Title : Fast Algorithmic Methods for Optimization and Learning (FAMOL)

The many challenges posed by machine learning and the processing of big and noisy data require the development of new mathematical tools and fast algorithms in optimization. In this course, we trace recent advances in fast first-order optimization algorithms. The acceleration of optimization methods is a current research topic and many teams in France (and around the world) are working in this area. First-order methods, such as gradient descent or stochastic gradient, have grown in popularity. A major development is due to Y. Nesterov who proposed in 1983 an accelerated gradient algorithm having a faster global convergence rate than gradient descent. In line with Nesterov's method, the FISTA algorithm, proposed by Beck and Teboulle in 2009, is a splitting algorithm for solving convex optimization problems with a « smooth+ nonsmooth » additive structure. It has proven to be very successful in the machine learning and signal and image processing communities. Gradient-based optimization algorithms can also be studied from the perspective of Ordinary Differential Equations (ODE). This point of view allows us to propose new algorithms obtained by time discretization of these ODEs and to improve their performances. We will make the link with the Ravine method, introduced by Gelfand and Tsetlin in 1961, and longtime ignored. Indeed, Nesterov's accelerated gradient method and the Ravine method can be deduced from each other by reversing the order of the extrapolation and gradient operations in their definitions. This is why the Ravine method is often used by practitioners, and sometimes confused with Nesterov's Accelerated Gradient.
The damping term is the crucial ingredient of the fast optimization properties of the inertial dynamical systems. In addition to the viscous damping on which the dynamics described above are based, we will show that the geometric damping driven by the Hessian of the function to be minimized is a crucial ingredient to improve their performance, and significantly reduce the oscillations that come with inertial systems.
The mini-course will be supported by historical facts and open questions.

Part 1 (Tuesday 11 October, 2022: 14h00-17h00).
- Gradient descent method (discrete) and Gradient Flow (continuous): some historical aspects
- Polyak's momentum and the heavy ball with friction (strongly convex and general case).
- Nesterov's Accelerated Gradient Method (NAG): convergence analysis. The Su-Boyd-Candès Model, with asymptotic vanishing damping coefficient, as continuous surrogate of NAG. The composite problem: ISTA and FISTA.
- The limit form of the Su-Boyd-Candès dynamic when the viscous parameter becomes large: the temporal scaling techniques.

Part 2 (Wednesday 12 October, 2022: 09h00-12h00).
- First-order optimization algorithms via inertial dynamics with the Hessian-driven damping
- The Ravine Accelerated Gradient (RAG) and its link with the Nesterov Accelerated Gradient (the discrete case)
- High-resolution ODE of (NAG) and (RAG) makes appear the Hessian.
- Convergence rates of (NAG) and (RAG): some improvements.

- Conclusion, perspectives and open questions of the two parts.

References

[1] S. Adly, H. Attouch, On the limit form of the Su-Boyd-Candès dynamic version of Nesterov's accelerated gradient method when the viscous parameter becomes large, Preprint (2022).
[2] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, First order optimization algorithms via inertial systems with Hessian driven damping, Mathematical Programming, (2020).
[3] H. Attouch and J. Fadili. From the Ravine method to the Nesterov method and vice versa: A dynamical system perspective. To appear in SIOPT. arXiv:2201.11643v1, 2022.
[4] A. Beck and M. Teboulle, A Fast Iterative Shrinkage-Thresholding Algorithm for linear inverse problems, SIAM J. Imaging Sciences (2009).
[5] Y. E. Nesterov, A method for solving the convex programming problem with convergence rate O(1/k2), Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.
[6] A. Chambolle, T. Pock , A first-order primal-dual algorithm for convex problems with applications to imaging, HAL CNRS (2009).
[7] P.-E. Maingé, F. Labarre, Accelerated methods with fastly vanishing subgradients for structured non-smooth minimization, Numerical Algorithms, (2021).
[8] W. Su, S. Boyd, and E. Candès, A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights. Journal of Machine Learning Research, 17(153):1-43, 2016.
[9] B. Shi, S.S. Du, M.I. Jordan, W.J. Su. Understanding the Acceleration Phenomenon via High-Resolution Differential Equations, ArXiv 2018 : https://arxiv.org/abs/1810.08907v3