On the Interplay between Acceleration and Identification for the Proximal Gradient algorithm

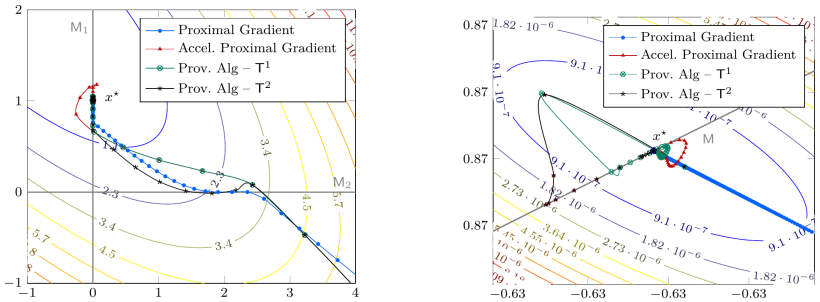
Gilles Bareilles
Univ. Grenoble Alpes

Keywords : Proximal Gradient, Accelerated methods, Identification.

Many problems in machine learning or linear inverse problems are formulated as composite non-smooth optimization problems. The non-smoothness is chosen to enforce some desirable structure on the solution, but it also affects the iterates produced by optimization methods. In this talk, we discuss the conditions under which the Proximal Gradient algorithm and its accelerated counterpart [1, 2] are able to collect this structure and how they do so.

Identification We are interested in solving $\min_{x\in\mathbb{R}^n} f(x) + g(x)$, where f is a differentiable convex function and g is a convex function, generally non differentiable.
As previously mentioned, the non-differentiable function g is chosen to enforce some structure on the optimal solution. For example, coordinate sparsity is enforced by taking g as the ℓ_1-norm, and yields more interpretable models for linear regression (in which case, the problems boils down to the lasso). Besides, the iterates of a given optimization method progressively collect that optimal structure, in a more or less stable way. When they reach (and then stay in) that optimal structure, we say that *identification* happens.

(Accelerated) Proximal Gradient Besides ensuring identification, it is desirable that iterates collect the optimal structure in a *stable* manner. Indeed, a stable identification behavior means that stopping the algorithm before convergence still provides some relevant structural information. Based on [3], we show that the Proximal Gradient does identify in a stable way, but its accelerated counter part has a more erratic behavior. We also propose two *Provisional algorithms* that benefit from the $\mathcal{O}(1/k^2)$ rate of accelerated methods while being more stable.



References

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

[2] Y. Nesterov, *A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$*, Dokl. akad. nauk Sssr, vol. 269 (1983).

[3] G. Bareilles and F. Iutzeler, *On the Interplay between Acceleration and Identification for the Proximal Gradient algorithm*, preprint arXiv:1909.08944 (2019).