Robust sparse recovery by stochastic optimization
M2 MSIAM 2019-2020 Master Thesis Proposal

Supervisor: Anatoli IOUDITSKI, anatoli.juditsky@univ-grenoble-alpes.fr

Description: A simple version of the problem we consider is as follows:

Let $\theta \in \Theta \subset \mathbf{R}^n$ be unknown signal. Given observations $\varphi_t \in \mathbf{R}^n$ and $y_t \in \mathbf{R}$ such that

$$y_t = \varphi_t^T \theta + e_t, \;\; t = 0, 1, ...m,$$

where $(e_t)_{t \geq 0}$ are scalar noises, we aim at estimating θ.

In the problem setting we are interested in signal θ is assumed high-dimensional and *approximately sparse* – can be well approximated by a vector with only $s \ll n$ nonvanishing entries, and regressors φ_t and noises e_t are random and independent across t.

The above problem of sparse recovery has recently received much attention. Classical ("off line") estimation techniques for this problem – Lasso and Dantzig Selector estimates – are now well studied [1] and are subject of statistical textbooks. On the other hand, algorithms of stochastic optimization for solving the above problem have received much less attention [2]. Recursive structure of such algorithms allowing for "on line" treatment of the data made them popular in the 60-80's, when memory and processing power limitations did not allow implementing numerically more challenging techniques. Study of these algorithms regained momentum recently motivated by new machine learning applications usually characterized by a very large scale of parameters and of available data that which, once again, rules out the use of classical techniques, non recursive (off line), which require excessive data storage and cannot treat the data flow efficiently.

The aim of this thesis is to study the properties of stochastic approximation algorithms of the proximal type when applied to the sparse recovery problem. Precisely, we will consider implementations of stage-wise stochastic approximation [3,4] tuned for sparse and approximately sparse solutions, satisfying the following requirements:

- they should obey optimal statistical bounds on the estimation performance, typically measured by the risk of estimation $\widehat{\theta}_t$ (e.g., $\mathbf{E}\{\|\theta - \widehat{\theta}_t\|^2\}$ where $\|\cdot\|$ is a given norm) or prediction risk (e.g., $\mathbf{E}\{(y_t - \varphi_t^T \widehat{\theta}_t)^2\}$) with the final objective of removing existing limitations of the algorithms developed in [2];
- algorithms should be robust to heavy-tail distributions of observation disturbances e_t and erroneous observations of regressors φ_t;
- they should adapt to heavily parallelized implementation – their numerical performance should "scale" according to available memory or processing resources and architecture.

The Master thesis will be supervised by Anatoli Iouditski at Laboratoire Jean Kuntzmann (LJK) Grenoble, and is expected to lead to a PhD thesis on the same topic.

[1] Candès, E. J. (2006). Compressive sampling. In *Proceedings of the international congress of mathematicians*, **3**, 1433–1452.
Agarwal, A., Negahban, S., and Wainwright, M. J. (2012). Stochastic optimization and sparse statistical recovery: Optimal algorithms for high dimensions. In *Advances in Neural Information Processing Systems* 1538–1546.
[3] J., Nesterov, Y. (2014). Deterministic and stochastic primal-dual subgradient algorithms for uniformly convex minimization. *Stochastic Systems*, **4**(1), 44–80.
[4] Lan, G. (2012). An optimal method for stochastic composite optimization. Mathematical Programming, 133(1), 365-397.