First-Order Methods of Solving Nonconvex Optimization Problems: Algorithms, Convergence, and Optimality

by

Songtao Lu

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering

Program of Study Committee:
Zhengdao Wang and Mingyi Hong, Major Professor
Nicola Elia
Aleksander Dogandžić
Kris De Brabanter

Iowa State University

Ames, Iowa

2018

Abstract

First-order methods of solving large scale nonconvex problem have been applied in many machine learning area, such as matrix factorization, dictionary learning, matrix sensing/completion, deep neural networks, etc. Take matrix factorization as an example. They have lots of important applications in data analytics problems such as document clustering, community detection and image segmentation. In this disseration, we study some novel nonconvex variable splitting methods for solving some matrix factorization problems, such as symmetric non-negative matrix factorization (SymNMF) and stochastic SymNMF.

In the problem of SymNMF, the proposed algorithm, called nonconvex splitting SymNMF (NS-SymNMF), is guaranteed to converge to the set of Karush-Kuhn-Tucker (KKT) points of the nonconvex SymNMF problem. Furthermore, it achieves a global sublinear convergence rate. We also show that the algorithm can be efficiently implemented in a distributed manner. Further, sufficient conditions are provided which guarantee the global and local optimality of the obtained solutions. Extensive numerical results performed on both synthetic and real data sets suggest that the proposed algorithm converges quickly to a local minimum solution.

Furthermore, we consider a stochastic SymNMF problem in which the observation matrix is generated in a random and sequential manner. The propose stochastic nonconvex splitting method not only guarantees convergence to the set of stationary points of the problem (in the mean-square sense), but further achieves a sublinear convergence rate. Numerical results show that for clustering problems over both synthetic and real world datasets, the proposed algorithm converges quickly to the set of stationary points.

When the objective function is nonconvex, it is well-known the most of the first-order algorithms converge to the first-order stationary solution (SS1) with a global sublinear rate. Whether the first-order algorithm can converge to the second-order stationary points (SS2) with some provable rate attracts a lot of attention recently. In particular, we study the alternating gradient descent (AGD) algorithm as an example, which is a simple but popular algorithm and has been applied to problems in optimization, machine learning, data mining, and signal processing, etc. The algorithm updates two blocks of variables in an alternating manner, in which a gradient step is taken on one block, while keeping the remaining block fixed.

In this work, we show that a variant of AGD-type algorithms will not be trapped by "bad" stationary solutions such as saddle points and local maximum points. In particular, we consider a smooth unconstrained optimization problem, and propose a perturbed AGD (PA-GD) which converges (with high probability) to the set of SS2 with a global sublinear rate. To the best of our knowledge, this is the first alternating type algorithm which takes $\mathcal{O}(\text{polylog}(d)/\epsilon^{7/3})$ iterations to achieve SS2 with high probability [where polylog(d) is polynomial of the logarithm of dimension d of the problem].