East Coast Optimization Meeting 2020

Dates

April 2-3, 2020

Location

GMU, Fairfax

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'll discuss how optimization problems arise in machine learning and what makes them challenging. Second, we'll 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'll 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'll discuss the opportunities of embracing nonconvex optimization, but also warn about the resulting challenges. We'll 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 computational complexity in parallel computing. The main part starts with basic parallelizable operators such

as finite sum and those with certain separable structures, which are simple to parallelize. Next, it presents basic operator splitting schemes and reviews various ADMM and primal-dual methods as special cases. After a short break, the tutorial continues with coordinate-friendly structures and discusses parallel methods in classic learning and conic optimization. The last part focuses on decentralized methods for solving a problem de?ned over a set of network-connected agents. They use parallel local updates to achieve global objectives. The tutorial closes with a discussion on open questions and future research.

PUBLIC LECTURE:

Title. A Method for Vertical Federated Learning

Abstract. Federated learning (FL) is a form of distributed learning in which training data are scattered and must be secured by local agents. FL is to enable model training with multi-agent data where privacy and data security prevent us from using the existing machine learning algorithms. Besides security, agents in the FL setting typically have heterogeneous capacities of computing powers and network connections. This talk considers the scenario where different agents have different data describing the same set of subjects, also known as Vertical FL. A method that integrates asynchronous block-coordinate descent, differential privacy, and local embedding is presented. The method has provable convergence performance along with data privacy guarantees. It has the potential to facilitate the collaborations of banks, hospitals, and cross-domain business to improve their services for their joint customers. This talk is based on a joint work with Tianyi Chen and Xiao Jin at RPI and Yuejiao Sun at UCLA.

Invited Speakers

1. Stefanie Guenther (Lawrence Livermore National Laboratory)

   Title: Simultaneous Layer-Parallel Training for Deep Residual Networks

   Abstract: Deep residual networks (ResNets) have shown great promise to model complex data relations with applications in image classification, speech recognition, or text processing, among others. Despite the rapid methodological developments, compute times for ResNet training however can still be tremendous, measured in the order of hours or even days. While common approaches to decrease the training runtimes mostly involve data-parallelism, the sequential propagation through the network layers creates a scalability barrier where training runtimes increase linearly with the number of layers. This talk presents an approach to enable concurrency across the network layers and thus overcome this scalability barrier. The proposed method is inspired by the fact that training networks of residual type can mathematically be interpreted as an optimal control problem. In this context, the discrete network layers are interpreted as the discretization of a time-continuous dynamical system. Recent advances in parallel-in-time integration and optimization methods can thus be leveraged in order to speed up training runtimes. In particular, an iterative multigrid-reduction-in-time approach will be discussed, which recursively divides the time domain (i.e. the layers) into multiple time chunks that can be processed in parallel on multiple compute units. Additionally, the multigrid iterations enable a simultaneous optimization framework where weight updates are based on inexact gradient information.

2. Patrick O'Neil (BlackSky)

   Title: xxxxx

   Abstract: xxxxx

3. Chris Teixeira (MITRE)

   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 breadth of prescriptive analytics that can be used when machine learning applications aren't applicable. Prescriptive analytics provides stakeholders with a robust, data-driven approach to exploring "what if" scenarios to the challenges they face.

4. Madeleine Udell (Cornell University)

   Title: xxxx

   Abstract: xxxx

Contributed Talks

Abstracts

List of Participants

1. xxx