

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Method Description</b>	<b>2</b>
<b>3</b>	<b>Algorithm correctness</b>	<b>3</b>
3.1	Zero Error . . . . .	3
3.2	Nonzero Error . . . . .	4
3.3	Undifferentiable convex function . . . . .	4
<b>4</b>	<b>Error's Value</b>	<b>4</b>
<b>5</b>	<b>Number of iterations</b>	<b>7</b>
<b>6</b>	<b>Details</b>	<b>9</b>
6.1	Function calcularting . . . . .	11
6.2	The Derivative Value: Segment Solution . . . . .	12
6.3	The Derivative Value: Current Point On Segment . . . . .	12
<b>7</b>	<b>Tests</b>	<b>13</b>
7.1	About function <i>solve</i> . . . . .	13
7.2	Tests for iterations number . . . . .	13
7.3	Comparison Of Stop Conditions . . . . .	14
7.4	Comparison With Other Methods . . . . .	16
<b>8</b>	<b>Conclusion</b>	<b>17</b>

# 1 Introduction

In this paper we research one method of optimization on a square in  $\mathbb{R}^2$ . The method was offered by Nesterov (see ex. 4 from [1]).

In the paper [2] there are some results for this method. Namely, there are some estimates for iterations number and accuracy for task on segment (see the next section or [2]). Also there are comparison of this method with method of ellipsoids. Discussed method showed better results on time.

In this paper we want to continue this research. In the next section there is description of method and pseudocode for it. In the section 3. there are results where this method works correctly. The section 4. includes two ways to estimate accuracy for a task on segment and it is one of two main results in this paper. Following important result is theorem 5.3 for iterations number that is asymptotically twice as good as the estimate in [2].

Section Tests include different tests that show some theoretical results in practice and include comparison of this method with gradient descent.

# 2 Method Description

Let's consider a following task:

$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

where  $f$  is a convex function,  $Q$  - is a square on the plane.

Let's consider a following method. One solves task of minimization for a function  $g(x) = f(x, y_0 = \frac{a}{2})$  on a segment  $[0, a]$  with an accuracy  $\delta$  on function. After that one calculates a sub-gradient in a received point and chooses the rectangle which the sub-gradient "does not look" in. Similar actions are repeated for a vertical segment. As a result we have the square decreased twice. Let's find a possible value of error  $\delta_0$  for task on segment and a sufficient iteration's number  $N$  to solve the initial task with accuracy  $\epsilon$  on function.

Let's describe an algorithm formally. See pseudo-code 1.

---

**Algorithm 1** Algorithm of the method

---

```
1: function METHOD(convex function  $f$ , square  $Q = [a, b] \times [c, d]$ )
     $x_0 := \text{solve}(g = f(\cdot, \frac{c+d}{2}), [a, b], \delta)$ 
     $g = \text{subgradient}(f, (x_0, \frac{c+d}{2}))$ 
2:   if  $g[1] > 0$  then
3:      $Q := [a, b] \times [c, \frac{c+d}{2}]$ 
4:   else
5:      $Q := [a, b] \times [\frac{c+d}{2}, d]$ 
6:   end if
     $y_0 := \text{solve}(g = f(\frac{a+b}{2}, \cdot), [c, d], \delta)$ 
     $g := \text{subgradient}(f, (\frac{a+b}{2}, y_0))$ 
7:   if  $g[0] > 0$  then
8:      $Q := [a, \frac{a+b}{2}] \times [c, d]$ 
9:   else
10:     $Q := [\frac{a+b}{2}, b] \times [c, d]$ 
11:  end if
12:  if StopRec() == False then
13:    Method( $f$ ,  $Q$ )
14:  end if return  $(\frac{a+b}{2}, \frac{c+d}{2})$ 
15: end function
```

---

### 3 Algorithm correctness

Let's  $\mathbf{x}_0$  is solution of the task on segment,  $Q_1$  is choosed rectangle,  $Q_2$  is not choosed rectangle.

#### 3.1 Zero Error

**Lemma 3.1.** *If the optimization task on segment is solved with zero error and the  $f$  is convex and differentiable at a point-solution, rectangle with solution of initial task was choosed correct.*

*Proof.* From sub-gradient definition,  $\mathbf{x}^* \in \{x | (g(\mathbf{x}_0), \mathbf{x}_0 - \mathbf{x}^*) \geq 0\}$ . Lemma's statement follows from it and a fact that the first (or the second for vertical segment) gradient's component in point  $\mathbf{x}_0$  is zero.  $\square$

### 3.2 Nonzero Error

**Theorem 3.1.** *Let's the  $f$  has continuous derivative on the square. Then there is a neighbourhood of a solution of optimization task on segment such as a choice of rectangle will not change if one use any point from the neighbourhood.*

*Proof.* Let's consider a case when we work with horizontal segment. Case with vertical segment is considered analogously. Then we are interesting in  $f'_y(x_0, y_0)$ . If  $\mathbf{x}_0$  is not solution of initial task, then  $f'_y(x_0, y_0) \neq 0$ .

From a continuity of the derivative:

$$\lim_{\delta \rightarrow 0} f'_y(x_0 + \delta, y_0) = f'_y(x_0, y_0)$$

Therefore,

$$\exists \delta_0 : \forall \mathbf{x}_\delta \in B_{\delta_0}(x_0) \times y_0 \Rightarrow \text{sign}(f'_y(\mathbf{x}_\delta)) = \text{sign}(f'_y(\mathbf{x}_0))$$

From it and lemma 4.1 theorem's statement follows.

□

### 3.3 Undifferentiable convex function

The method does not work for all convex functions even for zero error on segment.

**Example 1.** There is an example in [2].

## 4 Error's Value

From derivative continuously we have following obvious result:

**Lemma 4.1.** *If  $f$  has continuous derivative. If  $|f'_y(x, y_0)| > 0$  for all  $x$  on horizontal segment, then the second gradient's component has same sign at all points of segment. If  $|f'_x(x_0, y)| > 0$  for all  $y$  on vertical segment, then the first gradient's component has same sign at all points of segment.*

**Example 2.** All functions  $f$  of the following type meet conditions of written above lemma:

$$f(x, y) = \psi(x) + \phi(y),$$

where  $\psi, \phi$  are convex and differentiable functions.

**Example 3.** Let's illustrate that we can not always take any point from segment. Let's consider following task:

$$\min \left\{ (x - y)^2 + x^2 \mid Q = [0, 1]^2 \right\}$$

On segment  $[0, 1] \times \{\frac{1}{2}\}$  this task has solution  $f^* = f(\frac{1}{4}, \frac{1}{2}) = \frac{1}{8}$ . Derivative on  $y$  at this point is  $f'_y(\frac{1}{4}, \frac{1}{2}) = \frac{1}{2}$  but at the point  $(1, \frac{1}{2})$  is equal to  $-1$ . We can see that in this case rectangle will be selected non-correctly.

Let's find possible value of error's value, i.e. let's find a number  $\delta_0$  such as if an error for a solution on a segment is less than  $\delta_0$  then rectangle for the segment is defined correctly. Rectangles are defined correctly for a horizontal optimization task, if:

$$\forall \delta : |\delta| < \delta_0 \Rightarrow f'_y(\mathbf{x}_0)f'_y(x_0 + \delta, y_0) > 0 \quad (1)$$

Analogically, for a vertical segment:

$$\forall \delta : |\delta| < \delta_0 \Rightarrow f'_x(\mathbf{x}_0)f'_x(x_0, y_0 + \delta) > 0 \quad (2)$$

**Theorem 4.1.** *Let function  $f$  be convex and has  $L$ -Lipschitz continuous gradient and a point  $\mathbf{x}_0$  is a solution of optimization's task on a current segment.*

*The current segment is horizontal and  $\exists M > 0 : \Rightarrow |f'_y(\mathbf{x}_0)| \geq M$  or the current segment is vertical and  $\exists M > 0 : \Rightarrow |f'_x(\mathbf{x}_0)| \geq M$ . Then rectangle is defined correctly if the possible value of error is less than  $\frac{M}{L}$ .*

*Proof.* Condition (1) is met if there is a derivative  $f'_y(x_0 + \delta, y_0)$  in a neighbourhood of  $f'_y(\mathbf{x}_0)$  with radius  $|f'_y(\mathbf{x}_0)|$ :

$$|f'_y(\mathbf{x}_0) - f'_y(x_0 + \delta, y_0)| < |f'_y(\mathbf{x}_0)|$$

The  $L$ -Lipschitz continuity gives following inequality:

$$|f'_y(\mathbf{x}_0) - f'_y(x_0 + \delta, y_0)| \leq L|\delta|$$

Therefore the following possible value is sufficient to select rectangle correctly:

$$\delta_0 < \frac{M}{L} \leq \frac{|f'_y(\mathbf{x}_0)|}{L}$$

Statement for vertical segment is proved similarly.  $\square$

In written above theorem the estimate needs some lower bound for derivative in point-solution on segment and this task can be hard in practise. Using other interpretation of the condition (1) one can take more useful result.

**Theorem 4.2.** *Let function  $f$  be convex and has  $L$ -Lipschitz continuous gradient and a point  $\mathbf{x}_0$  is a solution of optimization's task on a current segment.*

*The current segment is horizontal and  $M = |f'_y(\mathbf{x}_{current})|$  or the current segment is vertical and  $M = |f'_y(\mathbf{x}_{current})|$ . Then rectangle is defined correctly if a distance between  $\mathbf{x}_{cur}$  and accurate solution on segment is less than  $\frac{M}{L}$ .*

*Proof.* Condition (1) is met if there is a derivative  $f'_y(x_0 + \delta, y_0)$  in a neighbourhood of  $f'_y(\mathbf{x}_0)$  with radius  $|f'_y(\mathbf{x}_{cur})|$ :

$$|f'_y(\mathbf{x}_0) - f'_y(\mathbf{x}_{cur})| < |f'_y(\mathbf{x}_{cur})|$$

The  $L$ -Lipschitz continuity gives following inequality:

$$|f'_y(\mathbf{x}_0) - f'_y(\mathbf{x}_{cur})| \leq L|\delta|$$

Therefore the following possible value is sufficient to select rectangle correctly:

$$\delta_0 < \frac{M}{L} \leq \frac{|f'_y(\mathbf{x}_0)|}{L}$$

Statement for vertical segment is proved similarly.  $\square$

Theorems 4.1 and 4.2 give stop conditions in the case when gradient in point-solution on segment and close points is large. But what should we do if gradient in this point is small?

**Theorem 4.3.** *Let  $f$  be convex and has  $L$ -Lipschitz continuous gradient. Then for accuracy on function  $\epsilon$  following condition in point  $\mathbf{x}$  is sufficient:*

$$\|\nabla f(\mathbf{x})\| \leq \frac{\epsilon}{a\sqrt{2}},$$

where  $a$  is size of current square.

*Proof.* Let's consider following inequality for convex functions (see prove in [3]):

$$f(\mathbf{x}^*) - f(\mathbf{x}) \geq (\nabla f(\mathbf{x}), \mathbf{x}^* - \mathbf{x})$$

Using Cauchy–Bunyakovsky–Schwarz inequality one has following inequality

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}^*) &\leq -(\nabla f(\mathbf{x}), \mathbf{x}^* - \mathbf{x}) \leq \\ &\leq \|\nabla f(\mathbf{x})\| \|\mathbf{x}^* - \mathbf{x}\| \leq \|\nabla f(\mathbf{x})\| a\sqrt{2} \end{aligned}$$

This inequality proves theorem's statement.  $\square$

Let's make a couple of remarks.

Firstly, we can replace the Lipschitz condition on the square by the Lipschitz condition on the segments in the theorems 4.1 and 4.2.

Thirdly, all theorems in this section use  $f'_y$  on horizontal segment and  $f'_x$  on vertical segment. As a result, if one checks condition of stop on each iteration method will use full gradient on each iteration. It can slow down method but this estimates can be better than in [2]. As a result, method can work faster.

## 5 Number of iterations

Following estimates are correct if each iterations was correct (a rectangle is selected correctly on each iterations).

**Theorem 5.1.** *If function  $f$  is convex and  $L_f$ -Lipschitz continuous, then for to solve initial task with accuracy  $\epsilon$  on function one has to take a center of a current square as approximate solution and make following iteration's numbers:*

$$N = \left\lceil \log_2 \frac{L_f a}{\sqrt{2}\epsilon} \right\rceil \quad (3)$$

where  $a$  is a size of the initial square  $Q$ .

This estimate is a little improved estimate from [2] because of we use a center of a current square as approximate solution. The prove is similar to prove of not improved estimate.

There are functions which estimates from written above theorem are very accurate for.

**Example 5.** Let's consider following task with positive constant  $A$ :

$$\min \{A(x+y)|Q=[0,1]^2\}$$

If one take a center of a current solution as approximate solution one have value  $\frac{A}{2^N}$  after  $N$  iterations. Therefore, for accuracy  $\epsilon$  one has to  $\lceil \log_2 \frac{A}{\epsilon} \rceil$ . For this function  $L_f = 2A$ . Therefore, estimate (3) is accurate for such tasks with little error that not more one iteration.

But any convex function is locally Lipschitz continuous at all  $x \in \text{int } Q$ . Therefore, we have following theorem.

**Theorem 5.2.** *If function  $f$  is convex and a solution  $x^* \in \text{int } Q$ , then for to solve initial task with accuracy  $\epsilon$  on function one has to take a center of a current square as approximate solution and make following iteration's numbers:*

$$N = \left\lceil \log_2 \max \left\{ \frac{a}{\epsilon_0(x^*)}, \frac{L_f a}{\sqrt{2}\epsilon} \right\} \right\rceil \quad (4)$$

where  $a$  is a size of the initial square  $Q$ ,  $\epsilon_0(x^*)$  is size of neighbourhood of  $x^*$  which  $f$  is  $L_f$ -Lipschitz continuous in,  $\Delta f = f(x_0) - f(x^*)$ ,  $x_0$  is a center of square  $Q$ .

We can improve written above estimates if to add new conditions:

**Theorem 5.3.** *Let function  $f$  be convex and has  $L$ -Lipschitz continuous gradient.*

*If solution is a internal point, then for to solve initial task with accuracy  $\epsilon$  on function one has to take a center of a current square as approximate solution and make following iteration's numbers:*

$$N = \left\lceil \frac{1}{2} \log_2 \frac{La^2}{4\epsilon} \right\rceil \quad (5)$$

where  $a$  is a size of the initial square  $Q$ .

*Proof.* For all convex functions there is following inequality (one may find proof in [3]):

$$f(\mathbf{x}) - f(\mathbf{x}^*) - (f'(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^*) \leq \frac{L}{2} \|\mathbf{x} - \mathbf{x}^*\|^2$$

If  $\mathbf{x}^*$  is a solution and an internal point, then  $f'(\mathbf{x}^*) = 0$ :



$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq \frac{L}{2} \|\mathbf{x} - \mathbf{x}^*\|^2$$

After  $N$  iterations we have the estimate:

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq \frac{L}{4} \left( \frac{a}{2^N} \right)^2$$

Using it we have estimate (5).  $\square$

Estimate 3 works in all cases and better then estimate 5 when following condition is met:

$$\frac{2L_f^2}{L_g} \leq \epsilon,$$

где  $L_f, L_g$  is Lipschitz constants for function and for gradient. On the other hand, the estimate 5 works if point with zero gradient exists in the square.

## 6 Details

This method was created to solve dual task. Namely, following task:

$$-\phi(\lambda_1, \lambda_2) \rightarrow \min_{\lambda \geq 0}, \quad (6)$$

$$\text{where } \phi = \min_{\mathbf{x}} (f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \lambda_2 g_2(\mathbf{x})) \quad (7)$$

In the section we will discuss how transform this task to task of the task of minimization on square, what is derivative and lipschitz constants and how to calculate its fastly.

Fistly, let's transform this task to the task of minimization on square. According to [1] (see ex. 4.1), we can add following restraint for the dual variables:

$$\|\lambda\|_1 \leq \lambda_{\max} = \frac{1}{\gamma} \left( f(\bar{\mathbf{x}}) - \min_{\mathbf{x}} f(\mathbf{x}) \right), \quad (8)$$

$$\text{where } \bar{\mathbf{x}} : g_i(\bar{\mathbf{x}}) < 0, \gamma = \min_i g_i(\bar{\mathbf{x}}) \quad (9)$$

And we understand that there is the dual task's solution in a square  $Q = [0, \lambda_{\max}]^2$ . And we have following optimization task:

$$-\phi(\lambda_1, \lambda_2) \rightarrow \min_{\lambda \in Q} \quad (10)$$

Secondly, the gradient of function  $\phi$  will be calculated according well-known Demyanov-Danskin-Rubinov Theorem, see [5].

**Theorem 6.1.** *Let  $\phi(\lambda) = \min_{x \in X} \Phi(x, \lambda)$  for all  $\lambda \geq 0$ , where  $\Phi$  is a smooth convex function with respect to  $\lambda$  and  $x(\lambda)$  is the only maximum point. Then*

$$\nabla \phi(\lambda) = F'_\lambda(x(\lambda), \lambda)$$

If theorem's conditions is met for our case then we have derivative value:

$$\phi'_{\lambda_k}(\lambda) = g_k(\mathbf{x}(\lambda)) \quad (11)$$

Additionally, we need a Lipschitz constant for gradient. In the work [6] there is following theorem:

**Theorem 6.2.** *Let  $f(x)$  be a  $\mu_f$ -strongly convex function, the function  $g(x)$  satisfies the Lipschitz condition with a constant  $M_g$ . Then the function  $\phi(\lambda) = \min_{\mathbf{x}} (f(\mathbf{x} + \lambda_1 g_1(\mathbf{x}) + \lambda_2 g_2(\mathbf{x}))$  defined in 18, where  $x(\lambda) = \arg \min_x (f(x) + \lambda g(x))$ , has Lipschitz smooth gradient with constant  $L_{\phi'} = \frac{M_g^2}{\mu_f}$*

This theorem can be proved easy for the 2-dimensional space. In this case  $g$  is a vector-function.

In the following subsection we can use following notations:

$$\Phi(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \lambda_2 g_2(\mathbf{x})$$

$$\mathbf{x}(\lambda) = \arg \min_{\mathbf{x}} \Phi(\mathbf{x}, \lambda)$$

$$\phi(\lambda) = \min_{\mathbf{x}} \Phi(\mathbf{x}, \lambda) = \Phi(\mathbf{x}(\lambda), \lambda)$$

The next subsections are devoted to efficient calculation of function's and derivative's values.

## 6.1 Function calcularting

Let  $\phi(\lambda) = \Phi(x(\lambda), \lambda)$  is our object function,  $\tilde{\phi}(\lambda) = \Phi(\mathbf{x}, \lambda)$  is its approximation. We define following error functions:

$$\begin{aligned}\epsilon(\lambda) &= \left| \phi(\lambda) - \tilde{\phi}(\lambda) \right| - \text{approximation error} \\ \delta_\phi(\lambda_1, \lambda_2) &= \phi(\lambda_1) - \phi(\lambda_2)\end{aligned}$$

The GSS method and the dichotomy method are worked using signum of  $\delta_\phi$ . Therefore, when we use approximation for  $\delta_\phi$  the signum of  $\delta_{\tilde{\phi}}$  must match with the signum of  $\delta_\phi$ . It is met if following sufficient condition was met:

$$\left| \delta_\phi(\lambda_1, \lambda_2) - \delta_{\tilde{\phi}}(\lambda_1, \lambda_2) \right| \leq \left| \delta_\phi(\lambda_1, \lambda_2) \right| \quad (12)$$

Obviously, we can not calculate left part. But we have following obvious inequality:

$$\begin{aligned}\left| \delta_\phi(\lambda_1, \lambda_2) - \delta_{\tilde{\phi}}(\lambda_1, \lambda_2) \right| &\leq \\ \epsilon(\lambda_1) + \epsilon(\lambda_2)\end{aligned}$$

Therefore for 12 the following condition is sufficient:

$$\epsilon(\lambda_1) + \epsilon(\lambda_2) \leq \left| \delta_\phi(\lambda_1, \lambda_2) \right| \quad (13)$$

Let  $\phi$  is  $L_\phi$ -lipschitz continious. Then if the condition

$$\epsilon(\lambda_1) + \epsilon(\lambda_2) \leq L_\phi \left| \lambda_1 - \lambda_2 \right| \quad (14)$$

was met we can use the signum of  $\delta_{\tilde{\phi}}$  for our one dimensional methods.

If functions  $f$ ,  $g_1$  and  $g_2$  are lipschitz continious then  $\Phi$  is  $L_\Phi$ -lipschitz continious too and there is following estimate for error  $\epsilon$ :

$$\epsilon(\lambda) \leq L_\Phi \left| \mathbf{x}(\lambda) - \mathbf{x} \right|$$

## 6.2 The Derivative Value: Segment Solution

Let  $\lambda$  is a solution of one-dimensional minimization task for current segment. And in this point it should to calculate the derivative with respect to  $\lambda_k$  for to define what rectangle must be selected. If we used its approximation it is important that signum of approximation  $\phi'_{\lambda_k}$  and of  $\widetilde{\phi'_{\lambda_k}}$  are match. Analogically to previous subsection, we have following sufficient condition for it:

$$|\phi'_{\lambda_k} - \widetilde{\phi'_{\lambda_k}}| \leq |\widetilde{\phi'_{\lambda_k}}|$$

$$|g_k(\mathbf{x}(\lambda)) - g_k(\mathbf{x})| \leq |g_k(\mathbf{x})|$$

If  $g_k$  is  $L_{g_k}$ -lipschitz continious function:

$$L_{g_k} \|\mathbf{x}(\lambda) - \mathbf{x}\| \leq |g_k(\mathbf{x})|$$

## 6.3 The Derivative Value: Current Point On Segment

We have following condition for segment solution accuracy (see theorem 4.2):

$$\|\lambda_{cur} - \lambda^*\| = \delta \leq \frac{|\phi'_{\lambda_k}(\lambda_{cur})|}{L_{\phi'}} = \epsilon,$$

where  $\lambda_{cur}$  current approximation. In other notation:

$$\delta - \epsilon < 0$$

If we calculate approximation  $\tilde{\epsilon}$  instead  $\epsilon$  the signum of  $\delta - \epsilon$  will match with the signum of  $\delta - \tilde{\epsilon}$ . The following condtion is sufficient for it:

$$\left| (\delta - \epsilon) - (\delta - \tilde{\epsilon}) \right| \leq |\delta - \tilde{\epsilon}|$$

$$|\epsilon - \tilde{\epsilon}| \leq |\delta - \tilde{\epsilon}|$$

For our task  $\phi'_{\lambda_k}(\lambda)$  is equal to  $g_k(\mathbf{x}(\lambda))$  and  $\widetilde{\phi'_{\lambda_k}(\lambda)}$  is equal to  $g_k(\mathbf{x})$ . Then if  $g_k$  is  $L_{g_k}$ -lipschitz continious function we have following condition:

$$\frac{L_{g_k}}{L_{\phi'}} \|\mathbf{x}^* - \mathbf{x}\| \leq \left| |\lambda_1 - \lambda_2| - \frac{|g_k(\mathbf{x})|}{L_{\phi'}} \right|$$

$$L_{g_k} \|\mathbf{x}^* - \mathbf{x}\| \leq \left| L_{\phi'} |\lambda_1 - \lambda_2| - |g_k(\mathbf{x})| \right|$$

## 7 Tests

In this section we show estimate on number iterations of practice and compare work time our new method with work time classic methods of optimization such as gradient descent and method ellipsoid<sup>1</sup>. All code was made in Anaconda 5.3.1 Python 3.6 (see cite [4])

### 7.1 About function *solve*

If you look at pseudocode 1 you can find function *solve*. This function solves task of minimization on segment. A cost of one iteration depends on a choice of its implementation. In our tests we will use golden search selection. It is the best method for one-dimension task in our case because on each iteration the GSS need to calculated value of function in only one new point. It is important because we will use this method for dual tasks and calculating function in such tasks are hard enough.

### 7.2 Tests for iterations number

There is results of tests for iterations number. We consider task when global solution is internal point and external point. In 1 you can find values of estimates (see theorems 3, 5) for this experiments.

Estimate	Internal	External
Through $L_f$	40	30
Through $L_{f'}$	20	14

Table 1: Value for estimates through function's and gradient's lipschitz constant

Let's consider some quadratic form:

$$f(x, y) = (Ax + By)^2 + Cy^2 + Dx + Ey + F, C > 0.$$

One minimize this function on square that involves global solution. We took parameteres randomly.

---

<sup>1</sup>You can find all code in the repository [7]

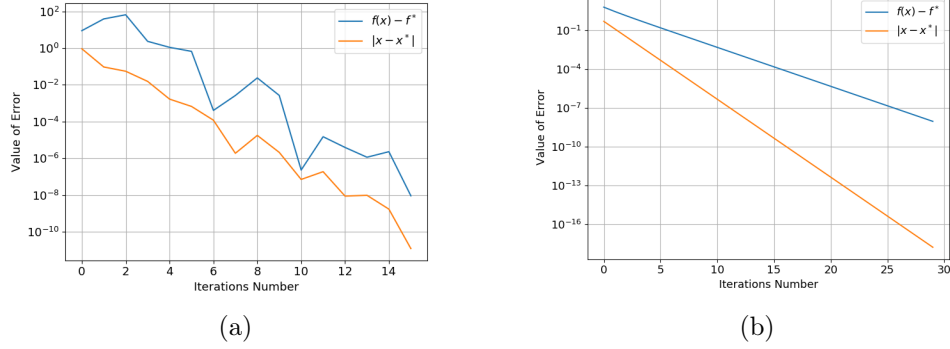


Figure 1: Test for iterations number (a) when the estimate 5 is reached; (b) when the estimate 5 is not reached.

We can see on 1(a) that the estimate 5 is reached in this case. But let's consider following task:

$$f(x, y) = (x + y)^2 + x^2, Q = [1, 2]^2, (x^*, y^*) = (1, 1)$$

$$\min_{(x, y) \in Q} f(x, y)$$

And we see on 1(b) that the estimate 5 is not reached in this case and in this case we should use estimate 3.

### 7.3 Comparison Of Stop Conditions

In current section we will compare different stop conditions for the solving task on separating segments. Firstly, let's define all stop conditions.

- Constant estimate (**Const**):  $\delta \leq \frac{\epsilon}{2La\sqrt{5} \log_2 \frac{2Ma\sqrt{2}}{\epsilon}}$  from [2],
- Constant gradient estimate (**ConstGrad**):  $\delta \leq \frac{M_{der}}{L}$  from 4.1,
- Estimate through current gradient (**CurGrad**):  $\frac{|f'(\mathbf{x}_{cur})|}{L}$  from 4.2.

Obviously, that in practic we can not use the second strategy because to calculate this estimate for one segment is harder than to solve all task. This estimate is the improved the third estimate. And there is interesting question: how much better does the second strategy work than the third?

We will test on the simple quadratic functions. Parameters for this functions will be generated randomly. We will research dependence of work time on required accuracy  $\epsilon$ . For each epsilon we will generate  $N$  tested functions and measure work time of each method. In our experiments  $N$  is equal to 1000 times. Results of experiments you can see on the picture 7.3.

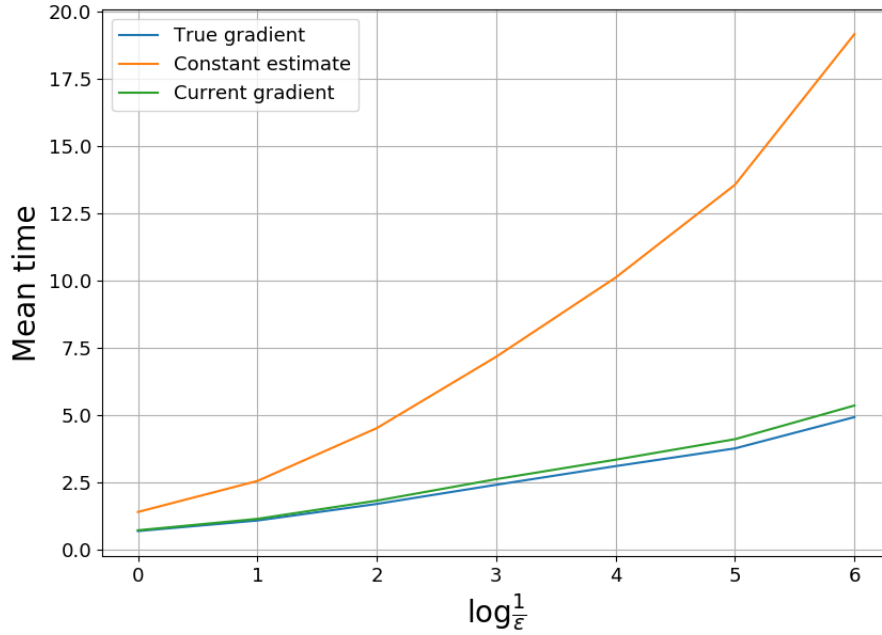


Figure 2: Comparison of different strategies for a segment

Firstly, we see that the first strategy is the slowest strategy from three strategy. There is good explanation for it. The estimate for the separatable segments is proportional to  $\epsilon$ . The method GSS makes the steps number proportional to the logarithm of estimate. Therefore, the number of iteration on segment for the first strategy is proportional to  $\ln \epsilon$ . It decreases speed of this strategy.

Secondly, we can see that the second and the third strategies have similar results. That's why we can state that the second estimate is good replacement for the third. Of course, it is the simple tasks. But if we will talk about dual tasks the main complexity is from calculating an objective function and the number of calculating is similar for both strategy. That's why we can state

that the second estimate is good estimate for dual task too.

From this part we conclude that the second strategy for segment is optimal.

## 7.4 Comparison With Other Methods

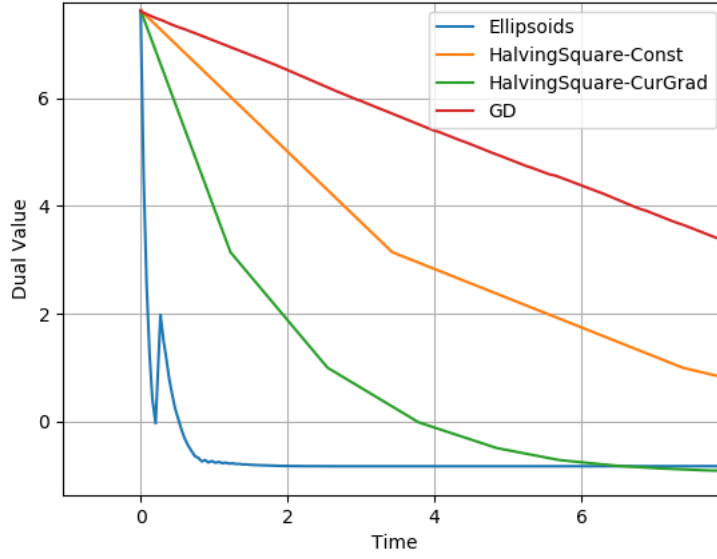


Figure 3: Comparison of different optimization method

Let's compare our method with ellipsoid method and gradient descent with optimal constant step  $\frac{1}{L_g}$  on some dual task.

We consider following prime task:

$$f(\mathbf{x}) = \ln \left( 1 + \sum_{k=1}^n e^{a_k x_k} \right) + \|\mathbf{x}\|_2^2 \rightarrow \min_{\mathbf{x}} \quad (15)$$

$$\text{s.t. } g_1(\mathbf{x}) = \|\mathbf{x}\|_2^2 - R_1 \leq 0 \quad (16)$$

$$g_2(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}\|_2^2 - R_2 \leq 0 \quad (17)$$

It is task to minimize LogSumExp-function on the intersection of sphere



with center in zero-point and radius  $R_1$  and of sphere with center  $\mathbf{c}$  and radius  $R_2$ . Parameter  $\mathbf{a}$  is generated randomly.

We introduce the following notation:

$$\phi(\lambda_1, \lambda_2) = -\min_{\mathbf{x}} (f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \lambda_2 g_2(\mathbf{x})) \quad (18)$$

In such notations the dual task for the task 15 looks like:

$$\phi(\lambda_1, \lambda_2) \rightarrow \min_{\lambda_1, \lambda_2} \quad (19)$$

$$\text{s.t } \lambda_1, \lambda_2 \geq 0 \quad (20)$$

Obviously,  $\min_{\mathbf{x}} f(\mathbf{x}) \geq 0$ . Therefore, according to 8 we can add following conditions on the dual variables:

$$|\lambda_k| \leq \lambda_{\max} = \frac{f(\bar{\mathbf{x}})}{\gamma}, k = 1, 2$$

And we have following task:

$$\phi(\lambda_1, \lambda_2) \rightarrow \min_{0 \leq \lambda_k \leq \lambda_{\max}}$$

How to calculate function and derivatave value for such task was discussed in the section 6.

And now we can compare speed of different methods for the dual task. We will compare ellipsoids method, gradient descent and our method HalvingSquare with two stop-condition for the task on segment. We will compare for different  $N = 10^k, k = \overline{1, 4}$ .

We can see on 7.4 following results. Firstly, gradient descent is the slowest method in this test. Secondly, our method is better than ellipsoid method for this task. Thirdly, the time gain of our method in comparison with ellipsoid method increases when  $N$  increases.

## 8 Conclusion

## References

- [1] Gasnikov A. Universal gradient descent // MIPT — 2018, 240 p.
- [2] Pasechnykh D.A., Stonyakin F.S. One method for minimization a convex Lipchitz continuous function of two variables on a fixed square // arXiv.org e-Print archive. 2018. — URL: <https://arxiv.org/pdf/1812.10300.pdf>
- [3] Nesterov U.E. Methods of convex optimization // M.MCNMO — 2010, 262 p.
- [4] Anaconda[site]. At available: <https://www.anaconda.com>
- [5] Danskin, J.M.: The theory of Max-Min, with applications. J. SIAM Appl. Math.14(4) (1966)
- [6] Fedor S. Stonyakin, Mohammad S. Alkousa, Alexander A. Titov, and Victoria V. Piskunova1 On Some Methods for Strongly Convex Optimization Problems with OneFunctional Constraint // ...
- [7] Repository with code: <https://github.com/ASEDOS999/Optimization-Halving-The-Square>