
Continuous Optimization – Algorithms, Applications, and Analysis
(Org: **Ahmet Alacaoglu, Michael Friedlander** and/et **Jiajin Li** (University of British Columbia))

AHMET ALACAOGLU, UBC

Towards Weaker Variance Assumptions for Stochastic Optimization: A Blast From the Past

In this talk, we focus on a classical assumption for analyzing stochastic gradient algorithms where the squared norm of the stochastic subgradient (or the variance for smooth problems) is allowed to grow as fast as the squared norm of the optimization variable. We contextualize this assumption in view of its inception in the 1960s, its seemingly independent appearance in the recent literature, its relationship to weakest-known variance assumptions for analyzing stochastic gradient algorithms, and its relevance even in deterministic problems for non-Lipschitz nonsmooth convex optimization. We build on and extend a connection recently made between this assumption and the Halpern iteration in view of nonasymptotic convergence rates for stochastic optimization. For convex nonsmooth, and potentially stochastic, optimization we provide horizon-free algorithms with last-iterate rates. For problems beyond simple constrained optimization, such as convex problems with functional constraints, we obtain rates for optimality measures that do not require boundedness of the feasible set. (Joint work with Yura Malitsky and Stephen J. Wright)

HEINZ BAUSCHKE, UBC Okanagan

On the Bredies-Chenchene-Lorenz-Naldi algorithm

Monotone inclusion problems are central in optimization and variational analysis, often solved using splitting methods featuring resolvents or proximal mappings. In 2022, Bredies, Chenchene, Lorenz, and Naldi introduced an elegant framework that unifies well-known algorithms, including Douglas-Rachford and Chambolle-Pock, with strong convergence results under certain conditions.

In this talk, I will report on joint work with Walaa Moursi, Shambhavi Singh, and Xianfu Wang. We extend the analysis of Bredies et al., providing new strong convergence results for linear relations. For the Chambolle-Pock algorithm, we prove convergence to the projection onto an intersection of linear subspaces. We also discuss algorithms by Ryu and by Malitsky and Tam.

YING CUI, University of California Berkeley

Variational Theory and Algorithms for a Class of Asymptotically Approachable Nonconvex Problems

We investigate a class of composite nonconvex functions, where the outer function is the sum of univariate extended-real-valued convex functions and the inner function is the limit of difference-of-convex functions. A notable feature of this class is that the inner function can be merely lower semicontinuous instead of continuously differentiable. It covers a range of important yet challenging applications, including the composite value functions of nonlinear programs and the value-at-risk constraints. We propose an asymptotic decomposition of the composite function that guarantees epi-convergence to the original function, leading to necessary optimality conditions for the corresponding minimization problem. The proposed decomposition also enables us to design a numerical algorithm such that any accumulation point of the generated sequence, if exists, satisfies the newly introduced optimality conditions. These results expand on the study of so-called amenable functions introduced by Poliquin and Rockafellar in 1992, which are compositions of convex functions with smooth maps, and the prox-linear methods for their minimization.

JELENA DIAKONIKOLAS, University of Wisconsin-Madison

Faster solutions to variational inequalities with highly nonuniform component or block Lipschitz constants

Block coordinate methods have a rich history in optimization, particularly for minimization problems, where they offer computational advantages over full vector (single block) methods whenever the problem at hand is compatible with blockwise updates.

In contrast, the potential of block coordinate updates remains underexplored in the realm of variational inequalities—a class of equilibrium problems. To date, a rigorous demonstration of computational advantages for block coordinate updates in this context has largely been lacking.

I will present a novel block coordinate method addressing a standard class of variational inequalities with monotone Lipschitz operators. This method achieves a provably lower computational cost than traditional full vector update methods—by a factor scaling with the number of blocks—in settings where blockwise Lipschitz constants are highly nonuniform, which is the very setting where block coordinate methods are known to excel. I will also discuss how this method can be adapted for problems involving finite sum operators, where it functions as a variance reduction method. In this context, and for cases with highly nonuniform Lipschitz constants among the components, the method leads to complexity improvements over state-of-the-art approaches by a factor scaling with the square-root of the number of components in the finite sum.

I will conclude by highlighting some intriguing open questions.

The talk is based on a recent preprint, available here: <https://arxiv.org/abs/2411.00979>.

DIMA DRUSVYATSKIY, University of Washington

MICHAEL FRIEDLANDER, University of British Columbia

Density Estimation from Moments

We present a maximum entropy method for estimating probability densities from a limited set of moment measurements, with applications to x-ray Thomson scattering in high-energy physics. A stable dual formulation using indirect linear algebra operations yields robust density estimates. (Joint work with Nicholas Barnfield, Thomas Chuna, Tobias Dornheim, Tim Hoheisel, and Matthew Rose-Kamp.)

NAOMI GRAHAM, UBC

TIM HOHEISEL, McGill University

Stability in nonsmooth optimization via graphical differentiation

We present some applications of implicit function theorems from variational analysis to stability analysis of prominent optimization problems with a focus on regularized least-squares. These contain results on well-posedness, Lipschitz stability and smoothness of the optimal solution function

CHO HO (PETER) LAM, Huawei Technologies Canada

Faster Infeasibility Analysis for Linear Programs

Presolving a linear program is important for fast solution and can sometimes detect infeasibility before the reduced model is solved. But presolving typically interferes with finding an Irreducible Infeasible Subset (IIS) of row constraints and variable bounds, the main way to analyze infeasibility. Early attempts to backtrack the set of logical model reductions when the presolver detects infeasibility, with the goal of finding an IIS, were abandoned as impractical. However, the OptVerse solver has now implemented a very fast backtrack capability that greatly speeds IIS isolation whether or not the presolver detects infeasibility. In both cases, the backtracker isolates a small subset of the model that is then subjected to typical IIS isolation procedures. The speed advantage is demonstrated experimentally vs. other major LP solvers.

JIAJIN LI, University of British Columbia

Unveiling Spurious Stationarity and Hardness Results for Bregman Proximal-Type Algorithms

Bregman proximal-type algorithms, such as mirror descent, are popular in optimization and data science for effectively exploiting problem structures and optimizing them under tailored geometries. However, most of existing convergence results rely on the gradient Lipschitz continuity of the kernel, which unfortunately excludes most commonly used cases, such as the Shannon entropy. In this paper, we reveal a fundamental limitation of these methods: Spurious stationary points inevitably arise when the kernel is not gradient Lipschitz. The existence of these spurious stationary points leads to an algorithm-dependent hardness result: Bregman proximal-type algorithms cannot escape from a spurious stationary point within any finite number of iterations when initialized from that point, even in convex settings. This limitation is discovered through the lack of a well-defined stationarity measure based on Bregman divergence for non-gradient Lipschitz kernels. Although some extensions attempt to address this issue, we demonstrate that they still fail to reliably distinguish between stationary and non-stationary points for such kernels. Our findings underscore the need for new theoretical tools and algorithms in Bregman geometry, paving the way for further research.

TIANYI LIN, Columbia University

Lower bound construction in nonsmooth optimization

In this talk, we discuss the lower bound construction in non-smooth optimization under a Lipschitz condition. First, we review the classical results when the function to be minimized is convex where the hard instance is piecewise linear. Second, we explain why the Goldstein stationarity is more favorable than the Clarke stationarity from a computational viewpoint. Finally, we construct a new hard instance and prove that (1) a lower bound of $\Omega(d)$ and (2) no finite-time guarantee under linear span condition, for any deterministic algorithm that has access to 1-order and 0-order oracles to find an approximate Goldstein stationary point up to sufficiently small parameter and tolerance.

ZHAOSONG LU, University of Minnesota

Variance-reduced first-order methods for stochastic optimization with deterministic constraints

We consider stochastic optimization problems with deterministic constraints. Existing methods typically focus on finding an approximate stochastic solution that ensures the expected constraint violations and optimality conditions meet a prescribed accuracy. However, such an approximate solution can possibly lead to significant constraint violations in practice. To address this issue, we propose variance-reduced first-order methods that treat the objective and constraints differently. Under suitable assumptions, our proposed methods achieve stronger approximate stochastic solutions with complexity guarantees, offering more reliable constraint satisfaction than existing approaches.

DOMINIQUE ORBAN, GERAD/Polytechnique Montreal

Complexity of trust-region methods in the presence of unbounded Hessian approximations

We extend traditional complexity analyses of trust-region methods for unconstrained, possibly nonconvex, optimization. Whereas most complexity analyses assume uniform boundedness of model Hessians, we work with potentially unbounded model Hessians. Boundedness is not guaranteed in practical implementations, in particular ones based on quasi-Newton updates. Our analysis is conducted for a family of trust-region methods that includes most known methods as special cases. We examine two regimes of Hessian growth: one bounded by a power of the number of successful iterations, and one bounded by a power of the number of iterations. This allows us to formalize and confirm the profound intuition of Powell (2010), who studied convergence under a special case of our assumptions, but whose proof contained complexity arguments. Specifically, for $0 \leq p < 1$, we establish sharp $O(\epsilon^{-2/(1-p)})$ evaluation complexity to find an ϵ -stationary point when model Hessians are $O(k^p)$, where k is the iteration counter. For $p = 1$, which is the case studied by Powell, we establish a sharp $O(\exp(c\epsilon^{-2}))$ evaluation complexity for a certain constant $c > 0$. This is as Powell suspected and is far worse than other bounds surmised elsewhere in the literature. We establish similar bounds when model Hessians are $O(|S_k|^p)$, where $|S_k|$ is the number of iterations where the step was accepted, up to iteration k . To the best of our knowledge, ours is the first work to provide complexity bounds when model Hessians grow linearly with $|S_k|$ or at most linearly with k , which covers multiple quasi-Newton approximations.

NICHOLAS RICHARDSON, University of British Columbia

Density Separation with Tensor Factorization

The nonnegative Tucker-1 tensor factorization is used to separate mixtures of probably densities. A kernel density estimation transforms raw samples into compressed and discretized densities. An implementation of a specialized block coordinate descent algorithm that enforces the required simplex constraints is guaranteed to converge to Nash equilibria. Numerical experiments on real geological data, using our open source Julia implementation, illustrate the model's effectiveness at separating the density mixtures. Portability of the method to other applications like spacial transcriptomics and audio decomposition is discussed.

MARK SCHMIDT, UBC

Global-Local Smoothness: Line Search can really help! Really!

Iteration complexities for first-order optimization algorithms are typically stated in terms of a global Lipschitz constant of the gradient, and optimal results can typically be achieved using fixed step sizes. But many objective functions that arise in practice have regions with small Lipschitz constants where larger step sizes can be used. Many local Lipschitz assumptions have thus been proposed, which lead to results showing that adaptive step sizes and/or line searches yield improved convergence rates over fixed step sizes. However, these faster rates tend to depend on the path taken by the algorithm, which makes it difficult to compare the iteration complexities of different methods.

We consider a simple characterization of global and local smoothness that only depends on properties of the function. This allows explicit lower and upper bounds on iteration complexities in terms of problem-dependent constants, which allows us to compare iteration complexities between algorithms. Under this assumption it is straightforward to show the advantages of line searches over fixed step sizes, of line searches with unbounded step sizes over those that bound the maximum step size, and even that in some settings gradient descent with line search has a better iteration complexity than accelerated gradient methods with fixed step sizes.

HENRY WOLKOWICZ, University of Waterloo

The ω -condition number for optimal preconditioning of linear systems

Preconditioning is essential in iterative methods for solving linear systems. It is also the implicit objective in updating approximations of Jacobians in optimization methods, e.g., in quasi-Newton methods. We study a nonclassic matrix condition number, the ω -condition number, the ratio of the arithmetic and geometric means of the singular values. We do this in the context of optimal conditioning for: (i) low rank updating of generalized Jacobians; (ii) iterative methods for linear systems: (iia) clustering of eigenvalues and (iib) convergence rates. In particular, we show the advantages over the classical kappa-condition number. (work with Woosuk L. Jung and David Torregrosa-Gelén)

ALP YURTSEVER, Umeå University

Block Coordinate DC Programming

We introduce an extension of the Difference of Convex Algorithm (DCA) in the form of a block coordinate approach for problems with separable structure. For n coordinate-blocks and k iterations, our main result proves a non-asymptotic convergence rate of $O(n/k)$ for the proposed method. Furthermore, leveraging the connection between DCA and Expectation Maximization (EM), we propose a block coordinate EM algorithm.