Implicit regularization for constant step-size SGD: why it works for non-smooth, non-convex, low-rank models.

Despite the development of many new optimization techniques, constant step-size SGD remains the tool of choice in modern machine learning. It performs unreasonably well in practice, both in convex settings where there are poor guarantees, and also in non-convex settings, where there are often no guarantees. In this talk, we will give an argument as to why it works for low-rank models, illustrating the argument for the particular case of phase retrieval.

Phase retrieval is the problem of solving systems of rank-1 quadratic equations, and can be solved by optimizing a non-convex, non-smooth objective. In order to obtain guarantees, however, a spectral initialization method is first used in order to get a constant error estimate. We will show how SGD works even from arbitrary initialization. The key insight is that the low-rank nature of the signal allows us to study the dynamics as a Markov process with a two-dimensional phase space, and in this phase space, the effective step size decreases as the dimension increases. Our analysis also introduces and makes use of several new ideas from stochastic process theory, which we believe will be of use much more broadly. (Received January 27, 2019)