Stochastic first order optimization for finite sum problems

Damiano Zeffiro
University of Padua

June 3, 10:00
Room 2BC30

Abstract

The goal of first order optimization is to minimize efficiently an objective given an oracle that can return a subgradient and the value of the objective for any point in the domain. In the stochastic setting this oracle is exact only in expected value. In this seminar we will see a few recent improvements in the design and the analysis of efficient stochastic first order optimization methods. We focus on the case where the objective is the sum of a large number of components, discussing a few applications and complexity bounds. In particular we will see how the stochastic variance reduced gradient method (SVRG) outperforms the stochastic gradient descent by reducing the variance of the stochastic gradients with few additional oracle calls. We will also discuss recent results for stochastic FW variants, analyzing in some details an application of empirical process theory to the study of one such variant.

References

[1] Zeyuan Allen-Zhu. Katyusha: The first direct acceleration of stochastic gradient methods. *The Journal of Machine Learning Research*, 18(1):8194–8244, 2017.

[2] Zeyuan Allen-Zhu and Elad Hazan. Variance reduction for faster non-convex optimization. In *International conference on machine learning*, pages 699–707, 2016.

[3] Zeyuan Allen-Zhu and Yang Yuan. Improved svrg for non-strongly-convex or sum-of-non-convex objectives. In *International conference on machine learning*, pages 1080–1089, 2016.

[4] Dan Garber and Elad Hazan. Faster rates for the frank-wolfe method over strongly-convex sets. In *32nd International Conference on Machine Learning, ICML 2015*, 2015.

[5] Donald Goldfarb, Garud Iyengar, and Chaoxu Zhou. Linear convergence of stochastic frank wolfe variants. *arXiv preprint arXiv:1703.07269*, 2017.

[6] Elad Hazan and Haipeng Luo. Variance-reduced and projection-free stochastic optimization. In *International Conference on Machine Learning*, pages 1263–1271, 2016.

[7] Rie Johnson and Tong Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in neural information processing systems*, pages 315–323, 2013.

[8] Simon Lacoste-Julien and Martin Jaggi. On the global linear convergence of frank-wolfe optimization variants. In *Advances in Neural Information Processing Systems*, pages 496–504, 2015.

[9] Haihao Lu and Robert M Freund. Generalized stochastic frank-wolfe algorithm with stochastic" substitute”gradient for structured convex optimization. *arXiv preprint arXiv:1807.07680*, 2018.

[10] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[11] Yurii E Nesterov. A method for solving the convex programming problem with convergence rate o (1/kˆ 2). In *Dokl. akad. nauk Sssr*, volume 269, pages 543–547, 1983.

[12] Sashank J Reddi, Ahmed Hefny, Suvrit Sra, Barnabas Poczos, and Alex Smola. Stochastic variance reduction for nonconvex optimization. In *International conference on machine learning*, pages 314–323, 2016.

[13] Sashank J Reddi, Suvrit Sra, Barnabás Póczos, and Alex Smola. Stochastic frank-wolfe methods for nonconvex optimization. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1244–1251. IEEE, 2016.