Improving our understanding of SGD with Lyapunov energy arguments

Daniel CORTILD, Bernoulli Institute - Groningen
Guillaume GARRIGOS, LPSM - Paris
Robert GOWER, Flatiron Institute - New York
Lucas KETELS, LPSM & Bernoulli Institute - Paris & Groningen
Juan PEYPOUQUET, Bernoulli Institute - Groningen

One of the cornerstones in Hedy Attouch's analysis of dynamical systems is the use of Lyapunov energy arguments. In this talk, we aim to show how those elegant ideas can be used to analyze the Stochastic Gradient algorithm (SGD) [2]. The talk will cover two distinct results, each illustrating the power of Lyapunov analysis in a different fashion.

In the first part, we present sharp and **new complexity results** for SGD, under minimal variance assumptions, and with the focus on large step sizes.
The standard analysis of SGD usually relies on assuming some sort of uniform bound on the gradient noise, which is usually impossible to verify in practice. Recent efforts made it possible to simply assume that the gradient noise is bounded at the solution, which is trivially true for many practical problems. The drawback of this approach is that it usually restricts the stepsizes to be short, typically strictly less than 1/L.
We show that we can push further and obtain complexity results up to 2/L, thanks to a tight Lyapunov analysis [1]. On top of this, we show that the obtained complexity bounds are sharp, by using the performance estimation problem (PEP) framework which allows to relax our problem into a solvable semidefinite program.

In a second part, we will illustrate how focusing on a Lyapunov energy of the dynamic can help to design **new methods**.
We present the so-called *Polyak stepsize*, which is a formula providing an automatic and powerful choice for the stepsize : it leads to near-optimal convergence rates, and automatic adaptivity to local smoothness of the problem [3]. As a matter of fact, this rule was discovered by maximizing the decrease of a certain Lyapunov !
Then we will discuss how much *fraudulous* is this algorithm, since such formula cannot be implemented in general. We will see that we can circumvent this difficulty in the interpolation regime (the formula becomes trivial) or in a teacher / student distillation scenario (a proxy for the forumla is available).

[1] D. Cortild, G. Garrigos, L. Ketels, J. Peypouquet. *Stochastic gradient descent without variance assumption : A tight lyapunov study for a wide range of step-sizes (to appear)*, 2025.

[2] G. Garrigos, R. M. Gower. *Handbook of Convergence Theorems for (Stochastic) Gradient Methods*, 2023. doi :10.48550/arXiv.2301.11235.

[3] R. M. Gower, G. Garrigos, N. Loizou, D. Oikonomou, K. Mishchenko, F. Schaipp. *Analysis of an Idealized Stochastic Polyak Method and its Application to Black-Box Model Distillation*, 2025. doi :10.48550/arXiv.2504.01898.

Contact : garrigos@lpsm.paris