Accelerated Random Search

Sebastian U. Stich
CORE/INMA
Université catholique de Louvain, Belgium
sebastian.stich@uclouvain.be

We study randomized accelerated methods for convex optimization, i.e. schemes where the search directions are not the negative gradients, but samples from a discrete or continuous probability distribution.

Whilst the classical Gradient descent converges at rate $O(k^{-1})$, it is well-known that optimal schemes converge at rate $O(k^{-2})$. In 1983, Nesterov proposed a framework that allows to construct optimal schemes by means of approximations, so called estimate sequences. Recently, Nesterov (2011,2012) showed that the same ideas can be used to accelerate randomized schemes, where either the search directions are sampled from an isotropic normal distribution or the set of coordinate directions. Lee and Sidford (2013) extended the estimate sequence framework to this setting, but they restricted themselves to distributions over coordinate directions. To achieve the optimal convergence rate, the choice of the step sizes is crucial. The two aforementioned schemes therefore require access to either directional derivatives or to gradient oracles (random variables whose expectation equals the gradient of the objective function at the query point).

In this work, we study the impact of the step sizes on the convergence rate. The analysis is not limited to discrete or isotropic distributions, but also applies to certain non-isotropic continuous distributions. We consider (i) inexact gradient oracles and (ii) (gradient-free) inexact line search oracles. These modified schemes can still accelerate, but they do not necessarily attain the optimal rate, that is, they converge at rate $O(k^{-\alpha})$, for $1 \leq \alpha \leq 2$. These modifications are motivated by two concrete examples: (i) high-dimensional optimization, where directional derivatives might only be computed approximately for efficiency reasons, and (ii) derivative-free optimization, where no gradient information is available but only function values.