
OPTIMAL GRADIENT SLIDING FOR SADDLE POINT PROBLEMS

DRAFT

Antyshev Tikhon
MIPT
antyshev.tg@phystech.edu

Ekaterina Borodich
MIPT
borodich.ed@phystech.edu

ABSTRACT

In this paper we expand the algorithm proposed in [1] to Saddle Point Problems from [2]. In this paper, we expand algorithm proposed in the paper [1] using a FOAM algorithm described in [2] in order to create Optimal Gradient Sliding for Saddle Point Problems.

Keywords Gradient Sliding, Saddle Point Problem

1 Introduction

We consider a variant of minimax optimization problem in the form:

$$\min_{x \in \mathbb{R}^{d_x}} \max_{y \in \mathbb{R}^{d_y}} f(x) + F(x, y) \quad (1)$$

where

Saddle point problems are a class of optimization problems that arise in a variety of applications, including game theory, economics, and machine learning. In these problems, the objective function has a saddle point, which is a point where the gradient is zero in one direction and nonzero in another direction. Saddle point problems are notoriously difficult to solve, and existing optimization methods often suffer from slow convergence or instability.

In recent years, there has been growing interest in developing new optimization methods for solving saddle point problems. One promising approach is to use first-order optimization methods that combine gradient descent with proximal operators and dual variables. These methods have been shown to be effective in solving convex optimization problems, and can be adapted to handle saddle point problems.

1.1 Main Contributions

1.2 Related Works

2 Preliminaries

Assumption 1.

Assumption 2.

3 Gradient Sliding

Lemma 1.

rewrite
and add
stuff

Algorithm 1 FOAMed AE**Input:** $x^0 = x_f^0 \in \mathbb{R}^{d_x}$, $y^0 = y_f^0 \in \mathbb{R}^{d_y}$ **Parameters:** $\tau \in (0, 1)$, η_x, η_y , θ , $\alpha > 0$, $K \in \{1, 2, \dots\}$ + FOAM parameters1: **for** $k = 0, 1, 2, \dots, K - 1$ **do**2: $(x_g^k, y_g^k) = \tau(x^k, y^k) + (1 - \tau)(x_f^k, y_f^k)$ 3: $(x_f^{k+1}, y_f^{k+1}) = \text{FOAM}(C_\theta^k(x, y), x_g^k, y_g^k, \varepsilon_k)$ ▷ where $C_\theta^k(x, y) = f(x_g^k) + \langle \nabla f(x_g^k), x - x_g^k \rangle + \frac{1}{2\theta} \|x - x_g^k\|^2 + F(x, y)$ 4: $x^{k+1} = x^k + \eta\alpha(x_f^{k+1} - x^k) - \eta\nabla_x \left(f(x_f^{k+1}) + F(x_f^{k+1}, y_f^{k+1}) \right)$ 5: $y^{k+1} = y^k + \eta\alpha(y_f^{k+1} - y^k) + \eta\nabla_y F(x_f^{k+1}, y_f^{k+1})$ 6: **end for****Output:** (x^k, y^k)

4 FOAMed Sliding

In this section we present modified Accelerated Extragradient algorithm for solving SP problems.

An important note is that the following theorem is correct if $F(x, y)$ satisfies Danskin's Theorem differentiability condition.

Theorem 2. yada-yada it converges with [insert convergence rate]

Proof. In order to be able to use all of the previously proven convergence theorems we will need to modify their criterion. Original Accelerated Extragradient algorithm would require FOAM to achieve the following condition:

$$\|\nabla A_\theta^k(x_f^{k+1})\|^2 \leq \frac{L_f^2}{3} \|x_g^k - \arg \min_{x \in \mathbb{R}^{d_x}} A_\theta^k(x)\|$$

and in our case:

$$A_\theta^k(x) = f(x_g^k) + \langle \nabla f(x_g^k), x - x_g^k \rangle + \frac{1}{2\theta} \|x - x_g^k\|^2 + \max_{y \in \mathbb{R}^{d_y}} F(x, y)$$

The FOAM algorithm is capable to produce an epsilon-accurate solution of the minimax problem for C_θ^k , the epsilon-accurate solution is understood in the following terms:

$$\|x - x^*\|^2 + \|y - y^*\|^2 \leq \varepsilon$$

After differentiating $A_\theta^k(x)$ we will receive:

$$\nabla A_\theta^k(x) = \nabla f(x_g^k) + 2L_f(x - x_g^k) + \nabla_x F(x, \bar{y})$$

Since $F(x, y)$ satisfies Danskin's Theorem and convergence Theorwm requires $\theta = \frac{1}{2L_f}$. Suppose that the FOAM algorithm returned values (x_f, y_f) , which are the epsilon-accurate solution to the minimax problem, then they should be achieved with such accuracy ε_k , that:

$$\|\nabla f(x_g^k) + 2L_f(x_f - x_g^k) + \nabla_x F(x_f, \bar{y})\|^2 \leq \frac{L_f^2}{3} \|x_g^k - x^*\|^2$$

$$\begin{aligned} \|\nabla f(x_g^k) + 2L_f(x_f - x_g^k) + \nabla_x F(x_f, \bar{y})\|^2 &= \|\nabla f(x_g^k) + 2L_f(x_f - x^* + x^* - x_g^k) + \nabla_x F(x_f, \bar{y}) - \nabla_x F(x^*, \bar{y}) + \nabla_x F(x^*, \bar{y})\|^2 \leq \\ &\leq \|\nabla f(x_g^k) + 2L_f(x^* - x_g^k) + \nabla_x F(x^*, \bar{y})\|^2 + \|2L_f(x_f - x^*)\|^2 + \|\nabla_x F(x_f, \bar{y}) - \nabla_x F(x^*, \bar{y})\|^2 + \\ &+ 2\|\nabla f(x_g^k) + 2L_f(x^* - x_g^k) + \nabla_x F(x^*, \bar{y})\| \|2L_f(x_f - x^*)\| + 2\|\nabla f(x_g^k) + 2L_f(x^* - x_g^k) + \nabla_x F(x^*, \bar{y})\| \|\nabla_x F(x_f, \bar{y}) - \nabla_x F(x^*, \bar{y})\| \\ &\quad + 2\|2L_f(x_f - x^*)\| \|\nabla_x F(x_f, \bar{y}) - \nabla_x F(x^*, \bar{y})\| \leq \\ &\leq \|\nabla f(x_g^k) + 2L_f(x^* - x_g^k) + \nabla_x F(x^*, \bar{y})\|^2 + 4L_f^2 \varepsilon_k + L_F^2 \varepsilon_k + 2\|\nabla f(x_g^k) + 2L_f(x^* - x_g^k) + \nabla_x F(x^*, \bar{y})\| \cdot \\ &\quad \cdot (2L_f \sqrt{\varepsilon_k} + L_F \sqrt{\varepsilon_k}) + 4L_f L_F \varepsilon_k = (\|\nabla f(x_g^k) + 2L_f(x^* - x_g^k) + \nabla_x F(x^*, \bar{y})\| + (2L_f + L_F) \sqrt{\varepsilon_k})^2 \leq \end{aligned}$$

$$\begin{aligned}
&\leq \frac{L_f^2}{3} \|x_g^k - x^*\|^2 \\
&\quad \|\nabla f(x_g^k) + 2L_f(x^* - x_g^k) + \nabla_x F(x^*, \bar{y})\| + (2L_f + L_F)\sqrt{\varepsilon_k} \leq \frac{L_f}{\sqrt{3}} \|x_g^k - x^*\| \\
\varepsilon_k &\leq \left(\frac{L_f}{\sqrt{3}(2L_f + L_F)} \|x_g^k - x^*\| - \frac{1}{2L_f + L_F} \|\nabla f(x_g^k) + 2L_f(x^* - x_g^k) + \nabla_x F(x^*, \bar{y})\| \right)^2
\end{aligned}$$

□

[1] Dmitry Kovalev, Aleksandr Beznosikov, Ekaterina Borodich, Alexander Gasnikov, and Gesualdo Scutari. Optimal gradient sliding and its application to distributed optimization under similarity. 2022.

- 4

5 Proof of Lemmas**6 Proof of Theorems**