FOCM 2014 - Workshop B3
Continuous Optimization

B3 - December 15, 14:30 – 15:00

Randomized Methods for Zeroth-Order Optimization

Alexander Rakhlin

University of Pennsylvania, USA

rakhlin@gmail.com

We present several methods for zeroth order stochastic convex optimization and analyze their complexity. The proposed algorithms are based on random walks on convex bodies. We find that such methods can deal with the noisy information in a more stable manner.

Joint work with A. Belloni (Duke University, USA), T. Liang (University of Pennsylvania, USA) and H. Narayanan (University of Washington, USA).

B3 - December 15, 15:00 – 15:30

A Trust Region Algorithm with a Worst-Case Global Function Evaluation Complexity of $\mathcal{O}(\epsilon^{-3/2})$ for Nonconvex Smooth Optimization

Frank E. Curtis

Lehigh University, United States

frank.e.curtis@gmail.com

We present a trust region algorithm for solving nonconvex optimization problems that, in the worst-case, is able to drive the norm of the gradient of the objective below a prescribed threshold $\epsilon > 0$ after at most $\mathcal{O}(\epsilon^{-3/2})$ function evaluations, gradient evaluations, or iterations. Our work has been inspired by the recently proposed Adaptive Regularisation framework using Cubics (i.e., the ARC algorithm), which attains the same worst-case complexity bound. Our algorithm is modeled after a traditional trust region algorithm, but employs modified step acceptance criteria and a novel trust region updating mechanism that allows it to achieve this desirable property. Importantly, our method also maintains standard global and fast local convergence guarantees.

Joint work with Daniel P. Robinson (Johns Hopkins University) and Mohammadreza Samadi (Lehigh University).

B3 - December 15, 15:35 – 16:25

Classical unconstrained optimization based on "occasionally accurate" random models

Katya Scheinberg

Lehigh University, USA

katyascheinberg@gmail.com

We will present a very general framework for unconstrained optimization which includes methods using random models for deterministic and stochastic optimization. We make assumptions on the stochasticity that are different from the typical assumptions of stochastic and simulation-based optimization. In particular we assume that our models, search directions and function values satisfy some good quality conditions with some probability, but can be arbitrarily bad otherwise. Recently several convergence and expected convergence rates results have been developed under this framework when applied to standard optimization methods, such as line search, trust region method, direct search methods and adaptive regularization with cubics. We will present these results and outline the general analysis techniques based on theory of stochastic processes.

Joint work with A. Bandeira, C. Cartis, R. Chen, M. Menickelly and L.N. Vicente.

B3 - December 15, 17:00 – 17:30

Lipschitzean Piecewise Smooth Minimization (LiPSMin)

Andreas Griewank
Humboldt University, Germany
griewank@math.hu-berlin.de

We present a new method for the minimization of continuous piecewise smooth objectives. It is based on successive piecewise linearization in abs-normal form combined with quadratic overestimation. The inner problems are solved in by a bundle method in finitely many steps, and the outer iterates converge with a linear rate that depends on the curvature of the selection functions. We present numerical results on the usual test problems. Related successive piecewise linearization strategies are being developed for equation solving and the numerical integration of dynamical systems with Lipschitzian right hand sides. This is a fundamental shift of paradigm from the customary local approximation of problem functions by linear or quadratic Taylor expansions.

Joint work with Andrea Walther (University of Paderborn, Germany), Sabrina Fiege (University of Paderborn, Germany) and Torsten Bosse (Argonne National Laboratory, USA).

B3 - December 15, 17:30 – 18:00

Optimization Formulations and Algorithms for Constrained Phase Retrieval

Stefan Wild
Argonne National Laboratory, USA
wild@anl.gov

Far-field diffraction patterns measured in coherent X-ray diffractive imaging capture only the intensities of the diffracted waves; the phases must be recovered numerically. We discuss formulations of the phase retrieval problem as large-scale, structured, nonconvex optimization problems that incorporate physical knowledge as constraints, either in the optimization variable (direct) space or in the diffraction data (reciprocal) space. We provide an overview of classical algorithms for the phase retrieval problem as well as ADM- and multilevel-style approaches for solving our optimization-based formulations. Numerical results will highlight the challenges of constrained phase retrieval problems and the benefits of making algorithmic improvements.

Joint work with Sven Leyffer (Argonne), Todd Munson (Argonne), Ash Tripathi (Argonne).

Relative Entropy Relaxations for Signomial Optimization

Venkat Chandrasekaran

California Institute of Technology, United States of America

venkatc@caltech.edu

Signomial programs (SPs) are optimization problems consisting of an objective and constraints specified by signomials, which are sums of exponentials of linear functionals of a decision variable. SPs are non-convex optimization problems in general, and instances of NP-hard problems can be reduced to SPs. We describe a hierarchy of convex relaxations to obtain successively tighter lower bounds of the optimal value in SPs. This sequence of lower bounds is computed by solving increasingly larger-sized relative entropy optimization problems, which are convex programs specified in terms of linear and relative entropy functions. Our approach relies crucially on the observation that the relative entropy function – by virtue of its joint convexity with respect to both arguments – provides a convex parametrization of certain sets of globally nonnegative signomials with efficiently computable nonnegativity certificates via the arithmetic-geometric-mean (AM/GM) inequality. By appealing to representation theorems from real algebraic geometry, we demonstrate that our sequence of lower bounds converges to the global optimum for broad classes of SPs. Finally, we discuss numerical experiments that demonstrate the effectiveness of our approach.

Joint work with Parikshit Shah (University of Wisconsin at Madison, United States of America).

Toward a Broader View of Theory of Computing – Part 3

Narendra Karmarkar

Indian Institute of Technology, Bombay, India

narendrakarmarkar@yahoo.com

Beginning with the projectively invariant method for linear programming, interior point methods have led to powerful algorithms for many difficult computing problems, in combinatorial optimization, logic, number theory and non-convex optimization. Algorithms for convex optimization benefitted from many pre-established ideas from classical mathematics, but non-convex problems require new concepts. This three part series outlines some of these concepts

In this session, we extend the Riemannian geometry underlying projective method for linear programming to more general metric g(x), to solve non-convex optimization problems. A number of computational methods in optimization and solving equations use locally approximate linear or quadratic models based on Taylor expansion of the functions at a point . We introduce a method of creating locally exact models by defining appropriate space curvature adapted to the function(s) under consideration. We introduce concepts of degree and algebraicity relative to a metric or an affine connection. If all covariant derivatives of the function except for a finite number vanish, zero set of the function is considered to be algebraic w.r.t. that affine connection. The metric is adapted to the input instance in such a way as to make the objective function "relatively algebraic" with low degree

A property of more fundamental significance than convexity is path-wise connectivity of the level sets of the objective function. We introduce graded version of this property. A connected subset is (k, l) –

connected if any pair of points in the set can be joined by segment of a curve of degree not exceeding k, dimension of the affine space spanned by points on the segment not exceeding l, and lying entirely in the subset. Convex sets have the simplest type of connectivity – they are (1,1) – connected .A simple example of non-convex set is the set of m by n matrices of rank not exceeding k. This set is (2,3) – connected. Sets of higher connectivity indices are important in going beyond convex optimization. In case a curved space is introduced as above, the degree is relative to the affine connection and dimension is that of the geodesic submanifold spanned by points on the segment. Geodesically convex sets have (1,1)-connectivity relative to the metric.

B3 - December 16, 14:35 – 15:25

ANALYSIS AND DESIGN OF OPTIMIZATION ALGORITHMS VIA INTEGRAL QUADRATIC CONSTRAINTS

Benjamin Recht
University of California, Berkeley, USA
brecht@berkeley.edu

I will present a new method to analyze and design iterative optimization algorithms built on the framework of Integral Quadratic Constraints (IQC) from robust control theory. IQCs provide sufficient conditions for the stability of complicated interconnected systems, and these conditions can be checked by semidefinite programming. I will discuss how to adapt IQC theory to study optimization algorithms, proving new inequalities about convex functions. Using these inequalities, I will derive upper bounds on convergence rates for the gradient method, the heavy-ball method, Nesterov's accelerated method, and related variants by solving small, simple semidefinite programming problems. I will close with a discussion of how these techniques can be used to search for optimization problems with desired performance characteristics, establishing a new methodology for algorithm design.

Joint work with Laurent Lessard (University of California, Berkeley) and Andrew Packard (University of California, Berkeley).

B3 - December 16, 15:30 – 16:00

GEODESIC DISTANCE MAXIMIZATION VIA CONVEX OPTIMIZATION

Maryam Fazel
University of Washington, USA
mfazel@uw.edu

Given a graph with fixed edge weights, finding the shortest path, also known as the geodesic, between two nodes is a well-studied network flow problem. We introduce the Geodesic Distance Maximization Problem (GDMP), i.e., the problem of finding the edge weights that maximize the length of the geodesic, subject to convex constraints on the weights.

We show that GDMP is a convex optimization problem for a wide class of flow costs, and provide a physical interpretation using the dual. We present applications in various fields, including network interdiction, optical lens design, and control of forest fires. GDMP can be generalized from graphs to continuous fields, where the Eikonal equation (a fundamental partial differential equation governing flow propagation) naturally arises in the dual problem. For the case of propagation on a regular grid, the problem can be cast as a second-order cone program; however standard solvers fail to scale to the large

grid sizes of interest. We develop an Alternating Direction Method of Multipliers (ADMM) algorithm by exploiting specific problem structure to solve large-scale GDMP, and demonstrate its effectiveness in numerical examples.

Joint work with De Meng (University of Washington, USA), Pablo A. Parrilo (MIT, USA) and Stephen P. Boyd (Stanford University, USA).

B3 - December 16, 16:00 – 16:30

Integral Geometry and Phase Transitions in Conic Optimization

Martin Lotz
The University of Manchester, United Kingdom
martin.lotz@manchester.ac.uk

Integral geometry and geometric probability, going back to the work of Blaschke and Santaló, deal with measures on spaces of geometric objects, and can answer questions about the probability that random geometric objects intersect. This talk discusses various applications of (spherical) integral geometry in optimization: from the complexity theory of conic optimization to the analysis of convex approaches to solving inverse problems. In particular, it is shown how integral geometry naturally gives rise to a complete explanation of phase transition phenomena for the applicability of convex regularization to data recovery problems. We also propose extensions of the theory that allow to precisely study the probability distribution of singular values and condition numbers of conically restricted operators, and discuss relations to approaches based on geometric functional analysis.

Joint work with Dennis Amelunxen (City University Hong Kong), Michael B. McCoy (Cofacet), Joel A. Tropp (Caltech).

B3 - December 16, 17:00 – 17:30

Communication-Efficient Distributed Dual Coordinate Ascent

Martin Takac
Lehigh University, USA
takac.mt@gmail.com

Communication remains the most significant bottleneck in the performance of distributed optimization algorithms for large-scale machine learning. In this paper, we propose a communication-efficient framework, CoCoA, that uses local computation in a primal-dual setting to dramatically reduce the amount of necessary communication. We provide a strong convergence rate analysis for this class of algorithms, as well as experiments on real-world distributed datasets with implementations in Spark. In our experiments, we find that as compared to state-of-the-art mini-batch versions of SGD and SDCA algorithms, CoCoA converges to the same .001-accurate solution quality on average 25x as quickly.

Joint work with Martin Jaggi (ETH Zurich), Virginia Smith (UC Berkeley, USA), Jonathan Terhorst (UC Berkeley, USA), Sanjay Krishnan (UC Berkeley, USA), Thomas Hofmann (ETH Zurich) and Michael I. Jordan (UC Berkeley, USA).

B3 - December 16, 17:30 – 18:00

Stochastic Dual Coordinate Ascent with Arbitrary Sampling of Coordinates

Peter Richtarik

University of Edinburgh, United Kingdom

peter.richtarik@ed.ac.uk

We design a novel stochastic dual coordinate ascent (SDCA) method for minimizing an L2-regularized empirical loss function. The method operates by updating a random **set** of coordinates of the dual problem at every iteration. We allow for an **arbitrary probability law** (sampling) to govern the choice of the set of coordinates to be updated in an iteration and show how this enters the complexity bound. By varying the sampling we obtain SDCA with importance sampling, mini-batch SDCA and distributed SDCA as special cases. in a statistically interesting regime for the choice of the regularization parameter, we obtain a linear speedup up to the square root of the number of coordinates (examples). However, our method also enjoys further **data-dependent speedup.** Lastly, unlike traditional analysis of SDCA, in our analysis we control the decrease of the duality gap directly.

Joint work with Zheng Qu (Edinburgh), Tong Zhang (Baidu).

B3 - December 16, 18:00 – 18:30

A short proof of infeasibility and generating all infeasible semidefinite programs

Gabor Pataki

University of North Carolina at Chapel Hill, USA

gabor@unc.edu

The fundamental Farkas lemma of linear programming allows us to easily verify the infeasibility of an LP. For semidefinite programs (SDPs) the straightforward generalization of Farkas' lemma is not *exact*, as it does not always prove infeasibility.

We present a short and exact certificate of infeasibility of SDPs using a standard form reformulation. From the reformulation the infeasibility is easy to read off, and it allows us to systematically generate all infeasible SDPs. We prove that – loosely speaking – there are only finitely many representative infeasible SDPs in every dimension. The URL of the paper is http://arxiv.org/abs/1406.7274

I will also recall related earlier work: we call a feasible semidefinite system *well behaved* if it has strong duality with its dual for all objective functions. I will present an exact characterization of well behaved systems, which allows to systematically generate all of them.

In particular, we can systematically generate all linear maps under which the image of the semidefinite cone is closed: this is a problem of independent interest in convex geometry. The URL of the latter paper is: http://arxiv.org/abs/1112.1436

Joint work with Minghui Liu (University of North Carolina at Chapel Hill).

B3 - December 16, 18:30 – 19:00

On the graphical derivative of solution maps to parameterized equilibria with conic constraints

Héctor Ramírez

Universidad de Chile, Chile

hramirez@dim.uchile.cl

In this talk we present new calculations of the graphical derivative for the solution map to parameterized generalized equations/KKT systems associated with conic constraints. We first compute new second-order generalized differential constructions based on the graphical derivative of the normal cone mapping appearing in the KKT system. These computations are derived provided the feasible set appearing in the KKT system is convex. They provide verifiable conditions for isolated calmness of the corresponding solution map. Then, the application of a "dilatation" technique permitted to extend this computation to the nonconvex case. The latter requires, however, an additional condition of geometric nature imposed on the considered cone. This is related to the σ-term associated with projection onto this cone and has a local character. Under this condition our formula for the graphical derivative has the same form as the formula resulting in VI over polyehdral sets, and so, it can be viewed as its generalization to a class of nonpolyhedral cones. The main results obtained in this general conic programming setting are specified for and illustrated by the second-order cone programming.

Joint work with Boris Mordukhovich (Wayne University, USA) and Jiri Outrata (Czech Academic of Sciences, Czech Republic).

B3 - December 17, 14:30 – 15:00

PROJECTION: A UNIFIED APPROACH TO SEMI-INFINITE LINEAR PROGRAMS WITH APPLICATIONS TO CONVEX OPTIMIZATION

Amitabh Basu
Johns Hopkins University, United States of America
basu.amitabh@jhu.edu

We extend Fourier-Motzkin elimination to semi-infinite linear programs. Applying projection leads to new characterizations of important properties for primal-dual pairs of semi-infinite programs such as zero duality gap. Our approach yields a new classification of variables that is used to determine the existence of duality gaps. Our approach has interesting applications in finite-dimensional convex optimization, such as completely new proofs of Slater's condition for strong duality.

Joint work with Kipp Martin (University of Chicago, USA) and Chris Ryan (University of Chicago, USA).

B3 - December 17, 15:00 – 15:30

A HYBRID PROXIMAL EXTRAGRADIENT SELF-CONCORDANT PRIMAL BARRIER METHOD FOR MONOTONE VARIATIONAL INEQUALITIES

Renato Monteiro
Georgia Tech, USA
monteiro@isye.gatech.edu

This talk presents a hybrid proximal extragradient (HPE) self-concordant primal barrier method for solving a monotone variational inequality over a closed convex set endowed with a self-concordant barrier and whose underlying map has Lipschitz continuous derivative. In contrast to the iteration of a previous authors' method which has to compute an approximate solution of a linearized variational inequality, the one of the present method solves a simpler Newton system of linear equations. The method performs two types of iterations, namely: those which follow ever changing interior paths and those which correspond to large-step HPE iterations. Due to its first-order nature, the present method is shown to have a better iteration-complexity than its 0-th order counterparts such as Korpelevich's algorithm and Tseng's

modified forward-backward splitting method, although its work per iteration is larger than the one for the latter methods.

Joint work with Mauricio R Sicre (Universidade Federal da Bahia, Salvador, Brazil) and Benar Svaiter (IMPA, Rio de Janeiro, Brazil).

B3 - December 17, 15:30 – 16:00

A New Primal-Dual Predictor-Corrector Interior-Point Method

Daniel Robinson
Johns Hopkins University, USA
daniel.p.robinson@gmail.com

I present a new primal-dual predictor-corrector interior-point method for solving nonconvex optimization problems. The method is based on a new primal-dual shifted barrier function that uses Lagrange multiplier estimates to expand the feasible region for each subproblem. This expansion of the feasible region allows for improved numerical performance. Akin to predictor-corrector methods for linear programming, the step computation involves the combination of an aggressive superlinearly convergent step with a more conservative step aimed at minimizing the merit function.

Joint work with Philip E. Gill (University of California, San Diego) and Vyacheslav Kungurtsev (Czech Technical University in Prague).

B3 - December 17, 16:00 – 16:30

Gordon's Inequality and condition numbers in conic optimization

Dennis Amelunxen
City University of Hong Kong, SAR of China
damelunx@cityu.edu.hk

The probabilistic analysis of condition numbers has traditionally been approached from different angles; one is based on Smale's program in complexity theory and features integral geometry, while the other is motivated by geometric functional analysis and makes use of the theory of Gaussian processes, notably through Slepian's and Gordon's Inequalities. In this talk we aim at providing a unifying viewpoint on these approaches, and we will showcase how the different methods can be combined. Among other things, we will introduce the concept of conically restricted linear operators, whose associated "spectrum" provides a fresh light on conic condition numbers and intriguing new conjectures about their probabilistic behavior.

Joint work with Martin Lotz (The University of Manchester, UK).

B3 - December 17, 17:00 – 17:30

Stable polynomials, matroids, and sums of squares

Cynthia Vinzant
North Carolina State University, United States
vinzant@umich.edu

In 2004, Choe, Oxley, Sokal and Wagner established a tight connection between matroids and multiaffine real stable polynomials. Recently, Brändén used this theory and a polynomial coming from the Vámos matroid to disprove the generalized Lax conjecture. I will discuss the fascinating connections between these fields and how sums of squares can be used to test both for real stability and for determinantal representability of polynomials.

B3 - December 17, 17:30 – 18:00

A certificate for non-negativity of polynomials over unbounded sets

Juan C Vera
Tilburg University, Netherlands
j.c.veralizcano@tilburguniversity.nl

Certificates of non-negativity are fundamental tools in optimization. Recently, sum-of-square certificates for non-negativity of polynomials over semialgebraic sets, have been used to obtain powerful numerical techniques to address the solution of polynomial optimization problems. Usually these certificates (e.g. Schmudgen and Putinar Positivstellensatz) require the semialgebraic set to be compact.

We present a new certificate of non-negativity for polynomials that no require the semialgebraic set to be bounded. We use this certificate to generalize classical results regarding the non-negativity of quadratic polynomials over sets defined by a quadratic. We also use it to obtain a convergent hierarchy of linear matrix inequalities for polynomial optimization problems with unbounded feasible sets.

Joint work with Javier Pena (Carnegie Mellon, USA) and Luis Zuluaga (Lehigh, USA).

B3 - December 17, 18:00 – 18:30

Nonnegative polynomials and sums of squares on real projective varieties.

Mauricio Velasco
Universidad de los Andes, Colombia
mvelasco@uniandes.edu.co

A fundamental problem in polynomial optimization is the determination of the set of polynomials of degree d in n variables which are nonnegative on a given set $X \subseteq \mathbb{R}^n$. When X is basic semialgebraic, general certificates of positivity are available by the Positivstellensatz. In this talk we focus in the case when X is a real projective variety. The main contribution is that, for some varieties, we are able to give sharp bounds on the degree of the certificate. These bounds depend only on the classical algebro-geometric invariants of X.

Specifically, in the talk I will discuss the following two cases,

(1) The classification of all varieties X and integers d for which every polynomial of degree d which is nonnegative on X can be written as a sum of squares.

(2) Positivity certificates for curves. Let $X \subseteq \mathbb{P}^n$ be a totally real curve of arithmetic genus g whose real connected components are one-dimensional. If p is a nonnegative polynomial of degree $2s$ on X and

$$k \geq \max\left(\deg(X) - n + 1, \left\lfloor \frac{2g-1}{\deg(X)} \right\rfloor + 1 \right)$$

then there exists a multiplier q of degree $2k$ such that pq is a sum of squares.

B3 - Poster

SEMI STOCHASTIC GRADIENT DESCENT

Jakub Konečný
University of Edinburgh
kubo.konecny@gmail.com

In this work we study the problem of minimizing the average of a large number (n) of smooth convex loss functions. We propose a new method, S2GD (Semi-Stochastic Gradient Descent), which runs for one or several epochs in each of which a single full gradient and a random number of stochastic gradients is computed, following a geometric law. The total work needed for the method to output an ε-accurate solution in expectation, measured in the number of passes over data, or equivalently, in units equivalent to the computation of a single gradient of the loss, is $O(\log(1/\varepsilon))$. This is achieved by running the method for $O(\log(1/\varepsilon))$ epochs, with a single gradient evaluation and $O(\kappa)$ stochastic gradient evaluations in each, where κ is condition number of the problem. The SVRG method of Johnson and Zhang (SVRG) arises as a special case. To illustrate our theoretical results, S2GD only needs the workload equivalent to about 2.1 full gradient evaluations to find an 10^{-6}-accurate solution for a problem with $n = 10^9$ and $\kappa = 10^3$. Furthermore, we present a minibatching scheme, which admits simple possibility of parallelism and even improves the complexity bound under certain conditions.

B3 - Poster

NEAREST NEIGHBORS METHODS FOR SUPPORT VECTOR MACHINES

Sergio Camelo Gómez
Universidad de los Andes, Colombia
sa.camelo38@uniandes.edu.co

A key issue in the practical applicability of the Support Vector Machine methodology is the identification of the support vectors in very large data sets. In this article we propose methods based on sampling and nearest neighbours that allow for an efficient implementation of an approximate solution to the classification problem. The main idea will be that under some conditions, the support vectors of the SVM problem are, with high probability, k-neighbours of the support vectors of a random subsample from the original data. This means that if a random subsample is expanded repeatedly with the k-neighbours of its support vectors, updating the support vectors at each iteration, an approximate solution of the complete data set can be obtained at low cost. We prove some theoretical results that motivate the methodology and evaluate the performance of the proposed method with different examples.

B3 - Poster

Mateo Díaz
Universidad de los Andes, Colombia
m.diaz565@uniandes.edu.co

Compressive sensing is an emerging area with many interesting applications. For all applications it is important to estimate the number of measurements required for perfect recovery. A recent paper of Chandrasekaran et al. uses convex geometry arguments to bound these measurements. In this work we present experimental results that show this bound is far from sharp. We also present empirical evidence to show that as the dimension of the sparse vector grows, making a fixed fraction of measurements, the probability of perfect recovery approaches a 0-1 law. This interesting behavior opens questions about the geometry of the faces of the cross-polytope.

Joint work with Mauricio Junca (Universidad de los Andes, Colombia).

B3 - Poster

SEMIDEFINITE APPROXIMATIONS OF CONICAL HULLS OF MEASURED SETS.

Julian Romero
University of Waterloo, Canada
jaromero@uwaterloo.ca

Let C be a proper convex cone generated by a compact set which supports a measure μ. A construction due to A.Barvinok, E.Veomett and J.B. Lasserre produces, using μ, a sequence $(P_k)_{k \in \mathbb{N}}$ of nested spectrahedral cones which contains the cone C^* dual to C. The main results of this talk are quantitative measurements on the quality of these approximations for cones over symmetric traveling salesman polytopes and for cones of nonnegative polynomials. Measurements already known for the quality of these approximations in the traveling salesman case are improved and new measurements for the bipartite case and the cone of nonnegative polynomials are shown.

Joint work with Mauricio Velasco (Universidad de los Andes).