UDC 519.853.62

Optimization with Markovian Noise

B. Prokhorov, A. Beznosikov
Moscow Institute of Physics and Technology

May 17, 2024

Abstract

In this work we study lower bounds for optimal convergence rates in stochastic optimization with smooth strongly convex objective function and first-order markovian oracle. In these settings we provided lower bounds matching previously known upper bounds for a wide family of target functionals. Optimal rates remain unknown only for the special case of small noise.

1 Problem Statement and Previous Work

We study the minimization problem

$$\min_{x \in \mathbb{R}^d} f(x) := \mathbb{E}_{Z \sim \pi}[F(x, Z)]$$

where the access to the function f and its gradient is available only through the (unbiased) noisy oracle $F(x, Z)$ and $\nabla F(x, Z)$, respectively.

Our research aims to expand the results of [1] by providing tight lower bounds. Following the article, we consider additional assumptions:

- (A1,A2) f is (μ, L)-smooth.
- (A3) Z is uniformly geometrically ergodic with mixing time τ.
- (A4) F is (σ, δ)-bounded:

$$\|\nabla F(x, Z) - \nabla f(x)\|^2 < \sigma^2 + \delta^2 \|\nabla f(x)\|^2$$

The main result of [1] is the following

Theorem. Under A1-A4 the problem can be solved (in terms of $\mathbb{E}\left[\|x^{(k)} - x^*\|^2\right] \leq \varepsilon$) in

$$\tilde{\mathcal{O}}\left(\tau\left[(1+\delta^2)\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon} + \frac{\sigma^2}{\mu^2\varepsilon}\right]\right) \quad \text{oracle calls}\,. \tag{1}$$

And the bound is tight for $\delta = 1$.

2 Results

The main result of our work is the

Theorem 1. For any $\delta \geq 1$ there is a problem instance f and markov chain Z satisfying A1-A4 such that any first order method needs at least

$$\tilde{\mathcal{O}}\left(\tau\left[(1+\delta^2)\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}+\frac{\sigma^2}{\mu^2\varepsilon}\right]\right) \quad \text{oracle calls} \tag{2}$$

in order to achieve $\mathbb{E}\left[\|x^{(k)}-x^*\|^2\right] \leq \varepsilon$.

Thus we obtain optimal convergence rates for $\delta \geq 1$. For $\delta < 1$ provided rate may be not tight if τ is big enough, since markovian noise may be treated as adversarial. In 1983 Polyak shown in his book [3], simple gradient descent converges linearly if $\delta < 1$.

Optimal rates when $\delta < 1$ with adversarial noise are known only for $\delta \lesssim \sqrt{\frac{\mu}{L}}$ (see [2]). And it is furthermore unknown if acceleration is possible for markovian noise.

Another interesting regime is when $\delta \to 1^-$ for adversarial (or markovian) noise. As previously mentioned, it can be shown [3] that oracle complexity for gradient descent is of order

$$\tilde{\mathcal{O}}\left(\frac{1}{1-\delta}\frac{L}{\mu}\log\frac{1}{\varepsilon}\right) \tag{3}$$

We also establish the lower bound for this regime with the same dependence on δ.

Theorem 2. For any $\delta < 1$ there is a problem instance f such that any first order method needs at least

$$\tilde{\mathcal{O}}\left(\frac{1}{1-\delta}\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}\right) \tag{4}$$

in order to achieve $\left[\|x^{(k)}-x^*\|^2\right] \leq \varepsilon$ with adversarial oracle.

3 References

[1] Aleksandr Beznosikov et al. *First Order Methods with Markovian Noise: from Acceleration to Variational Inequalities*. 2023. arXiv: `2305.15938 [math.OC]`.

[2] Nikita Kornilov et al. *Intermediate Gradient Methods with Relative Inexactness*. 2023. arXiv: `2310.00506 [math.OC]`.

[3] Boris Polyak. *Introduction to Optimization*. July 2020.