

Internship: New Stochastic Sketching Methods for Big Data Ridge Regression.

co-supervisors: Dr. Robert M. Gower* and Dr. Francis Bach

December 6, 2016

The need to solve linear regression occurs throughout scientific computing, such as ridge regression in machine learning and data assimilation in oceanography, weather forecasting ...etc. The regression problem we will consider in this project is:


$$\min_x \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda}{2} \|x\|_2^2, \quad (1)$$

where the number of rows and columns of $A \in \mathbb{R}^{m \times n}$ are in the order of tens of millions.

With data sets now being collected automatically and electronically, there has been a surge of new stochastic incremental methods that can gracefully scale with the dimensions of the data in (1). Yet it is still unclear if the new stochastic methods are capable of outperforming the classic Conjugate Gradients (CG) methods [4], even though the CG method was developed in the 1950's! This project aims to put this doubt to rest, by designing and implementing new stochastic methods based on randomized sketching [5, 6] together with iterative projection [2] and stochastic preconditioning [3] that will significantly outperform the CG method. If successful, the implementation of the methods could become a reference package for solving large scale linear regression.

Requisites: The prospective student will have completed a course in programming, linear algebra and preferably optimization. The student should have an aptitude for coding.

Timeline: The student will start by reading [2, 3] and possibly [1, 6], followed by a design phase, and finally implementing and testing the new methods. The student will meet the co-supervisor Dr. Gower on a weekly basis to discuss the project.

Output: The main output of this project will be a software package written in either Python, C, C++ or , that will be freely distributed on github or google.code, and a paper detailing the new methods. Depending on the students preference, there will also be an opportunity to analyse the new methods and produce new theory.

*contact: robert.gower@inria.fr, robert.gower@ens.fr

References

- [1] Petros Drineas, Michael W Mahoney, and S Muthukrishnan. “Sampling algorithms for l2 regression and applications”. In: *Proceedings of the seventeenth annual ACMSIAM symposium on Discrete algorithm* (2006), pp. 1127–1136.
- [2] Robert Mansel Gower and Peter Richtárik. “Randomized Iterative Methods for Linear Systems”. In: *SIAM Journal on Matrix Analysis and Applications* 36.4 (2015), pp. 1660–1690.
- [3] Robert Mansel Gower and Peter Richtárik. “Randomized Quasi-Newton Updates are Linearly Convergent Matrix Inversion Algorithms”. In: *arXiv:1602.01768* (2016).
- [4] M. R. Hestenes and E. Stiefel. “Methods of Conjugate Gradients for Solving Linear Systems”. In: *Journal of research of the National Bureau of Standards* 49.6 (1952).
- [5] M. Pilanci and M.J. Wainwright. “Randomized Sketches of Convex Programs With Sharp Guarantees”. In: *Information Theory, IEEE Transactions on* 61.9 (2015), pp. 5096–5115.
- [6] Mert Pilanci and Martin J. Wainwright. “Iterative Hessian sketch : Fast and Accurate Solution Approximation for Constrained Least-Squares”. In: *Journal of Machine Learning Research* 17 (2016), pp. 1–33. arXiv: 1411.0347v1.