STABILITY OF MACHINE LEARNING OPTIMIZERS FROM A NUMERICAL ANALYSIS POINT OF VIEW

B.Bensaid [a,b*], G.Poette [a], R.Turpault [b]

[a] CEA/CESTA, DAM, F-33114 Le Barp, France.
[b] Institut de Mathématiques de Bordeaux (IMB), Université de Bordeaux, CNRS, Bordeaux INP, F33405, Talence, France.

ABSTRACT

In this work we analyse the most popular machine learning optimizers using ODEs tools and numerical analysis concepts. This leads to a new way to design new optimizers that possess naturally stability and convergence properties.

INTRODUCTION

Deep learning models have recently been used to substitute parts of simulation codes with neural networks [1]. In this context, huge networks do not seem appropriate from a computational point of view. With this in mind, this work aims at building efficient shallow networks. The guiding principle of this research starts from this assessment: some recent papers [2] point the optimizer out as a lever to improve performances. There are stochastic and deterministic optimizers. Before analysing stochastic algorithms we suggest studying first the deterministic ones (as stochastic optimizers are often based on deterministic ones).

FROM LYAPUNOV INSTABILITY TO ODE ANALYSIS

In a previous work [3], we have analysed some optimizers of the literature such as Gradient Descent (GD), Momentum [5] and Adam [6]. This study pointed out the lack of convergence properties and Lyapunov stability of the classical optimizers as illustrated by figure 1. On these maps for each minimum we assign a colour to all the initial points that converge to this minimum and the brown part refers to divergent trajectories. Starting very close to a given minimum, GD can diverge whereas a variant of Adam called AWB converges to a point far away from the initial guess. This is a very concerning issue because the optimizer can deteriorate the initialization and the a priori knowledge integrated in the neural networks. To understand how to fix this problem it has been useful to introduce some ODE approximations in the sense that the iterates converge to the solution of the ODE when the learning rate (time step) tends to 0. In this framework, it is possible to build some functions that decrease along the trajectories of the ODE. By choosing an adaptive time step in order to decrease the discretisation of these Lyapunov functions, it comes up a new way to design algorithms that have better stability properties and performances without tuning hyperparameters.

FROM ODE TO LYAPUNOV CONTROL OPTIMIZERS

In this work (under review in [4]), we go a step further by preserving the continuous dissipation rate (given by the ODEs) of Lyapunov functionals in a weak form. Imposing this supplementary inequality in addition to a convergent discretisation of the ODE gives **much faster algorithms with better predictive performances**

*Correspondence to bilel.bensaid@u-bordeaux.fr

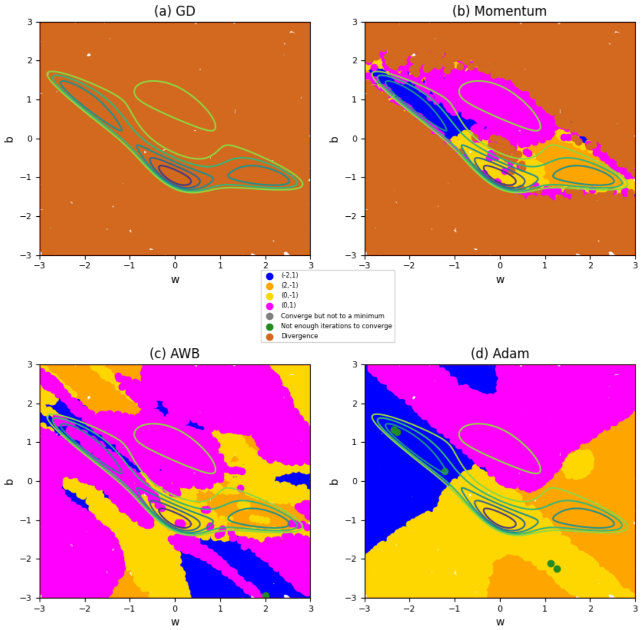

FIGURE 1: Sensitivity analysis of the classical optimizers to the initial point

compared to the state of the art, on many classical machine learning tasks such as MNIST, FASHION-MNIST,... Combining ideas from Lyapunov control theory and dissipation ODEs [7], this framework enables to **prove stability and convergent properties and to compute convergence rates in the non-convex** setting.

REFERENCES

[1] G.Kluth and al *Deep learning for NLTE spectral opacities* Physics of Plasmas, vol 27, 2020.

[2] P.Novello,G.Poëtte and al *Goal-Oriented Sensitivity Analysis of Hyperparameters in Deep Learning* Journal of Scientific Computing, vol 94, 2023

[3] B.Bensaid, G.Poëtte, R.Turpault *Deterministic Neural Networks Optimization from a Continuous and Energy Point of View* Journal of Scientific Computing, vol 96, 2023.

[4] B.Bensaid, G.Poëtte, R.Turpault *Lyapunov Optimization for Neural Networks Training* Journal of Machine Learning Research, 2023

[5] B.T.Polyak *Some methods of speeding up the convergence of iteration methods* USSR Computational Mathematics and Mathematical Physics, vol 4, 1964

[6] D.Kingma and B.Jimmy *Adam: A Method for Stochastic Optimization* International Conference on Learning Representations, 2014

[7] A.Haraux and M.Jendoubi *The Convergence Problem for Dissipative Autonomous Systems: Classical Methods and Recent Advances* SpringerBriefs in Mathematics, 2015