ISSN 2078–5097. Вісн. Львів. ун-ту. Сер. прикл. матем. та інф. 2017. Вип. 25. С. 3–8
Visnyk of the Lviv University. Series Appl. Math. and Informatics. Issue 25. P. 3–8

О Б Ч И С Л Ю В А Л Ь Н А   М А Т Е М А Т И К А

УДК 519.6

ПРО ОДИН МЕТОД СПУСКУ МІНІМІЗАЦІЇ ФУНКЦІЙ

М. Бартіш, Н. Огородник

Львівський національний університет імені Івана Франка,
вул. Університетська, 1, Львів, 79000, e-mail: ogorodnyk.nataly@gmail.com

Використовуючи ідею побудови методів спуску, побудовано новий варіант “градієнтного” методу. У запропонованому методі використано принцип побудови методів покоординатного спуску. Теоретично обгрунтовано збіжність методу. Виконано числові експерименти, зроблено висновки про ефективність і можливість застосування запропонованих алгоритмів.

Ключові слова: методи спуску, метод покоординатного спуску, безумовна мінімізація.

1. ВСТУП

У різних галузях людського життя часто виникає потреба у розв’язуванні задач оптимізації. Такі задачі часто трапляються в науці, техніці, промисловості та багатьох інших галузях. Здебільшого математичні моделі зводяться до розв’язування типових задач оптимізації, а зростання потужності комп’ютерів сприяє розширенню сфер успішного застосування методів оптимізації у розв’язанні все складніших задач все більших розмірностей. Різні методи виявляють свою ефективність на різних класах задач. Методи мають свої переваги та недоліки: вибір початкового наближення, швидкість збіжності, трудомісткість окремої ітерації тощо. Сьогодні не існує універсального алгоритму, тому і надалі актуальною залишається проблема побудови ефективних методів [1-3].

Ми досліджуватимемо методи спуску, а саме метод блочного покоординатного спуску, який використовує часткові похідні мінімізуючої функції у координатах вектора спуску. Проведено теоретичні дослідження методу, доведено збіжність. При практичному застосуванні алгоритм довів свою ефективність щодо кількості обчислень.

2. ФОРМУЛЮВАННЯ ЗАДАЧІ

Розглянемо задачу безумовної мінімізації функції багатьох змінних

$$f(x) \to \min, \quad x \in R^n. \tag{1}$$

Введемо позначення $N = \{1, 2, \ldots, n\}$. З цієї множини виділимо підмножини $N_l = \{l_1, l_2, \ldots, l_l\}$, де $l_i \in N$; $N_l \subseteq N$ і $\bigcup_{l=1}^{p} N_l = N$. $N_l \neq N_j$, якщо $l \neq j$, $l, j = 1, 2, \ldots, p$, $p \leq n$. Введемо вектор $a_l(x) = \left(a_l^1(x), \ldots, a_l^n(x)\right)^T$, де $a_l^i(x) = f'_{x_i}(x)$, якщо $i \in N_l$ і $a_l^i = 0$, якщо $i \notin N_l$.

Для розв’язування (1) запропоновано метод вигляду

$$x_{k,\,j+1} = x_{kj} - \alpha_{kj} a_l\left(x_{kj}\right), \tag{2}$$

$$f\left(x_{k,\,j+1}\right) - f\left(x_{kj}\right) \le -\varepsilon\alpha_{kj}\left\|a_l\left(x_{kj}\right)\right\|^2, \tag{3}$$

де $\varepsilon \in (0,1)$, $j = 1,2,\ldots,p$, $k = 0,1,\ldots$, $x_{k(p+1)} = x_{k+1} = x_{k+1,1}$, для якого справджуються такі леми.

Лема 1. Нехай $a_l \ne 0$, тоді $a_l\left(\bar{x}\right)$ є вектором спадання функції $f(x)$ у точці $\bar{x}$.

Отже, метод (2) – (3) є методом спуску.

Лема 2. Нехай $f \in C^1\left(R^n\right)$ і $f'(x)$ задовольняє умову Ліпшиця. $\left\|f'(x) - f'(y)\right\| \le L\|x-y\|$. Тоді для $\forall x_{kj} \in R^n$ умова (3) виконується при

$$0 < \alpha_{kj} \le \frac{(1-\varepsilon)}{L}.$$

Доведення. З формули Лагранжа за деякого $\varepsilon \in (0,1)$ і прирості $\alpha_{kj}h$ отримаємо

$$f\left(x_{kj} + \alpha_{kj}h\right) - f\left(x_{kj}\right) = \left(f'\left(x_{kj} + \theta\alpha_{kj}h\right), \alpha_{kj}h\right) =$$

$$= \alpha_{kj}\left(f'\left(x_{kj}\right), h\right) + \alpha_{kj}\left(f'\left(x_{kj} + \theta\alpha_{kj}h\right) - f'\left(x_{kj}\right), h\right) \le$$

$$\le \alpha_{kj}\left(f'\left(x_{kj}\right), h\right) + \alpha_{kj}L\left\|\theta\alpha_{kj}h\right\|\|h\| \le \alpha_{kj}\left(f'\left(x_{kj}\right), h\right) + L\alpha_{kj}\|h\|^2.$$

Приймемо $h = -a_l$. Тоді, враховуючи побудову вектора a_l

$$\left(f'\left(x_{kj}\right), h\right) = \left(f'\left(x_{kj}\right), -a_l\right) \le -\left\|a_l\right\|^2,$$

$$f\left(x_{k,\,j+1}\right) - f\left(x_{kj}\right) \le -\alpha_{kj}\left\|a_{kj}\right\|^2 + L\alpha_{kj}\left\|a_{kj}\right\|^2 \le$$

$$\le \alpha_{kj}\left\|a_{kj}\right\|^2\left(-1+1-\varepsilon\right) = -\varepsilon\alpha_{kj}\left\|a_{kj}\right\|^2$$

Лема 3. Якщо функція $f \in C^2\left(R^n\right)$ *і* $f''(x)$ задовольняє умову $\left(f''(x)h, h\right) \le D\|h\|^2$ $\forall x, h \in R^n, \infty > D > 0,$ тоді для довільних $x_k \in R^n$, $\varepsilon \in (0,1)$ умова (3) виконується при

$$0 < \alpha_{kj} \le \frac{2(1-\varepsilon)}{D}.$$

Доведення. З формули Тейлора і доведення попередньої леми отримаємо

$$f\left(x_{kj} + \alpha_{kj}h_{kj}\right) - f\left(x_{kj}\right) =$$

$$= \left(f'\left(x_{kj}\right), \alpha_{kj}h_{kj}\right) + \frac{1}{2}\left(f''\left(x_{kj} + \theta\alpha_{kj}h_{kj}\right)\alpha_{kj}h_{kj}, \alpha_{kj}h_{kj}\right) \le$$

$$\le \alpha_{kj}\left(f'\left(x_{kj}\right), h_{kj}\right) + \frac{{\alpha_{kj}}^2}{2}D\left\|h_{kj}\right\|^2 \le -\alpha_{kj}\left\|a_{kj}\right\|^2 + (1-\varepsilon)\alpha_{kj}\left\|a_{kj}\right\|^2 =$$

$$= \left(-1+1-\varepsilon\right)\alpha_{kj}\left\|a_{kj}\right\|^2 = -\varepsilon\alpha_{kj}\left\|a_{kj}\right\|^2.$$

Теорема 1. Нехай $f(x) \in C^1(R^n)$ і обмежена знизу на R^n, градієнт задовольняє умову Ліпшиця $\|f'(x) - f'(y)\| \le L\|x - y\| \quad \forall x, y \in R^n$. Тоді за довільного початкового наближення $x_0 \in R^n$ для методу (2) – (3) матимемо $\lim\limits_{k\to\infty} f'(x_k) = 0$.

Доведення. Із (3) отримаємо

$$f(x_{k,j+1}) - f(x_{kj}) \le -\varepsilon\alpha_{kj}\|a_{kj}\|^2 \le 0.$$

Отже, послідовність $\{f(x_k)\}$ монотонно спадна, а за умовою теореми функція $f(x)$ обмежена знизу, тоді послідовність збіжна і

$$f(x_k) - f(x_{k+1}) \underset{k\to\infty}{\to} 0.$$

Якщо взяти до уваги, що в разі виконання умов теореми справджується умова $\alpha_{kj} \ge \alpha > 0$, то отримаємо

$$\|f'(x_{kj})\|^2 \le \frac{f(x_{kj}) - f(x_{k,j+1})}{\varepsilon\alpha_{kj}} \le \frac{f(x_{kj}) - f(x_{k,j+1})}{\varepsilon\alpha} \underset{k\to\infty}{\to} 0.$$

Отже, гранична точка x_* є стаціонарною. Теорему доведено.

Зауваження. У випадку опуклої функції $f(x)$ ми отримаємо

$$\lim_{k\to\infty} f'(x_k) = f'(x_*) = 0$$

і x_* – точка глобального мінімуму.

У випадку $p = 1$ метод (2) – (3) буде градієнтним методом, а у випадку $N_l = \{l\}$ – методом покоординатного градієнтного спуску.

4. ЧИСЛОВІ ЕКСПЕРИМЕНТИ

Ми розглянули приклади та провели порівняння алгоритму (2) – (3) з градієнтним методом

$$x_{k+1} = x_k - \alpha_k f'(x_k). \tag{10}$$

Обчислення проводили до виконання умови $\|x_{k+1} - x_k\| \le \varepsilon$, $\varepsilon = 10^{-5}$. У табл. подано кількість ітерацій – h і K – кількість обчислень еквівалентних обчисленню функції $f(x)$, які були затрачені для отримання наближеного розв'язку наведених задач, іншими обчисленнями ми нехтували, оскільки вони суттєво не впливали на оцінку ефективності роботи методів. Зі збільшенням точності обчислень ε, кількість обчислень K суттєво зростає, позаяк для багатьох функцій градієнтний метод при наближенні до точки розв'язку значно сповільнюється. Тестові задачі взяли з [4].

Задача 1. Штрафна функція

$$f(x) = \sum_{i=1}^{n-1}(x_i - 1)^2 + \left(\sum_{i=1}^{n} x_i^2 - 0.25\right)^2.$$

Значення x^* і $f(x^*)$ залежать від значення n; $x_0 = (-1.2, 1, \ldots, -1.2, 1)^T$.

Таблиця 1

Порівняльні характеристики роботи методів (10), (2) – (3)

	(10)		(2) – (3)		
n	*h*	*K*	*p*	*h*	*K*
4	22	172	2	21	420
10	19	557	2	43	898
10			5	17	408
20	18	465	5	36	859
20			10	19	550
100	16	2003	20	48	1805
100			50	20	1424
1000	10	11216	100	22	11041
1000			500	13	10351

Задача 2. Збурена квадратична функція

$$f\left(x\right)=\sum_{i=1}^{n} i x_i^2+\frac{1}{100}\left(\sum_{i=1}^{n} x_i\right)^2 .$$

$x^*=\left(0,0,\ldots,0\right)^T$ і $f\left(x^*\right)=0$; $x_0=\left(-1.2,1,\ldots,-1.2,1\right)^T$.

Таблиця 2

Порівняльні характеристики роботи методів (10), (2) – (3)

	(10)		(2) – (3)		
n	*h*	*K*	*p*	*h*	*K*
4	21	154	2	14	292
10	45	657	2	25	523
10			5	28	650
20	77	1965	5	32	749
20			10	34	972
100	369	39803	20	44	1704
100			50	42	2914
1000	2420	2465186	100	68	8142
1000			500	52	30581

Задача 3. QF1 квадратична функція

$$f\left(x\right)=\frac{1}{2}\sum_{i=1}^{n} i x_i^2+x_n .$$

$x^*=\left(0,0,\ldots,0,a\right)^T$ (a – значення, яке змінюється залежно від n) і $f\left(x^*\right)$ – змінюється залежно від n; $x_0=\left(1,1,\ldots,1\right)^T$.

Таблиця 3

Порівняльні характеристики роботи методів (10), (2) – (3)

	(10)		(2) – (3)		
n	*h*	*K*	*p*	*h*	*K*
4	18	114	2	14	295
10	42	572	2	25	593
10			5	18	428
20	77	1888	5	27	633
20			10	18	520
100	239	28253	20	34	1321
100			50	24	1685
1000	2155	2196218	100	57	6844
1000			500	26	13984

Задача 4. Функція Райдана 1

$$f\left(x\right)=\sum_{i=1}^{n}\frac{i}{10}\left(e^{x_i}-x_i\right).$$

$x^*=\left(0,0,\ldots,0\right)^T$, $f\left(x^*\right)$ – змінюється залежно від n; $x_0=\left(1,1,\ldots,1\right)^T$.

Таблиця 4

Порівняльні характеристики роботи методів (10), (2) – (3)

	(13)		(2) – (3)		
n	*h*	*K*	*p*	*h*	*K*
4	84	425	2	56	171
10	84	935	2	40	123
10			5	22	138
20	86	1831	5	42	261
20			10	54	609
100	262	27110	20	55	1270
100			50	26	1430
1000	2300	2316114	100	56	6754
1000			500	24	12927

5. ВИСНОВКИ

Запропоновано схему побудови нового класу методів розв'язування задач безумовної мінімізації функції багатьох змінних. Провели теоретичні та числові дослідження цього алгоритму. Запропонований метод побудований за принципом градієнтного методу, проте замість вектора-градієнта використоно лише декілька часткових похідних мінімізуючої функції цього вектора. Доведено, що метод має лінійну збіжність.

На підставі проведених обчислень і порівнянні отриманих результатів з'ясували, що запропонований метод спуску за кількістю обчислень та ітерацій переважає метод, зі збільшенням розмірності функції ефективність запропонованого алгоритму, в сенсі кількості обчислень, зростає.

Подані результати числових обчислень узгоджуються з доведеними теоретичними твердженнями.

Список використаної літератури

1. *Бейко І. В.* Задачі, методи та алгоритми оптимізації / І. В. Бейко, П. М. Зінько, О. Г. Наконечний. – Київ: ВПЦ Київський університет, 2012. – 800 с.
2. *Васильев Ф. П.* Методы оптимизации / Ф. П. Васильев. – Москва: Факториал Пресс, 2002. – 824 с.
3. *Пшеничный Б. Н.* Численные методы в экстремальных задачах / Б. Н. Пшеничный, Ю. М. Данилин. – Москва: Наука, 1975. – 320 с.
4. *Andrei N.* An unconstrained optimization test functions collection / Neculai Andrei // Advanced modelling and optimization. – 2008. – Vol. 10, № 1. – P. 147–161.

Стаття: надійшла до редколегії 16.11.2016
доопрацьована 19.04.2017
прийнята до друку 14.06.2017

ON SOME DESCENT METHO FOR FUNCTION MINIMIZATION

M. Bartish, N. Ogorodnyk

Ivan Franko National University of Lviv,
Universytetska Str., 1, Lviv, 79000, e-mail: ktop@lnu.edu.ua

We consider gradient method for constructing new class of algorithm for function minimization. We use some combination of partial derivatives for constructing new descent vector. The receiving vector is used for building the new descent methods. We proved the theoretical convergence of new methods. Some numerical experiments have been done and conclusions of efficiency have been made . We suggested the possibility of using of new methods.

Key words: unconstrained optimization, gradient method, coordinate descent method.