Tutorial proposal: Nonconvex optimization: Challenges and Recent Successes

Anima Anandkumar Furong Huang Majid Janzamin
Hanie Sedghi

January 18, 2016

1 Topic overview

Most machine learning tasks require solving non-convex optimization problems. For these problems, as the data dimensions grow, the number of critical points can grow exponentially. Local search methods such as gradient descent can get stuck in one of these critical points. Finding the globally optimal solution is computationally hard in the worst case for non-convex optimization. Instead, over the last few years, focus has shifted in characterizing transparent conditions for non-convex problems which are tractable. In many instances, these conditions turn out to be mild and natural for machine learning applications.

This tutorial will provide an overview of the recent theoretical success stories in non-convex optimization. This includes learning latent variable models, dictionary learning, robust principal component analysis, and so on. Simple iterative methods such as spectral methods, alternating projections, and so on, can be proven to learn consistent models with polynomial sample and computational complexity. This tutorial will present main ingredients towards establishing these results. For instance, one approach is to first obtain an approximate solution which lands in the basin of the globally optimal solution, and then run gradient descent to reach the global optimum. Another approach is to discard the original objective function, and to instead optimize over different objective functions which are better behaved. For example, the use of moment fitting objective instead of maximum likelihood for learning latent variable models allows us to establish consistency in estimation.

For problems where finding the globally optimal solution is not tractable, it is desirable to at least reach a local optimum. It turns out that even this is computationally hard in the worst case. In this tutorial, we will cover the important problem of escaping saddle points. Saddle points are critical points which are not local minima, meaning there exist directions where the objective value decreases (for minimization problems). Saddle points can slow down gradient descent arbitrarily. Alternatively, if Newton's method is run, it converges to an arbitrary critical point, and does not distinguish between a local minimum and a saddle point. Solutions such as Trust region methods and recent work on noisy stochastic gradient descent will be covered.

There are certainly many challenging open problems in the area of non-convex optimization. While guarantees have been established in individual instances, there is no common unifying theme of what makes non-convex problem tractable. Many challenging instances such as optimization for training multi-layer neural networks or analyzing novel regularization techniques such as dropout for non-convex optimization still remain wide open. On the practical side, conversations between theorists and practitioners can help identify what kind of conditions are reasonable for specific applications, and thus lead to the design of practically motivated algorithms for non-convex optimization with rigorous guarantees. The tutorial will cover these issues as well.

2 Target audience

This tutorial will fill a very important gap in bringing researchers from disparate communities and bridging the gap between theoreticians and practitioners. To facilitate discussion between theorists and practitioners, we aim to make the tutorial easily accessible to people currently unfamiliar with the intricate details of these methods. We expect around 200 participants.

3 Content details

1. Overview of challenges in nonconvex optimization and recent advances (60 min): covered by Prof. Anima Anandkumar.
   (a) Critical points and curse of dimensionality
   (b) Saddle points vs. local optima: difficulty in escaping them.
   (c) Success stories in guaranteed nonconvex optimization: spectral methods, dictionary learning, robust PCA etc.
2. Escaping saddle points. (40 min): Covered by Furong Huang.
   (a) Behavior of gradient descent and Newton's method around saddle points.
   (b) Trust region methods and Nestorov's cubic regularization.
   (c) Noisy gradient descent as a simple solution with guarantees
3. Delving into spectral methods (50 min): Covered by Majid Janzamin and Hanie Sedghi.
   (a) Review: matrix eigen-decompositions
   (b) Orthogonal tensor decompositions and power method.

(c) The method-of-moments via orthogonal tensor decompositions
(d) Unsupervised learning of latent variable models.
(e) Training of neural networks using tensor methods.

4 Format

The tutorial will be presented as lectures with slides.

5 Organizers' and presenters' expertise

1. Anima Anandkumar
   `a.anandkumar@uci.edu`
   `http://newport.eecs.uci.edu/anandkumar/`
   Assistant Professor of Electrical Engineering & Computer Science at UC Irvine

2. Furong Huang
   `furongh@uci.edu`
   `https://sites.google.com/site/furongfionahuang/`
   Graduate student at Electrical Engineering & Computer Science at UC Irvine

3. Majid Janzamin
   `mjanzami@uci.edu`
   `http://newport.eecs.uci.edu/anandkumar/MajidJanzamin/`
   Graduate student at Electrical Engineering & Computer Science at UC Irvine

4. Hanie Sedghi
   `honey.sedghi@gmail.com`
   `http://www-scf.usc.edu/h̃sedghi/`
   Research scientist at Paul Allen AI institute.

Anima Anandkumar has given numerous tutorials at machine learning conferences and summer schools, including tutorials at ICML 2013, AAAI 2014, MLSS 2014, and so on. Sample set of slides and videos are available at `http://newport.eecs.uci.edu/anandkumar/MLSS.html`. Furong Huang has given a tutorial at the machine learning conference, organized at Janelia Research Campus at Howard Hughes medical institute. in Oct 2015. In addition, she will be speaking at MLConf 2016 at NYC, which is a popular industry conference in machine learning. Majid Janzamin and Hanie Sedghi have given numerous talks at various venues and have also taught many sessions on spectral methods.

All four presenters have worked on theoretical analysis and algorithms for non-convex optimization (and other related techniques). The following are a limited list of presenters' publications in the tutorial area.

1. Anima Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. arXiv:1210.7559.

2. Majid Janzamin, Hanie Sedghi, Anima Anandkumar. Beating the Perils of Non-Convexity: Guaranteed Training of Neural Networks using Tensor Methods on arXiv, June 2015.

3. Rong Ge, Furong Huang, Chi Jin, Yang Yuan, Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition, Proc. of COLT 2015.

4. H. Sedghi, M. Janzamin, A. Anandkumar, Provable Tensor Methods for Learning Mixtures of Generalized Linear Models, Proc. of AISTATS 2016.

5. Anima Anandkumar, Prateek Jain, Yang Shi, U.N. Niranjan, Tensor vs Matrix Methods: Robust Tensor Decomposition under Block Sparse Perturbations, Proc. of AISTATS 2016.

6. Furong Huang and Anima Anandkumar. Convolutional Dictionary Learning through Tensor Factorization. NIPS 2015 International Workshop "Feature Extraction:Modern Questions and Challenges"