Журнал обчислювальної та прикладної математики | 2013, №4(114) | Journal of Computational & Applied Mathematics

УДК 516.9

ТРИКРОКОВИЙ МЕТОД З ПОРЯДКОМ ЗБІЖНОСТІ $1+\sqrt{2}$ ДЛЯ РОЗВ'ЯЗУВАННЯ ЗАДАЧ МІНІМІЗАЦІЇ

В. Я. Бартіш, Н. П. Огородник

Резюме. В роботі запропоновано трикроковий алгоритм розв'язування задач мінімізації нелінійної функції багатьох змінних, який використовує ідею побудови трикрокових методів і базується на методі із швидкістю збіжності $1+\sqrt{2}$. Досліджено швидкість збіжності запропонованого алгоритму, встановлено порядок збіжності. Проведено числові експерименти на функціях різних типів, які підтвердили перевагу трикрокового алгоритму над базовим в сенсі кількості обчислень.

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. Для розв'язування актуальних проблем побудовано широкі класи алгоритмів. Різні методи виявляють свою ефективність на різних класах задач і на даний час не існує універсального алгоритму. При визначенні ефективності алгоритму враховується критерій точності результату, швидкість збіжності та складність алгоритму [1, 2].

Запропоновано алгоритм розв'язування задач мінімізації функцій, який базується на методі із швидкістю збіжності $1+\sqrt{2}$ [3]. На кожному кроці за допомогою даного методу обчислюються два проміжні наближення до розв'язку, а наступне наближення алгоритму шукається як точка мінімуму на прямій, що з'єднує отримані точки. Показано, що швидкість збіжності даного алгоритму є $1+\sqrt{2}$ із знаменником збіжності меншим, ніж знаменник збіжності базового методу.

Проведено тестування даного алгоритму на низці відомих задач, зроблено його порівняння з базовим.

1. Формулювання задачі

Розглянемо задачу безумовної мінімізації функції багатьох змінних

$$f(x) \to \min, \tag{1}$$

де $x \in R^n$ і $f(x) \in C^3(R^n)$.

Одним з класичних методів розв'язування задачі (1) є метод Ньютона. До недоліків методів Ньютона відносять трудомісткість кожної ітерації. У [3] для розв'язування задачі (1) запропоновано метод, який побудовано

на основі методу Ньютона, і у якому більш ефективно використовується інформація про матрицю других похідних, а саме

$$x_{k+1} = x_k - (f''(\theta_k))^{-1} f'(x_k), \tag{2}$$

$$\theta_{k+1} = x_{k+1} - \frac{1}{2}(f''(\theta_k))^{-1} f'(x_{k+1}), \theta_0 = x_0, k = 0, 1, \ldots. \tag{3}$$

Показано, що швидкість збіжності даного методу дорівнює $1 + \sqrt{2}$.

Використовуючи ідею побудови методів з [4], на основі методу (2)–(3) досліджено новий трикроковий метод вигляду

$$x_1 = x_0 - (f''(\theta_0))^{-1} f'(x_0), \theta_0 = x_0, \tag{4}$$

$$u_k = x_k - (f''(\theta_{k-1}))^{-1} f'(x_k), \theta_k = \frac{x_k + u_k}{2}, \tag{5}$$

$$v_k = x_k - (f''(\theta_k))^{-1} f'(x_k), \tag{6}$$

$$x_{k+1} = \arg\min_{\gamma} f(v_k + \gamma(u_k - v_k)), k = 1, 2, \ldots. \tag{7}$$

На кожному кроці додатково проводимо мінімізацію функції однієї змінної, формула (7).

2. Обгрунтування збіжності

Теорема 1. *Нехай виконуються умови*

1. $f(x) \in C^3(D)$, $D = \{x \in R^n : f(x) \le f(x_0)\}$;

2. $\forall x, y \in D$: $\|f'''(x)\| \le M_3$ *і* $f'''(x)$ *задовольняє умову Ліпшіца*

$$\|f'''(x) - f'''(y)\| \le N\|x - y\|,$$

де $0 < N, M_3 < \infty$;

3. $\forall x \in D$, $h \in R^n$

$$m\|h\|^2 \le (f''(x)h, h) \le M\|h\|^2,$$

де $0 < m < M < \infty$.

Тоді послідовність $\{x_k\}$ *коректно визначена за (4)–(7) та існує константа* $0 < C < \infty$ *така, що має місце оцінка*

$$(f(x_{k+1}) - f(x^*)) \le C\beta_{k+1}(f(x_k) - f(x^*))^2 (f(x_{k-1}) - f(x^*)),$$

де x^* *— розв'язок задачі (1),* $\beta_{k+1} \in (0, 1]$, $k = 1, 2, \ldots$.

Якщо початкове наближення x_0 *вибране так, що виконується умова*

$$\mu = \sqrt{C}(f(x_0) - f(x^*)) < 1, \tag{8}$$

то

$$f(x_k) - f(x_*) \le \prod_{i=1}^{k} \beta_i^{l_{k+1-i}} \mu^{D_k} (f(x_0) - f(x_*)), k = 1, 2, \ldots, \tag{9}$$

де $D_{k+1} = 2D_k + D_{k-1} + 2$, *причому* $D_0 = 0$, $D_1 = 1$, $l_{k+1} = 2l_k + l_{k-1}$, $l_0 = 0$, $l_1 = 1$.

Доведення. Згідно з умовою 3) функція $f(x)$ сильно опукла. Отже, розв'язок задачі (1) існує та єдиний, при цьому справедлива оцінка [2]

$$\|f''(x)^{-1}\| \le \frac{1}{m}.$$

Припустимо, що наближення x_k, $k>1$ до розв'язку задачі (1), знайдено. Тоді враховуючи (6), маємо

$$\|v_k - x^*\| = \|x_k - (f''(\theta_k))^{-1} f'(x_k) - x^*\| \le \|(f''(\theta_k))^{-1}\| \times$$
$$\times \|f''(\theta_k)(x_k - x^*) - f'(x_k)\| \le \frac{1}{m}\|f''(\theta_k)(x_k - x^*) - (f'(x_k) - f'(x^*))\|. \quad (10)$$

Використовуючи розклад $f''(\theta_k)$ та $f'(x_k)$ в околі точки x^* та умову 2) теореми, отримаємо (при умові, що $\tau_1, \tau_2 \in (0,1)$)

$$\|v_k - x^*\| \le \frac{1}{m}\|f''(x^*)(x_k - x^*) + f'''(x^* + \tau_1(\theta_k - x^*))(\theta_k - x^*)(x_k - x^*) -$$
$$- f''(x^*)(x_k - x^*) - \frac{1}{2} f'''(x^* + \tau_2(x_k - x^*))(x_k - x^*)^2\| \le$$
$$\le \frac{N}{m}\|x^* + \tau_1(\theta_k - x^*) - x^*\|\|\theta_k - x^*\|\|x_k - x^*\| + \frac{N}{2m}\|x^* + \tau_2(x_k - x^*) - x^*\| \times$$
$$\times \|x_k - x^*\|^2 + \frac{1}{m}\|f'''(x^*)(\theta_k - x^*)(x_k - x^*) - \frac{1}{2} f'''(x^*)(x_k - x^*)^2\| \le$$
$$\le \frac{N}{m}\|\theta_k - x^*\|^2\|x_k - x^*\| + \frac{N}{2m}\|x_k - x^*\|^3 +$$
$$+ \frac{M_3}{2m}\|x_k - x^*\|\|u_k - x^*\|. \quad (11)$$

Використавши (5) та роклад $f''(\theta_{k-1})$ і $f'(x_k)$ в околі x^*, запишемо оцінку

$$\|u_k - x^*\| = \|x_k - (f''(\theta_{k-1}))^{-1} f'(x_k) - x^*\| \le \|(f''(\theta_{k-1}))^{-1}\| \times$$
$$\times \|f''(\theta_{k-1})(x_k - x^*) - f'(x_k)\| \le \frac{1}{m}\|f''(\theta_{k-1})(x_k - x^*) - (f'(x_k) - f'(x^*))\| \le$$
$$\le \frac{1}{m}\|f''(x^*)(x_k - x^*) + f'''(x^* + \tau_1(\theta_{k-1} - x^*))(\theta_{k-1} - x^*)(x_k - x^*) -$$
$$- f''(x^*)(x_k - x^*) - \frac{1}{2} f'''(x^* + \tau_2(x_k - x^*))(x_k - x^*)^2\| \le$$
$$\le \frac{1}{m}\|f'''(x^* + \tau_1(\theta_{k-1} - x^*))\|\|\theta_{k-1} - x^*\|\|x_k - x^*\| +$$
$$+ \frac{1}{2m}\|f'''(x^* + \tau_2(x_k - x^*))\|\|x_k - x^*\|^2 \le \frac{M_3}{m}\|\theta_{k-1} - x^*\|\|x_k - x^*\| +$$
$$+ \frac{M_3}{2m}\|x_k - x^*\|^2. \quad (12)$$

Використавши умови теореми, отримаємо

$$\frac{m}{2}\|x - x^*\|^2 \le f(x) - f(x_*) \le \frac{M}{2}\|x - x^*\|^2.$$

Тоді, врахувавши (7), маємо

$$\|u_k - x^*\|^2 \le \frac{2}{m}(f(u_k) - f(x^*)) \le \frac{2}{m}(f(x_k) - f(x^*)) \le \frac{M}{m}\|x_k - x^*\|^2$$

або

$$\|u_k - x^*\| \le \sqrt{\frac{M}{m}}\|x_k - x^*\|.$$

Використаємо (5) та запишемо

$$\|\theta_k - x^*\| = \|\frac{u_k + x_k}{2} - x^*\| \le \frac{1}{2}(\|u_k - x^*\| + \|x_k - x^*\|) \le$$

$$\le \frac{1}{2}(\sqrt{\frac{M}{m}} + 1)\|x_k - x^*\|. \tag{13}$$

Продовжимо оцінку (11), використавши (12) та (13) та умову 2):

$$\|v_k - x^*\| \le \frac{N}{m}\|\theta_k - x^*\|^2\|x_k - x^*\| + \frac{N}{2m}\|x_k - x^*\|^3 +$$

$$+\frac{M_3}{2m}\|x_k - x^*\|(\frac{M_3}{m}\|\theta_{k-1} - x^*\|\|x_k - x^*\| + \frac{M_3}{2m}\|x_k - x^*\|^2) \le$$

$$\le \frac{N}{4m}(\sqrt{\frac{M}{m}} + 1)^2\|x_k - x^*\|^3 + \frac{N}{2m}\|x_k - x^*\|^3 +$$

$$+(\frac{M_3}{2m})^2(\sqrt{\frac{M}{m}} + 1)\|x_k - x^*\|^2\|x_{k-1} - x^*\| + (\frac{M_3}{2m})^2\|x_k - x^*\|^3 \le$$

$$\le (\frac{N}{4m}(\sqrt{\frac{M}{m}}+1)^2 + \frac{N}{2m} + (\frac{M_3}{2m})^2(\sqrt{\frac{M}{m}}+1) + (\frac{M_3}{2m})^2)\|x_{k-1} - x^*\|\|x_k - x^*\|^2 =$$

$$= \tilde{C}\|x_{k-1} - x^*\|\|x_k - x^*\|^2,$$

де $\tilde{C} = \frac{N}{4m}(\sqrt{\frac{M}{m}} + 1)^2 + \frac{N}{2m} + (\frac{M_3}{2m})^2(\sqrt{\frac{M}{m}} + 1) + (\frac{M_3}{2m})^2 < \infty$.

Тоді

$$f(v_k) - f(x_*) \le \frac{M}{2}(\tilde{C})^2\|x_{k-1} - x^*\|^2\|x_k - x^*\|^4 \le$$

$$\le \frac{M}{2}(\tilde{C})^2(\frac{2}{m})^3(f(x_{k-1}) - f(x_*))(f(x_k) - f(x_*))^2 =$$

$$= C(f(x_{k-1}) - f(x_*))(f(x_k) - f(x_*))^2. \tag{14}$$

З формул (7) і (14) випливає

$$f(x_{k+1}) - f(x^*) \le \beta_{k+1}C(f(x_{k-1}) - f(x_*))(f(x_k) - f(x_*))^2,$$

де $\beta_{k+1} \in (0, 1]$.

За допомогою методу математичної індукції, враховуючи (8), доведемо, що оцінка (9) виконується для довільного k. Спершу запишемо оцінку для $f(x_1) - f(x^*)$:

$$f(x_1) - f(x^*) \le \frac{M}{2}\|x_1 - x^*\|^2 = \frac{M}{2}\|x_0 - x^* - (f''(x_0))^{-1}f'(x_0)\|^2 \le$$

$$\le \frac{M}{2m^2}\|f''(x_0) - f''(x^* + \tau(x_0 - x^*))\|^2\|x_0 - x^*\|^2 \le \frac{MM_3}{m^2}\|x_0 - x^*\|^4 \le$$

$$\le \frac{MM_3}{m^2}(\frac{2}{m})^2\|f(x_0) - f(x^*)\|^2 \le \sqrt{C}(f(x_0) - f(x^*))\|^2 \le \mu^{D_1}(f(x_0) - f(x^*)),$$

де $\sqrt{C} = max\{\frac{MM_3}{m^2}(\frac{2}{m})^2, \tilde{C}\sqrt{\frac{M}{m}}\frac{2}{m}\} < \infty$.

Якщо $k=1$, одержимо

$$f(x_2)-f(x^*)\le \beta_2 C(f(x_0)-f(x^*))(f(x_1)-f(x^*))^2\le$$
$$\le \beta_2 C(f(x_0)-f(x^*))\beta_1^{2l_1}\mu^{2D_1}(f(x_0)-f(x^*))^2\le$$
$$\le \mu^2\mu^{2D_1}\beta_2^{l_1}\beta_1^{2l_1}(f(x_0)-f(x^*))\le$$
$$\le \prod_{i=1}^{2}\beta_i^{l_{3-i}}\mu^{D_2}(f(x_0)-f(x^*)).$$

Припустимо, що дана оцінка виконується для деякого $k>1$, тобто

$$f(x_k)-f(x^*)\le \prod_{i=1}^{k}\beta_i^{l_{k+1-i}}\mu^{D_k}(f(x_0)-f(x^*)).$$

Тоді для $k+1$, отримаємо

$$f(x_{k+1})-f(x^*)\le \beta_{k+1}C(f(x_{k-1})-f(x^*))(f(x_k)-f(x^*))^2\le,$$
$$\le \beta_{k+1}C(\prod_{i=1}^{k}\beta_i^{l_{k+1-i}})^2\mu^{2D_k}(f(x_0)-f(x^*))^2\prod_{i=1}^{k-1}\beta_i^{l_{k-i}}\mu^{D_{k-1}}(f(x_0)-f(x^*))=$$
$$=\beta_{k+1}\prod_{i=1}^{k}\beta_i^{2l_{k+1-i}}\prod_{i=1}^{k-1}\beta_i^{l_{k-i}}\mu^{2D_k+D_{k-1}+2}(f(x_0)-f(x^*))=$$
$$=\prod_{i=1}^{k+1}\beta_i^{l_{k+2-i}}\mu^{D_{k+1}}(f(x_0)-f(x^*)).$$

Отже, за методом математичної індукції маємо, що послідовність наближень, побудованих за формулами (4)–(7), збігаєтся до x_*. □

3. Чисельні експерименти

Нами розглянуто ряд прикладів і проведено порівняння запропонованого трикрокового алгоритму (4)–(7) з базовим методом із швидкістю збіжності $1+\sqrt{2}$ (2)–(3). Для забезпечення вибору довільного початкового наближення у методи було введено крокові множники.

Обчислення проводили до виконання умови

$$\|x_{k+1}-x_k\|\le\varepsilon,$$

де $\varepsilon=10^{-8}$. В таблиці наведено кількість ітерацій N та кількість обчислень, еквівалентних обчисленню функції K, затрачених для отримання наближеного розв'язку наведених задач з [5].

1. Штрафна функція

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

$$x^0=(10,10,\dots,10);\qquad x^0=(5,5,\dots,5);\qquad n=2,3,\dots.$$

Розв'язок залежить від значення n.

2. Розширена функція Вайта і Холста

$$f(x) = \sum_{i=1}^{n/2} [100(x_{2i} - x_{2i-1}^3)^2 + (1 - x_{2i-1})^2];$$

$$x^0 = (-1, 0.8, -1, 0.8 \ldots, -1, 0.8); \qquad x^0 = (0, 0, \ldots, 0); \qquad n = 2, 4, \ldots;$$

$$x^* = (1, 1, \ldots 1); \qquad f(x^*) = 0.$$

3. Розширена функція Пауела

$$f(x) = \sum_{i=1}^{n/4} [(x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4];$$

$$x^0 = (3, -1, 0, 1, \ldots, 3, -1, 0, 1); \qquad x^0 = (30, -10, 0, 10, \ldots, 30, -10, 0, 10);$$

$$n = 4, 8, 12, 16, \ldots; \qquad x^* = (0, 0, \ldots, 0); \qquad f(x^*) = 0.$$

4. Штрафна функція 1

$$f(x) = 10^{-5} \sum_{i=1}^{n} (x_i - 1)^2 + (\sum_{i=1}^{n} x_i^2 - 0.25)^2;$$

$$x^0 = (10, 10, \ldots, 10); \qquad x^0 = (5, 5, \ldots, 5); \qquad n = 2, 3, \ldots.$$

Розв'язок залежить від значення n.

Таблиця 1

ф-ція	x_0	$n = 4$				$n = 100$			
		(2)–(3)		(4)–(7)		(2)–(3)		(4)–(7)	
		N	K	N	K	N	K	N	K
1	1	10	145	7	467	12	61901	6	31878
1	2	9	131	7	433	10	51601	6	31876
2	1	23	585	19	893	23	234729	19	194459
2	2	19	483	12	605	19	193923	12	122824
3	1	35	495	2	119	38	195801	2	10679
3	2	39	551	2	119	43	221551	2	10679
4	1	11	159	2	120	12	62514	2	10680
4	2	10	145	2	149	13	67051	2	10680

Висновки

В даній статті запропоновано новий метод розв'язування задач безумовної мінімізації функцій багатьох змінних. Метод побудовано на основі методу із швидкістю збіжності $1 + \sqrt{2}$, використовуючи ідею побудови трикрокових алгоритмів. В якості проміжних наближень у даному трикроковому методі використано наближення отримані за методом (2)–(3). Отже, на кожній ітерації алгоритму потрібно додатково проводити одновимірну

мінімізацію. При невеликих розмірностях функції, кількість обчислень затрачених на одновимірну мінімізацію може впливати на ефективність запропонованого методу (як видно з таблиці). Із збільшенням розмірності функції ($n > 10$) кількість обчислень затрачених на одновимірну мінімізацію стає несуттєвою в порівнянні із кількістю обчислень, затрачених на знаходження другої похідної запропонований алгоритм починає працювати краще. Проведено теоретичні та числові дослідження даного методу. Доведено, що швидкість збіжністі методу (4)–(7) є $1+\sqrt{2}$.

В числових експериментах запропонований трикроковий метод показав суттєві переваги для великої кількості тестових функцій, особливо для функцій з виродженим в точці розв'язку гессіаном (функції 3 і 4).

Література

1. Васильев Ф. П. Методы оптимизации / Ф. П. Васильев. — М. : Факториал Пресс, 2002. — 824 с.
2. Пшеничный Б. Н., Данилин Ю. М.Численные методы в экстремальных задачах — М. : Наука, 1975. — 320с.
3. Бартіш М. Я., Ковальчук О., Огородник Н. Про один ітераційний метод розв'язування функціональних рівнянь // ДАН УРСР. Сер. А. — 1968. — №5. — С. 387–391.
4. Бартіш М. Трикрокові методи розв'язування задач безумовної мінімізації // Вісн. Львів. ун-ту. Сер. прикл. матем. та інформ. — 2007. — Вип. 13. — С. 3–10.
5. Andrei N. An unconstrained optimization test functions collection // Advanced modelling and optimization. — 2008. — V. 10, №1. — P. 147–161.

Факультет прикладної математики та інформатики,
Львівський національний університет імені Івана Франка,
вул. Університетська, 1, Львів, 70000, Україна.

Надійшла 01.10.2013