samsi
NSF·Duke·NCSU·UNC

Optimization Program Opening Workshop
August 29 - September 2, 2016

SPEAKER TITLES/ABSTRACTS

Genevera Allen
Rice University

"Algorithmic Regularization Paths: A New Approach to Variable Selection"

Variable selection has become a cornerstone of statistical machine learning and is ubiquitously used for the analysis of high-dimensional data. While there is a robust literature on providing guarantees for the performance of variable selection techniques, there is one setting for which there are no such guarantees and existing methods perform poorly: that of high-dimensional high-correlation (HDHC) data. Such HDHC data commonly arises from high-throughput biomedical technologies and in graph selection problems for highly connected graphs. In this paper, we develop a radically different type of variable selection method that will prove to be superior in HDHC settings. Our so-called Algorithmic Regularization Paths generate a sequence of sparse models as the iterates of an algorithm inspired by the Alternating Direction Methods of Multiplers (ADMM) algorithm for solving the Lasso. In this paper, we introduce our method, discuss its origin, study its theoretical properties to better understand how it works, and draw connections to existing regularization techniques. Empirical studies show that our Algorithmic Regularization Paths yield better performance in terms of prediction accuracy, variable selection, model selection, and computing time than all existing approaches in HDHC settings.

Joint work with Yue Hu & Michael Weylandt.

Mihai Anitescu
Argonne National Laboratory

"Mathematical Challenges of Energy Systems"

In this talk we discuss existing difficulties in modeling and computing energy systems and the important role that mathematical modeling and optimization can play in solving them. We argue that in some cases a change of the problem framework may be desirable and that this may be made while keeping the solution computationally achievable. We outline a number of existing and emerging fundamental research challenges and discuss some recent promising avenues in the area. A distinguishing feature of power grid applications is that optimization is ubiquitous and that it must accommodate simultaneously multiple complexity drivers. These include not only discrete variables, nonconvexity, or stochasticity but also ordinary and — with the increased usage of natural gas — partial differential equations. We discuss the productivity and performance implications of this fact for the modeling and computational environments.

Amr El-Bakry
Exxon

"The Evolving Optimization Needs and Challenges in the Oil and Gas Industry"

Optimization has been a key modeling and algorithmic framework for inference and decision-making support in the oil and gas industry. In this talk, I will review examples of past and current optimization in the upstream sector and discuss some developing needs and challenges arising from the current state of the industry

Maryam Fazel
University of Washington

"Variational Functions of Gram Matrices: Properties and Optimization Algorithms"

We study a class of convex penalty functions, called Variational Gram Functions that can promote pairwise relations such as orthogonality among a set of vectors in a vector space. These functions of a Gram matrix find various applications in machine learning, e.g., in hierarchical classification and multitask learning. We give a general condition for convexity of these functions, and use it to prove the convexity of a few known functions as well as some new ones. We discuss efficient optimization algorithms for loss-minimization problems regularized with these penalty functions. Numerical experiments on a hierarchical classification problem show the effectiveness of the penalties and the associated optimization algorithms.

Michael Ferris
University of Wisconsin

"Modeling and Optimization within Interacting Systems"

We consider models built up from a collection of optimizations within an interacting physical, economic or virtual system. We show how optimization and equilibrium concepts can be deployed and resulting models solved within an extended mathematical programming framework. Examples are drawn from sustainable land use modeling, power system design and economic operation, discrete Nash equilibria and risk analysis. The interplay between stochasticity, complementarity and hierarchical optimization will be highlighted.

Mark Girolami
Warwick, UK

"Probabilistic Numerics: A New Concept?"

The vast amounts of data in many different forms becoming available to politicians, policy makers, technologists, and scientists of every hue presents tantalising opportunities for making advances never before considered feasible.

Yet with these apparent opportunities has come an increase in the complexity of the mathematics required to exploit this data. These sophisticated mathematical representations are much more challenging to analyse, and more and more computationally expensive to evaluate. This is a particularly acute problem for many tasks of interest such as making predictions since these will require the extensive use of numerical solvers for linear algebra, optimization, integration or differential equations. These methods will tend to be slow, due to the complexity of the models, and this will potentially lead to solutions with high levels of uncertainty.

This talk will introduce our contributions to an emerging area of research defining a nexus of applied mathematics, statistical science and computer science, called “probabilistic numerics”. The aim is to consider numerical problems from a statistical viewpoint, and as such provide numerical methods for which numerical error can be quantified and controlled in a probabilistic manner. This philosophy will be illustrated on problems ranging from predictive policing via crime modelling to computer vision, where probabilistic numerical methods provide a rich and essential quantification of the uncertainty associated with such models and their computation.

Matthias Heinkenschloss
Rice University

"PDE Constrained Optimization with Uncertain Data"

Many science and engineering problems lead to optimization problems governed by partial differential equations (PDEs), and in many of these problems some of the problem data are not known exactly. I focus on a class of such optimization problems where the uncertain data are modeled by random variables or random fields, and where decision variables (controls/designs) are deterministic and have to be computed before the uncertainty is observed. It is important that the uncertainty in problem data is adequately incorporated into the formulation of the optimization problem. The numerical solution of all such formulations require a sampling of random variables. Since the number of PDEs that have to be solved depends linearly on the number of samples, it is important key the overall number of samples low.

I will review problem formulations, problem discretization/sampling approaches, and their implications for the numerical solution of PDE constrained optimization with uncertain data.

Philipp Hennig
Tuebingen

“# Numerical Optimization as Probabilistic Inference”

Like all numerical methods, optimizers *estimate* an unknown quantity — the locus of roots of a gradient — given the result of tractable, `observable’ computations — gradient values at evaluation nodes. Optimization can thus be seen as an instance of inference in the statistical sense. Probabilistic numerics formalizes this connection by phrasing computation as the manipulation of probability measures. I will outline some results showing that classic methods, like iterative linear solvers and quasi-Newton methods, can be captured naturally in this framework, and present recent work on using this formalism to address sources of uncertainty arising in the big data setting. I will conclude with an outlook on open problems.

Ekaterina Kostina
Heidelberg University

“Numerical Methods for Optimization-based Model Validation”

Mathematical models are of great importance for natural sciences and engineering. Besides providing scientific insight into processes, mathematical models are fundamental for process simulation, optimization and control. However, the results from simulation and optimization can be only trusted as a basis for decision and control if an underlying model describes a given process quantitatively and qualitatively correctly. This implies a model which is validated by experimental data with sufficiently good estimates for model parameters. The development and quantitative validation of complex nonlinear differential equation models is a difficult task that requires the support by numerical methods for sensitivity analysis, parameter estimation, and the optimal design of experiments.

The talk will adress new developments in optimization methods for validation of models, in particular design of robust optimal experiments based on a second order sensitivity analysis of parameter estimates and design of optimal experiments for PDE models.

Sven Leyffer
Argonne National Laboratories

“Mixed-Integer PDE-Constrained Optimization”

Many complex applications can be formulated as optimization problems constrained by partial differential equations (PDEs) with integer decision variables. Examples include the remediation of contaminated sites and the maximization of oil recovery; the design of next generation solar cells; the layout design of wind-farms; the design and control of gas networks; disaster recovery; and topology optimization.

We will present emerging applications of mixed-integer PDE-constrained optimization, review existing approaches to solve these problems, and highlight their computational and mathematical challenges. We introduce a new set of benchmark set for this challenging class of problems, and present some early numerical experience using both mixed-integer nonlinear solvers and heuristic techniques.

Youssef Marzouk
MIT

"Measure Transport and Optimization-based Samplers for Bayesian Computation"

We will discuss how transport maps, i.e., deterministic couplings between probability measures, can enable useful new approaches to Bayesian computation and the Bayesian solution of large-scale inverse problems.

First, we discuss a variational approach to Bayesian inference that constructs a deterministic transport map from a reference distribution to the posterior, without resorting to MCMC. Independent and unweighted posterior samples can then be obtained by pushing forward reference samples through the map. This approach involves solving a stochastic optimization problem over a space of candidate transport maps. Crucially, the conditional independence structure of the target and/or the presence of low-dimensional likelihood-informed subspaces lead to useful restrictions on the form of the transport, and thus make the representation and computation of high-dimensional maps tractable.

We will also discuss other recent algorithms for optimization-based sampling, such as implicit sampling and randomize-then-optimize (RTO)---and interpret these optimization-based samplers through the lens of measure transport. Rather than optimizing over spaces of candidate transports, optimization-based samplers repeatedly solve minimization problems that yield the action of a particular transport. From this analysis, several new variants of these optimization-based samplers---e.g., with local proposals, mixture proposals, and adaptation---follow naturally.

This is joint work with Daniele Bigoni, Alessio Spantini, and Zheng Wang.

Rahul Mazumder
MIT

"Sparse Multivariate Statistics with Discrete Optimization"

Several statistical estimation tasks arising in modern multivariate statistics are naturally posed as discrete optimization problems. While continuous convex optimization methods have played a highly influential role in these tasks, the role of modern discrete optimization methods, namely, integer programming has been relatively less explored, despite the tremendous advances in the field over the past 10-15 years. In this talk I will describe how techniques in modern optimization: mixed integer optimization, first order methods in nonlinear optimization provides a systematic algorithmic lens to address some key problems in sparse multivariate statistics. I will illustrate this new approach with examples in variable selection in regression, robust statistical regression, function estimation and factor analysis.

Angelia Nedich
Arizona State University

"Distributed Algorithms for Optimization and Nash Games on Graphs"

The talk will focus on distributed computational models for solving some specially structured optimization problems and Nash games on time-varying graphs. The first part will be focused on a multi-agent optimization problems where the objective is to minimize a sum of agents' local objectives. The agent system is embedded in a communication network, modeled as a time-varying graph, over which the agents communicate locally with their one-hop neighbors. Using the graph models and optimization techniques, distributed algorithms will be discussed together with their convergence and convergence rate properties. / In the second part, a class of aggregative Nash games will be considered, where / each player's objective is a function of the aggregate of all the players' decisions. The player's information is however constrained to local players only, as given by a connectivity network in which the players are embedded. Distributed synchronous and asynchronous algorithms for reaching a Nash point will be presented. / In both parts, the results will demonstrate the dependence of the methods' performance on the problem data and the network properties, mainly, the network capability to diffuse the information.

Noemi Petra
University of California, Merced

"Bayesian Inversion Applied to an Ice Sheet Flow Problem and to Power Grid"

We consider the estimation of the uncertainty in the solution of (large-scale) inverse problems within the framework of Bayesian inference. We approximate the posterior probability density with a Gaussian. Computing the mean of this Gaussian, i.e., the maximum a posteriori (MAP) estimate of the posterior distribution, requires the solution of a regularized least-squares optimization problem governed by systems of differential equations. The optimization problem is solved with an efficient adjoint-based Newton-conjugate gradient method, which uses first and second derivative information of the negative log posterior. The posterior covariance matrix is given (in the linear-Gaussian case) by the inverse of the Hessian of the negative log posterior. To alleviate the computational cost, we exploit the compact nature of the Hessian of the data misfit term and construct a low rank approximation at a cost (in number of PDE solves) that does not depend on the problem size, thus providing scalability to problem sizes of practical interest. We apply this method to quantify the uncertainties in the inference of the basal sliding coefficient in an ice sheet inverse problem and a similar technique for the inversion of the inertia parameter of generators in a power grid problem.

Cynthia Phillips
Sandia National Laboratories

"Highly Scalable Parallel Branch and Bound"

I will discuss PEBBL (Parallel Enumeration and Branch-and-Bound Library), a highly scalable solver for general combinatorial optimization problems. I will discuss some of the performance and usability features of PEBBL, and demonstrate its performance on a problem motivated by boosting in machine learning applications. On an open database of spam classification data, PEBBL scales essentially linearly to over 6000 processors. I will discuss using this as a base for parallel integer programming. The goal of the talk is to introduce PEBBL as a possible building block for solvers for more complex problems to be built (or started) during the SAMSI optimization year.

Fred Roosta
University of California, Berkeley

"Sub-sampled Newton Methods: Uniform and Non-Uniform Sampling"

Many data analysis applications require the solution of optimization problems involving a sum of large number of functions. We consider the problem of minimizing a sum of n functions over a convex constraint set. Algorithms that carefully sub-sample to reduce n can improve the computational efficiency, while maintaining the original convergence properties. For second order methods, we first consider a general class of problems and give quantitative convergence results for variants of Newtons methods where the Hessian or the gradient is uniformly sub-sampled. We then show that, given certain assumptions, we can extend our analysis and apply non-uniform sampling which results in modified algorithms exhibiting more robustness and better dependence on problem specific quantities, such as the condition number.

Volker Schulz
University of Trier

"New Results on Optimization in Shape Manifolds"

Manifolds as such are often envisioned as curvy shapes itself. However, in this talk, we discuss manifolds of shapes. That means that each point on this manifold is itself a shape. Although this might seem, at the first glance, as an almost too sophisticated abstraction, we will see that this point of view provides a very convenient framework for shape optimization and is able to resolve puzzles which have been severe obstacles for the development of efficient numerical algorithms in this field of research. The concept of shape manifolds has been introduced in differential geometry without a focus on shape optimization and thus in a setting which requires far too much smoothness for PDE constrained shape optimization, where one has to live with Sobolev spaces. In this talk, novel results on the interplay of shape manifolds and shape optimization are presented as well as their consequences for efficient and robust shape optimization algorithms in the context of system models defined by partial differential equations.

Jewell Thomas
MaxPoint

"Scalability and Data Synthesis in Computational Advertising and Marketing"

When you visit the New York Times website and see an advertisement that encourages you to purchase, say, a specific brand of hand sanitizer, this most likely means that an algorithm designed by a marketing technology company such as ours---acting on behalf of a seller of hand sanitizer---decided to place a bid on that specific ad space at that moment in time, and that this bid exceeded the other bids that were placed by other algorithms put in place by other companies. These bids are not randomly distributed across users, but rather bring to bear multiple sources of information about your characteristics as a user, and some knowledge about how likely users with these characteristics are to take some desired action as a result of the ad (such as hovering the mouse over or clicking on the ad, downloading a coupon, or visiting a particular retailer that sells the advertised product).

The success of companies such as ours therefore depends on solving an extremely challenging optimization problem: we must bring as many sources of information to bear on these real-time decisions as possible, weight different sources of information appropriately, and synthesize multiple sources of information in a way that can scale to very large numbers of users (we evaluate over 100 billion opportunities per day) in extremely narrow time windows (less than 6ms from the time the user navigates to the page, on average).

We present a handful of case studies of optimization problems we face in our business, and describe some implemented approaches to these problems. The goal of our presentation is to underscore practical challenges to optimization techniques applied to "real-world" data and to invite collaboration on and discussion of some of the most pressing challenges in internet-scale data science.

Jonathan Weare
University of Chicago

"Toward more Faithful Simulations of Cascade Failures in Power Systems"

I will report on our efforts to develop tools allowing accurate simulation of cascade failure events, in which a significant portion of the lines in a power network drop out of service within rapid succession. An event of this type was responsible for the 2003 blackout of much of the northeastern United States and parts of Canada. Recently introduced more detailed models of power system dynamics, in principal, suggest a simulation based approach to the study of cascade failures, avoiding ad-hoc modeling of the failure events themselves. That is, by direct and repeated integration of the power system dynamics one can hope to simulate cascade failures many times and gather statistics that can be analyzed to reveal, for example, weakness in a grid that make it vulnerable to potential blackouts. Leveraging structural similarities between these power system models and molecular dynamics models, I will describe efficient numerical integration schemes for that purpose. In some cases even using efficient numerical integration schemes, direct simulation of the power system on time scales long enough to observe a cascade failure is impractical. I will also discuss the use of rare event simulation techniques which modify the dynamics of the network so that cascade failures become more frequent, but still allow estimation of statistics for the unbiased model.

Stephen Wright
University of Wisconsin

"Optimization Methods for Computational Statistics and Data Analysis"

The talk will focus on optimization methods that are proving to be useful in statistical and data-analytic applications. Such problems are distinguished by fairly elementary loss and regularization functions and a large amount of data. Algorithms need to take account of the statistical context, the expense of computing function and derivative information (or approximations to it), nonsmoothness, and (increasingly) nonconvexity. The talk will sketch canonical problem formulations, fundamental algorithmic techniques, and issues of current research focus.

Wotao Yin
University of California, Los Angeles

"Coordinate Update Algorithms for Optimization Problems in Machine Learning and Signal Processing"

This talk focuses on a class of algorithms, called coordinate update algorithms, which are useful at solving large-sized problems involving linear and nonlinear mappings, and smooth and nonsmooth functions. They decompose a problem to simple subproblems, where each subproblem updates one, or a small block of, variables each time. They have found applications throughout signal/imaging processing, differential equations, and machine learning. We abstract many problems to the fixed-point problem x^{k+1}=Tx^k. This talk discusses the favorable structures of the operator T that enable highly efficient coordinate update iterations. It can be carried out in sequential, parallel, or async-parallel fashions. We introduce new scalable coordinate-update algorithms to many problems involving coupling constraints Ax=b, composite nonsmooth functions f(Ax), and large-scale data. We will present a software package and its numerical examples. This is joint work with Zhimin Peng and Tianyu Wu (UCLA), Yangyang Xu (IMA), and Ming Yan (MSU).

Hui Zou
University of Minnesota

"A Selective Overview of High-dimensional Statistics and Optimization"

In this talk I will present several examples where the interplay of statistics and optimization is crucial for the success of the statistical methods and theory.