SGD with Variance Reduction beyond Empirical Risk Minimization

Massil ACHAB, CMAP, Ecole Polytechnique

Stéphane GAIFFAS, CMAP, Ecole Polytechnique

Agathe GUILLOUX, LSTA, UPMC

Emmanuel BACRY, CMAP, Ecole Polytechnique

We introduce a doubly stochastic proximal gradient algorithm for optimizing a finite average of smooth convex functions, whose gradients depend on numerically expensive expectations.
Our main motivation is the acceleration of the optimization of the regularized Cox partial-likelihood (the core model used in survival analysis), but our algorithm can be used in different settings as well.
The proposed algorithm is doubly stochastic in the sense that gradient steps are done using stochastic gradient descent (SGD) with variance reduction, where the inner expectations are approximated by a Monte-Carlo Markov-Chain (MCMC) algorithm. We derive conditions on the MCMC number of iterations guaranteeing convergence, and obtain a linear rate of convergence under strong convexity and a sublinear rate without this assumption.
We illustrate the fact that our algorithm improves the state-of-the-art solver for regularized Cox partial-likelihood on several datasets from survival analysis.

Références

[1] Johnson, R. and Zhang, T., *Accelerating Stochastic Gradient Descent using Predictive Variance Reduction*, Advances in Neural Information Processing Systems 26, 2013.

[2] Atchade, Y. F. and Fort, G. and Moulines, E., *On stochastic proximal gradient algorithms*, ArXiv e-prints, 2014.

Massil ACHAB, Centre de Mathématiques Appliques, Ecole Polytechnique, 91128 Palaiseau, France
massil.achab@cmap.polytechnique.fr
Stéphane GAIFFAS, Centre de Mathématiques Appliques, Ecole Polytechnique, 91128 Palaiseau, France
stephane.gaiffas@cmap.polytechnique.fr
Agathe GUILLOUX, Laboratoire de Statistique Théorique et Appliquée, Paris 6, 75005, France
agathe.guilloux@upmc.fr
Emmanuel BACRY, Centre de Mathmatiques Appliquées, Ecole Polytechnique, 91128 Palaiseau, France
emmanuel.bacry@polytechnique.fr