

## **From Stochastic to Deterministic: Stochastic Gradient Descent in Non-convex and high Dimensional Problems**

Stochastic gradient descent (SGD) stands as a cornerstone of optimization and modern machine learning. However, a complete understanding of why SGD performs so well remains a major challenge. In this talk, I will present a theory for SGD in high dimensions when both the number of samples and problem dimensions are large. We show that the dynamics of one pass SGD applied to generalized linear models and multi-index problems with data possessing a general covariance matrix become deterministic in the large sample and dimensional limit. In particular, the limiting dynamics are governed by a set of low-dimensional ordinary differential equations (ODEs). Our setup encompasses a wide spectrum of optimization problems, including linear regression, logistic regression, and two-layer neural networks. In addition, it unveils the implicit bias inherent in SGD. For each of these problems, the deterministic equivalent of SGD enables us to derive a close approximation of the statistical risk (with explicit and vanishing error bounds). Furthermore, we leverage our theorem to establish explicit conditions on the step size, ensuring the convergence and stability of SGD within high-dimensional settings.

This is a joint work with E. Collins-Woodfin, C. Paquette, and E. Paquette, for more details see <https://arxiv.org/abs/2308.08977>.