УДК 519.6

В. Я. Бартіш, Н. П. Огородник

ТРИКРОКОВИЙ ІТЕРАЦІЙНО-РІЗНИЦЕВИЙ МЕТОД З ПОРЯДКОМ ЗБІЖНОСТІ $1+\sqrt{2}$

V. Ya. Bartish, N. Ph. Ogorodnyk. *Iterative difference three step method with* $1+\sqrt{2}$ *convergence rate*, Mat. Stud. **43** (2015), 220–224.

The new three step method for solving unconstrained minimization problems is proposed. The method use the idea of building of three-step methods and is based on a method with the rate of convergence $1+\sqrt{2}$. The rate of convergence for the new method is investigated. Numerical investigation is conducted on the test functions. The result of numerical experiments shows that three step method is more effective in sense of amount of calculations. The efficiency of the method is growing with increasing of function's dimension.

В. Я. Бартиш, Н. Ф. Огородник. *Трёхшаговый итерационно-разностный метод с порядком сходимости* $1+\sqrt{2}$ // Мат. Студії. – 2015. – Т.43, №2. – С.220–224.

Исследуется трёхшаговый итерационно-разностный метод решения задач безусловной минимизации функций многих переменных. В трёхшаговом методе в качестве базисного предлагается использовать разностный метод с скоростью сходимости $1+\sqrt{2}$. Исследована скорость сходимости метода. Проведен вычислительный эксперимент. Метод продемонстрировал свою эффективность по количеству вычислений в сравнении с базисным методом. Эффективность метода повышается с увеличением размерности функции, особенно для функции с вырожденной в точке решения матрицей Гессе.

1. Вступ. Математичні моделі багатьох фізичних чи економічних процесів зводяться до розв'язування задач оптимізації, зокрема, задач безумовної мінімізації. В літературі велика увага приділяється ітераційним методам розв'язування таких задач (див., наприклад, [1], [2]). Різні методи виявляють свою ефективність на різних класах задач, при цьому методи мають свої переваги і недоліки такі, як вибір початкового наближення, швидкість збіжності, трудомісткість окремої ітерації, тощо. Оскільки універсального алгоритму нема та й не може бути, то й надалі актуальною залишатиметься проблема побудови ефективних, у тому чи іншому сенсі, методів.

В статті розглянуто ітераційно-різницевий метод розв'язування задачі безумовної мінімізації, який використовують, коли матрицю других похідних складно або не можливо обчислити, а саме, запропоновано трикроковий ітераційний метод, побудований на основі різницевого методу зі швидкістю збіжності $1+\sqrt{2}$ з [3]. У порівнянні з основним методом, у трикроковому методі необхідно проводити додатково одновимірну оптимізацію, що не істотно впливає на кількість обчислень особливо для задач великих розмірностей. Під час чисельних екпериментів, метод показав свою ефективність у сенсі кількості обчислень на функціях різних типів.

2010 *Mathematics Subject Classification:* 90C52.
Keywords: tree step method; minimization problem.
doi:10.15330/ms.43.2.220-224

2. Формулювання задачі. Розглянемо задачу

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

де $x \in \mathbb{R}^n$, $f \in C^1(\mathbb{R}^n)$. Для розв'язування задачі (1) можна використовувати низку різних методів ([4]). У статті [3] запропоновано різницевий метод з порядком збіжності $1+\sqrt{2}$. Метод має наступний вигляд

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

$$\theta_k = x_k - (f'(x_{k-1}, \theta_{k-1}))^{-1} f'(x_k),\ k \in \{1, 2, \ldots\}, \tag{3}$$

де f' — градієнт функції (тобто, вектор-стовпець часткових похідних), а $f'(\cdot,\cdot)$ — матриця відповідних поділених різниць. Вважаємо, що $\theta_0 = \tilde{x_0}$ і $\|\tilde{x_0} - x_0\| \leq \alpha \|f'(x_0)\|$, де $\alpha \in (0, 1]$. У цій статті ми пропонуємо метод розв'язування задачі (1), який використовує метод (2)–(3) для відшукання двох проміжних наближень, а наступне наближення запропонованого методу шукаємо як мінімум на прямій, що з'єднує отримані проміжні наближення ([5]). Тобто, розглядаємо трикрокову ітераційно-різницеву схему такого вигляду

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

$$\theta_k = x_k - (f'(x_{k-1}, \theta_{k-1}))^{-1} f'(x_k), \tag{5}$$

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

$$x_{k+1} = u_k + \gamma_k(\theta_k - u_k), \tag{7}$$

$$\text{де } f(u_k + \gamma_k(\theta_k - u_k)) = \min\{f(u_k + \gamma(\theta_k - u_k))\colon \gamma \in \mathbb{R}\},\ k \in \{1, 2, \ldots\},$$

а θ_0 вибираємо так само як і в схемі (2)–(3). Послідовність $\{x_k\}$, отримана за схемою (4)–(7), має кращі властивості в сенсі швидкості збіжності, ніж послідовність $\{x_k\}$, отримана за (2)–(3). Обчислювальні затрати на кожній ітерації запропонованого методу істотно не зростають у порівнянні з такими ж затратами за схемою (2)–(3).

Надалі використовуватимемо наступні означення поділених різниць першого і другого порядку ([1])

$$f'(x, y)(x - y) = f'(x) - f'(y), \tag{8}$$

$$f'(x, y, z)(x - y) = f'(x, z) - f'(y, z). \tag{9}$$

Власне, поділені різниці першого порядку обчислюємо за формулою ([1]) $f'(x,y)_{i,j} = \big(f'_{x_i}(x_1, ..., x_j, y_{j+1}, ..., y_n) - f'_{x_i}(x_1, ..., x_{j-1}, y_j, ..., y_n)\big)/(x_j - y_j)$.

3. Обґрунтування збіжності.

Теорема. *Нехай $f \in C^1(\mathbb{R}^n)$, $x_0 \in \mathbb{R}^n$ — початкове наближення, $D = \{x \in \mathbb{R}^n \colon f(x) \leq f(x_0)\}$, і виконуються умови:*
1) f сильно опукла і $(\exists m > 0)(\forall x, y \in D)\colon (f'(x,y)(x-y), x-y) \geq m\|x-y\|^2$;
2) $f'(x)$ має поділені різниці першого і другого порядків в D, які задовольняють умови

$$(\exists M \in (0, +\infty))(\exists N \in (0, +\infty))(\forall x, y, z \in D)\colon \|f'(x,y)\| \leq M,\ \|f'(x,y,z)\| \leq N;$$

3) $(\exists B \in (0, +\infty))(\forall x, y \in D)$(існує обернена матриця $(f'(x,y))^{-1}$): $\|(f'(x,y))^{-1}\| \leq B$.
Тоді існує стала $C \in (0, +\infty)$ така, що виконується нерівність

$$f(x_k) - f(x_*) \leq \beta_k C^2 (f(x_{k-1}) - f(x_*))^2 (f(x_{k-2}) - f(x_*)), \quad \text{де } \beta_k \in (0, 1] \quad (k \geq 2).$$

Якщо ж для x_0 виконується умова $\mu = C(f(x_0) - f(x_*)) < 1$, то

$$f(x_k) - f(x^*) \leq \Big(\prod_{i=1}^{k} \beta_i^{l_{k+1-i}}\Big)\mu^{D_k}(f(x_0) - f(x^*)) \quad (k \geq 2),$$

де $D_0 = 0$, $D_1 = 1$, $D_k = 2D_{k-1} + D_{k-2} + 2$ і $l_0 = 0$, $l_1 = 1$, $l_k = 2l_{k-1} + l_{k-2}$ $(k \geq 2)$.

Доведення. Існування і єдиність розв'язку задачі (1) випливають з умови сильної опуклості функції $f(x)$ ([2]).

Використовуючи умову 2) і розклад функції $f(x)$ в ряд Тейлора, отримаємо

$$f(x) - f(x_*) = (f'(\tilde{z}), x - x^*) = (f'(\tilde{z}, x^*)(\tilde{z} - x^*), x - x^*) \leq M\|x - x^*\|^2,$$

де $\tilde{z} = x + \xi(x^* - x)$, $\xi \in (0,1)$. З сильної опуклості f випливає ([2]), що $(\forall x, y \in D)$: $f(x) - f(y) \geq (f'(y), x - y) + m\|x - y\|^2$. Отже, при $y = x^*$ отримаємо,

$$\|x - x^*\|^2 \leq (f(x) - f(x^*))/m. \tag{10}$$

Припустимо, що наближення x_k до розв'язку задачі (1) знайдено. Тоді, використовуючи (6), (8), (9) отримаємо

$$u_k - x^* = x_k - x^* - f'(x_k, \theta_k)^{-1} f'(x_k) = (f'(x_k, \theta_k))^{-1}(f'(x_k, \theta_k) - f'(x_k, x^*))(x_k - x^*) =$$
$$= (f'(x_k, \theta_k))^{-1} f'(x_k, \theta_k, x^*)(\theta_k - x^*)(x_k - x^*).$$

Скориставшись умовами 2) і 3), звідси отримаємо

$$\|u_k - x^*\| \leq BN\|\theta_k - x^*\|\|x_k - x^*\|. \tag{11}$$

Використовуючи знову умови 2) і 3), маємо

$$\|\theta_k - x^*\| = \|x_k - x^* - (f'(x_{k-1}, \theta_{k-1}))^{-1} f'(x_k)\| =$$
$$= \|(f'(x_{k-1}, \theta_{k-1}))^{-1}\|\|f'(x_{k-1}, \theta_{k-1}) - f'(x_k, x^*)\|\|x_k - x^*\| \leq$$
$$\leq B(\|f'(x_{k-1}, \theta_{k-1})\| + \|f'(x_k, x^*)\|)\|x_k - x^*\| \leq 2BM\|x_k - x^*\|.$$

З іншого боку, використаємо (9), умови 2) і 3), а також отриману вище оцінку

$$\|\theta_k - x^*\| \leq B\|f'(x_{k-1}, \theta_{k-1}) - f'(x_{k-1}, x^*) + f'(x_{k-1}, x^*) - f'(x_k, x^*)\|\|x_k - x^*\| \leq$$
$$\leq B\|f'(x_{k-1}, \theta_{k-1}, x^*)(\theta_{k-1} - x^*) + f'(x_{k-1}, x^*, x_k)(x_{k-1} - x_k)\|\|x_k - x^*\| \leq BN(\|\theta_{k-1} - x^*\| +$$
$$+ \|x_{k-1} - x_k\|)\|x_k - x^*\| \leq BN(\|\theta_{k-1} - x^*\| + \|x_k - x^*\| + \|x_{k-1} - x^*\|)\|x_k - x^*\| \leq$$
$$\leq BN(2BM\|x_{k-1} - x^*\| + 2\|x_{k-1} - x^*\|)\|x_k - x^*\| = 2BN(1 + BM)\|x_{k-1} - x^*\|\|x_k - x^*\|.$$

Продовжимо оцінку (11), $\|u_k - x^*\| \leq 2(BN)^2(1 + BM)\|x_{k-1} - x^*\|\|x_k - x^*\|^2$. Тоді,

$$f(u_k) - f(x^*) \leq M\|u_k - x^*\|^2 \leq 4M(BN)^4(1 + BM)^2\|x_{k-1} - x^*\|^2\|x_k - x^*\|^4. \tag{12}$$

Нехай $C^2 \geq \frac{4M(BN)^4(1+BM)^2}{m^3}$. З нерівності (12) за допомогою нерівності (10) отримуємо, що $f(u_k) - f(x^*) \leq C^2(f(x_{k-1}) - f(x^*))(f(x_k) - f(x^*))^2$.

З (7) зрозуміло, що $f(x_{k+1}) - f(x_*) \leq \beta_{k+1}(f(u_k) - f(x_*))$, причому $\beta_{k+1} \in (0, 1]$, тоді $f(x_{k+1}) - f(x_*) \leq \beta_{k+1} C^2(f(x_{k-1}) - f(x^*))(f(x_k) - f(x^*))^2$. Використовуючи метод математичної індукції, доведемо, що оцінка теореми виконується для довільного k. Спершу, використовуючи (7), запишемо оцінку для $f(x_1) - f(x^*)$

$$f(x_1) - f(x^*) \leq M\|x_1 - x^*\|^2 = M\|x_0 - x^* - (f'(x_0, \theta_0))^{-1} f'(x_0)\|^2 \leq$$
$$\leq M\|f'(x_0, \tilde{x}_0)^{-1}\|^2\|f'(x_0, \tilde{x}_0) - f'(x_0, x^*)\|^2\|x_0 - x^*\|^2 \leq$$
$$\leq MB^2\|f'(x_0, \tilde{x}_0, x^*)\|^2\|\tilde{x}_0 - x^*\|^2\|x_0 - x^*\|^2 \leq$$

$$\le M(BN)^2\|f'(x_0)\|^2\|x_0-x^*\|^2 = M(BN)^2\|f'(x_0,x^*)\|^2\|x_0-x^*\|^4 \le M^3(BN)^2\|x_0-x^*\|^4 \le$$
$$\le \frac{M^3(BN)^2}{m^2}(f(x_0)-f(x^*))^2 \le C(f(x_0)-f(x^*))^2 = \mu^{D_1}(f(x_0)-f(x^*)).$$

При $k=1$ отримаємо

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

Припустимо, що оцінка $f(x_k)-f(x^*) \le \prod_{i=1}^{k}\beta_i^{l_{k+1-i}}\mu^{D_k}(f(x_0)-f(x^*))$ виконується для деякого $k>1$. Тоді, для кроку $k+1$ отримаємо

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

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

Відзначимо, що вибір початкового наближення x_0, яке б задовольняло умову теореми, є досить складною проблемою, тому на практиці доцільно використовувати схему вигляду

$$x_1 = x_0 - \alpha_0(f'(x_0,\theta_0))^{-1}f'(x_0), \quad \theta_k = x_k - \alpha_k(f'(x_{k-1},\theta_{k-1}))^{-1}f'(x_k), \tag{13}$$
$$u_k = x_k - \lambda_k(f'(x_k,\theta_k))^{-1}f'(x_k), \quad x_{k+1} = u_k + \gamma_k(\theta_k - u_k), \tag{14}$$
$$\text{де } f(u_k+\gamma_k(\theta_k-u_k)) = \min\{f(u_k+\gamma(\theta_k-u_k))\colon \gamma\in\mathbb{R}\},\ k\in\{1,2,\ldots\},$$

а параметри α_k, λ_k повинні забезпечувати монотонне спадання функції. В схемі (13)–(14), за певних умов, доведена у теоремі збіжність буде справджуватися локально.

4. Апробація методу. Нами розглянуто низку прикладів і проведено порівняння методу (13)–(14) з базовою схемою (2)–(3), у якій також введено крокові множники α_k, λ_k, для забезпечення можливості вибору довільного початкового наближення. Обчислення проводилися до виконання умови $\|x_{k+1}-x_k\| \le \varepsilon$, $\varepsilon = 10^{-8}$. У таблиці 1 наведено кількість ітерацій N та K — кількість обчислень, еквівалентних до кількості обчислень значення функції $f(x)$, які були затрачені для отримання наближеного розв'язку наведених тестових завдань. Іншими операціями нехтуємо, оскільки в обидвох схемах, при кожній ітерації вони практично однакові. Тестові завдання запозичено з [6].

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

$$x^0=(10,10,\ldots,10);\ x^0=(-10,-10,\ldots,-10);\ n\in\{2,3,\ldots\}.$$

Розв'язок залежить від значення n.

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

$$x^0 = (-1, 0.8, -1, 0.8 \ldots, -1, 0.8);\ x^0 = (0.5, -0.5, \ldots, 0.5, -0.5);\ n \in \{2, 4, \ldots\};$$
$$x^* = (1, 1, \ldots 1);\ f(x^*) = 0.$$

3. Розширена функція Пауела
$$f(x) = \sum_{i=1}^{n/4}[(x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4];$$
$$x^0 = (3, -1, 0, 1, \ldots, 3, -1, 0, 1);\ x^0 = (30, -10, 0, 10, \ldots, 30, -10, 0, 10);$$
$$n \in \{4, 8, 12, 16, \ldots\};\ x^* = (0, 0, \ldots, 0);\ f(x^*) = 0.$$

4. Штрафна функція 1 $f(x) = 10^{-5}\sum_{i=1}^{n}(x_i - 1)^2 + (\sum_{i=1}^{n} x_i^2 - 0.25)^2$;
$$x^0 = (10, 10, \ldots, 10);\ x^0 = (-10, -10, \ldots, -10);\ n \in \{2, 3, \ldots\}.$$

Розв'язок залежить від значення n.

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	11	242	7	311	13	131326	6	61214
1	2	13	287	7	303	14	141429	6	61212
2	1	29	936	19	971	28	283118	19	194219
2	2	17	591	12	462	17	171953	13	132724
3	1	38	836	4	194	41	414182	3	30856
3	2	44	1030	5	374	47	474856	5	51158
4	1	12	264	3	134	16	161632	3	30583
4	2	12	264	3	137	16	161632	3	30615

В наведених прикладах запропонована нами трикрокова схема демонструє свої переваги особливо для тестових функцій з виродженим в точці розв'язку гессіаном (функції 3 і 4). Зі збільшенням розмірності функції ($n > 10$) кількість обчислень затрачених на одновимірну мінімізацію стає малою у порівнянні з кількістю обчислень, затрачених на знаходження похідних і розділених різниць, що робить трикроковий алгоритм ефективнішим у сенсі кількості обчислень.

Література

1. Vasiljev F.P. Numerical methods for solving extremal problems. – M.: Nauka, 1988. – 552 p. (in Russian)
2. Pshenichnyj B.N., Danilin Yu.M. Numerical methods in extremal problems. – M.: Nauka, 1975. – 319 p. (in Russian)
3. Bartish M.Ya., Shcherbyna Yu.M. *On a difference method for solving nonlinear operator equations*// Dop. AN USSR., ser. A. – 1972. – №7. – P. 579–582. (in Ukrainian)
4. Beyko I.V., Zin'ko P.M., Nakonechnyj O.G. Problems methods and algorithms optimization. – VGC.: Kyiv. Univ., 2012. – 800 p. (in Ukrainian)
5. Bartish M.Ya., Kovalchuk O.V., Ogorodnyk N.P. *Three-step methods for soving unconstrained minimization problems*// Visnyk Lviv. Univ., ser. Appl. Math. Inf. – 2007. – №13. – P. 3–10. (in Ukrainian)
6. Koko J. *A conjugate gradient method with quasi-Newton approximation*// Aplicationes mathematicae. – 2000. – №27. – P. 153–165.

Ivan Franko National University of Lviv
gut.natalochka@gmail.com
ktop@franko.lviv.ua

Надійшло 22.01.2014