M2 internship : Theoretical and practical aspects of spectral estimation for optimal optimization algorithms design

The interested candidate should contact both supervisors by email, attaching a CV and briefly detailing their motivation for the internship.

Supervision: Rémi Gribonval, Inria Lyon (remi.gribonval@inria.fr)
Mathurin Massias, Inria Lyon (mathurin.massias@inria.fr)

Location: The internship will take place at INRIA Lyon (located at ENS de Lyon, 46 allée d'Italie, 69007 Lyon), in the Ockham team.

Profile: We are looking for a highly motivated student with a background in mathematics (continuous optimization, machine learning, linear algebra, probability and statistics). Strong abilities in computer sciences are required. Experience with Python is a plus. If the candidate is successful, this internship may be pursued by a PhD.

Salary: The intern will be granted the usual stipend of ∼600 euros/month.

Keywords: Optimization, gradient descent, optimal transport, polynomial approximation

1 Context

The convergence analysis of smooth optimization algorithms greatly depends on the spectral properties of the objective function's Hessian (Nocedal and Wright, 1999). In particular, upper and lower bounds on the Hessian spectrum are key quantities that govern the convergence rates of many first and second order algorithms, such as gradient descent or Newton method. Such convergence rates are generally worst-case over a class of functions sharing the same spectral bounds Nesterov (2018).

A finer analysis, proposed by Pedregosa and Scieur (2020), developed optimal and average convergence rates of first order methods (gradient descent and Polyak acceleration) on *random* classes of functions, based on their spectrum (e.g., postulating a Marchenko-Pastur distribution). The analysis relies on polynomial approximation tools, with principles guiding towards an adequate tight choice of coefficients (Fischer, 2011). This work is however limited to quadratic objectives, for which the Hessian and a fortiori its spectrum remain constant across iterations. A natural question is to extend this approach to design provably (near)optimal algorithms when the Hessian spectrum varies.

2 Internship

The internship will study various questions related to imperfect knowledge of the spectrum.

- On quadratics, the algorithm of Pedregosa and Scieur (2020) requires to estimate parameters of a density (e.g., the scale parameter of an exponential law, or the bounds of a Marchenko Pastur distribution). The authors have proposed to accomplish this via moment matching. One open question is how does an imperfect estimation of the spectrum affects the theoretical and practical performance of the algorithm? In particular, which perturbation impact the algorithm the most (in an adversarial fashion), and how can they be measured – for example, using optimal transport tools to measure distances between distributions. How can one connect the ideal algorithm for a given spectrum, and the optimal for the law it's sampled from?

- The candidate will then investigate the case where the spectrum varies smoothly over time, such that reestimating the spectrum periodically may become a requirement. How often should it happen in theory, and what are computationally efficient ways to do so? The candidate will perform an empirical study of this problem on adequate neural networks (Sagun et al., 2017).
- The dimension d of modern optimization problems is large enough that optimization algorithms rarely perform more than $\mathcal{O}(d)$ iterations in practice. The candidate will investigate the different regimes of optimization, in particular the early stage, and their relationship with the Hessian spectrum.

References

Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

Yurii Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018.

Fabian Pedregosa and Damien Scieur. Acceleration through spectral density estimation. In *International Conference on Machine Learning*, pages 7553–7562. PMLR, 2020.

Bernd Fischer. *Polynomial based iteration methods for symmetric linear systems*. SIAM, 2011.

Levent Sagun, Utku Evci, V Ugur Guney, Yann Dauphin, and Leon Bottou. Empirical analysis of the Hessian of over-parametrized neural networks. *arXiv preprint arXiv:1706.04454*, 2017.