Субградієнтний метод з кроком Поляка у перетвореному просторі змінних

Віктор Стовба
доктор філософії, науковий співробітник

vik.stovba@gmail.com

Інститут кібернетики імені В.М. Глушкова НАН України

Популярна лекція

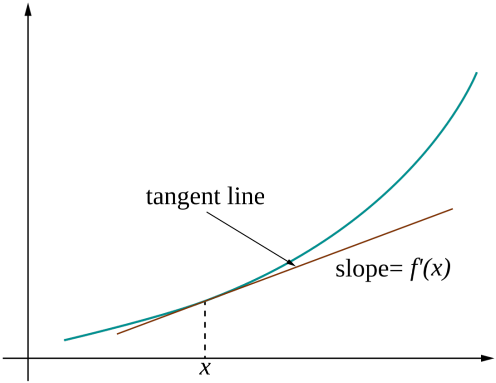
6 липня 2024 року

Зміст

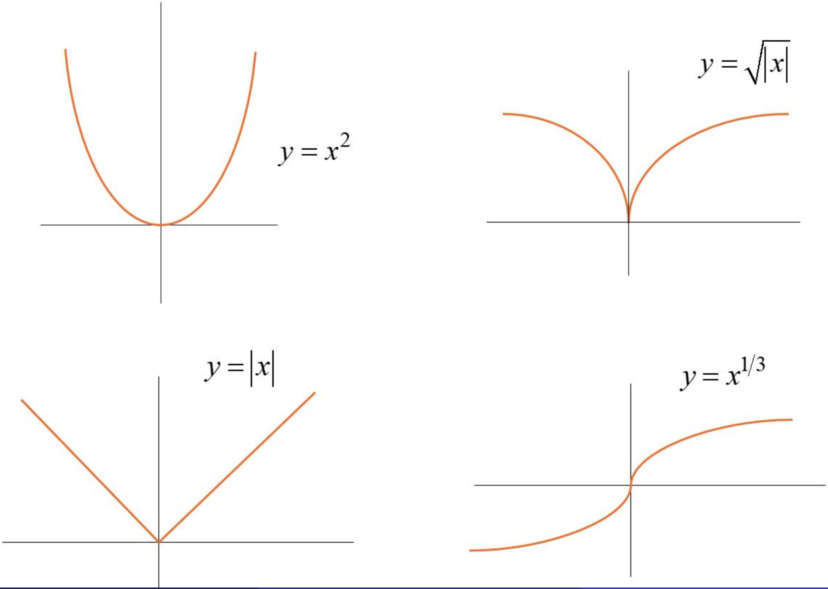
Зміст

Похідна

- Похідна: $f'(x_0) = \lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$



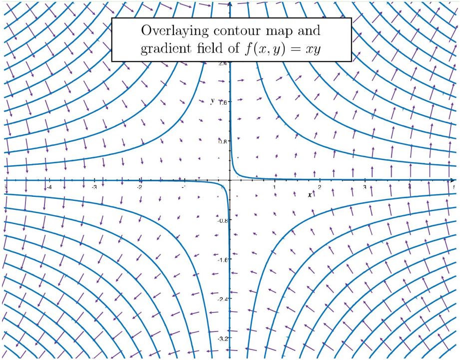
Неперервність та диференційовність



Градієнт

Градієнт $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ функції $f : \mathbb{R}^n \to \mathbb{R}$ в точці $x = (x_1, x_2, \ldots, x_n)$ n-вимірного простору визначається як вектор

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^T.$$



Зміст

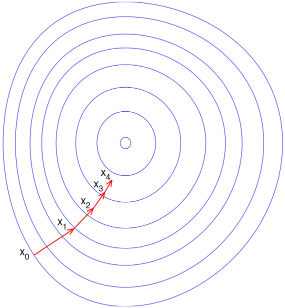
Градієнтний спуск

Також відомий як метод найшвидшого спуску

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{1}$$

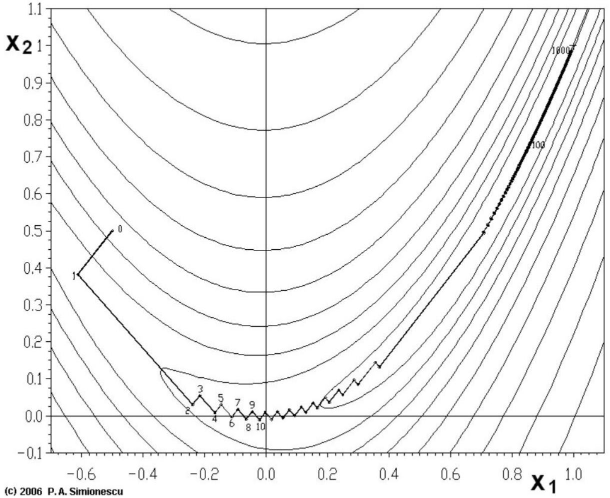
Маємо монотонну послідовність $f(x_0) \geq f(x_1) \geq f(x_2) \geq \dots$

Збігається за певних припущень: f опукла, $\nabla f(x_k)$ ліпшицевий та певному виборі α_k



Градієнтний спуск: функція Розенброка

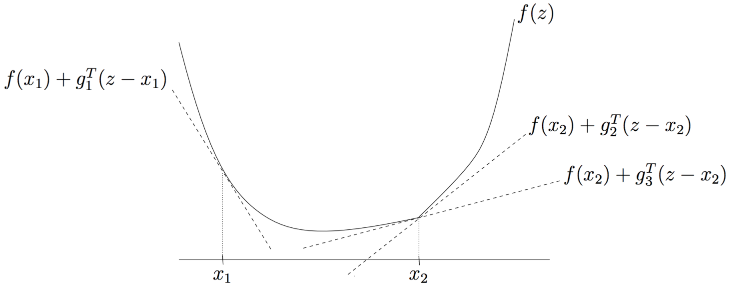
$$f(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$$



Субградієнт

Вектор $g \in \mathbb{R}^n$ є **субградієнтом** функції $f(x)$ в точці x_0, якщо для довільного вектора $x \in dom\ f$ виконується нерівність

$$f(x) \geq f(x_0) + g^T(x - x_0) \qquad (2)$$



Субградієнтний метод [1]

для мінімізації опуклих недиференційовних функцій

$$x_{k+1} = x_k - \alpha_k g_f(x_k) \tag{3}$$

- g_k – довільний субградієнт функції f в точці x
- крок $\alpha_k > 0$:
  - $\alpha_k = \alpha$
  - $\gamma / \|g_f(x_k)\|$
  - $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, $\sum_{k=1}^{\infty} \alpha_k = \infty$
- не є методом спуску, тому фіксуємо рекорди

$$f_k^{best} = \min_{i=1,\dots,k} f(x_k)$$

Проблеми

Які є проблеми?

1. Напрямок руху
2. Розмір кроку
3. Умови зупинки

За що змагаємось?

- Збіжність та швидкість збіжності
- Мінімізація функцій багатьох змінних $(n = 1K, 10K, 100K, 10M)$
- Швидкодія (ціна ітерації)

... а також за простоту реалізації, простір для модифікацій тощо

Зміст

Трохи історії

- **Агмон** [2], **Моцкін** та **Шонберг** [3] у 1954 році використали цей крок в релаксаційному методі для знаходження розв'язку сумісної системи лінійних нерівностей
- **Єрьомін** [4] у 1965 році узагальнив цей релаксаційний метод для системи опуклих нерівностей
- **Поляк** [5] в 1969 році використав цей крок для мінімізації негладких функціоналів

Корені вибору такого кроку прямують до Альтмана, Альбера...

Постановка задачі

Розглянемо таку задачу:

$$\text{знайти} \quad x^* = \arg\min_{x \in R^n} f(x) \quad \text{при відомому } f^\star, \tag{4}$$

де $f(x)$ – опукла функція та $f^\star = f(x^\star) = \min_{x \in R^n} f(x)$.

Субградієнтний метод з кроком Поляка

Метод А

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

- h_k – крок Поляка (ams-крок)
- $g_f(x)$ – субградієнт функції $f(x)$ в точці x
- $m \geq 1$ – скалярний параметр (зміщення по опуклості)

Що за параметр m?

$$(x - x^\star, g_f(x)) \geq m(f(x) - f^\star), \ \ \forall x \in R^n, \tag{6}$$

де $f(x)$ – довільна опукла функція, $g_f(x)$ – її субградієнт

$m \geq 1$ – максимальне зміщення по опуклості функції $f(x)$, яке не відкидає локалізацію точки $x^\star$ і дозволяє врахувати спеціальні види функцій

- довільні опуклі, кусочно-лінійні негладкі функції ($m = 1$)
- квадратичні гладкі функції ($m = 2$)
- однорідні диференційовні з показником σ ($m = \sigma$)

Зменшення відстані до точки мінімуму

Теорема 1

Послідовність $\{x_k\}_{k=0}^{k^*-1}$, згенерована методом (5), задовольняє нерівності

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

Примітка: теорема 1 гарантує, що в субградієнтному методі з кроком Поляка відстань до точки мінімуму зменшується **монотонно**.

Швидкість збіжності методу (5)

Теорема 2 (для довільних опуклих функцій)

Для послідовності $\{x_k\}_{k=0}^{k^*-1}$, згенерованої методом (5), справедлива нерівність $\underline{\lim}_{k\to\infty} \sqrt{k}(f(x_k) - f^*) = 0$.

Теорема 3 (для опуклих функцій з гострим мінімумом)

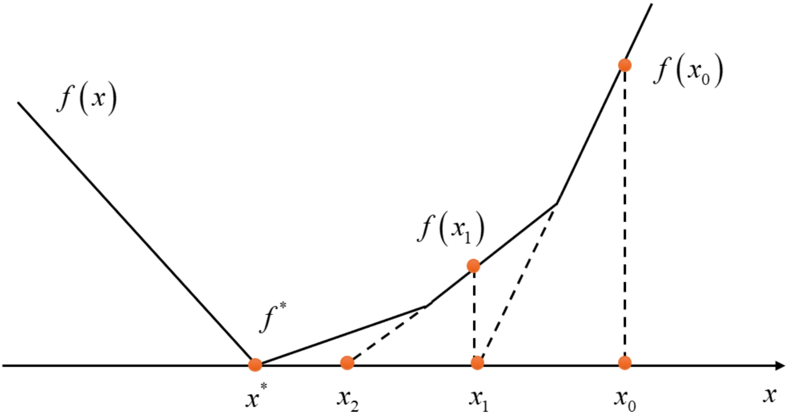
Нехай функція $f(x)$ має гострий мінімум, тобто для неї виконується нерівність $f(x) - f^* \geq \alpha \|x - x^*\|$. Тоді метод (5) збігається зі швидкістю геометричної прогресії зі знаменником $q_1 = 1 - (\frac{m\alpha}{C_1})^2$, де C_1 – константа, що обмежує норму субградієнта $g_f(x)$.

Геометричний зміст методу (5)

- Функція $f(x)$ апроксимується функцією $\tilde{f}(x) = f(x_k) + (g_f(x_k), x - x_k)$ і крок h_k обирається так, щоб $\tilde{f}(x_{k+1}) = f^*$ та виконувалась умова (6)

- для опуклої функції $f(x)$ крок h_k визначає величину **максимального зміщення** з точки x_k в напрямку нормалізованого антисубградієнта $-\dfrac{g_f(x_k)}{\|g_f(x_k)\|}$, для якого умова (6) гарантує, що $(-g_f(x_k), x_{k+1} - x^*) \geq 0$

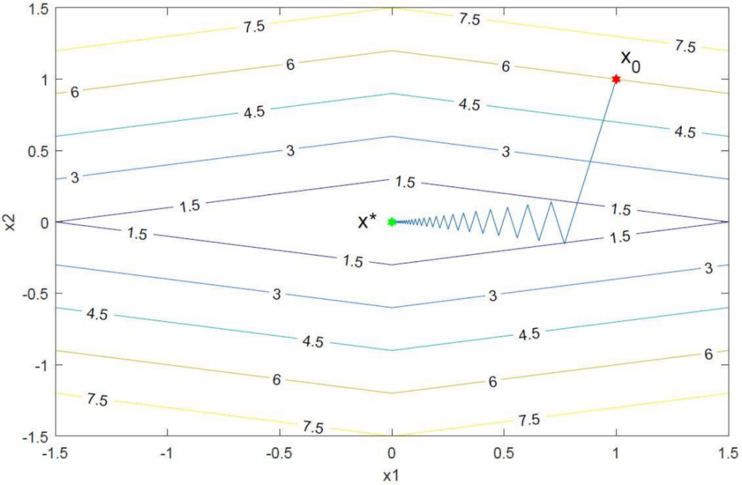
Геометричний зміст кроку Поляка



Зміст

Кусочно-лінійна функція

$$f(x_1, x_2) = |x_1| + 5|x_2|: \; m = 1, \; f^* = 0, \; x_0 = (1, 1)^T$$



Кусочно-лінійна та квадратична функції

- Для f_1: $m = 1$ Для f_2: $m = 1$ (в дужках) та $m = 2$
- $f^* = 0$, $x_0 = (1, 1)^T$

ε_f	$f_1(x_1, x_2) = \lvert x_1\rvert + t\lvert x_2\rvert$			$f_2(x_1, x_2) = x_1^2 + tx_2^2$		
	$t = 100$	$t = 50$	$t = 25$	$t = 10K$	$t = 1K$	$t = 100$
1.0e-01	14930	3721	925	6(139)	6(97)	6(9)
1.0e-02	26443	6600	1645	10(295)	10(127)	10(29)
1.0e-03	37956	9478	2365	12(551)	12(194)	12(51)
1.0e-04	49469	12356	3084	16(646)	16(230)	16(68)
1.0e-05	60982	15234	3804	20(885)	20(324)	20(86)
1.0e-06	72495	18113	4523	22(1003)	22(390)	22(98)
1.0e-07	84008	20991	5243	26(1135)	26(418)	26(121)
1.0e-08	95521	2386	5962	30(1332)	30(485)	28(135)
1.0e-09	107034	26747	6682	32(1518)	32(513)	32(150)
1.0e-10	118547	29625	7401	36(1947)	36(543)	36(169)

Квадратична функція

$$f_3(x) = (x - e)^T D(x - e)$$

- $x, e \in \mathbb{R}^n, \quad e = (1, \ldots, 1)^T$
- $D = \{d_i\}_{i=1}^n$ – діагональна $n \times n$ матриця
  - $d_i = 1 + \alpha \cdot rand[0,1], \quad i = \overline{1, n}$
  - $\alpha \in \{2, 1, 0.5, 0.1, 0.01\}$
- $x^\star = (1, \ldots, 1)^T, \quad f^\star = 0$
- $m = 1, 2, \quad q = itr_1 / itr_2$

Квадратична функція

$$n = 10\ 000\ 000, \quad \varepsilon = 10^{-20}$$

No	λ_{max}	λ_{min}	$m = 1$		$m = 2$		q
			itr_1	$f(x_k^\star)$	itr_2	$f(x_k^\star)$	
1	3	1	**45**	9.6065e-21	**42**	2.7832e-21	1.07
2	2	1	**45**	6.0550e-21	**27**	3.5420e-21	1.67
3	1.5	1	**45**	4.5327e-21	**19**	2.1762e-21	2.37
4	1.1	1	**45**	3.4582e-21	**11**	1.7910e-22	4.09
5	1.01	1	**45**	3.2485e-21	**7**	1.7799e-23	6.43

Зміст

Перетворення простору: трохи історії

Друга ідея Шора [6]: використання лінійних неортогональних перетворень простору для покращення властивостей функції, що мінімізується, в перетвореному просторі змінних та прискорення методів

Приклади: метод еліпсоїдів, r-алгоритми тощо

Як здійснюється перетворення?

- Заміна змінних: $x = By$, де B – невироджена $n \times n$-матриця
- Нова функція $\varphi(y) = f(By)$ та її субградієнт $g_\varphi(y)$
- Субградієнти пов'язані співвідношенням $g_\varphi(y) = B^T g_f(x)$
- $\varphi^* = \varphi(y^*) = f(By^*)$, $y^* = Ax^*$
- Аналог нерівності (6):

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

Субградієнтний метод з кроком Поляка у перетвореному просторі

Метод В

$$x_{k+1} = x_k - h_k B \frac{B^T g_f(x_k)}{\|B^T g_f(x_k)\|}, \quad h_k = \frac{m(f(x_k) - f^*)}{\|B^T g_f(x_k)\|}, k = 0, 1, \dots \quad (7)$$

Тут h_k – крок Поляка у перетвореному просторі змінних, $m \geq 1$ – скалярний параметр (зміщення по опуклості)

Примітка: якщо $B = I$, то субградієнтний метод з кроком Поляка у перетвореному просторі змінних співпадає з субградієнтним методом з кроком Поляка у вихідному просторі

Зменшення відстані до точки мінімуму

Теорема 4

Послідовність $\{x_k\}_{k=0}^{k^*-1}$, згенерована методом (7), задовольняє нерівності

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

Примітка: теорема 4 гарантує, що в субградієнтному методі з кроком Поляка в перетвореному просторі відстань до точки мінімуму зменшується монотонно, якщо використовується нерівність (6).

Швидкість збіжності методу (7)

Теорема 5 (для довільних опуклих функцій)

Для послідовності $\{y_k\}_{k=0}^{k^*-1}$, згенерованої методом (7), справедлива нерівність $\varliminf_{k\to\infty} \sqrt{k}(\varphi(y_k) - \varphi^*) = 0$.

Теорема 6 (для опуклих функцій з гострим мінімумом)

Нехай функція $\varphi(x)$ має гострий мінімум, тобто для неї виконується нерівність $\varphi(x) - \varphi^* \geq \alpha\|y - y^*\|$. Тоді метод (7) збігається зі швидкістю геометричної прогресії зі знаменником $q_2 = 1 - (\frac{m\alpha}{C_2})^2$, де C_2 – константа, що обмежує норму субградієнта $g_\varphi(y)$.

Ітеративний процес (для обох методів)

Ініціалізація: f^*, $m \geq 1$, $x_0 \in \mathbb{R}^n$, $n \times n$-матриця B (для методу з перетворенням), $\varepsilon > 0$.

Перехід: на k-й ітерації маємо точку $x_k \in \mathbb{R}^n$. Для переходу на ітерацію $k+1$ виконуємо такі кроки.

1. Обчислюємо $f(x_k)$ та $g_f(x_k)$. Якщо $f(x_k) - f^* \leq \varepsilon$ – зупинка: $k^* = k$, $x^*_\varepsilon = x_k$.
2. Обчислюємо наступну точку

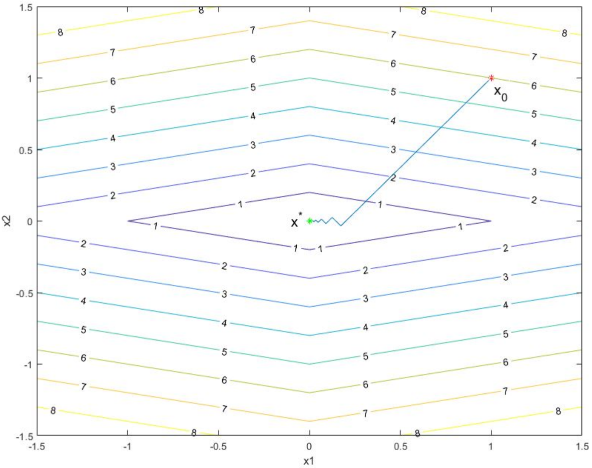
$$x_{k+1} = x_k - h_k B \frac{B^T g_f(x_k)}{\|B^T g_f(x_k)\|}, \quad h_k = \frac{m(f(x_k) - f^*)}{\|B^T g_f(x_k)\|}, k = 0, 1, \ldots$$

3. Переходимо на ітерацію $k+1$ з точкою x_{k+1}.

Зміст

Кусочно-лінійна функція

$$f(x_1, x_2) = |x_1| + 5|x_2|: \; m = 1, \; f^* = 0, \; x_0 = (1, 1)^T, \; B = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}$$

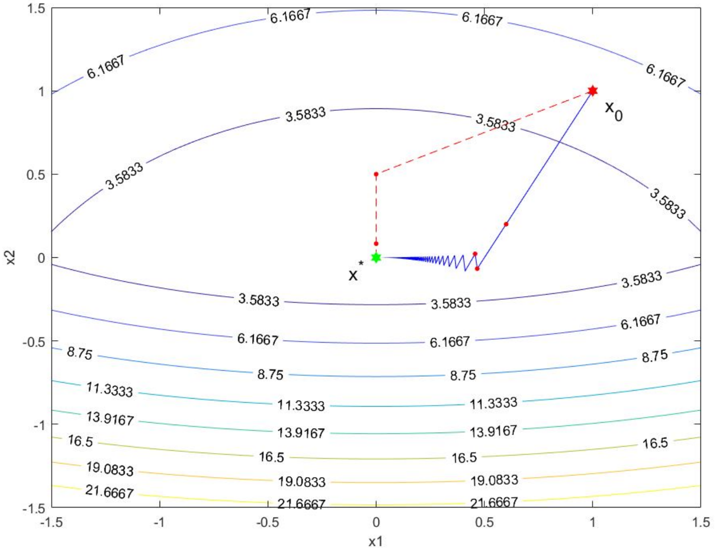


Кусочно-квадратична функція

$$f_3(x_1, x_2) = \max\{x_1^2 + (2x_2 - 2)^2 - 3, x_1^2 + (x_2 + 1)^2\}$$

- $m = 1, 2,\ f^* = 0,\ x_0 = (1, 1)$
- $\mathrm{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}$
- Блакитна траєкторія: метод A, $m = 1$
- Червона траєкторія: метод B, $m = 1$

Кусочно-квадратична функція



Квадратична функція

$$f_4(x_1, x_2) = x_1^2 + tx_2^2,\ t > 1$$

В дужках – **метод B**

ε_f	Методи A та B					
	$m = 1$			$m = 2$		
	$t = 100$	$t = 1K$	$t = 10K$	$t = 100$	$t = 1K$	$t = 10K$
1.0e-01	9 (6)	119 (10)	255 (23)	6 (2)	6 (3)	6 (4)
1.0e-05	81 (13)	311 (23)	1167 (71)	20 (2)	20 (5)	20 (6)
1.0e-10	173 (21)	564 (42)	1943 (162)	36 (2)	36 (8)	36 (8)
1.0e-15	266 (30)	878 (63)	2406 (232)	52 (2)	52 (10)	52 (10)
1.0e-20	341 (38)	1240 (82)	3635 (304)	68 (2)	70 (12)	70 (12)

Проблеми та шляхи вирішення

Які є проблеми в субградієнтному методі з кроком Поляка в перетвореному просторі?

1. Підбір параметра m
   - $m = 1$ підходить завжди, можна збільшувати
   - аналітично (вдається не завжди)
   - знати клас функції, що мінімізується
2. Підбір матриці перетворення простору B
   - Залежить від функції, що мінімізується
   - Перетворення у напрямку осей або їхньої лінійної комбінації
   - Емпірично

Основна перевага – простота та робота з функціями великої розмірності

Зміст

Використані джерела

1. **Shor N.Z.** Minimization Methods for Non-differentiable Functions. Springer-Verlag. 1985.
2. **Agmon S.** The relaxation method for linear inequalities. Canad. J. Math. 1954. 6. Pp. 382-392.
3. **Motzkin T., Schoenberg I.J.** The relaxation method for linear inequalities. Canad. J. Math. 1954. 6. Pp. 393-404.
4. **Еремин И.И.** Обобщение релаксационного метода Агмона-Моцкина. УМН. 1965. 2(122). С. 183-187.
5. **Поляк Б.Т.** Минимизация негладких функционалов. Вычислительная математика и математическая физика. 1969. 9(3). С. 509-521.
6. **Sergienko I.V., Stetsyuk P.I.** On N.Z. Shor's three scientific ideas. Cybernetics and Systems Analysis. 2012. 48 (1). Pp. 2-16.

Рекомендовані джерела

- **Стецюк П.І., Донець Г.П., Ненахов Е.І. та ін.** Субградієнтні алгоритми та задачі на комбінаторних конфігураціях. Київ: Унів. вид-во ПУЛЬСАРИ, 2019. 235 с.
- **Стецюк П.И.** Методы эллипсоидов и r-алгоритмы. Кишинэу, Эврика, 2014. 488 с.
- **Shor N.Z.** Nondifferentiable optimization and polynomial problems. Boston; Dordrecht; London: Kluwer Academic Publishers. 1998. 412 p.
- **Поляк Б.Т.** Введение в оптимизацию. М.: Наука. Главная редакция физико-математической литературы. 1983. 384 с.

Подяки

Доповідь підтримана

- проєктом дослідних робіт молодих вчених №07-02/03-2023
- грантом НАН України для дослідницьких груп/лабораторій молодих вчених №02/01-2024(5)
- грантом Volkswagen Foundation №97775
- проєктом ВЦП КНУ імені Тараса Шевченка при НАН України "Нові субградієнтні та екстраградієнтні методи для негладких задач регресії" 0124U002162

Дякую за увагу!

Якщо вас зацікавив даний напрямок досліджень або ви хотіли б обрати його як тему своєї дипломної роботи, прошу звертатись за адресами:

- Стецюк Петро Іванович (*stetsyukp@gmail.com*)
- Стовба Віктор Олександрович (*vik.stovba@gmail.com*)
- Корабльов Микола Миколайович (*m.m.korablov@gmail.com*)