AND ARGONNE NATIONAL LABORATORY

2022 ACMNTW Workshop on Optimization and Machine Learning

Thursday, May 19th, 2022

James L. Allen Center

2169 Campus Drive, Evanston, IL 60208

Schedule of Events

9:00 A.M.	Welcome and Continental Breakfast
9:30 A.M.	Siqian Shen, University of Michigan
10:10 A.M.	Carla Michini, University of Wisconsin
10:50 A.M.	Kibaek Kim, Argonne National Laboratory
11:30 A.M.	Lunch, Allen Center Dining Room
1:00 P.M.	Brian Bullins, Toyota Technological Institute at Chicago
1:40 P.M.	Haihao Lu, University of Chicago
2:20 P.M.	Coffee Break
3:00 P.M.	Zhaoran Wang, Northwestern University
3:40 P.M.	Tammy Kolda, mathsci.ai and Northwestern University
4:30 P.M.	Poster Session
6:00 P.M.	Dinner, Farmhouse Evanston

All lectures will be held in Allen Center Room 140, while the poster session will be held in Allen Center Room 153. Breakfast and coffee can be found in the first-floor lounge.

Please see pages 17 and 18 of this program for details on event parking.

Dinner will be served on the second floor of Farmhouse Evanston, located at 703 Church Street in Evanston. This is a .8 mile walk from the Allen Center, and approximately a six minute drive.

Hosted by Northwestern University

Co-Sponsored by the University of Wisconsin- Madison

Lectures

9:30 A.M.
Optimization Methods for Discrete Binary Quantum Control
Siqian Shen, University of Michigan

Quantum control aims to manipulate quantum systems toward specific quantum states or desired operations. Designing highly accurate and effective control steps is vitally important to various quantum applications, including energy minimization and circuit compilation. We study optimal control problems for quantum systems and additionally examine the problems under the uncertainty of controllers' performance. For deterministic binary quantum control, we develop a generic model and several extensions to handle additional side constraints and reduce switches. We modify the popular gradient ascent pulse engineering (GRAPE) algorithm, develop a new alternating direction method of multipliers (ADMM) algorithm to solve the continuous relaxation of the penalized model, and then apply rounding techniques to obtain binary control solutions. We propose a modified trust-region method to further improve the solutions. For stochastic quantum control, we propose a chance-constrained formulation to guarantee high-fidelity of the control performance and solve its corresponding conditional value at risk (CVaR) approximation with continuous or binary control. We develop two alternative ways, including an interpolated formulation and a two-step minimization algorithm based on our previous modified GRAPE algorithm, to solve the CVaR model that has a non-differentiable and non-convex objective function. We conduct numerical studies on a family of quantum pulse optimization examples, and the results demonstrate that (i) our algorithms can obtain high-quality control results; (ii) the CVaR model improves solution robustness under controller uncertainty.

This is a joint work with Xinyu Fei, a PhD student in U-M IOE, Dr. Lucas T. Brady at Ames Quantum AI Lab in NASA, and Drs. Jeffrey Larson and Sven Leyffer in Argonne National Lab.

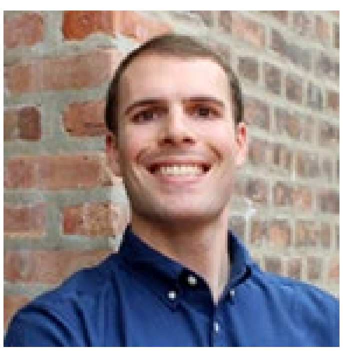
10:10 A.M.
Shattering Inequalities for Learning Optimal Decision Trees
Carla Michini, University of Wisconsin

Recently, mixed-integer programming (MIP) techniques have been applied to learn optimal decision trees. Empirical research has shown that optimal trees typically have better out-of-sample performance than heuristic approaches such as CART. However, the underlying MIP formulations often suffer from slow runtimes, due to weak linear programming relaxations. We propose a new MIP formulation for learning optimal decision trees with multivariate branching rules and no assumptions on the feature types. Our formulation crucially employs binary variables expressing how each observation is routed throughout the entire tree. We then introduce a new class of valid inequalities for learning optimal multivariate decision trees. Each inequality encodes an inclusion-minimal set of points that cannot be shattered by a multivariate split, and in the context of a MIP formulation, the inequalities are sparse, involving at most the number of features plus two variables. We leverage these valid inequalities within a Benders-like decomposition, where the master problem determines how to route each observation to a leaf node to minimize misclassification error, and the subproblem checks whether, for each branch node of the decision tree, it is possible to construct a multivariate split that realizes the given routing of observations; if not, the subproblem adds at least one of our valid inequalities to the master problem. We demonstrate through numerical experiments that our MIP approach outperforms (in terms of training accuracy, testing accuracy, solution time, and relative gap) two other popular MIP formulations, and is able to improve both in and out-of-sample performance, while remaining competitive in terms of solution time to a wide range of popular approaches from the literature.

10:50 A.M.
A scalable algorithm for solving ACOPF on GPUs
Kibaek Kim, Argonne National Laboratory

GPU has been successfully and widely used for solving machine learning problems, but not for more generic optimization problems. We present the challenges of using GPUs for solving constrained optimization and our approach to solving large-scale nonlinear optimization for alternating current optimal power flow (ACOPF) on GPUs. Our approach is based on the Lagrangian decomposition that reformulates the problem into many small subproblems. In particular, we present the numerical performance of our algorithm for solving large-scale ACOPF (up to 70,000-bus systems) on GPUs.

1:00 P.M.
Beyond First-Order Methods for Large-Scale Optimization
Brian Bullins, Toyota Technological Institute at Chicago



In recent years, stochastic gradient descent (SGD) has taken center stage for training large-scale models in machine learning. Although methods which go beyond first-order information may achieve better iteration complexity in theory, the per-iteration costs often render them unusable when faced with the current growth in both the available data and the size of the models, particularly when such models now have hundreds of billions of parameters.

In this talk, I will present results, both theoretical and practical, for dealing with two key challenges in this setting, whereby I will show how second-order optimization may be as scalable as first-order methods. First, given the non-convexity of deep neural networks, it has become important to develop a better understanding of non-convex guarantees. Thus, I will present a Hessian-based method which provably converges to first-order critical points faster than gradient descent, alongside guarantees for converging to second-order critical points. In addition, optimization methods which may parallelize have also become increasingly critical when facing enormous deep learning models, and so I will show how we may leverage stochastic second-order information to attain faster methods in the distributed optimization setting.

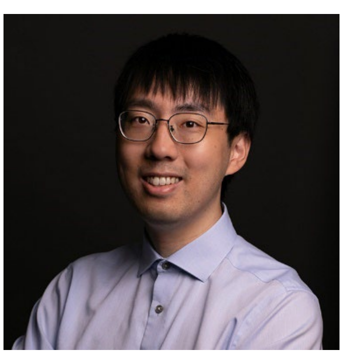
1:40 P.M.
First Order Methods for Linear Programming: Theory, Computation, and Applications

Haihao Lu, University of Chicago

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

3:00 P.M.
Demystifying (Deep) Reinforcement Learning with Optimism and Pessimism

Zhaoran Wang, Northwestern University



Coupled with powerful function approximators such as deep neural networks, reinforcement learning (RL) achieves tremendous empirical successes. However, its theoretical understandings lag behind. In particular, it remains unclear how to provably attain the optimal policy with a finite regret or sample complexity. In this talk, we will present the two sides of the same coin, which demonstrates an intriguing duality between optimism and pessimism.

– In the online setting, we aim to learn the optimal policy by actively interacting with the environment. To strike a balance between exploration and exploitation, we propose an optimistic least-squares value iteration algorithm, which achieves a \sqrt{T} regret in the presence of linear, kernel, and neural function approximators.

– In the offline setting, we aim to learn the optimal policy based on a dataset collected a priori. Due to a lack of active interactions with the environment, we suffer from the insufficient coverage of the dataset. To maximally exploit the dataset, we propose a pessimistic least-squares value iteration algorithm, which achieves a minimax-optimal sample complexity.

3:40 P.M.
Randomized Least Squares Optimization and its Incredible Utility for Large-Scale Tensor Decomposition
Tamara Kolda, mathsci.ai and Northwestern University

Randomized least squares is a promising method but not yet widely used in practice. We show an example of its use for finding low-rank canonical polyadic (CP) tensor decompositions for large sparse tensors. This involves solving a sequence of overdetermined least problems with special (Khatri-Rao product) structure. In this work, we present an application of randomized algorithms to fitting the CP decomposition of sparse tensors, solving a significantly smaller sampled least squares problem at each iteration with probabilistic guarantees on the approximation errors. We perform sketching through leverage score sampling, crucially relying on the fact that the problem structure enable efficient sampling from overestimates of the leverage scores with much less work. We discuss what it took to make the algorithm practical, including general-purpose improvements. Numerical results on real-world large-scale tensors show the method is faster than competing methods without sacrificing accuracy. (This is joint work with Brett Larsen, Stanford University.)

AND ARGONNE NATIONAL LABORATORY

Poster Session

ParMOO: Parallel MultiObjective Optimization in Python

Tyler Chang, Argonne National Laboratory
Coauthor: Stefan Wild (Argonne National Laboratory)

ParMOO is a parallel multiobjective optimization solver that seeks to exploit simulation-based structure in objective and constraint functions. To exploit structure, ParMOO models simulations separately from objectives and constraints. ParMOO is implemented in Python. In order to achieve scalable parallelism, we use libEnsemble to distribute batches of simulation evaluations across parallel resources. ParMOO is available for download on GitHub: https://github.com/parmoo/parmoo

Federated Condition Monitoring Signal Prediction with Improved Generalization

Seokhyun Chung, University of Michigan
Coauthor: Raed Kontar (University of Michigan)

Revolutionary advances in Internet of Things technologies have paved the way for a significant increase in computational resources at edge devices that collect condition monitoring (CM) data. This poses a significant opportunity for federated analytics which exploits edge computing resources to distribute model learning, reduce communication traffic and circumvent the need to share raw data. In this paper we study CM signal prediction where operating units, that have data storage and computational capabilities, jointly learn models without sharing their collected CM signals. Specifically, we first propose a framework for CM signal prediction and introduce a federated approach that tries to improve generalization by encouraging flat solutions through distributed computations. Then, a personalization approach is proposed to adapt the learned model to new clients without losing old knowledge. We examine our proposed framework on CM signals from aircraft turbofan engines under three realistic federated CM scenarios. Experimental results highlight the advantageous features of the proposed approach in improving generalization while decentralizing model inference.

Exploiting Prior Function Evaluations in Derivative-Free Optimization

Shima Dezfulian, Northwestern University
Coauthors: Frank E. Curtis (Lehigh University), Andreas Waechter (Northwestern University)

A derivative-free optimization (DFO) algorithm is presented. The distinguishing feature of the algorithm is that it allows for the use of function values that have been made available through prior runs of a DFO algorithm for solving prior related optimization problems. Applications in which sequences of related optimization problems are solved such that the proposed algorithm is applicable include certain least-squares and simulation-based optimization problems. A convergence guarantee of a generic algorithmic framework that exploits prior function evaluations is presented, then a particular instance of the framework is proposed and analyzed. The results of numerical experiments when solving engineering test problems show that the algorithm gains advantage over a state-of-the-art DFO algorithm when prior function values are available.

Trust-Region Stochastic Sequential Quadratic Programming for Equality-Constrained Optimization

Yuchen Fang, University of Chicago
Coauthors: Sen Na (University of California, Berkeley), Mladen Kolar (University of Chicago)

We design two stochastic SQP algorithms, with a trust-region method used to select random stepsizes, for solving equality-constrained optimization problems: a fully stochastic algorithm in which only a single sample is generated in each iteration; and an adaptive algorithm in which batch samples are generated to have a more precise model estimate. For both algorithms, we establish an almost sure global convergence. Under reasonable assumptions, the KKT residuals converge to zero almost surely.

Asynchronous Distributed Rendezvous For Continuously Moving Agents

Charikleia Iakovidou, Northwestern University
Coauthor: Ermin Wei (Northwestern University)

We consider an asynchronous distributed rendezvous problem with continuously moving agents, where each agent is associated with a unique, strongly convex, position-dependent cost function, and the optimal rendezvous point minimizes the sum of the individual cost functions. We assume that agents randomly and independently alternate between two states: i) an active state, during which they can sense their current positions, broadcast messages to their neighbors, access their local buffers where potentially outdated information from their neighbors is stored, and adjust their velocities;

and ii) a passive state, during which agents continue to move towards the direction they calculated in their most recent active state and listen for messages. We model the rendezvous problem as a distributed consensus optimization problem, and propose a fully asynchronous algorithm that is robust to outdated information and erroneous displacements caused by inactive states, and converges in expectation to an arbitrarily small neighborhood of the optimal solution with an appropriate selection of parameters.

Bilevel Optimization Algorithm With Importance Sampling

Simiao Jiao, University of Chicago
Coauthors: Boxin Zhao, Mladen Kolar (University of Chicago)

Bilevel Optimization problem (BO) is a powerful model which appears in many fields, such as Meta-learning or hyperparameter tuning. In this paper, we focus on the non-convex stochastic bilevel optimization problem where the outer and inner functions are nonconvex and strongly convex respectively, and both of them can be expressed as finite summation. Even though many SGD-based algorithms have been proposed, they are limited to use uniform sampling when calculating stochastic gradients. When the norms of gradients have high variation, those uniform sampling based method may face slow convergence. In this paper, we introduce the technique of importance sampling into Bilevel Optimization problem, and show that better convergence rate can be obtained. We also validate our theoretical results on both synthetic and real data experiments.

Efficient Sparse Matrix Factorization via Alternating Minimization

Geyu Liang, University of Michigan
Coauthor: Salar Fattahi (University of Michigan)

Many phenomena in physical, engineering and social sciences feature systems that exhibit low-rank and sparse patterns, which makes a challenging yet meaningful problem to recover matrices that represent them. In this talk, we will present an alternating minimization algorithm with shrinking thresholds to solve the sparse matrix factorization problems. The algorithm is verified, both theoretically and in practice, to converge linearly around any ground truth as long as there is sufficient sparsity. Furthermore, we will show that the sparsity level required is nearly constant and the radius of the basin for linear convergence only depends on the sparsity level.

Global Convergence of Sub-gradient Method for Robust Matrix Recovery: Small Initialization, Noisy Measurements, and Over-parameterization

Jianhao Ma, University of Michigan
Coauthor: Salar Fattahi (University of Michigan)

In this work, we study the performance of sub-gradient method (SubGM) on a natural nonconvex and nonsmooth formulation of low-rank matrix recovery with l1-loss, where the goal is to recover a low-rank matrix from a limited number of measurements, a subset of which may be grossly corrupted with noise. We study a scenario where the rank of the true solution is unknown and over-estimated instead. The over-estimation of the rank gives rise to an over-parameterized model in which there are more degrees of freedom than needed. Such over-parameterization may lead to overfitting, or adversely affect the performance of the algorithm. We prove that a simple SubGM with small initialization is agnostic to both over-parameterization and noise in the measurements. In particular, we show that small initialization nullifies the effect of over-parameterization on the performance of SubGM, leading to an exponential improvement in its convergence rate. Moreover, we provide the first unifying framework for analyzing the behavior of SubGM under both outlier and Gaussian noise models, showing that SubGM converges to the true solution, even under arbitrarily large and arbitrarily dense noise values, and—perhaps surprisingly—even if the globally optimal solutions do not correspond to the ground truth. At the core of our results is a robust variant of restricted isometry property, called Sign-RIP, which controls the deviation of the sub-differential of the l1-loss from that of an ideal, expected loss. As a byproduct of our results, we consider a subclass of robust low-rank matrix recovery with Gaussian measurements, and show that the number of required samples to guarantee the global convergence of SubGM is independent of the over-parameterized rank.

Stability Constrained Optimization Using Neural Lyapunov Control

Hideaki Nakao, Argonne National Laboratory

In many practical applications, the decisions one has made are fluctuated by the dynamics evolving over time. To provide reliable decisions, it is important to take the stability of the dynamical system into account in the optimization problem. Bridging between the static component of the decision-making and the dynamic stability assessment has been particularly challenging due to the involvement of differential constraints. In this paper, we reformulate the dynamic stability constraint using control Lyapunov functions which facilitate the connection of static optimization and optimal control and guarantee robustness against various scenarios of disturbances. To overcome the difficulty of finding the Lyapunov function, we take two approaches. First, we use deep learning to characterize the Lyapunov function using the rich class of representation capability of neural networks. Secondly, we decompose the Lyapunov function for each smaller subsystem and reformulate structured Lyapunov functions which can be computed in parallel. We test the algorithm on optimal power flow problems in power systems optimization, where the voltage angles are required to be stable.

Differentially Private Federated Learning: Algorithms and Open-Source Software Framework

Minseok Ryu, Argonne National Laboratory
Coauthors: Kibaek Kim, Youngdae Kim, Ravi Madduri (Argonne National Laboratory)

Federated learning (FL) enables learning from decentralized data owned by multiple clients without gathering them in a central server. In FL, the server iteratively updates global model parameters by gathering and processing the local model parameters trained by clients. The FL capability is especially important to domains such as biomedicine and smart grid, where data may not be shared freely or stored at a central location because of policy challenges. To guarantee data privacy during the FL process, differential privacy (DP) techniques can be applied by randomizing the communication of local model parameters. The DP techniques, however, could compromise the learning performances to ensure strong data privacy. To address the issue, we develop a privacy-preserving FL algorithm constructed by introducing the objective perturbation and multiple local update techniques to the inexact alternating direction method of multipliers algorithm. Moreover, we have developed the Argonne Privacy-Preserving Federated Learning (APPFL) framework that allows users to use a range of existing algorithms, implement new algorithms, and simulate and deploy various FL algorithms with privacy-preserving techniques. In APPFL, a remote procedure call is used as a protocol for communicating model parameters between the server and clients, which enables FL under a heterogeneous computing architecture (e.g., Google Cloud Server, high-performance computing (HPC) clusters, and laptops). In addition, APPFL supports a simulation of FL in an HPC architecture with message passing interface. Finally, we present numerical demonstrations of APPFL using various data.

Convex Chance-Constrained Programs with Wasserstein Ambiguity

Haoming Shen, University of Michigan
Coauthor: Ruiwei Jiang (University of Michigan)

Chance constraints yield non-convex feasible regions in general. In particular, when the uncertain parameters are modeled by a Wasserstein ball, arXiv:1806.07418 and arXiv:1809.00210 showed that the distributionally robust (pessimistic) chance constraint admits a mixed-integer conic representation. This paper identifies sufficient conditions that lead to convex feasible regions of chance constraints with Wasserstein ambiguity. First, when uncertainty arises from the left-hand side of a pessimistic individual chance constraint, we derive a convex and conic representation if the

Wasserstein ball is centered around a Gaussian distribution. Second, when uncertainty arises from the right-hand side of a pessimistic joint chance constraint, we show that the ensuing feasible region is convex if the Wasserstein ball is centered around a log-concave distribution (or, more generally, an α-concave distribution with $\alpha \geq -1$). In addition, we propose a block coordinate ascent algorithm for this class of chance constraints and prove its convergence to global optimum. Furthermore, we extend the convexity results and conic representation to optimistic chance constraints.

Accelerating Stochastic Sequential Quadratic Programming for Equality Constrained Optimization using Predictive Variance Reduction

Jiahao Shi, University of Michigan
Coauthors: Albert S. Berahas, Zihong Yi, and Baoyu Zhou (University of Michigan)

In this paper, we propose a stochastic method for solving equality con- strained optimization problems that utilizes predictive variance reduction. Specifically, we develop a method based on the sequential quadratic programming paradigm that employs variance reduction in the gradient approximations. Under reasonable assumptions, we prove that a measure of first-order stationarity evaluated at the iterates generated by our proposed algorithm converges to zero in expectation from arbitrary starting points, for both constant and adaptive step size strategies. Finally, we demonstrate the practical performance of our proposed algorithm on constrained binary classification problems that arise in machine learning.

A Trust Region Method for the Optimization of Noisy Functions

Shigeng Sun, Northwestern University
Coauthor: Jorge Nocedal (Northwestern University)

Classical trust region methods were designed to solve problems in which function and gradient information are exact. This paper considers the case when there are bounded errors (or noise) in the above computations and proposes a simple modification of the trust region method to cope with these errors. The new algorithm only requires information about the size of the errors in the function evaluations and incurs no additional computational expense. It is shown that, when applied to a smooth (but not necessarily convex) objective function, the iterates of the algorithm visit a neighborhood of stationarity infinitely often, and that the rest of the sequence cannot stray too far away, as measured by function values. Numerical results illustrate how the classical trust region algorithm may fail in the presence of noise, and how the proposed algorithm ensures steady progress towards stationarity in these cases.

High Probability Complexity Bounds for Optimization Algorithm Based on Stochastic Oracles

Miaolan Xie, Argonne National Laboratory and Cornell University
Coauthors: Billy Jin and Katya Scheinberg (Cornell University)

We consider an optimization method for continuous optimization under a stochastic setting where the function values and gradients are available only through inexact probabilistic zeroth and first-order oracles. These oracles capture multiple standard settings including expected loss minimization and zeroth-order optimization. Moreover, our framework is very general and allows the function and gradient estimates to be biased. The proposed algorithm is simple to describe, easy to implement. Under fairly general conditions on the oracles, we derive a high probability tail bound on the iteration complexity of the algorithm.

Nearly Optimal Linear Convergence of Stochastic Primal-Dual Methods for Linear Programming

Jinwen Yang, University of Chicago
Coauthor: Haihao Lu (University of Chicago)

There is a recent interest on first-order methods for linear programming (LP). In this paper, we propose a stochastic algorithm using variance reduction and restarts for solving sharp primal-dual problems such as LP. We show that the proposed stochastic method exhibits a linear convergence rate for solving sharp instances with a high probability. In addition, we propose an efficient coordinate-based stochastic oracle for unconstrained bilinear problems, which has $\mathcal O(1)$ per iteration cost and improves the complexity of the existing deterministic and stochastic algorithms. Finally, we show that the obtained linear convergence rate is nearly optimal (upto $\log$ terms) for a wide class of stochastic primal dual methods.

Strong valid inequalities for a class of concave submodular minimization problems under cardinality constraints

Qimeng Yu, Northwestern University
Coauthor: Simge Küçükyavuz (Northwestern University)

We study the polyhedral convex hull structure of a mixed-integer set which arises in a class of cardinality-constrained concave submodular minimization problems. This class of problems has an objective function in the form of $f(a^\top x)$, where f is a univariate concave function, a is a non-negative vector, and x is a binary vector of appropriate dimension. Such minimization problems frequently appear in applications that involve risk-aversion or economies of scale. We propose three classes of strong valid linear inequalities for this convex hull and specify their facet conditions

when a has two distinct values. We show how to use these inequalities to obtain valid inequalities for general a that contains multiple values. We further provide a complete linear convex hull description for this mixed-integer set when a contains two distinct values and the cardinality constraint upper bound is two. Our computational experiments on the mean-risk optimization problem demonstrate the effectiveness of the proposed inequalities in a branch-and-cut framework.

Risk-Averse Reinforcement Learning via Dynamic Time-Consistent Risk Measures

Xian Yu, University of Michigan
Coauthor: Siqian Shen (University of Michigan)

Traditional reinforcement learning (RL) aims to maximize the expected total reward, while risk management is needed to ensure reliable performance in a risk-averse setting. In this paper, we consider the problem of maximizing a dynamic risk measure of a sequence of rewards under an infinite-horizon Markov Decision Process (MDP) with finite states and actions. We adapt the Expected Conditional Risk Measures (ECRMs) to the infinite-horizon risk-averse MDP setting and prove its time consistency. Using a convex combination of expectation and CVaR as a special one-step conditional risk measure, we reformulate the risk-averse MDP as a risk-neutral counterpart with augmented action space and manipulation on the immediate rewards. We further prove that the related Bellman operator is a contraction mapping, which guarantees the convergence of any value-based RL algorithms. Accordingly, we develop a risk-averse deep Q-learning framework, and our numerical studies based on two simple MDPs show that the risk-averse setting can reduce the variance and enhance robustness of the results.

Federated Gaussian Process and Stochastic Gradient Descent in the Correlated Setting

Xubo Yue, University of Michigan
Coauthor: Raed Al Kontar (University of Michigan)

In this paper, we propose FGPR: a Federated Gaussian process (GP) regression framework that uses an averaging strategy for model aggregation and stochastic gradient descent for local client computations. Notably, the resulting global model excels in personalization as FGPR jointly learns a global GP prior across all clients. The predictive posterior then is obtained by exploiting this prior and conditioning on local data which encodes personalized features from a specific client. Theoretically, we show that FGPR converges to a critical point of the full log-likelihood function, subject to statistical error. Through extensive case studies we show that FGPR excels in a wide range of applications and is a promising approach for privacy-preserving multi-fidelity data modeling.

Adaptive Client Sampling in Federated Learning via Online Learning with Bandit Feedback

Boxin Zhao, University of Chicago, Ant Financial Group
Coauthors: Ziqi Liu (Ant Financial Group), Chaochao Chen (Zhejiang University), Mladen Kolar (University of Chicago), Zhiqiang Zhang (Ant Financial Group), Jun Zhou (Ant Financial Group)

Due to the high cost of communication, federated learning (FL) systems need to sample a subset of clients that are involved in each round of training. As a result, client sampling plays an important role in FL systems as it affects the convergence rate of optimization algorithms used to train machine learning models. Despite its importance, there is limited work on how to sample clients effectively. In this paper, we cast client sampling as an online learning task with bandit feedback, which we solve with an online stochastic mirror descent (OSMD) algorithm designed to minimize the sampling variance. To handle the tuning parameters in OSMD that depend on the unknown problem parameters, we use the online ensemble method and doubling trick. We prove a dynamic regret bound relative to the theoretically optimal sampling sequence. The regret bound depends on the total variation of the comparator sequence, which naturally captures the intrinsic difficulty of the problem. To the best of our knowledge, these theoretical contributions are new and the proof technique is of independent interest. Through both synthetic and real data experiments, we illustrate advantages of the proposed client sampling algorithm over the widely used uniform sampling and existing online learning based sampling strategies. The proposed adaptive sampling procedure is applicable beyond the FL problem studied here and can be used to improve the performance of stochastic optimization procedures such as stochastic gradient descent and stochastic coordinate descent.

An SDP Relaxation for the Sparse Integer Least Square Problem

Dekun Zhou, University of Wisconsin-Madison
Coauthor: Alberto Del Pia (University of Wisconsin-Madison)

We study the sparse integer least square problem (SILS), which is the NP-hard variant of the least square problem, where we only consider sparse {0,1,-1}-vectors. We propose an l1-based SDP relaxation to SILS, and provide sufficient conditions for our SDP relaxation to solve SILS. The class of data input which guarantee that SDP solves SILS is broad enough to cover many cases in real-world applications, such as privacy preserving identification, and multiuser detection. We also show with numerical tests

that our SDP relaxation performs generally better than other l1-based methods, such as Lasso and Dantzig Selector.

AND ARGONNE NATIONAL LABORATORY

Further Details

- As noted in the schedule, breakfast, lunch, and dinner will be provided at no cost to attendees.

- Parking passes useable in the North Campus Parking Garage will be available for all participants who request them in advance. If you did not request a pass while submitting your event RSVP, please reach out to mcccenters@northwestern.edu in advance of the event to submit a request. More information on parking can be found on the next page of this program.

- If you registered to present a poster, an easel, backboard, and pushpins will be provided to display it during the poster session. Backboards will be 40" by 60", although posters are not required to be this size. When you arrive, please present your poster to the event organizer staffing the check-in desk, and it will be set up for you prior to the poster session.

Program Committee

Jorge Nocedal, Northwestern University
Jeff Linderoth, University of Wisconsin-Madison

Event Volunteers

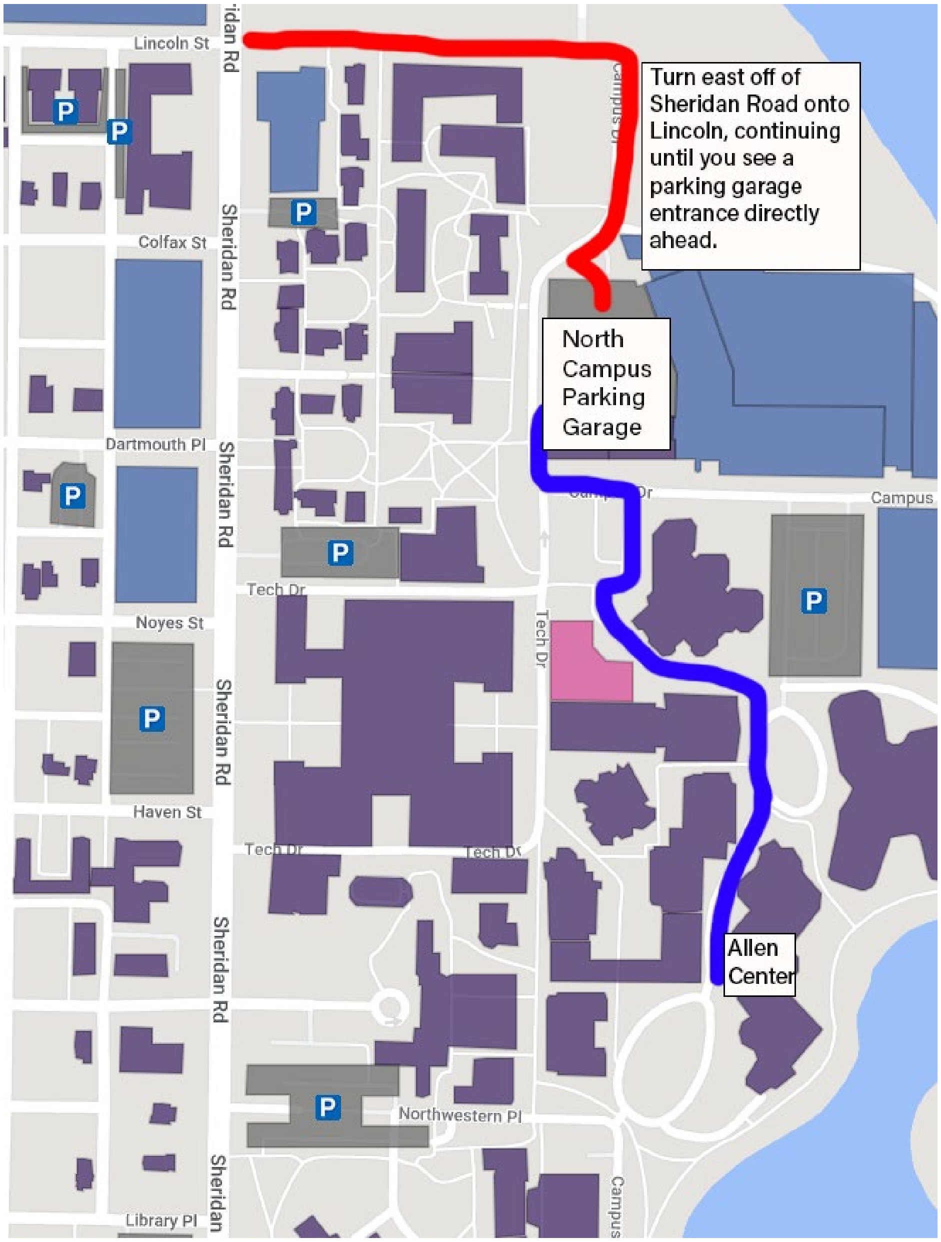
Guyi Chen, Northwestern University
Shigeng Sun, Northwestern University

Event Contact

Stephen Pedersen
847-491-3761
mcccenters@northwestern.edu

Event Parking

Parking is available at the North Campus Parking Garage on **[2311 N Campus Drive, Evanston, IL 60208]**. A parking permit is required to enter the garage. An event representative will be waiting outside the north entrance of the garage to provide one to you before parking and [walking to the event space].



Guest Wireless Access at Northwestern

While visiting Northwestern, campus guests can connect to the Guest-Northwestern wireless network. Guest wireless is a free service intended for use by visitors requiring short-term access to the wireless network to perform general Internet activities. It is an unsecured connection and should not be used to send or receive sensitive information.

If your institution uses Eduroam, please use the same credentials that you normally would to access Northwestern University's eduroam network. If not, please complete the following steps to register for guest internet access.

1. Choose "Guest-Northwestern" in the list of available networks on your device.
2. Accept the University's acceptable use policy.

Click here for more information on the University's guest wireless network, or go to https://www.it.northwestern.edu/oncampus/guest-wireless/ .

Hotel Options

All hotels are located less than one mile from the Allen Center, and close to our dinner venue in downtown Evanston.

Graduate Hotel Evanston: 1625 Hinman Avenue, Evanston IL 60201
Phone: (847) 475-3300
Sales Email: info@graduateevanston.com

Hilton Orrington: 1710 Orrington Avenue, Evanston, IL 60201
Phone: (847) 866-8700

Holiday Inn Chicago North-Evanston: 1501 Sherman Avenue, Evanston, IL 60201
Phone: (847) 491-6400
Sales Email: reservations@evanstonuniversityplaza.com