THE 2023 MIDWEST OPTIMIZATION MEETING

October 21, 22, 2023

University of Michigan, Ann Arbor, MI, USA

The meeting is sponsored by *Michigan Center for Applied and Interdisciplinary Mathematics* (MCAIM, https://sites.lsa.umich.edu/mcaim/) and *National Science Foundation* (NSF).

Organizers: Albert Berahas, Ilya Kolmanovsky, Boris Mordukhovich, Nguyen-Truc-Dao Nguyen, Ebrahim Sarabi, Vasileios Tzoumas, Bingwu Wang.

Acknowledgements: Anthony Bloch (MCAIM liaison), Wanping Dong, Miguel Fernandez, Khristine Henson-Jones, Zhongming Jiang, Scott Klunder, Atharva Navsalkar, Trang Nguyen, Christina Ruhfel, Lauren Russell, Jiahao Shi, Yu Shi, Alissa Smith, Ray Xu, Hongyu Zhou.

Website: https://emunix.emich.edu/∼bwang/MOM2023/index.html

The Schedule of the 2023 Midwest Optimization Meeting:

DAY 1: Saturday, October 21, 2023 **DAY 2:** Sunday, October 22, 2023

EAST HALL 1360 (Session 1), **1068** (Session 2), **1084** (Session 3)

DAY 1		**Session 1**	**Session 2**	**Session 3**
7:45AM	Welcoming desk open (badges)			
8:30AM	Coffee			
9:00-10:00	Plenary Talk	Terry Rockafellar		
10:00-10:20	Coffee Break			
10:20-10:50		Nghia Tran	Marcia Fampa	Shagun Gupta
10:50-11:20		Ruiwei Jiang	Matt Menickelly	Jiahao Shi
11:20-11:50		Walaa Moursi	Lewei Zhao	Sachin Garg
11:50-12:20		Yunier Bello Cruz	Olga Brezhneva	Jianhao Ma
12:20-1:50	Boxed Lunch			
1:50-2:50	Plenary Talk	Stephen J. Wright		
3:00-3:30		Zhaosong Lu	Miaolan Xie	Thanh Phat Vo
3:30-4:00		Courtney Paquette	Oliver Hinder	Khoa Vu
4:00-4:30		Yunsoo Ha	Dat Tran	Viet Le
4:30-4:50	Coffee Break			
4:50-5:50	Plenary Talk	Quim Martins		
6:30-9:00	Dinner/Banquet, Cash Bar			

DAY 2		**Session 1**	**Session 2**	**Session 3**
7:45AM	Welcoming desk open			
8:30AM	Coffee			
9:00-10:00	Plenary Talk	Irena Lasiecka		
10:00-11:00	Plenary Talk	Daniel Liberzon		
11:00-11:20	Coffee Break			
11:20-11:50		Bao Truong	Samara Chamoun	Nouman Khan
11:50-12:20		Ali Mohammad Nezhad	Trang Nguyen	
12:20-1:50	Boxed Lunch			
1:50-2:20		Tuyen Tran	Jordan Leung	Fadi Hamad
2:20-2:50		Duy Nhat Phan		Syeda Atik
2:50-3:20		Thi Phung	Wanping Dong	Sayma Sultana
3:20-3:30	Closing Remarks			
3:30-4:00	Coffee			

The Program of the 2023 Midwest Optimization Meeting:

DAY 1: Saturday, October 21, 2023 **DAY 2:** Sunday, October 22, 2023
EAST HALL 1360 (Session 1), **1068** (Session 2), **1084** (Session 3)

DAY 1:

- **7:45AM:** Welcoming desk open
- **8:30AM:** Coffee
- **9:00–10:00,** Plenary Talk:
  - **Generalized Nash equilibrium from a variational analysis perspective**

    Terry Rockafellar, Department of Mathematics, University of Washington–Seattle

    In a generalized Nash equilibrium as a multi-agent model of continuous optimization, the objective function and the constraint set for each agent can depend on the decisions made by the other agents. Understanding how such an equilibrium may respond to shifts in model parameters is a key topic for many reasons, but raises many questions. Answers can be found with the tools of variational analysis in a reformulated framework of equilibrium with extended-real-valued functions and their subgradients, where assumptions of convexity or quasi-convexity can be relaxed with global optimality replaced by local optimality.

    Concepts of tilt stability and full stability can be introduced and, to some extent, be characterized by conditions on partial coderivatives. However, single-valued Lipschitz continuity of the solution mapping must generally be relaxed to set-valued Aubin continuity. Important issues come up then about what should really be meant by a“good” equilibrium model.
- **10:00–10:20:** Coffee Break
- **10:20–10:50:**
  - (session 1)

    Strong minima of low-Rank optimization problems

    Nghia Tran, Oakland University

    The theory developed for low complexity exact recovery is usually based on sharp minima. In this talk, we show that strong minima occurs naturally and plays important roles in exact/robust recovery of a low-rank matrix. Many geometric and quantitative characterizations for strong minima of low-rank minimization problems are obtained. Consequently, we derive the least bound for number of observations in exact recovery via the rank of the original matrix that we wish to recover. This bound is much smaller than what is known in the literature.

- (session 2)
  Branch-and-bound for D-Optimality with fast local search and variable-bound tightening

  Marcia Fampa, Federal University of Rio de Janeiro

  We develop a branch-and-bound algorithm for the D-optimality problem, a central problem in statistical design theory, based on two convex relaxations, employing variable-bound tightening and fast local-search procedures, testing our ideas on randomly-generated test problems.

  Joint work with Gabriel Ponte (UFRJ) and Jon Lee (University of Michigan).

- (session 3)
  Balancing communications and computations in gradient tracking algorithms for decentralized optimization

  Shagun Gupta,
  Operations Research and Industrial Engineering, University of Texas at Austin

  In this talk, we present a unified framework for the various gradient tracking algorithms present in literature. We introduce flexibility with respect to communication structure and number of communication and computation steps performed each iterations over this framework. We theoretically differentiate among the different gradient tracking algorithms. We provide convergence conditions and theoretically show improvement with the introduced flexibility. We illustrate the improvements on quadratic functions and binary classification problems.

- **10:50–11:20:**

  - (session 1)
    Convex chance-constrained programs with Wasserstein ambiguity

    Ruiwei Jiang, Industrial and Operations Engineering, University of Michigan

    As a natural approach to modeling system safety conditions, chance constraints (CC) seek to satisfy a set of uncertain inequalities individually or jointly with high probability, but the feasible regions they produce are non-convex in general. In this talk, we discuss various sufficient conditions that allow CC with Wasserstein ambiguity to produce a convex feasible region.

  - (session 2)
    Exploiting structure in (derivative-free) composite nonsmooth optimization

    Matt Menickelly, Argonne National Laboratory

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

  - (session 3)
    Subsampled second-order SQP method with modified line search for equality constrained stochastic optimization

    Jiahao Shi, Industrial Operations and Engineering, University of Michigan

    In this paper, we propose a subsampled second order algorithms for solving equality constrained optimization problems in which the objective is defined by the expectation of a stochastic function and constraint functions are deterministic. We use line search methods in the algorithm and present the local and global convergence results. To fill the gap of the step size selection scheme in the analysis, we modify the algorithm with a new proposed line search condition. Finally, we develop an inexact algorithm to reduce computational cost.

- **11:20–11:50:**
  - (session 1)
    Range of the displacement operator of PDHG with applications to quadratic and conic programming
    Walaa Moursi, University of Waterloo

    Primal-dual hybrid gradient (PDHG) is a first-order method for saddle-point problems and convex programming introduced by Chambolle and Pock. Recently, Applegate et al. analyzed the behavior of PDHG when applied to an infeasible or unbounded instance of linear programming, and in particular, showed that PDHG is able to diagnose these conditions. Their analysis hinges on the notion of the infimal displacement vector in the closure of the range of the displacement mapping of the splitting operator that encodes the PDHG algorithm. In this paper, we develop a novel formula for this range using monotone operator theory. The analysis is then specialized to conic programming and further to quadratic programming (QP) and second-order cone programming (SOCP). A consequence of our analysis is that PDHG is able to diagnose infeasible or unbounded instances of QP and of the ellipsoid-separation problem, a subclass of SOCP.
  - (session 2)
    Optimization applications in radiation treatment planning
    Lewei Zhao, Department of Radiation Oncology, Stanford University

    Radiation therapy plays a pivotal role in the treatment of cancer, offering the potential for precise and effective tumor targeting while sparing healthy tissues. It is an optimization problem that we want to maximizing the killing to tumor and minimizing the side effect to normal tissue. In recent years, the field of radiation therapy has witnessed a paradigm shift, thanks to the integration of mathematical optimization methods. This talk will be an overview of treatment plan optimization, emphasizing the role of mathematical modeling and computational algorithms in refining radiation therapy plans to achieve the best possible balance between tumor coverage and normal tissue sparing. A series of work including applications of evolutionary algorithm, alternating direction method of multipliers, primal dual active set method, multi-criterial optimization and so on will be introduced.
  - (session 3)
    Exploiting Second-order information in conjunction with variance reduction promotes robustness
    Sachin Garg, EECS Department, University of Michigan

    Stochastic gradient methods (SG) have witnessed great success in solving convex finite-sum (n elements) optimization tasks such as empirical risk minimization but provide a sublinear convergence rate to the optimal point. Incorporating variance-reduced techniques accelerates the convergence of stochastic first-order methods, providing global linear convergence for smooth and strongly convex functions. However, the convergence rate of first-order variance-reduced stochastic methods deteriorates with an increase in the mini-batch sample size used for the stochastic gradient. This suggests the use of small batches for optimal convergence rate, resulting in highly sequential algorithms. Moreover, the literature on the benefits of variance reduction techniques to accelerate the popular second-order methods [Der22] is limited and requires further investigation.

In this work, we propose a stochastic second-order method with variance-reduced gradients which achieves the same fast convergence regardless of the mini-batch size, and takes advantage of variance reduction to provide improved local convergence rates (per data pass) compared to both first-order stochastic variance-reduced methods and stochastic second-order methods. Specifically, we prove that in the big data regime where the number of components n is much larger than the condition number κ, our algorithm achieves local linear convergence rate $(C\frac{k}{n}\log(n/\kappa))^t$ with high probability after t iterations, independent of the choice of the mini-batch size. We compare our method with Stochastic Variance Reduced Gradient (SVRG) [JZ13] and Subsampled Newton (SN) [RM19], and provide evidence that the proposed method achieves a faster convergence rate, and is more robust to the noise in the stochastic Hessian.

References:
[JZ13] Rie Johnson and Tong Zhang, *Accelerating stochastic gradient descent using predictive variance reduction*, Advances in neural information processing systems 26 (2013).
[RM19[Farbod Roosta-Khorasani and Michael W Mahoney, *Sub-sampled Newton methods*, Mathematical Programming 174.1 (2019), pp. 293-326.
[Der22[Michal Derezinski, *Stochastic Variance-Reduced Newton: Accelerating Finite-Sum Minimization with Large Batches*, arXiv preprint arXiv:2206.02702 (2022).

- **11:50–12:20:**
  - (session 1)
    On the circumcentered-reflection method for solving convex feasibility problems

    Yunier Bello Cruz, Northern Illinois University

    In this talk, we will introduce the Centralized Circumcentered-Reflection Method (cCRM), a novel iteration scheme designed to identify a point at the intersection of two closed convex sets without resorting to a product space reformulation. Unlike its predecessor, the CRM, the cCRM ensures convergence in general convex sets, showing both linear and superlinear convergence under certain conditions. Through a comparative analysis with the Method of Alternating Projections (MAP), we will demonstrate the superior performance of CRM and cCRM, backed by promising numerical experiments and theoretical proofs.
  - (session 2)
    Explicit formulas for degenerate systems of quadratic equations and QP problems

    Olga Brezhneva, Department of Mathematics, Miami University

    We describe an application of the p-regularity theory to nonlinear equations with quadratic mappings and quadratic programming (QP) problems. A special structure of the nonlinear equation and the construction of the 2-factor operator are used to obtain an exact formula for a solution to the nonlinear equation. A similar approach with using the 2-factor operator is applied to reduce the QP problem to a system of linear equations. The solution of this system represents a local minimizer of the QP problem along with its corresponding Lagrange multiplier. An explicit formula for the solution of the linear system is provided (this is the joint work with Agnieszka Prusinska (Poland) and Alexey Tret'yakov (Poland)).

- (session 3)

  Efficient optimization in the overparameterized regime

  Jianhao Ma, University of Michigan

  Overparameterization plays an extremely important role in modern machine learning. However, it is still unclear whether overparameterization affects underlining optimization. In this work, we identify some sufficient and necessary conditions under which overparameterization can benefit the optimization.

- **12:20–1:50:** Boxed lunch

- **1:50–2:50,** Plenary Talk:

  - **On Squared-variable formulations**

    Stephen J. Wright, University of Wisconsin-Madison

    We revisit a formulation technique for inequality constrained optimization problems that has been known for decades: the substitution of squared variables for nonnegative variables. Using this technique, inequality constraints are converted to equality constraints via the introduction of a squared-slack variable. Such formulations have the superficial advantage that inequality constraints can be dispensed with altogether. But there are clear disadvantages, not least being that first-order optimal points for the squared-variable reformulation may not correspond to first-order optimal points for the original problem, because the Lagrange multipliers may have the wrong sign. Extending previous results, this paper shows that points satisfying approximate second-order optimality conditions for the squared-variable reformulation also, under certain conditions, satisfy approximate second-order optimality conditions for the original formulation, and vice versa. Such results allow us to adapt complexity analysis of methods for equality constrained optimization to account for inequality constraints. On the algorithmic side, we examine squared-variable formulations for several interesting problem classes, including bound-constrained quadratic programming and linear programming. We show that algorithms built on these formulations are surprisingly competitive with standard methods. For linear programming, we examine the relationship between the squared-variable approach and primal-dual interior-point methods.

    This talk represents joint work with Lijun Ding.

- **3:00–3:30:**

  - (session 1)

    First-order methods for bilevel optimization

    Zhaosong Lu, Industrial Engineering, University of Minnesota

    Bilevel optimization has been widely used in a variety of areas such as adversarial training, hyperparameter tuning, image reconstruction meta-learning, neural architecture search, and reinforcement learning. In this talk, I will present first-order methods for solving a class of bilevel optimization through the use of single or sequential minimax optimization. The first-order operation complexity of the proposed methods will be discussed.

– (session 2)

Reliable adaptive stochastic optimization for messy data with high probability guarantees

Miaolan Xie, Cornell University, ORIE

To handle real-world data that is noisy, biased and even corrupted, we consider a simple adaptive framework for stochastic optimization where the step size is adaptively adjusted according to the algorithm's progress instead of manual tuning or using a pre-specified sequence. Function value, gradient and possibly Hessian estimates are provided by probabilistic oracles and can be biased and arbitrarily corrupted, capturing multiple settings including expected loss minimization in machine learning, zeroth-order and low-precision optimization. This framework is very general and encompasses stochastic variants of line search, quasi-Newton, cubic regularized Newton and SQP methods for unconstrained and constrained problems. Under reasonable conditions on the oracles, we show high probability bounds on the sample (and iteration) complexity of the algorithms.

– (session 3)

Coderivative-based Newtonian methods in nonsmooth and nonconvex optimization

Thanh Phat Vo, Department of Mathematics, Wayne State University

This talk proposes and justifies two globally convergent Newton-type methods to solve unconstrained and constrained problems of nonsmooth optimization by using tools of variational analysis and generalized differentiation. Both methods are coderivative-based and employ generalized Hessians (coderivatives of subgradient mappings) associated with objective functions, which are either of class C1,1, or are represented in the form of convex composite optimization, where one of the terms may be extended-real-valued. The proposed globally convergent algorithms are of two types. The first one extends the damped Newton method and requires positive-definiteness of the generalized Hessians for its well-posedness and efficient performance, while the other algorithm is of the regularized Newton type being well-defined when the generalized Hessians are merely positive-semidefinite. The obtained convergence rates for both methods are at least linear, but become superlinear under the semismooth property of subgradient mappings. Problems of convex composite optimization are investigated with and without the strong convexity assumption on smooth parts of objective functions by implementing the machinery of forward-backward envelopes. Numerical experiments are conducted for Lasso problems and for box constrained quadratic programs with providing performance comparisons of the new algorithms and some other first-order and second-order methods that are highly recognized in nonsmooth optimization (Based on the joint work with Pham Duy Khanh, Boris Mordukhovich and Dat Ba Tran).

- **3:30–4:00:**

– (session 1)

High-dimensional optimization

Courtney Paquette, Mathematics and Statistics, McGill University

We analyze the dynamics of streaming stochastic gradient descent (SGD) in the high-dimensional limit when applied to generalized linear models and multi-index models (e.g. logistic regression, phase retrieval) with general data-covariance. In particular, we demonstrate a deterministic equivalent of SGD in the form of a system of ordinary differential equations that describes a wide class of statistics, such as the risk and other measures of sub-optimality. This equivalence holds with overwhelming probability when the model parameter count grows proportionally to the number of data. This framework allows us to obtain learning rate thresholds for stability of SGD as well as convergence guarantees. In addition to the deterministic equivalent, we introduce an SDE with a simplified diffusion coefficient (homogenized SGD) which allows us to analyze the dynamics of general statistics of SGD iterates. Finally, we illustrate this theory on some standard examples and show numerical simulations which give an excellent match to the theory.

– (session 2)
Making SGD parameter-free
Oliver Hinder, Industrial Engineering, University of Pittsburgh
We develop an algorithm for parameter-free stochastic convex optimization (SCO) whose rate of convergence is only a double-logarithmic factor larger than the optimal rate for the corresponding known-parameter setting. In contrast, the best previously known rates for parameter-free SCO are based on online parameter-free regret bounds, which contain unavoidable excess logarithmic terms compared to their known-parameter counterparts. Our algorithm is conceptually simple, has high probability guarantees, and is also partially adaptive to unknown gradient norms, smoothness, and strong convexity. At the heart of our results is a novel parameter-free certificate for SGD step size choice, and a time-uniform concentration result that assumes no a-priori bounds on SGD iterates.

– (session 3)
Local maximal monotonicity in variational analysis and optimization
Khoa Vu, Wayne State University
The talk is devoted to a systematic study and characterizations of notions of local maximal monotonicity and their strong counterparts for set-valued mappings that appear in variational analysis, optimization, and their applications. We obtain novel resolvent characterizations of these notions together with efficient conditions for their preservation under summation in broad infinite-dimensional settings. Further characterizations of these notions are derived by using generalized differentiation of variational analysis in the framework of Hilbert spaces.

- **4:00–4:30:**

– (session 1)
Common random numbers and complexity in simulation optimization with adaptive sampling
Yunsoo Ha, Industrial and Systems Engineering, North Carolina State University
Adaptive sampling has an important history in improving the efficiency of stochastic optimization algorithms by balancing the stochastic error with the distance to optimality. In addition, simulation optimization has the useful feature of controlling the random number generation during the optimization process. In this talk we show that combining adaptive sampling with common random numbers has the potential to significantly improve the sample complexity. We discuss corresponding results in derivative-based and derivative-free settings within a trust-region framework.

– (session 2)
Inexact gradient methods for nonconvex functions with applications in nonsmooth optimization
Dat Tran, Department of Mathematics, Wayne State University

The talk introduces novel inexact gradient methods (IGD) for minimizing C1 -smooth functions. We show that the the methods achieve fundamental convergence properties including the stationarity of accumulation points, the global convergence and convergence rates under the KL property the objective functions. The newly developed IGD is applied to designing two novel gradient-based methods of nonsmooth optimization such as the inexact proximal point methods (GIPPM) and the inexact augmented Lagrangian method (GIALM) for convex programs with linear equality constraints. These two methods inherit global convergence properties from IGD and are confirmed by numerical experiments to have practical advantages over some well-known algorithms of nonsmooth convex optimization.

– (session 3)
Quadratic bundles of quadratic penalty functions and applications
Viet Le, Wayne State University

Quadratic bundles of functions, a newly proposed type of second-order derivative by Rockafellar (2022), is established as the collection of all possible epi-limits of second-order subderivatives at generalized twice-differentiable points. This talk involves some calculus rules and explicit computations for quadratic bundles of the quadratic penalty functions and possible applications. This class belongs to a wider well-known function class in Variational Analysis - convex piecewise linear-quadratic functions - which has been of much interest in recent developments. To achieve our goals, we first formulate the quadratic bundles for the indicator functions of polyhedral sets, and utilize some properties of quadratic bundles and the structure of quadratic penalty functions. Further investigations on applications are also provided in this talk.

- **4:30–4:50:** Coffee Break
- **4:50–5:50,** Plenary Talk:

– **Enabling large-scale multidisciplinary design optimization through the coupled-adjoint method**
Quim Martins, University of Michigan

Despite the progress in high-fidelity numerical simulations enabled by high-performance computing, challenges have remained in the use of these simulations for design optimization. This talk focuses on the developments that made it possible to perform high-fidelity design optimization of aircraft and other engineering systems. The challenges addressed include handling a large number of design variables, robust and efficient large-scale simulations, effective geometry and mesh handling, and efficient discipline coupling. To tackle these issues, we combine gradient-based optimization algorithms with adjoint gradient computation and develop an adaptive coupled Newton-Krylov approach to solve the coupled numerical simulations efficiently and robustly. The applications focus on aircraft design, including unconventional configurations. We first tackle wing design by coupling computational fluid dynamics to structural finite-element solvers and simultaneously optimizing the aerodynamic shape and structural sizing. The methods we developed to tackle this problem are generalized in the OpenMDAO framework, an open-source framework for multidisciplinary analysis and optimization. This and other open-source tools developed in this work open the door to further advances in algorithms and their application to aircraft design and beyond.

- **6:30–9:00:** Dinner/Banquet, Cash Bar

DAY 2:

- **7:45AM:** Welcoming desk open
- **8:30AM:** Coffee
- **9:00–10:00,** Plenary Talk:
  - **Can we control a flutter in flow-structure interactions? How and where?**

    Irena Lasiecka, University of Memphis

    Flow-structure interactions are ubiquitous in nature. Problems such as attenuation of turbulence or flutter in an oscillating structure [Tacoma bridge], flutter in tall buildings, fluid flows in flexible pipes, in nuclear engineering flows about fuel elements and heat exchanger vanes -are prime examples of relevant applications. Mathematically, the models are represented by a 3 D compressible, irrotational Euler Equation coupled to a nonlinear dynamic elasticity on a 2 D manifold. Strong boundary-type coupling at the interface between the two media is at the center of the analysis. This provides for a rich mathematical structure, opening the door to several unresolved problems in the area of nonlinear PDE's, dynamical systems, related harmonic analysis and differential geometry. This talk aims at providing a brief overview of recent developments in the area along with a presentation of some recent advances addressing the issues of control, optimization and long time behavior of such models. Of particular interest is the issue of optimizing flutter speed in a tur- bulent environment.

    Part of this talk is based on recent work with D. Bonheur, F. Gazzola and J. Webster: *Annales de L'Institute Henri Poincare Analyse, 2022* and with A. Balakrishna, J. Webster *Math. Models Methods Appl. Sciences, 2023.* The work was partially performed while the author was a member of the MSRI program "Mathematical problem in fluid dynamics" at the University of California Berkeley during the Spring 2021 and Fall 2023 (NSF DMS-1928930).

- **10:00–11:00,** Plenary Talk:
  - **Entropy and minimal data rates for state estimation and model detection**

    Daniel Liberzon, University of Illinois at Urbana-Champaign

    We will discuss estimation entropy for continuous-time nonlinear dynamical systems, which is a variant of topological entropy formulated in terms of the number of functions that approximate all system trajectories up to an exponentially decaying error. We will derive upper and lower bounds on the estimation entropy. We will describe an iterative procedure that uses quantized and sampled state measurements to generate state estimates that converge to the true state at the desired exponential rate. The average bit rate utilized by this procedure matches the derived upper bound on the estimation entropy, and no other algorithm of this type can perform the same estimation task with bit rates lower than the estimation entropy. As an application of this estimation procedure, we will study the problem of determining, from quantized state measurements, which of two competing models of a dynamical system is the true model. We will show that under a mild assumption of exponential separation of the candidate models, detection always happens in finite time. Recent work on entropy of switched and interconnected systems will also be briefly discussed.

- **11:00–11:20:** Coffee Break
- **11:20–11:50:**
  - (session 1)
    Exact and approximate vector Ekeland variational principles
    Bao Truong, Northern Michigan University

    The talk concerns with exact and approximate Ekeland variational principles for vector-valued functions and bifunctions, that are derived via linear and nonlinear scalarization processes by an approximate scalar formulation of the Ekeland variational principle and a revised version of Dancs-Hegedus-Medvegyev's fixed point theorem. Both results are also interesting in themselves and involve really mild assumptions.
  - (session 2)
    Optimal control for coupled sweeping process: nonsmooth time-dependent sweeping set and joint endpoints constraints
    Samara Chamoun, Department of Mathematics, Michigan State University

    In this talk, we launch the study of nonsmooth optimal control problems involving a controlled sweeping process governed by a nonsmooth time-dependent sweeping set and coupled with controlled differential equations and having joint endpoints constraints. This general model incorporates different controlled submodels as a particular case including second order sweeping processes, a subclass of integro-differential sweeping processes, evolution variational inequalities (EVI), and dynamical variational inequalities (DVI). Essentially, after establishing the existence of optimal solutions for our general fixed time Mayer problem, we shall establish the full form of the nonsmooth Pontryagin maximum principle. This will be accomplished via a modification of the exponential-penalty approximation method and by using advanced tools from nonsmooth and variational analysis. The utility of the theoretical results will be illustrated with a numerical application.

    The talk is based on a joint work with Prof. Vera Zeidan.
  - (session 3)
    Cooperative multi-agent constrained POMDPs: strong duality and primal-dual reinforcement learning with approximate information states
    Nouman Khan, EECS, University of Michigan

    We study the problem of decentralized constrained POMDPs in a team-setting where the multiple non-strategic agents have asymmetric information. Strong duality is established for the setting of infinite-horizon expected total discounted costs when the observations lie in a countable space, the actions are chosen from a finite space, the constraint costs are bounded, and the objective cost is bounded from below. Following this, connections with the common-information and approximate information-state approaches are established. The approximate information-states are characterized independent of the Lagrange-multipliers vector so that adaptations of the multiplier (during learning) will not necessitate new representations. Finally, a primal-dual multi-agent reinforcement learning (MARL) framework based on centralized training distributed execution (CTDE) and three time-scale stochastic approximation is developed with the aid of recurrent and feedforward neural-networks as function-approximators.

- **11:50–12:20:**
  - (session 1)
    Improved effective Lojasiewicz inequality and applications
    Ali Mohammad Nezhad,
    Statistics and Operations Research, University of North Carolina at Chapel Hill
    Given a closed and bounded semi-algebraic set $A \in \mathbb{R}^n$ and semi-algebraic continuous functions $f, g\colon A \to \mathbb{R}$, such that $f^{-1}(0) \subset g^{-1}(0)$, there exist N and $c > 0$ such that the inequality (Lojasiewicz inequality) $|g(x)|^N \leq c|f(x)|$ holds for all $x \in A$. In this talk, we prove a nearly tight upper bound on the Lojasiewicz exponent for semi-algebraic functions. We show that the Lojasiewicz exponent N is bounded by $(8d)^{2(n+7)}$, where d is the maximum degree of polynomials in the given semi-algebraic descriptions of A and the graphs of f and g. Unlike the previous best known bound in this setting, our bound is independent of the cardinalities of the semi-algebraic descriptions of f, g, and A. We exploited this fact to improve the best known error bounds for polynomial and non-linear semi-definite systems.
  - (session 2)
    Optimization of control free time sweeping processes and applications
    Trang Nguyen, Wayne State University
    This project addresses a free time optimal control problem of Mayer type for sweeping processes in which the perturbation is non-smooth. We develop a constructive finite-difference approximation procedure that allows us to establish necessary optimality conditions for the original continuous-time controlled sweeping process, and then show how these optimality conditions are applied to solve several applications in the real life.
- **12:20–1:50:** Boxed lunch
- **1:50–2:20:**
  - (session 1)
    The boosted DC algorithm for clustering with constraints
    Tuyen Tran, Loyola University Chicago
    This talk aims to investigate the effectiveness of the recently proposed Boosted Difference of Convex functions Algorithm (BDCA) when applied to clustering with constraints problems. We present the mathematical basis for the BDCA and Difference of Convex functions Algorithm (DCA), along with a penalty method based on distance functions. We then develop algorithms for solving these problems and implement them. The implementations are numerically compared using various examples. We find that the BDCA method converges in fewer iterations than the corresponding DCA-based method. However, BDCA requires self-adaptivity to yield faster run-times for small problems.
  - (session 2)
    Time-distributed optimization for linear model predictive control
    Jordan Leung, Aerospace Engineering, University of Michigan
    Time-distributed optimization is a strategy for reducing the computational effort required to implement Model Predictive Control (MPC). Time-distributed MPC (TDMPC) methods maintain a running estimate of the solution to an optimal control problem (OCP) and improve this estimate using a limited number of optimizer iterations during each sampling

period. This talk explores the closed-loop properties of a constrained linear system controlled using TDMPC. Continuity properties of the OCP and convergence properties of the optimizer are used to derive a bound on the number of iterations per sampling period required to enforce asymptotic stability. Moreover, a region of attraction estimate in the plant-optimizer space is explicitly characterized.

– (session 3)
A consistently adaptive trust-region method
Fadi Hamad, Department of Industrial Engineering, University of Pittsburgh
Adaptive trust-region methods attempt to maintain strong convergence guarantees without depending on conservative estimates of problem properties such as Lipschitz constants. However, on close inspection, one can show existing adaptive trust-region methods have theoretical guarantees with severely suboptimal dependence on problem properties such as the Lipschitz constant of the Hessian. For example, TRACE developed by Curtis et al. obtains a $O(\Delta_f L^{3/2}\epsilon^{-3/2}) + \tilde{O}(1)$ iteration bound where L is the Lipschitz constant of the Hessian. Compared with the optimal $O(\Delta_f L^{1/2}\epsilon^{-3/2})$ bound this is suboptimal with respect to L. We present the first adaptive trust-region method which circumvents this issue and requires at most $O(\Delta_f L^{1/2}\epsilon^{-3/2}) + \tilde{O}(1)$ iterations to find an ϵ-approximate stationary point, matching the optimal iteration bound up to an additive logarithmic term. Our method is a simple variant of a classic trust-region method and in our experiments performs competitively with both ARC and a classical trust-region method.

- **2:20–2:50:**

– (session 1)
Difference-of-convex algorithm with extrapolation for nonconvex, nonsmooth optimization problems
Duy Nhat Phan, UDRI Applied Sensing Devision, University of Dayton Research Institute
In this talk, we focus on the problem of minimizing the sum of a nonconvex differentiable function and a difference of convex (DC) function, where the differentiable function is not restricted to the global Lipschitz gradient continuity assumption. This problem covers a broad range of applications in machine learning and statistics, such as compressed sensing, signal recovery, sparse dictionary learning, matrix factorization, etc. We first take inspiration from the Nesterov acceleration technique and the DC algorithm to develop a novel algorithm for the considered problem. We then study the subsequential convergence of our algorithm to a critical point. Furthermore, we justify the global convergence of the whole sequence generated by our algorithm to a critical point and establish its convergence rate under the Kurdyka-Lojasiewicz condition. Numerical experiments on the nonnegative matrix completion problem are performed to demonstrate the efficiency of our algorithm and its superiority over well-known methods.

– (session 3)
Are turn-by-turn navigation systems of regular vehicles ready for edge-assisted autonomous vehicles?
Syeda Atik, Wayne State University
Private and public transportation will be dominated by Autonomous Vehicles (AV), which are safer than regular vehicles. However, ensuring good performance for the autonomous features requires fast processing of heavy tasks. Providing each AV with powerful computing resources may result in increased AV cost and decreased driving range. An alternative

solution is to install low-power computing hardware on each AV and offload the heavy tasks to powerful nearby edge servers. In this case, the AV's reaction time depends on how quickly the navigation tasks are completed in the edge server. To reduce task completion latency, the edge servers must be equipped with enough network and computing resources to handle the vehicle demands, which show large spatio-temporal variations. Thus, deploying the same resources in different locations may lead to unnecessary resource over-provisioning. In this paper, we leverage simulations using real traffic data to discuss the implications of deploying heterogeneous resources in different city areas to sustain peak versus average demand of edge-assisted AVs. Our analysis indicates that a reduction in network bandwidth and computing cores of up to 60% and 50%, respectively, is achieved by deploying edge resources for the average demand rather than peak demand. We also investigate how the peak-hour demand affects the safe travel time of AVs and find that it can be reduced by approximately 20% if they would be rerouted to areas with a lower edge-resource load. Thus, future research must consider that traditional turn-by-turn navigation systems may not provide the fastest routes for edge-assisted AVs.

- **2:50–3:20:**

  – (session 1)
  Optimal control of several motion models
  Thi Phung, Wayne State University

  This paper is devoted to the study of the dynamic optimization of several controlled crowd motion models in the general planar settings, which is an application of a class of optimal control problems involving a general nonconvex sweeping process with perturbations. A set of necessary optimality conditions for such optimal control problems involving the crowd motion models with multiple agents and obstacles is obtained and analyzed. Several effective algorithms based on such necessary optimality conditions are proposed and various nontrivial illustrative examples together with their simulations are also presented.

  – (session 2)
  Exploiting negative curvature in conjunction with adaptive sampling: Theoretical Results and a Practical Algorithm
  Wanping Dong, Industrial and Operations Engineering, University of Michigan

  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.

  – (session 3)
  Automated identification of gender discriminatory texts on GitHub
  Sayma Sultana, Wayne State University

In an industry dominated by straight men, many developers representing other gender identities and sexual orientations often encounter hateful or discriminatory messages. Such communications pose barriers to participation for women and LGBTQ+ persons. Although many communities have created codes of conduct (CoC) to discourage such communications, minorities are often afraid of reporting CoC violations fearing backlashes and their own career prospects. Manually checking all communications for such violations is infeasible due to the sheer volume of communications generated in a large-scale Free Open Source Software (FLOSS) community. To address this challenge, this study aims to develop an automated mechanism to identify Gender identity or Sexual orientation Discriminatory (GSD) texts from software developers' communications. On this goal, we first developed manual labeling rubric to to identify GSD texts. With a keyword based filtering followed by stratified sampling using an off-the-shelf classifier, we created a sample of 11,007 issue comments written on GitHub. Our manual labeling of this dataset following our rubric identified $1,499$ GSD texts. Using this dataset, we trained and evaluated GSD4SE (Gender identity or Sexual orientation based D iscrimination identification for (4) Software Engineering texts), as a supervised learning based GSD detection tool. GSD4SE incorporates six preprocessing steps, and ten state-of-the-art algorithms. In our ten-fold cross validation based evaluations, a transformer based model using the BERT-base encoding [23] boosts the performance with 84.1% precision, 80.7% recall, 82.1% f-score for the GSD class. This model achieves 95.2% accuracy, and 79.5% Matthews Correlation Coefficient. Our dataset and tool establish a foundation for further research in this direction.

- **3:20–3:30:** Closing Remarks
- **3:30–4:00:** Coffee