Accelerated Minibatch Three Point Method

Kutenkov Roman, Kuznetsov Ivan

Scientific advisor: Solodkin Vladimir

March 11, 2025

Table of contents

Problem statement

Problem

We consider unconstrained minimization problem

$$\min_{x \in \mathbb{R}^d} f(x)$$

with smooth target function, that is bounded from below. Possible settings are: non-convex, convex and λ-strongly convex.

DFO methods

DFO stands for **D**erivative **F**ree **O**ptimization, in this context derivatives of the object function are inaccessible. It may be due to gradient being either noisy, impractical to evaluate or unavaliable.

Related work

Stochastic Three Points (STP) method

Our algorithm is based on STP method, proposed in [1], which requires distribution $\mathcal{D}$ and stepsizes $\{\alpha_k\}_{k=0}^{\infty}$. At k-th iteration of STP a random direction s^k is sampled from $\mathcal{D}$ and then the next iterate is given by $x^{k+1} = \arg\min\{f(x^k), f(x^k \pm \alpha_k s^k)\}$.

Stochastic Momentum Three Points (SMTP) method

SMTP is a modification of STP, that introduces momentum to each iteration, it is described in [2]. Iterate of this method is more complicated, however in essence it follows the same principle of moving along random direction if it reduces objective's value, but now with momentum as additional parameter added to mix.

Related work

Both papers provide proofs of convergence in non-convex, convex and λ-strongly convex cases for specifically selected stepsizes. Convergence is understood in terms of mathematical expectation $\mathbb{E}$ because of the random nature of algorithms. To give you an idea of what it looks like here is the result for λ-strongly convex case:

Let $\varepsilon > 0$ and $\alpha_k = \frac{f(x^k+\alpha_k s^k)-f(x^k)}{tL}$, where $0 < t \leq \sqrt{\frac{4\lambda\mu_{\mathcal{D}}^2}{L^2}\varepsilon}$.
Then $\mathbb{E}(f(x^K) - f(x^*)) \leq \varepsilon$ after $K = \frac{1}{\mu_{\mathcal{D}}^2}\frac{L}{\lambda}\log\left[\frac{2(f(x^0)-f(x^*))}{\varepsilon}\right]$ iterations of STP.

There $\mu_{\mathcal{D}}$ is a constant related to distribution $\mathcal{D}$. For SMTP there is an analogous theorem, but with additional momentum parameter β.

Motivation

1. Heavy ball outperforms GD, yet there no proofs
2. Nesterov acceleration outperforms heavy ball and is proven
3. Introducing linear coupling allows for potential improvment

Plans and expectations

Linear coupling

Linear coupling introduces additional parameters, depending on their choice proposed algorithm could outperform SMTP, that outperforms STP in practice.

Experiments

So far experiments were conducted only on quadratic L-smooth and λ-strongly convex function. Presumably due to simplicity of the problem iterates of proposed algorithm mostly coincide with those of STP. Maybe some modification is possible to fix this.

Additional randomness

There is a potential for introducing randomness in choices of algorithm that may give additional results.

Bibliography

- [1] El Bergou, Eduard Gorbunov, and Peter Richtarik – *stochastic three points method for unconstrained smooth minimization*
- [2] El Bergou, Eduard Gorbunov, Peter Richtarik, Adel Bibi, Ozan Sener – *a stochastic derivative free optimization method with momentum*