A Mean-Field Optimal Control Approach to Deep Learning

- **Benoît Bonnet** (LAAS - CNRS, Toulouse, France)
- Cristina Cipriani (Technical University Munich, Department of Mathematics, Munich, Germany)
- Massimo Fornasier (Technical University Munich, Department of Mathematics, Munich, Germany)
- Hui Huang (University of Calgary, Department of Mathematics and Statistics, Calgary, Canada)

Keywords: Deep learning, mean-field optimal control, necessary optimality conditions.

Abstract: In this talk, I will discuss some of the results of [3] in which we propose a reformulation of a certain class of deep learning problems as optimal control problems in the space of probability measures. The roots of our work trace back to the pioneering articles of Weinan E and co-authors [6, 7]. The incentive for developing such a point of view is twofold. Firstly, the latter builds on the so-called residual block regularisation of neural networks proposed in [8], which has since been known to improve the stability thereof as the number of layers increases. Secondly, embedding residual networks into continuous-time dynamical systems grants access to the broad literature of mathematical control theory, with the help of which one may hope to improve the overall explainability of learning algorithms (see e.g. the works [1, 9, 10, 11] in this direction).

After exposing the conceptual path leading to the reformulation of deep residual learning procedures as mean-field optimal optimal control problems, I shall present a general family of first-order optimality conditions that we derived for this class of problem, and show that the latter can be established by following either of two possible paths. On the one hand, one can derive such optimality extrinsically as a consequence of an abstract Lagrange multiplier rule in the Banach space of Radon measures, in which the subspace of probability measures appears as a convex constraint set. On the other hand, one may also adopt the intrinsic viewpoint developed in [2, 4, 5], where a mean-field counterpart of the classical Pontryagin Maximum Principle – involving the existence of a state-costate pair solution of an Hamiltonian flow in Wasserstein spaces – is derived. This general discussion on optimality conditions will be supplemented by simple numerical illustrations on toy-ish classification problems in low dimension.

References:

[1] A. Agrachev and A. Sarychev. Control on the Manifolds of Mappings as a Setting for Deep Learning. *arXiv preprint arXiv:2008.12702*, 2021.

[2] B. Bonnet. A Pontryagin Maximum Principle in Wasserstein Spaces for Constrained Optimal Control Problems. *ESAIM COCV* 25(52), 2019

[3] B. Bonnet, C. Cipriani, M. Fornasier and H. Huang. A Measure Theoretical Approach to the Mean-field Maximum Principle for Training NeurODEs. *arxiv preprint arXiv:2107.08707*, 2022.

[4] B. Bonnet and H. Frankowska. Necessary Optimality Conditions for Optimal Control Problems in Wasserstein Spaces. *Applied Mathematics and Optimization*, 84:1281–1330, 2021.

[5] B. Bonnet and F. Rossi. The Pontryagin Maximum Principle in the Wasserstein Space. *Calculus of Variations and Partial Differential Equations*, 58:11, 2019.

[6] W. E. A Proposal on Machine Learning via Dynamical Systems. *Communication in Mathematics and Statistics*, 1(5):1-11, 2017.

[7] W. E, J. Han and Q. Li. A Mean-Field Optimal Control Formulation of Deep Learning. *Research in the Mathematical Sciences*, 6(10), 2019.

[8] K. He, X. Zhang, S. Ren and S. Jian. Deep Residual Learning for Image Recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 770-778, 2016.

[9] J.-F. Jabir, D. Šiška and L. Szpruch. Mean-Field Neural ODEs via Relaxed Optimal Control. *arxiv preprint arXiv:1912.05475*, 2021.

[10] A. Scagliotti. Deep Learning Approximation of Diffeomorphisms via Linear Control Systems. *arxiv preprint arXiv:2110.12393*, 2022.

[11] P. Tabuada and B. Gharesifard. Universal Approximation Power of Deep Residual Networks via Nonlinear Control. *arXiv preprint arXiv:2007.06007*, 2020.