High-dimensional Stochastic Gradient Descent: effective dynamics and critical scaling

Gérard Ben Arous

Courant Institute of Mathematical Sciences, New York University, U.S.A

benarous@cims.nyu.edu

I will survey recent joint work with Reza Gheissari (Northwestern) and Aukosh Jagannath (Waterloo) and upcoming work with the same authors and with Jiaoyang Huang (Wharton, University of Pennsylvania).

The trajectories of SGD (properly rescaled) typically converge to the gradient flow of the population loss, in finite dimensions. How can one efficiently state such a theorem when the dimension is diverging?

We prove limit theorems for the trajectories of summary statistics (i.e., finite-dimensional functions) of SGD as the dimension goes to infinity. Our approach allows one to choose the summary statistics that are tracked, the initialization, and the step-size. It yields both ballistic (ODE) and diffusive (SDE) limits, with the limit depending dramatically on the former choices.

Interestingly, we find a critical scaling regime for the step-size below which the effective ballistic dynamics matches gradient flow for the population loss, but at which, a new correction term appears which changes the phase diagram.

About the fixed points of this effective dynamics, the corresponding diffusive limits can be quite complex and even degenerate. We demonstrate our approach on popular examples including estimation for spiked matrix and tensor models and classification via two-layer networks for binary and XOR-type Gaussian mixture models.

These examples exhibit surprising phenomena for these limiting dynamical systems, including multimodal timescales to convergence as well as convergence to sub-optimal solutions with probability bounded away from zero from random (e.g.,Gaussian) initializations.

The next open question are then: how does one find relevant summary statistics for less explicit models than these examples? And how are we sure that the induced effective dynamics perform well?

Joint work with Aukosh Jagannath (University of Waterloo, Canada), Reza Gheissari (Northwestern University, U.S.A) and Jiaoyang Huang (Wharton School, University of Pennsylvania, U.S.A).