УДК 519.95
DOI https://doi.org/10.24144/2616-7700.2019.1(34).94-101

П. І. Стецюк[1], В. О. Стовба[2], О. О. Жмуд[3]

[1] Інститут кібернетики ім. В.М. Глушкова НАН України, Київ ,
завідувач відділу методів негладкої оптимізації,
доктор фізико-математичних наук
stetsyukp@gmail.com
ORCID: https://orcid.org/0000-0003-4036-2543

[2] Інститут кібернетики ім. В.М. Глушкова НАН України, Київ,
аспірант
vik.stovba@gmail.com
ORCID: https://orcid.org/0000-0003-3023-5815

[3] Інститут кібернетики ім. В.М. Глушкова НАН України, Київ,
аспірант
zhmud17@gmail.com
ORCID: https://orcid.org/0000-0002-4591-1110

ПРО ШВИДКІСТЬ ЗБІЖНОСТІ СУБГРАДІЄНТИХ МЕТОДІВ З КРОКОМ ПОЛЯКА

Розглядаються субградієнтні методи (метод **A** та метод **B**) для знаходження точки мінімуму опуклої функції, якщо відоме її мінімальне значення. Обидва методи використовують відомий крок Поляка, який залежить від скалярного параметра $m \geq 1$ і гарантує монотонне зменшення відстані до точки мінімуму. Якщо $m = 1$, то методи **A** та **B** застосовні для довільної опуклої функції. Параметр $m > 1$ дозволяє врахувати спеціальні класи опуклих функцій – опуклі квадратичні функції ($m = 2$), диференційовні однорідні функції з показником σ ($m = \sigma$) та інші. Для обох методів доведено теореми про їх збіжність зі швидкістю $O\left(1/\sqrt{k}\right)$ для довільної опуклої функції та збіжність зі швидкістю геометричної прогресії для опуклої функції з гострим мінімумом.

Метод **A** є субградієнтним методом з кроком Поляка у вихідному просторі змінних. Параметр $m = 1$ відповідає класичному кроку Поляка для довільної опуклої функції. Параметр $m = 2$ можна використовувати при мінімізації квадратичних функцій, для яких величина кроку Поляка збільшується в два рази порівняно з класичним кроком Поляка при $m = 1$. Для одновимірних квадратичних функцій метод **A** знаходить точку мінімуму за один крок з довільної стартової точки, що відповідає одній ітерації методу Ньютона. Доведення теорем про швидкість збіжності методу **A** базується на характеристиках опуклої функції в вихідному просторі змінних, серед яких ключову роль відіграє константа c_1, яка обмежує норму субградієнта.

Метод **B** – субградієнтний метод з перетворенням простору, де крок Поляка обчислюється у перетвореному лінійним оператором просторі змінних. Метод визначається невиродженою матрицею B. Параметр $m = 1$ відповідає класичному кроку Поляка для довільної опуклої функції у перетвореному просторі змінних. Якщо $m = 2$, то для одновимірної квадратичної функції метод **B** знаходить точку мінімуму за один крок при довільних матриці B та стартовій точці. Доведення теорем про швидкість збіжності методу **B** базуються на характеристиках функції в перетвореному просторі змінних, аналогічних характеристикам функцій у початковому просторі для методу **A**.

Ключові слова: субградієнтний метод, крок Поляка, перетворення простору, швидкість збіжності, гострий мінімум.

1. Вступ. Нехай $f(x)$ – опукла функція, $x \in \mathbb{R}^n$. Позначимо її мінімальне значення як $f^* = f(x^*)$ та припустимо, що точка мінімуму x^* єдина. Субградієнт $g_f(x)$ задовольняє наступну умову:

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

Тут (x, y) – скалярний добуток векторів $x \in \mathbb{R}^n$ та $y \in \mathbb{R}^n$.

Якщо $f(x)$ є неперервно-диференційовною в точці $\overline{x}$, то субградієнт $g_f(\overline{x})$ однозначно визначається і збігається з $\nabla f(\overline{x})$ – градієнтом функції $f(x)$ в точці $\overline{x}$. В точках, де функція $f(x)$ негладка, субградієнт $g_f(\overline{x})$ визначається неоднозначно.

Нерівність (1) випливає з означення субградієнта $g_f(x)$ опуклої функції $f(x)$. Справді, субградієнт $g_f(x)$ в точці x задовольняє нерівність

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

Нерівність (2) виконується також для точки мінімуму x^*, для якої вона перетворюється на нерівність $f(x^*) - f(x) \geq (g_f(x), x^* - x)$, яку, враховуючи рівність $f(x^*) = f^*$, можна записати як нерівність (1).

Якщо значення f^* відоме, то для знаходження наближення до точки x^* можна використати субградієнтний метод Б.Т. Поляка [1]:

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

Крок h_k називається кроком Поляка (або кроком Агмона-Моцкіна-Шонберга). Вперше крок Поляка використовувався для мінімізації кусково-лінійних опуклих функцій. Наприклад, у 1954 році Аґмон (Agmon S.) [2] та Моцкін (Motzkin T. S.) і Шонберг (Schoenberg I. J.) [3] використали цей крок у релаксаційному методі для знаходження хоча б одного розв'язку сумісної системи лінійних нерівностей. У 1965 р. І. І. Єрьомін [4] узагальнив цей релаксаційний метод для системи опуклих функцій.

У цій статті розглядаються два субградієнтних методи з кроком Поляка (метод **A** та метод **B**) для більш загального випадку опуклої функції $f(x)$, коли її субградієнт $g_f(x)$ задовольняє наступну умову:

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

де параметр $m \geq 1$. Параметр m вводиться для того, щоб врахувати спеціальні класи опуклих функцій: наприклад, для квадратичних гладких функцій $m = 2$, для функцій виду $f(x) = \sum_{i=1}^{k} \left| \sum_{j=1}^{n} a_{ij} x_j - b_i \right|^p$, де $p > 1$, $m = p$. Параметр m можна активно використовувати для диференційованих однорідних з показником σ опуклих функцій. Для них виконується рівність $\sigma\left(f(x) - f^*\right) = (x - x^*, g_f(x))$, отже, параметр m можна вибрати $m = \sigma > 1$.

2. Субградієнтний метод Поляка (метод A). Нехай опукла функція $f(x)$ задовольняє умову (4) та значення f^* відоме. Тоді для знаходження точки $x_\varepsilon^* \in \mathbb{R}^n$ такої, що $f(x_\varepsilon^*) \leq f^* + \varepsilon$, можна використати такий ітеративний метод, який будемо називати методом **A**.

Крок 0. Відомі f^* та $m \geq 1$. Вибираємо стартову точку $x_0 \in \mathbb{R}^n$, величину $\varepsilon > 0$ та переходимо до наступної ітерації з величиною x_0.

Нехай точка $x_k \in \mathbb{R}^n$ знайдена на k-й ітерації.

Крок 1. Обчислимо $f(x_k)$ та $g_f(x_k)$. Якщо $f(x_k) - f^* \leq \varepsilon$, тоді STOP ($k^* = k,\ x^*_\varepsilon = x_k$).

Крок 2. Обчислимо наступну точку

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

Крок 3. Переходимо до **Кроку 1** зі значеннями $(k+1)$ та x_{k+1}.

Теорема 1. *[5] Послідовність $\{x_k\}_{k=0}^{k^*}$, породжена методом **A**, задовольняє такі нерівності:*

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

Теорема 1 гарантує, що в субградієнтному методі Поляка відстань до точки мінімуму монотонно зменшується. Крім того, задовольняються такі нерівності:

$$(x^* - x_{k+1}, -g_f(x_k)) \geq 0, \quad k = 0, 1, \ldots \tag{6}$$

Дійсно, нерівності (6) випливають з того факту, що, використовуючи (4), маємо

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

$$= (x_k - x^*, g_f(x_k) - h_k \|g_f(x_k)\|) = (x_k - x^*, g_f(x_k)) - m(f(x_k) - f^*) \geq 0.$$

Нерівності (6) означають, що для опуклої функції $f(x)$, яка задовольняє умову (4), h_k визначає таку величину максимального зміщення з точки x_k в напрямку нормованого антисубградієнта, для якого умова (1) гарантує, що кут між антисубградієнтом в 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 у напрямку нормованого антисубградієнта, то нова гіперплощина буде локалізувати x^* у такому напівпросторі по відношенню до точки x_{k+1}.

Якщо $m = 1$, то метод **A** співпадає з субградієнтним методом (3). Оцінимо швидкість збіжності методу **A** аналогічно до того, як це було виконано Б. Т. Поляком для методу (3) [6, ст. 132]. Для цього нам знадобиться наступна лема.

Лема 1. *[6, ст. 121] Субградієнти опуклої функції $f(x)$ є обмеженими на довільній обмеженій множині або множині виду $\{x :\ f(x) \leq \alpha\}$.*

Теорема 2. *Якщо опукла функція $f(x)$ задовольняє нерівність (4), тоді справедлива рівність* $\lim\limits_{k\to\infty} \sqrt{k}\,(f(x_k) - f^*) = 0$.

Доведення. З нерівності (5) теореми 1 випливає, що послідовність $h_k^2 = \frac{m^2(f(x_k)-f^*)^2}{\left\|g_f(x_k)\right\|^2}$ збіжна. Враховуючи те, що послідовність x_k обмежена, з леми 1 маємо $\|g_f(x_k)\| \leq c_1$, де c_1 – константа, що обмежує норму субградієнта $g_f(x)$. Звідси випливає збіжність послідовності $(f(x_k) - f^*)^2$. Нехай $\lim\limits_{k\to\infty} \sqrt{k}\,(f\,(x_k) - f^*) > 0$. Тоді $f\,(x_k) - f^* > {}^{a}\big/\sqrt{k}$ при досить великих k та $a > 0$, а це суперечить збіжності послідовності $(f(x_k) - f^*)^2$. Звідси випливає, що $\lim\limits_{k\to\infty} \sqrt{k}\,(f\,(x_k) - f^*) = 0$. Теорема 2 доведена.

У теоремі 2 наведена збіжність методу **A** зі швидкістю $O\left(1/\sqrt{k}\right)$. Якщо функція $f(x)$ має гострий мінімум, то можна досягти геометричної швидкості збіжності для методу **A**.

Теорема 3. *Нехай для функції $f(x)$ виконується нерівність $f(x) - f^* \geq \geq \alpha \|x - x^*\|$. Тоді метод **A** сходиться зі швидкістю геометричної прогресії зі знаменником $q_1 = 1 - \left(\frac{m\alpha}{c_1}\right)^2$, де c_1 – константа, що обмежує норму субградієнта $g_f(x)$.*

Доведення. Враховуючи те, що виконується нерівність (5) та нерівність для $f(x)$ в умові теореми, маємо:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{m^2(f(x_k) - f^*)^2}{\|g_f(x_k)\|^2} \leq \|x_k - x^*\|^2 - \frac{m^2\alpha^2 \|x_k - x^*\|^2}{\|g_f(x_k)\|^2}.$$

Враховуючи нерівність $\|g_f(x_k)\| \leq c_1$, маємо:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{m^2\alpha^2 \|x_k - x^*\|^2}{\|g_f(x_k)\|^2} \leq \|x_k - x^*\|^2 - m^2\frac{\alpha^2}{c_1^2} \|x_k - x^*\|^2 =$$
$$= q_1 \|x_k - x^*\|^2,$$

де $q_1 = 1 - \left(\frac{m\alpha}{c_1}\right)^2$, що і означає збіжність методу зі швидкістю геометричної прогресії зі знаменником q_1. Теорема 3 доведена.

Використання параметра $m > 1$ дає можливість застосовувати метод **A** для спеціальних класів опуклих функцій. Наприклад, використовуючи значення параметра $m = 2$, метод **A** знаходить точку мінімуму одновимірної квадратичної функції за один крок з довільної стартової точки, що відповідає одній ітерації методу Ньютона. Якщо застосувати значення параметру $m = 1$ метод **A** збігатиметься зі швидкістю методу дихотомії.

3. Субградієнтний метод Поляка у перетвореному просторі (метод B). Розглянемо субградієнтний метод Поляка у перетвореному просторі змінних $y = Ax,\ \ A = B^{-1}$, де B є невиродженою $n \times n$-матрицею. Зробимо заміну змінних $x = By$. Субградієнт $g_\varphi(x)$ опуклої функції $\varphi(y) = f(By)$ в точці $y = Ax$ задовольняє наступну умову:

$$(y - y^*, g_\varphi(y)) \geq m(\varphi(x) - \varphi^*),\ \forall y \in \mathbb{R}^n, \tag{7}$$

де $g_\varphi(y) = B^T g_f(x)$, $\varphi^* = \varphi(y^*) = y^* = Ax^*$. Дійсно, оскільки $A = B^{-1}$ та $x = By$, нерівність (4) можна переписати у вигляді

$$(A(x - x^*), B^T g_f(x)) \geq m(f(By) - f(By^*)),\quad \forall By \in \mathbb{R}^n,$$

звідки отримуємо нерівність (7).

Субградієнтний метод з кроком Поляка у перетвореному просторі (визначений невиродженою матрицею B) має наступний вигляд:

$$x_{k+1} = x_k - h_k B \frac{B^T g_f(x_k)}{\|B^T g_f(x_k)\|}, \quad h_k = \frac{m(f(x_k) - f^*)}{\|B^T g_f(x_k)\|}, \quad k = 0, 1, 2 \dots \tag{8}$$

Тут h_k є кроком Поляка (Аґмона-Моцкіна-Шонберга) в перетвореному просторі змінних $y = Ax$. Це випливає з того, що в перетвореному просторі змінних метод (8) записується як субградієнтний процес

$$y_{k+1} = y_k - h_k \frac{g_\varphi(y_k)}{\|g_\varphi(y_k)\|}, \quad h_k = \frac{m(\varphi(y_k) - \varphi^*)}{\|g_\varphi(y_k)\|}, \quad k = 0, 1, 2 \dots \tag{9}$$

Крок Поляка у перетвореному просторі змінних має ті ж самі властивості, що й крок Поляка в вихідному просторі. Вони визначаються мінімальним значенням функції $\varphi^* = f^*$ та нерівністю (7) для φ^*.

Щоб знайти точку $x^*_\varepsilon \in \mathbb{R}^n$, для якої $f(x^*_\varepsilon) \le f^* + \varepsilon$, можна використати субградієнтний метод з кроком Поляка у перетвореному просторі з критерієм зупинки $f(x_k) - f^* \le \varepsilon$, що описується такою ітеративною процедурою.

Крок 0. Відомі f^* та $m \ge 1$. Вибираємо стартову точку $x_0 \in \mathbb{R}^n$, невироджену $n \times n$-матрицю B і величину $\varepsilon > 0$. Переходимо до наступної ітерації з величиною x_0.

Нехай точка $x_k \in \mathbb{R}^n$ знайдена на k-й ітерації.

Крок 1. Обчислимо $f(x_k)$ та $g_f(x_k)$. Якщо $f(x_k) - f^* \le \varepsilon$, тоді STOP ($k^* = k,\ x^*_\varepsilon = x_k$).

Крок 2. Обчислимо наступну точку

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

Крок 3. Переходимо до **Кроку 1** зі значеннями $(k+1)$ та x_{k+1}.

Теорема 4. *[7] Послідовність $\{x_k\}_{k=0}^{k^*}$, породжена методом $\boldsymbol{B}$, задовольняє нерівності*

$$\|A(x_{k+1} - x^*)\|^2 \le \|A(x_k - x^*)\|^2 - \frac{m^2(f(x_k) - f^*)^2}{\|B^T g_f(x_k)\|^2}, \quad k = 0, 1, 2, \dots, k^* - 1. \tag{10}$$

Теорема 4 гарантує, що в субградієнтному методі з кроком Поляка у перетвореному просторі змінних відстань до точки мінімуму зменшується монотонно. Крім того, задовольняються такі нерівності:

$$(A(x^* - x_{k+1}), -B^T g_f(x_k)) \ge 0,\ k = 0, 1, \dots, \tag{11}$$

які можуть бути переписані як нерівності

$$(y^* - y_{k+1}, -g_\varphi(y_k)) \ge 0,\ k = 0, 1, \dots. \tag{12}$$

Дійсно, нерівності (11) випливають з того факту, що, використовуючи (7) і (9), маємо:

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

$$= (x_k - x^*, g_f(x_k)) - h_k \left\| B^T g_f(x_k) \right\| = (x_k - x^*, g_f(x_k)) - m(f(x_k) - f^*) \geq 0.$$

Нерівності (12) означають, що для опуклої функції $\varphi(y)$, яка задовольняє умову (7), крок h_k визначає величину максимального переміщення 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.

У теоремах 5 та 6, які сформульовані аналогічно до теорем 2 і 3, наводяться оцінки збіжності методу **B** для довільної опуклої функції та опуклої функції, що має гострий мінімум. Для їх доведення нам знадобиться наступна лема.

Лема 2. *Субградієнти опуклої функції $\varphi(y)$ є обмеженими на довільній обмеженій множині або множині виду $\{y : \varphi(y) \leq \alpha\}$.*

Теорема 5. *Якщо опукла функція $\varphi(y)$ задовольняє нерівність* (7)*, тоді справедлива рівність* $\lim\limits_{k\to\infty} \sqrt{k}\,(\varphi(y_k) - \varphi^*) = 0$.

Доведення. З нерівності (10) теореми 4 випливає, що послідовність $h_k^2 = \frac{m^2(f(x_k)-f^*)^2}{\left\|B^T g_f(x_k)\right\|^2} = \frac{m^2(\varphi(y_k)-\varphi^*)^2}{\|g_\varphi(y_k)\|^2}$ збіжна. Зважаючи на те, що послідовність y_k збіжна, з леми 2 випливає нерівність $\|g_\varphi(y_k)\| \leq c_2$, де c_2 – константа, що обмежує норму субградієнта $g_\varphi(y)$. Звідси маємо, що послідовність $(\varphi(y_k) - \varphi^*)^2$ збіжна. Нехай $\lim\limits_{k\to\infty} \sqrt{k}\,(\varphi(y_k) - \varphi^*) > 0$. Тоді $\varphi(y_k) - \varphi^* > a/\sqrt{k}$ при досить великих k та $a > 0$, а це суперечить збіжності послідовності $(\varphi(y_k) - \varphi^*)^2$. Звідси випливає, що $\lim\limits_{k\to\infty} \sqrt{k}\,(\varphi(y_k) - \varphi^*) = 0$. Теорема 5 доведена.

Теорема 6. *Нехай для функції $\varphi(y) = f(By) = f(x)$ виконується нерівність $\varphi(y) - \varphi^* \geq \alpha \|y - y^*\|$. Тоді метод **B** сходиться зі швидкістю геометричної прогресії зі знаменником $q_2 = 1 - \left(\frac{m\alpha}{c_2}\right)^2$, де c_2 – константа, що обмежує норму субградієнта $g_\varphi(y)$.*

Доведення. Зважаючи на те, що $\varphi(y) = f(By) = f(x)$ та $g_\varphi(y_k) = B^T g_f(x_k)$, нерівність (10) можна переписати так:

$$\|y_{k+1} - y^*\|^2 \leq \|y_k - y^*\|^2 - \frac{m^2(\varphi(y_k) - \varphi^*)^2}{\|g_\varphi(y_k)\|^2}.$$

Звідси, враховуючи нерівність для $\varphi(y)$ в умові теореми, маємо:

$$\|y_{k+1} - y^*\|^2 \leq \|y_k - y^*\|^2 - \frac{m^2(\varphi(y_k) - \varphi^*)^2}{\|g_\varphi(y_k)\|^2} \leq \|y_k - y^*\|^2 - \frac{m^2\alpha^2 \|y_k - y^*\|^2}{\|g_\varphi(y_k)\|^2}.$$

Зважаючи на нерівність $\|g_\varphi(y_k)\| \leq c_2$, маємо:

$$\|y_{k+1} - y^*\|^2 \leq \|y_k - y^*\|^2 - \frac{m^2\alpha^2 \|y_k - y^*\|^2}{\|g_\varphi(y_k)\|^2} \leq \|y_k - y^*\|^2 - m^2\frac{\alpha^2}{c_2^2} \|y_k - y^*\|^2 =$$

$$= q_2 \|y_k - y^*\|^2$$

де $q_2 = 1 - \left(\frac{m\alpha}{c_2}\right)^2$, що і означає збіжність зі швидкістю геометричної прогресії зі знаменником q_2. Теорема 6 доведена.

4. Висновки. В статті розглянуто два субградієнтних методи (**A** та **B**) для знаходження точки мінімуму опуклої функції з відомим мінімальним значенням. Метод **A** використовує крок Поляка у вихідному просторі змінних, метод **B** – у перетвореному просторі змінних. Доведено теореми про швидкість збіжності обох методів для довільної опуклої функції та опуклої функції з гострим мінімумом. Зауважимо, що метод **B**, будучи субградієнтним у перетвореному просторі змінних, дає можливість моделювати у вихідному просторі методи з кроком Поляка не по субградієнту, а по лінійній комбінації двох або більшої кількості субградієнтів.

Список використаної літератури

1. Поляк Б. Т. Минимизация негладких функционалов. *Журнал вычислительной математики и математической физики.* 1969. Том № 9 (3). С. 509–521.
2. Agmon S. The relaxation method for linear inequalities. *Canadian Journal of Mathematics.* Vol. 6. 1954. P. 382–392. doi: https://doi.org/10.4153/CJM-1954-037-2
3. Motzkin T., Schoenberg I. J. The relaxation method for linear inequalities. *Canadian Journal of Mathematics.* Vol. 6. 1954. P. 393–404. doi: https://doi.org/10.4153/CJM-1954-038-x
4. Еремин И. И. Обобщение релаксационного метода Агмона-Моцкина. *Успехи математических наук.* 1965. т. XX., Вып. 2 (122). С. 183–187.
5. Стецюк П .И. Ускорение субградиентного метода Поляка. *Теорія оптимальних рішень.* 2012. №11. С. 151–160.
6. Поляк Б. Т. Введение в оптимизацию. М.: Наука, 1983. 384 с.
7. Stetsyuk P., Stovba V., Chernousova Z. Subgradient method with Polyak's step in transformed space. *Optimization and Applications. OPTIMA 2018.* 2018. Vol. 974. Springer, Cham. P. 49–63. doi: https://doi.org/10.1007/978-3-030-10934-9_4

Stetsyuk P. I., Sovba V. O., Zhmud O. O. On speed of convergence of subgradient methods with Polyak step.

Subgradient methods (method A and method B) are considered for finding the minimum point of a convex function for the known optimal value of the function. Both methods use known Polyak's step, which depends on scalar parameter $m \geq 1$ and guarantees monotonous decrease of the distance to the minimum point. If $m = 1$ the methods A and B are applicable to arbitrary convex function. Parameter $m > 1$ permits to take into account special classes of convex functions – convex quadratic functions ($m = 2$), differentiable homogeneous of degree σ ($m = \sigma$), etc. For both methods theorems are proven about the convergence rate $O\left(1/\sqrt{k}\right)$ for arbitrary convex function and convergence with geometric progression rate for a convex function with acute minimum.

Method A is a subgradient method with Polyak's step in original space of variables. The parameter $m = 1$ corresponds to classical Polyak's step for arbitrary convex function. Parameter $m = 2$ can be used for quadratic functions minimization, for which the value of Polyak's step is doubled comparing to the classical Polyak's step with $m = 1$. For one-dimentional functions method A finds the minimum point in one step from the arbitrary start point, which corresponds to one iteration of the Newton's method. Proof of theorems about convergence rate of the method A is based on convex function characteristics in original space of variables, among which a constant c_1 plays a crucial role – it bounds a subgradient norm.

Method B is a subgradient method with space transformation, where the Polyak's step is calculated in space transformed by linear operator. The method is defined by nonsingular matrix B. The parameter $m = 1$ corresponds to classical Polyak's step for arbitrary convex function in transformed space of variables. If $m = 2$, for one-dimentional quadratic function, the method B finds the minimum point in one step with both arbitrary matrix B and start point. Proof of theorems about convergence rate of the method B is based on function characteristics in transformed space of variables, which are analogous to the method A characteristics in original space.

Keywords: subgradient method, the Polyak's step, space transformation, convergence rate, acute minimum.

References

1. Polyak, B. T. (1969). Minimizatsiya negladkih funktsionalov [Minimization of non-smooth functionals]. *Computational Mathematics and Mathematical Physics, 9(3),* 509–521 [in Russian].
2. Agmon, S. (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics, 6,* 382–392. doi: https://doi.org/10.4153/CJM-1954-037-2
3. Motzkin, T. S., & Schoenberg, I. J. (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics, 6,* 393–404. doi: https://doi.org/10.4153/CJM-1954-038-x
4. Eremin, I. I. (1965). Obobschenie relaksatsionnogo metoda Motskina-Agmona [Generalization of the relaxation method of Moitskin-Agmon]. it Uspekhi Matematicheskikh Nauk 20(2(122)), 183–187 [in Russian].
5. Stetsyuk, P. I. (2012). Acceleration of Polyak's subgradient method. *Theory of Optimum Solutions,* 11, 151–160.
6. Polyak, B. T. (1983). Vvedenie v optimizatsiyu [Introduction to optimization]. Nauka [in Russian].
7. Stetsyuk, P., Stovba, V., & Chernousova, Z. (2018). Subgradient Method with Polyak's Step in Transformed Space. In *International Conference on Optimization and Applications,* 49–63. Springer, Cham. doi: https://doi.org/10.1007/978-3-030-10934-9_4

Одержано 08.05.2019