

DISSERTATION PROPOSAL

Qihang Lin

Thursday, March 28, 2013

11:00 am

324 GSIA (West Wing)

Large-Scale Optimization for Machine Learning and Sequential Decision Making

The development of modern technology has enabled collecting data of unprecedented size, dimension and complexity. Structured regression techniques provide us a suite of tools for understanding and exploring these big data from many areas in business, science and engineering. A structured regression model is usually formulated as an optimization problem using a large number of decision variables. To solve such a large-scale optimization, traditional techniques often suffer from low scalability and unaffordable computational time. To address this challenge, in this thesis, we propose low memory-required optimization methods for different structured regression problems based on homotopy, smoothing, stochastic sampling and other techniques. Our methods demonstrate nice scalability for large instances and show better performance than existing methods both theoretically and practically.

First, we derive a gradient homotopy method for Lasso problem, the most popular structured regression model. This method fully utilizes the sparse nature of the optimal solution and keeps the whole path of solution in a low-dimensional subspace. Due to the local strongly convexity of Lasso problem under restricted isometric property, our method achieves a linear convergence rate, which prevails over the theoretical optimal convergence rate under a black-box gradient assumption. Second, we develop a smoothing proximal gradient algorithm for solving general structured regressions. This method resolves the difficulty that the proximal mapping central in most first-order methods does not have a closed form solution due to the sophisticated structure inducing term.

When the size of the data grows beyond the storage capacity, existing deterministic first-order methods cannot be applied due to the difficulty of computing the exact gradient. As an alternative, we propose stochastic first-order methods for structured regression that utilize a stochastic gradient constructed using only a random sample of the whole data that can fit into memory. Our methods are proved to converge to optimality despite the gradient noise and in fact obtain an uniformly optimal convergence rate.

The challenge of large-scale optimizations arises from not only the large volume of data but also the exponential growth rate of the number of variables in a multistage decision making, known as the curse of dimensionality. To explore the scalability of first-order methods in the latter case, we study the optimal trade execution problem under coherent dynamic risk measure, which is a large-scale multistage stochastic optimization problem in financial engineering. We formulate this problem as a saddle-point problem and solve it with a primal-dual first-order method. Due to a compact dual representation of coherent risk measures, the proximal mapping allows a quick solution, resulting an efficient implementation of this first-order method.