PhD Open Days

Towards a Physics Understanding of Deep Learning

PhD in Physics

Carlos Couto (carloscouto@tecnico.ulisboa.pt)

FCT Fundação para a Ciência e a Tecnologia

Work supported by the PhD grant 2021.06468.BD.

CeFEMA

CAMGSD

instituto de telecomunicações

Summary

Deep learning (DL) has achieved tremendous success. However, a theoretical understanding of deep neural networks (DNNs) is still lacking. Recently, connections between DNNs and Physics have been proposed to address this issue. Particularly promising are the relations of DNN learning with the **Renormalization Group** (RG) and with **dissipative methods to improve gradient descent algorithms** (GD).

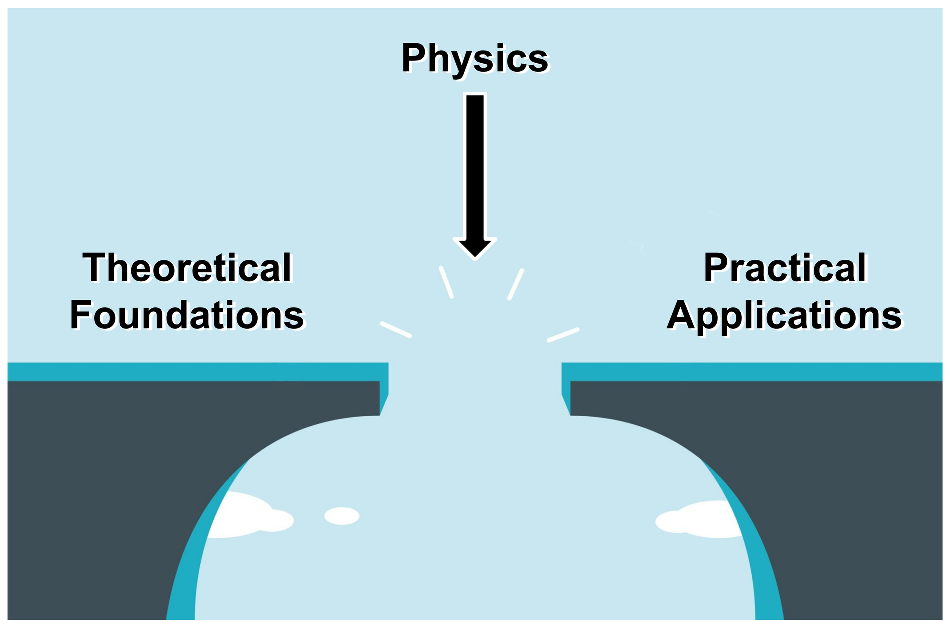

Figure 1: Physics may help in bridging the gap between the foundations and practical applications of machine learning.

Three key areas where Physics may help shed new light are:

- The connection between overparameterization and gradient descent;
- Accelerating gradient descent training via complex Hamiltonian flows;
- Connecting deep neural networks with the renormalization group scheme.

Overparameterization & gradient descent

- Neural network models with a **very large number of parameters often generalize better** than those with a lesser number of parameters. One possible contributor to this effect is that, **by adding additional parameters, we "open" new paths for the GD to reach a lower minimum**, in the case where it was previously stuck on a local minimum.

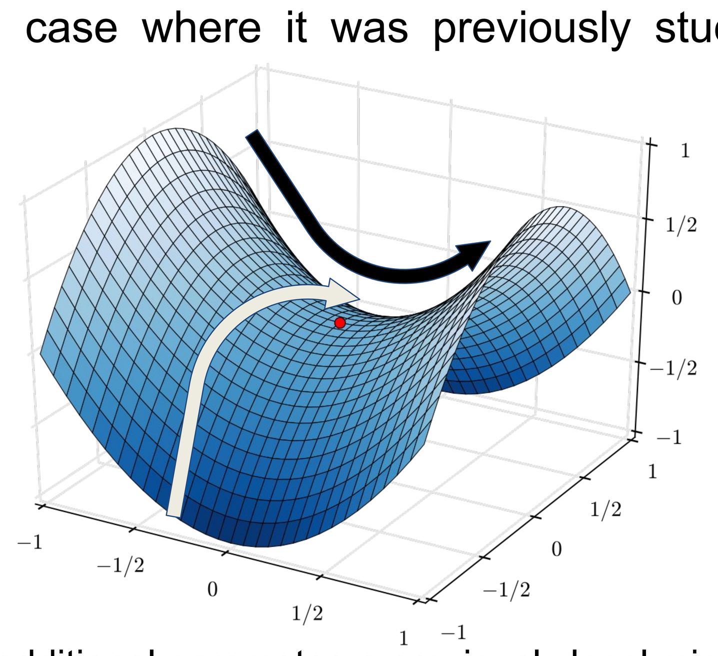

Figure 2: By adding an additional parameter, a previously local minimum (along the black arrow) is now a saddle point. This allows the GD method to achieve a smaller minimum (via the lighter arrow).

- Our goal is to formulate the learning process as an evolution of a probability distribution in weight-space. Here, we will study the **dynamical signatures underlying successful learning** and the **role of overparameterization in reaching a near-optimal solution**.

Complex Hamiltonian Dynamics

- Neural networks used in the real world are often made up of **millions of parameters**. As such, the training procedure is **very expensive both in time and energy**. Any **improvements** to the training procedure which achieve either faster or better results **are of extreme importance**.
- Recently, modifications of the (first-order) GD method in machine learning have been considered by using second-order dissipative Hamiltonian systems [MJ], yielding **better and faster convergence** of neural networks.
- We can generalize these modifications to a non-Hermitian Hamiltonian setting. Our goal here is to compare the Hamiltonian dynamics in complex time with [MJ] to check if further improvements are viable.

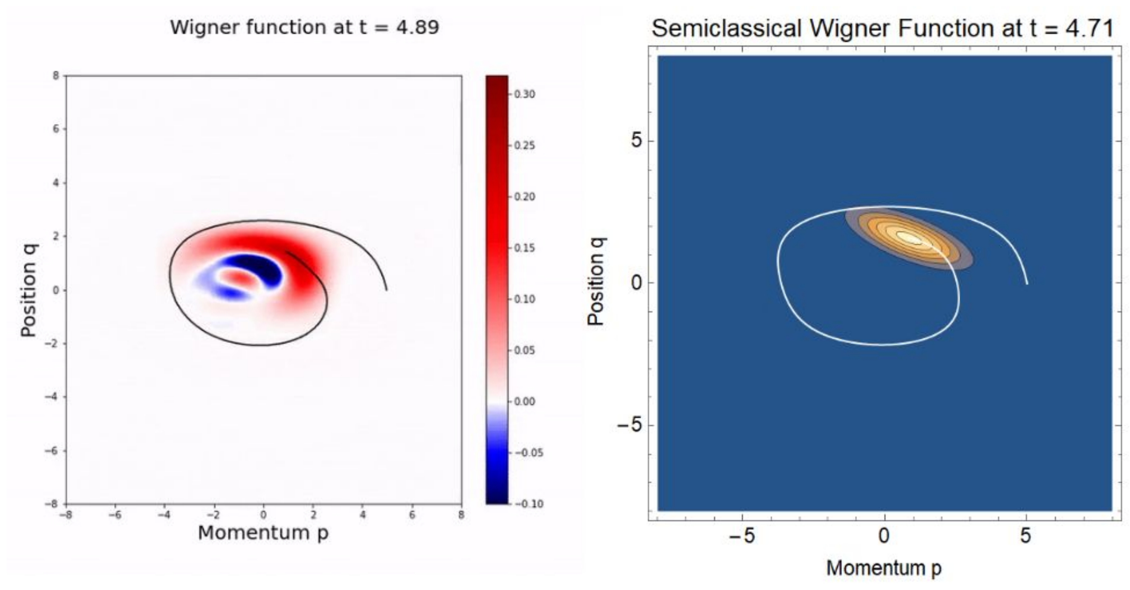

Figure 3: Preliminary work showing numerical and analytical simulations of non Hermitian Hamiltonian dynamics for a gaussian coherent state.

Renormalization Group

- The **renormalization group** is an iterative coarse-graining scheme that allows for the **extraction of relevant features** (i.e. operators) as a physical system is examined at different length scales.

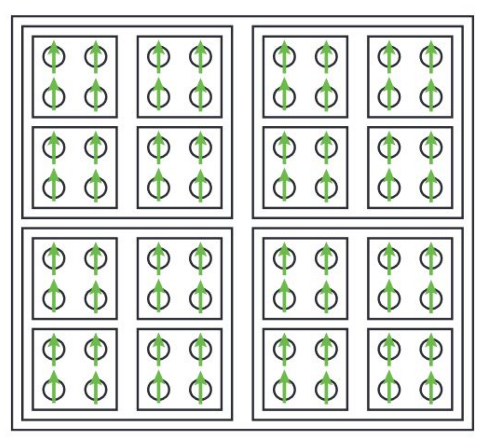

Figure 4: Block spin RG. At each iteration we group four spins into a new effective spin. By iterating this procedure we average over a larger and larger number of spins. Image taken from [MS].

- Our goal here is to understand how closely connected these two concepts are and to try to generalize the work done in [MS, KKC] to study the **link between RG and data compression** in classification problems.

References

[MJ] - Jordan, M. I. (2018). Dynamical, symplectic and stochastic perspectives on gradient-based optimization. In Proc. Int. Cong. of Math., 2018, Rio de Janeiro, Vol 1, 523-550.

[MS] - Mehta, P., & Schwab, D. J. (2014). An exact mapping between the variational renormalization group and deep learning. arXiv preprint arXiv:1410.3831.

[KKC] - Koch, E. D. M., Koch, R. D. M., & Cheng, L. (2020). Is Deep Learning a Renormalization Group Flow?. IEEE Access, 8, 106487-106505.