Master Research Internship 2018/2019
Supervised by Franck Iutzeler

UNIVERSITÉ Grenoble Alpes

Structured Acceleration of Optimization methods for Machine Learning problems

Context: The growth and diversification in artificial intelligence systems has led to tremendous changes in the optimization methods used in machine learning. An important novelty in modern algorithms for learning is the use of *dimension reduction.* It consists in identifying pertinent directions in the variable search space and concentrating most computational efforts onto these directions. A successful and typical example of dimension reduction is the screening of variables for the lasso problem; e.g. the strong rules [6] used in the popular GLMNET R package, or more recently the scikit-learn compatible package Celer (package url: https://mathurinm.github.io/celer/) [4].

The success of these methods is partly due to the fact that the optimization problem at hand is *strongly structured* and this structure is *harnessed* to produce computationally efficient methods. Indeed, structure is often present in Machine Learning as it is *brought by regularization* (for instance, ℓ_1-norm regularization, elastic net, etc. enforce sparsity of the models). In addition, it is well-known that the iterates of most popular optimization methods actually become sparse in finite time for ℓ_1-norm regularized problems, in which case they are said to *identify* some sparsity pattern; at this point, the convergence of the algorithm gets faster [3].

Recently, identification results have been extended to more general classes of regularizers (e.g. 1D Total Variation, trace norm, see [2]) for which no numerical dimension reduction methods are available yet. Efficiently harnessing more general types of identification for improving the convergence of numerical optimization algorithms is the general idea of this internship.

Objective: Accelerating first-order methods such as the (proximal) gradient algorithm can be done

- by adding inertia (predicting future iterates using past ones) as in Nesterov's fast gradient [5] and FISTA [1]
- by quasi-Newton (adding second order information to go in a better direction than the gradient) as in BFGS
- etc.

However, these techniques are still *structure-blind* and can actually slow down identification by making the iterates leave an otherwise stable subspace. The goal of this internship is thus to develop and analyze *structure-adapted acceleration schemes* for first-order methods.

While this internship is mostly theoretical, some implementation of the proposed methods on usual machine learning objectives is expected.

Practical Details: 6 months internship between March and Sept. 2019 (adaptable)

- Supervised by Franck Iutzeler http://www.iutzeler.org.
- DAO team, Laboratoire Jean Kuntzmann (applied maths), Univ. Grenoble Alpes.
- Prerequisites: Strong background in linear algebra and optimization with an interest in machine learning.
- Proficiency in Python is appreciated.
- Candidates should be motivated to continue with a Ph.D. thesis.
- Apply by mail at franck.iutzeler@univ-grenoble-alpes.fr

[1] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[2] J. Fadili, J. Malick, and G. Peyré. Sensitivity analysis for mirror-stratifiable convex functions. *arXiv:1707.03194*, 2017.

[3] J. Liang, J. Fadili, and G. Peyré. Local linear convergence of forward–backward under partial smoothness. In *Advances in Neural Information Processing Systems*, pages 1970–1978, 2014.

[4] M. Massias, J. Salmon, and A. Gramfort. Celer: a fast solver for the lasso with dual extrapolation. In *International Conference on Machine Learning*, pages 3321–3330, 2018.

[5] Y. 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.

[6] R. Tibshirani, J. Bien, J. Friedman, T. Hastie, N. Simon, J. Taylor, and R. J. Tibshirani. Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2):245–266, 2012.