Paper Review: Escaping from Saddle Points — Online Stochastic Gradient for Tensor Decomposition

1 Motivation

Convex gradient descent is a powerful and extensively studied optimization method and comes with some good properties that guarantee convergence to a global minimum relatively quickly. However, non-convex problems typically lose this convergence guarantee. The issues with non-convex optimization come from the fact that local minimums are not guaranteed to be global minimum, where cases that $\nabla f(x) = 0$ can occur at local minimums, local maximums, as well as saddle points. One particular non-convex problem that suffers from many saddle points is tensor decomposition.

Tensor decomposition is the process of finding a series of order-p, rank-1 tensors A_i that sum to form the objective tensor $T = \sum_{i=1}^{r} A_i$ with various additional constraints on the tensors A_i depending the type of decomposition [8]. The types of decomposition stems from the different definitions of rank and orthogonality, of which we will focus on complete orthogonality: Given decomposed tensors $A_i = u_i^{(1)} \bigotimes u_i^{(2)} ... \bigotimes u_i^{(p)}$ where u is an m-dimensional vector, then $A_i \perp_c A_j$ if $u_i^k \perp u_j^k, \forall k \in [1, 2, ...p], i \neq j, ||A|| = 1$. Tensor decomposition has applications in ICA, topic models, and community detection. As a non-convex optimization problem, it can have many saddle points. We need an algorithm that can exploit problem properties that help to avoid these saddle points.

2 Related Research

We will explore the algorithm for noisy stochastic gradient descent provided in *"Escaping from Saddle Points – Online Stochastic Gradient for Tensor Decomposition"* [8] to see how it can be used in tensor decomposition as well as identifying non-convex problems that can be characterized as "strict-saddle". Of course, there are multiple ways tensor decomposition has been defined and approached, so to identify the benefits and potential issues with the algorithms presented in the main paper, we will look at other popular approaches such as Alternating least Squares [7] [1], FastICA [5], and minimizing reconstruction error. For more general solutions to non-convex problems with multiple saddle points, we will look at alternative ways of escaping saddle points in applications unrelated to tensor decomposition [2] [3] [4].

3 Project Objectives

The main purpose of this project is to investigate ways of overcoming the convergence issues for non-convex problems. With the proliferation of deep learning, a typically non-convex problem, finding efficient ways to bridge the gap between convex and non-convex problems will become increasingly valuable. Focusing on tensor decomposition also forces us to become familiar with higher order data structures and the notation used by the main paper, defined in [6].

The structure of the paper will be as follows: Provide a framework of notation and problem examples that can be used as a benchmark for evaluating algorithms provided in the relevant papers. Then conduct a low order implementation of those algorithms that can be compared to the results of similar algorithms in EECS 559. This is possible because tensors of orders 1 and 2 are simply vectors and matrices, whose optimization is covered extensively in this course. Finally we will compare performance of noisy stochastic gradient descent with the competing algorithms to see how these algorithms generalize to higher dimensions.

References

[1] Ankur Moitra Allen Liu. Tensor completion made practical, 2020. NIPS'20: 34th International Conference on Neural Information Processing Systems.

[2] Rong Ge Anima Anandkumar. Efficient approaches for escaping higher order saddle points in non-convex optimization. *JMLR: Workshop and Conference Proceedings*, 49:1–22, 2016.

[3] Rong Ge Chi Jin et al. How to escape saddle points efficiently. *ICML'17: Proceedings of the 34th International Conference on Machine Learning*, 70:1724–1732, 2017.

[4] Yann Dauphin et al. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2:2933–2941, 2014.

[5] Aapo Hyvarinen. Fast ica for noisy data using gaussian moments. *In Circuits and Systems*, 5:57–61, 1999.

[6] Tamara Kolda. Orthogonal tensor decompositions. *SIAM J. MATRIX ANAL. APPL.*, 23(1):243–255, 2001.

[7] Andre L. F. de Almeida Pierre Comon, Xavier Luciani. Tensor decompositions, alternating least squares and other tales. *Journal of Chemometrics*, 23:393–405, 2009.

[8] Chi Jin Yang Yuan Rong Ge, Furong Huang. Escaping from saddle points – online stochastic gradient for tensor decomposition. *JMLR: Workshop and Conference Proceedings*, 40:1–46, 2015.