Accelerated algorithms for large-scale parameter tracking
M2 MSIAM 2017-2018 Master Thesis Proposal

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

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

Let $\theta_t \in \Theta \subset \mathbf{R}^n$, $t = 0, 1, 2, ...$ be unknown time-varying parameter. Given observations $\varphi_t \in \mathbf{R}^n$ and $y_t \in \mathbf{R}$ such that

$$y_t = \varphi_t^T \theta_t + e_t, \ \ t = 0, ..., T$$

where $(e_t)_{t\geq 0}$ are scalar noises, we aim at estimating the parameter vector θ_T at instant T (or, say, predicting θ_{T+k}, for some $k \in \mathbf{Z}_+$, etc). It is usually assumed that θ_t is *slowly varying*, e.g., it satisfies

$$\theta_t = \theta_{t-1} + \gamma w_t, \ \ t \geq 0$$

where $w_t \in \mathbf{R}^n$ is a random or deterministic perturbation and $\gamma > 0$ is "small."

The above problem arises in numerous control, signal and image processing applications, and has received much attention. The "classical" algorithm used to build *recursively* the estimate $\widehat{\theta}_t$ of θ_t is of the form (LMS, RLS, Kalman filter, etc)

$$\widehat{\theta}_t = \widehat{\theta}_{t-1} + \mu L_t \rho(y_t - \varphi_t^T \widehat{theta}_t) \tag{1}$$

where μ is the scalar algorithm gain, and various methods differ in the way the scalar *score function* $\rho(\cdot)$ and the vector gains L_t are computed. The recursive structure of the algorithm allowing *on-line* treatment of the data made it popular in the 60-80's, when memory and processing power requirements did not allow implementing numerically more challenging techniques.

This problem currently sees a rise of interest motivated by new machine learning applications. These applications are usually characterized by a very large scale of parameters and of available data that, once again, rules out the use of classical techniques, either non recursive (*off-line*) which require excessive data storage and cannot treat the data flow efficiently, or compute the estimate $\widehat{\theta}_t$ recursively but use the choice of the vector gain L_t in the recursion (1) which is too hard to compute when the problem dimension is high.

The aim of this thesis is to develop and study theoretically and numerically some alternatives to (1) which rely upon "simple" accelerated first-order algorithms of deterministic and stochastic optimization (see, e.g., [1,2]) which also satisfy the following requirements

- they should obey theoretically sound bounds on the tracking performance, typically measured by the risk of prediction (e.g., $\text{Risk}[\widehat{\theta}_t] = \mathbf{E}\{(y_t - \varphi_t^T \widehat{\theta}_t)^2\}$;
- they should adapt to heavily parallelized implementation – their numerical performance should "scale" according to available memory or processing resource 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] Nesterov, Y. (2013). Gradient methods for minimizing composite functions. Mathematical Programming, 140(1), 125-161.
[2] Lan, G. (2012). An optimal method for stochastic composite optimization. Mathematical Programming, 133(1), 365-397