Numerical experiments based on modified CUTEr test problems demonstrate the practical performance of the proposed method.

■ Quantum Computing and Optimization III

Room: **RB 91** (4:30–6:00 pm) Chair: *M. Mohammadisiahroudi*

1. New perspectives on quantum interior point methods

Brandon Augustino[1,*]

[1]Lehigh University; *bra216@lehigh.edu;

Existing quantum interior point methods seek to speedup their classical counterpart by using quantum linear systems algorithms to solve the Newton linear system at each iteration. While this approach yields polynomial speedups in the dimension, it introduces a dependence on a condition number bound which may prohibit any overall advantage. In this talk we present two new approaches to quantizing IPMs. The first approach draws on a recently established relationship between interior point methods and simulated annealing, while the second seeks to derive a quantum central path.

2. A Quantum Framework for Topological Data Analysis

Bernardo Ameneyro[1,*], *George Siopsis*[1], *Vasileios Maroulas*[1], *Rebekah Herrman*[1]

[1]University of Tennessee; *bameneyr@vols.utk.edu;

Topological data analysis methods capture the shape of data, which can be useful for classification problems. First they extract topological features from the data — like the number of connected components, holes and voids — via persistent homology, tracking them across different scales or resolutions. The topological features are then displayed in persistence diagrams that show the birth and death of each feature. These diagrams are a great way to summarize data before performing machine learning algorithms. But to do this one must compute distances between persistence diagrams, such as the Wasserstein distance. To estimate these one needs to minimize a cost function over all possible ways to match points from two persistence diagrams. We present a quantum algorithm for persistent homology, as well as a QAOA approach for estimating the distance between persistence diagrams, and the advantages that they promise once quantum computers are more advanced.

3. An Inexact Feasible Quantum Interior Point Method for Linear and Quadratic Optimization

Zeguan Wu[1,*], *Mohammadhossein Mohammadisiahroudi*[1], *Brandon Augustino*[1], *Xiu Yang*[1], *Tamás Terlaky*[1]

[1]Lehigh University; *zew220@lehigh.edu;

Quantum linear system algorithms (QLSAs) have the potential to speed up algorithms that rely on solving linear systems. Interior Point Methods (IPMs) yield a fundamental family of algorithms for solving optimization problems. IPMs require the solution of a Newton linear system at each iteration, thus QLSAs can potentially speed up IPMs. Such quantum-assisted IPM (QIPM) provides inexact solution for the Newton linear system due to the noise in contemporary quantum computers. Typically, an inexact search direction leads to an infeasible iterate. In our work, we propose an inexact feasible QIPM (IF-QIPM) and show its advantage in solving linear optimization and linearly constrained convex quadratic optimization problems.

■ Machine Learning and Discrete Optimization

Room: **RB 241** (4:30–6:00 pm) Chair: *Mohammad Hesam Shaelaie*

1. Power System Informed Reinforcement Learning for Unit Commitment

Robert Ferrando[1,*], *Laurent Pagnier*[2], *Zhirui Liang*[2], *Yury Dvorkin*[1], *Michael Chertkov*[1]

[1]University of Arizona; *rferrando@math.arizona.edu;
[2]Johns Hopkins University;

With increased renewable penetration, and the stochasticity inherent in renewable technologies, it is imperative for power system operations to be aware of uncertainty. In this work, we consider unit commitment (UC), a mixed-integer program for day-ahead electricity market clearing, under many load and wind scenarios sampled from a probability distribution. We model the binary on/off statuses of generators as a Markov decision process, and seek to recover the probabilities of transitioning between unit commitment statuses. Furthermore, we propose to use reinforcement learning (RL) to recover the optimal sequence of unit commitment decisions outside of the optimization pipeline, thus reducing the UC problem from mixed-integer to convex and enhancing its computational tractability. The RL algorithm is validated on the New York Independent System Operator (NYISO) 1814-bus system, supplemented with five offshore wind farms.

2. Scarf's algorithm and stable matchings

Chengyue He[1,*], *Yuri Faenza*[1], *Jay Sethuraman*[1]

[1]Columbia University; [*]ch3480@columbia.edu;

Scarf's lemma is a geometric tool that shows the existence of many combinatorial objects, including stable matchings in bipartite graphs, fractional kernels in digraphs, and cores of balanced NTU-games. In its most abstract form, Scarf's lemma shows the existence of a special vertex —a dominating vertex—in any anti-blocking polytope. Scarf's original proof is algorithmic, as it gives a finite pivoting algorithm that finds a dominating vertex. However, the generality of Scarf's results comes at a computational price: the problem of finding a dominating vertex of a down-monotone polytope is PPAD-complete.
The broad applicability of Scarf's lemma and algorithm calls for a more fine-grained analysis of its efficiency. We show what are, to the best of our knowledge, the first proofs of the polynomial-time convergence of Scarf's algorithm in significant settings. Our results apply to the bipartite matching polytope and give new polynomial-time algorithms to find a stable matching in a marriage model, for which all known algorithms use LP solvers or are combinatorial in nature.
Next, we focus on limits of Scarf's algorithm. We show that Scarf's algorithm may output a stable matching other than the one side-optimal, but, under some natural restriction, in general only a matching from an exponentially small subset of all stable matchings.

3. Toward a unified framework for branch & bound and reinforcement learning

Mohammad Hesam Shaelaie[1,*], *Ted Ralphs*[1], *Lawrence V. Snyder*[1]

[1]Lehigh University; [*]mos519@lehigh.edu;

In this research, we are aiming to develop a framework unifying Branch and Bound (BB) and Reinforcement Learning (RL). Each of these algorithmic approaches has its own strengths and weaknesses. Our effort is to design a unified framework (UF) to benefit from the strengths while addressing the weaknesses of each algorithm. By taking a close look at the fundamental elements of each framework, we found out that they carry interesting similarities, despite unique differences in the context of discrete optimization. BB searches the feasible region systematically and employs powerful pruning methods, but this comes at a high cost per iteration. In contrast, RL has a relatively low iteration cost but is also quite myopic; we explore the tradeoffs inherent in these two approaches and explain the idea of UF between these two; also, we will show how BB and RL can be viewed in the UF and how they can exchange information.

■ Fairness in Optimization

Room: **RB 271** (4:30–6:00 pm) Chair: *Man Yiu Tsang*

1. A Unified Framework for Analyzing and Optimizing a Class of Convex Inequity Measures

Man Yiu Tsang[1,*], *Karmel S. Shehadeh*[1]

[1]Lehigh University; [*]mat420@lehigh.edu;

We present a unified framework for analyzing a new parameterized class of convex inequity measures suitable for optimization contexts. We provide theoretical analyses and derive a dual representation of these measures. Importantly, this dual representation renders a unified mathematical expression and an alternative geometric characterization for convex inequity measures. Moreover, we present generic solution approaches for equity-promoting optimization problems with a convex inequity measure objective or constraint. Finally, we provide stability results on the choice of convex inequity measures in the objective of optimization models. Our numerical results show the computational efficiency of our approaches over existing approaches.

2. Bridging Two Fairness Perspectives: Group Parity Metrics and Social Welfare Functions

Violet (Xinying) Chen[1,*], *John Hooker*[2], *Derek Leben*[2]

[1]Stevens Institute of Technology; [*]vchen3@stevens.edu;
[2]Carnegie Mellon University;

Statistical parity metrics have been widely studied and endorsed as a means of achieving group fairness, but they present both technical and philosophical problems. Technically, they are often incompatible with each other and criteria for individual fairness. Philosophically, they disregard the actual utility consequences of decisions, provide no guidance on which groups should be protected, and lack convincing justification for selecting one metric rather than another. This paper explores whether a broader conception of social justice, based on a social welfare function (SWF), can address these issues. We derive implications of the well-known alpha fairness SWF for demographic parity, equalized odds, and predictive rate parity. The goal is to replace conflicting intuitions about parity with a justice principle that reflects decades of study in welfare economics. One model investigates whether achieving alpha fairness in the population as a whole results in parity between groups. Another determines the extent to which alpha fairness can justify a special policy for a protected group. We find that alpha fairness can justify demographic parity and equalized odds under fairly weak conditions, while it provides no justification for predictive rate parity. Proportional fairness (Nash bargaining), a special case of alpha fairness, achieves demographic parity and equalized odds across all groups simultaneously under certain conditions.

3. Equity-promoting Integer Programming Approaches For Medical Resident Rotation Scheduling

Shutian Li[1,*], *Karmel S. Shehadeh*[1], *Beth Hochman*[2], *Jacob Krimbill*[2], *Alexander P. Kossar*[2]

[1]Lehigh University; [*]shl919@lehigh.edu; [2]Columbia University Medical Center;

Newly graduated physicians from medical schools often join certified residency programs to fulfill specialty board certification requirements. Residents rotate through various clinical settings during their residency to gain the necessary training. Manually constructing the annual rotation schedule is challenging and laborious. Moreover, manual methods often produce inequitable schedules. To address these challenges, we propose new equity-promoting integer programming approaches for rotation scheduling. Numerical experiments based on a case study from a residency program illustrate the potential of the proposed approaches in automating the resident-to-rotation scheduling process and improving equity among residents and their satisfaction.

■ Bilevel Optimization

Room: **RB 85** (9:45 – 11:15am) Chair: *Federico Battista*

1. Generating Improving Solutions/Directions for Mixed Integer Bilevel Linear Problems

Federico Battista[1,*], *Ted Ralphs*[1]

[1]Lehigh University; [*]feb223@lehigh.edu;

Mixed integer bilevel linear problems (MIBLPs) involve the optimization of the strategy of a leader and the subsequent reaction of a follower. Despite the theoretical difficulty, branch-and-cut frameworks represent a successful tool in practice. Given a solution of a relaxation, the cut generation for MIBLPs often involves the resolution of a mixed integer linear subproblem to identify either an improving solution or an improving direction. In this talk we discuss the connections between the two approaches and we show how to exploit the information arising from the cut generation to enhance the feasibility check of a solution.

2. Robust solutions to mixed integer bilevel (two-stage) linear optimization with bounded rationality

Yu Xie[1,*], *Ted K. Ralphs*[1], *Oleg Prokopyev*[2]

[1]Lehigh University; [*]yux616@lehigh.edu; [2]University of Pittsburgh;

Mixed integer bilevel (or two-stage) linear optimization problems, in which a leader makes an initial decision and a follower reacts, arise in various important applications. It is typically assumed that the follower is rational, i.e., the follower's reaction is a globally optimal solution to the corresponding second stage optimization problem. Our work focuses on the case of bounded rationality, in which the follower is allowed to react in a sub-optimal manner. For example, the follower may not be capable of producing globally optimal solutions due to some computational limitations, making the response hard to predict. Therefore, a leader may need to seek decisions that are robust against a range of possible follower actions. We propose an approach based on what we call a pessimistic risk function, which captures the risk-averse objective of the leader. The proposed approach involves reformulations of the original problem and a solution method based on a branch-and-cut framework. We present preliminary computational results demonstrating the robustness of the solutions produced using this approach, compared to the traditional alternative.

3. On the stability of a bilevel optimisation program with a convex lower level problem

Emanuele Pizzari[1,2,*], *Massimiliano Caramia*[2]

[1]Lehigh University; [*]emp323@lehigh.edu; [2]University of Rome Tor Vergata;

Several managerial issues require the involvement of different decision-makers. Multi-level optimisation is a suitable programming approach whenever a specific hierarchy is present between these decision-makers. We focus on the case with two decision-makers, i.e., bilevel optimisation. In this kind of model, the upper level is the leader, while the lower level is the follower. The leader acts first, while the follower reacts to their decision. Thus, the leader is affected by the follower's decision, whether in the constraints and/or the objective function. Finding a solution to these models is quite complex; even in its simplest form, a bilevel optimisation model is NP-hard. Moreover, the model could be unstable. Indeed, the follower could have several equivalent solutions to adopt; these solutions may have different effects on the leader. The two extremes in this set of solutions are the optimistic and the pessimistic solutions, i.e., the best and the worst outcomes for the leader, respectively. If the difference between these solutions is quite high, the decision-maker obtains very little information value from solving the model. Therefore, our objective is to provide an algorithm to scout the possible points of instability and provide the leader with the option of adjusting their initial solution in order to decrease and possibly eliminate the instability of the model. The now stable solution is then compared to the previous unstable solutions to gauge the effectiveness. The analysis is conducted on the scenario where the follower problem is convex.

■ Advances in Mixed-Integer Programming and Max-Cut Problems

Room: **RB 91** (9:45 – 11:15am) Chair: *Nimita Shinde*

1. Warm Starting Series of Mixed Integer Linear Programs with Fixed Dimensions via Disjunctive Cuts

Sean Kelley[1,*], *Aleksandr M. Kazachkov*[2], *Ted Ralphs*[1]

[1]Lehigh University; [*]sek519@lehigh.edu; [2]University of Florida;

Many applications involve solving a sequence of closely related mixed integer linear programs (MILPs) that all share the same variables and number of constraints. In such cases, disjunctive cuts generated while solving one instance may prove to be effective at improving solver performance when parameterized and applied to unsolved instances later in the sequence. In this work, we detail how to find and parameterize disjunctive cuts and test disjunctive cuts' effectiveness when applied to later instances in series of MILPs with fixed dimension.

2. An SU(2)-symmetric Semidefinite Programming Hierarchy for Quantum MaxCut

Cunlu Zhou[1,*], *Jun Takahashi*[1], *Chaithanya Rayudu*[1], *Robbie King*[2], *Kevin Thompson*[3], *Ojas Parekh*[3]

[1]University of New Mexico; [*]cunlu.zhou@gmail.com; [2]Caltech; [3]Sandia National Laboratories;

Understanding and approximating extremal energy states of local Hamiltonians is a central problem in quantum physics and complexity theory. Recent work has focused on developing approximation algorithms for local Hamiltonians, and in particular the "Quantum Max Cut" (QMaxCut) problem, which is closely related to the antiferromagnetic Heisenberg model. In this talk, I will present a new hierarchy of semidefinite programming (SDP) relaxations for the QMaxCut, which is a tailored Navascues-Pironio-Acin (NPA) hierarchy for QMaxCut by taking its SU(2) symmetry into account. We prove that for sufficiently large levels of the hierarchy the optimal value converges to the correct QMaxCut value. Our results may also be viewed as a characterization of the algebra of SWAP operators. We give several analytic proofs and computational results showing exactness/inexactness of our hierarchy at the lowest level on several families of graphs.

I will also discuss relationships between SDP approaches for QMaxCut and frustration-freeness in condensed matter physics. We numerically demonstrate that the SDP-solvability practically becomes an efficiently-computable generalization of frustration-freeness. Furthermore, we show the potential of SDP algorithms to perform as an approximate method to compute physical quantities and capture physical features of some Heisenberg-type statistical mechanics models even away from the frustration-free regions.

3. Memory-efficient approximation algorithm for Max-Cut, Max-k-Cut and Correlation Clustering

Nimita Shinde[1,*], *Vishnu Narayanan*[2], *James Saunderson*[3]

[1]Lehigh University; [*]nimitashinde25@gmail.com; [2]Indian Institute of Technology Bombay; [3]Monash University;

Semidefinite programs are a special class of convex optimization problems that have a wide range of applications. We discuss three fundamental graph partitioning problems Max-Cut, Max-k-Cut, and the Max-Agree variant of correlation clustering for a given graph G=(V, E). For large-scale instances of problems, the memory required to solve SDPs becomes a key computational bottleneck. We discuss the application of a Gaussian sampling-based technique to the three combinatorial optimization problems.
We show that by applying our approach to the three problems, we achieve nearly the same approximation guarantees as the best-known results, while the memory used by our method(in addition to the memory required to store the problem instance) is O(|V|) for Max-Cut and O(|V|+|E|) for Max-k-Cut and Max-Agree that uses O(|V|+|E|).

■ Advances in Nonconvex Optimization

Room: **RB 241** (9:45 – 11:15am) Chair: *Wei Liu*

1. Gaussian smoothing gradient descent for high-dimensional nonconvex optimization

Andrew Starnes[1,*]

[1]Lirio; [*]astarnes@lirio.com;

This talk will focus on Gaussian smoothing gradient descent (GSGD) approaches for minimizing high-dimensional nonconvex functions. In this setting we will detail how certain conditions on the smoothing parameter enable GSGD to avoid the all too frequent situation where an optimization algorithm becomes trapped in a local extrema. In addition, even though there has been significant efforts aimed at solidifying the theoretical foundation of GSGD, often restrictive assumptions about the function are made, e.g., strongly convex or Lipschitz continuity of both the target function and its derivative. We will show that having Lipschitz continuity of either the function or its derivative will still allow for convergence to a minimum, which is a global minimum if the target function is convex. Moreover, results showing how close the minimizers of the smoothed functions are to minimizers of the target function will also be provided. Several numerical will exemplify the theoretical results and will be used to explain the advantages and challenges associated with GSGD and its variants. Along the way, we will explicitly explain the "deep" connections of Gaussian smoothing to a variety of topics including the heat equation (or general Hamilton-Jacobi equations), homotopy continuation, and Laplacian smoothing.

2. Multisecant Quasi-Newton methods

Mokhwa Lee[1,*]

[1]Stony Brook University; [*]mokhwa.lee@stonybrook.edu;

When dealing with a large-scale optimization problem, classical second-order methods, such as Newton's method, are no longer practical because it requires iteratively solving a large-scale linear system of order n. For this reason, Quasi-Newton(QN) methods, like BFGS or Broyden's method, are introduced because they are more efficient than Newton's method. This project focuses on multi-secant extensions of the BFGS method, to improve its Hessian approximation properties. Unfortunately, doing so sacrifices the matrix estimate's positive semidefiniteness, and steps are no longer assured to be descent directions. Therefore, we apply a perturbation strategy to construct an almost-secant positive-definite Hessian estimate matrix. This strategy has a low computational cost, involving only rank-2 updates with variable and gradient successive differences. We also explore several ways of improving this method, accepting and rejecting older updates according to several non-degeneracy metrics. Future goals include extending these techniques to limited memory versions.

3. Lower Complexity Bound of First-order Methods for Affinely Constrained Nonconvex Nonsmooth Problems

Wei Liu[1,2,*], *Qihang Lin*[2], *Yangyang Xu*[1]

[1]RPI; [*]liuwei175@lsec.cc.ac.cn; [2]University of Iowa;

In the study of affinely constrained non-convex non-smooth problems, the upper-bound complexity results of first-order methods have been extensively researched. However, this paper takes a different approach by determining lower-bound complexity results of first-order methods for large-scale affinely constrained non-convex non-smooth problems. Our results apply to first-order methods whose iterates are in the linear span of past first-order information. We prove that for the affinely constrained non-convex non-smooth problems, first-order methods have a lower-bound complexity result of $O(\epsilon^{-2})$ of finding a ϵ-stationary point.
Additionally, we design a new inexact proximal gradient algorithm and show that it is worst-case optimal within its function class. Hence our lower bounds are tight to a class of affinely constrained non-convex non-smooth problems.

■ Optimization and Applications II

Room: **RB 271** (9:45 – 11:15am) Chair: *Tuyen Tran*

1. Real-Time Personalized Order Holding

Mohammad Reza Aminian[1,*], *Will Ma*[2], *Linwei Xin*[1]

[1]University of Chicago; [*]maminian@chicagobooth.edu; [2]Columbia University;

This paper introduces a parsimonious online decision-making problem to capture this trade-off. In our model, there is a fixed capacity c on the total number of orders that can be held in the pool. Orders i arrive over a continuous time horizon, each with a lifespan li and a holding reward hi. Lifespan li specifies the maximum duration for which order i can be held, due to the promised delivery time. Holding reward hi (called height interchangeably) specifies the rate at which value is accumulated by the e-retailer for holding the order, capturing the chance of multiorder. We note that our model also abstractly captures the benefits accrued from holding orders that may be canceled, or delaying the fulfillment decision for that order. If the system is at full capacity and a new order arrives, then a decision must be made on which customer's order to dispatch, possibly the order that has just arrived. The objective is to maximize the total holding reward.
We propose several online algorithms that make these dispatching decisions without requiring knowledge of future order arrivals. They all yield prescriptive heuristics on managing the virtual pool, and come with "competitive ratio" (CR) guarantees on the reward collected relative to that of the optimal dispatching decisions in hindsight (with full knowledge of the arrival sequence).

2. Evaluation of Manual Assembly Line Balancing Problem in Flexible Manufacturing and Uncertain Operator Performance

Liam J Cahalane[1,*], *Thomas Mazzuchi*[1], *Shahram Sarkani*[1]

[1]The George Washington University; *lcahala@gwu.edu;

Uncertainty in manual assembly operations can significantly impact manufacturing processes, resulting in late deliveries and potential loss of revenue. This research addresses the challenge of optimizing uncertain task durations of a manual assembly line in flexible manufacturing to minimize lateness and maximize revenue. We propose an approach to assigning operators to assembly tasks, without prior knowledge of their performance, which is periodically reevaluated to ensure optimal assignment and balance of the assembly line. Our approach considers the uncertain and variable nature of manual assembly operations, and leverages the Assembly Line Balancing Problem framework to achieve optimal efficiency. We demonstrate the effectiveness of our approach through numerical simulations and provide recommendations for implementation in manufacturing processes.

3. Solving multifacility Location Problems Based on Mixed Integer Programming

Tuyen Tran[1,*], *Anuj Bajaj*[2], *Mau Nam Nguyen*[3], *Boris Mordukhovich*[4]

[1]Loyola University Chicago; *ttran18@luc.edu; [2]Mercy Hurst University; [3]Portland State University ; [4]Wayne State University;

The talk introduces a new approach to solve multifacility location problems based on mixed integer programming and algorithms for minimizing differences of convex (DC) functions. This class of multifacility location problems is very difficult to solve because of its intrinsic discrete, nonconvex, and nondifferentiable nature. We first reformulate the problem under consideration as a continuous optimization problem then develop a new DC type algorithm involving Nesterov's smoothing. We also implement our method with MATLAB, numerical tests are done on both artificial and real data sets.

■ Advances in Constrained Composite Optimization

Room: **RB 85** (11:30am – 1:00pm) Chair: *Dimitri Papadimitriou*

1. A single-loop Lagrangian method for nonconvex optimization with nonlinear constraints

Dimitri Papadimitriou[1,*]

[1]3nLab; *dpapadimitriou@3nlab.org;

The augmented Lagrangian-based (ALM) method is one of the most common approaches for solving linearly and nonlinearly constrained problems. However, for nonconvex objectives, the handling of nonlinear inequality constraints remains challenging. Under very general assumptions, we propose a single loop ALM-based algorithmic scheme with Backtracking Line Search to solve nonconvex optimization problems including both nonlinear equality and inequality constraints. When some variables belong to the real Hilbert space and others to the integer space, we obtain a computational method for solving nonconvex mixed-integer/-binary nonlinear problems for which the nonconvexity is not limited to the integrality constraints, i.e., both nonlinear equality and inequality constraints may be nonconvex. We demonstrate the local convergence properties as well as the iteration complexity (i.e., the number of iterations required to obtain an approximate KKT point) of the proposed algorithmic scheme. The performances of the proposed method are then numerically evaluated when solving a multi-constrained network design problem with nonconvex constraints. For this purpose, extensive numerical executions on a set of instances extracted from the SNDlib repository are performed to analyze the behavior and performance of the algorithm as well as to identify potential improvements. Finally, analysis of the results and their comparison against those obtained when solving its convex relaxation using mixed-integer programming (MIP) solvers are reported.

2. A primal-dual splitting method for nonlinear composite problems

Bang Vu[1,*], *Dimitri Papadimitriou*[1]

[1]3nLab; *bangcvvn@3nlab.org;

This paper investigates a structured nonconvex optimization problem including a nonlinear composition of form f(x) + g(x) + h(c(x)-b). For this purpose, wee propose a primal-dual splitting method. At each iteration, the nonsmooth components (f and g) are evaluated by means of their proximity operators: the nonconvex smooth component (h) via its gradient, the nonlinear operator (c) via its values, separately. We characterize the convergence properties of the proposed methods in terms of gradient mapping and feasibility and show that, under mild conditions, the gradient mapping and feasibility of the generated iteration converge to 0. A stochastic variant of the proposed method is also proposed for solving various large-scale problems ranging from data science to statistical learning.

3. Inexact proximal augmented Lagrangian method for compositional problems with convex constraints

Yangyang Xu[1,*]

[1]RPI; *xuy21@rpi.edu;

In this talk, I will present a first-order method for solving nonconvex composition problems with convex affine and nonlinear constraints. Our method is based on the framework of the proximal augmented Lagrangian method. Under Slater's condition, I will show that our method can produce an ϵ KKT point of the considered problem with $O(1/\epsilon^{2.5})$ oracle complexity if the objective is smooth, and with $O(1/\epsilon^{2.5})$ oracle complexity if the objective is in a compositional form. Numerical results on linear and quadratic constrained quadratic programming and on a robust nonlinear least square fitting problem will be shown.

■ Healthcare Optimization

Room: **RB 91** (11:30am – 1:00pm) Chair: *M. Mohammadisiahroudi*

1. Treatment Planning Optimization for Proton Therapy

Mohammadhossein Mohammadisiahroudi[1,*], *Wei Zou*[2], *Yuriy Zinchenko*[3], *Tamás Terlaky*[1]

[1]Lehigh University; *mom219@lehigh.edu; [2]University of Pennsylvania; [3]University of Calgary;

Radiation therapy (RT) has been well-established as an essential tool to treat cancer. In RT, a team optimizes treatment plans by modulating external beam intensity to achieve the best medical outcome. Finding optimal machine configurations, to get maximal coverage of the tumor and minimize damage to critical and healthy organs, is an important optimization problem. Proton Therapy is a new technology, using protons instead of X-Ray, which enables high-precision treatment of the tumor close to critical organs. In this research, we experiment with different optimization models for proton therapy and evaluate their performance. In addition, an efficient successive relaxation technique is proposed to solve the mixed integer formulation. The numerical results show a high level of conformality to DVH goals.

2. Supervised Inverse Optimization

Felix Parker[1,*], *Kimia Ghobadi*[1]

[1]Johns Hopkins University; *fparker9@jhu.edu;

Inverse Optimization (IO) aims to infer the parameters of an optimization problem from observed outputs of the decision-making process. Many practical applications of IO involve multiple observed decisions which can be sub-optimal or infeasible, yet existing methods for this setting make strong assumptions that often do not hold in practice. In particular, these methods assume that observations are approximately optimal for the true forwards optimization problem, even though the parameters of this optimization problem are unknown. The result is that traditional IO approaches may be overly conservative and biased towards what was done in the past. At the same time, machine learning approaches cannot guarantee optimal or even feasible decisions. To address these limitations, we propose Supervised Inverse Optimization (SIO), which aims to meld existing approaches for IO with supervised learning by incorporating scores for each observation, and learning the objective function parameters through a model that optimizes IO and regression objectives. The observation scores represent partial information about the objective function, which is often available, and reduces the reliance on assumptions, which improves performance. The model is highly flexible and can model problems with linear or non-linear objectives, objectives with unknown structure, multiple scores per observation, and many other problem settings. SIO is valuable in real-world applications including diet recommendation, resource allocation, optimizing clinical pathways, and improving radiation therapy plans. We demonstrate this with empirical results for the data-driven diet recommendation problem and other challenging IO problems.

3. BiomedGPT: A Unified and Generalist Biomedical Generative Pre-trained Transformer for Vision, Language, and Multimodal Tasks

Jun Yu[1,*]

[1]Lehigh University; *juy220@lehigh.edu;

Recent large-scale Transformers make a unified model capable of performing miscellaneous tasks, via pretraining on a diverse collection of multi-modal datasets followed by finetuning on a specific downstream task. Existing pre-trained models have shown promising performance on a variety of tasks in healthcare. However, a tailored architecture is always required for each task in previous works. This paper aims to bridge the gap of unifying different medical tasks into an omnipotent model. Under this challenging setting, we present the first Biomedical Generative Pre-trained Transformer (dubbed BiomedGPT) for versatile tasks in vision, language, and multi-modality data, including bioimaging classification, medical captioning, pathological visual question answering, etc. Compared with previous SOTAs, our experiments on approximately 20 public medical datasets show the possibility of comprehensive representations in the medical area. The further zero-shot analysis indicates BiomedGPT can effectively transfer knowledge to unseen tasks or domains. Through the ablation study, we also showcase the efficacy of our multi-modal and multi-task pretraining approach in transferring knowledge. Overall, our work presents a significant step forward in developing unified and generalist models for biomedicine, with far-reaching implications for improving healthcare outcomes.

■ Advances in Continuous Optimization for Learning Problems

Room: **RB 241** (11:30am – 1:00pm) Chair: *César A. Uribe*

1. Applications of DC Programming to bilevel hierarchical clustering problem

Tuyen Tran[1,*], *Mau Nam Nguyen*[2], *Samuel Reynolds*[2], *Wondi Geremew*[3]

[1]Loyola University Chicago; *ttran18@luc.edu; [2]Portland State University; [3]Stockton University;

Multilevel hierarchical clustering has a long history and enormous important applications in data mining and statistics. In this talk, we consider a different formulation of the bilevel hierarchical clustering problem, a commonly used model in designing optimal multicast networks and a discrete optimization problem which can be shown to be NP-hard. Our approach is to reformulate the problem as a continuous optimization problem by making some relaxations on the discreteness conditions. Then, Nesterov's smoothing technique and a numerical algorithm for minimizing difference of convex functions called the DCA are applied to cope with the nonsmoothness and nonconvexity of the problem. Numerical examples are provided to illustrate our method.

2. On the Performance of Gradient Tracking with Local Updates

Edward Duc Hien Nguyen[1,*], *Sulaiman A. Alghunaim*[2], *Kun Yuan*[3], *César A. Uribe*[1]

[1]Rice University; *en18@rice.edu; [2]Kuwait University; [3]Peking University;

We study the decentralized optimization problem where a network of n agents seeks to minimize the average of a set of heterogeneous non-convex cost functions distributedly. State-of-the-art decentralized algorithms like Exact Diffusion and Gradient Tracking (GT) involve communicating every iteration. However, communication is expensive, resource intensive, and slow. This work analyzes a locally updated GT method (LU-GT), where agents perform local recursions before interacting with their neighbors. While local updates have been shown to reduce communication overhead in practice, their theoretical influence has not been fully characterized. We show LU-GT has the same communication complexity as the Federated Learning setting but allows for decentralized (symmetric) network topologies and prove that the number of local updates does not degrade the quality of the solution achieved by LU-GT.

3. On First-Order Meta-Reinforcement Learning with Moreau Envelopes

César A. Uribe[1,*], *Taha Toghani*[1], *Sebastian Perez-Salazar*[1]

[1]Rice University; *cauribe@rice.edu;

Meta-Reinforcement Learning (MRL) is a promising framework for training agents that can quickly adapt to new environments and tasks. In this work, we study the MRL problem under the policy gradient formulation, where we propose a novel algorithm that uses Moreau envelope surrogate regularizers to jointly learn a meta-policy that is adjustable to the environment of each individual task. Our algorithm, called Moreau Envelope Meta-Reinforcement Learning (MEMRL), learns a meta-policy that can adapt to a distribution of tasks by efficiently updating the policy parameters using a combination of gradient-based optimization and Moreau Envelope regularization. Moreau Envelopes provide a smooth approximation of the policy optimization problem, which enables us to apply standard optimization techniques and converge to an appropriate stationary point. We provide a detailed analysis of the MEMRL algorithm, where we show a sublinear convergence rate to a first-order stationary point for non-convex policy gradient optimization. We finally show the effectiveness of MEMRL on a multi-task 2D-navigation problem.

■ Optimization and Applications III

Room: **RB 271** (11:30am – 1:00pm) Chair: *Billy Jin*

1. Probabilistic Submodularity of Maximizing Anticoordination in Finite Network Games

Soham Das[1,*], *Ceyhun Eksinq*[1]

[1]Texas A&M University; *soham.das@tamu.edu;

We consider the control of decentralized learning dynamics for agents in an anti-coordination network game. In the anti-coordination network game, there is a preferred action in the absence of neighbors' actions, and the utility an agent receives from the preferred action decreases as more of its neighbors select the preferred action, potentially causing the agent to select a less desirable action. The decentralized dynamics that correspond to the elimination of dominated strategies converge for the considered game. Given a convergent action profile, we measure anti-coordination by the number of edges in the underlying graph that have at least one agent on either end of the edge not taking the preferred action. The maximum anti-coordination (MAC) problem seeks to find an optimal set of agents to control under a finite budget so that overall network disconnect is maximized on game convergence. We show that MAC is submodular in expectation with high probability in bipartite graphs, where the probability value is lower bounded by a function of the graph topology. The proof relies on characterizing well-behavedness of a class of MAC instances and proving probabilistic submodularity of the one step influence function for the dynamics. Thereby we design a coupling between the dynamics and another distribution preserving selection protocol, for which we can show the diminishing returns property. Utilizing this result, we obtain a performance guarantee for the greedy optimization of MAC. Finally, we provide a computational study to show the effectiveness of greedy node selection strategies to solve MAC on bipartite networks.

2. Stochastic Programming Inspired Modeling Techniques for Shaping Dynamic Trajectories

Joshua Pulsipher[1,*]

[1]University of Waterloo; *pulsipher@uwaterloo.ca;

Infinite-dimensional optimization (InfiniteOpt) captures problems that arise in both stochastic and dynamic optimization where the decision variables are often indexed over a continuous domain (i.e., are functions). In previous work, I developed a unifying modeling abstraction for InfiniteOpt problems which establishes unified modeling objects to capture formulations in stochastic optimization, dynamic optimization, PDE-constrained optimization, and combinations. This is conveniently implemented in the Julia package InfiniteOpt.jl. Through the lens of this abstraction, we have found that continuous-time dynamic optimization problems can be interpreted as a special case of two-stage stochastic programs. In this talk, I will present how this perspective has inspired the transfer modeling objects from stochastic programming to be used in a deterministic dynamic context. Namely, I will discuss how time-valued analogs of risk measures and chance constraints (called event constraints in this context) enable unique ways to shape dynamic trajectories. The use of these new modeling objects will be demonstrated through illustrative case studies.

3. Optimal Tradeoffs for Two-Stage Bipartite Matching with Advice

Billy Jin[1,*], *Will Ma*[2]

[1]Cornell University; *bzj3@cornell.edu;
[2]Columbia University;

We study two-stage vertex-weighted bipartite matching problem with advice. We evaluate an algorithm by its robustness, which is its performance relative to that of the optimal offline matching, and its consistency, which is its performance relative to that of the advice. We characterize the tight robustness-consistency tradeoff for this problem.

Index of Authors