# One Method for Convex Optimization on Square

I. Kuruzov[1] F. Stonyakin[1, 2]

[1]Moscow Institut of Physics and Technology

[2]V.I. Vernandsky Crimean Federal University

QIPA, 2019

# Method's Description

Problem:

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

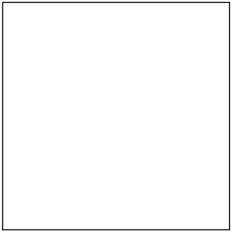
where $f$ is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$ is a square.
The method was proposed by Yu. Nesterov:

# Method's Description

Problem:

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

where $f$ is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$ is a square.
The method was proposed by Yu. Nesterov:

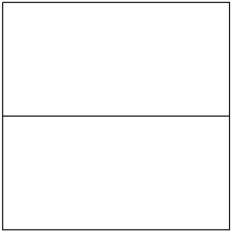


# Method's Description

Problem:

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

where $f$ is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$ is a square.
The method was proposed by Yu. Nesterov:

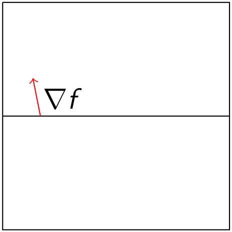


# Method's Description

Problem:

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

where $f$ is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$ is a square.
The method was proposed by Yu. Nesterov:

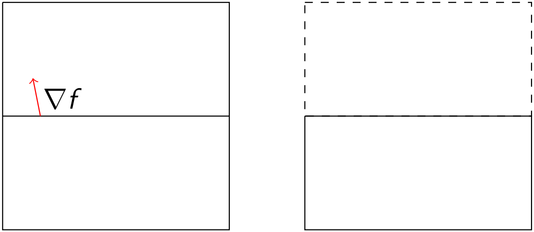


# Method's Description

Problem:

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

where $f$ is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$ is a square.
The method was proposed by Yu. Nesterov:

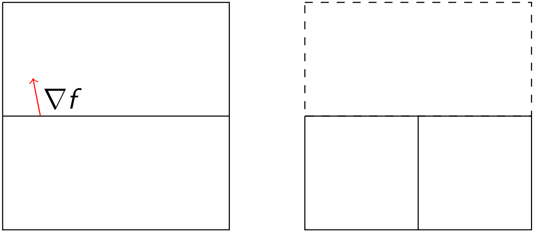


# Method's Description

Problem:

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

where $f$ is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$ is a square.
The method was proposed by Yu. Nesterov:

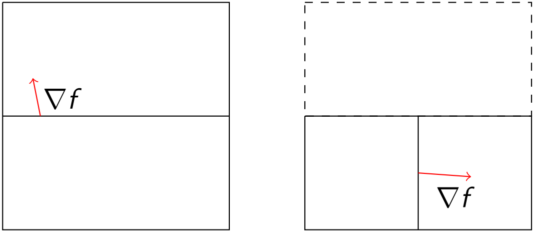


# Method's Description

Problem:

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

where $f$ is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$ is a square.
The method was proposed by Yu. Nesterov:

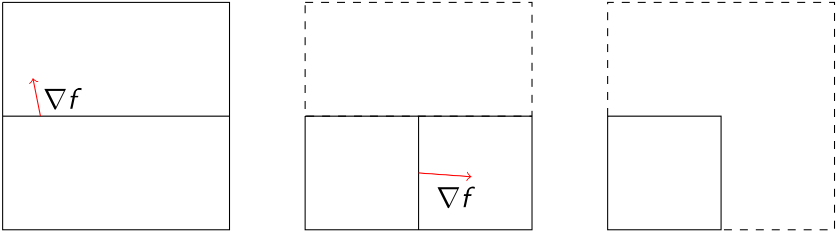


# Method's Description

Problem:

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

where $f$ is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$ is a square.
The method was proposed by Yu. Nesterov:



# Plan

Let $f$ be $L$-Lipschitz continious function with $M$-Lipschitz continious gradient.
$\delta$ is a distance between the one-dimensional task's solution and its approximation.

## Constant Estimate

Then if each one-dimensional task was solved with the following accuracy

$$\boxed{\delta \leq \frac{\epsilon}{2Ma(\sqrt{2}+\sqrt{5})(1-\frac{\epsilon}{La\sqrt{2}})}}$$

then this method converge to minimum of $f$ on square with accuracy $\epsilon$ on function.

Let $f$ be $L$-Lipschitz continious function with $M$-Lipschitz continious gradient.
$\delta$ is a distance between the one-dimensional task's solution and its approximation.

## Constant Estimate

Then if each one-dimensional task was solved with the following accuracy

$$\boxed{\delta \leq \frac{\epsilon}{2Ma(\sqrt{2}+\sqrt{5})(1-\frac{\epsilon}{La\sqrt{2}})}}$$

then this method converge to minimum of $f$ on square with accuracy $\epsilon$ on function.

There are examples where there is not convergence on argument.

# Strategies
## Current Gradient

Purpose: to select rectangle with $x^*$.

Purpose: to select rectangle with $x^*$.
Let $x_* = \arg\min_{x \in \text{segment}} f(x)$ is a solution of one-dimensional task, $x_{current}$ is its current approximation on segment.
Purpose:

$$\text{sign}\, f'_y(x_*) = \text{sign}\, f'_y(x_{current})$$

# Strategies
## Current Gradient

Purpose: to select rectangle with $x^*$.
Let $x_* = \arg\min_{x \in \text{segment}} f(x)$ is a solution of one-dimensional task, $x_{current}$ is its current approximation on segment.
Purpose:

$$\text{sign}\, f'_y(x_*) = \text{sign}\, f'_y(x_{current})$$

More strong condition

$$|f'_y(x_*) - f'_y(x_{current})| \leq |f'_y(x_{current})|$$

Purpose: to select rectangle with $x^*$.
Let $x_* = \arg\min_{x \in \text{segment}} f(x)$ is a solution of one-dimensional task, $x_{current}$ is its current approximation on segment.
Purpose:

$$\operatorname{sign} f'_y(x_*) = \operatorname{sign} f'_y(x_{current})$$

More strong condition

$$|f'_y(x_*) - f'_y(x_{current})| \leq |f'_y(x_{current})|$$

### Current Gradient

If $f$ has $M$-Lipschitz continious gradient then if each task on segment was solved with the accuracy satisfying condition

$$\boxed{\delta \leq \frac{|f'_y(x_{current})|}{M}}$$

## Small Gradient

$f$ is convex function with $M$-Lipschitz continious gradient. Then $\mathbf{x}$ is solution of inital task with accuracy $\epsilon$ on function, if

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

where $a$ is square's size.

# Convergence

Let $f$ be convex function. $Q$ is square with size $a$.

## Convergence

If function $f$ is $L$-Lipschitz continious then for to approach accuracy $\epsilon$ on function the following iterations number is sufficient:

$$N = \left\lceil \log_2 \frac{La\sqrt{2}}{\epsilon} \right\rceil . \quad (1)$$

# Convergence

## Convergence

If

1. Function $f$ has $M$-Lipschitz continious gradient
2. $\exists \mathbf{x}^* \in Q : \nabla f(\mathbf{x}^*) = \mathbf{0}$
3. Strategy gives a convergence on argument

then for to approach accuracy $\epsilon$ on function the following iterations number is sufficient:

$$N = \left\lceil \frac{1}{2} \log_2 \frac{Ma^2}{4\epsilon} \right\rceil . \tag{2}$$

Problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

$$\text{s.t. } g_k(\mathbf{x}) \leq 0, k = \overline{1, m}$$

where $f$ is $\mu_f$-strong convex $L_f$-Lipschitz continious function with $M_f$-Lipschitz continious gradient , $g_k$ is convex $L_{g_k}$-Lipschitz continious function with $M_{g_k}$-Lipschitz continious gradient for all $k = \overline{1, m}$.

Dual problem

$$\min_{\lambda \in \mathbb{R}^m_+} \Phi(\lambda),$$

where

$$\Phi(\lambda) = -\min_{\mathbf{x} \in \mathbb{R}^n} \left( f(\mathbf{x}) + \langle \lambda, g(\mathbf{x}) \rangle \right).$$

$$x(\lambda) = \arg\min_{\mathbf{x} \in \mathbb{R}^n} \left( f(\mathbf{x}) + \langle \lambda, g(\mathbf{x}) \rangle \right)$$

## Slater's Condition

If $x_0 \in \mathbb{R}^n : g(x_0) < 0$ then

$$\|\lambda^*\|_1 \leq \frac{1}{\gamma}\left(f(x_0) - f(x^*)\right) = a, \text{где } \gamma = \min_k \left[-g_k(x_0)\right].$$

### Slater's Condition

If $x_0 \in \mathbb{R}^n : g(x_0) < 0$ then

$$\|\lambda^*\|_1 \leq \frac{1}{\gamma}\left(f(x_0) - f(x^*)\right) = a, \text{где } \gamma = \min_k \left[-g_k(x_0)\right].$$

$$\min_{\lambda \in \mathbb{R}^m_+} \Phi(\lambda) = \min_{\lambda \in Q} \Phi(\lambda),$$

where $Q = [0, a]^m$.

$$\Phi(\lambda) = -\min_{\mathbf{x}\in\mathbb{R}^n} \left(f(\mathbf{x}) + \langle \lambda, g(\mathbf{x})\rangle\right)$$

Gradient (Demyanov-Danskin-Rubinov theorem):

$$\nabla\Phi(\lambda) = -g(\mathbf{x}(\lambda))$$

Lipschitz constant for function:

$$L = \max \|g(\mathbf{x})\|$$

Lipschitz constant for gradient:

$$M = \frac{L_g^2}{\mu_f}$$

Questions:

1. How does one do iteration on segment?

2. How does one test the stop condition for one-dimensional task $\delta \leq \frac{|\Phi_2'(\lambda)|}{L}$?

3. How does one select a rectangle?

We are interesting in only signum:

1. $\Phi_1'(\lambda) = g_1\left(\mathbf{x}(\lambda)\right)$
2. $\delta - \frac{|\Phi_2'(\lambda)|}{M} = \delta - \frac{|g_2(\mathbf{x}(\lambda))|}{M}$
3. $\Phi_2'(\lambda) = g_2\left(\mathbf{x}(\lambda)\right)$

We are interesting in only signum:

1. $\Phi_1'(\lambda) = g_1\left(\mathbf{x}(\lambda)\right)$
2. $\delta - \frac{|\Phi_2'(\lambda)|}{M} = \delta - \frac{|g_2(\mathbf{x}(\lambda))|}{M}$
3. $\Phi_2'(\lambda) = g_2\left(\mathbf{x}(\lambda)\right)$

Use one trick:

$$\forall a, b \neq 0\, |a - b| \leq |b| \to \operatorname{sign} a = \operatorname{sign} b$$

We are interesting in only signum:

1. $\Phi_1'(\lambda) = g_1\left(\mathbf{x}(\lambda)\right)$
2. $\delta - \frac{|\Phi_2'(\lambda)|}{M} = \delta - \frac{|g_2(\mathbf{x}(\lambda))|}{M}$
3. $\Phi_2'(\lambda) = g_2\left(\mathbf{x}(\lambda)\right)$

Use one trick:

$$\forall a, b \neq 0 \, |a - b| \leq |b| \to \text{sign } a = \text{sign } b$$

Stop-condition for calculating $x(\lambda)$:

1. $L_{g_1}\|\mathbf{x} - \mathbf{x}(\lambda)\| \leq |g_1(\mathbf{x})|$
2. $\frac{L_{g_2}}{M}\|\mathbf{x} - \mathbf{x}(\lambda)\| \leq \left|\delta - \frac{|g_2(\mathbf{x})|}{M}\right|$
3. $L_{g_2}\|\mathbf{x} - \mathbf{x}(\lambda)\| \leq |g_2(\mathbf{x})|$

$$f(\mathbf{x}) = \log_2\left(1 + \sum_{k=1}^{n} e^{\alpha x_k}\right) + \beta\|\mathbf{x}\|_2^2 \to \min_{\mathbf{x}\in\mathbb{R}^n}$$

$$\text{s.t. } g_k(\mathbf{x}) = \langle b_k, \mathbf{x}\rangle + c_k$$

Dual problem:

$$-\min_{\mathbf{x}\in\mathbb{R}^n}\left(f(\mathbf{x}) + \langle\lambda, g(\mathbf{x})\rangle\right) \to \min_{\lambda\in[0,a]^2}$$

1. Ellipsoids Method with $\epsilon$-subgradient

Convergence:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \max_{\lambda \in Q} |\Phi(\lambda)| \exp\left(-\frac{k}{8}\right) + \epsilon$$

# Experiments

## Other methods

1. Ellipsoids Method with $\epsilon$-subgradient
Convergence:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \max_{\lambda \in Q} |\Phi(\lambda)| \exp\left(-\frac{k}{8}\right) + \epsilon$$

2. Primal Gradient Method with $(\delta, L)$-oracle.
Convergence:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \frac{LR^2}{2} \exp\left(-k\frac{\mu}{L}\right) + \delta$$

1. Ellipsoids Method with $\epsilon$-subgradient
Convergence:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \max_{\lambda \in Q} |\Phi(\lambda)| \exp\left(-\frac{k}{8}\right) + \epsilon$$

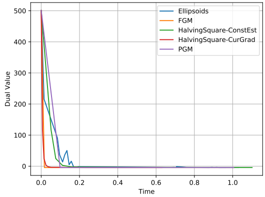
2. Primal Gradient Method with $(\delta, L)$-oracle.
Convergence:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \frac{LR^2}{2} \exp\left(-k\frac{\mu}{L}\right) + \delta$$

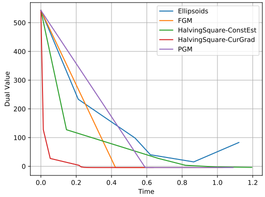
3. Fast Gradient Method with $(\delta, L)$-oracle.
Convergence:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \min\left(\frac{4LR^2}{k^2}, LR^2 \exp\left(-\frac{k}{2}\sqrt{\frac{\mu}{L}}\right)\right) + C_k\delta,$$

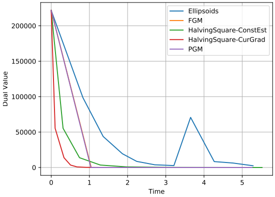
# Experiments
## Results



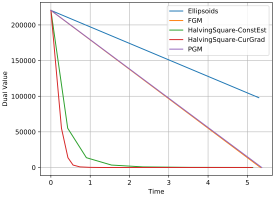

$n = 100, \epsilon = 1e-3$




$n = 100, \epsilon = 1e-10$




$n = 10000, \epsilon = 1e-3$




$n = 10000, \epsilon = 1e-10$

# Generalization

In the case of dimension $>2$:

$$\text{Square} \to n\text{-dimensional hypercube}$$

$$\text{Separating segment} \to n-1\text{-dimensional hypercube}$$

In the case of dimension >2:

$$\text{Square} \to n\text{-dimensional hypercube}$$

$$\text{Separating segment} \to n-1\text{-dimensional hypercube}$$

Convergence:

$$N = \left\lceil \log_2 \frac{La\sqrt{2n}}{\epsilon} \right\rceil. \tag{3}$$

# Generalization

In the case of dimension >2:

$$\text{Square} \to n\text{-dimensional hypercube}$$

$$\text{Separating segment} \to n-1\text{-dimensional hypercube}$$

Convergence:

$$N = \left\lceil \log_2 \frac{La\sqrt{2n}}{\epsilon} \right\rceil. \tag{3}$$

One can use this method recursively!

# Summary

- One Method for Two-dimensional optimization
- Its application in dual problems
- Comparison of different modifications and different methods
- Generalization

Thank you for your attention!