Losing momentum in continuous-time stochastic optimisation

Jonas Latz, Kexin Jin, Chenguang Liu, Alessandro Scagliotti

Abstract

The training of modern machine learning models often consists in solving high-dimensional non-convex optimisation problems that are subject to large-scale data. In this context, momentum-based stochastic optimisation algorithms have become particularly widespread. The stochasticity arises from data subsampling which reduces computational cost. Both, momentum and stochasticity help the algorithm to converge globally. In this work, we propose and analyse a continuous-time model for stochastic gradient descent with momentum. This model is a piecewise-deterministic Markov process that represents the optimiser by an underdamped dynamical system and the data subsampling through a stochastic switching. We investigate longtime limits, the subsampling-to-no-subsampling limit, and the momentum-to-no-momentum limit. We are particularly interested in the case of reducing the momentum over time. Under convexity assumptions, we show convergence of our dynamical system to the global minimiser when reducing momentum over time and letting the subsampling rate go to infinity. We then propose a stable, symplectic discretisation scheme to construct an algorithm from our continuous-time dynamical system. In experiments, we study our scheme in convex and non-convex test problems. Additionally, we train a convolutional neural network in an image classification problem. Our algorithm attains competitive results compared to stochastic gradient descent with momentum.

References

[1] H. Attouch, X. Goudou, P. Redont (2000) *The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system.* Communications in Contemporary Mathematics, 2(01):1–34.

[2] K. Jin, J. Latz, C. Liu, A. Scagliotti (2022) *Losing momentum in continuous-time stochastic optimisation*, arXiv:2209.03705.

[3] J. Latz (2021) *Analysis of stochastic gradient descent in continuous time*, Statistics and Computing 31, 39.