

Contents

1	Description Of A Task	2
2	Algorithm correctness	3
2.1	Zero Error	3
2.2	Nonzero Error	3
2.3	Undifferentiable convex function	3
3	Error's Value	4
4	Number of iterations	8
5	Tests	10
5.1	About function <i>solve</i>	10
5.2	Tests for iterations number	10
5.3	Comparison	13
5.3.1	Sinuses	14
5.3.2	Quadric functions	14

1 Description Of A Task

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 δ 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 δ_0 for task on segment and a sufficient iteration's number N to solve the initial task with accuracy ϵ 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  $Q = [a, b] \times [c, d]$ )
     $x_0 := solve(g = f(\cdot, \frac{c+d}{2}), [a, b], \delta)$ 
     $g = subgradient(f, (x_0, \frac{c+d}{2}))$ 
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
     $y_0 := solve(g = f(\frac{a+b}{2}, \cdot), [c, d], \delta)$ 
     $g := subgradient(f, (\frac{a+b}{2}, y_0))$ 
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

```

2 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.

2.1 Zero Error

Lemma 2.1. *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, $\mathbf{x}^* \in \{x | (\mathbf{g}(\mathbf{x}_0), \mathbf{x}_0 - \mathbf{x}^*) \geq 0\}$. 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

2.2 Nonzero Error

Theorem 2.1. *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 1 theorem's statement follows. \square

2.3 Undifferentiable convex function

The method does not work for all convex functions even for zero error on segment. Let's consider following example.

Example 1.

$$f(x, y) = \max\{x - 2y, y - 2x\}, Q = [-1, 1]^2 \quad (1)$$

Function f is convex as maximum of affine functions on x and y . A solution of task on horizontal segment $[-1, 1] \times \{0\}$ is point $(0, 0)$. Its subdifferential is

$$\partial f(0, 0) = \text{conv} \left\{ (1, -2)^\top, (-2, 1)^\top \right\}.$$

So if one takes a subgradient $(-2, 1)^\top$ then a bottom rectangle will be choosed. But optimal value is point $(1, 1)$ and there is not it in choosed rectangle. Therefore, this method cannot give a solution of initial task with error less than $\frac{1}{2}$.

3 Error's Value

From derivative continuously we have following obvious result:

Lemma 3.1. *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 ψ, ϕ 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 selected non-correctly.

Let's find possible value of error's value, i.e. let's find a number δ_0 such as if an error for a solution on a segment is less δ_0 then rectangle for the

segment is defined correctly. Rectangles are defined correctly for a horizontal optimization task, if:

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

Analogically, for a vertical segment:

$$\forall \delta : |\delta| < \delta_0 \Rightarrow f'_x(\mathbf{x}_0)f'_x(x_0, y_0 + \delta) > 0 \quad (3)$$

Theorem 3.1. *Let the function f be convex and differentiable and a current rectangle be $[a, b] \times [c, d]$.*

For horizontal segment: *There is f''_{xy} on the segment. Rectangle is defined in point $(x_0 + \delta, y_0)$ correctly if one meet following condition:*

$$\delta_0 \leq \frac{|f'_y(\mathbf{x}_0)|}{\max_{t \in [a, b]} |f''_{xy}(t, y_0)|} \quad (4)$$

For vertical segment: *There is f''_{yx} on the segment. Rectangle is defined in point $(x_0, y_0 + \delta)$ correctly if one meet following condition:*

$$\delta_0 \leq \frac{|f'_x(\mathbf{x}_0)|}{\max_{t \in [c, d]} |f''_{yx}(x_0, t)|} \quad (5)$$

Proof. Let's prove this statement for horizontal segment.

Rewrite the condition (2) using Taylor formula:

$$\forall \delta : |\delta| < \delta_0 \Rightarrow f'_y(\mathbf{x}_0) (f'_y(\mathbf{x}_0) + f''_{xy}(\mathbf{x}_0 + (\theta\delta, 0)^\top) \delta) > 0,$$

where $\theta \in (0, 1)$

Using the written above inequality we have a following inequality for δ_0 :

$$\delta_0 \leq \frac{|f'_y(\mathbf{x}_0)|}{\max_{\theta \in [-1, 1]} |f''_{xy}(x_0 + \theta\delta_0, y_0)|}$$

It and an obvious inequality $\max_{\theta \in [-1, 1]} |f''_{xy}(x_0 + \theta\delta_0, y_0)| < \max_{t \in [a, b]} |f''_{xy}(t, y_0)|$ proves (3). Inequality (4) are proved similar. \square

Example 4. All positive semidefinite quadratic form meet conditions of written above theorem:

$$B(x, y) = Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F,$$

$$B''_{xy} = B''_{yx} = 2B < \infty$$

where $A \geq 0, AC - B^2 \geq 0$. Also estimate for δ has some sense (estimate is non zero if \mathbf{x}_0 is not optimal solution). Also one can easy show that this estimate is accurate for task from example 3.

Theorem 3.2. *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 $\exists M > 0 : \Rightarrow |f'_y(\mathbf{x}_0)| \geq M$ 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 (2) 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 order to give an estimate for all functions with a continuous gradient let's consider following object. We define a modulus of continuity $\omega(f, \delta)$ for function f on a set E as:

$$\omega(f, \delta) = \left\{ \sup |f(x_1) - f(x_2)| \mid x_1, x_2 \in E, |x_1 - x_2| < \delta \right\}$$

Theorem 3.3. *Let function f be convex and has a continuous gradient and E is a current segment.*

$$\omega(f'_y, \delta_0) < |f'_y(\mathbf{x}_0)| \tag{6}$$

$$\omega(f'_x, \delta_0) < |f'_x(\mathbf{x}_0)| \quad (7)$$

If a modulus of continuity $\omega(f'_y, \delta_0)$ on a horizontal (vertical) segment E meet the condition (6) (the condition (7)) and possible error less than δ_0 then rectangle is selected correctly.

Proof. Let δ be such as $|\delta| < \delta_0$

$$\begin{aligned} |f'_y(\mathbf{x}_0) - f'_y(x_0 + \delta_1, y_0)| &\leq \sup_{\delta: |\delta| < \delta_0} |f'_y(\mathbf{x}_0) - f'_y(\mathbf{x}_0 + \delta)| \leq \\ &\leq \sup_{x_1, x_2 \in E: |x_1 - x_2| < \delta_0} |f'_y(\mathbf{x}_1) - f'_y(\mathbf{x}_2)| = \omega(f'_y, \delta_0) < |f'_y(\mathbf{x}_0)| \end{aligned}$$

The condition (2) 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 $|f'_y(\mathbf{x}_0)|$:

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

Therefore, the condition (2) is met. Statement for vertical segment is proved similarly. \square

Theorems 3.1 and 3.2 gives conditions of stop in the case when gradient in point-solution on segment is large. But what should we do if gradient in this point is small?

Theorem 3.4. *Let f be convex and has L -Lipschitz continuous gradient. Then for accuracy on function ϵ following condition in point \mathbf{x} is sufficient:*

$$\|\nabla f(\mathbf{x})\| \leq \frac{\sqrt{2}\epsilon}{3a},$$

where a is size of current square.

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

$$f(\mathbf{x}) + (\nabla f(\mathbf{x}), \mathbf{x}^* - \mathbf{x}) + \frac{1}{2L} \|\nabla f(\mathbf{x})\|_2^2 \leq f(\mathbf{x}^*)$$

Let's use Cauchy–Bunyakovsky–Schwarz inequality. As result, we have:

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq \|\nabla f(\mathbf{x})\|_2 \left(\|\mathbf{x}^* - \mathbf{x}\|_2 - \frac{1}{2L} \|\nabla f(\mathbf{x})\|_2 \right) \leq$$

$$\begin{aligned}
&\leq \|\nabla f(\mathbf{x})\|_2 \left(\|\mathbf{x}^* - \mathbf{x}\|_2 + \frac{1}{2L} \|\nabla f(\mathbf{x})\|_2 \right) \leq \\
&\leq \frac{3\|\nabla f(\mathbf{x})\|_2 \|\mathbf{x}^* - \mathbf{x}\|_2}{2} \leq \frac{3\|\nabla f(\mathbf{x})\|_2 a}{\sqrt{2}}
\end{aligned}$$

This inequallite proves statement of theorem. \square

Let's make a couple of remarks.

Firstly, theorems 3.1 and 3.2 need some lower bound for $f'_x(\mathbf{x}_0)$ and $f'_y(\mathbf{x}_0)$ on segments. We can take this constant using Taylor's formula. For example, on horizontal segment we have:

...

If the left part is positive it is constant M from theorem 3.2 and it can be used in theorem 3.1.

Secondly, we can replace the Lipschitz condition on the square by the Lipschitz condition on the segments in the theorem 3.2. Similarly, we can replace condition for the second derivative in theorem 3.1.

4 Number of iterations

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

Theorem 4.1. *If function f is convex and L -Lipschitz continuous, then for to solve initial task with accuracy ϵ on function one has to take a center of a current square as proximal solution and make following iteration's numbers:*

$$N = \left\lceil \log_2 \frac{La}{\sqrt{2}\epsilon} \right\rceil \quad (8)$$

where a is a size of the initial square Q .

Proof.

$$|f(\mathbf{x}^*) - f(\mathbf{x})| < L|\mathbf{x}^* - \mathbf{x}|$$

After N iterations we have a square with size $\frac{a}{2^N}$. That's why if we choose a squares center as proximal solution we have following estimates:

$$|\mathbf{x}^* - \mathbf{x}| \leq \frac{a}{\sqrt{2}} 2^{-N}$$

$$|f(\mathbf{x}^*) - f(\mathbf{x})| < \frac{1}{\sqrt{2}} La 2^{-N}$$

Therefore, for accuracy epsilon following number of iterations is sufficient:

$$N > \log_2 \frac{La}{\sqrt{2}\epsilon}$$

It proves the estimate (8). \square

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 proximal solution one have value $\frac{A}{2^N}$ after N iterations. Therefore, for accuracy ϵ one has to $\lceil \log_2 \frac{A}{\epsilon} \rceil$. For this function $L = 2A$. Therefore, estimate (8) 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 4.2. *If function f is convex and a solution $x^* \in \text{int } Q$, then for to solve initial task with accuracy ϵ on function one has to take a center of a current square as proximal solution and make following iteration's numbers:*

$$N = \left\lceil \log_2 \max \left\{ \frac{a}{\epsilon_0(x^*)}, \frac{La}{\sqrt{2}\epsilon} \right\} \right\rceil \quad (9)$$

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

We can improve written above estimates:

Theorem 4.3. *Let function f be convex and has L_g -Lipschitz continuous gradient.*

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

$$N = \left\lceil \frac{1}{2} \log_2 \frac{L_g a^2}{4\epsilon} \right\rceil \quad (10)$$

where a is a size of the initial square Q .

Proof. For all convex functions there is following inequality (one may find proofs in many books on convex analysis or convex optimization):

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

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

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

After N iterations we have the estimate:

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

Using it we have estimate (10). □

5 Tests

In this section we show estimate on number iterations of practice and compare work time of gradient descent and our new method - halving square¹.

5.1 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 gradient descent with step $h_k = \frac{h}{\sqrt{k}} = \frac{a}{4\sqrt{k}}$, where a is size of current segment.

5.2 Tests for iterations number

Let's make tests for the estimate (8).

Functions $-\sin \frac{\pi x}{a}$ and $-\sin \frac{\pi y}{b}$ are convex on square $Q = [0, 1]^2$ when $a, b \geq 1$. Therefore, a function $-A \sin \frac{\pi x}{a} - B \sin \frac{\pi y}{b}$ is convex for all $A, B \geq 0$ as cone combination of convex function.

Functions x^n are convex and monotonously non-decreasing on $[0, 1]$ for all $n \in \mathbb{N}$ that's why functions $(-A \sin \frac{\pi x}{a} - B \sin \frac{\pi y}{b} + A + B + D)^n$ are convex for all $D \geq 0$.

¹You can find all code in [this repository](#)

Therefore, following function is convex:

$$f(x, y) = -A_1 \sin \frac{\pi x}{a_1} - B_1 \sin \frac{\pi y}{b_1} + \sum_{n=2}^N \left(-A_n \sin \frac{\pi x}{a_n} - B_n \sin \frac{\pi y}{b_n} + A_n + B_n + D_n \right)^n,$$

where $A_i, B_i, D_i \geq 0$ and $a_i, b_i \geq 1$ for all $i = \overline{1, n}$.

The function f is differentiable infinite times and we can use it to test the method.

Let's take $a_1 = \dots = a_n = a$ and $b_1 = \dots = b_n = b$:

$$f(x, y) = -A_1 \sin \frac{\pi x}{a} - B_1 \sin \frac{\pi y}{b} + \sum_{n=2}^N \left(-A_n \sin \frac{\pi x}{a} - B_n \sin \frac{\pi y}{b} + A_n + B_n + D_n \right)^n,$$

where $A_i, B_i, D_i \geq 0$ for all $i = \overline{1, n}$ and $a, b \geq 1$.

We have following estimates for derivatives on horizontal and vertical segments:

$$\begin{aligned} |f'_x| \Big|_{x=x_0} &\geq \left(A_1 + \sum_{n=2}^N n A_n D_n^{n-1} \right) \frac{\pi}{a} \left| \cos \frac{\pi x_0}{a} \right| \\ |f'_y| \Big|_{y=y_0} &\geq \left(B_1 + \sum_{n=2}^N n B_n D_n^{n-1} \right) \frac{\pi}{b} \left| \cos \frac{\pi y_0}{b} \right| \end{aligned}$$

Therefore this functions met conditions of lemma 3.1 and we can take gradient at any point of segment. One can see examples of functions f on [fig. 1](#)

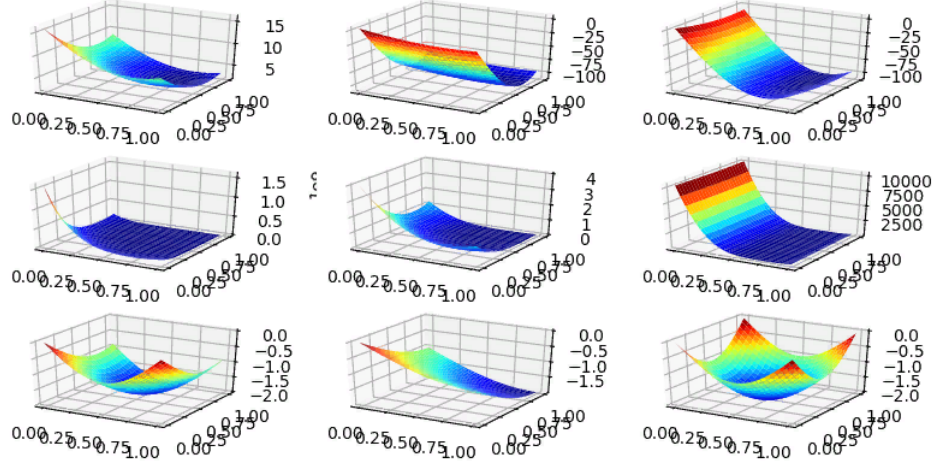


Figure 1: Examples of tests functions

We will use a and b from $[1, 2]$ and $N = 2$. Let's solve task of minimization function f with different parameters on square $[0, 1]^2$ through new method and compares number of iteration with estimate (8).

Result of the test you can see on fig. 2.

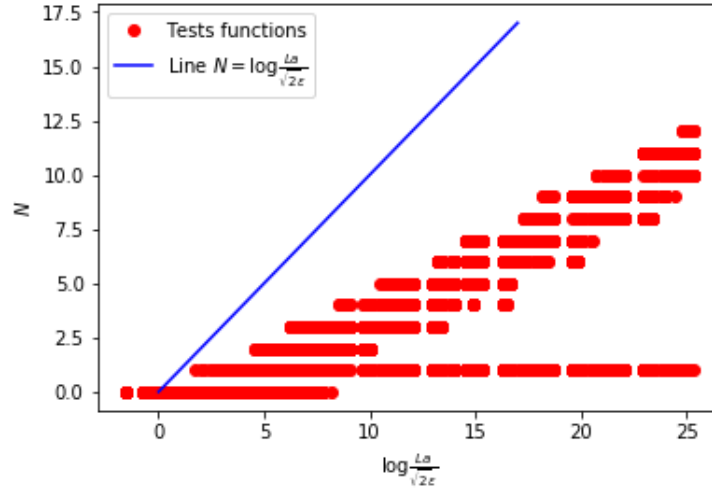


Figure 2: Tests' results for iterations number

On graphic N is number of done iterations for accuracy ϵ , $\log_2 \frac{La}{\sqrt{2}\epsilon}$ is the

parameter of tests tasks. Line $N = \log_2 \frac{La}{\sqrt{2}\epsilon}$ is our estimates (8).

We can see that there are tests points under this line. It confirms our estimates (8).

5.3 Comparison

Let's compare experimentally work's time of gradient descent and method of halving square. We will compare it on optimization task for functions from previous subsection and random quadric function with positive semidifinite matrix:

$$f(x, y) = -A_1 \sin \frac{\pi x}{a} - B_1 \sin \frac{\pi y}{b} + \left(-A_2 \sin \frac{\pi x}{a} - B_2 \sin \frac{\pi y}{b} + A_2 + B_2 + D_2 \right)^2, \quad (11)$$

where $A_i, B_i, D_i \geq 0$ for all $i = \overline{1, 2}$ and $a, b \in [1, 2]$.

$$f(x, y) = (Ax + By)^2 + Cx^2 + Dx + Ey + F, C \geq 0 \quad (12)$$

Halving square will stop when current value will be no further than ϵ from optimal value. Gradient descent will stop when distance between current and previous value will be not more than ϵ .

We will use gradient descent with a step $h_k = \frac{h}{\sqrt{k}} = \frac{a}{4\sqrt{k}}$, where a is size of the square. Also we add limit $N_{max} = 1000$ on the iterations number for method of halving square, solving on segment and gradient descent.

When we compare two method we distinguish the following situations (similar name of situations on graphics):

- Type 1 (T1) - tasks which gradient descent completed but halving square was not complete,
- Type 2 (T2) - gradient descent is faster than halving square,
- Type 3 (T3) - work's times are approximately equal or time is too short to measure,
- Type 4 (T4) - gradient descent is slower than halving square,
- Type 5 (T5) - gradient descent was not complete but halving square completed,
- Type 6 (T6) - both methods were not completed successfully

5.3.1 Sinuses

As already mentioned in previous subsection, functions (11) meet conditions of lemma 3.1, therefore, we can take any point from segment on each iteration in method of halving square. Using all written above we have following results (see fig. 3).

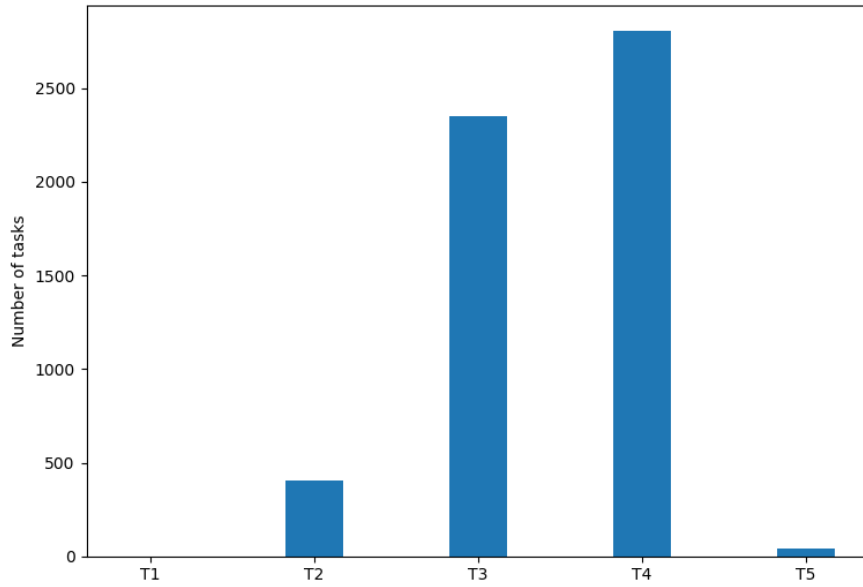


Figure 3: Tests' results for comparison on sinuses

We can see that new method is not less efficient on the most tasks and more efficient on the task's half than gradient descent.

5.3.2 Quadric functions

Quadric functions do not always meet conditions of lemma 3.1, therefore, we can not think that error for task on segment can be very big. We can try to estimate this error but let's use for this task gradient descent with $|x - x_{prev}| < \delta$, where δ is much less than size of segment. In our experiments we will use $\delta = \frac{a}{400}$. Using it we have following results (see fig. 4).

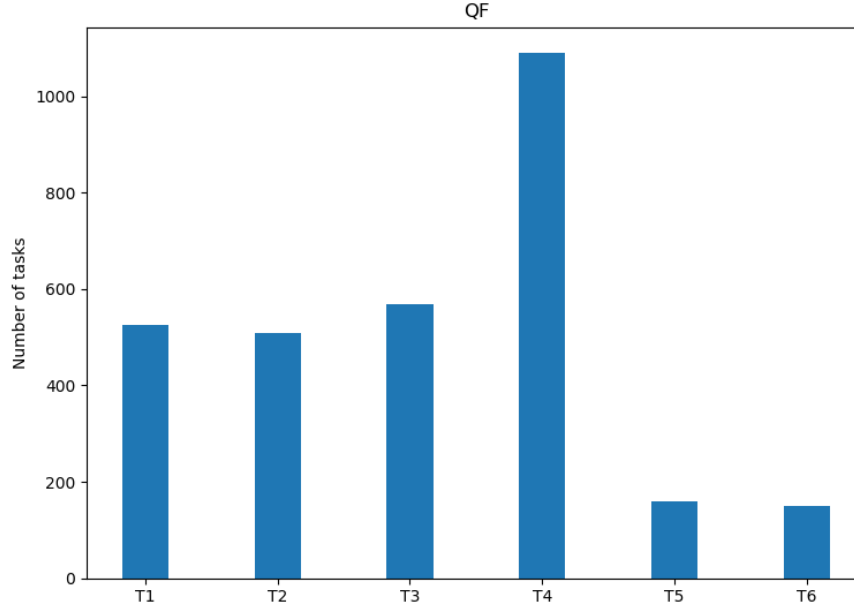


Figure 4: Tests' results for comparison on quadric function

We can see that halving square was not completed successfully on big parts of task - about $\frac{1}{6}$ from all tasks. We can decrease δ in its modification, but work's time increases in this case.

Also gradient descent was better than halving square in more than half tests. Decreasing δ lead to decreasing work's time but also it can lead to that method will not completed successfully more often.

Therefore, halving square with proposed function *solve* is not efficient.