Report 1

Roman Voronov, Nursultan Abdullaev

1 Problem statement and its importance

Problem Statement:
The paper addresses optimization problems represented as $F(x) = f(x) + R(x)$, where f is a smooth, strongly convex function, and R is a closed convex regularizer. Traditional methods, like subspace descent, often struggle with these problems, especially when they involve non-separable proximal terms. Additionally, while stochastic gradient-type methods have gained traction, there's a need for novel approaches that use different stochastic gradient information.
Importance:
SEGA, as introduced in this paper, offers a pivotal advancement in optimization techniques. It not only provides unbiased gradient estimators , enhancing the accuracy of optimization, but also showcases versatility by accommodating general sketches from various distributions. This flexibility positions SEGA as a potentially transformative tool in optimization tasks.

2 Examples where the problem occurs

- **Coordinate Descent in Complex Optimization Problems:** Traditional optimization methods like coordinate descent demonstrate proficiency in many contexts. However, they often stumble when confronted with optimization problems that encompass non-separable proximal terms. These terms introduce intricacies that make the problems harder to solve using standard techniques. For example, in problems where the objective function includes terms that span multiple variables simultaneously (non-separable terms), standard subspace descent methods might not suffice.
- **Challenges with Specific Regularizers:** In various real-world applications, the regularizer R can take the form of an indicator function of a convex set or a sparsity-inducing penalty like the l_1-norm. Such regularizers are pivotal in scenarios where model interpretability is crucial or where computational efficiency is paramount. However, they also introduce non-smooth elements to the optimization problem. This non-smoothness can make the optimization landscape rugged and challenging for many algorithms to navigate.
- **Randomized Coordinate Descent's Limitations:** The randomized coordinate descent method, a variant of the coordinate descent, employs a specific type of oracle. While this method is powerful and efficient in various settings, it faces challenges when the oracle corresponds to a distribution over standard basis vectors. Such a distribution might not offer the flexibility or granularity needed to capture intricate details of certain optimization landscapes.

In essence, while traditional optimization techniques have their strengths, they also possess inherent limitations in specific scenarios. The paper underscores these gaps and introduces SEGA as a versatile and robust solution to address them.

3 Approach of authors

The approach to solve the optimization problem is following:

1. Get the gradient sketch for the current point. The transformation is sampled from D

2. Update the estimate of the gradient with accordance to gradient sketch using sketch-and-project process. **Note:** the estimation we have is **biased**

3. Make a unbiased gradient estimation from biased by introducing some random variable.

4. Perform proximal step with respect to regularizer

As authors have proven in their paper, as x^k approaches optimum, biased and unbiased estimations of gradient at x^k become better at approximating the true gradient at x^k. Therefore, the variance of unbiased gradient estimation becomes zero as x^k approaches zero. This is the reason why this name is taken for the algorithm

4 Works the approach of authors is based on

- **JackSketch algorithm** by *Robert M Gower, Peter Richtarik, and Francis Bach.*

  The structure of SEGA is inspired by this algorithm

- **Randomized iterative methods for linear systems** by *Robert M Gower and Peter Richtarik.* Here the sketch-and-project method is introduced.

- **Stochastic gradient-type methods** like SGD, SAGA, SAG, SVRG and other.

  SEGA improves those algorithms, so latter are the base algorithms for the former one.

5 Improvements with the basic versions

- While other stochastic gradient-type methods work only with objectives expressed as finite sums or expectations, SEGA works with any objective function

- SEGA does not work directly with unbiased gradient estimator. SEGA works with gradient sketches that are not assumed to be unbiased before additional steps. As a result, SEGA has an ability to work with any regularizer in objective function, unlike other methods that work with separable regularizers.

- SEGA based on linear transformations of true gradient. Those linear transformations are taken from some predefined distribution D. In other stochastic gradient-type methods, we cannot define distribution by ourself, they are predefined by algorithms. For example, D of randomized coordinate distance is predefines and corresponds to distribution over standart basis vectors. It means, that information in SEGA can be compatible with the information in other methods for some specific D.