Subgradient method with Polyak's step in transformed space

Petro Stetsyuk[1], Viktor Stovba[1], and Zhanna Chernousova[2]

[1] V.M. Glushkov Institute of Cybernetics, Kyiv, Ukraine
[2] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

OPTIMA-2018
Petrovac, Montenegro
October 1–5, 2018

Outline

1. Polyak's subgradient method ($m = 1$)
2. Method A ($m \geqslant 1$)
3. Method B ($m \geqslant 1$, matrix B)
4. Comparison of the methods A and B

Contents

Formulation of the problem

We consider the following problem

$$\text{to find} \quad x^\star = \arg\min_{x \in R^n} f(x) \quad \text{if } f^\star \text{ is known,} \tag{1}$$

where $f(x)$ is a convex function and $f^\star = f(x^\star) = \min_{x \in R^n} f(x)$.

Main inequality for the subgradient

For convex function $f(x)$ the following inequality is valid

$$(x - x^\star, g_f(x)) \geq f(x) - f^\star, \quad \forall x \in R^n, \tag{2}$$

where $g_f(x)$ is a subgradient (gradient) of the function $f(x)$.

Inequality (2) is used for the calculation of the step in subgradient method, which B.T. Polyak offered in 1969.

Polyak B.T. Minimization of unsmooth functionals // *USSR Comput. Math. Math. Phys.*. 1969. Vol. 9, No. 3, pp. 14-29.

Polyak's subgradient method

Polyak's subgradient method has the iterative form

$$x_{k+1} = x_k - h_k \frac{g_f(x_k)}{\|g_f(x_k)\|}, \quad h_k = \frac{f(x_k) - f^\star}{\|g_f(x_k)\|}, \quad k{=}0, 1, \ldots \tag{3}$$

Step h_k is called Polyak's step.

Decrease of the distance to the minimum point

Theorem 1 (Polyak, 1969)

The sequence $\{x_k\}_{k=0}^{\infty}$, *generated by the method (3), satisfies the inequalities*

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - \frac{(f(x_k) - f^\star)^2}{\|g_f(x_k)\|^2}, \quad k = 0, 1, 2, \ldots$$

Remark. Theorem 1 guarantees that in Polyak's method the distance to the minimum point decreases monotonically.

m-inequality for the subgradient

We consider methods A and B for convex functions $f(x)$, subgradients $g_f(x)$ of which satisfy the following condition:

$$(x - x^\star, g_f(x)) \geq m(f(x) - f^\star), \quad \forall x \in R^n, \tag{4}$$

where parameter $m \geq 1$.

The method A uses the step h_k in **original space**, the method B uses h_k in **transformed space** of variables.

Contents

Method A

Initialization. $f^\star$, $m \geq 1$, $x_0 \in R^n$, $\varepsilon > 0$.

A1. Calculate $f(x_k)$ and $g_f(x_k)$.
If $f(x_k) - f^* \leq \varepsilon$, then STOP ($k^* = k$, $x_\varepsilon^* = x_k$).

A2. Calculate the next point

$$x_{k+1} = x_k - h_k \frac{g_f(x_k)}{\|g_f(x_k)\|}, \quad h_k = \frac{m(f(x_k) - f^*)}{\|g_f(x_k)\|}.$$

A3. Go to the $(k+1)$-th iteration with x_{k+1}.

Decrease of the distance to the minimum point

Theorem 2 (Stetsyuk, 2012)

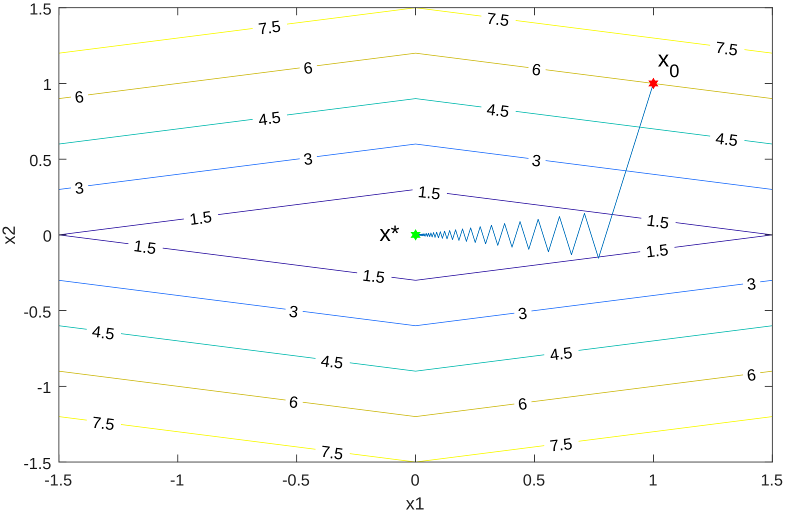
The sequence $\{x_k\}_{k=0}^{k^*-1}$*, generated by the method A, satisfies the inequalities*

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - \frac{m^2(f(x_k) - f^\star)^2}{\|g_f(x_k)\|^2}, \quad k = 0, 1, 2, \ldots \quad (5)$$

Remark. Theorem 2 guarantees that in Polyak's subgradient method the distance to the minimum point decreases monotonically if the inequality (4) is used.

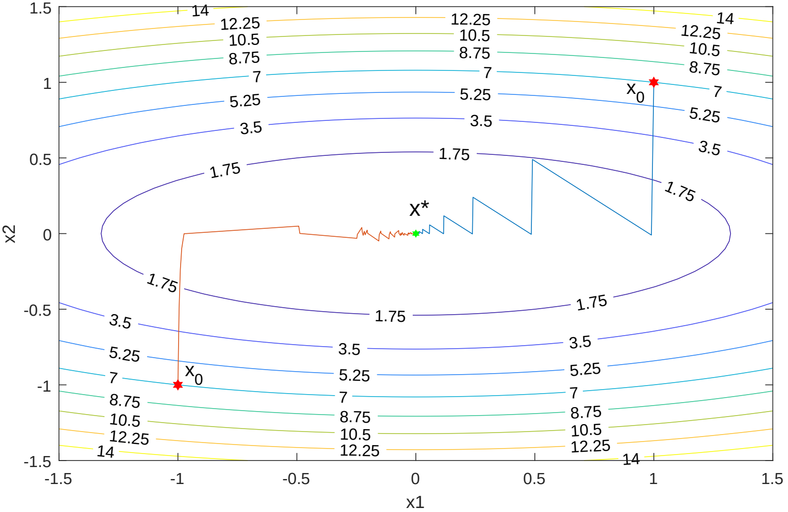
Method A for $f_1(x_1, x_2) = |x_1| + 5|x_2|$

$$m = 1, \ f^\star = 0, \ x_0 = (1, 1)^T$$



Method A for $f_2(x_1, x_2) = x_1^2 + 6x_2^2$

Blue - $m = 2,\ x_0 = (1, 1)^T$ Red - $m = 1,\ x_0 = (-1, -1)^T$



Contents

Method B

Initialization. $f^{\star}$, $m \geq 1$, $x_0 \in R^n$, $n \times n$ matrix B, $\varepsilon > 0$.

A1. Calculate $f(x_k)$ and $g_f(x_k)$.
If $f(x_k) - f^* \leq \varepsilon$, then STOP ($k^* = k$, $x_\varepsilon^* = x_k$).

A2. Calculate the next point

$$x_{k+1} = x_k - h_k B \frac{B^T g_f(x_k)}{\|B^T g_f(x_k)\|}, \quad h_k = \frac{m(f(x_k) - f^*)}{\|B^T g_f(x_k)\|}.$$

A3. Go to the $(k+1)$-th iteration with x_{k+1}.

Here h_k is the Polyak's step in the transformed space of variables.

If B is an identity matrix, the method B turns into the method A.

Decrease of the distance to the minimum point

Theorem 3 (Stetsyuk, Stovba, Chernousova, 2018)

The sequence $\{x_k\}_{k=0}^{k^*-1}$*, generated by the method B, satisfies the inequalities*

$$\|A(x_{k+1}-x^\star)\|^2 \leq \|A(x_k-x^\star)\|^2 - \frac{m^2(f(x_k)-f^\star)^2}{\|B^T g_f(x_k)\|^2}, \ k=0,1,\ldots \quad (6)$$

Remark. Theorem 3 guarantees that in the method B the distance to the minimum point decreases monotonically in the trasformed space if the inequality (4) is used.

Method B for $f_1(x_1, x_2) = |x_1| + 10|x_2|$

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix}, \; x_0 = (1, 1)^T$$

ε_f	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$	$\alpha = 5$
1.0e-02	262	114	62	19	17	13
1.0e-04	492	216	119	44	31	22
1.0e-06	**722**	**319**	**177**	**70**	**45**	**31**
1.0e-08	952	421	234	95	60	40
1.0e-10	1183	523	292	121	74	49

Contents

Functions used to compare the methods A and B

1. Piecewise quadratic function

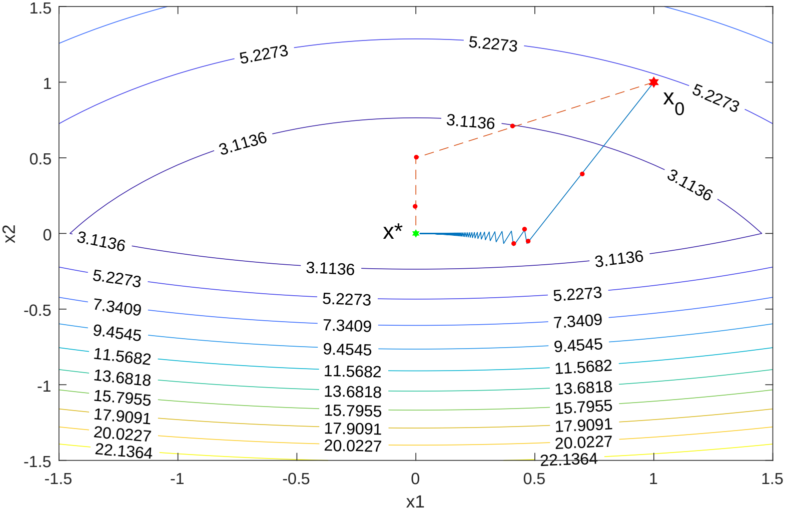
$$f_3(x_1, x_2) = \max\left\{x_1^2 + (2x_2 - 2)^2 - 3, x_1^2 + (x_2 + 1)^2\right\},$$
$$x^* = (0, 0),\ f^* = 1.$$

2. Quadratic function

$$f_4(x) = \|Ax - b\|^2,\ x^* = (1, 1, ..., 1)^T,\ f^* = 0.$$

Methods A and B for piecewise quadratic function

$x_0 = (1,1)^T$, $B = diag(1; 0.5)$, method A - blue, method B - red



Methods A and B for function $f_4(x) = \|Ax - b\|^2$

$$A = \|a_{ij}\|_{i,j=1}^{l,n} = \left\| \begin{array}{ccc} 100 & 0 & 0\ldots0 \\ 0 & 100 & 0\ldots0 \\ & A_1 & \end{array} \right\|, \quad b_i = \sum_{j=1}^{n} a_{ij},\, i = \overline{1,l},$$

A_1 - 498×100 random matrix, $m = 2$.

ε	Method A		Method B	
	itnA	$\|x_\varepsilon^* - x^*\|$	**itnB**	$\|x_\varepsilon^* - x^*\|$
1.0e-04	695	8.0612e-04	86	1.2175e-04
1.0e-06	1085	8.2861e-05	108	1.1773e-05
1.0e-08	1491	8.3470e-06	130	1.1353e-06
1.0e-10	**1901**	**8.3881e-07**	**150**	**1.3526e-07**
1.0e-12	2313	8.3994e-08	172	1.3018e-08
1.0e-14	2725	8.4440e-09	194	1.2523e-09

Conclusions

1. Using parameter $m \geq 1$ in Polyak's subgradient method gives a possibility to minimize special classes of convex functions more effectively.

2. Application of space transformation can accelerate convergence of subgradient method with Polyak's step in the transformed space of variables.

Thanks

Volkswagen Foundation
grant No 90 306

Questions?

THANK YOU
FOR YOUR ATTENTION

vik.stovba@gmail.com