Oracle Complexity in Nonsmooth Nonconvex Optimization

Guy Kornowski Ohad Shamir
Weizmann Institute of Science
{guy.kornowski,ohad.shamir}@weizmann.ac.il

October 9, 2021

Abstract

It is well-known that given a smooth, bounded-from-below, and possibly nonconvex function, standard gradient-based methods can find ϵ-stationary points (with gradient norm less than ϵ) in $\mathcal{O}(1/\epsilon^2)$ iterations. However, many important nonconvex optimization problems, such as those associated with training modern neural networks, are inherently not smooth, making these results inapplicable. In this paper, we study nonsmooth nonconvex optimization from an oracle complexity viewpoint, where the algorithm is assumed to be given access only to local information about the function at various points. We provide two main results: First, we consider the problem of getting *near* ϵ-stationary points. This is perhaps the most natural relaxation of *finding* ϵ-stationary points, which is impossible in the nonsmooth nonconvex case. We prove that this relaxed goal cannot be achieved efficiently, for any distance and ϵ smaller than some constants. Our second result deals with the possibility of tackling nonsmooth nonconvex optimization by reduction to smooth optimization: Namely, applying smooth optimization methods on a smooth approximation of the objective function. For this approach, we prove an inherent trade-off between oracle complexity and smoothness: On the one hand, smoothing a nonsmooth nonconvex function can be done very efficiently (e.g., by randomized smoothing), but with dimension-dependent factors in the smoothness parameter, which can strongly affect iteration complexity when plugging into standard smooth optimization methods. On the other hand, these dimension factors can be eliminated with suitable smoothing methods, but only by making the oracle complexity of the smoothing process exponentially large.