cea

Institut de Mathématiques de Bordeaux

Deterministic optimization in Deep Learning from a continuous and energetic point of view

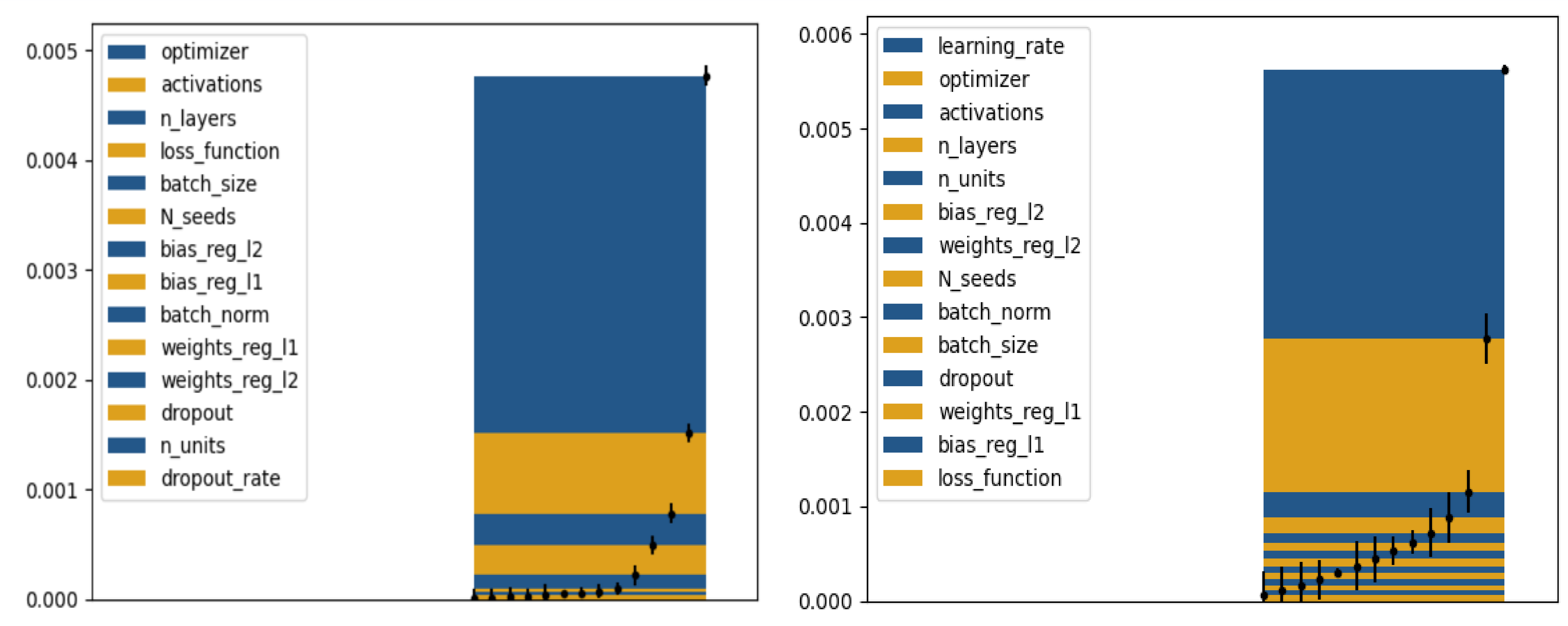
Bensaid Bilel[1,2], Turpault Rodolphe[1] and Gael Poette[2]

[1]Institut de Mathematiques de Bordeaux, F-33405 Talence, France & [2]CEA, CESTA, DAM, F-33114 Le Barp, France

Context and Issue

- Substitute parts of simulation codes with Deep Learning Models $\Rightarrow$ Build efficient **shallow** networks
- Sensitivity analysis $\Rightarrow$ the **OPTIMIZER** appears as the main leverage to improve performances
- **Objective** :

Better understanding of the deterministic algorithms



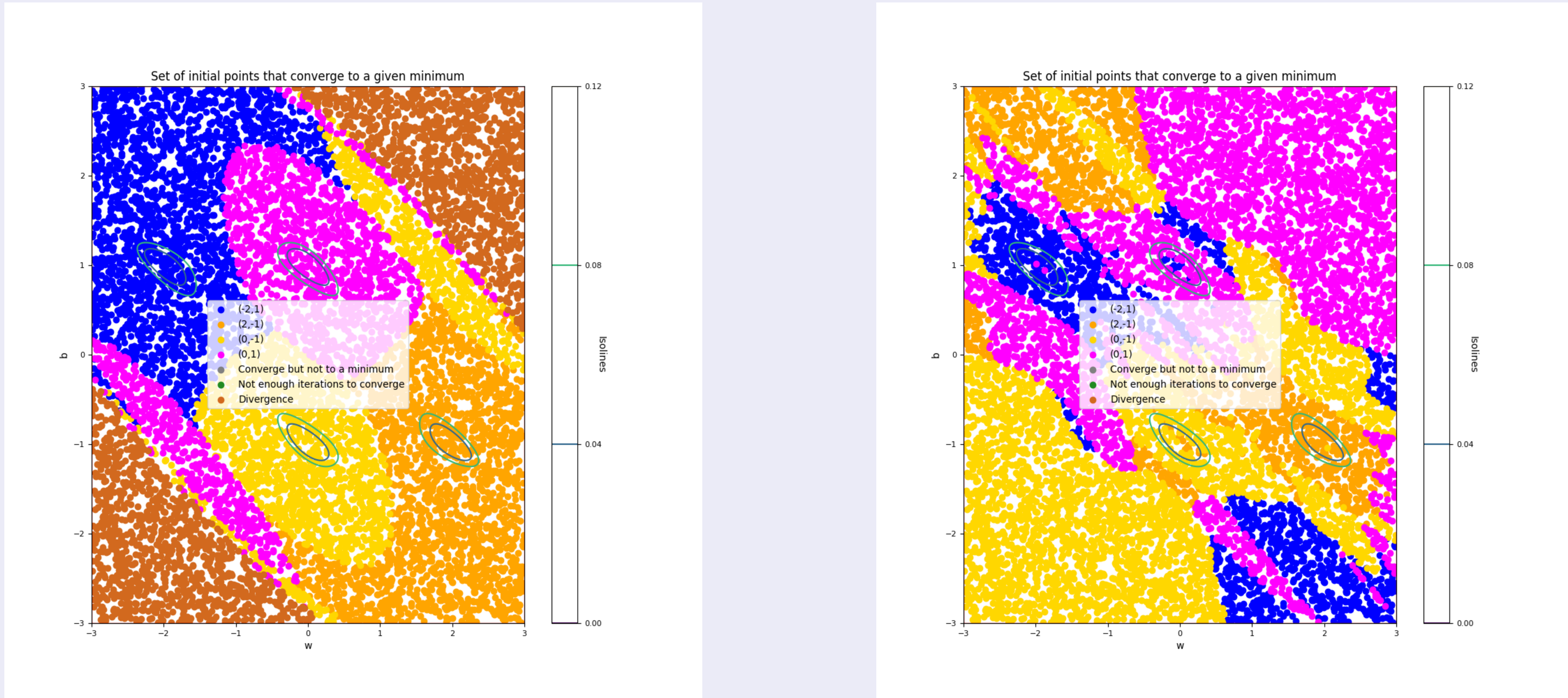
(a) Runge function (b) Bateman equation

Global sensitivity analysis of hyperparameters [2]

Starting Point

- NEW TEST METHODOLOGY : Construction of neural networks which enable the **localization of critical points**;
- **Sensitivity to the initializer** $\Rightarrow$ Display for each mnimum θ^* the regions :
  - $\mathcal{A}$: a given optimizer;
  - ϵ : a small positive value (fixed to 10^{-3}).

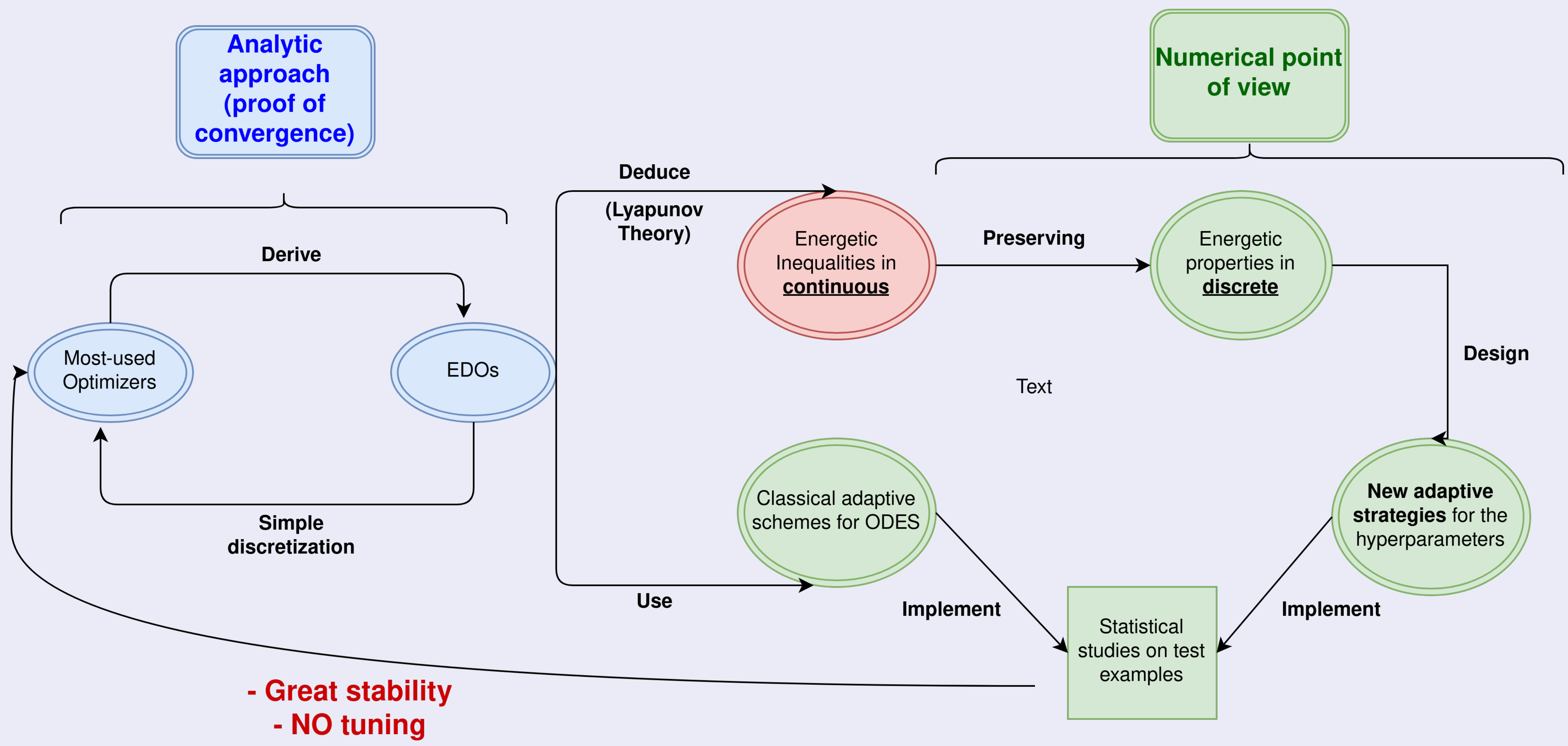
$$\boxed{\mathcal{E}_{\theta^*} = \{\theta_0, \|\mathcal{A}(\theta_0) - \theta^*\| < \epsilon\}}$$



(a) Gradient Descent (b) Adam without bias steps

Sensitivity to the initial conditions

Mathematical Approach



Results

- Maximum and saddle points **instabilities** in continuous-time;
- **Convergence/Moderate sensitivity to the initializer without tuning** :

(a) Adaptive Gradient Descent (b) Adaptive Momentum

Efficiency of the new schemes

References

[1] G. Kluth & al. Deep learning for nlte spectral opacities. *Physics of Plasmas*,27(5) :052707, 2020

[2] Paul Novello, Gael Poette & al. Explainable Hyperparameters Optimization using Hilbert-Schmidt Independence Criterion.June 2021.