SCALABLE SPLITTING ALGORITHMS FOR LARGE SCALE OPTIMIZATION

PoliMi Advisor: Silvia Villa,
https://www.mate.polimi.it/view=pp&id=552&lg=it/?view=pp&id=552&lg=it#ann,
silvia.villa@polimi.it

In the last years, convex optimization has been revolutionized by data availability and data nature [2]. On the one hand, optimization problems of an unprecedented scale need to be solved. On the other hand, the computation of the solution must be robust with respect to perturbations and not necessarily very accurate, since data are incomplete or affected by noise. These two aspects fueled the design of provably convergent first-order optimization methods. These algorithms use only first order information about the objective function, such as gradient estimates, and thus are easy to implement, robust, and they have a low cost per each iteration. Moreover, they can handle nonsmooth objective functions, arising for example from sparsity priors, using proximal tools [3]. First order methods combined with monotone operator theory and the splitting technology represent state of the art optimization methods to solve large scale regularized problems [1].

However, for huge scale problems, the implementation of such algorithms still faces challenges that render them often inapplicable in their original form.

The starting point of this thesis is the observation that the success of state-of-the-art splitting methods is due to their ability to take advantage of the structure of optimization problems arising in data science. These are often the sum of data fit and regularization terms, with different smoothness or sparsity promoting properties, often composed with large-size linear operators. Splitting methods are able to isolate the contribution of each summand, and activate each element in the sum independently. The objective of this thesis is to exploit further structural properties of data driven optimization problems to derive more efficient algorithms. The idea is to jointly consider modeling and optimization with two main outcomes. First, the design of optimization methods intrinsically related to the modeling assumptions, and second, the use of the model structure to derive a sharper convergence analysis of the proposed algorithms with respect to the worst case one. Results in this direction can be found in [4, 5]. More precisely, in this thesis we plan to analyze from a theoretical point of view some heuristics that improve the scalability of optimization algorithms and are currently used in practice. Examples include multiscale approaches [7], possibly combined with stochastic [8] and asynchronous implementations [6].

References

[1] H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, second ed., Springer, New York, 2017.

[2] V. Cevher, S. Becker, M. Schmidt, Convex optimization for big data: scalable, randomized, and parallel algorithms for big data analytics, *IEEE Signal Processing Magazine* **31** (2014), 32-43.

[3] P.L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, (H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Editors), 185-212. Springer, New York, 2011.

[4] D. Drusvyatskiy and A.S. Lewis, Error bounds, quadratic growth, and linear convergence of proximal methods, *arXiv:1602.06661*, 2016.

[5] G. Garrigos, L. Rosasco and S. Villa, Convergence of the forward-backward algorithm: beyond the worst case with the help of geometry, *arXiv:1703.09477*, 2017.

[6] J. Liu and S. Wright, Asynchronous stochastic coordinate descent: parallelism and convergence properties, *SIAM Journal on Optimization*, **25** (2015), 351-376.

[7] P. Parpas, A multilevel proximal gradient algorithm for large scale optimization, *SIAM Journal on Scientific Computing*, to appear.

[8] L. Rosasco, S. Villa, and B. C. Vu, Stochastic forward-backward splitting for monotone inclusions, Journal Optimization Theory and Applications, **169** (2016), 388-406.