Gradient flows, random perturbations, and large deviations

Panayotis Mertikopoulos
Department of Mathematics, NKUA
pmertik@math.uoa.gr

Abstract

Gradient flows are among the simplest and most fundamental dynamical systems in continuous time. Their discretization gives rise to *gradient descent* — the most widely studied method in continuous optimization — and when gradients are subject to noise and randomness, ***stochastic*** *gradient descent* (SGD) — the workhorse of artificial neural networks and modern machine learning architectures.

Topologically, the long-run behavior of a gradient flow is very simple to describe under mild conditions, every trajectory converges to a component of critical points. But which one? And what happens in the stochastic case? Which critical points are more likely to be observed in the long run, and by how much? And how long does it take for SGD to reach the vicinity of a given critical point (e.g., the function's global minimum)?

I will describe how little I know about these questions, and I will outline an approach yielding some partial answers based on the theory of large deviations and randomly perturbed dynamical systems. This talk is otherwise intended as a "work-in-progress" call for ideas, input, and lively discussions.