Національна академія наук України
Інститут кібернетики імені В.М. Глушкова

Субградієнтний метод з кроком Поляка у перетвореному просторі

Стовба Віктор Олександрович

Дисертація на здобуття наукового ступеня
доктора філософії за спеціальністю
113 Прикладна математика

17 травня 2021 року

Науковий керівник:
доктор фізико-математичних наук
Стецюк Петро Іванович

Мета, об’єкт, предмет

Мета роботи: розробка субградієнтних методів з кроком Поляка та перетворенням простору, обґрунтування швидкості збіжності цих методів

Об’єкт дослідження: субградієнтні методи з перетворенням простору, мінімізація яружних функцій

Предмет дослідження: субградієнтний метод з кроком Поляка та метод еліпсоїдів, їхнє застосування для розв’язання систем лінійних рівнянь та визначення параметрів лінійної регресії з використанням L_p-норми

Загальна структура дисертації

Розділ 1. Огляд літератури за темою дисертації

Розділ 2. Субградієнтний метод з кроком Поляка та зміщенням по опуклості

Розділ 3. Субградієнтний метод з кроком Поляка та перетворенням простору

Розділ 4. Метод еліпсоїдів для лінійної регресії та систем лінійних рівнянь

Розділ 5. Програмні реалізації субградієнтних методів з кроком Поляка (С++)

Постановка задачі

Знайти

$$x^* = arg \min_{x \in \mathbb{R}^n} f(x) \text{ при відомому } f^* = f(x^*),$$

де $f(x)$, $x \in \mathbb{R}^n$ – опукла функція, $x^* \in X^*$ – точка мінімуму

Для довільної опуклої функції при $m = 1$ виконується нерівність

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

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

- квадратичні ($m = 2$)
- однорідні диференційовні з показником σ ($m = \sigma$)

Розділ 2. Субградієнтний метод з кроком Поляка та параметром $m > 1$

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

$\boldsymbol{h_k}$ – крок Поляка (Аґмона – Моцкіна – Шонберга)

Що зроблено: обґрунтовано монотонне зменшення відстані до точки мінімуму та швидкість збіжності субградієнтного методу з кроком Поляка та параметром $m > 1$ для загальної опуклої функції та опуклої функції з гострим мінімумом

Теореми про швидкість збіжності

Теорема 1. Для послідовності точок $\{x_k\}_{k=0}^{k^*-1}$, породженої методом (2), та довільної опуклої функції $f(x)$ справедлива рівність $\underline{\lim}_{k\to\infty} \sqrt{k}\,(f(x_k) - f^*) = 0$, тобто метод збігається зі швидкістю $O\left(\frac{1}{\sqrt{k}}\right)$.

Теорема 2. Нехай функція $f(x)$ має гострий мінімум, тобто для неї виконується нерівність $f(x) - f^* \geq \alpha\|x - x^*\|$. Тоді метод (2) збігається 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 $g_f(x)$.

При $\boldsymbol{m = 1}$ монотонне зменшення та швидкість збіжності обґрунтовано Б.Т. Поляком

$$f_3(x_1, x_2) = x_1^2 + t x_2^2, \quad t > 1$$

ε_f	**Метод А**					
	$m = 1$			$m = 2$		
	$t = 100$	$t = 1000$	$t = 10000$	$t = 100$	$t = 1000$	$t = 10000$
10^{-1}	9	119	255	6	6	6
10^{-5}	81	311	1167	20	20	20
10^{-10}	173	564	1943	36	36	36
10^{-15}	266	878	2406	52	52	52
10^{-20}	341	1240	3635	68	70	70

Використання параметра $m > 1$ дозволяє прискорити метод А для спеціальних функцій

$$f_5(x) = \sum_{i=1}^{k}\left|\sum_{j=1}^{n} a_{ij}x_j - b_i\right|^p, \quad p \geq 1, \quad m = p$$

$$A = \begin{pmatrix} 100 & 0 & 0 \dots & 0 \\ 0 & 100 & 0 \dots & 0 \\ & & A_1 & \end{pmatrix}$$

A_1 – матриця розмірності 498×100 заповнена випадковими числами з інтервалу [0,3]

Поза дужками – метод A $(m = 1)$, в дужках – метод A $(m = p)$

ε_f \ p	1	1.2	1.4	1.6	1.8	2
10^{-2}	110	357 (**146**)	204 (**98**)	223 (**128**)	164 (**164**)	158 (**353**)
10^{-4}	189	523 (**242**)	349 (**176**)	378 (**240**)	271 (**304**)	252 (**695**)
10^{-6}	267	728 (**334**)	554 (**260**)	481 (**362**)	427 (**460**)	334 (**1085**)
10^{-8}	344	1062 (**456**)	760 (**340**)	634 (**484**)	525 (**618**)	416 (**1491**)
10^{-10}	414	1302 (**513**)	906 (**424**)	819 (**606**)	661 (**778**)	544 (**1901**)

Розділ 3. Субградієнтний метод з кроком Поляка у перетвореному просторі та параметром $m \geq 1$

Заміна змінних: $x = By$, де $\boldsymbol{B}$ – невироджена $n \times n$-матриця, $A = B^{-1}$

Функція $\varphi(y) = f(By)$, причому $\varphi^* = \varphi(y^*) = f(By^*) = f(x^*) = f^*$, $g_\varphi(y) = B^T g_f(x)$

$$y_{k+1} = y_k - h_k \frac{g_\varphi(y_k)}{\|g_\varphi(y_k)\|}, \ h_k = \frac{m(\varphi(y_k) - \varphi^*)}{\|g_\varphi(y_k)\|}, \ k = 0,1,2,\ldots \qquad (3)$$

$\boldsymbol{h_k}$ – крок Поляка в перетвореному просторі змінних

$$x_{k+1} = x_k - h_k B \frac{B^T g_f(x_k)}{\|B^T g_f(x_k)\|}, \ h_k = \frac{m(f(x_k) - f^*)}{\|B^T g_f(x_k)\|}, \ k = 0,1,2,\ldots \qquad (4)$$

Що зроблено: обґрунтовано монотонне зменшення відстані до точки мінімуму та швидкість збіжності субградієнтного методу з кроком Поляка у перетвореному просторі змінних та параметром $m \geq 1$ для загальної опуклої функції та опуклої функції з гострим мінімумом

Теореми про монотонне зменшення відстані до точки мінімуму

Теорема 3. Послідовність $\{x_k\}_{k=0}^{k^*-1}$, породжена методом (4) при $m = 1$, задовольняє нерівності

$$\|A(x_{k+1} - x^*)\|^2 \leq \|A(x_k - x^*)\|^2 - \frac{(f(x_k) - f^*)^2}{\|B^T g_f(x_k)\|^2}, \; k = 0,1,2, \ldots.$$

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

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

Теореми про швидкість збіжності

Теорема 5. Для послідовності точок $\{y_k\}_{k=0}^{k^*-1}$, породженої методом (3), та довільної опуклої функції $\varphi(y)$ справедлива рівність $\lim\limits_{k\to\infty} \sqrt{k}\,(\varphi(y_k) - \varphi^*) = 0$, тобто метод збігається зі швидкістю $O\left(\frac{1}{\sqrt{k}}\right)$.

Теорема 6. Нехай функція $\varphi(y)$ має гострий мінімум, тобто виконується нерівність $\varphi(y) - \varphi^* \geq \alpha\|y - y^*\|$, тоді метод (3) збігається зі швидкістю геометричної прогресії зі знаменником $q_3 = 1 - \left(\frac{m\alpha}{C_3}\right)^2$, де C_3 – константа, що обмежує норму субградієнта $g_\varphi(y)$.

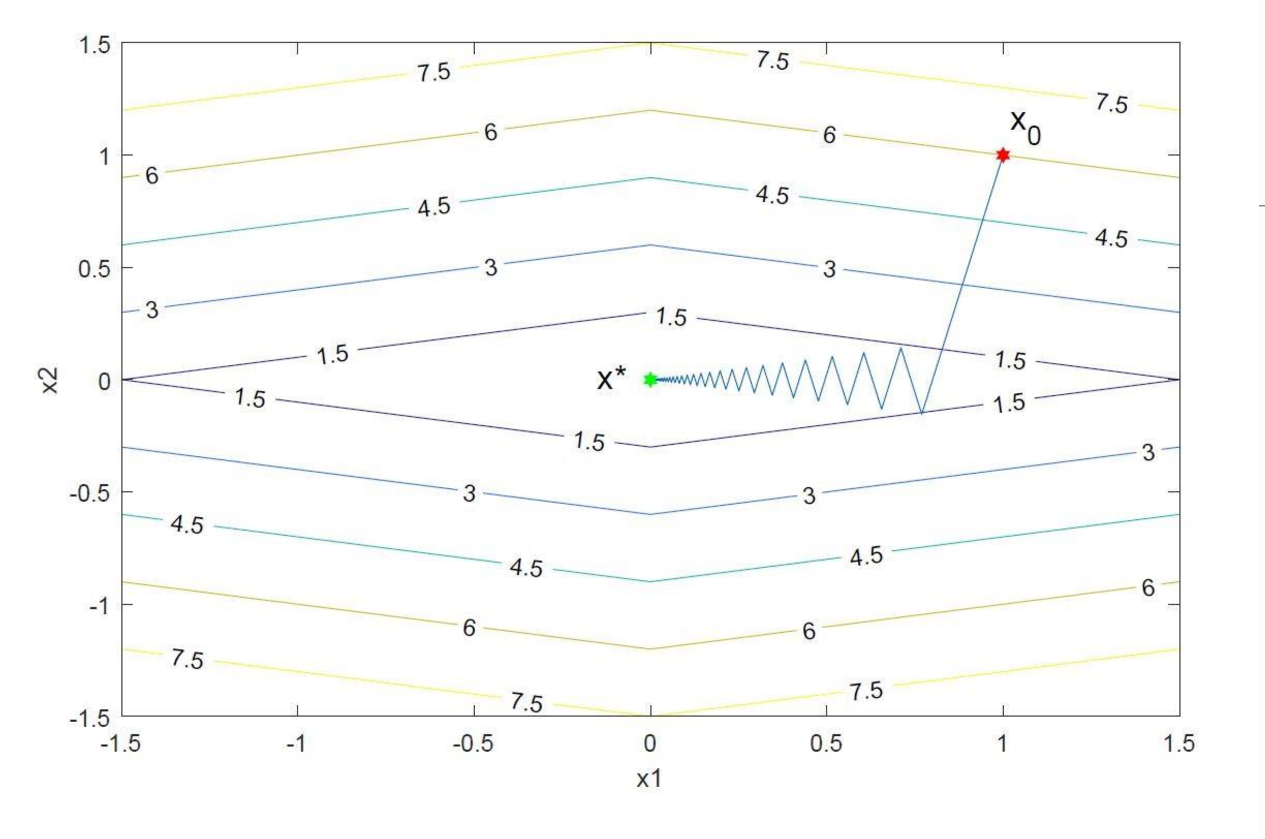
Теореми 5 і 6 справедливі при $\boldsymbol{m = 1}$

$$f_3(x_1, x_2) = x_1^2 + tx_2^2$$

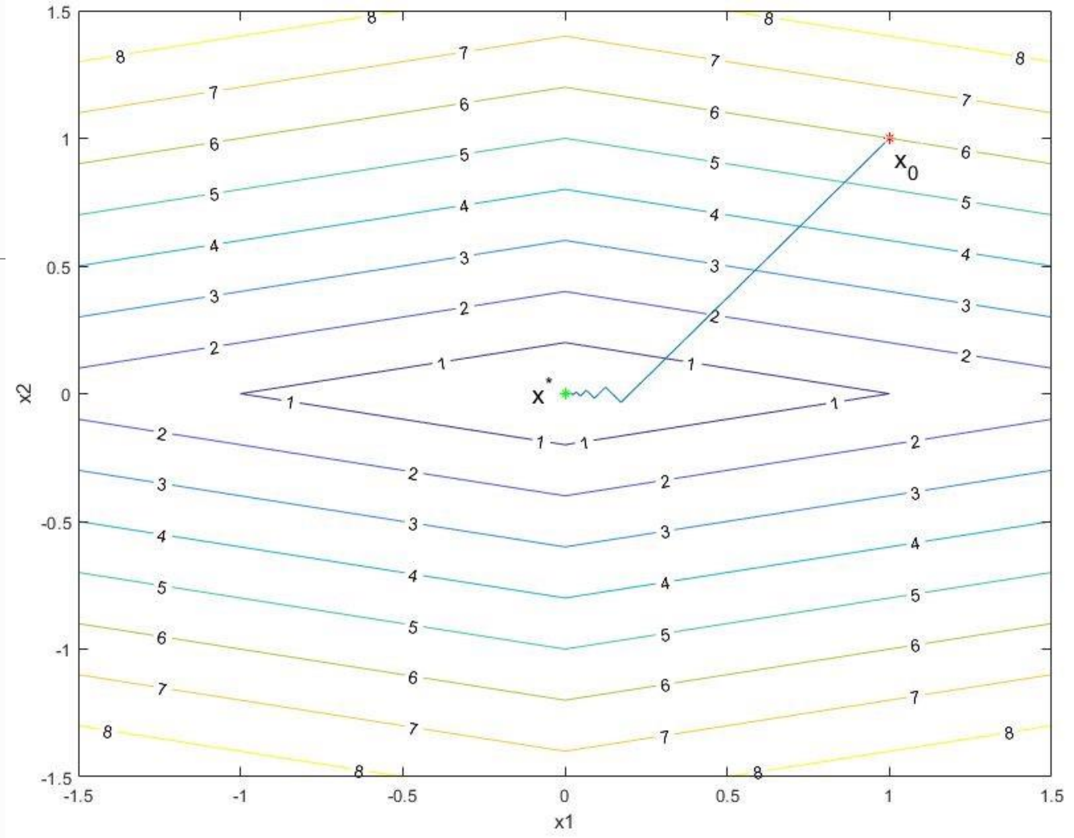
$$B = \begin{pmatrix} 1 & 0 \\ 0 & 0.1 \end{pmatrix}$$

	Метод B					
	$m = 1$			$m = 2$		
ε_f \ t	100	1000	10000	100	1000	10000
10^{-1}	6	10	23	2	3	4
10^{-5}	13	23	71	2	5	6
10^{-10}	21	42	162	2	8	8
10^{-15}	30	63	232	2	10	10
10^{-20}	38	82	304	2	12	12

Використання параметра $m > 1$ та операції перетворення простору дозволяє прискорити роботу класичного методу для спеціальних функцій з суттєво витягнутими поверхнями рівня



Метод A $(m = 1)$



Метод B $(m = 1)$

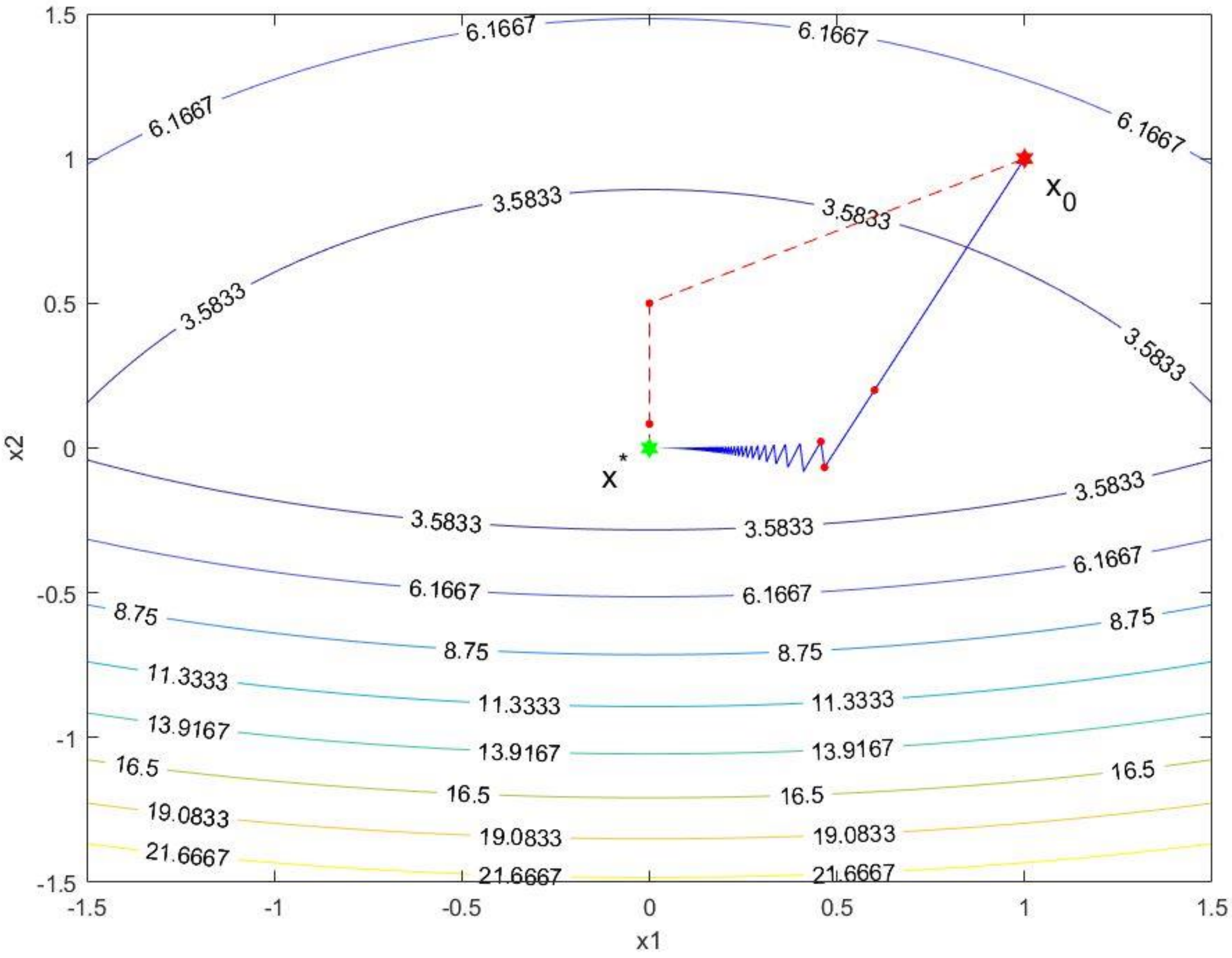
$$f_8(x_1, x_2) = |x_1| + 5|x_2|,\ B = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}$$

$$f_9(x_1, x_2) = max\begin{Bmatrix} x_1^2 + (2x_2 - 2)^2 - 3, \\ x_1^2 + (x_2 + 1)^2 \end{Bmatrix}$$

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}$$

Метод A $(m = 1)$ – субградієнтний метод з кроком Поляка

Метод B $(m = 1)$ – субградієнтний метод з кроком Поляка у перетвореному просторі



$$f_{10}(x) = x^T H x + b^T x + c,\ \ H\ - 1000 \times 1000 \qquad x^* = rand[0,10],\ \ f^* = 0,\ \ \varepsilon_f = 10^{-10}$$

$B = diag(1, \dots, 1) + (\beta - 1)\chi\chi^T, \quad \beta = \sqrt{\frac{\lambda_{min}}{\lambda_{max}}}, \quad \chi$ – в.в., що відповідає λ_{max}

№	$\lambda = \frac{\lambda_{max}}{\lambda_{min}}$	Метод A		Метод B	
		$m = 1$	$m = 2$	$m = 1$	$m = 2$
1	1.7865e+06	7399	14	20	5
2	3.6504e+04	1892	18	21	6
3	5.8373e+03	932	22	23	8
4	4.4003e+03	879	26	25	9
5	4.1499e+03	1449	28	27	10

Обчислювальні експерименти

- $f_1(x_1, x_2) = x_1^4 + 10000x_2^4$ $(m = 4)$
- $f_2(x_1, x_2) = (x_1 + 1.001x_2)^4 + (1.001x_1 + x_2)^4$ $(m = 4)$
- $f_3(x_1, x_2) = x_1^2 + tx_2^2$ $(m = 2)$
- $f_4(x) = \|Ax - b\|^2$ $(m = 2)$
- $f_5(x) = \sum_{i=1}^{k} \left|\sum_{j=1}^{n} a_{ij}x_j - b_i\right|^p, p \geq 1$ $(m = p)$
- $f_6(x) = \sum_{i=1}^{p} |(a_i, x) - b_i|$ $(m = 1)$
- $f_7(x) = \max\limits_{i=\overline{1,p}} |(a_i, x) - b_i|$ $(m = 1)$
- $f_8(x_1, x_2) = |x_1| + t|x_2|, t > 1$ $(m = 1)$
- $f_9(x_1, x_2) = max\{x_1^2 + (2x_2 - 2)^2 - 3,\ x_1^2 + (x_2 + 1)^2\}$ $(m = 1)$
- $f_{10}(x) = x^T Hx + b^T x + c$ $(m = 2)$

Розділ 4. Лінійна регресія та L_p-норма

Модель парної лінійної регресії в матричному вигляді

$$y = Xb + \varepsilon,$$

де $y = (y_1, \dots, y_n)^T$, $b = (b_0, b_1)^T$, $X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{1n} \end{pmatrix}$ – $n \times 2$-матриця факторів

Мінімізація опуклої при $p \geq 1$ функції

$$f_p(b) = \|Xb - y\|_p = \|\varepsilon\|_p = \left(\sum_{i=1}^{n} |\varepsilon_i|^p\right)^{\frac{1}{p}}$$

Значенням $p = 1, 2, \infty$ відповідають метод найменших модулів, метод найменших квадратів та мінімаксний (Чебишевський) метод відповідно

Що зроблено: побудовано метод еліпсоїдів для мінімізації функції $f_p(b)$ при *великих* значеннях p

Формули для обчислення значення функції та її субградієнта при великих значеннях p

Функція

$$f_p(b) = \|Xb - y\|_p \qquad (5)$$

обчислюється як

$$f_p(b) = |\varepsilon_{max}| \cdot \|z\|_p, \quad \varepsilon_{max} = \max_{i=\overline{1,n}} |\varepsilon_i|, \quad z_i = \frac{\varepsilon_i}{\varepsilon_{max}} \leq 1,\ i = \overline{1,n} \qquad (6)$$

Субградієнт

$$g_f(b) = \|Xb - y\|_p^{1-p} \sum_{i=1}^{n} sign(x_i b - y_i) \cdot |x_i b - y_i|^{p-1} x_i^T \qquad (7)$$

обчислюється як

$$g_f(b) = \|z\|_p^{1-p} \sum_{i=1}^{n} sign(\varepsilon_i) \cdot |z_i|^{p-1} x_i^T, \qquad (8)$$

x_i – i-й вектор-рядок матриці X

$$(\boldsymbol{x_i}, \boldsymbol{y_i}), \quad x_i = y_i = i - 1, \quad i = \overline{1,17},$$

$$(x_{18}, y_{18}) = (17,0), \quad (x_{19}, y_{19}) = (18,0), \quad (x_{20}, y_{20}) = (19,0)$$

$$y = cx + d$$

p	itn	$\overline{c_p^*}$	$\overline{d_p^*}$	$f_p\left(\overline{c_p^*}, \overline{d_p^*}\right)$
10	112	4.4854e – 02	6.8507	9.3126
10^2	120	4.2751e – 03	7.8780	8.1090
10^3	130	4.2764e – 04	7.9878	8.0109
10^4	141	4.2763e – 05	7.9988	8.0011
10^5	151	4.2766e – 06	7.9999	8.0001
10^6	150	4.2789e – 07	8.0000	8.0000
∞	225	6.0668e – 14	8.0000	8.0000

$$f_\infty(b) = \max_{i=1,n} |x_i b - y_i|$$

Розділ 4. Задача знаходження L_p-розв'язку системи лінійних рівнянь з двосторонніми обмеженнями на змінні

Система лінійних рівнянь

$$Ax \approx b \quad \text{за умов} \quad l \leq x \leq u,$$

де $\boldsymbol{A}$ – $m \times n$-матриця, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{x}, \boldsymbol{l}, \boldsymbol{u} \in \mathbb{R}^n$, причому $l_i < u_i$, $i = \overline{1, n}$

Мінімізація опуклої функції

$$f_p(x) = \|Ax - b\|_p \quad \text{за умов} \quad l \leq x \leq u$$

Що зроблено: запропоновано алгоритм Юдіна-Неміровського (використовує H-форму методу еліпсоїдів)

Для мінімізації функції $f_p(x)$ можна використовувати **формули (6), (8)** для великих p

Розділ 5. Програмні реалізації

- функція **PolyakA** – субградієнтний метод з кроком Поляка та параметром $m \geq 1$
- функція **PolyakB** – субградієнтний метод з кроком Поляка у перетвореному просторі та параметром $m \geq 1$
- тест **run1** (розділи 2 і 3): функції *squad*, *sabs*, *maxx*, *four1*, *four2*
- тест **run2** (розділ 2): функція *psum*
- тест **run3** (розділ 4): функції *func1*, *func2*, *func3*
- допоміжні функції та бібліотеки *vector*, *random*, *algorithm*

Висновки

- Обґрунтовано монотонне зменшення відстані до точки мінімуму субградієнтного методу з кроком Поляка та параметром $m > 1$

- Обґрунтовано монотонне зменшення відстані до точки мінімуму субградієнтного методу з кроком Поляка у перетвореному просторі

- Обґрунтовано монотонне зменшення відстані до точки мінімуму субградієнтного методу з кроком Поляка у перетвореному просторі змінних та параметром $m > 1$

Висновки

- Обґрунтовано швидкість збіжності субградієнтного методу з кроком Поляка та параметром $m > 1$, яка рівна $O\left(\frac{1}{\sqrt{k}}\right)$ для довільної опуклої функції та швидкості геометричної прогресії зі знаменником $q_1 = 1 - \left(\frac{m\alpha}{C_1}\right)^2$ для опуклої функції з гострим мінімумом

- Обґрунтовано швидкість збіжності субградієнтного методу з кроком Поляка у перетвореному просторі, яка рівна $O\left(\frac{1}{\sqrt{k}}\right)$ для довільної опуклої функції та швидкості геометричної прогресії зі знаменником $q_2 = 1 - \frac{\alpha^2}{C_2}$ для опуклої функції з гострим мінімумом

- Обґрунтовано швидкість збіжності субградієнтного методу з кроком Поляка у перетвореному просторі та параметром $m > 1$, яка рівна $O\left(\frac{1}{\sqrt{k}}\right)$ для довільної опуклої функції та швидкості геометричної прогресії зі знаменником $q_3 = 1 - \left(\frac{m\alpha}{C_3}\right)^2$ для опуклої функції з гострим мінімумом

Висновки

- Проведено обчислювальні експерименти з мінімізації яружних функцій з сильно та суттєво витягнутими поверхнями рівня з використанням субградієнтного методу з кроком Поляка у вихідному та перетвореному просторах змінних та параметром $m \geq 1$

- Розроблено програмні реалізації субградієнтного методу з кроком Поляка у вихідному та перетвореному просторах змінних та параметром $m \geq 1$ мовою C++, а також низку тестових прикладів для проведення експериментів

Висновки

- Побудовано метод еліпсоїдів для визначення параметрів лінійної регресії при великих значеннях параметра $p \geq 1$. Досліджено стійкість алгоритму до появи аномальних спостережень при різних значеннях параметра p та проведено його порівняння з класичними методами

- Запропоновано алгоритм на основі H-форми методу еліпсоїдів для знаходження L_p-розв'язку системи лінійних рівнянь з двосторонніми обмеженнями на змінні. Алгоритм є стійким при розв'язанні погано обумовлених систем лінійних рівнянь

Публікації

12 публікацій

- 2 наукові статті – в фахових виданнях України
- 1 стаття англійською мовою – в зарубіжному виданні (Scopus)
- 1 робота – як розділ колективної монографії
- 5 тез доповідей – в збірниках міжнародних науково-практичних конференцій
- 1 тези доповіді – в збірнику міжнародного наукового семінару
- 1 тези доповіді – в збірнику міжнародного воркшопу

Дякую за увагу!