One Method for Convex Optimization on Square

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

[1]Moscow Institute of Physics and Technology

[2]V.I.Vernadsky Crimean Federal University

COG, 2019

Method Description

Task:

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

where f is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$.

One iteration

- Find minimum with accuracy δ on central horizontal segment in square and calculate gradient at this point

Method Description

Task:

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

where f is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$.

One iteration

- Find minimum with accuracy δ on central horizontal segment in square and calculate gradient at this point
- Choose rectangle which anti-gradient looks in

Method Description

Task:

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

where f is a convex function, $Q = [a,b] \times [c,d] \in \mathbb{R}^2$.

One iteration

- Find minimum with accuracy δ on central horizontal segment in square and calculate gradient at this point
- Choose rectangle which anti-gradient looks in
- Analogically for vertical segment in the rectangle

Plan

1. Strategy for segment accuracy
2. Convergence
3. Convergence
4. Tests

Strategies

True Gradient

Let f be convex and has L-Lipschitz continuous gradient.

$$\operatorname{sign} f_y'(x_0) = \operatorname{sign} f_y'(x_{current})$$

For this it is sufficient:

$$|f_y'(x_0) - f_y'(x_{current})| \leq |f_y'(x_0)|$$

Strategies
True Gradient

Let f be convex and has L-Lipschitz continuous gradient.

$$\operatorname{sign} f'_y(x_0) = \operatorname{sign} f'_y(x_{current})$$

For this it is sufficient:

$$|f'_y(x_0) - f'_y(x_{current})| \leq |f'_y(x_0)|$$

$$\boxed{\delta < \frac{|f'_y(x_0)|}{L}}$$

Strategies
Current Gradient

$$\text{sign } f'_y(x_0) = \text{sign } f'_y(x_{current})$$

For this it is sufficient:

$$|f'_y(x_0) - f'_y(x_{current})| \leq |f'_y(x_{current})|$$

$$\operatorname{sign} f_y'(x_0) = \operatorname{sign} f_y'(x_{current})$$

For this it is sufficient:

$$|f_y'(x_0) - f_y'(x_{current})| \leq |f_y'(x_{current})|$$

$$\boxed{\delta < \frac{|f_y'(x_{current})|}{L}}$$

Strategies
Small Gradient

But what can we do when norm of gradient is small?

But what can we do when norm of gradient is small?

Small Gradient

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

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

where a is size of current square.

Convergence

Function f is convex. Size of square Q is equal to a. One takes a center of a current square as approximate solution.

Estimate through Lipschitz function constant

Let function f be L_f-Lipschitz continuous. Then for accuracy ϵ on function it is sufficient:

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

Convergence

Function f is convex. Size of square Q is equal to a. One takes a center of a current square as approximate solution.

Estimate through Lipschitz function constant

Let function f be L_f-Lipschitz continuous. Then for accuracy ϵ on function it is sufficient:

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

Estimate through Lipschitz gradient constant

Let function f have L_g-Lipschitz continuous gradient. Moreover, point with zero derivative is **internal point**. Then for accuracy ϵ on function it is sufficient:

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

Convergence

If following inequallity is met

$$\frac{2L_f^2}{L_g} \geq \epsilon$$

then the estimate through L_g is better than through L_f.

- Quadratic Function

$$f(x,y) = (Ax + By)^2 + Cy^2 + Dx + Ey$$

$$(x^*, y^*) \in Q$$

$$L_f = \max_{(x,y)\in Q} \|\nabla f(x,y)\|$$

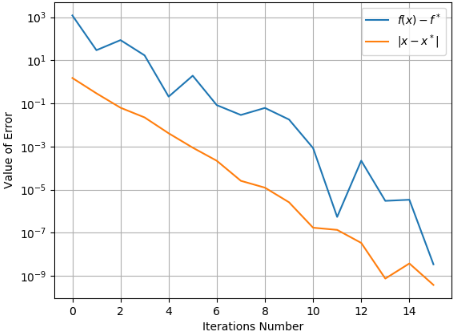
$$L_g = \max |\lambda\left(H(f)\right)|$$

Experiments

Iterations Number

Theoretical Iteration Number through function constant equals 40
Theoretical Iteration Number through gradient constant equals 20

Theoretical Iteration Number through function constant equals 40
Theoretical Iteration Number through gradient constant equals 20



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

Theoretical Iteration Number through function constant 30
Theoretical Iteration Number through gradient constant 14

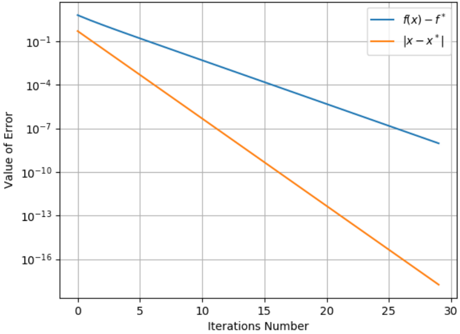
Experiments

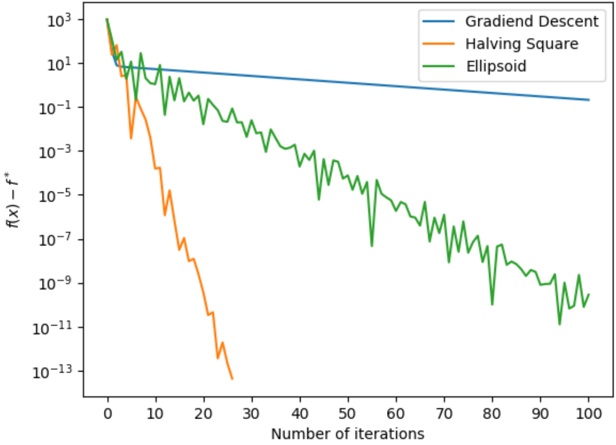
Iterations Number

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

Theoretical Iteration Number through function constant 30

Theoretical Iteration Number through gradient constant 14





Gradiend Descent
Halving Square
Ellipsoid
$f(x) - f^*$
10^{3}
10^{1}
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
0 10 20 30 40 50 60 70 80 90 100
Number of iterations

Work time

Gradiend Descent 4.5 ms
Halving Square 20.9 ms
Ellipsoid 28.3 ms

Summary

- Strategy for solution on segment
- Convergence Results
- Experiments for this method
  - Iterations number
  - Comparison with different methods