PhD fellowship: Optimization aspects in discretized neural differential equations

Dauphine | PSL
UNIVERSITÉ PARIS

A three-year PhD fellowship is available at Université Paris Dauphine-PSL (Paris, France), under the supervision of Antonin Chambolle (senior researcher, main supervisor) and Clément Royer (associate professor, co-supervisor). The position will be funded through the PEPR *(Programmes et Équipements Prioritaires de Recherche)* on Partial Differential Equations and Artificial Intelligence. The position is expected to start by October 1st, 2024.

Applicants should hold a Masters degree with a strong component in applied mathematics, including optimization and partial differential equations. Programming experience using Python, Matlab or Julia is strongly recommended. A good level in English is mandatory, but knowledge of French is not required.

Interested candidates should send a CV to: `clement.royer@lamsade.dauphine.fr`. Any inquiries regarding this position can be sent to the same email address. By recruiting on behalf of Université Paris-Dauphine, the co-advisors A. Chambolle and C. Royer commit to providing equal employment opportunities to all qualified applicants.

Context Neural differential equations are a recently proposed learning model in which a neural network is implicitly defined through the solution of a differential equation. Such a model highlights connections between residual architectures and ordinary differential equations (ODEs) on one hand, and between convolutional layers and partial differential equations (PDEs) on the other hand [5, 9]. In practice, those networks are instantiated using discretization techniques, which poses a number of challenges pertaining to optimization.

First, training such neural differential architectures amounts to solving a constrained optimization problem where the constraints are expressed as differential equations. Such problems bear a strong connection with those arising in scientific computing, where one typically minimizes an objective function (such as the error between a model and observations) under physical constraints represented by ODEs or PDEs [1]. Secondly, implementing neural differential architectures requires to set multiple hyperparameters (different from the parameters of the network that are learned through training). For general neural networks, those arise from the training procedures, that often involve algorithmic hyperparameters with significant impact on the performance. In the case of neural differential equations, additional hyperparameters arise from the discretization operator used to implement the PDE or ODE calculations. Tuning both sets of hyperparameters represents a modern challenge towards automated use of neural differential architectures, that can be posed as an optimization problem where testing a particular hyperparameter configuration is an expensive procedure. Such a paradigm corresponds to that of blackbox optimization, which has been used extensively in scientific computing to calibrate complex numerical models, including numerical discretization tools [2].

Thesis description This thesis aims at proposing efficient optimization techniques for the training and calibration of neural differential architectures. The first axis of the thesis will focus on the training problem. In order to exploit the special structure of discretized differential equations, we will rely on algorithmic techniques similar to those extensively used in PDE-constrained optimization and scientific computing [1]. By leveraging existing work on such algorithms in nonconvex optimization [6], we expect to develop methods that adapt to the various discretized architectures used in the literature [7], with both theoretical guarantees and practical appeal. Our approach will be validated on data science and imaging problems where standard training procedures can be improved by using a neural differential equation model [8].

The second axis of the thesis will investigate hyperparameter tuning of discretized neural architectures. In a departure from existing techniques, we will consider the calibration of both the numerical tools used to discretize the differential equation at hand, and that of the training algorithms. Building on recent advances in the field of derivative-free optimization [3], we will design efficient blackbox optimization techniques supported by theoretical guarantees and demonstrated practical performance. Although calibrating neural architectures will be at the heart of our project, the thesis could also consider solving similar hyperparameter tasks in the context of imaging problems, where discretization of regularizers also arises [4].

References

[1] H. Antil, D. P. Khouri, M.-D. Lacasse, and D. Ridzal, editors. *Frontiers in PDE-Constrained Optimization*, volume 163 of *The IMA Volumes in Mathematics and its Applications*. Springer, New York, NY, USA, 2016.

[2] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, 2017.

[3] E. Bergou, Y. Diouane, V. Kungurtsev, and C. W. Royer. A stochastic Levenberg-Marquardt method for using random models with complexity results and application to data assimilation. *SIAM/ASA J. Uncertain. Quantif.*, 10:507–536, 2022.

[4] A. Chambolle and T. Pock. Learning consistent discretizations of the total variation. *SIAM J. Imaging Sci.*, 14:778–813, 2021.

[5] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems 31*, pages 6571–6583, 2018.

[6] F. E. Curtis, D. P. Robinson, C. W. Royer, and S. J. Wright. Trust-region Newton-CG with strong second-order complexity guarantees for nonconvex optimization. *SIAM J. Optim.*, 31:518–544, 2021.

[7] Y. Lu, A. Zhong, Q. Li, and B. Dong. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 3276–3285. PMLR, 2018.

[8] E. Newman, L. Ruthotto, J. Hart, and B. van Bloemen Waanders. Train like a (var)pro: Efficient training of neural networks by variable projection. *SIAM J. Math. Data Sci.*, 3:1041–1066, 2021.

[9] L. Ruthotto and E. Haber. Deep neural networks motivated by partial differential equations. *J. Math. Imaging Vision*, 62:352–364, 2020.