Interplay between
stochastic algorithms and automatic differentiation

Edouard Pauwels (Toulouse School of Economics)
Samuel Vaiter (CNRS et Université Côte d'Azur, Laboratoire J.A. Dieudonné, Nice)

Internship proposal (Spring 2024, 5 to 6 months)

Context. Algorithmic differentiation [8] is one of the workhoses of modern AI/learning architectures [6], typically known as "backpropagation". For iterative algorithms, in a parametric optimization context algorithmic differentiation allows to propagate jointly iterates and their derivative with respect to external parameters. The asymptotics of this process was studied in [5, 1], in relation with implicit differentiation, see also [7]. In modern machine learning era, this has received a regain of interest in the context of differentatiation of gradient type algorithms with implicit differentiation [11], its connection with asymptotics of forward derivative propagation [9], quantitative rate estimates [10] and extension to nonsmooth optimization [2].

Goals of the internship. The goal of the internship is to understand the interplay between automatic differentiation and stochastics methods. In particular, we consider the two following directions.

1. In a parametric optimization context, asymptotic analysis of derivatives produced by iterative processes is limited to some form of contractivity of the algorithmic map [5, 2]. Most empirical risk minimization usually involve a large sum. Common algorithms in this setting, such as stochastic gradient descent (SGD), involve random subsampling approximations to handle this large sum aspect [4]. But this typically does not induce the required contractivity property. We aim at studying convergence of iterate derivatives of such algorithms, starting with SGD in a convex setting. The main techniques and background will consist of classical convergence arguments for deterministic and stochastic methods, as well as convergence of the derivatives of deterministic gradient type algorithms. The question is mostly of theoretical interest and we expect the outcomes of the internshipt to consist essentially of abstract convergence results for derivatives of stochastic approximation algorithms.

2. We would like to improve the mathematics foundations of the differentiation of Monte Carlo-based methods, arising either in probabilistic programming or in differentiable rendering. Under adequate hypotheses [3], it is possible to show interchange result for (generalized) derivatives and integrals, and more generally, the well-defined aspect of set-valued integrals. Nevertheless, even under well-specified integrals, it is not clear that it is possible to differentiate Monte Carlo sampling strategies directly. We propose to address it in two steps, first by proving a nonsmooth Reynold's theorem and then harness the structure (typically a stratification) to propose differentiable Monte Carlo strategies.

Candidate profile. We seek a second year Master student in applied mathematics or computer science. A prior background in optimisation is welcome, along with a taste for numerical simulations.
Location. The supervision of the internship is shared between Nice and Toulouse, the internship will mostly take place at Toulouse School of Economics.
Application procedure. Send your CV and a copy of your last year transcripts to both Edouard Pauwels (`edouard.pauwels@tse-fr.eu`) and Samuel Vaiter (`samuel.vaiter@cnrs.fr`).

References

[1] Beck, T. (1994). Automatic differentiation of iterative processes. Journal of Computational and Applied Mathematics.
[2] Bolte, J., Pauwels, E., Vaiter, S. (2022). Automatic differentiation of nonsmooth iterative algorithms. NeurIPS.
[3] Bolte, J., Le, T., Pauwels, E. (2023). Subgradient sampling for nonsmooth nonconvex minimization. SIAM J. Optim.
[4] Bottou, L., Curtis, F. E., Nocedal, J. (2018). Optimization methods for large-scale machine learning. Siam Review.
[5] Gilbert, J. C. (1992). Automatic differentiation and iterative processes. Optimization Methods and Software.
[6] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep learning.
[7] Griewank, A., Faure, C. Piggyback differentiation and optimization (2003). In Large- scale PDE-constrained optimization. Springer.
[8] Griewank, A., Walther, A. (2008). Evaluating derivatives: principles and techniques of algorithmic differentiation. SIAM.
[9] Lorraine, J., Vicol, P., Duvenaud, D. (2020). Optimizing millions of hyperparameters by implicit differentiation. AISTATS.
[10] Mehmood, S., Ochs, P. (2020). Automatic differentiation of some first-order methods in parametric optimization. AISTATS.
[11] Pedregosa, F. (2016). Hyperparameter optimization with approximate gradient. ICML.