# Solving nonsmooth decentralized optimization with affine constraints using gradient sliding

Jiawei Chen[1,2], Zhenzhen Song[1], Alexander Rogozin[1,2], Nhat Trung Nguyen[2], Demyan Yarmoshik[2], Irina Podlipnova[2], Alexander Gasnikov[2]

## Introduction

This paper explores decentralized nonsmooth convex optimization with affine constraints. We extend existing research by incorporating a nonsmooth stochastic oracle, solved by the well know gradient sliding method. Our result show sliding algorithm achieves sub-optimal solution for these optimization problems under certain conditions, addressing limitations of prior methods. This work enhances the theoretical understanding of distributed optimization and offers practical solutions for applications in sensor networks and machine learning.

## Porblem Formulation

We consider following optimization problem:

$$\min_{x\in\mathcal{X}\subset\mathbb{R}^d} \frac{1}{m}\sum_{i=1}^{m} f_i(x) \quad \text{s.t.} \quad Bx = 0. \tag{1}$$

We assume that

1. $f_i$ has a unbiased stochastic subgradient $f_i^{'}(x,\xi)$ with bounded variance $\sigma^2$.
2. $f_i$ is convex.

## Connected Networks

The graph Laplacian matrix $W \in \mathbb{R}^{m\times m}$ of connected networks:

$$[W]_{ij} = \begin{cases} -1, & \text{if } (i,j)\in E, \\ \deg(i), & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

where $\deg(i)$ is the degree of the $i$-th node.

## Solution to the problem

**1. Composite optimization:**

$$\min_{x\in\mathcal{Q}}\{\phi(x) = g(x) + f(x)\}, \tag{3}$$

where $f(x)$ is a nonsmooth convex function and $g(x)$ is a smooth convex function.

**2. Convex optimization with two affine constraints:**

We introduce a minimization problem with two affine constraints :

$$\min_{\mathbf{x}\in\mathcal{Q}} \mathbf{F}(\mathbf{x}) s.t. \quad \mathbf{B}\mathbf{x} = 0, \mathbf{C}\mathbf{x} = 0, \tag{4}$$

where $F(\mathbf{x}) = \frac{1}{m}\sum_{i=1}^{m} f_i(x_i)$ and $\mathbf{x} = \mathrm{col}(x_1,\ldots,x_m)$, and also introduce $\mathbf{B} = B\otimes\mathbf{I}_m, \mathbf{C} = C\otimes\mathbf{I}_d$, $B\in\mathbb{R}^{p\times d}$, $C\in\mathbb{R}^{m\times m}$. Then we introduce this problem:

$$\tilde{\mathbf{F}}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) + \frac{R_{\mathbf{y}}^2}{\varepsilon}\|\mathbf{A}\mathbf{x}\|^2, \tag{5}$$

**3. Decentralized gradient sliding method**

$$\min_{\substack{x_1=\cdots=x_m \\ x_1,x_2,\cdots,x_m\in\mathcal{X}}} f(\mathbf{x}) = \frac{1}{m}\sum_{i=1}^{m} f_i(x_i) s.t. \quad \mathbf{B}\mathbf{x} = 0. \tag{6}$$

## Convergence analysis

Gradient sliding algorithm requires

$$\widetilde{O}\left(\frac{(M^2+\sigma^2)D_{\mathcal{X}}^2}{\varepsilon^2}\right) \text{ calculations of } f'(x,\xi) \text{ per node.} \tag{7}$$

and

$$O\left(\sqrt{\frac{\chi^2(W)MD_{\mathcal{X}}^2}{\varepsilon^2}}\right) \text{ communications,} \tag{8}$$

and

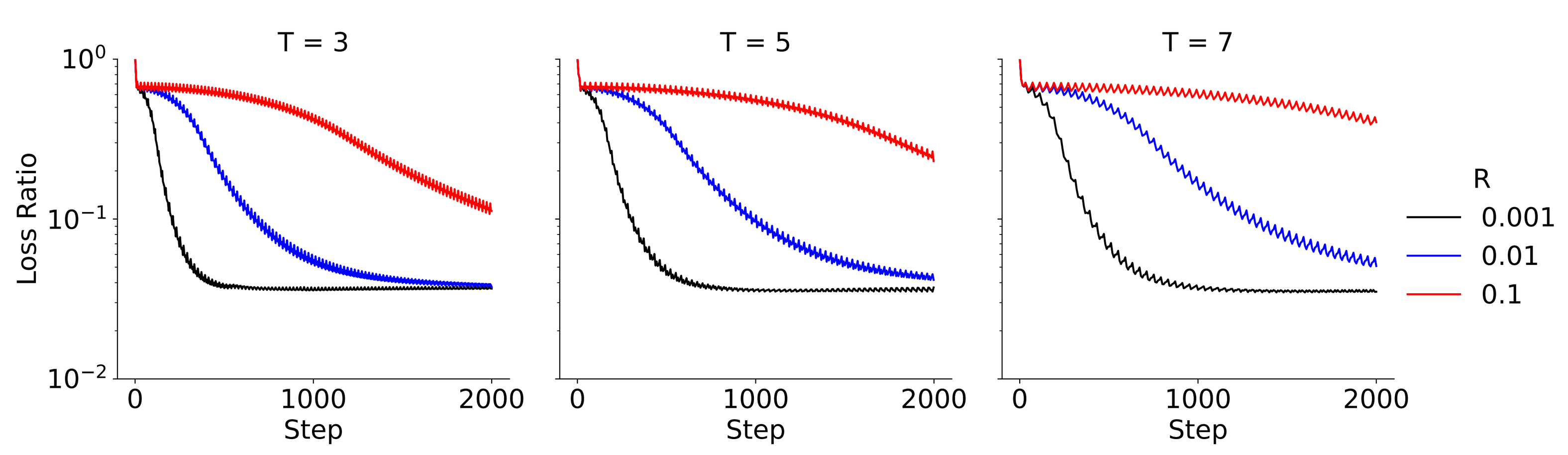
$$O\left(\sqrt{\frac{\chi(B^\top B)MD_{\mathcal{X}}^2}{\varepsilon^2}}\right) \text{ multiplications by } B^\top B \text{ per node.} \tag{9}$$

## Experiments

We conducted numerical experiments on the following optimization problem:

$$\min_x f(x) := \frac{1}{n}\sum_{i=1}^{n} f_i(x) \quad \text{subject to } Bx = 0, \text{where } f_i(x) = \sqrt{\frac{1}{m}\|C_i x - d_i\|^2},$$

with $C_i\in\mathbb{R}^{m\times d}$, $d_i\in\mathbb{R}^{m\times 1}$, and $x\in\mathbb{R}^d$.




## Conclusions and Future Work

Our approach relies on the gradient sliding algorithm, which requires parameter estimation before implementation, slightly weakening its theoretical performance. In our experiments, we showed that the effect of choice for different parameters $R$ and $T$.

Future work will focus on extending the algorithm to handle biased stochastic oracle and non-convex objectives, as well as exploring adaptive strategies to dynamically adjust the parameters of the algorithm based on the network topology and the structure of the optimization problem.

## Affiliation

1. Skolkovo Institute of Science and Technology, Moscow, Russia
2. Moscow Institute of Physics and Technology, Moscow, Russia

## QR code