

Посвящаем эту работу памяти
Федора Павловича Васильева

ОБ ИЗБЫТОЧНОСТИ НЕВЫРОЖДЕННОСТИ ГЕССИАНА ДЛЯ ГЕОМЕТРИЧЕСКОЙ СКОРОСТИ СХОДИМОСТИ МЕТОДА НЬЮТОНА ПРИ МИНИМИЗАЦИИ ВЫПУКЛЫХ ФУНКЦИЙ¹⁾

© 2024 г. Ю. Г. Евтушенко^{1,2,*}, А. А. Третьяков^{1,3,**}

¹ 119333 Москва, ул. Вавилова, 44, ФИЦ ИУ РАН, Россия,

² 141701 Долгопрудный, М.о., Институтский переулок, 9, Московский физико-технический институт
(государственный университет), Россия

³ 08-110 Siedlce, Siedlce University, Faculty of Exact and Natural Sciences, Poland

*e-mail: yuri-evtushenko@yandex.ru

**e-mail: prof.tretyakov@gmail.com

Поступила в редакцию 10.08.2023 г.

Переработанный вариант 07.11.2023 г.

Принята к публикации 07.11.2023 г.

В статье устанавливается новое свойство выпуклых функций, позволяющее добиться геометрической скорости сходимости метода Ньютона в процессе минимизации. А именно, установлено, что даже в случае вырождения гессиана в решении, ньютоновская система разрешима в окрестности точки минимума, т.е. градиент целевой функции принадлежит образу матрицы вторых производных и поэтому можно применять аналоги метода Ньютона. Библ. 10.

Ключевые слова: выпуклая функция, метод Ньютона, разрешимость, сходимость, скорость сходимости, регулярность.

DOI: 10.31857/S0044466924040045, **EDN:** ZKJRII

В задаче поиска безусловного минимума рассматриваются функции $f(\cdot)$, определенные и достаточно гладкие в окрестности $U(x^*)$ точки минимума функции n вещественных переменных. Всяду далее множество точек минимума функции f обозначается как $X^* = \text{Argmin} f$ и предполагается непустым. Необходимое условие минимума функции f в точке x^* задается равенством $f'(x^*) = 0$, при этом матрица вторых производных функции в точке минимума является положительно полуопределенной. Следует отметить, что исследованиям по методу Ньютона посвящено значительное число научных работ, среди которых укажем [1–7]. Со многими работами можно ознакомиться в обзорной статье [8]. В данной работе показывается, что несмотря на возможную вырожденность матрицы Гессе в точке x^* , в окрестности этой точки градиент целевой функции принадлежит образу ее второй производной и, следовательно, ньютоновская система относительно направления спуска разрешима в точках этой окрестности. Это топологическое свойство выпуклости и экстремальности позволяет по новому взглянуть на численные методы ньютоновского типа и обосновать скорость сходимости этих методов без обременительного предположения относительно невырожденности матрицы Гессе в точке x^* . Для задачи безусловной оптимизации получены свойства, уточняющие неравенство Лоясевича (см. [9–10]). Именно, неравенство обобщено для всего спектра производных до определенного порядка. В работе рассматриваются функции, производные которой равномерны по аргументу и в совокупности по порядку ограничены в некоторой окрестности $U(x^*)$ точки x^* , т.е. для данной окрестности существует положительная константа M такая, что значения ее производных любого порядка не превосходят по абсолютному значению эту константу. Кроме того, объектом рассмотрения в работе будут достаточно гладкие в окрестности точки минимума функции, т.е. функции, имеющие неограниченное число порядков производных. Всяду далее без ограничения общности считаем $f(x^*) = 0$ и обозначаем $f^{(0)}(x) = f(x)$. Для функции одной переменной справедлива

¹⁾ Работа выполнена при финансовой поддержке РФФИ (проект № 21-71-30005).

Лемма 1. Если x^* -точка изолированного локального минимума достаточно гладкой в $U(x^*)$ функции $f : R \rightarrow R$, производные которой равномерно по аргументу и в совокупности по порядку ограничены в некоторой окрестности $U(x^*)$, то существует четная степень $2p, p = p_f \in \mathbb{N}$, для которой справедливы неравенства

$$f^{(2p)}(x^*) > 0, \quad f^{(k)}(x^*) = 0, \quad \frac{f^{(k)}(x)}{(x - x^*)^{2p-k}} \geq C_k, \quad k = 0, 1, 2, \dots, 2p - 1,$$

при всех $x \in \dot{U}(x^*)$, где положительные константы $C_k = C_{k,f}, k = 0, 1, 2, \dots, 2p - 1$, не зависят от x .

Доказательство. Поскольку x^* — точка локального минимума функции f , то $f'(x^*) = 0$. Сначала покажем, что в условиях леммы невозможна ситуация, когда производные $f^{(k)}(x^*) = 0$ при любом порядке $k, k = 1, 2, \dots$. Действительно в этом случае для фиксированной точки $x = x^* + t \in U(x^*)$ в силу ограниченности производных в окрестности x^* из формулы Тэйлора при нулевых производных до любого фиксированного порядка k следует неравенство $|f(x^* + \theta t)| \leq \frac{M(|\theta t|)^k}{k!}$, где M — обозначенная выше верхняя грань для множества значений производных функции f в указанной окрестности и $\theta \in (0, 1)$. Отсюда и из условия изолированности локального минимума x^* значение

$$f(x^* + t) = \int_0^1 f'(x^* + \theta t) d\theta \leq M / (k + 1)!$$

При достаточно больших k это означает противоречие с возможным предположением $f(x) \neq 0, x \neq x^*$ что противоречит изолированности локального минимума x^* . Следовательно, существует конечное

$$k > 1 : f^{(k)}(x^*) \neq 0, \quad f^{(l)}(x^*) = 0, \quad l = 1, 2, \dots, k - 1.$$

При этом порядок k может быть только четным: $k = 2p, p \in \mathbb{N}$ и $f^{(2p)}(x^*) > 0$, поскольку x^* — точка локального минимума функции f . Теперь для завершения доказательства достаточно обозначить через C_0 значение $\frac{1}{(2p + 1)!} f^{(2p)}(x^*)$, тогда из формулы Тэйлора и ограниченности производной порядка $2p + 1$ в окрестности $U(x^*)$ будет вытекать первое неравенство леммы. Разложение в окрестности x^* по формуле Тэйлора производной $f^{(k)}(x)$ дает остальные неравенства леммы при любом $k = 1, 2, \dots, 2p - 1$, если через C_k обозначить $\frac{1}{(2p + 1 - k)!} f^{(2p)}(x^*)$.

Следствие 1. При выполнении условий леммы 1 функция f локально выпукла.

Действительно, если $f^{(2)}(x^*) > 0$, то в малой окрестности x^* вторая производная будет оставаться положительной, что означает локальную выпуклость $f(x)$. Если же $f^{(2)}(x^*) = 0$, то, как показано в лемме 1, $f^{(2p)}(x^*) > 0$ для некоторого $p = p_f \in \mathbb{N}, p > 1$ и при этом $f^{(k)}(x^*) = 0, k = 1, 2, \dots, 2p - 1$. Тогда $f^{(2)}(x) \geq C_2(x - x^*)^{2p-2}$ для любого $x \in U(x^*)$, положительная константа C_2 определена ранее в лемме 1. Последнее неравенство также означает локальную выпуклость f .

Для функции n переменных производную k -го порядка по направлению h в точке x будем обозначать как $f_h^{(k)}(x)$.

Следствие 2. Если достаточно гладкая функция $f : R^n \rightarrow R$, производные которой равномерно по аргументу и в совокупности по порядку ограничены в некоторой окрестности $U(x^*)$ изолированной точки минимума x^* , то для каждого $h \in R^n : \|h\| = 1$, существует четная степень $2p_{h,f}, p_{h,f} \in \mathbb{N}$, для которой справедливы неравенства

$$f^{(2p_{h,f})}(x^*) > 0, \quad f^{(k)}(x^*) = 0, \quad \frac{f^{(k)}(x^* + th)}{t^{2p_{h,f}-k}} \geq C_k,$$

где $k = 0, 1, 2, \dots, 2p_{h,f} - 1$ при всех $t \in (0, \delta_h]$, положительные константы $C_k = C_{k,h,f}, k = 0, 1, 2, \dots, 2p_{h,f} - 1$, не зависят от $t, x^* + \delta_h \in U(x^*)$, при этом сама функция локально выпукла вдоль направления h .

Здесь элемент $p = p_{h,f} \in \mathbb{N}$ определяется направлением $h, \|h\| = 1$ и не зависит от малых t , для которых $x^* + \delta_h \in U(x^*)$, но значение $\delta_h > 0$ зависит от h и в общем случае эта величина может быть бесконечно малой относительно h . Далее в лемме 2 будет показано, что для случая выпуклой функции f можно гарантировать существование верхней границы для величины $p_{h,f} \in \mathbb{N}$ и положительной нижней границы для δ_h на множестве векторов $h : \|h\| = 1$.

Из леммы 1 вытекает, что в точках достаточно малой выколотой окрестности решения корректно определен оператор Ньютона $\psi(x) = x - f^{(2)}(x)^{-1}f(x)$. Для случая произвольной размерности пространства R^n лемма 1 означает выпуклость функции f вдоль любой прямой, проходящей через точку x^* , при условии достаточной гладкости функции на пересечении прямой и окрестности $U(x^*)$. Кроме того, для любой прямой, проходящей через точку x^* вдоль вектора h , справедливо неравенство $f_h^{(2m)}(x^*) \geq C_2$ для некоторой степени $2p$, $p = p_h \in \mathbb{N}$, и некоторой константы $C_2 = C_{2,h} > 0$. Неравенство Лоясевича гарантирует выполнимость данного неравенства в случае аналитичности в окрестности $U(x^*)$ функции при некоторых p, C_2 , не зависящих от h . Выпуклость в указанной окрестности функции позволяет требовать лишь достаточную гладкость функции и получить свойство “устойчивости” неравенства Лоясевича, что расширяет область применения метода Ньютона. Имеет место

Лемма 2. Если выпуклая достаточно гладкая функция $f : R^n \rightarrow R$, производные которой равномерно по аргументу и в совокупности по порядку ограничены в некоторой окрестности $U(x^*)$ изолированной точки минимума x^* , то существует четная степень $2p$, $p = p_f \in \mathbb{N}$, константа $C = C_f > 0$ и величина $\delta = \delta_f > 0$, для которых

$$f(x^* + th) \geq Ct^{2p}$$

при всех $h : \|h\| = 1, t \in (0, \delta]$.

Доказательство. Докажем существование такого $p = p_f \in \mathbb{N}$, обладающего свойством: если $f_h^{(i)}(x^*) = 0, i = 0, 1, 2, \dots, 2p'(h) - 1, f_h^{(2p'(h))}(x^*) > 0$, то $p'(h) \leq p$. Предположим противное, тогда для некоторых последовательностей $h_k : \|h_k\| = 1, t_k : t_k \rightarrow +0$ имеет место неравенство $f(x^* + t_k h_k) < C_k t_k^{m_k}$, где $C_k > 0$ — элементы некоторой ограниченной, а $m_k \in \mathbb{N}$ — возрастающей последовательностей. Без ограничения общности можно считать, что $h_k \rightarrow h, C_k \rightarrow C, k \rightarrow \infty$. Обозначим $\text{conv}\{h_k, h_{k+1}, h_{k+2}, \dots, h_{k+n}\}$ через V_k . В случае $\dim V_k < n$ указанный набор векторов изменяется путем прибавления к $n - \dim V_k$ векторам линейно независимых приращений длины порядка $C_k t_k^{2p}, C_k \rightarrow 0, k \rightarrow \infty$, после чего можно считать $\dim V_k = n$. Пусть далее $h'_k \in \text{int} V_k, h''_k = h + \alpha_k (h'_k - h), \alpha_k \in (0, 1), \alpha_k = \inf\{\alpha > 0 : h + \alpha(h'_k - h) \in V_k\}$. Из выпуклости и непрерывности f следует $f(x^* + t_k h''_k) \leq 2C_k t_k^{m_k}$. С другой стороны, из леммы 1 следует, что для некоторого $p = p_h \in \mathbb{N}$ существует степень $k \in \mathbb{N}, k \leq p$, для которой производная $f_h^{(2k)}(x^*) \geq C_2 > 0$ для некоторого $C_2 > 0$. Поскольку $h_k \rightarrow h, k \rightarrow \infty$, то при достаточно больших номерах k будут выполняться неравенства $f(x^* + t_k h''_k) \geq C_2 t_k^{2k}$, что противоречит предположению.

Следствие 3. Из доказательства леммы 2 следует существование степени $2p, p = p_f \in \mathbb{N}$, а также констант $C_k = C_{k,f}, k = 1, 2, \dots, 2p' - 1$, для которых

$$f_h^{(2p')}(x^*) > 0, f_h^{(k)}(x^*) = 0, k = 0, 1, \dots, 2p' - 1, f_h^{(k)}(x^* + th) \geq C_k t^{2p-k}, 0 < t \leq \delta,$$

при этом $p' = p'(h) \leq p$ при всех $h : \|h\| = 1$.

Для точки $x = x^* + th \in U(x^*), \|h\| = 1$, определим базис $G = G(h) = \{g_1, g_2, \dots, g_n\}$ пространства R^n и соответствующий индекс $q = q(h)$ следующим образом. Рассмотрим матрицы

$$a_k = \frac{1}{(k-1)!} f^{(k)}(x^*) [h]^{k-2}, k = 2, 3, \dots$$

Тогда в точке $x = x^* + th$ для достаточно гладкой функции f справедливы соотношения

$$f^{(2)}(x)th = \sum_{k \geq 2} (k-1) a_k h t^{k-1}, f'(x) = \sum_{k \geq 2} a_k h t^{k-1}, f(x) = \sum_{k \geq 2} \frac{a_k [h]^2 t^k}{k}. \tag{1}$$

Далее определим номер $k_1 \geq 2$ как минимальный среди тех номеров k , для которых одномерное пространство $L_1 = L_{k_1} = \text{Lin}\{a_k h\} \neq \{0\}$. Прямую L_1 переобозначим через L^1 , вектор g_1 определим как $\frac{a_{k_1} h}{\|a_{k_1} h\|}$ и далее номер $k_2 > k_1$ определим как минимальный номер k , для которого $\text{Lin}\{a_k h\}$ не содержится в L^1 . Тогда $L_{k_2} = \text{Lin}\{a_{k_2} h\}, \dim(L_{k_2} \oplus L^1) = 2$, одномерное подпространство L_2 определяется как $\text{Pr}_{(L^1)^\perp} L_{k_2}$, вектор g_2 определяется как нормированный вектор $\text{Pr}_{L_2} a_{k_2} h$. При этом $\text{Pr}_{L_2} a_{k_2} h \subseteq \text{Pr}_{L_2} \text{Im} a_{k_2}$. Далее пространство L^2 определим $L^1 \oplus L_2$. Далее для каждого $j = 3, 4, \dots, q = q(h)$ аналогично определяются прямые $L_j = \text{Pr}_{(L^{j-1})^\perp} L_{k_j}, j = 3, 4, \dots, q$, и соответствующие единичные векторы $g_j, j = 1, 2, \dots, q$. Номер $q = q(h) = \dim \text{Lin}\{a_1 h, a_2 h, \dots\} \leq n$, прямая сумма $L^{j-1} \oplus L_j$ обозначается как L^j и является

подпространством в R^n размерности j . На конечном этапе построено подпространство L^q размерности $q = q(h) \leq n$. Обозначим ортогональное дополнение подпространства L^q через $H : H = (L^q)^\perp$ в случае $q < n$ и $H = \{0\}$ при $q = n$. Определим базис $G(h)$ пространства R^n как набор построенных единичных взаимно ортогональных векторов g_1, g_2, \dots, g_q , направленных вдоль построенных ортогональных прямых $L_l, l = 1, 2, \dots, q$, и произвольного ортогонального базиса $g_{q+1}, g_{q+2}, \dots, g_n$ подпространства H .

Справедлива

Теорема 1. Если для достаточно гладкой функции $f :: R^n \rightarrow R$ производные равномерно по аргументу и в совокупности по порядку ограничены в некоторой окрестности $U(x^*)$ изолированной точки минимума x^* , то для любого фиксированного $h : \|h\| = 1$, система

$$f^{(2)}(x^* + th)s = f'(x^* + th) \tag{2}$$

разрешима относительно s при достаточно малых t .

Доказательство. Решение системы s в базисе G при любом фиксированном векторе h определим по координатам следующим образом. Положим $\text{Pr}_H s = 0$, далее координата s_q в соответствии с (1) определяется из условия $\text{Pr}_{L_q} \left((k_q - 1)a_{k_q} t^{k_q - 2} s \right) = \text{Pr}_{L_q} f'(x) = \text{Pr}_{L_q} \sum_{k \geq k_q} a_k h t^{k-1}$. При этом $s_q = \text{Pr}_{L_q} h t / (k_q - 1) + o(t)$. Подстановка координаты s_q в систему (2) позволяет получить координату $s_{q-1} = \text{Pr}_{L_{q-1}} h t / (k_{q-1} - 1) + o(t)$. Далее последовательно определяются координаты $s_l, l = q - 2, q - 3, \dots, 1$. Решение системы (2) вообще говоря, не единственно, но из (1) следует, что координаты любого решения s в базисе $G = G(h)$ имеют вид

$$s_j = \frac{\text{Pr}_{L_j} h t}{k_j - 1} + o(t), \quad j = 1, 2, \dots, q. \tag{3}$$

Замечание 1. Для получения единственного решения системы (2) определим задачу

$$|s|^2 \rightarrow \min_s, \quad f^{(2)}(x)s = f'(x), \tag{4}$$

решение которой существует при каждом фиксированном h при всех достаточно малых t , при условиях, указанных в теореме 1. При этом длина интервала для t , в пределах которого решение существует, зависит от h .

Пример 1. Для невыпуклой функции $f(x) = x_1^4 + (x_2 - x_1^2)^2$ вывод теоремы 1 нарушается для точек параболы $x_2 = 4x_1^2$. Таким образом, длина интервала, для которого имеет место теорема 1, стремится к нулю, по мере приближения вектора h к $(1, 0)$.

Далее будет показано, что в случае выпуклой функции можно указать общий для всех векторов единичной сферы радиус окрестности переменной t , в пределах которой задача (4) разрешима. Из вида целевой функции задачи (4) следует, что ее решение $s = f^{(2)}(x)^+ f'(x)$ где $(\cdot)^+$ означает псевдообратный оператор на своем образе. Координаты решения этой задачи в базисе G удовлетворяют условию $\text{Pr}_H s = 0$.

Теорема 2. При выполнении условий леммы 2 при всех x , достаточно близких к x^* , задача (4) разрешима и ее решение удовлетворяет условиям

$$(f(x), s) \geq M_1 \|x - x^*\|^{2p}, \quad (f''(x)s, s) \geq M_2 \|x - x^*\|^{2p}, \tag{5}$$

где степень $2p$ определена в лемме 2, константы $M_1 = M_{1,f}, M_2 = M_{2,f} > 0$ не зависят от x .

Доказательство. Из леммы 2 и (1) следует, что при всяком $h : \|h\| = 1$, номера k_j , определяющие базис G , удовлетворяют условию: существует индекс $l_1 \in \{k_1, k_1 + 1, \dots, 2p\}$, для которого $a_{l_1} [h]^2 \geq C > 0$, степень $2p$ определена в лемме 2 и от h не зависит, константа $C > 0$ также не зависит от h и определяется в следствии к лемме 2. Последнее неравенство эквивалентно условию $\frac{a_{l_1} [h]^2}{l_1 - 1} \geq C_1 t^{2p}$ при малых t . Обозначая $\frac{C}{2p - 1}$ через M_1 , получим неравенство в утверждении теоремы. Аналогично из следствия к лемме 2 получается и второе неравенство.

Следствие 4. При выполнении условий леммы 2 для выпуклой функции f при всех x , достаточно близких к x^* , имеет место разрешимость относительно s системы

$$f^{(2)}(x)s = f'(x), \tag{6}$$

что равносильно справедливости включения

$$f'(x) \in \text{Im} f^{(2)}(x) \tag{7}$$

независимо от величины ранга матрицы $f^{(2)}(x)$.

Далее будем обозначать множество $\{x \in R^n : f(x) \leq f(x^*) + \varepsilon\}$ как X_ε^* , диаметр множества A как $\text{diam}A$. Справедлива

Теорема 3. При выполнении условий леммы 2 существует натуральное p и константа $\bar{C} < \infty$, для которых справедливо неравенство

$$\text{diam}X_\varepsilon^* < \bar{C}\varepsilon^{\frac{1}{2p}}. \tag{8}$$

Замечание 2. Указанные ранее свойства справедливы в предположении, что множество точек минимума функции $f(x)$ представляет изолированную точку. Если отказаться от данного предположения, то при выполнении остальных предположений теоремы 1 справедливо представление

$$\text{Argmin}f = \{x^*\} + L, \tag{9}$$

где L — собственное подпространство R^n , которое определяется условием $f^k(x^*)[h]^k = 0$ для всякого натурального k .

Из теорем 1, 2 не следует положительная определенность матрицы $f^{(2)}(x)$ в окрестности точки минимума x^* . Тем не менее они гарантируют применимость метода Ньютона для поиска точки минимума гладкой функции при условии подходящей начальной точки, поскольку задача (1) позволяет получить вектор перехода в итерационной схеме без предположения выпуклости целевой функции. Кроме того, в данном случае удается получить монотонность по аргументу метода Ньютона, а при более сильных предположениях — линейную скорость сходимости. В случае же выпуклости целевой функции полученное свойство справедливо без дополнительных предположений.

Рассмотрим сначала применение результата теоремы 1 для доказательства монотонности по аргументу схемы метода Ньютона. Именно, определим оператор Ньютона $\psi(x, s) = x - f^{(2)}(x)^+ f(x)$, где $f^{(2)}(x)^+ f(x) = s$ — решение задачи (4). Из процесса построения решения s в теореме 1, и замечания к теореме следует, что для любого фиксированного $h : \|h\| = 1$, при достаточно малых значениях t данный оператор корректно определен без предположения выпуклости функции f . Для получения оценки скорости сходимости дополнительно к предположениям теоремы 1 естественно определяется следующее свойство.

Определение 1. Функция f слабо регулярна в точке x^* , если существует константа $d = d_f > 0$, для которой $\max_{j \leq q} |h_j| \geq d$ для всякого $h, \|h\| = 1$. Здесь как и ранее $q = q(h) = \dim \text{Lin}\{a_2h, a_3h, \dots\}, h_j$ — координата вектора h с номером j базисе G .

Теорема 4. Если функция f слабо регулярна в точке x^* , то при выполнении условий теоремы 1 существует $\lambda \in (0, 1)$, для которого $\|\psi(x^* + th, s) - x^*\| \leq \lambda t$ при всех достаточно малых t .

Доказательство. Для любого решения системы (1) в силу предположения слабой регулярности в точке x^* существует номер j , для которого $\|\text{Pr}_{L_{k_j}} h\| > d$. Отсюда

$$\begin{aligned} \|th - s\| &\leq \left\| t \sum_{l=1, l \neq k_j}^q \text{Pr}_{L_l} h - \text{Pr}_{L_l} s \right\| + t \|\text{Pr}_H h\| + |t \text{Pr}_{L_{k_j}} h - \text{Pr}_{L_{k_j}} s| \leq \\ &\leq t(1 - \alpha) + t\alpha \left(1 - \frac{1}{k_j - 1} \right) \leq t \left(1 - \frac{d}{k_j - 1} \right) = \lambda t, \\ \lambda = \lambda(h) &= 1 - \frac{d}{k_j - 1} \in (0, 1), \alpha = \alpha(h) \geq d. \end{aligned}$$

Замечание 3. Полученное неравенство не означает линейной сходимости метода Ньютона, поскольку знаменатель λ , который участвует в оценке, вообще говоря, зависит от h . В случае выпуклой функции полученный результат справедлив в более сильном виде: $\|\psi(x, s) - x^*\| \leq \lambda \|x - x^*\|$ при всех x из достаточно малой окрестности x^* , поскольку для выпуклых функций справедливо следующее свойство.

Определение 2. Функция f равномерно регулярна в точке x^* , если существуют номер $m = m_f \in \mathbb{N}$ и константа $d = d_f > 0$, для которых найдется индекс $j \leq q : k_j \leq m$, такой что $\|\text{Pr}_{L_j} h\| \geq d$.

Имеет место

Теорема 5. Если функция f равномерно регулярна в точке x^* , то при выполнении условий теоремы 1 существует $\lambda \in (0,1)$, для которого $\|\psi(x,s) - x^*\| \leq \lambda \|x - x^*\|$ при всех x из достаточно малой окрестности x^* .

Доказательство. Из теоремы 4 следует, что полученный знаменатель $\lambda = \lambda(h)$ ограничен сверху равномерно по h , $\|h\| = 1 : \lambda(h) = 1 - \frac{d}{k_j - 1} \leq \lambda' < 1$, где $j : k_j \leq m$. Такой номер j существует, как показано в теореме 1. Кроме того, из условия равномерной регулярности следует, что $\|h_H^\perp\| > d$ для некоторого $d > 0$, не зависящего от h , где h_A – проекция вектора h на подпространство A . Отсюда следует оценка для знаменателя λ , не зависящая от h .

Замечание 4. Из леммы 2 следует, выпуклая функция равномерно регулярна в точке x^* : $k_j = k_j(h) \leq m = 2p, \|h\| = 1$. Условие равномерной регулярности является необходимым и достаточным условием линейной скорости сходимости метода Ньютона при выполнении условий теоремы 1 без предположения о выпуклости функции f . Итерационная схема метода Ньютона имеет вид

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

а оценка скорости сходимости метода будет линейная:

$$\|x_{k+1} - x^*\| \leq \lambda \|x_k - x^*\|, \lambda \in (0,1), k = 0,1,2,\dots,$$

при начальном приближении достаточно близком к решению x^* в случае равномерно регулярной в точке минимума функции.

Следствие 5. При выполнении условий теоремы 2 для выпуклой функции скорость сходимости итерационной схемы метода Ньютона является линейной при любом выборе начальной точки x из достаточно малой окрестности решения. Действительно, из теоремы 2 следует, что выпуклая функция является равномерно регулярной в решении x^* , откуда вытекает линейная оценка скорости сходимости.

Для получения сходимости метода Ньютона с более высокой скоростью, чем линейная, рассмотрим модифицированный оператор $\psi_1(x,s)$, по координатной записи которого в базисе G представляет собой $\psi_1(x,s)_j = x - (k_j - 1)s_j, j = 1,2,\dots,q; \psi_1(x,s)_H = x_H - g(x,h)$, где вектор-функция $g(x,h) : H \rightarrow H, s = f^{(2)}(x)^+ f(x)$. Оператор $\psi_1(x,s)_H$ позволяет определить модифицированную схему Ньютона $x_{k+1} = \psi_1(x_k, s_k)$. Из теоремы 1 следует

Теорема 6. Модифицированная схема Ньютона гарантирует сверхлинейную оценку скорости сходимости при хорошем начальном приближении в том и только том случае, когда функция $g(x,h)$ удовлетворяет условию $\|g(x,h) - h_H\| = o(t)$.

СПИСОК ЛИТЕРАТУРЫ

1. Бомаджо Б., Лебедев К. А. Метод Ньютона для нахождения экстремумов сильно выпуклых функций // Международный научно-исследовательский журнал. 2015. Выпуск 6–2 (37). С. 11–14.
2. Заботин В. И., Черняев Ю. А. Метод Ньютона для задачи минимизации выпуклой дважды гладкой функции на предвыпуклом множестве // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 3. С. 340–345. DOI: 10.7868/S0044466918030031.
3. Budzko D., Cordero A., Torregrosa J. R. Modification of Newton's Method to extend the convergence domain // SeMA Journal. 2014. Т. 66. № 1. С. 43–53. DOI: 10.1007/s40324-014-0020-y.
4. Nesterov Y. Accelerating the cubic regularization of Newton's method on convex problems // Mathematical Programming. 2008. Т. 112. № 1. С. 159–181. DOI: 10.1007/s10107-006-0089-x.
5. Polyak B., Tremba A. New versions of Newton method: step-size choice, convergence domain and under-determined equations // Optimization Methods and Software. 2019. С. 1272–1303. DOI: 10.1080/10556788.2019.1669154.
6. Nesterov Y., Polyak B. T. Cubic regularization of Newton method and its global performance // Mathematical Programming. 2006. Т. 108. № 1. С. 177–205.
7. Поляк Б. Т. Градиентные методы минимизации функционалов // Ж. вычисл. матем. и матем. физ. 1963. Т. 3. № 4. С. 643–653.
8. Поляк Б. Т. Метод Ньютона и его роль в оптимизации и вычислительной математике // Труды Института системного анализа Российской академии наук. 2006. Т. 28. С. 44–62.
9. Colding T. H., Minicozzi W. P. Lojasiewicz inequalities and applications // arXiv:1402.5087. 2014.
10. Lojasiewicz S. Division d'une distribution par une fonction analytique de variables reelles // C. R. Acad. Sci. 1958. Т. 246. № 5. С. 683–686.

ON THE REDUNDANCY OF HESSIAN NONSINGULARITY FOR LINEAR CONVERGENCE RATE OF THE NEWTON METHOD APPLIED TO THE MINIMIZATION OF CONVEX FUNCTIONS

Yu. G. Evtushenko^{a,b,*}, A. A. Tret'yakov^{a,c,**}

^a*Federal Research Center "Information Science and Control," Russian Academy of Sciences,
Vavilova St., 44, Moscow, 119333 Russia*

^b*Moscow Institute of Physics and Technology,
Institutskiy per., 9, Dolgoprudny, Moscow oblast, 141701 Russia*

^c*Siedlce University, Faculty of Exact and Natural Sciences,
Siedlce, 08-110 Poland*

**e-mail: yuri-evtushenko@yandex.ru*

***e-mail: prof.tretyakov@gmail.com*

Received 10 August, 2023

Revised 07 November, 2023

Accepted 07 November, 2023

Abstract. A new property of convex functions that makes it possible to achieve the linear rate of convergence of the Newton method during the minimization process is established. Namely, it is proved that, even in the case of singularity of the Hessian at the solution, the Newtonian system is solvable in the vicinity of the minimizer; i.e., the gradient of the objective function belongs to the image of the matrix of second derivatives and, therefore, analogs of the Newton method may be used.

Keywords: convex function, Newton method, solvability, convergence, convergence rate, regularity.