КІБЕРНЕТИКА та КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

УДК 519.85

DOI:10.34229/2707-451X.22.4.4

В.О. СТОВБА, О.О. ЖМУД

ПРО СУБГРАДІЄНТНІ МЕТОДИ З КРОКОМ ПОЛЯКА ТА ПЕРЕТВОРЕННЯМ ПРОСТОРУ

Розглянуто модифікації субградієнтного методу з кроком Поляка та перетворенням простору для мінімізації спеціальних класів опуклих функцій. Наведено обґрунтування збіжності та швидкості збіжності описаних модифікацій. Описано рекомендації стосовно підбору параметра m *та матриці перетворення простору* B.

Ключові слова: *субградієнтний метод, перетворення простору, яружні функції.*

Вступ. Методи феєрівського типу – це важливий клас методів мінімізації негладких функцій. Вони беруть свій початок у роботах Агмона [1], Моцкіна та Шонберга [2] – це різновид методів субградієнтного спуску з регулюванням кроку в напрямку нормованого субградієнта. В подальшому ці методи розвивались у роботах Б.Т. Поляка [3, 4], звідки й отримали назву субградієнтних методів з кроком Поляка (або Аґмона – Моцкіна – Шонберга). Також було запропоновано узагальнення методів для системи опуклих нерівностей [5] та знаходження спільної точки опуклих множин [6].

Окремою ланкою розвитку методів з кроком Поляка стало використання операції перетворення простору змінних (зокрема, одноразового) для їхнього прискорення при мінімізації яружних функцій [7–10]. В роботі [11] запропонований метод «amsg2p», в якому перетворення простору здійснюється за допомогою двох останніх субградієнтів та агрегатного вектора.

В роботі зроблено огляд модифікацій субградієнтного методу з кроком Поляка, запропонованих в роботі [12], які використовують разове перетворення простору та скалярний параметр $m \geq 1$, який дозволяє мінімізувати спеціальні класи опуклих функцій. Внесено виправлення в обґрунтування збіжності та швидкості збіжності описаних модифікацій для довільних опуклих функцій та опуклих функцій з гострим мінімумом. Наведено рекомендації стосовно вибору параметра m та матриці перетворення простору B.

1. Субградієнтний метод з кроком Поляка. Розглянемо субградієнтний метод, що був запропонований Б.Т. Поляком у роботі [4]. Нехай $f(x)$ – опукла функція, $x \in \mathbb{R}^n$. Позначимо її мінімальне значення як $f^* = f\left(x^*\right)$ та, без обмеження загальності, припустимо, що точка мінімуму x^* єдина. Субградієнт $g_f(x)$ функції $f(x)$ задовольняє умові

$$\left(x - x^*, g_f(x)\right) \geq f(x) - f^*, \ \forall x \in \mathbb{R}^n, \qquad (1)$$

де (x, y) – скалярний добуток векторів x та y. Нерівність (1) випливає з означення субградієнта опуклої функції $f(x)$. Справді, субградієнт $g_f(x)$ у точці x задовольняє нерівність

$$f(y) - f(x) \geq \left(g_f(x), y - x\right), \ \forall y \in \mathbb{R}^n. \tag{2}$$

Ця нерівність виконується також для точки мінімуму x^*:

$$f\left(x^*\right) - f(x) \geq \left(g_f(x), x^* - x\right),$$

яку, зважаючи на те, що $f\left(x^*\right) = f^*$, можна записати як нерівність (1).

Якщо $f(x)$ неперервно-диференційовна в точці $\bar{x}$, то субградієнт $g_f\left(\bar{x}\right)$ визначається однозначно та збігається з $\nabla f\left(\bar{x}\right)$ – градієнтом функції $f(x)$ у точці $\bar{x}$. У точках, де функція $f(x)$ негладка, субградієнт $g_f\left(\bar{x}\right)$ визначається неоднозначно.

Якщо f^* відома, то для знаходження наближення до точки x^* можна використати субградієнтний метод з кроком Поляка [4]:

$$x_{k+1} = x_k - h_k \frac{g_f(x_k)}{\left\| g_f(x_k) \right\|}, \quad h_k = \frac{f(x_k) - f^*}{\left\| g_f(x_k) \right\|}, \quad k = 0, 1, 2, \ldots . \tag{3}$$

Тут крок h_k – це крок Поляка (Агмона – Моцкіна – Шонберга).

Геометричний зміст методу (3) такий. Функція $f(x)$ апроксимується лінійною функцією $f(x) = f(x_k) + \left(g_f(x_k), x - x_k\right)$, і крок вибирається так, щоб значення апроксимуючої функції стало рівним f^* (тобто $f(x_{k+1}) = f^*$). Для опуклої функції $f(x)$ крок h_k визначає таку величину максимального зміщення з точки x_k у напрямку нормалізованого антисубградієнта, для якого умова (1) гарантує, що кут між антисубградієнтом у точці x_k та напрямком від точки x_{k+1} до точки мінімуму x^* не буде тупим.

Це означає такий факт. Субградієнт у точці x_k визначає гіперплощину, яка локалізує точку x^* у напівпросторі у напрямку антисубградієнта. Якщо її перемістити на величину максимального зміщення, визначену кроком h_k, у напрямку нормованого антисубградієнта, то нова гіперплощина локалізуватиме x^* у такому напівпросторі відносно точки x_{k+1}.

Теорема 1 [3]. Послідовність $\{x_k\}_{k=0}^{k^*-1}$, породжена методом (3), задовольняє такі нерівності:

$$\left\| x_{k+1} - x^* \right\|^2 \leq \left\| x_k - x^* \right\|^2 - \frac{\left(f(x_k) - f^*\right)^2}{\left\| g_f(x_k) \right\|^2}, \quad k = 0, 1, 2, \ldots, \tag{4}$$

де k^* – перша ітерація, на якій виконується нерівність $f\left(x_{k^*}\right) - f^* \leq \varepsilon$.

Дана теорема гарантує, що в субградієнтному методі з кроком Поляка відстань до точки мінімуму монотонно зменшується. Крім того, задовольняються такі нерівності:

$$\left(x^* - x_{k+1}, -g_f(x_k)\right) \geq 0, \quad k = 0, 1, 2, \ldots . \tag{5}$$

Вони означають, що для опуклої функції $f(x)$, яка задовольняє умову (1), h_k визначає величину максимального зміщення у напрямку нормованого антисубградієнта. Цим гарантується, що кут між антисубградієнтом і напрямком від точки x_{k+1} до точки мінімуму не буде тупим.

Обґрунтування оцінки швидкості збіжності субградієнтного методу з кроком Поляка для довільних опуклих функцій та опуклих функцій з гострим мінімумом було проведено Б.Т. Поляком у [3].

Теорема 2 [3]. Якщо опукла функція $f(x)$ задовольняє нерівність (1), тоді справедлива рівність $\lim\limits_{k\to\infty} \sqrt{k}\left(f(x_k) - f^*\right) = 0$.

Теорема 3 [3]. Нехай функція $f(x)$ має гострий мінімум, тобто для неї виконується нерівність $f(x) - f^* \ge \alpha\left\|x - x^*\right\|$. Тоді субградієнтний метод з кроком Поляка збігається зі швидкістю геометричної прогресії зі знаменником $q = 1 - \alpha^2/C^2$, де C – константа, що обмежує норму субградієнта $g_f(x_k)$.

2. Субградієнтний метод з кроком Поляка та параметром $m \ge 1$. Розглянемо такий клас опуклих функцій $f(x)$, для субградієнтів $g_f(x)$ яких виконується нерівність

$$\left(x - x^*, g_f(x)\right) \ge m\left(f(x) - f^*\right), \quad \forall x \in \mathbb{R}^n, \text{ де } m \ge 1, \tag{6}$$

яка узагальнює нерівність (1). Тут скалярний параметр m задає величину максимального зміщення по опуклості функції $f(x)$ та вводиться для того, щоб врахувати спеціальні класи опуклих функцій. Наприклад, для кусково-лінійних негладких функцій $m = 1$, для квадратичних гладких $m = 2$, для функцій вигляду $f(x) = \sum_{i=1}^{k}\left|\sum_{j=1}^{n} a_{ij}x_j - b_i\right|^p$, де $p > 1$, значення $m = p$. Параметр m можна використовувати для диференційованих однорідних з показником σ опуклих функцій. Для них використовується рівність $\sigma\left(f(x) - f^*\right) = \left(x - x^*, g_f(x)\right)$, отже параметр m можна вибрати $m = \sigma > 1$. Якщо розглядається довільна опукла функція, то параметр m рівний одиниці.

Якщо опукла функція $f(x)$ задовольняє умові (6) та значення f^* відоме, то для знаходження точки $x_\varepsilon^* \in \mathbb{R}^n$ такої, що $f\left(x_\varepsilon^*\right) \le f^* + \varepsilon$, можна використати такий ітеративний метод.

Ініціалізація. Нехай f^* та $m \ge 1$ відомі. Вибираємо початкову точку $x_0 \in \mathbb{R}^n$, величину $\varepsilon > 0$ та переходимо до наступної ітерації з величиною x_0.

Ітеративний процес. Нехай точку $x_k \in \mathbb{R}^n$ знайдено на k-й ітерації. Для виконання $(k+1)$-ї ітерації виконаємо такі дії.

A1. Обчислимо $f(x_k)$ та $g_f(x_k)$. Якщо $f(x_k) - f^* \le \varepsilon$, тоді ЗУПИНКА $\left(k^* = k,\ x_\varepsilon^* = x_k\right)$.

A2. Обчислимо наступну точку:

$$x_{k+1} = x_k - h_k \frac{g_f(x_k)}{\left\|g_f(x_k)\right\|}, \quad h_k = \frac{m\left(f(x_k) - f^*\right)}{\left\|g_f(x_k)\right\|}. \tag{7}$$

А3. Переходимо до $(k+2)$-ї ітерації з точкою x_{k+1}.

Крок h_k тут задає величину максимального зміщення у напрямку нормованого антисубградієнта, при якому для опуклої функції $f(x)$ умова (6) гарантує, що кут між антисубградієнтом і напрямком від точки x_{k+1} до точки мінімуму не буде тупим.

Варто зазначити, що при $m=1$ метод (7) співпадає з класичним субградієнтним методом з кроком Поляка (3).

Теорема 4 [13]. Послідовність $\{x_k\}_{k=0}^{k^*-1}$, породжена методом (7), задовольняє такі нерівності:

$$\left\|x_{k+1}-x^*\right\|^2 \le \left\|x_k-x^*\right\|^2 - \frac{m^2\left(f(x_k)-f^*\right)^2}{\left\|g_f(x_k)\right\|^2}, \quad k=0,1,2,\dots. \tag{8}$$

Крім того, задовольняються нерівності

$$\left(x^*-x_{k+1},-g_f(x_k)\right)\ge 0, \quad k=0,1,2,\dots. \tag{9}$$

В роботі [14] доведено, що субградієнтний метод з кроком Поляка та параметром $m\ge 1$ володіє такою ж теоретичною оцінкою швидкості збіжності, як і класичний субградієнтний метод з кроком Поляка, а саме: збіжність зі швидкістю $O\left(1/\sqrt{k}\right)$ для довільних опуклих функцій та збіжність зі швидкістю геометричної прогресії для опуклих функцій з гострим мінімумом. Однак це доведення містить певні технічні неточності. Також окрім монотонного зменшення відстані до точки мінімуму в методі (7) легко довести також збіжність цього методу. Зважаючи на вищесказане, наведемо уточнене формулювання та доведення теореми про збіжність та швидкість збіжності методу (7) для довільних опуклих функцій.

Теорема 5. Нехай $f(x)$ – опукла функція, тоді в методі (7) $x_k\to x^*$ при $k\to\infty$. Якщо функція $f(x)$ задовольняє нерівності (6), тоді справедлива рівність $\lim\limits_{k\to\infty}\sqrt{k}\left(f(x_k)-f^*\right)=0$.

Доведення. Оскільки послідовність $\{x_k\}$ обмежена ($\left\|x_k-x^*\right\|\le\left\|x_0-x^*\right\|$), виконується нерівність $\left\|g_f(x_k)\right\|\le C_1$. З нерівності (8) маємо:

$$\frac{m^2\left(f(x_k)-f^*\right)^2}{C_1^2}\le\left\|x_k-x^*\right\|^2-\left\|x_{k+1}-x^*\right\|^2.$$

Просумувавши за індексом k

$$\left(\frac{m}{C_1}\right)^2\sum_{k=0}^{K}\left(f(x_k)-f^*\right)^2\le\left\|x_0-x^*\right\|^2-\left\|x_{K+1}-x^*\right\|^2,$$

маємо:

$$\sum_{k=0}^{\infty}\left(f(x_k)-f^*\right)^2\le\left(\frac{C_1}{m}\right)^2\left\|x_0-x^*\right\|^2. \tag{10}$$

Звідси випливає рівність $\lim\limits_{k\to\infty}\left(f\left(x_k\right)-f^*\right)=0$. З обмеженої послідовності $\left\{x_k\right\}$ можна виділити збіжну підпослідовність $x_{k_i}\to x^*$. З нерівності (8) маємо, що послідовність $\left\{\left\|x_k-x^*\right\|\right\}$ монотонно спадна, а послідовність $\left\{\left\|x_{k_i}-x^*\right\|\right\}\to 0$. Звідси випливає, що $x_k\to x^*$ при $k\to\infty$.

Припустимо, що $\underline{\lim\limits_{k\to\infty}}\sqrt{k}\left(f\left(x_k\right)-f^*\right)>0$. Тоді $f\left(x_k\right)-f^*>a/\sqrt{k}$ при досить великих k та $a>0$, а це суперечить збіжності ряду $\sum\limits_{k=0}^{\infty}\left(f\left(x_k\right)-f^*\right)^2$ (нерівність (10)). Звідси випливає, що $\underline{\lim\limits_{k\to\infty}}\sqrt{k}\left(f\left(x_k\right)-f^*\right)=0$. Теорему 5 доведено.

Варто зауважити, що з рівності $\underline{\lim\limits_{k\to\infty}}\sqrt{k}\left(f\left(x_k\right)-f^*\right)=0$ не випливає нерівність $f\left(x_k\right)-f^*\le C/\sqrt{k}$, тобто збіжність методу (7) зі швидкістю $O\left(1/\sqrt{k}\right)$. Ця рівність виконується для послідовності рекордів функції $f\left(x\right)$ на перших k ітераціях: $f_k^{\min}=\min\limits_{i=0,k}f\left(x_i\right)$. Дійсно, користуючись нерівністю (10), маємо:

$$\left(K+1\right)\left(f_K^{\min}-f^*\right)\le\sum_{k=0}^{K}\left(f\left(x_k\right)-f^*\right)^2\le\left(\frac{C_1}{m}\right)^2\left\|x_0-x^*\right\|^2,$$

звідки отримуємо оцінку

$$f_K^{\min}-f^*\le\frac{\left\|x_0-x^*\right\|^2C_1}{\sqrt{K+1}\,m}\le\frac{\left\|x_0-x^*\right\|^2\max\limits_{k=0,K}\left\|g_f\left(x_k\right)\right\|}{\sqrt{K+1}\,m}=\frac{C}{\sqrt{K+1}},$$

де $C=\dfrac{1}{m}\left\|x_0-x^*\right\|^2\max\limits_{k=0,K}\left\|g_f\left(x_k\right)\right\|$.

У теоремі 6 стверджується збіжність методу (7) із швидкістю геометричної прогресії при мінімізації опуклих функцій з гострим мінімумом.

Теорема 6 [14]. Нехай функція $f\left(x\right)$ має гострий мінімум, тобто для неї виконується нерівність $f\left(x\right)-f^*\ge\alpha\left\|x-x^*\right\|$. Тоді метод (7) збігається зі швидкістю геометричної прогресії зі знаменником $q_1=1-\left(\dfrac{m\alpha}{C_1}\right)^2$, де C_1 – константа, що обмежує норму субградієнта $g_f\left(x\right)$.

3. Субградієнтний метод з кроком Поляка у перетвореному просторі. В основу цього методу покладена одна з ідей Н.З. Шора [15] для прискорення методів, використовуючи лінійні перетворення простору. Зробимо заміну змінних $x=By$, де B – невироджена $n\times n$-матриця (тобто для неї існує обернена матриця $A=B^{-1}$) та введемо функцію $\varphi\left(y\right)=f\left(By\right)$. Субградієнт $g_\varphi\left(y\right)$ опуклої функції $\varphi\left(y\right)$ у точці $y=Ax$ задовольняє умові

$$\left(y-y^*,g_\varphi\left(y\right)\right)\ge\varphi\left(y\right)-\varphi^*,\ \forall y\in\mathbb{R}^n, \qquad (11)$$

аналогічну до умови (1) для функції $f(x)$. Тут $g_\varphi(y) = B^T g_f(x)$, $\varphi^* = \varphi(y^*) = f(By^*) = f(x^*) = f^*$. Дійсно, оскільки $A = B^{-1}$ та $x = By$, нерівність (1) можна переписати у вигляді

$$\left(BB^{-1}\left(x - x^*\right), g_f(x)\right) = \left(A\left(x - x^*\right), B^T g_f(x)\right) \geq f(By) - f\left(By^*\right),\ \forall By \in \mathbb{R}^n,$$

звідки випливає нерівність (11).

Субградієнтний метод з кроком Поляка у перетвореному просторі змінних, визначений невиродженою матрицею B, має такий вигляд:

$$x_{k+1} = x_k - h_k B \frac{B^T g_f(x_k)}{\left\|B^T g_f(x_k)\right\|}, \quad h_k = \frac{f(x_k) - f^*}{\left\|B^T g_f(x_k)\right\|}, \quad k = 0,1,2,\ldots. \tag{12}$$

Тут h_k – крок Поляка (Аґмона – Моцкіна – Шонберга) у перетвореному просторі змінних $y = Ax$. Це випливає з того, що в перетвореному просторі змінних метод (12) можна записати як субградієнтний процес

$$y_{k+1} = y_k - h_k \frac{g_\varphi(y_k)}{\left\|g_\varphi(y_k)\right\|}, \quad h_k = \frac{\varphi(y_k) - \varphi^*}{\left\|g_\varphi(y_k)\right\|}, \quad k = 0,1,2,\ldots. \tag{13}$$

Крок Поляка у перетвореному просторі змінних має ті ж властивості, що й крок Поляка в початковому просторі. Вони визначаються мінімальним значенням функції та нерівністю (11), пов'язаною з φ^*.

Якщо матрицю перетворення простору B покласти рівною одиничній матриці відповідної розмірності, метод (12) збігається з класичним субградієнтним методом з кроком Поляка (у початковому просторі змінних).

Умови використання субградієнтного методу з кроком Поляка у перетвореному просторі збігаються з умовами для класичного методу: опуклість функції $\varphi(y)$ та виконання умови (11). Метод можна використати для знаходження точки $x_\varepsilon^* \in \mathbb{R}^n$, для якої $f\left(x_\varepsilon^*\right) \leq f^* + \varepsilon$. Він описується такою ітеративною процедурою.

Ініціалізація. Маємо відоме значення f^*. Виберемо початкову точку $x_0 \in \mathbb{R}^n$, невироджену $n \times n$-матрицю B та величину $\varepsilon > 0$. Переходимо до наступної ітерації з величиною x_0.

Ітеративний процес. Нехай точку $x_k \in \mathbb{R}^n$ знайдено на k-й ітерації. Для виконання $(k+1)$-ї ітерації виконаємо такі дії.

В1. Обчислимо $f(x_k)$ та $g_f(x_k)$. Якщо $f(x_k) - f^* \leq \varepsilon$, тоді ЗУПИНКА $\left(k^* = k,\ x_\varepsilon^* = x_k\right)$.

В2. Обчислимо наступну точку:

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

В3. Переходимо до $(k+2)$-ї ітерації з точкою x_{k+1}.

Теорема 7 [16]. Послідовність $\{x_k\}_{k=0}^{k^*-1}$, породжена методом (12), задовольняє такі нерівності:

$$\left\|A\left(x_{k+1}-x^*\right)\right\|^2 \le \left\|A\left(x_k-x^*\right)\right\|^2 - \frac{\left(f\left(x_k\right)-f^*\right)^2}{\left\|B^T g_f\left(x_k\right)\right\|^2}, \quad k=0,1,2,\dots. \tag{14}$$

Теорема 7 гарантує, що у субградієнтному методі з кроком Поляка в перетвореному просторі змінних відстань до точки мінімуму зменшується монотонно в перетвореному просторі. Крім того, задовольняються нерівності

$$\left(A\left(x^*-x_{k+1}\right), -B^T g_f\left(x_k\right)\right) \ge 0, \quad k=0,1,2,\dots, \tag{15}$$

які можна переписати як нерівності

$$\left(y^*-y_{k+1}, -g_\varphi\left(y_k\right)\right) \ge 0, \quad k=0,1,2,\dots. \tag{16}$$

Нерівності (16) означають, що для опуклої функції $\varphi(y)$, яка задовольняє умові (11), крок h_k визначає величину максимального зміщення в напрямку нормованого антисубградієнта. Цим гарантується, що кут між антисубградієнтом і напрямком від точки y_{k+1} до точки мінімуму y^* не буде тупим у перетвореному просторі змінних.

Оскільки субградієнтний метод з кроком Поляка у перетвореному просторі співпадає з цим же методом у вихідному просторі змінних, для нього справедливі аналогічні теореми про збіжність та швидкість збіжності для випадку довільних опуклих функцій (теорема 8) та опуклих функцій з гострим мінімумом (теорема 9).

Теорема 8. Нехай $\varphi(y)$ – опукла функція, тоді в методі (13) $y_k \to y^*$ при $k \to \infty$. Якщо функція $\varphi(y)$ задовольняє нерівності (11), тоді справедлива рівність $\varliminf_{k\to\infty} \sqrt{k}\left(\varphi\left(y_k\right)-\varphi^*\right)=0$.

Доведення теореми 8 проводиться аналогічно до доведення теореми 2.

Теорема 9 [14]. Нехай функція $\varphi(y)=f(By)=f(x)$ має гострий мінімум, тобто для неї виконується нерівність $\varphi(y)-\varphi^* \ge \alpha\left\|y-y^*\right\|$. Тоді субградієнтний метод з кроком Поляка у перетвореному просторі збігається зі швидкістю геометричної прогресії із знаменником $q_2 = 1-\frac{\alpha^2}{C_2}$, де C_2 – константа, що обмежує норму субградієнта $g_\varphi(y)$.

3. Субградієнтний метод з кроком Поляка у перетвореному просторі з параметром $m \ge 1$. Для прискорення субградієнтного методу з кроком Поляка у перетвореному просторі для спеціальних класів опуклих функцій можна використати параметр $m \ge 1$. Будемо мінімізувати такі опуклі функції $\varphi(y)=f(By)=f(x)$, для субградієнтів $g_\varphi(y)$ яких виконується узагальнена нерівність

$$\left(y-y^*, g_\varphi(y)\right) \ge m\left(\varphi(y)-\varphi^*\right), \quad \forall y \in \mathbb{R}^n, \tag{17}$$

аналогічна нерівності (11). Легко показати, що значення параметра m інваріантне відносно лінійного перетворення, що задається матрицею B. Звідси випливає, що значення m для функцій $f(x)$ та $\varphi(y)$ співпадають. Субградієнтний метод з кроком Поляка у перетвореному просторі, визначений невиродженою матрицею B, та параметром $m \ge 1$, має такий вигляд:

$$x_{k+1} = x_k - h_k B \frac{B^T g_f\left(x_k\right)}{\left\|B^T g_f\left(x_k\right)\right\|}, \quad h_k = \frac{m\left(f\left(x_k\right)-f^*\right)}{\left\|B^T g_f\left(x_k\right)\right\|}, \quad k=0,1,2,\dots. \tag{18}$$

Тут h_k – крок Поляка (Аґмона – Моцкіна – Шонберга) у перетвореному просторі змінних $y = Ax$, в якому метод (18) можна записати у вигляді

$$y_{k+1} = y_k - h_k \frac{g_\varphi(y_k)}{\left\| g_\varphi(y_k) \right\|}, \quad h_k = \frac{m\left(\varphi(y_k) - \varphi^*\right)}{\left\| g_\varphi(y_k) \right\|}, \quad k = 0,1,2,\ldots. \tag{19}$$

Теорема 10 [16]. Послідовність $\{x_k\}_{k=0}^{k^*-1}$, породжена методом (18), задовольняє нерівності

$$\left\| A\left(x_{k+1} - x^*\right) \right\|^2 \le \left\| A\left(x_k - x^*\right) \right\|^2 - \frac{m^2\left(f(x_k) - f^*\right)^2}{\left\| B^T g_f(x_k) \right\|^2}, \quad k = 0,1,2,\ldots. \tag{20}$$

Метод (19) формулюється аналогічно до методу (7), тому для нього справедливі аналогічні теореми про збіжність та теоретичні оцінки швидкості збіжності для довільної опуклої функції (теорема 11) та опуклої функції з гострим мінімумом (теорема 12). Приведемо їх далі.

Теорема 11. Нехай $\varphi(y)$ – опукла функція, тоді в методі (18) $y_k \to y^*$ при $k \to \infty$. Якщо функція $\varphi(y)$ задовольняє нерівності (17), тоді справедлива рівність $\underline{\lim}_{k \to \infty} \sqrt{k}\left(\varphi(y_k) - \varphi^*\right) = 0$.

Доведення теореми 11 проводиться аналогічно до доведення теореми 5.

Теорема 12 [14]. Нехай функція $\varphi(y) = f(By) = f(x)$ має гострий мінімум, тобто для неї виконується нерівність $\varphi(y) - \varphi^* \ge \alpha \left\| y - y^* \right\|$. Тоді метод (18) збігається зі швидкістю геометричної прогресії із знаменником $q_3 = 1 - \left(\frac{m\alpha}{C_3}\right)^2$, де C_3 – константа, що обмежує норму субградієнта $g_\varphi(y)$.

Рекомендації стосовно підбору параметра m та матриці перетворення простору B. Незважаючи на те, що теоретичні оцінки швидкості збіжності класичного субградієнтного методу з кроком Поляка та вищенаведених модифікацій однакові, використання параметра m та разового перетворення простору для мінімізації різних класів опуклих функцій дозволяє суттєво прискорити класичний метод, що демонструють результати обчислювальних експериментів [12]. Однак актуальним питанням є схема підбору доцільних значень цих параметрів.

Параметр m. Цей параметр – множник, що дозволяє збільшити розмір кроку для спеціальних класів опуклих функцій, забезпечивши прискорену збіжність та невідсікання області, які можуть містити точку мінімуму x^*. Певний перелік таких класів та значень параметра m, які доцільно обирати, наведено в розділі 2. В загальному випадку підбір значень параметра можна виконати емпіричним шляхом: почати зі значення $m = 1$ та поступово його збільшувати, слідкуючи за кількістю ітерацій та збіжністю методу.

Матриця B. Оскільки операція перетворення простору в субградієнтному методі з кроком Поляка спрямована на покращення властивостей функції, що мінімізується (формально на перехід у простір, де властивості функції «кращі», ніж у вихідному), вид матриці перетворення простору B напряму залежить від виду цієї функції. Зокрема, яружні функції характеризуються великим ступенем витягнутості в певних напрямках (одному чи багатьох), тому структура матриці B для таких функцій буде діагональною, де кожен діагональний елемент зменшуватиме ступінь витягнутості у відповідному напрямку. Наприклад, для мінімізації двовимірної функції

$f(x_1, x_2) = x_1^2 + tx_2^2$, $t > 1$, витягнутої у напрямку вісі x_2, доцільно використовувати матрицю вигляду $B = \begin{pmatrix} 1 & 0 \\ 0 & 1/t \end{pmatrix}$. Недіагональні елементи матриці перетворення простору задіюються, коли функція витягнута в напрямку, який є лінійною комбінацією осей $x_1, \ldots, x_n$. В певних випадках, знаючи структуру цільової функції, це дає можливість підібрати матрицю перетворення простору емпіричним шляхом.

Наостанку варто відмітити, що перевагою субградієнтних методів з кроком Поляка у перетвореному просторі є простота й ефективність роботи для мінімізації яружних функції, що залежать від великої кількості змінних. Матриця перетворення простору B – це незмінний вхідний параметр, тому вартість ітерації досить низька. Однак в тих випадках, коли пріоритетом є точність розв’язку, а не обчислювальна вартість, можна скористатись субградієнтними методами з перетворенням простору у напрямку субградієнта (метод еліпсоїдів) [17] та різниці двох послідовних субградієнтів (r-алгоритми) [18]. В цих алгоритмах перерахунок матриці перетворення простору B виконується на кожній ітерації, що дозволяє динамічно підбирати таке перетворення, що зменшує ступінь яружності функції, що мінімізується.

Висновки. Розглянуто модифікації субградієнтного методу з кроком Поляка з використанням параметра $m \geq 1$ та перетворенням простору змінних. Наведено обґрунтування збіжності та швидкості збіжності описаних модифікацій для довільних опуклих функцій та опуклих функцій з гострим мінімумом. Наведено рекомендації стосовно підбору параметра $m \geq 1$ та матриці перетворення простору B.

Подяки. Автори вдячні Миколі Георгійовичу Журбенку за критичні зауваження по матеріалу цієї статті.

Робота підтримана грантом Volkswagen Foundation (грант № 97775).

Список літератури

1. Agmon S. The relaxation method for linear inequalities. *Canad. J. Math.* 1954. No. 6. P. 382−392. https://doi.org/10.4153/CJM-1954-037-2
2. Motzkin T., Schoenberg I. The relaxation method for linear inequalities. *Canad. J. Math.* 1954. No. 6. P. 393−404. https://doi.org/10.4153/CJM-1954-038-x
3. Поляк Б.Т. Введение в оптимизацию. М.: Наука. 1983.384 с.
4. Polyak B.T. Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*. 1969. **9** (3). P. 14–29. https://doi.org/10.1016/0041-5553(69)90061-5
5. Еремин И.И. Обобщение релаксационного метода Агмона-Моцкина. *УМН.* 1965. **20** (122). С. 183–187.
6. Брэгман Л.М. Релаксационный метод нахождения общей точки выпуклых множеств и его применение для решения задач выпуклого программирования. *ЖВМ и МФ.* 1967. **7** (3). С. 200–217. https://doi.org/10.1016/0041-5553(67)90040-7
7. Стецюк П.И. r-алгоритмы и эллипсоиды. *Кибернетика и системный анализ*. 1996. № 1. С. 113–134. https://doi.org/10.1007/BF02366587
8. Стецюк П.И. Ортогонализующие линейные операторы в выпуклом программировании. I. *Кибернетика и системный анализ*. 1997. № 3. С. 97–119. https://doi.org/10.1007/BF02733072
9. Стецюк П.И. Ортогонализующие линейные операторы в выпуклом программировании. II. *Кибернетика и системный анализ*. 1997. № 5. С. 111–124. https://doi.org/10.1007/BF02667194
10. Гершович В.И., Шор Н.З. Метод эллипсоидов, его обобщения и приложения. *Кибернетика*. 1982. № 5. С. 61–69. https://doi.org/10.1007/BF01068741
11. Стецюк П.И. Метод amsg2p для овражных выпуклых функций. *Информационный бюллетень Ассоциации математического программирования*. 2011. № 12. С. 57–58.

12. Стовба В.О. Субградієнтний метод з кроком Поляка у перетвореному просторі : дис. на здобуття наук. ступеня д-ра філософії : 113 Прикладна математика / Інститут кібернетики імені В.М. Глушкова НАН України, Київ, 2021. 184 с.
13. Стецюк П.І. Ускорение субградиентного метода Поляка. *Теорія оптимальних рішень*. 2012. № 11. С. 151–60. http://dspace.nbuv.gov.ua/handle/123456789/85030
14. Стецюк П.І., Стовба В.О., Жмуд О.О. Про швидкість збіжності субградієнтних методів з кроком Поляка. *Наук. вісник Ужгород. ун-ту. Сер. матем. і інформ*. 2019. **1** (34). С. 94–101. https://doi.org/10.24144/2616-7700.2019.1(34).94-101
15. Сергиенко И.В., Стецюк П.И. О трех научных идеях Н.З. Шора. *Кибернетика и системный анализ*. 2012. № 1. С. 4–22. https://doi.org/10.1007/s10559-012-9387-x
16. Stetsyuk P., Stovba V., Chernousova Z. Subgradient method with Polyak's step in transformed space. *Optimization and Applications (OPTIMA 2018)*. 2018. Vol. 974. P. 49–63. https://doi.org/10.1007/978-3-030-10934-9_4
17. Шор Н.З. Метод отсечения с растяжением пространства для решения задач выпуклого программирования. *Кибернетика.* 1977. № 1. С. 94–95. https://doi.org/10.1007/BF01071394
18. Шор Н.З., Журбенко Н.Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов. *Кибернетика*. 1971. № 3. С. 51–59. https://doi.org/10.1007/BF01070454

Одержано 17.12.2022

Стовба Віктор Олександрович,
доктор філософії, молодший науковий співробітник
Інституту кібернетики імені В.М. Глушкова НАН України, Київ,
https://orcid.org/0000-0003-3023-5815
vik.stovba@gmail.com

Жмуд Олександр Олексійович,
провідний інженер-програміст
Інституту кібернетики імені В.М. Глушкова НАН України, Київ.
https://orcid.org/0000-0002-4591-1110
zhmud17@gmail.com

УДК 519.85

В.О. Стовба [*]**, О.О. Жмуд**

Про субградієнтні методи з кроком Поляка та перетворенням простору

Інститут кібернетики імені В.М. Глушкова НАН України, Київ
[*] *Листування:* vik.stovba@gmail.com

Вступ. Мінімізація яружних опуклих функцій, як гладких, так і негладких, виникає в багатьох задачах планування, керування, аналізу стійкості динамічних систем, штучному інтелекті та машинному навчанні. Тому розвиток нових та покращення наявних методів є важливою задачею, зважаючи на те, що дедалі частіше функції, що мінімізуються, залежать від великої кількості змінних.

Особливої уваги заслуговують методи, що використовують операцію лінійного перетворення простору, які дозволяють покращити властивості цільової функції і значно прискорити її мінімізацію. Відомі методи такого типу, зокрема методи еліпсоїдів та r-алгоритми, потребують перерахунку матриці перетворення на кожній ітерації. Тому розвиток методів з разовим перетворенням простору, що мають меншу ціну ітерації, є безумовно актуальною задачею.

Мета роботи. Огляд модифікацій субградієнтного методу з кроком Поляка, що використовують скалярний параметр $m \geq 1$ та операцію разового перетворення простору. Доповнити обґрунтування збіжності та швидкості збіжності описаних модифікацій. Навести рекомендації стосовно підбору параметру m та матриці перетворення простору B.

Результати. Субградієнтний метод з кроком Поляка у перетвореному просторі та параметром $m>1$ є ефективним методом мінімізації гладких та негладких опуклих функцій, які мають яружну структуру.

Підбір доцільного значення параметра $m \geq 1$ та матриці перетворення простору B дозволяє суттєво прискорити цей метод і використовувати його для мінімізації функцій, що залежать від великої кількості змінних.

Висновки. Розвиток швидких методів мінімізації негладких опуклих функцій багатьох змінних з яружною структурою дозволяє ефективно розв'язувати сучасні задачі штучного інтелекту, зокрема задачі машинного навчання, розпізнавання образів, аналізу даних великих об'ємів тощо.

Ключові слова: субградієнтний метод, перетворення простору, яружні функції.

MSC 46N10, 65K05, 90C25, 15A04

Viktor Stovba [*], Oleksandr Zhmud

On Subgradient Methods with Polyak's Step and Space Transformation

V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv
[*] *Correspondence:* vik.stovba@gmail.com

Introduction. Minimization of ravine convex functions, both smooth and non-smooth, arises in many problems of planning, control, stability analysis of dynamic systems, artificial intelligence, and machine learning. Therefore, the development of new and improvement of existing methods is an important task, taking into account the fact that more and more frequently the functions to be minimized depend on a large number of variables.

Special attention should be paid to the methods using the operation of linear transformation of space, which allow to improve the properties of the objective function and significantly accelerate its minimization. Known methods of this type, in particular, ellipsoid methods and *r*-algorithms, require recalculation of the transformation matrix at each iteration. Therefore, the development of methods with a one-time transformation of space, which have a lower cost of iteration, is definitely an urgent task.

The purpose of the article is to review modifications of subgradient method with Polyak's step that use a scalar parameter $m \geq 1$ and one-time space transformation operation. Supplement the justification of the convergence and convergence rate of the modifications described. Give recommendations regarding determination of parameter m and space transformation matrix B.

Results. Subgradient method with Polyak's step in transformed space and parameter $m>1$ is an effective method for minimizing smooth and non-smooth convex functions that have a ravine structure. Determination of an appropriate value of the parameter $m \geq 1$ and space transformation matrix B allows to significantly accelerate this method and use it for minimization functions that depend on a large number of variables.

Conclusions. The development of fast methods for minimization of non-smooth convex functions of many variables with a ravine structure makes it possible to effectively solve modern problems of artificial intelligence, in particular, the problems of machine learning, image recognition, big data analysis, etc.

Keywords: subgradient method, space transformation, ravine functions.