

Область диференційовності асимптотичної швидкості збіжності методу найшвидшого спуску

П. Ф. Жук

Інститут математики НАН України, Київ; petro_zhuk@ukr.net

The subject of this work is to study the asymptotic convergence rate of the steepest descent method for minimizing the quadratic functional in a finite-dimensional space. Its main result is determination of domain of differentiability of this rate. It is proven that the convergence rate is continuously differentiated nearly everywhere (in the Lebesgue's measure sense).

Исследовано асимптотическую скорость сходимости метода наискорейшего спуска при минимизации квадратичного функционала в конечномерном пространстве. Основным результатом является определение области дифференцируемости этой скорости. В частности, доказано, что асимптотическая скорость сходимости метода наискорейшего спуска почти везде (по мере Лебега) непрерывно дифференцируема.

1. Вступ

Швидкість збіжності відомих ітераційних методів варіаційного типу істотно залежить від вибору початкового наближення (див., наприклад, [1, 2]). Ці залежності мають теоретичну і практичну цінність, оскільки дають змогу детальніше оцінити можливості того чи іншого методу. Для методу найшвидшого спуску таку залежність можна дослідити на базі поняття асимптотичної швидкості збіжності (це поняття природним чином впливає з асимптотичних властивостей методу, встановлених у [3]–[5]). Так, у праці [6] була знайдена істотна (за мірою Лебега) область значень асимптотичної швидкості збіжності

методу як функції від початкового наближення. У [7, 8] охарактеризована область неперервності і побудована множина точок розриву цієї асимптотичної швидкості збіжності. У даній праці, яка є продовженням [6]– [8], визначена область диференційовності асимптотичної швидкості збіжності методу найшвидшого спуску як функції від початкового наближення. Показано, зокрема, що асимптотична швидкість майже всюди (за мірою Лебега) неперервно диференційовна.

2. Постановка задачі

Будемо використовувати позначення із [6]: \bar{S} – стандартний симплекс в n -вимірному просторі R^n ; S^* – множина точок симплекса \bar{S} , в яких перша і остання координати ненульові; $T: \bar{S} \rightarrow \bar{S}$ – відображення, що породжується оператором переходу методу найшвидшого спуску; $V: \bar{S} \rightarrow R$ – асимптотична швидкість збіжності методу найшвидшого спуску; V_{\min} і V_{\max} – істотний мінімум і істотний максимум функції V ; $\Delta = [V_{\min}, V_{\max}]$; $\bar{\Delta} = [V_{\min}, V_{\max}]$ – істотна область значень функції V на симплексі \bar{S} ; $M = \{x \in S^* \mid V(x) \in \Delta\}$; $\bar{M} = \{x \in S^* \mid V(x) \in \bar{\Delta}\}$.

Задача: охарактеризувати точки симплекса \bar{S} , в яких функція V диференційовна.

3. Область диференційовності функції V

Основним результатом даної статті є наступна теорема:

Теорема 3.1. *Нехай D і D^* – це множини точок симплекса \bar{S} , в яких функція V відповідно диференційовна і неперервно диференційовна. Якщо $n \geq 2$, то $D^* = D = M$.*

Для доведення теореми сформулюємо декілька допоміжних тверджень.

Лема 3.1. *Якщо $n \geq 2$, то $M \subseteq D^*$.*

Доведення. Нехай точка $x \in M$. Із означення функції V (див. [6, формула (43)]) і співвідношень (46) з [6] випливає, що достатньо довести неперервну диференційовність у точці x функції ξ , яка визначена співвідношеннями (45) з [6]. Розглянемо послідовність функцій $\xi_0(y) = y_1$, $\xi_1(y) = \xi_0(T^2y)$, \dots , $\xi_k(y) = \xi_0(T^{2k}y)$, \dots , де $y = (y_1, \dots, y_n) \in S^*$. Оскільки $T^{2k}(y) \rightarrow y^\infty$ при $k \rightarrow \infty$, то і

$\xi_k(y) \rightarrow \xi(y)$ при $k \rightarrow \infty$. Крім того, функції ξ_k , $k = 0, 1, \dots$, – неперервно диференційовні на множині S^* та

$$\xi'_k(y) = \xi'_0 J(y, T^{2k}), \quad \xi'_0 = (1, 0, \dots, 0). \quad (1)$$

Доведемо, що в деякому околі $O(x)$ точки x послідовність функцій ξ'_k , $k = 0, 1, \dots$, рівномірно збігається; звідси, очевидно, випливає твердження леми. Оскільки $\xi_{k+m}(y) = \xi_0(T^{2(k+m)}y) = \xi_k(T^{2m}y)$, $\xi'_{k+m}(y) = \xi'_k(T^{2m}y)J(y, T^{2m})$, $T^{2m}y \rightarrow x^\infty$ при $m \rightarrow \infty$, то достатньо показати, що послідовність функцій ξ'_k , $k = 0, 1, \dots$, рівномірно збігається в деякому, скільки завгодно малому, околі $O(x^\infty)$ точки x^∞ .

Виберемо окіл $O(x^\infty)$ так, щоб:

- 1) його замикання $\bar{O}(x^\infty) \subseteq M$;
- 2) $\forall y \in O(x^\infty)$ і $k = 0, 1, \dots$ мало місце розвинення

$$J(T^{2k}y, T^2) = J(y^\infty, T^2) + \varepsilon_k(y), \quad \|\varepsilon_k(y)\| = \max_{i,j} |\varepsilon_{ij,k}(y)| \leq \alpha q^k, \quad (2)$$

де $y^\infty = T^\infty y$, α, q ($\alpha > 0$, $0 < q < 1$) – деякі числа, що не залежать від y і k .

Покажемо, що окіл $O(x^\infty)$ з такими властивостями дійсно існує.

Оскільки за умовою $x \in M$, то мають місце оцінки (55) з [6] і існує окіл $O_1(x^\infty)$ точки x^∞ і числа $m > 0$, $0 < q < 1$ такі, що $\forall y \in O_1(x^\infty)$ виконуються оцінки (56) з [6]. Нехай $O_2(x^\infty)$ – це окіл точки x^∞ , що задовольняє лему 2 з [6]. Оскільки відображення T^2 є нескінченно диференційовним на множині S_ω , то, зменшуючи за необхідності окіл $O_2(x^\infty)$, отримуємо, що $\forall y \in O_2(x^\infty)$ і $k = 0, 1, \dots$ має місце розвинення

$$J(T^{2k}y, T^2) = J(y^\infty, T^2) + \varepsilon_k(y), \quad \|\varepsilon_k(y)\| \leq \alpha_1 \|T^2y - y^\infty\|, \quad (3)$$

де число α_1 не залежить від y і k . З оцінок (56) [6] випливає, що $\forall y \in O_2(x^\infty)$ справджується

$$\|T^{2k}y - y^\infty\| \leq \frac{mq^k}{1-q}, \quad k = 0, 1, \dots$$

Тому з (3) отримуємо розвинення (2) з $\alpha = \alpha_1 m / (1 - q)$ і околom $O(x^\infty) = O_2(x^\infty)$. Таким чином, окіл $O(x^\infty)$, що задовольняє властивість 2, існує.

Далі, оскільки $x \in M$, то і $x^\infty \in M$. Множина M є відкритою в топології сиплекса \bar{S} (див. [6, лема 3]), тому, зменшуючи за необхідності окіл $O(x^\infty)$, можна отримати окіл точки x^∞ з властивостями 1, 2.

Доведемо, що послідовність функцій ξ'_k , $k = 0, 1, \dots$, рівномірно збігається в околі $O(x^\infty)$, що має властивості 1 і 2.

Використовуючи формулу (1), достатньо довести, що в околі $O(x^\infty)$ рівномірно збігається послідовність матриць $J(y, T^{2k})$, або, що еквівалентно, наступне твердження: для будь-якого вектора $w \in R^n$ послідовність вектор-функцій

$$w^{(0)} = w, \quad w^{(k)}(y) = J(y, T^{2k})w, \quad k = 1, 2, \dots, \quad (4)$$

рівномірно збігається в околі $O(x^\infty)$. Оскільки $J(y, T^{2k+2}) = J(T^{2k}y, T^2)J(y, T^{2k})$, то з (4) випливає, що

$$w^{(k+1)}(y) = J(T^{2k}y, T^2)w^{(k)}(y), \quad k = 0, 1, \dots \quad (5)$$

Використаємо розвинення матриці $J(T^{2k}y, T^2)$ за формулою (2). Матрицю $J(y^\infty, T^2)$ обчислимо безпосередньо, виходячи з означення оператора T . Оскільки $y^\infty = (y_1^\infty, 0, \dots, 0)$, то

$$J(y^\infty, T) = \begin{bmatrix} \gamma_{11}(y) & \gamma_{12}(y) & \gamma_{13}(y) & \dots & \gamma_{1,n-1}(y) & \gamma_{1n}(y) \\ 0 & \gamma_{22}(y) & 0 & \dots & 0 & 0 \\ 0 & 0 & \gamma_{33}(y) & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & \gamma_{n-1,n-1}(y) & 0 \\ 0 & 0 & 0 & \dots & 0 & \gamma_{nn}(y) \end{bmatrix}, \quad (6)$$

де для $i = \overline{2, n}$

$$\gamma_{11}(y) = -1, \quad \gamma_{1i}(y) = -\frac{(\lambda_{n+1} - \lambda_i)^2}{(\lambda_{n+1} - \lambda_1) y_1^\infty}, \quad \gamma_{ii}(y) = \frac{(\mu(y^\infty) - \lambda_i)^2}{\beta(y^\infty)}.$$

Матриця Якобі $J(Ty^\infty, T)$ також має вигляд (6) з

$$\gamma_{11}(y) = -1, \quad \gamma_{1i}(y) = -\frac{(\lambda_{n+1} - \lambda_i)^2}{(\lambda_{n+1} - \lambda_1)(1 - y_1^\infty)},$$

$$\gamma_{ii}(y) = \frac{(\mu(Ty^\infty) - \lambda_i)^2}{\beta(y^\infty)}, \quad i = \overline{2, n}.$$

Оскільки $\forall y \in M$ значення функції $V(y) \in \Delta$, то (див. [6, формула (54)]) $\forall y \in \bar{O}(x^\infty)$ має місце нерівність $\gamma_{ii}(y) < 1, i = 2, 3, \dots, n$. Отже,

$$0 \leq \gamma = \max_{i=2,3,\dots,n} \sup_{y \in O(x^\infty)} \gamma_{ii}(y) < 1. \quad (10)$$

Покладемо, що

$$\zeta = \frac{c}{1-\gamma}, \quad \varphi(w^{(k)}(y)) = |w_{1k}(y)| + \zeta \sum_{i=2}^n |w_{ik}(y)|.$$

Оскільки $c \geq 1$, то з оцінки (10) маємо, що $\zeta \geq 1$. Далі використаємо рівняння (8):

$$\begin{aligned} \varphi(w^{(k+1)}(y)) &= |w_{1,k+1}(y)| + \zeta \sum_{i=2}^n |w_{i,k+1}(y)| \leq \\ &\leq |w_{1k}(y)| + \sum_{i=2}^n (|\gamma_{1i}(y)| + \zeta \gamma_{ii}(y)) |w_{ik}(y)| + \\ &+ \sum_{j=1}^n |\varepsilon_{1j,k}(y) w_{jk}(y)| + \zeta \sum_{i=2}^n \sum_{j=1}^n |\varepsilon_{ij,k}(y) w_{jk}(y)|. \quad (11) \end{aligned}$$

Оскільки $|\gamma_{1i}(y)| + \zeta \gamma_{ii}(y) \leq c + \zeta \gamma = \zeta$ для $\forall y \in O(x^\infty)$ і $i = \overline{2, n}$, то з оцінок (9), (11) випливає, що $\forall y \in O(x^\infty), k = 0, 1, \dots$, справджується нерівність

$$\varphi(w^{(k+1)}(y)) \leq (1 + \zeta n \alpha q^k) \varphi(w^{(k)}(y)),$$

тобто послідовність функцій $\{\varphi(w^{(k)}(y)), k = 0, 1, \dots\}$ рівномірно обмежена на околі $O(x^\infty)$. Тому існує число c_1 таке, що $\forall y \in O(x^\infty)$ має місце оцінка

$$|w_{ik}(y)| \leq c_1, \quad i = \overline{1, n}, \quad k = 0, 1, \dots \quad (12)$$

Покладемо $\omega_k = \max_{i=2,3,\dots,n} \sup_{y \in O(x^\infty)} |w_{ik}(y)|$. Із співвідношень (8)–(10),

(12) випливає оцінка

$$\omega_{k+1} \leq \gamma \omega_k + c_2 q^k, \quad k = 0, 1, \dots, \quad (13)$$

де $c_2 = \alpha n c_1$. З оцінок (10), (13) випливає, що $\gamma < 1$ і

$$\sum_{k=0}^m \omega_k \leq \frac{\omega_0(1-q) + c_2}{(1-\gamma)(1-q)}, \quad m = 0, 1, \dots,$$

тобто ряд $\sum_{k=0}^{\infty} \omega_k$ збігається і $w_{ik}(y) \rightarrow 0$ при $k \rightarrow \infty$ рівномірно по $y \in O(x^\infty)$, $i = \overline{2, n}$. Але тоді з першого співвідношення (8) випливає збіжність ряду

$$\sum_{k=0}^{\infty} \sup_{y \in O(x^\infty)} |w_{1,k+1}(y) - w_{1k}(y)|,$$

отже, збіжність на множині $O(x^\infty)$ послідовності функцій $\{w_{1k}(y), k = 0, 1, \dots\}$ рівномірна. Лему доведено. \square

Розглянемо множину $M^\infty = T^\infty M$. Неважко бачити, що $M^\infty = M \cap S_{1,n+1}$ і є інтервалом. Позначимо його кінці через $x^{1,\infty}$ і $x^{2,\infty}$, припускаючи, що $\xi(x^{1,\infty}) < \xi(x^{2,\infty})$. Тоді $x^{2,\infty} = Tx^{1,\infty}$, і при $x \in \{x^{1,\infty}, x^{2,\infty}\}$ маємо $V(x) = V_{\max}$, $x \in \bar{M} \setminus M$, $\rho_{i^0,\infty}(x) = \rho_{1,\infty}(x)$.

Обчислимо $J(x, T^\infty)$ при $x \in M^\infty$. Оскільки $x = T^2x = T^\infty x = x^\infty$, то з виразу (6) отримуємо, що

$$\begin{aligned} J(x, T^\infty) &= \lim_{k \rightarrow \infty} J(x, T^{2k}) = \\ &= \prod_{k=1}^{\infty} J(x, T^2) = \begin{bmatrix} 1 & \gamma_{12}^*(x) & \cdots & \gamma_{1n}^*(x) \\ & 0 & & \end{bmatrix}, \quad (14) \\ \gamma_{1i}^*(x) &= \frac{\gamma_{1i}(x)}{1 - \gamma_{ii}(x)}, \quad i = \overline{2, n}, \end{aligned}$$

де величини $\gamma_{ij}(x)$ визначені в (7).

Розглянемо поведінку вектор-функції ξ' на відрізку $\bar{M}^\infty = [x^{1,\infty}, x^{2,\infty}]$. Зі співвідношень (1), (14) отримуємо

$$\xi'(x) = \xi'_0 J(x, T^\infty) = (1, \gamma_{12}^*(x), \dots, \gamma_{1n}^*(x)). \quad (15)$$

Виконавши обчислення за формулами (14), бачимо, що при $i \neq i^0$ функції γ_{1i}^* існують і неперервні на відрізку \bar{M}^∞ , тоді як функція $|\gamma_{1,i^0}^*|$ необмежено зростає при $x \rightarrow x^{j,\infty}$, $j = 1, 2$. Таким чином, можна припустити, що функція V не диференційовна на множині $\mathfrak{R} = \bar{M} \setminus M$. Доведемо це.

Лема 3.2. *Якщо $n \geq 2$, то $D \cap \mathfrak{R} = \emptyset$.*

Доведення. Достатньо довести, що на множині \mathfrak{R} функція ξ недиференційовна. Розглянемо кінці відрізка \bar{M}^∞ , наприклад, $x = x^{1,\infty}$.

Припустимо протилежне: функція ξ є диференційовною у точці x . Тоді, якщо точка $x + \Delta x \in \bar{S}$, то має місце розвинення

$$\xi(x + \Delta x) - \xi(x) = \xi'(x)\Delta x + \alpha(\Delta x)\Delta x, \quad (16)$$

де $\alpha(\Delta x) \rightarrow 0$ при $\Delta x \rightarrow 0$. Якщо $\Delta x = (\Delta x_1, 0, \dots, 0)$, то з означення функції ξ випливає, що $\xi(x + \Delta x) - \xi(x) = \Delta x_1$, тобто маємо рівність

$$\xi'_{x_1}(x) = 1. \quad (17)$$

Нехай тепер $\Delta x = (\Delta x_1, 0, \dots, 0)$ з $\Delta x_1 < 0$. Оскільки множина M всюди щільна в симплексі \bar{S} (див. [7, наслідок 2.1 і теорема 3.1]), то існують вектори $\Delta x^{(m)}$, $m = 1, 2, \dots$, такі, що $\Delta x^{(m)} \rightarrow 0$ при $m \rightarrow \infty$ та $x + \Delta x + \Delta x^{(m)} \in M$, $m = 1, 2, \dots$. Оскільки $\forall y \in M$ буде $\xi(y) > \xi(x)$, то $\xi(x + \Delta x + \Delta x^{(m)}) > \xi(x)$, $m = 1, 2, \dots$. Отже, зі співвідношень (16), (17) матимемо $\Delta x_1 + \xi'(x)\Delta x^{(m)} + \alpha(\Delta x + \Delta x^{(m)})(\Delta x + \Delta x^{(m)}) > 0$, $m = 1, 2, \dots$, що суперечить умові $\Delta x_1 < 0$ при достатньо малих $|\Delta x_1|$, тобто функція ξ не є диференційовною у точці $x = x^{1,\infty}$.

Доведення у загальному випадку подібне до наведеного вище. Нехай $x \in \mathfrak{R}$ і функція ξ є диференційовною в точці x . Тоді ξ є диференційовною у будь-якій точці $T^k x$, $k = 1, 2, \dots$. Отже, без обмеження загальності можна вважати, що x знаходиться у будь-якому малому околі деякого кінця відрізка \bar{M}^∞ (наприклад, кінця $x^{1,\infty}$). Оскільки $U_{i^0} \cap \mathfrak{R} = \emptyset$ (див. [7, лема 3.3]), то компонента i^0 вектора x є вивродженою. Тобто без обмеження загальності можна вважати, що $x \in S_{i_1 \dots i_m}$, де $i_1 = 1$, $i_m = n + 1$, $i^0 \notin \{i_1, \dots, i_m\}$.

З леми 1 (застосованою до симплекса $\bar{S}_{i_1 \dots i_m}$) випливає, що звуження функції ξ на $\bar{S}_{i_1 \dots i_m}$ є неперервно диференційовним у точці x . Згідно з формулою (15) для достатньо близьких до $x^{1,\infty}$ точок x має місце нерівність

$$\xi'_{x_1}(x) > 0. \quad (18)$$

Отже, існує точка $x \in \mathfrak{R}$, для якої виконуються співвідношення (16), (18). Але тоді, як і в першому випадку, маємо суперечність. Лему доведено. \square

Доведення теореми. Оскільки $D \subseteq \bar{M}$ (див. [7, теорема 3.1]), то з леми 2 маємо, що $D^* \subseteq D \subseteq M$. Але тоді з леми 1 випливає, що $D^* = D = M$. Теорему доведено. \square

Наслідок 3.1. *Асимптотична швидкість збіжності V методу найшвидшого спуску майже всюди (за мірою Лебега) неперервно диференційовна.*

Доведення. Позначимо через ϑ міру Лебега на просторі R^n . Оскільки множина U точок, що не вироджуються, є підмножиною множини M (див. [7, теорема 3.1]), то $\vartheta(\bar{S} \setminus M) \leq \vartheta(\bar{S} \setminus U)$. Множина $\bar{S} \setminus U$ складається з точок, що вироджуються, тому $\vartheta(\bar{S} \setminus U) = 0$ (див. [7, лема 2.1]). Отже, маємо $\vartheta(\bar{S} \setminus D^*) = \vartheta(\bar{S} \setminus M) = 0$, що і потрібно було довести. \square

- [1] Neymeyr K., Zhou M. Iterative minimization of the Rayleigh quotient by block steepest descent iterations // Numerical Linear Algebra with Applications. – 2014. – **21**, N 5. – P. 604–617.
- [2] Narushima Y., Wakamatsu T., Yabe H. Extended Barzilai-Borwein method for unconstrained minimization problems // Pacific Journal of Optimization. – 2010. – **6**, N 3. – P. 591–613.
- [3] Akaike H. On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method // Ann. Inst. Statist. Math. Tokyo. – 1959. – **11**. – P. 1–16.
- [4] Forsythe G.E. On the asymptotic directions of the s-dimensional optimum gradient method // Numerische Mathematik. – 1968. – **11**. – P. 57–76.
- [5] Самарский А.А., Николаев Е.С. Методы решения сеточных уравнений. – М.: Наука, 1978. – 592 с.
- [6] Жук П.Ф., Мусина А.О. Асимптотична швидкість збіжності методу найшвидшого спуску // Аналітична механіка та її застосування: Зб. пр. Ін-ту математики НАН України. – Київ: Ін-т математики НАН України, 2012. – **9**, № 1. – С. 128–135.
- [7] Жук П.Ф., Мусина А.А. Асимптотическая скорость сходимости двухслойного итерационного метода вариационного типа // Укр. мат. журн. – 2013. – **65**, № 12. – С. 1622–1635.
- [8] Жук П.Ф., Мусина А.А. Об операторе перехода метода наискорейшего спуска // Математическое моделирование. – 2014. – **26**, № 8. – С. 65–80.