

Dissertation Defense

Qihang Lin

Wednesday, August 28, 2013

1:30 pm

324 GSIA (West Wing)

Large-Scale Optimization for Machine learning and Sequential Decision Making

Thanks to the development of the modern digital technology, people today are able to generate, collect and store data of unprecedented volume, dimension and complexity. This growing trend of big data brings new needs for powerful tools to explore and analyze different data sets. Among different tools for data analysis, structured regression is one of the most popular techniques. It has been successfully applied to data from many disciplines in business, science and engineering.

A structured regression model is formulated as an optimization problem using a large number of decision variables. Traditional techniques often suffer from low scalability and unaffordable computational time when applied to solve optimization problems in such a large scale. To address this challenge, in the first main part of this thesis, we propose several optimization methods with low memory requirements for different structured regression problems based on homotopy, smoothing, stochastic sampling and other techniques. In particular, our contributions include:

1. A gradient homotopy method for the Lasso problem, the most popular structured regression model. This method converges to the optimality in a linear rate. This rate improves the traditional sub-linear rate, which is theoretically not improvable if an algorithm is only allowed to use black-box gradient information. The reason for us to achieve a better complexity is that our method fully utilizes the local strong convexity property of the Lasso problem under the restricted eigenvalue assumption, which provides much more information than gradients.
2. A smoothing proximal gradient algorithm for solving general structured regressions. One difficulty of applying first-order methods to structured regressions with sophisticated penalty terms is that the proximal mapping, which must be computed in each iteration, does not have a closed-form solution. The algorithm we proposed constructs a smooth approximation to the penalty terms so that the proximal mapping can be computed in closed form, and thus, first-order methods can be applied efficiently.
3. A unified theoretical framework for analyzing the geometric nature of different first-order methods. We show that most popular versions of the accelerated gradient methods essentially construct an estimate sequence of the optimization problem in different ways. This explains why these methods achieve the same optimal convergence rate even with different updating schemes.
4. A stochastic first-order method for structured regression, which utilizes a stochastic gradient constructed using only a random sample of the whole data set. This method is

favorable when the size of the data grows beyond the storage capacity and deterministic first-order methods cannot be applied due to the overwhelming computational cost involved with computing the exact gradient. Our method achieves the “uniformly” optimal complexity, which means, when the noise of the stochastic gradient is reduced to zero, our method achieves the optimal complexity as a deterministic first-order method. By contrast, the traditional stochastic gradient methods would still have sub-optimal complexity

The challenges of large-scale optimization arise not only from the large volume of data but also from the exponential rate of growth of the number of variables in multistage decision making models. In the second main part of this thesis, we explore the scalability of first-order methods in the latter case. We focus on the optimal trade execution problem under coherent dynamic risk measure. This is a large-scale multistage stochastic optimization problem that arises in financial engineering. Relying on the dual representation of the coherent risk measure, we can formulate this problem as a saddle-point problem and solve it with a primal-dual first-order method. The truncated simplex structure of the primal and dual domains allows us to obtain a closed-form solution to the relevant proximal mapping sub-problem, resulting in an efficient implementation of this first-order method. Our models and algorithms are tested on limit order book real data and demonstrate promising numerical properties. Furthermore, we generalize the same primal-dual first-order method to the case where the multistage decision making problem is modeled with a scenario tree in a non-Markovian fashion. In the last part of the thesis, we discuss future research directions in optimization techniques for solving high-dimensional structured regression problems and multistage decision making problems from both the theoretical and computational aspects.

KEYWORDS:

Machine Learning, Sparse Learning, Lasso, Optimization, Regression, First-Order Method, Proximal Gradient Method, Stochastic Optimization, Stochastic Gradient, Convex Programming, Dynamic Coherent Risk Measure, Saddle-Point Problem, Markov Decision Process, Trade Execution, Multistage Decision Making, Scenario Tree, Limit Order Book.