samsi
NSF·Duke·NCSU·UNC

Optimization Program WISO Workshop
February 8-10, 2017

SPEAKER TITLES/ABSTRACTS

Larry Biegler
Carnegie Mellon University

"Model Discrimination and Parameter Estimation for Complex Reactive Systems"

Parameter estimation and model discrimination of reaction kinetics with limited (informative) measurement data remains an important and challenging problem. To address this problem we propose a strategy based on maximum likelihood principles, non-linear optimization techniques, as well as collocation methods that discretize the differential-algebraic model. Within this approach we estimate variances of the noise in system variables (e.g. concentrations) from measurements, assess the sensitivity of model parameters, determine state profiles and model parameters, and discriminate among physics-based candidate models. Key components of this strategy are large-scale nonlinear programming (NLP) and post-optimality analysis. The resulting approach is presented on two case studies drawn from real-world processes for the manufacture of specialty chemicals.

John Burns
Virginia Tech

"Parameter Identification for Dynamical Systems with Structured Uncertainty"

In this talk we consider problems of parameter identification for models defined by dynamical systems with structured uncertainty. This type of problem often occurs in science and engineering and we show how parameter estimation can be improved by taking advantage of the specific structure of the model form and by using physics to define constraints on the corresponding optimization problem. The results are applicable to problems where, in the process of modeling a complex system, one ignores "small" parameters to obtain simplified models and when infinite dimensional systems are approximated to obtain finite dimensional models. We show for these common sources of model discrepancies, hierarchical modeling can be employed to aid in the development of prior knowledge about the model form caused by uncertain or ignored parameters. We focus on a special class of problems arising from parameter estimation and illustrate how one can use information about the structure of the dynamical system to help deal with model discrepancy. Examples from population dynamics and thermal fluid systems are given to illustrate the ideas

Jack Dongarra
University of Tennessee

"The Road to Exascale and Legacy Software for Dense Linear Algebra"

In this talk we will look at the current state of high performance computing and look at the next stage of extreme computing. With extreme computing, there will be fundamental changes in the character of floating point arithmetic and data movement. In this talk we will look at how extreme scale computing has caused algorithm and software developers to changed their way of thinking on how to implement and program certain applications.

Yonina Eldar
Technion-Israel Institute of Technology

"Phase Retrieval and Analog to Digital Compression"

The problem of phase retrieval, namely – the recovery of a function given the magnitude of its Fourier transform - arises in various fields of science and engineering, including electron microscopy, crystallography, astronomy, and optical imaging. Due to the loss of Fourier phase information, this problem is generally ill-posed. In this talk we review several modern methods for treating the phase retrieval problem which are based on advanced optimization tools and statistical analysis. We then show how these concepts can be used to tackle a very different set of nonlinear problems: low-rate analog to digital conversion without assuming any structure on the signal being sampled. This is possible by careful design of the measurement scheme, together with advanced nonlinear recovery methods. We end by demonstrating our sub-Nyquist methods via several prototypes developed in our lab for cognitive radio and ultrasound imaging.

Jianqing Fan
Princeton University

"Sparse Learning and Distributed PCA with Control of Statistical Errors and Computing Resources"

High-dimensional sparse learning and analysis of Big Data data pose significant challenges on computation and communication. Scalable statistical procedures need to take into account both statistical errors and computing resource constraints. This talk illustrate this idea by using two important examples in statistical machine learning. The first one is to solve sparse learning via a computational framework named iterative local adaptive majorize-minimization (I-LAMM) to simultaneously control algorithmic complexity and statistical error when fitting high dimensional sparse models via a family of folded concave penalized quasi-likelihood. The algorithmic complexity and statistical errors are explicitly given and we show that the algorithm achieves the optimal statistical error rate under the weakest signal strength assumptions. The second problem is to study distributed PCA with communication constraints: each node machine computes the top eigenvectors and communicates to the central server; the central server then aggregates the information transmitted from the node machines and conduct another PCA based on the aggregated information. We investigate the bias and variance for such a distributed PCA. We derive the rate of convergence for distributed PCA, which depends explicitly on effective rank, eigen-gap, and the number of machines, and show that the distributed PCA performs as well as the whole sample PCA, even without full access of whole data.

Michael I. Jordan
University of California, Berkeley

"On Gradient-Based Optimization: Accelerated, Distributed, Asynchronous and Stochastic"

Many new theoretical challenges have arisen in the area of gradient-based optimization for large-scale statistical data analysis, driven by the needs of applications and the opportunities provided by new hardware and software platforms. I discuss several recent results in this area, including: (1) a new framework for understanding Nesterov acceleration, obtained by taking a continuous-time, Lagrangian/Hamiltonian perspective, (2) a general theory of asynchronous optimization in multi-processor systems, (3) a computationally-efficient approach to variance reduction, and (4) a primal-dual methodology for gradient-based optimization that targets communication bottlenecks in distributed systems.

Arkadi Nemirovski
Georgia Institute of Technology

"On Statistical Inferences via Convex Optimization"

The talk presents applications of Convex Optimization Theory (saddle points, duality,...) to designing provably near-optimal ``computation-friendly" statistical inferences, computational friendliness meaning that the inferences and their risks are yielded by efficient computation. Examples include: (1) pairwise and multiple hypothesis testing in Gaussian, Poisson, and Discrete observation schemes, provided the hypotheses are specified by convex constraints on the parameters of observed distributions, (2) recovery of linear and linear-fractional functions of a ``signal" known to belong to finite union of given convex sets, via indirect observations of the signal, (3) recovery, in Euclidean norm, of linear image of a signal known to belong to a given convex set via indirect observations of the signal in Gaussian noise.

The talk is based on joint research with A. Iouditski, University Grenoble-Alpes, and A. Goldenshluger, Haifa University.

Alexander Shapiro
Georgia Institute of Technology

"Statistical Inference of Empirical Estimates of Stochastic Programs"

In this talk we discuss statistical properties of empirical (sample) estimates of the optimal value and optimal solutions of stochastic programs. Examples include the classical Maximum Likelihood method, Sample Average Approximation (SAA) approach to two and multistage stochastic programming and risk averse stochastic optimization. We give a survey of consistency (Law of Large Numbers), asymptotics (Central Limit Theorem) and Large Deviations type results. In particular, this has an application to investigation of sample complexity of two and multi-stage stochastic programming.

Bernd Sturmfels
University of California, Berkeley

"Tensors and their Eigenvectors"

Eigenvectors of square matrices are central to linear algebra. Eigenvectors of tensors are a natural generalization. The spectral theory of tensors was pioneered by Lim and Qi over a decade ago, and it has found numerous applications. This lecture offers a first introduction, with emphasis on algebraic aspects.

Martin Wainwright
University of California, Berkeley

"Statistics meets Optimization: Fast Randomized Algorithms for large data sets"

Large-scale data sets are now ubiquitous throughout engineering and science, and present a number of interesting challenges at the interface between statistics and optimization. In this talk, we discuss the use of randomized dimensionality reduction techniques, also known as sketching, for obtaining fast but approximate solutions to large-scale convex programs. Using information-theoretic techniques, we first reveal a surprising deficiency of the most widely used sketching technique. We then show how a simple iterative variant leads to a much faster algorithm, and one which adapts to the intrinsic dimension of the solution space. Moreover, we show how it generalizes naturally to a randomized version of the Newton algorithm with provable guarantees.

Margaret Wright
Courant Institute-New York

"Old, New, Borrowed, and Blue in the Marriage of Statistics and Optimization"

Statistics and optimization have been closely connected for decades, sometimes accompanied, led, or followed by numerical analysis and computer science. The speaker will look closely at revealing instances in their joint history, with special emphasis on what each has brought and is bringing to the partnership.

Stephen Wright
University of Wisconsin-Madison

"Randomness in Coordinate Descent"

Coordinate descent is a basic approach to nonlinear optimization that has become popular again because of its relevance in data analysis. Despite the simplicity of this approach, different variants show unexpected behavior even on the simplest problems. Here we examine three variants of coordinate descent, applied to convex quadratic programming, focusing on problems in which the cyclic variant is much slower than randomized cyclic and fully random variants. An analysis of the convergence behaviors of these variants, is surprisingly technical, though it uses only elementary mathematical tools. This talk will outline the results and survey related work.