Losing momentum in continuous-time stochastic optimisation

Jonas Latz
University of Manchester
`jonas.latz@manchester.ac.uk`

Coauthor(s): Kexin Jin, Chenguang Liu, Alessandro Scagliotti

The training of deep neural networks and other modern machine learning models usually consists in solving non-convex optimisation problems that are high-dimensional and subject to large-scale data. Here, momentum-based stochastic optimisation algorithms have become especially popular in recent years. The stochasticity arises from data subsampling which reduces computational cost. Moreover, both, momentum and stochasticity are supposed to help the algorithm to overcome local minimisers and, hopefully, converge globally. Theoretically, this combination of stochasticity and momentum is badly understood.

In our work [1], 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 particle movement by an underdamped dynamical system and the data subsampling through a stochastic switching of the dynamical system. In our analysis, we investigate long-time 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: intuitively, the momentum helps to overcome local minimisers in the initial phase of the algorithm, but prohibits fast convergence to a global minimiser later. Under convexity assumptions, we show convergence of our dynamical system to the global minimiser when reducing momentum over time and let 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 numerical experiments, we study our discretisation scheme in convex and non-convex test problems. Additionally, we train a convolutional neural network to solve the CIFAR-10 image classification problem. Here, our algorithm reaches competitive results compared to stochastic gradient descent with momentum.

[1] Jin, Kexin, Jonas Latz, Chenguang Liu, & Alessandro Scagliotti (2022). Losing momentum in continuous-time stochastic optimisation. arXiv:2209.03705.