East Coast Optimization Meeting 2021

Theme: Optimization for Machine Learning

Dates

April 1-2, 2021

Location

Virtually hosted by Center for Mathematics and Artificial Intelligence
George Mason University
Fairfax, VA

Organizing Committee

Harbir Antil (George Mason University)
Drew P. Kouri (Sandia National Laboratories)
Denis Ridzal (Sandia National Laboratories)

Sponsors

National Science Foundation
Department of Mathematical Sciences, George Mason University
College of Science, George Mason University
Association for Women in Mathematics, GMU Chapter
Society for Industrial and Applied Mathematics

Keynote Speakers

1. Frank E. Curtis (Lehigh University)

TUTORIAL:

Title. Optimization Methods for Large-Scale Machine Learning

Abstract. This tutorial provides a review of numerical optimization algorithms for machine learning. First, we will discuss how optimization problems arise in machine learning and what makes them challenging. Second, we will discuss why large-scale machine learning represents a distinctive setting in which the stochastic gradient (SG) method has traditionally played a central role while conventional gradient-based nonlinear optimization techniques typically falter. With this in mind, we will consider a comprehensive theory of a straightforward, yet versatile SG algorithm, discuss its practical behavior, and highlight opportunities for designing algorithms with improved performance. This will lead to a discussion about the next generation of optimization methods for large-scale machine learning, including an investigation of two main streams of research on techniques that diminish noise in the stochastic directions and methods that make use of second-order derivative approximations.

Read More: `https://epubs.siam.org/doi/abs/10.1137/16M1080173?journalCode=siread`

PUBLIC LECTURE:

Title. Nonconvex Optimization: Opportunities and Challenges

Abstract. Much of the history of mathematical optimization has focused on theory and algorithms for solving convex problems, with much conventional wisdom holding the opinion that if a problem is not formulated to be convex, then it has been formulated incorrectly. However, in the past few years, researchers and practitioners have loosened the shackles. Now is a heyday of nonconvex optimization. In this talk, we will discuss the opportunities of embracing nonconvex optimization, but also warn about the resulting challenges. We will also discuss related topics, such as how one should compare numerical algorithms in the presence of nonconvexity and stochasticity.

2. Wotao Yin (University of California, Los Angeles)

TUTORIAL:

Title. Parallel, Distributed, and Decentralized Optimization Methods

Abstract. This talk reviews a series of recent parallel, distributed, and decentralized optimization methods. The precursor is a brief discussion of basic parallel and distributed computing concepts. The main part starts with the two most basic distributed methods: primal decomposition and dual decomposition. Next, it covers basic operator splitting

schemes and reviews various ADMM and primal-dual methods as special cases. After a short break, the tutorial continues with decentralized methods for solving a problem defined over a set of network-connected agents. They use parallel local updates to achieve global objectives. The tutorial uses a software framework called BlueFog for numerical illustrations.

PUBLIC LECTURE:

Title. Learning to Optimize

Abstract. Many applications require repeatedly solving a certain type of optimization problem, each time with new but similar data. "Learning to optimize" or L2O is an approach to develop algorithms that solve these similar problems much faster. L2O-generated algorithms have achieved significant success in signal processing and inverse-problem applications. On LPs, SAT problems, and MIPs, L2O shows promising progress in some aspects. This talk introduces the motivation for L2O and overviews different types of L2O approaches for continuous optimization. We will cover model-based approaches, which are derived from general-purpose optimization algorithms but involve (possibly many) tunable parameters, as well as model-free approaches, which use recurrent neural networks and other deep learning architectures to build algorithms. We will briefly go through plug-and-play and safeguarded L2O approaches, which incorporate learned algorithms into classic optimization frameworks.

Invited Speakers

1. Stefanie Guenther (Lawrence Livermore National Laboratory)

   Title: Simultaneous Layer-Parallel Training for Deep Residual Networks

   Abstract: Deep residual network have shown great promise for a variety of machine learning applications. Despite the rapid methodological developments, many challenges associated with very deep networks remain to be solved, such as for example the scalability barrier created by serial forward and backward propagation where training runtimes increase linearly with the number of layers, the high dimensionality of the resulting learning problem, as well as the question of initialization of the network weights.

   In this talk, we leverage recent advances in optimal control to address these challenges. In particular, a class of layer-parallel training methodologies is presented that enable concurrency across the network model. The approach is based on a continuous interpretation of deep residual learning as a problem of optimally controlling a continuous dynamical system, which will be summarized in the first part of the talk. Then, a parallel multigrid scheme is proposed that replaces the serial network propagation such that runtimes remain bounded when increasing network depth along with computational resources. The multigrid scheme further allows for coarse-grid representations of the training problem enabling effective initialization strategies. Advanced optimization strategies such as simultaneous optimization algorithms and decoupled state and control discretization approaches drawn from optimal control will be discussed.

2. Patrick O'Neil (BlackSky)

   Title: Analyzing Thin Film Morphologies using Machine Learning and Topological Data Analysis

   Abstract: This talk will detail a new open source computer package, M2PY, for analyzing thin film morphologies using unsupervised machine learning and techniques in topological data analysis. In particular, we will discuss the mathematical underpinnings of the package, including semantic segmentation using traditional approaches based on Gaussian mixture models as well as methods using homological persistence. We will also present an application of M2PY by studying a sample of $\pi-$Conjugated polymers.

3. Chris Teixeira (The MITRE Corporation)

   Title: When Machine Learning Fails

   Abstract. Analytics has a spectrum of applications that can be provided to answering different challenges across many domains and industries. The variety of ways in which data can be used has become a staple to supporting the variety of decisions that need to be made. The power that machine learning brings to solving any problem could be leading it down a path that allows companies to rely on this one solution rather than considering other analytic tools that are available. This talk will focus on the variety of

challenges that can be solved with analytics, and in particular, the types of questions you should be asking of your team or customer to increase the likelihood of success for your project.

4. Madeleine Udell (Cornell University)

Title: Scalable Semidefinite Programming

Abstract: Semidefinite programming (SDP) is a powerful framework from convex optimization that has striking potential for data science applications. This talk develop new provably correct algorithms for solving large SDP problems by economizing on both the storage and the arithmetic costs. We present two methods: one based on sketching, and the other on complementarity. Numerical evidence shows that these methods are effective for a range of applications, including relaxations of MaxCut, abstract phase retrieval, and quadratic assignment, and can handle SDP instances where the matrix variable has over 10^{13} entries.

The first method modifies a standard primal-dual convex method - the conditional-gradient augmented Lagrangian method - to store (and work with) a sketched version of the primal. A low rank approximation to the primal solution can be recovered from this sketch.

The second method begins with a new approximate complementarity principle:

Given an approximate solution to the dual SDP, the primal SDP has an approximate solution whose range is contained in the null space of the dual slack matrix. For weakly constrained SDPs, this null space has very low dimension, so this observation significantly reduces the search space for the primal solution.

This result suggests an algorithmic strategy that can be implemented with minimal storage:

(1) Solve the dual SDP approximately (with any first-order solver);
(2) compress the primal SDP to the null space of the dual slack matrix;
(3) solve the compressed primal SDP.

Contributed Talks

(1) Stephanie Allen (University of Maryland, College Park)
Using inverse optimization to learn cost functions in generalized Nash games

(2) Andersen Ang (Department of Combinatorics and Optimization, University of Waterloo)
Nonnegative unimodal matrix factorization

(3) Thomas Brown (Center for Mathematics and Artificial Intelligence, George Mason University)
Using DNNs for chemical reactions

(4) Matthias Chung (Virginia Tech)
Big data inverse problems

(5) Manuela Girotti (Montreal Institute of Learning Algorithm (MILA))
Condition numbers for first-order optimization

(6) Eswar Kumar Hathibelagal Kammara (University of Maryland, Baltimore County)
Column partition based distributed algorithms for coupled convex sparse optimization: dual and exact regularization approaches

(7) Furong Huang (University of Maryland)
Escaping from saddle points using asynchronous coordinate gradient descent

(8) Brendan Keith (Brown University)
Orbital dynamics of binary black hole systems can be learned from gravitational wave measurements

(9) Christian Kümmerle (Johns Hopkins University)
A scalable second order method for ill-conditioned matrix completion from few samples

(10) Yongchun Li (Virginia Tech)
Exact and approximation algorithms for sparse PCA

(11) Chitaranjan Mahapatra (University of California, San Francisco)
A genetic algorithm for optimal estimation of ion-channel kinetics from macroscopic currents in urinary bladder smooth muscles

(12) Nguyen-Truc-Dao Nguyen (Wayne State University)
Optimization of fully controlled sweeping processes

(13) Akwum Onwunta (Center for Mathematics and Artificial Intelligence, George Mason University)
Novel deep neural networks for solving Bayesian statistical inverse problems

(14) Vivak Patel (University of Wisconsin – Madison)
On consistency and asymptotic normality of adaptive stochastic gradient methods

(15) April Sagan (Rensselaer Polytechnic Institute)
Provable low-rank plus sparse matrix recovery via nonconvex regularizers

(16) Ramesh Sau (Indian Institute of Science, Bangalore)
Finite element analysis of the constrained Dirichlet boundary control governed by the diffusion problem

(17) Nitin Vaidya (Georgetown University)
Byzantine fault-tolerant distributed Optimization and Learning
(18) Ryan Vogt (Lawrence Livermore National Laboratory)
Optimal control Of SFQ quantum computers with binary optimal control
(19) Yunan Yang (New York University)
The implicit regularization of metrics
(20) Huaiqian You (Lehigh University)
A computational framework to machine-learn nonlocal constitutive models
(21) Roozbeh Yousefzadeh (Yale University)
Deep learning generalization and the convex hull of training sets
(22) Ming Zhong (Johns Hopkins University)
Data-driven modeling of celestial motion from modern ephemeris

Abstracts

Using Inverse Optimization to Learn Cost Functions in Generalized Nash Games

Stephanie Allen[1], **John P. Dickerson**[1], **and Steven A. Gabriel**[1]

[1]University of Maryland, College Park

As demonstrated by Ratliff et al. (2014), inverse optimization can be used to recover the objective function parameters of players in multi-player Nash games. These games involve the optimization problems of multiple players in which the players can affect each other in their objective functions. In generalized Nash equilibrium problems (GNEPs), a player's set of feasible actions is also impacted by the actions taken by other players in the game; see Facchinei and Kanzow (2010) for more background on this problem. One example of such impact comes in the form of joint/"coupled" constraints as referenced by Rosen (1965), Harker (1991), and Facchinei et al. (2007) which involve other players' variables in the constraints of the feasible region. We extend the framework of Ratliff et al. (2014) to find inverse optimization solutions for the class of GNEPs with joint constraints. The resulting formulation is then applied to a simulated multi-player transportation problem on a road network. We see that our model recovers parameterizations that produce the same flow patterns as the original parameterizations and that this holds true across multiple networks, different assumptions regarding players' perceived costs, and the majority of restrictive capacity settings and the associated numbers of players.

Nonnegative Unimodal Matrix Factorization

Andersen Ang[1]

[1]Department of Combinatorics and Optimization, University of Waterloo

We introduce a new Nonnegative Matrix Factorization (NMF) model called Nonnegative Unimodal Matrix Factorization (NuMF), which adds on top of NMF the unimodal condition on the columns of the basis matrix. NuMF finds applications for example in analytical chemistry. We propose a simple but naive brute-force heuristics strategy based on accelerated projected gradient. It is then improved by using multi-grid for which we prove that the restriction operator preserves the unimodality. We also present two preliminary results regarding the uniqueness of the solution, that is, the identifiability, of NuMF. Empirical

results on synthetic and real datasets confirm the effectiveness of the algorithm and illustrate the theoretical results on NuMF.

Using DNNs for Chemical Reactions

Thomas Brown[1]

[1]Center for Mathematics and Artificial Intelligence, George Mason University

The use of machine learning techniques to solve PDEs and ODEs is currently receiving a lot of attention. We present our contribution in this area by using DNNs with a ResNet architecture to solve a system of ODEs related to chemical kinetics. For our networks, we want the flexibility to use the solver with no prior knowledge of the time domain, and so our approach differs from much of the work done in this area (for example, PINNs). After an explanation of our method, we will present some results and future directions.

Big Data Inverse Problems

Matthias Chung[1]

[1]Virginia Tech

Emerging fields such as data analytics, machine learning, and uncertainty quantification heavily rely on efficient computational methods for solving inverse problems. With growing model complexities and ever increasing data volumes, state of the art inference method exceeded their limits of applicability and novel methods are urgently needed. Hence, new inference method need to focus on the scalability to large dimension and to address eventual model complexities.

In this talk, we discuss massive least squares problems where the size of the forward model matrix exceeds the storage capabilities of computer memory or the data is simply not available all at once. We consider randomized row-action methods that can be used to approximate the solution. We introduce a sampled limited memory row-action method for least squares problems, where an approximation of the global curvature of the underlying least squares problem is used to speed up the initial convergence and to improve the accuracy

of iterates. Our proposed methods can be applied to ill-posed inverse problem, where we establish sampled regularization parameter selection methods. Numerical experiments on very large superresolution and tomographic reconstruction examples demonstrate the efficiency of these sampled limited memory row-action methods. This is joint work with Julianne Chung, Tanner Slagel, and Luis Tenorio.

Condition Numbers for First-Order Optimization

Manuela Girotti[1]

[1]Montreal Institute of Learning Algorithm (MILA)

The theoretical study of first-order optimization algorithms often relies on second-order metrics on the objective functions (L-smoothness, μ-strong convexity and the condition number that derives from them) to tune hyperparameters and obtain theoretical convergence guarantees. Most theoretical bounds on convergence rates found in the strongly convex optimization literature are expressed in function of this condition number. We first show that adding an intense perturbation of the Hessian to some objective functions can drastically affect their condition number while having a very limited impact on their gradients. We discuss how this leads to a major weakness of these metrics and how it negatively impacts the soundness of the theoretical framework in which we study first-order optimization. We then carefully analyze the cause of this issue and consider weaker alternative conditions which we show to be more robust. We explore some of the properties achievable under these conditions as a first step toward building a sounder theory.

Column Partition based Distributed Algorithms for Coupled Convex Sparse Optimization: Dual and Exact Regularization Approaches

Eswar Kumar Hathibelagal Kammara [1]

[1]University of Maryland, Baltimore County, Baltimore, MD

In this talk, we discuss column partition based distributed schemes for a class of large-scale convex sparse optimization problems, e.g., basis pursuit (BP), LASSO, basis pursuit

denoising (BPDN), and their extensions, e.g., fused LASSO. We are particularly interested in the cases where the number of (scalar) decision variables is much larger than the number of (scalar) measurements, and each agent has limited memory or computing capacity such that it only knows a small number of columns of a measurement matrix. These problems in consideration are densely coupled and cannot be formulated as separable convex programs using column partition. To overcome this difficulty, we consider their dual problems which are separable or locally coupled. Once a dual solution is attained, it is shown that a primal solution can be found from the dual of corresponding regularized BP-like problems under suitable exact regularization conditions. A wide range of existing distributed schemes can be exploited to solve the obtained dual problems. This yields two-stage column partition based distributed schemes for LASSO-like and BPDN-like problems; the overall convergence of these schemes is established using sensitivity analysis techniques. Numerical results illustrate the effectiveness of the proposed schemes.

Escaping from Saddle Points Using Asynchronous Coordinate Gradient Descent

Furong Huang[1]

[1]University of Maryland, Baltimore County

Large-scale non-convex optimization problems are expensive to solve due to computational and memory costs. To reduce the costs, a first-order (computationally efficient) and asynchronous parallel (memory efficient) algorithm is necessary to minimize non-convex functions in machine learning. However, asynchronous-first-order methods applied within non-convex settings run into two difficulties: parallelization delays, which affect convergence, and sub-optimal saddle points where the gradient is zero. We propose an asynchronous-coordinate-gradient-descent algorithm that is shown to converge to local minima with a bounded delay. Our algorithm overcomes delay issues by reformulating a non-convex function into a Hamiltonian. The convergence rate we achieve depends logarithmically with respect to dimension and quadratically with respect to the delay. Furthermore, the convergence rate of our algorithm matches the well-known convergence rate of gradient descent to first-order stationary points (ignoring logarithmic and delay terms). To the best of our knowledge, this is the first local optima convergence result of a first-order asynchronous algorithm for non-convex settings.

Orbital Dynamics of Binary Black Hole Systems can be Learned from Gravitational Wave Measurements

Brendan Keith[1]

[1]Brown University

We introduce a gravitational waveform inversion strategy that discovers mechanical models of binary black hole (BBH) systems. We show that only a single time series of (possibly noisy) waveform data is necessary to construct the orbital trajectories and corresponding equations of motion for a BBH system. Starting with a general class of universal differential equations, our strategy involves the construction of a space of plausible mechanical models and a physics-informed constrained optimization within that space to minimize the waveform error. We choose to parameterize the space of models with feed-forward neural networks. This leads to efficient, flexible, and highly accurate algorithms which can be easily implemented with modern software libraries. We apply our method to various BBH systems including extreme and comparable mass ratio systems in eccentric and non-eccentric orbits. We show the resulting differential equations are applicable to time durations larger than the training interval, and a variety of relativistic effects, such as perihelion precession, radiation reaction, and orbital plunge, are automatically accounted for. The methods outlined here provide a new, data-driven approach to studying the dynamics of binary black hole systems. Various applications in gravitational wave astronomy will be discussed as time permits.

A Scalable Second Order Method for Ill-Conditioned Matrix Completion from Few Samples

Christian Kümmerle[1]

[1]Johns Hopkins University

We propose an iterative algorithm for low-rank matrix completion with that can be interpreted as an iteratively reweighted least squares (IRLS) algorithm, a saddle-escaping smoothing Newton method or a variable metric proximal gradient method applied to a non-convex rank surrogate. It combines the favorable data-efficiency of previous IRLS approaches with an improved scalability by several orders of magnitude. We establish the first local convergence guarantee from a minimal number of samples for that class of algorithms, showing that the method attains a local quadratic convergence rate. Furthermore, we show that the linear systems to be solved are well-conditioned even for very ill-conditioned ground truth matrices. We provide extensive experiments, indicating that unlike many state-of-the-art approaches, our method is able to complete very ill-conditioned matrices with a condition

number of up to 1010 from few samples, while being competitive in its scalability (Joint work with Claudio Mayrink Verdun).

Exact and Approximation Algorithms for Sparse PCA

Yongchun Li[1]

[1]Virginia Tech

Sparse PCA (SPCA) is a fundamental model in machine learning and data analytics, which has witnessed a variety of application areas such as finance, manufacturing, biology, healthcare. To select a prespecified-size principal submatrix from a covariance matrix to maximize its largest eigenvalue for the better interpretability purpose, SPCA advances the conventional PCA with both feature selection and dimensionality reduction. Existing approaches often approximate SPCA as a semi-definite program (SDP) without strictly enforcing the important cardinality constraint that restricts the number of selected features to be a constant. To fill this gap, we propose two exact mixed-integer SDPs (MISDPs) by exploiting the spectral decomposition of the covariance matrix and the properties of the largest eigenvalues. We then analyze the theoretical optimality gaps of their continuous relaxation values and prove that they are stronger than that of the state-of-art one. We further show that the continuous relaxations of two MISDPs can be recast as saddle point problems without involving semi-definite cones, and thus can be effectively solved by first-order methods such as the subgradient method. Since off-the-shelf solvers, in general, have difficulty in solving MISDPs, we approximate SPCA with arbitrary accuracy by a mixed-integer linear program (MILP) of a similar size as MISDPs. The continuous relaxation values of two MISDPs can be leveraged to reduce the size of the proposed MILP further. To be more scalable, we also analyze greedy and local search algorithms, prove their first-known approximation ratios, and show that the approximation ratios are tight. Our numerical study demonstrates that the continuous relaxation values of the proposed MISDPs are quite close to optimality, the proposed MILP model can solve small and medium-size instances to optimality, and the approximation algorithms work very well for all the instances. Finally, we extend the analyses to Rank-one Sparse SVD (R1-SSVD) with non-symmetric matrices and Sparse Fair PCA (SFPCA) when there are multiple covariance matrices, each corresponding to a protected group.

A Genetic Algorithm for Optimal Estimation of Ion-Channel Kinetics from Macroscopic Currents in Urinary Bladder smooth muscles

Chitaranjan Mahapatra[1]

[1]University of California, San Francisco

Conductance-based compartment modeling requires tuning of many parameters to fit the excitable cell model to target electrophysiological data. Markov modeling provides an effective approach for modeling ion channel kinetics. There are several search algorithms for global fitting of macroscopic or single-channel currents across different experimental conditions. Automated parameter optimization via evolutionary algorithms (EAs) is a common approach to accomplish this task, using error functions to quantify differences between model and target. Here, a topology-mutating genetic algorithm that searches for the best state diagram and transition-rate parameters to model macroscopic ion-channel behavior is described. Important features of the algorithm include a topology-altering strategy, automatic satisfaction of equilibrium constraints (microscopic reversibility), and multiple protocol fitting using sequential goal programming rather than explicit weighting. Application of this genetic algorithm to design a voltage-gated calcium channel model exhibiting both fast and prolonged inactivation yields a six-state model that produces realistic activity dependent action-potential generation in current-clamp simulations of a urinary bladder smooth muscle cells.

Optimization of Fully Controlled Sweeping Processes

Nguyen-Truc-Dao Nguyen[1]

[1]Wayne State University

This talk addresses a new class of optimal control problems for perturbed sweeping processes which are governed by the maximal monotone mappings. We develop a constructive discrete approximation procedure, employ advanced tools of first-order and second-order variational analysis and generalized differentiation, derive numerical algorithms to compute second-order constructions, and necessary optimality conditions for discrete optimal solutions under fairly general assumptions formulated entirely in terms of the given data. The obtained results give us efficient suboptimality ("almost optimality") conditions for the original sweeping control problem that are illustrated by several numerical examples.

This is based on joint work with Tan Cao, Giovanni Colombo, Boris Mordukhovich.

Novel Deep Neural Networks for Solving Bayesian Statistical Inverse Problems

Akwum Onwunta[1]

[1]Center for Mathematics and Artificial Intelligence, George Mason University

We consider the simulation of Bayesian statistical inverse problems governed by large-scale linear and nonlinear partial differential equations (PDEs). Markov chain Monte Carlo (MCMC) algorithms are standard techniques to solve such problems. However, MCMC techniques are computationally challenging as they require several thousands of forward PDE solves. The goal of this paper is to introduce a fractional deep neural network based approach for the forward solves within an MCMC routine. Moreover, we discuss some approximation error estimates and illustrate the efficiency of our approach via several numerical examples.

On Consistency and Asymptotic Normality of Adaptive Stochastic Gradient Methods

Vivak Patel[1]

[1]University of Wisconsin – Madison

Adaptive stochastic gradient methods have become integral to the training of machine learning models. While these adaptive methods have have enjoyed varying numerical success across a number of problems, their analysis has not kept pace with their practice. In this talk, we will make progress on this issue by stating a generic consistency result for a large class of adaptive stochastic gradient methods on a broad class of convex problems. Moreover, once we have stated the consistency result, we will then state a general asymptotic normality result that relies on a novel strategy that is distinct from the moment generating strategy of Chung and Sacks, and the averaging strategy of Polyak and Juditsky. We will then apply our general results to specific adaptive stochastic gradient methods.

Provable Low-Rank Plus Sparse Matrix Recovery Via Nonconvex Regularizers

April Sagan[1]

[1]Rensselaer Polytechnic Institute

We study the use of nonconvex regularizers on a large class of problems where we seek to recover a low rank matrix and/or sparse vector from some set of measurements. While methods based off of convex relaxations suffer from a (possibly large) estimator bias, and other nonconvex methods require the rank or sparsity to be known a priori, we utilize nonconvex regularizers to minimize the rank and l_0 norm without the estimator bias from the convex relaxation. We present a novel analysis of the alternating proximal gradient descent algorithm applied to such problems, and bound the error between the iterates and the ground truth sparse and low rank matrices. The algorithm and error bound can be applied to sparse optimization, matrix completion, and robust principal component analysis as special cases of our results.

Finite Element Analysis of the Constrained Dirichlet Boundary Control Governed by the Diffusion Problem

Ramesh Sau[1]

[1]Indian Institute of Science, Bangalore

We study an energy space based approach for the Dirichlet boundary optimal control problem governed by the Laplace equation with control constraints. The optimality system results in a simplified Signorini type problem for control which is coupled with boundary value problems for state and costate variables. We propose a finite element based numerical method using the linear Lagrange finite element spaces with discrete control constraints at the Lagrange nodes. The analysis is presented in a combination for both the gradient and the L^2 cost functional. *A priori* error estimates of optimal order in the energy norm is derived up to the regularity of the solution for both the cases. Theoretical results are illustrated by some numerical experiments.

Byzantine Fault-Tolerant Distributed Optimization and Learning

Nitin Vaidya[1]

[1]Georgetown University

Consider a network of agents wherein each agent has a private cost function. In the context of distributed machine learning, the private cost function of an agent may represent the loss function corresponding to the agents local data. The objective here is to identify parameters that minimize the total cost over all the agents. In machine learning for classification, the cost function is designed such that minimizing the cost function should result model parameters that achieve higher accuracy of classification. Similar problems arise in the context of other applications as well, including swarm robotics.

This talk will discuss Byzantine fault-tolerant (or secure) algorithms for distributed optimization with applications to machine learning. In the context of machine learning, the goal here is to be able to learn the model parameters correctly while tolerating adversarial agents that may supply incorrect information. When a large number of agents participate in distributed optimization, security compromise of some of the agents becomes increasingly likely. We constructively show that such secure algorithms for distributed optimization exist. The talk will provide intuition behind the design and correctness of the algorithms.

Optimal Control Of SFQ Quantum Computers With Binary Optimal Control

Ryan Vogt[1]

[1]Lawrence Livermore National Laboratory

At the moment the largest stable quantum computers are on the order of one-hundred qubits. Theoretical work has indicated that quantum computers must be at least the size of one-million qubits, and preferably one-hundred million qubits, so that quantum algorithms can solve challenging problems in such areas as combinatorial optimization, quantum chemistry, and linear systems, that classical computers cannot. One instance of a quantum computer, single-flux quantum (SFQ) quantum computers, offer a promising future as the first system that can realize the goal of one million qubits. These computers send instructions with microwave pulses that influence the quantum system to perform calculations on the quantum computer. We introduce a binary optimal control problem, where we decide whether or not to send microwaves pulses in time to realize a quantum computation (gate) on the scale of nanoseconds. We show that we are able to realize several standard gates such as the H, X, Y, and Z gate successfully ($> 99.9\%$ probability) for a single qubit, and

emphasize how our approach is readily applicable, and scalable, to many qubit systems. This is a joint work with Anders Petersson at Lawrence Livermore National Laboratory.

The Implicit Regularization of Metrics

Yunan Yang[1]

[1]New York University

Recently, we proposed the quadratic Wasserstein distance from optimal transport theory for inverse problems, tackling the classical least-squares method's longstanding difficulties such as nonconvexity and noise sensitivity. As we advance, we discover that the advantage of changing the data misfit is more general in a broader class of data-fitting problems by examining the preconditioning and "implicit" regularization effects of different mathematical metrics as the objective function in optimization, as the likelihood function in Bayesian inference, and as the measure of residual in numerical solution to PDEs.

A Computational Framework to Machine-Learn Nonlocal Constitutive Models

Huaiqian You[1]

[1]Lehigh University

We propose data-driven machine-learning algorithms for the identification of nonlocal constitutive models in the context of simulation of materials with a heterogeneous microstructure. New nonlocal kernel functions, embedding the microstructure behavior, are derived from high-fidelity data obtained by microscale simulations or measurements. Our model surrogates exhibit excellent generalization properties when tested on problems that are substantially different from the ones used for training. Numerical tests in one and two dimensions illustrate the consistency of our methods and the applicability of our algorithms to complex physical systems. Applications include wave propagation in heterogeneous materials and molecular dynamics.

Deep Learning Generalization and the Convex Hull of Training Sets

Roozbeh Yousefzadeh[1]

[1]Yale University

We study the generalization of deep learning models in relation to the convex hull of their training sets. A trained image classifier partitions its domain via decision boundaries and assigns a class to each of those partitions. The location of decision boundaries inside the convex hull of training set can be investigated in relation to the training samples. However, our analysis shows that in standard image classification datasets, all testing images are considerably outside that convex hull, in the pixel space, in the wavelet space, and in the internal representations learned by deep networks. Therefore, the performance of a trained model partially depends on how its decision boundaries are extended outside the convex hull of its training data. From this perspective which is not studied before, over-parameterization of deep learning models is necessary for shaping the extension of decision boundaries. At the same time, over-parameterization should be accompanied by a specific training regime, in order to yield a model that not only fits the training set, but also its decision boundaries extend desirably outside the convex hull. To illustrate this, we investigate the decision boundaries of a neural network, with various degrees of parameters, inside and outside the convex hull of its training set. Moreover, we use a polynomial decision boundary to study the necessity of over-parameterization and the influence of training regime in shaping its extensions outside the convex hull of training set.

Data-driven Modeling of Celestial Motion from Modern Ephemeris

Ming Zhong[1]

[1]Johns Hopkins University

The aspiration to accurately model the underlying mechanisms of celestial motion has led to affluent developments throughout the history of theoretical physics. Many fundamental theories have been developed to explain observations, and predict future positions, of planets and stars as faithfully as possible. Our data-driven learning approach, which was

first developed in Lu et al. (2019) and then extended in Zhong et al. (2020), can derive stable and accurate models of celestial motion. These models are based on a collective dynamics framework, and are learned from the NASA Jet Propulsion Laboratory's Modern Ephemeris. By modeling the major Astronomical Objects in the Solar system as pairwise interacting agents, we are able to learn interaction kernels from the Ephemeris data and generate dynamics accurate enough to produce a Precession Rate of 544 arc-seconds per Earth-century for the estimated orbit of Mercury. Compared to the theoretical rate of 532 based on Newtonian gravitation, our collective dynamics model is able to capture a portion of the general relativistic correction to the precession directly from observation data.