УДК 517.91

Є. М. Страхов, А. Т. Яровий
Одеський національний університет імені І. І. Мечникова

ДОСЛІДЖЕННЯ ЗБІЖНОСТІ БАГАТОКРОКОВОГО МЕТОДУ

Страхов Є. М., Яровий А. Т. Дослідження збіжності багатокрокового методу. У статті розглядається двокроковий алгоритм мінімізації функцій багатьох змінних при відсутності обмежень.
Ключові слова: багатокроковий метод, мінімізація, збіжність.

Страхов В. М., Яровой А. Т. Исследование сходимости многошагового метода. В статье рассматривается двухшаговый алгоритм минимизации функций многих переменных при отсутствии ограничений.
Ключевые слова: многошаговый метод, минимизация, сходимость.

Strahov E. M., Yarovoy A. T. Researching of a multi-stepped optimization algorithm. Considered a two-stepped multi-variable optimization algorithm for the problems without constraints.
Key words: multi-step algorithm, minimization, convergence.

Вступ. Задачі математичного програмування знаходять застосування у різноманітних областях людської діяльності, де є необхідним вибір одного з можливих способів дії, наприклад, при вирішенні проблем керування та планування виробничих процесів, у проектуванні і перспективному плануванні, у військовій справі і т. д. Задача математичного програмування полягає, як правило, у знаходженні екстремуму деякої функції на множині, що визначається лінійними та нелінійними обмеженнями у формі рівностей чи нерівностей. На сучасному етапі розвитку науки вже розроблено та досліджено багато підходів до розв'язку таких задач, одним з яких є методи спуску. Загальний принцип цих методів полягає у побудові послідовних напрямків спуску (тобто зменшення значень функції), виходячи з певної початкової точки (початкового наближення). Різні класи методів спуску визначаються, в першу чергу, способами побудови напрямків спуску.

Розрізняють методи нульового (при обчисленнях використовуються тільки значення функції в точках простору), першого (крім значень функції, обчислюється її перша похідна), другого (використовується друга похідна) порядків. Також методи поділяють на однокрокові та багатокрокові. Однокрокові алгоритми на (k+1)-й ітерації враховують лише ті значення функції та її похідних, які були отримані на попередньому кроці. Багатокрокові алгоритми використовують додатково інформацію, отриману на більш ранніх етапах процесу оптимізації. Класичним методом першого порядку є градієнтний метод.

Перед тим, як перейти до розв'язування конкретної екстремальної задачі, необхідно з'ясувати, якому з методів мінімізації віддати перевагу в даному випадку. На це питання однозначної відповіді немає, якщо тільки вона не є добре відомою тестовою задачею. Таким чином, перед нами постає питання порівняння методів. Один з можливих шляхів — порівнювати методи на основі досвіду та достатнього числа експериментів. Без цього не обходиться жоден спеціаліст, який використовує сучасну обчислювальну техніку. Другий шлях, який при цьому не виключає перший, — порівнювати якість методів на певних класах задач. Важливу роль тут відіграють апріорні характеристики методів. Ці характеристики повинні враховувати такі фактори, як об'єм та складність обчислень, швидкість збіжності, стійкість методу до похибок у обчисленнях, тривалість розрахунків та інші. Відзначимо, що дуже часто при розв'язуванні задач математичного програмування доводиться використовувати не один конкретний метод мінімізації, а їх комбінацію. Але, знову ж таки, через відсутність регулярного способу вибору напрямків, досвід та інтуїція все ще відіграють суттєву роль при чисельному розв'язуванні реальних задач. З іншого боку, ця "біла пляма" залишає широкий простір для подальших досліджень.

У даній статті пропонується новий двокроковий метод першого порядку, який можна застосовувати до задач мінімізації функцій при відсутності обмежень. Алгоритм, що буде розглянутий, у певному розумінні є комбінацією відомих класичних алгоритмів — методів змінної метрики та методу спряжених градієнтів. Будуть наведені основні характеристики цього алгоритму, виявлені деякі важливі його властивості, а також доведені теореми збіжності при різних умовах, що накладаються на цільову функцію.

1. Обчислювальна схема методу

Розглянемо задачу оптимізації деякої функції без обмежень

$$f(x) \to min, x \in R^n. \tag{1.1}$$

Для її розв'язання пропонується двокроковий метод з таким алгоритмом:

$$x_{k+1} = x_k + \beta_k s_k, k = 0, 1, ...,$$

$$s_0 = -f'(x_0), s_k = -H_k f'(x_k) + \xi_k s_{k-1}, \tag{1.2}$$

де $x_0, x_1, ..., x_k, ...$ — послідовні наближення, $s_0, s_1, ..., s_k, ...$ — напрямки спуску, β_k та ξ_k — числові параметри, H_k — матриці, що обчислюються рекурентним способом, як і у методах змінної метрики. У якості H_0 можна взяти довільну симетричну строго додатньо означену матрицю, наприклад, одиничну. Через певну кількість кроків проводиться операція відновлення матриці, тобто покладаємо $H_{k+1} = H_0$.

На практиці були розглянуті два варіанти обчислення матриць:

1) метод Девідона-Флетчера-Пауелла (ДФП)

$$H_{k+1} = H_k + \frac{r_k r_k^T}{r_k^T e_k} - \frac{H_k e_k [H_k e_k]^T}{e_k^T H_k e_k}; \tag{1.3}$$

2) метод Бройдена-Флетчера-Шенно (БФШ)

$$H_{k+1} = H_k + \frac{\rho_k r_k r_k^T - r_k e_k^T H_k - H_k e_k r_k^T}{(e_k r_k)}. \tag{1.4}$$

У цих формулах та надалі будемо позначати

$$r_k = x_{k+1} - x_k, e_k = f'(x_{k+1}) - f'(x_k).$$

Параметр ξ_k обчислювався наступним чином [2]:

$$\xi_k = \frac{(f'(x_k) - f'(x_{k-1}), f'(x_{k-1}))}{(s_{k-1}, f'(x_{k-1}))}. \tag{1.5}$$

Опишемо алгоритм обчислення параметру β_k. Нехай $\beta_0 = 1$. На k-му кроці спочатку покладаємо $\beta_k = \beta_{k-1}$. Якщо при цьому $f(x_{k+1}) < f(x_k)$, то або переходимо до наступної ітерації, або покладаємо $\beta_k = 2\beta_{k-1}$. Якщо значення $f(x)$ менше за попереднє, то процес подвоєння продовжуємо до тих пір, поки спадання не припиниться. Якщо ж $f(x_{k+1}) \geq f(x_k)$, то покладаємо $\beta_k = 0.5\beta_{k-1}$. Якщо $f(x_k + 0.5\beta_{k-1}s_k) < f(x_k)$, то переходимо до наступної ітерації. Якщо ж $f(x_k + 0.5\beta_{k-1}s_k) \geq f(x_k)$, то покладаємо $\beta_k = 0.25\beta_{k-1}$ і т.д.

2. Основні властивості алгоритму

Були встановлені такі властивості алгоритму (1.2)–(1.3).

1. Матриця H_k — симетрична.

Цей факт встановлюється по індукції. Представимо матрицю H_1 у вигляді $H_1 = H_0 + \Delta H_0$ згідно з формулою (1.3). Матриця H_0 – симетрична, обидві матриці, що утворюють ΔH_0, також симетричні (друга з них – згідно симетрії H_0), тому H_1 буде симетричною матрицею. Аналогічні міркування справедливі при довільному k=2,...,n.

1. Матриця H_k — строго додатньо означена.

Доведення проводиться методом індукції. Матриця H_0 — строго додатньо означена. Нехай H_k — строго додатньо означена матриця. Тоді при довільному $x \in R^n$

$$(H_{k+1}x, x) = (H_k x, x) + \frac{(r_k, x)^2}{(r_k, e_k)} - \frac{(H_k e_k, x)^2}{(H_k e_k, e_k)} =$$
$$= \frac{(H_k x, x)(H_k e_k, e_k) - (H_k e_k, x)^2}{(H_k e_k, e_k)} + \frac{(r_k, x)^2}{(r_k, e_k)}.$$

Згідно припущення щодо матриці H_k існує квадратний корінь $H_k^{1/2}$. Отже, з урахуванням симетричності матриці H_k

$$(H_k x, x) = (H_k^{1/2} H_k^{1/2} x, x) = (H_k^{1/2} x, H_k^{1/2} x) = (y, y).$$

Аналогічно

$$(H_k e_k, e_k) = (H_k^{1/2} e_k, H_k^{1/2} e_k) = (z, z),$$
$$(H_k e_k, x) = (H_k^{1/2} e_k, H_k^{1/2} x) = (z, y).$$

Враховуючи ці співвідношення та нерівність Коші–Буняковського, встановлюємо справедливість нерівності

$$(H_k x, x)(H_k e_k, e_k) - (H_k e_k, x)^2 = (y, y)(z, z) - (z, y)^2 \geq 0,$$

причому рівність має місце лише у випадку $z = y$, тобто, враховуючи невиродженість H_k, лише за умови $x = e_k$. Але при цьому $(r_k, x) = (r_k, e_k) = = (r_k, Ar_k) > 0$. Таким чином, при будь-якому $x \neq 0$:

$$(H_{k+1}x, x) = \frac{(y,y)(z,z) - (z,y)^2}{(H_k e_k, e_k)} + \frac{(r_k, x)^2}{(r_k, e_k)} > 0,$$

що доводить справедливість індуктивних міркувань.

Означення 1. *Вектори* $s_0, ..., s_{n-1}$ *називаються* A *-ортогональними (спряженими), якщо вони задовольняють умовам*

$$(s_i, As_j) = 0, i \neq j. \tag{2.1}$$

3. Напрямки спуску, визначені за схемою (1.2), є спряженими.

Розглянемо квадратичну функцію вигляду

$$f(x) = \frac{1}{2}(Ax, x) + (b, x) + c, \tag{2.2}$$

де A — симетрична строго додатньо означена матриця розміру nxn з постійними елементами: $(Ax, x) > 0$ при будь-якому $x \neq 0$, b — вектор, c — скаляр; градієнт дорівнює $f'(x) = Ax + b$.

З'ясуємо, чи будуть напрямки, визначені за схемою (1.2), спряженими для квадратичної функції вигляду (2.1). Для цього розглянемо

$$(s_k, As_j) = (-H_k^T f'(x_k) + \xi_k s_{k-1}, As_j) = (-H_k^T f'(x_k), As_j) + \xi_k (s_{k-1}, As_j).$$

Для того, щоб цей вираз дорівнював нулеві (тобто, щоб виконувалася умова спряженості), повинні виконуватися дві рівності

$$(f'(x_k), H_k As_j) = 0 \text{ та } (s_{k-1}, As_j) = 0 \text{ для будь-якого k.}$$

Друга рівність еквівалентна умові $(f'(x_{k-1}), H_{k-1}As_j) = 0$. Матриця H_k побудована так, щоб виконувалася умова $(f'(x_k), H_k As_j) = 0, 0 \leq j \leq k-1$. Так як ця умова повинна мати місце при будь-якому k, отримаємо, що рівність $(f'(x_{k-1}), H_{k-1}As_j) = 0$ також виконується. Отже, напрямки спуску, побудовані за алгоритмом (1.2), будуть спряженими для задачі мінімізації квадратичної функції.

3. Дослідження збіжності

Спочатку сформулюємо деякі допоміжні означення та теореми, які будуть використані при доведенні збіжності алгоритму.

Означення 2. *Функція* $f(x)$ *, визначена на опуклій множині* X*, називається опуклою, якщо для довільних* $x, y \in X$ *та всіх* $\alpha \in [0, 1]$ *виконується нерівність*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y).$$

Якщо для будь-якого $\alpha \in [0, 1]$ нерівність строга, то функцію називають строго опуклою.

Означення 3. *Функція* $f(x)$, *визначена на деякій множині* X, *називається сильно опуклою, якщо існує константа* $\rho > 0$ *така, що для будь-яких* $x, y \in X$ *таких, що* $[x, y] \subset X$, *і для будь-якого* $\alpha \in [0, 1]$ *буде виконуватися нерівність*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \alpha(1-\alpha)\rho\|x-y\|^2.$$

Величину ρ називають параметром сильної опуклості.

Означення 4. *Функція* $f(x)$, *означена та диференційована на деякій множині* X, *належить класу* $C^{1,1}(X)$, *якщо існує така константа* $L > 0$, *що для будь-яких* $x, y \in X$ *таких, що* $[x, y] \subset X$, *виконується нерівність*

$$\|f'(x) - f'(y)\| \leq L\|x - y\|.$$

Будемо далі припускати, що цільова функція $f(x)$ та допустима множина X опуклі.

Позначимо через α_k величину косинусу кута між напрямком антиградієнта $-f'(x_k)$ (тобто напрямком найшвидшого спадання функції $f(x)$ в точці x_k) і напрямком спуску s_k з цієї ж точки:

$$\alpha_k = \frac{(-f'(x_k), s_k)}{\|f'(x_k)\|\|s_k\|}. \tag{3.1}$$

Умова $\alpha_k > 0$ є необхідною умовою вибору напрямку спуску.

Множину $X^* = \{x^* : f'(x^*) = 0\}$ будемо називати множиною стаціонарних точок. Також визначимо множини $X_0 = \{x : f(x) \leq f(x_0)\}$ і $X_0^* = X^* \cap X_0$.

Для деякого числа $\varepsilon > 0$ визначимо множину

$$U_\varepsilon = \{x : \rho(x, X_0^*) < \varepsilon\}, \text{ де } \rho(x, X_0^*) = \inf_{x^* \in X_0^*} \|x - x^*\|^2.$$

Умова, що гарантує збіжність релаксаційної послідовності, полягає в наступному: для будь-якого $\varepsilon > 0$ знайдеться таке $\delta = \delta(\varepsilon) > 0$, що для всіх $x_k \in$ $\in X_0^* \backslash U_\varepsilon$ буде

$$\|f'(x_k)\| \geq \delta. \tag{3.2}$$

Припущення. Відносно функції $f(x)$ припустимо, що:

1) $f(x) \in C^{1,1}(R^n)$;

2) $X_0^* \neq \emptyset$;

3) $f(x)$ обмежена знизу на множині X_0.

Відносно послідовності $\{x_k\}$ припустимо, що:

4) x_0 – будь-яка точка;

5) $x_{k+1} = x_k + \beta_k s_k, k = 0, 1, ...$;

6) $\alpha_k \geq \alpha > 0, k = 0, 1, ...$;

7) виконується умова

$f(x_k + \beta_k s_k) \leq (1 - \lambda_k)f(x_k) + \lambda_k \omega_k$ при $0 < \lambda \leq \lambda_k \leq 1$.

Теорема 1. *[1] Якщо виконуються припущення 1) – 7) та умова (3.2), то*

$$\lim_{k\to\infty} \rho(x_k, X_0^*) = 0. \tag{3.3}$$

Перейдемо до оцінювання швидкості збіжності методів спуску для розв'язку задач безумовної мінімізації опуклих та сильно опуклих функцій. Ці оцінки для опуклих функцій вдається побудувати при доволі жорсткій умові обмеженості множини X_0.

Основне співвідношення, яке гарантує спадання функції $f(x)$ і якому повинна задовольняти величина кроку β_k у напрямку спуску [1]:

$$f(x_k + \beta_k s_k) \le (1 - \lambda_k) f(x_k) + \lambda_k \omega_k, \tag{3.4}$$

де $\omega_k = \inf\limits_{\beta \ge 0} f(x_k + \beta s_k), \lambda_k \le \frac{f(x_k) - f(x_{k+1})}{f(x_k) - \omega_k}$.

Теорема 2. *[1] Нехай:*
1) опукла функція $f(x)$ належить класу $C^{1,1}(R^n)$;
2) $diam X_0 = \eta < \infty$;
3) послідовність $\{x_k\}$ будується за формулою $x_{k+1} = x_k + \beta_k s_k$;
4) $\alpha_k > 0, k = 0, 1, ...$
Якщо виконується умова (3.4), то

$$f(x_m) - f(x^*) \le \mu_0 \left[1 + C\mu_0 \sum_{k=0}^{m-1} \lambda_k \alpha_k^2 \right]^{-1}, m = 1, 2, ... \tag{3.5}$$

для будь-якого $0 < C \le \frac{1}{2L\eta^2}$.

Теорема 3. *[1] Нехай:*
1) сильно опукла функція $f(x)$ належить класу $C^{1,1}(R^n)$;
2) послідовність $\{x_k\}$ будується за формулою $x_{k+1} = x_k + \beta_k s_k$;
3) $\alpha_k > 0, k = 0, 1, ...$
Якщо виконується умова (3.4), то

$$f(x_m) - f(x^*) \le \mu_0 exp\left\{ -\frac{\rho}{2L} \sum_{k=0}^{m-1} \lambda_k \alpha_k^2 \right\}, m = 1, 2, ...,$$

$$||x_m - x^*|| \le \frac{2}{\rho} \mu_0 exp\left\{ -\frac{\rho}{2L} \sum_{k=0}^{m-1} \lambda_k \alpha_k^2 \right\}, m = 1, 2, ...$$

Враховуючи співвідношення (3.5), ясно, що найбільш точні оцінки виникають при α_k і λ_k, близьких до одиниці. Однак це потребує високої точності обчислення градієнта функції $f(x)$ в точках x_k, а також високої точності одновимірної мінімізації, що часто знижує ефективність обраного процесу мінімізації.

Припустимо тепер, що при виборі напрямку спуску виконується умова $\alpha_k \ge \alpha > 0$, а точність обчислення β_k така, що $0 < \lambda \le \lambda_k < 1$. У цьому випадку з співвідношення (3.5) випливає оцінка вигляду

$$f(x_m) - f(x^*) < C\frac{1}{m}, m = 1, 2, ...,$$

а з теореми 3.3 — оцінки вигляду

$$f(x_m) - f(x^*) \leq \mu_0 exp\{-Cm\}, m = 1, 2, ...,$$

$$\|x_m - x^*\| \leq \frac{2}{\rho}\mu_0 exp\{-Cm\}, m = 1, 2, ...,$$

де у якості C виступають константи, що не залежать від номеру m та величини μ_0.

Перейдемо до розгляду питання про збіжність алгоритму (1.2)–(1.3). Припустимо, що крок β_k визначається співвідношенням

$$\beta_k = \arg\min\{f(x_k + \beta s_k) : \beta \geq 0\}, k = 0, 1, ... \tag{3.6}$$

Умова (3.6) визначає наступні дві особливості послідовності $\{x_k\}$.

Лема 1. *Для диференційованої функції $f(x)$ послідовність $\{x_k\}$, побудована за схемою (1.2), (1.3), (3.6), є такою, що виконуються наступні співвідношення:*

$$(f'(x_{k+1}), s_k) = 0, k = 0, 1, ..., \tag{3.7}$$

$$(f'(x_k), s_k) = -(H_k f'(x_k), f'(x_k)), k = 0, 1, ... \tag{3.8}$$

Доведення. З умови (3.6) випливає, що при $\beta_k > 0$ буде

$$\frac{d}{d\beta}f(x_k + \beta s_k)|_{\beta=\beta_k} = 0,$$

а при $\beta_k = 0$ буде

$$\frac{d}{d\beta}f(x_k + \beta s_k)|_{\beta=0} \geq 0.$$

Якщо $\beta_k > 0$, то

$$0 = \frac{d}{d\beta}f(x_k + \beta s_k)|_{\beta=\beta_k} - (f'(x_k + \beta_k s_k), s_k) = -(f'(x_{k+1}), s_k).$$

Доведення того, що співвідношення (3.7) справедливе і для $\beta_k = 0$, проведемо по індукції. Якщо $\beta_0 = 0$, то з $x_1 = x_0$ та $s_0 = -f'(x_0)$ отримуємо

$$0 \leq \frac{d}{d\beta}f(x_0 + \beta s_0)|_{\beta=0} - (f'(x_1), s_0) = \|f'(x_0)\|^2,$$

звідки випливає рівність

$$(f'(x_1), s_0) = 0.$$

Нехай справедливе співвідношення $(f'(x_k), s_{k-1}) = 0$. Доведемо, що

$$(f'(x_{k+1}), s_k) = 0$$

при $\beta_k = 0$. Так як $x_{k+1} = x_k$, то з (1.2) отримуємо

$$0 \le \frac{d}{d\beta} f(x_k + \beta s_k)|_{\beta=0} - (f'(x_{k+1}), s_k) = -(f'(x_k), s_k) =$$

$$= -(f'(x_k), -H_k f'(x_k) + \xi_k s_{k-1}) = (H_k f'(x_k), f'(x_k)).$$

Так як матриця H_k додатньо означена, то $(H_k f'(x_k), f'(x_k)) \ge 0$, а звідси

$$f'(x_{k+1}), s_k) = 0.$$

Рівність (3.8) є очевидним наслідком рівностей (1.2) та (3.7). Лему доведено.

Припустимо, що норма матриць H_k рівномірно обмежена зверху: $||H_k|| \le \le \zeta (k = 0, 1, ...)$, а всі власні числа цих матриць обмежені знизу числом $\nu > 0$. Тоді для довільних x буде

$$(H_k x, x) \ge \nu_k (x, x) \ge \nu ||x||^2, k = 0, 1, ...,$$

де ν_k — найменше власне число матриці H_k.

Також припустимо, що для деякого $C > 0$

$$|\xi_k| \le C \frac{||f'(x_k)||}{||s_{k-1}||}, k = 1, 2, ...$$

Враховуючи ці припущення та лему 1, розглянемо кут між напрямком спуску та антиградієнтом, який ми позначили через α_k .

$$\alpha_k = \frac{(-f'(x_k), s_k)}{||f'(x_k)|| ||s_k||} = \frac{(-f'(x_k), -H_k f'(x_k) + \xi_k s_{k-1})}{||f'(x_k)|| || - H_k f'(x_k) + \xi_k s_{k-1}||} \ge$$

$$\ge \frac{(f'(x_k), H_k f'(x_k)) - \xi_k (f'(x_k), s_{k-1})}{||f'(x_k)|| (|| - H_k f'(x_k)|| + ||\xi_k s_{k-1}||)}.$$

З леми 1 випливає, що $(f'(x_k), s_{k-1}) = 0$. Тоді, враховуючи зроблені припущення, отримаємо

$$\alpha_k \ge \frac{\nu ||f'(x_k)||^2}{||H_k|| ||f'(x_k)||^2 + |\xi_k| ||f'(x_k)|| ||s_{k-1}||} \ge \frac{\nu ||f'(x_k)||^2}{\zeta ||f'(x_k)||^2 + ||f'(x_k)||^2} = \frac{\nu}{\zeta + 1} > 0.$$

Отже, $\alpha_k > 0$. Тоді для алгоритму (1.2)–(1.3) виконуються всі умови теореми 1, з якої випливає збіжність методу. В цей же час теореми 2 та 3 дають оцінки швидкості збіжності для алгоритму (1.2)–(1.3).

Більш точні оцінки швидкості збіжності можна отримати для сильно опуклої двічі диференційованої функції, для якої виконуються умови

$$\forall x, y \in R^n : m||y||^2 \le (f''(x)y, y) \le M||y||^2, m > 0. \qquad (3.9)$$

В якості H_0 обирається симетрична строго додатньо означена матриця

$$\forall x, y \in R^n : m_0||y||^2 \leq (H_0 y, y) \leq M_0||y||^2, m_0 > 0. \tag{3.10}$$

Будемо розглядати процес (1.2)–(1.3), який здійснюється або з відновленням матриці H_k через скінчену кількість кроків, або без відновлення матриці.

Зауваження 1. *Якщо здійснюється процес з відновленням матриці H_k через скінчену кількість кроків, то для будь-якого з методів спряжених напрямків буде виконуватися умова*

$$\lim_{k\to\infty} ||f'(x_k)|| = 0. \tag{3.11}$$

Цей факт випливає з того, що кожний перший крок процесу після відновлення є кроком градієнтного спуску, для якого виконуються умови збіжності градієнтних методів, а на наступних кроках між відновленнями проводиться спуск до мінімуму функції в напрямку руху. Виконання умови (3.11) для строго опуклої функції означає, що будь-який метод спряжених напрямків, який здійснюється з відновленням матриці через скінчену кількість кроків, збігається до розв'язку x^*. Тому для процесів з відновленням для виявлення їх ефективності важливо отримати оцінку швидкості збіжності. Для процесів без відновлення матриці H_k сам факт збіжності алгоритму потребує обґрунтування, і так само важливо отримати оцінки для швидкості збіжності.

Теорема 4. *Нехай для мінімізації функції $f(x)$, яка задовольняє умовам (3.9), використовується процес (1.2), в якому побудова матриці H_k здійснюється за методом (1.3), причому через n кроків здійснюється відновлення H_k. Тоді, якщо значення β_k визначається з умови мінімуму функції у напрямку s_k, то послідовність $\{x_k\}$ незалежно від вибору початкової точки x_0 збігається до розв'язку зі зверхлінійною швидкістю.*

Загальна схема доведення. Використовується метод від супротивного. Припускається, що твердження теореми невірне, тобто для описаного ітераційного процесу при довільному виконується умова

$$||x_{k+1} - x^*|| \geq \lambda ||x_k - x^*||, \tag{3.12}$$

де $\lambda > 0$ — константа.

Для функції, яка задовольняє умовам (3.9), справедлива нерівність

$$||f'(x)|| = ||f'(x) - f'(x^*)|| \leq M||x - x^*||, \tag{3.13}$$

з якої випливає, що умова (3.12) еквівалентна наступній:

$$||f'(x_{k+1}|| \geq \delta ||f'(x_k||, \tag{3.14}$$

де $\delta > 0$ — константа. Далі, враховуючи умову (3.14), встановлюється справедливість наступних оцінок:

$$C||f'(x_k)|| \leq ||r_k|| \leq N||f'(x_k)||, \tag{3.15}$$

де C, N – константи, що не залежать від k, $C > 0$, та

$$(e_{\xi n+i}, r_{\xi n+j}) = o(\|e_{\xi n+i}\|\|r_{\xi n+j}\|), i \neq j, 0 \leq i, j \leq n-1. \qquad (3.16)$$

Далі можна показати, що при виконанні оцінок (3.15), (3.16) послідовність (1.2) збігається до розв'язку зі зверхлінійною швидкістю, що суперечить вихідному припущенню (3.12). Використовуючи цей факт, встановлюється справедливість теореми.

4. Практичні результати

Для перевірки практичних результатів для нового методу був проведений чисельний експеримент, у якому результати роботи запропонованого алгоритму порівнювалися з тими результатами, що були отримані для тих самих умов за допомогою методу спряжених градієнтів. Результати експерименту вказують на те, що запропонований метод є більш ефективним за такими показниками, як точність отриманого розв'язку та кількість проведених ітерацій.

На практиці розглядалися тестові функції з числом змінних $n = 2, 3, 4, 5, 7, 10$. Мінімізація проводилася при 10 різних початкових точках для кожної тестової функції. Для більш наглядного представлення процесу мінімізації були побудовані графіки відхилення i-го наближення від точного розв'язку в залежності від номеру ітерації.

Наведемо приклад застосування розглянутого методу мінімізації для тестової функції трьох змінних [5] при 8 початкових точках. Відновлення методів здійснювалося через кожні 4 ітерації. Пунктирною лінією на графіках позначений процес мінімізації за алгоритмом (1.2) з перерахунком матриць методом ДФП (1.3) (він позначений як М-ДФП), суцільною – метод спряжених градієнтів (СГ). Додатково застосовувався перерахунок матриць методом БФШ (1.4).

$$f(x) = 100\left[x_3 - \left(\frac{x_1 + x_2}{2}\right)^2\right]^2 + (1 - x_1)^2 + (1 - x_2)^2, n = 3.$$

Точний розв'язок — $x^* = (1; 1; 1)^T$, значення функції в цій точці $f(x^*) = 0$.

Розглядалися початкові точки $(-1.2; 2; 0)^T$, $(0; 0; 0)^T$, $(2; 2; 2)^T$, $(0; 0; 0.5)^T$, $(-0.5; 1.5; 0.5)^T$, $(-0.5; -0.5; -0.5)^T$, $(0; 1.2; -2)^T$, $(-10; -10; 10)^T$. Результати обчислень показали, що для перших семи вказаних точок метод (1.2)–(1.3) дає більш точний розв'язок (тобто значення цільової функції ближче до оптимального) при значно меншій кількості проведених ітерацій. При цьому на початкових ітераціях процесу суттєвої різниці у роботі алгоритмів, як правило, не простежується. Наведемо графік відхилення i-го наближення цільової функції від точного розв'язку для початкової точки $x_0 = (-1.2; 2; 0)^T$.

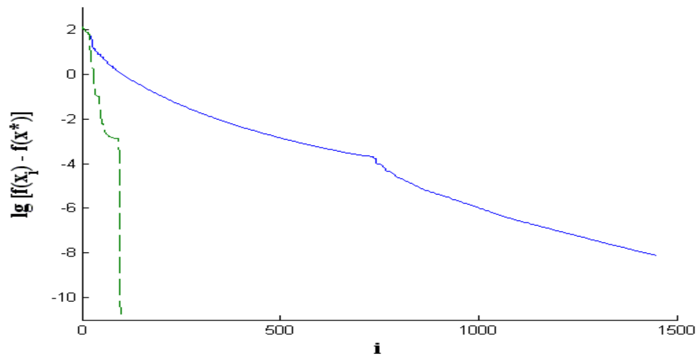

Рис. 1. $x_0 = (-1.2; 2; 0)^T$

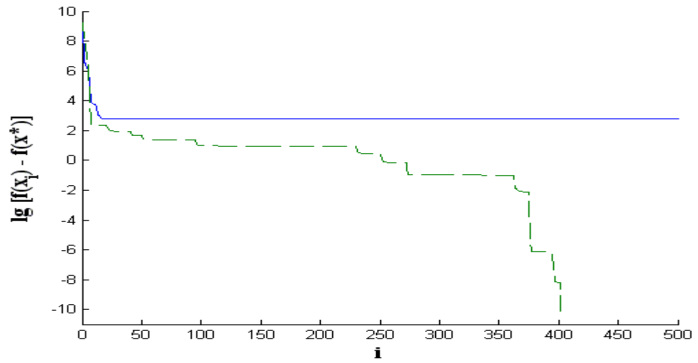

Рис. 2. $x_0 = (-10; -10; 10)^T$

На графіку рис. 1 видно, що метод СГ збігається до розв'язку повільніше (лінія більш полога), і при цьому точність отриманого розвязку, про яку можна судити за ординатою кінцевої точки, є гіршою порівняно з методом М-ДФП.

При початковій точці $x_0 = (-10; -10; 10)^T$ відбулося зациклювання методу СГ у неоптимальній точці $(3.0498; 3.0498; 9.3056)^T$, у той час як методи М-ДФП та М-БФШ прийшли до розв'язку за 403 та 203 ітерації відповідно. Цей процес також проілюструємо графічно.

На графіку рис. 2 бачимо, що з деякого моменту лінія методу СГ є паралельною до осі абсцис, тобто відбулося зациклювання алгоритму.

Слід також зазначити, що варіант перерахунку матриць (1.4) дає покращення результатів за кількістю ітерацій та значенням функції у 3 випадках з 8, а ще у трьох випадках він покращує тільки кінцеве значення функції.

Висновки. У статті був запропонований новий двокроковий метод мінімізації функцій багатьох змінних при відсутності обмежень. Розглянуті основні властивості цього алгоритму, доведені теореми збіжності. Приведені результати застосування алгоритму до задачі мінімізації функції трьох змінних.

1. **Карманов В. Г.** Математическое программирование [текст] / В. Г. Карманов. – Москва: ФИЗМАТЛИТ, 2004. – 264 с.
2. **Ларичев О. И.** Методы поиска локального экстремума овражных функций [текст] / О. И. Ларичев, Г. Г. Горвиц. – Москва: Наука, 1989. – 95 с.
3. **Поляк Б. Т.** Введение в оптимизацию [текст] / Б. Т. Поляк. – Москва: Наука, 1983. – 384 с.
4. **Пшеничный Б. Н.** Численные методы в экстремальных задачах [текст] / Б. Н. Пшеничный, Ю. М. Данилин. – Москва: Наука, 1975. – 320 с.
5. **Химмельблау Д.** Прикладное нелинейное программирование [текст] / Д. Химмельблау. – Москва: Мир, 1975. – 536 с.