On acceleration of subgradient method with Polyak's step

Petro Stetsyuk

stetsyukp@gmail.com

V.M.Glushkov Institute of Cybernetics, Kyiv, Ukraine

International Conference "Dynamical Systems: Stability,
Control, Optimization" (DSSCO'18), dedicated to
the 100th anniversary of Ye.A. Barbashin,
September 24–29, 2018, Minsk, Belarus

Outline

Outline

Formulation of the Problem

We consider the following problem

$$\text{to find} \quad x^{\star} = \underset{x \in R^n}{\operatorname{argmin}} f(x), \quad \text{if } f^{\star} \text{ is known,} \qquad (1)$$

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

Polyak's subgradient method

Polyak's subgradient method has the iterative form

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

Step h_k is called Polyak's step.

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

Decrease of the distance to the minimum point

Theorem 1 (Polyak,1969)

The sequence $\{x_k\}_{k=0}^{\infty}$, generated by the method (2), satisfies the inequalities

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

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

Outline

1. Polyak's subgradient method

2. **Polyak's method and ravine functions**

3. One-rank ellipsoidal operator

4. Accelerated method with Polyak's step

5. Comparison of both methods

Examples of slow convergence of Polyak's method

Example 1. Ravine piecewise linear function $(t \gg 1)$

$$f_1(x_1, x_2) = |x_1| + t|x_2|, \quad x^\star = (0, 0) \quad f^\star = 0.$$

The method (2) converges at a geometric rate with ratio $q(t) = \sqrt{1 - \frac{1}{t^2}}$ (Polyak, 1969). For large t the $q(t)$ is close to one and the method (2) converges very slowly.
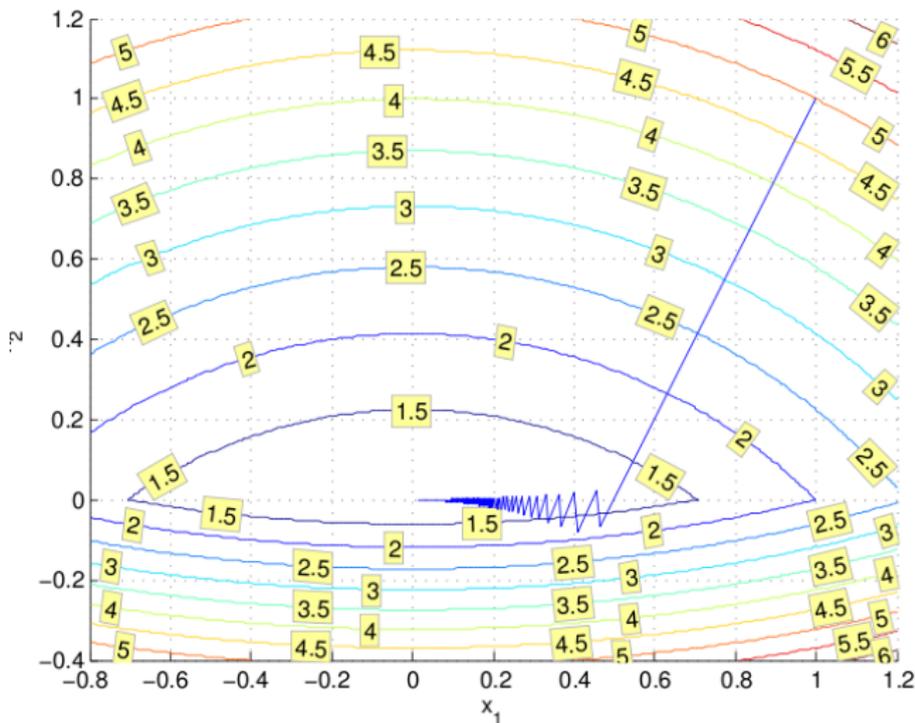
Example 2. Essentially ravine piecewise quadratic function

$$f_2(x_1, x_2) = \max \left\{ x_1^2 + (2x_2 - 2)^2 - 3, x_1^2 + (x_2 + 1)^2 \right\},$$

Degeneration at the minimum point $x^\star = (0, 0)$, $f^\star = 1$.

Polyak's method for f_1: $t = 10$, $x_0 = (1, 1)$
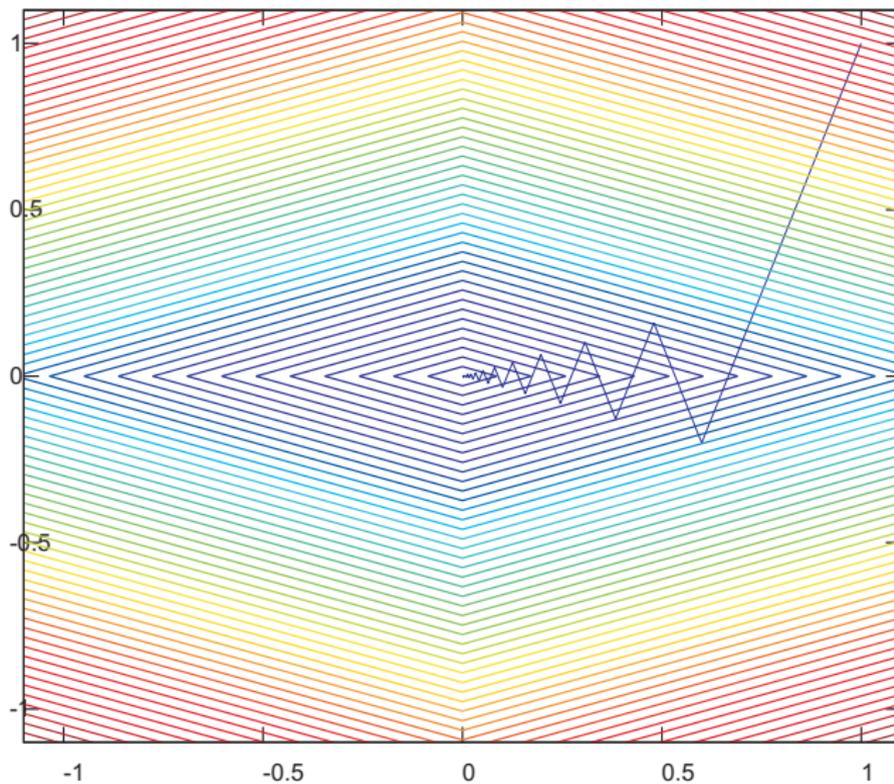
Polyak's method for f_2: $x_0 = (1, 1)$, >10000 iter.

Petro Stetsyuk On acceleration of subgradient method with Polyak's step

How can we handle the ravine?

1. A slow convergence of Polyak's method for ravine convex functions is caused by an obtuse angle between two successive subgradients. The closer the angle to 180 degrees, the slower the method's (2) convergence.

2. It is possible to accelerate the method (2) if we transform the space of variables so that to decrease the obtuse angle between two successive (consecutive) subgradients.

Outline

One-rank ellipsoidal operator

is a linear operator from R^n to R^n in the following form

$$T_1(\xi, \eta) = I_n - \left(\frac{1 - \sqrt{1 - (\xi, \eta)^2}}{1 - (\xi, \eta)^2} \eta - \frac{(\xi, \eta)}{1 - (\xi, \eta)^2} \xi \right) \eta^T, \qquad (3)$$

where I_n – identity $n \times n$-matrix, $\xi \in R^n$, $\eta \in R^n$ are vectors, such that $\|\xi\| = 1$, $\|\eta\| = 1$ and $(\xi, \eta)^2 \neq 1$.

In the transformed space $Y = T_1(\xi, \eta)X$ the angle between images of vectors ξ and η from $X = R^n$ is a right angle.

📄 Stetsyuk (1997) Orthogonalizing linear operators in convex programming, Cybernetics and System Analysis, **33**, No. 3.

One-rank ellipsoidal operator $(\xi^T \eta = (\xi, \eta) < 0)$
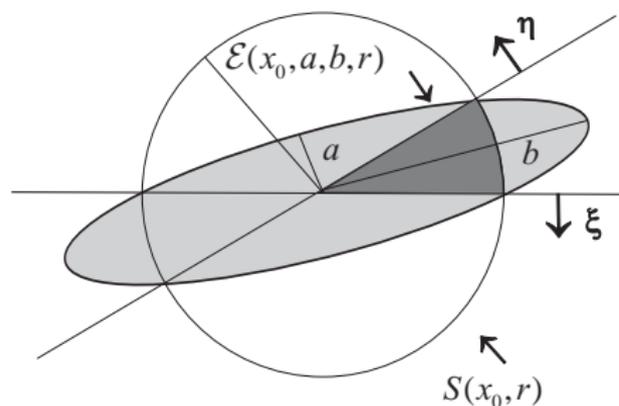
Makes a new ball from the optimal volume 2d-ellipsoid $\mathcal{E}(x_0, a, b, r)$, circumscribed around the intersection of the ball and two half-spaces, passing through the center of the ball $S(x_0, r)$.
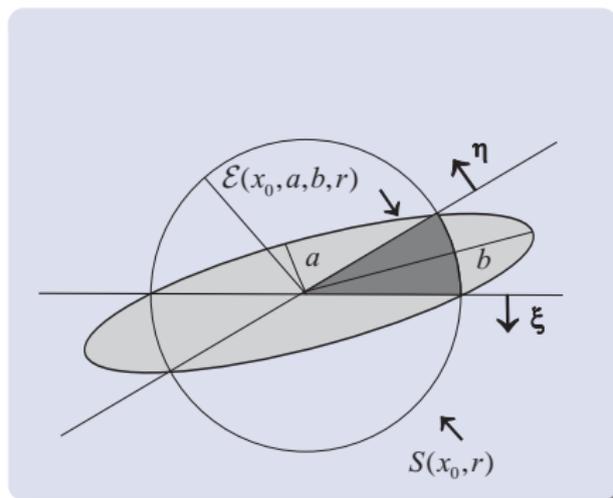
The ratio of 2d-ellipsoid volume to ball volume is equal to

$$q = \frac{\mathsf{vol}(\mathcal{E}(x_0, a, b, r))}{\mathsf{vol}(S(x_0, r))} = \left(\frac{a}{r}\right)\left(\frac{b}{r}\right) = \sqrt{1 - (\xi, \eta)^2}.$$
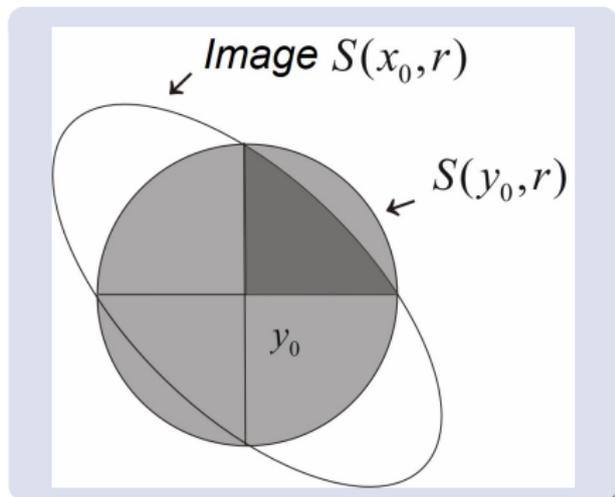
If the angle between vectors ξ and η becomes closer to 180 degrees then the ratio q becomes smaller.

2d-ellipsoid before and after transformation

In the transformed space $Y = T_1(\xi, \eta) X \equiv R^n$

$\mathcal{E}(x_0, a, b, r)$

η

a

b

ξ

$S(x_0, r)$

Image $S(x_0, r)$

$S(y_0, r)$

y_0

2d-ellipsoid $\mathcal{E}(x_0, a, b, r)$

becomes a ball $S(y_0, r)$

Similarity of $T_1(\xi, \eta)$ and Shor's r-algorithm

In the transformed space $Y = T_1(\xi, \eta) X \equiv E^n$

images of vectors ξ and η are orthogonal.

This feature allows to „extend" cone of feasiable directions of the function decrease for the subgradient process in the transformed space of variables, similar to Shor's r-algorithm.

Outline

Accelerated method with Polyak's step

We have $x_0 \in R^n$ and $n \times n$-matrix $B_0 = I_n$. Then we iterate

$$x_{k+1} = x_k - h_k B_k \frac{B_k^T \partial f(x_k)}{\|B_k^T \partial f(x_k)\|}, \quad h_k = \frac{f(x_k) - f^*}{\|B_k^T \partial f(x_k)\|}, \quad (4)$$

$$B_{k+1} = \begin{cases} B_k + (B_k \eta) \xi_{k+1}^T, & \text{if } \mu_k < 0, \\ B_k, & \text{otherwise,} \end{cases} \quad k = 0, 1, 2, \ldots \quad (5)$$

where

$$\mu_k = (\xi_k, \xi_{k+1}), \quad \xi_k = \frac{B_k^T \partial f(x_k)}{\|B_k^T \partial f(x_k)\|}, \quad \xi_{k+1} = \frac{B_k^T \partial f(x_{k+1})}{\|B_k^T \partial f(x_{k+1})\|},$$

$$\eta = \left(\frac{1}{\sqrt{1 - \mu_k^2}} - 1 \right) \xi_{k+1} - \frac{\mu_k}{\sqrt{1 - \mu_k^2}} \xi_k.$$

Decrease of the distance to the minimum point

Theorem 2 (Stetsyuk,1997)

Let $A_k = B_k^{-1}, A_{k+1} = B_{k+1}^{-1}$. The sequence $\{x_k\}_{k=0}^{\infty}$, generated by the method (4)–(5), satisfies the inequalities

$$\|A_{k+1}(x_{k+1}-x^{\star})\|^2 \leq \|A_k(x_k-x^{\star})\|^2 - \frac{(f(x_k)-f^{\star})^2}{\|B_k^T \partial f(x_k)\|^2}, \quad k = 0, 1, 2,$$

Remark. Theorem 2 guarantees that in accelerated subgradient method with Polyak's step the distance to the minimum point decreases monotonically in successively transformed spaces of variables.

On accelerated convergence of method (4)–(5)

If at k-th step the transformation of the space is realized, then

$$\det(B_{k+1}) = \det(B_k)\sqrt{1 - \mu_k^2} = \det(B_k)\sqrt{1 - \cos^2 \varphi_k}.$$

where φ_k is an obtuse angle between two successive subgradients ($\cos^2 \varphi_k = \mu_k < 0$).

It means that the volume of the ellipsoid localizing the point x^* decreases in $q_k = \sqrt{1 - \cos^2 \varphi_k}$.

This provides an accelerated convergence of the method (4)–(5) for ravine convex (smooth or nonsmooth) functions in comparison with the method (2).

Example 1: ravine function f_1

For piecewise linear function

$$f_1(x_1, x_2) = |x_1| + t|x_2|, \ \forall t > 1, \ \forall x_0 = (x_0^{(1)}, x_0^{(2)})$$

the method (4)–(5) finds the minimum point $x^* = (0, 0)$ in no more than three iterations:

1) in one iteration, if $|x_0^{(2)}| = t|x_0^{(1)}|$. No transformation.

2) in two iterations, if $|x_0^{(2)}| < t|x_0^{(1)}|$. One transformation.

3) in three iterations, if $|x_0^{(2)}| > t|x_0^{(1)}|$. One transformation.

Remark. If $|x_0^{(2)}| \neq t|x_0^{(1)}|$, then the method (2) converges at a geometric rate with ratio $q(t) = \sqrt{1 - 1/t^2}$ and requires a significant number of iterations for large values of t.

Example 2: essentially function f_2
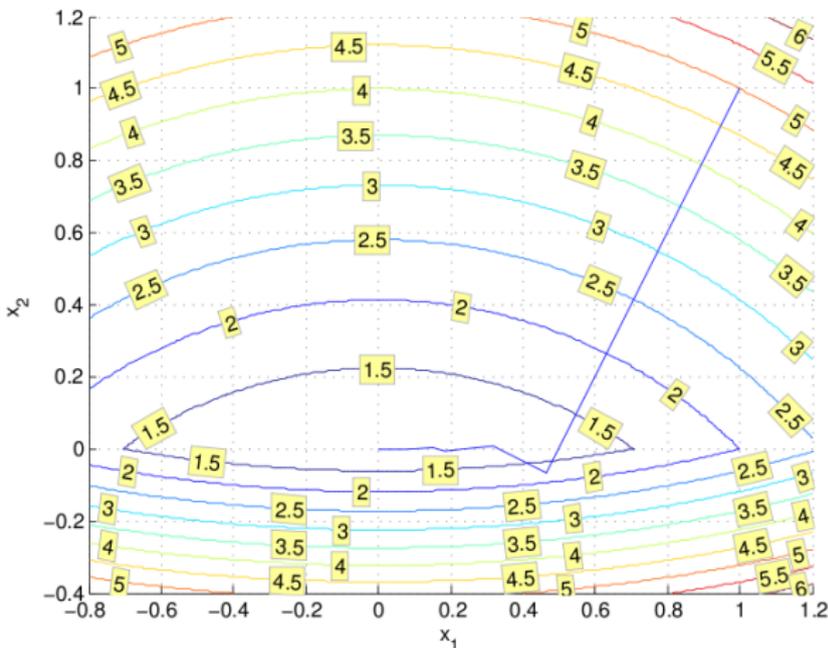
Piecewise quadratic function

$$f_2(x_1, x_2) = \max \left\{ x_1^2 + (2x_2 - 2)^2 - 3,\, x_1^2 + (x_2 + 1)^2 \right\},$$

degenerated at the minimum point $x^\star = (0, 0)$, $f^\star = 1$.

If $x_0 = (1, 1)$, then the method(4)–(5) finds:

- in 16 iterations – the point x_{16}, where $f_2(x_{16}) \leq 1 + 10^{-6}$;
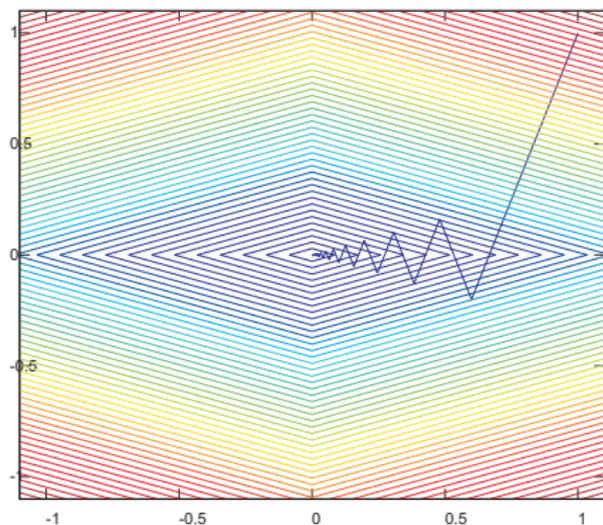- in 31 iterations – the point x_{31}, where $f_2(x_{31}) \leq 1 + 10^{-10}$.
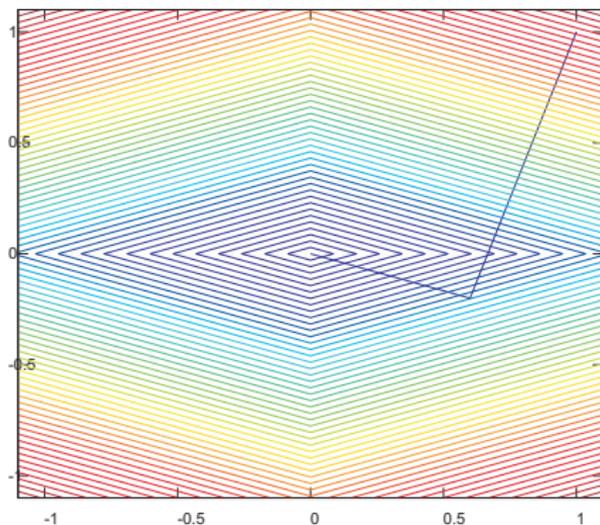
The method (4)–(5) for f_2: $x_0 = (1, 1)$, 31 iter.

Outline

Example 1. $f_1(x_1, x_2)$: piecewise linear function
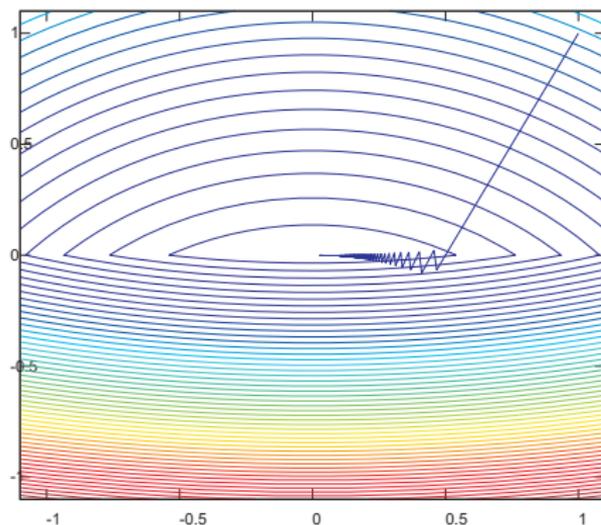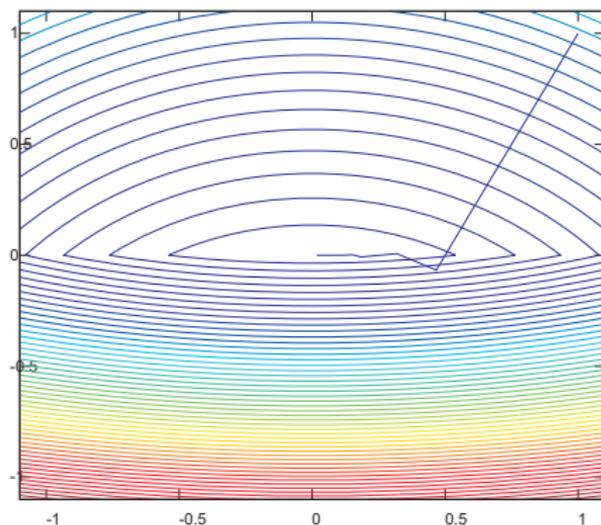

1. Trajectory of Polyak's method   2. ...accelerated Polyak's method

Example 2. $f_2(x_1, x_2)$: piecewise quadratic function

1. Trajectory of Polyak's method   2. ...accelerated Polyak's method

Piecewise linear function, different ravine degrees

eps	$t = 3$		$t = 9$		$t = 27$	
	itn1	**itn2**	**itn1**	**itn2**	**itn1**	**itn2**
1.e-01	31	15	220	37	1645	64
1.e-02	72	24	458	44	3257	73
1.e-03	113	29	695	49	4871	78
1.e-04	155	38	933	54	6481	80
1.e-05	196	43	1170	59	8083	84
1.e-06	237	50	1407	62	9633	91
1.e-07	279	54	1642	66	10000	93
1.e-08	320	59	1874	74	10000	100
1.e-09	362	62	2101	78	10000	108
1.e-10	403	65	2322	85	10000	113

$$f(x_1, \ldots, x_{10}) = \sum_{i=1}^{10} t^{(i-1)/9} |x_i - 1|, \ x_0 = (0, \ldots, 0)^T$$

Conclusion

On the basis of one-rank ellipsoidal operator,
accelerated version of subgradient methods can be
constructed for other techniques of step adjustment.

A remarkable feature of these methods is
automatic choice of space transformation parameters.

* Стецюк П.И. *Методы эллипсоидов и r-алгоритмы.* –
Кишинэу, Эврика, 2014. – 488 с. ▸ Stetsyuk2014

Volkswagen Foundation
grant No 90 306

Questions?

THANK YOU
FOR YOUR ATTENTION!