В. Н. Малозёмов, Г. Ш. Тамасян, О направлении наискорейшего спуска, *Вестн. С.-Петербург. ун-та. Сер. 10. Прикл. матем. Информ. Проц. упр.*, 2019, том 15, выпуск 4, 489–501

DOI: 10.21638/11701/spbu10.2019.406

УДК 519.85
MSC 49J52

Вестник СПбГУ. Прикладная математика. Информатика... 2019. Т. 15. Вып. 4

О направлении наискорейшего спуска

В. Н. Малозёмов, Г. Ш. Тамасян

Санкт-Петербургский государственный университет, Российская Федерация,
199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Малозёмов В. Н., Тамасян Г. Ш.* О направлении наискорейшего спуска // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2019. Т. 15. Вып. 4. С. 489–501.
https://doi.org/10.21638/11702/spbu10.2019.406

Статья посвящена памяти проф. В. Ф. Демьянова (1938–2014). Основные научные интересы В. Ф. Демьянова лежали в области численных методов оптимизации, где понятие направления наискорейшего спуска играет важную роль. Это понятие вводится как для гладких, так и для негладких функций, при отсутствии ограничений и при их наличии. В данной статье дается детальный анализ методов построения направления наискорейшего спуска. Во всех случаях дело сводится к решению задачи квадратичного программирования. Особое внимание уделяется негладким функциям, в изучение которых В. Ф. Демьянов внес значительный вклад. Рассматриваются функции поточечного максимума и квазидифференцируемые функции. Приводится пример квазидифференцируемой в некоторой точке функции, у которой имеются два направления наискорейшего спуска и два направления наискорейшего подъема.

Ключевые слова: направление наискорейшего спуска, негладкий анализ, квазидифференциал.

1. Дифференцируемая в точке x_0 функция n переменных $f(x)$ характеризуется разложением
$$f(x_0+h) = f(x_0) + \langle f'(x_0), h\rangle + o(\|h\|), \tag{1}$$
где $o(\|h\|)/\|h\| \longrightarrow 0$ при $h \longrightarrow \mathbf{0}$. Обозначим $S = \{g \in \mathbb{R}^n \mid \|g\| = 1\}$. Любой вектор $g \in S$ будем называть *направлением.* Производная функции $f(x)$ в точке $x = x_0$ по направлению g определяется естественным образом:
$$\frac{\partial f(x_0)}{\partial g} = \lim_{\alpha\to+0} \frac{f(x_0+\alpha g) - f(x_0)}{\alpha}.$$
Из формулы (1) следует, что при $\alpha > 0$
$$\frac{1}{\alpha}\big[f(x_0+\alpha g) - f(x_0)\big] = \langle f'(x_0), g\rangle + \frac{o(\|\alpha g\|)}{\|\alpha g\|}.$$
В пределе при $\alpha \longrightarrow +0$ получаем
$$\frac{\partial f(x_0)}{\partial g} = \langle f'(x_0), g\rangle.$$

Вектор $g_0 \in S$, на котором
$$\frac{\partial f(x_0)}{\partial g_0} = \min_{g\in S}\langle f'(x_0), g\rangle,$$
называется *направлением наискорейшего спуска* функции $f(x)$ в точке $x = x_0$.

При $f'(x_0) \neq \mathbf{0}$ вектор g_0 легко найти. По неравенству Коши—Буняковского при $g \in S$ имеем

$$-\langle f'(x_0), g\rangle = \langle f'(x_0), -g\rangle \leqslant \|f'(x_0)\|, \tag{2}$$

причем неравенство выполняется как равенство только тогда, когда $-g = \alpha f'(x_0)$ при некотором $\alpha > 0$. Так как $\|g\| = 1$, то $\alpha = \|f'(x_0)\|^{-1}$. Таким образом, в силу (2),

$$\langle f'(x_0), g\rangle \geqslant -\|f'(x_0)\| \qquad \forall\ g \in S,$$

и неравенство выполняется как равенство только при $g = g_0 := -\dfrac{f'(x_0)}{\|f'(x_0)\|}$. Указанный вектор g_0 и является направлением наискорейшего спуска функции $f(x)$ в точке $x = x_0$.

2. Возьмем симметричную положительно определенную матрицу D порядка n. Введем обобщенные скалярное произведение и норму:

$$\langle x, y\rangle_D = \langle Dx, y\rangle, \qquad \|x\|_D = \sqrt{\langle x, x\rangle_D}.$$

Пусть $f'(x_0) \neq \mathbf{0}$. Обобщенное направление наискорейшего спуска определим как решение следующей экстремальной задачи:

$$\langle f'(x_0), h\rangle \longrightarrow \min \quad \text{при ограничении} \quad \langle Dh, h\rangle = 1. \tag{3}$$

Это решение можно получить с помощью обобщенного неравенства Коши—Буняковского

$$\langle x, y\rangle_D \leqslant \|x\|_D \|y\|_D, \tag{4}$$

которое при ненулевых x, y обращается в равенство только тогда, когда $y = \alpha x$ при некотором $\alpha > 0$ (см., например, [1, с. 33]). Согласно (4), при $\|h\|_D = 1$ имеем

$$-\langle f'(x_0), h\rangle = \left\langle D\left(D^{-1} f'(x_0)\right), -h\right\rangle = \langle D^{-1} f'(x_0), -h\rangle_D \leqslant \|D^{-1} f'(x_0)\|_D$$

или

$$\langle f'(x_0), h\rangle \geqslant -\|D^{-1} f'(x_0)\|_D.$$

Неравенство выполняется как равенство только при $-h = \alpha D^{-1} f'(x_0)$ с некоторым $\alpha > 0$. Так как $\|h\|_D = 1$, то $\alpha = \|D^{-1} f'(x_0)\|_D^{-1}$. Единственным решением задачи (3) является вектор

$$h_D = -\frac{D^{-1} f'(x_0)}{\sqrt{\langle D^{-1} f'(x_0), f'(x_0)\rangle}}. \tag{5}$$

По формуле (5) находится направление D-наискорейшего спуска.

3. Наряду с задачей (3) рассмотрим задачу минимизации выпуклой квадратичной функции

$$Q(h) := \langle f'(x_0), h\rangle + \tfrac{1}{2}\langle Dh, h\rangle \longrightarrow \min_{h \in \mathbb{R}^n}. \tag{6}$$

Для нее критерий оптимальности имеет вид $Q'(h) = \mathbf{0}$ или $f'(x_0) + Dh = \mathbf{0}$. Значит, единственным решением задачи (6) будет вектор $h_0 = -D^{-1} f'(x_0)$. При $f'(x_0) \neq \mathbf{0}$ вектор h_0 отличается от направления D-наискорейшего спуска (5) только нормировкой, $h_D = h_0 / \|h_0\|_D$.

Получили следующий важный факт: *при* $f'(x_0) \neq \mathbf{0}$ (т. е. когда точка x_0 не является стационарной) *решения задач* (3) *и* (6) *различаются лишь нормировкой.*

Замечание 1. Пусть функция $f(x)$ дважды дифференцируема в точке x_0 и матрица $D = f''(x_0)$ положительно определена. Тогда вектор

$$h_0 = -D^{-1}f'(x_0) = -\left(f''(x_0)\right)^{-1} f'(x_0)$$

совпадает с вектором спуска в методе минимизации Ньютона—Рафсона [2].

4. Понятие направления наискорейшего спуска можно ввести и для задачи минимизации с ограничениями.

Рассмотрим, например, случай линейных ограничений-равенств

$$f(x) \longrightarrow \inf, \qquad Ax = b. \tag{7}$$

Здесь A — $(m \times n)$-матрица с линейно независимыми строками. Множество планов задачи (7) обозначим Ω. Пусть $x_0 \in \Omega$. Нас интересуют векторы сдвига h, удовлетворяющие условию $x_0 + h \in \Omega$ (не выводящие из множества планов). Такие векторы образуют множество $L = \{h \in \mathbb{R}^n \mid Ah = \mathbf{0}\}$. Направление наискорейшего спуска функции $f(x)$ в точке x_0 при ограничениях $Ax = b$ определим как решение экстремальной задачи

$$\begin{aligned} &\langle f'(x_0), h\rangle \longrightarrow \min, \\ &Ah = \mathbf{0}, \quad \|h\| = 1. \end{aligned} \tag{8}$$

Обозначим через P матрицу ортогонального проектирования на подпространство L. Как известно [1, с. 52–55],

$$P = E - A^T(AA^T)^{-1}A,$$

где E — единичная матрица n-го порядка. Условие $Pf'(x_0) = \mathbf{0}$ характеризует точку x_0 как стационарную для задачи (7). Она удовлетворяет необходимому условию минимума $f'(x_0) = A^T u$ при $u = (A^T A)^{-1}Af'(x_0)$.

Теорема 1. *Пусть проекция $Pf'(x_0)$ градиента $f'(x_0)$ на подпространство L отлична от нуля. Тогда вектор*

$$h_0 = -\frac{Pf'(x_0)}{\|Pf'(x_0)\|}$$

является единственным решением задачи (8).

Д о к а з а т е л ь с т в о. Так как $AP = \mathbb{O}$, то h_0 — план задачи (8). Обозначим $c = f'(x_0)$ и возьмем план h задачи (8), отличный от h_0. Нужно показать, что $\langle c, h\rangle > \langle c, h_0\rangle$.

По определению матрица P симметрична и $PP = P$, поэтому

$$\langle Pc, c\rangle = \langle PPc, c\rangle = \langle Pc, Pc\rangle = \|Pc\|^2$$

и

$$\langle c, h_0\rangle = -\frac{\langle Pc, c\rangle}{\|Pc\|} = -\|Pc\|.$$

Далее, из условий $h \neq h_0$ и $\|h\| = \|h_0\| = 1$ следует, что

$$\langle h, h_0\rangle < 1.$$

Наконец, $Ah = \mathbf{0}$. Значит,

$$Ph = h \ \text{ и } \ \langle Pc, h\rangle = \langle c, Ph\rangle = \langle c, h\rangle.$$

Теперь имеем

$$\langle c, h\rangle = \langle Pc, h\rangle = -\|Pc\|\langle h_0, h\rangle > -\|Pc\| = \langle c, h_0\rangle.$$

Теорема доказана. □

5. Наряду с задачей (8) рассмотрим задачу минимизации выпуклой квадратичной функции

$$Q(\lambda) := \tfrac{1}{2}\left\|A^T\lambda - f'(x_0)\right\|^2 \longrightarrow \min_{\lambda\in\mathbb{R}^m}. \qquad (9)$$

Так как

$$\begin{aligned}\left\|A^T\lambda - f'(x_0)\right\|^2 &= \left\langle A^T\lambda - f'(x_0), A^T\lambda - f'(x_0)\right\rangle = \\ &= \left\langle AA^T\lambda, \lambda\right\rangle - 2\left\langle A^T\lambda, f'(x_0)\right\rangle + \|f'(x_0)\|^2,\end{aligned}$$

то задача (9) сводится к следующей экстремальной задаче:

$$\tfrac{1}{2}\left\langle AA^T\lambda, \lambda\right\rangle - \langle Af'(x_0), \lambda\rangle \longrightarrow \min_{\lambda\in\mathbb{R}^m}.$$

Запишем критерий оптимальности: $AA^T\lambda - Af'(x_0) = \mathbf{0}$. Получим единственное решение $\lambda_0 = \left(AA^T\right)^{-1}Af'(x_0)$. В этом случае

$$A^T\lambda_0 - f'(x_0) = -\left(E - A^T\left(AA^T\right)^{-1}A\right)f'(x_0) = -Pf'(x_0).$$

Таким образом, решение задачи (9) порождает с точностью до нормировки направление наискорейшего спуска функции $f(x)$ в точке x_0 при ограничениях $Ax = b$.

6. Понятие направления наискорейшего спуска можно ввести и для некоторых классов негладких функций.

Рассмотрим функцию дискретного максимума

$$\varphi(x) = \max_{i\in M} f_i(x), \qquad x \in \mathbb{R}^n,$$

где M — конечное индексное множество. Для нее производная по направлению g в точке x_0 вводится обычным способом:

$$\frac{\partial\varphi(x_0)}{\partial g} = \lim_{\alpha\to+0}\frac{\varphi(x_0+\alpha g) - \varphi(x_0)}{\alpha}.$$

Обозначим $M(x) = \{i \in M \mid f_i(x) = \varphi(x)\}$.

Теорема 2 (В. Ф. Демьянов). *Если все функции $f_i(x)$, $i \in M$, непрерывно дифференцируемы в окрестности точки x_0, то производная функции максимума $\varphi(x)$ в точке x_0 по любому направлению $g \in S$ существует. При этом*

$$\frac{\partial\varphi(x_0)}{\partial g} = \max_{i\in M(x_0)}\langle f_i'(x_0), g\rangle. \qquad (10)$$

Доказательство этой теоремы опирается на следующий более общий результат.

Теорема 3. *В условиях теоремы 2 справедливо разложение*

$$\varphi(x_0 + h) = \varphi(x_0) + \max_{i \in M(x_0)} \langle f_i'(x_0), h \rangle + o(\|h\|),$$

где $o(\|h\|)/\|h\| \longrightarrow 0$ *при* $h \longrightarrow \mathbf{0}$.

Положительно однородный по h функционал

$$\ell(x_0, h) = \max_{i \in M(x_0)} \langle f_i'(x_0), h \rangle$$

называется *квазидифференциалом* функции $\varphi(x)$ в точке x_0.

Доказательство теоремы 3. Прежде всего отметим, что при малых h выполняется равенство

$$\varphi(x_0 + h) = \max_{i \in M(x_0)} f_i(x_0 + h). \tag{11}$$

Действительно, разности $r_i(x_0) = \varphi(x_0) - f_i(x_0)$ положительны при $i \notin M(x_0)$ (рис. 1). При малых h положительными будут и $r_i(x_0 + h)$, т. е.

$$f_i(x_0 + h) < \varphi(x_0 + h), \quad i \notin M(x_0).$$

Значит, максимум $f_i(x_0 + h)$ по всем $i \in M$, равный $\varphi(x_0 + h)$, может достигаться только на индексах из $M(x_0)$. Равенство (11) и отражает этот факт.

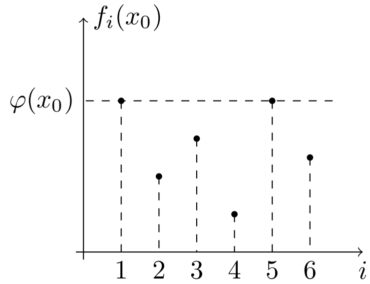

Рис. 1. График $f_i(x_0)$ как функции от i

Согласно (11) и определению $M(x_0)$, при малых h имеем

$$\varphi(x_0 + h) - \varphi(x_0) = \max_{i \in M(x_0)} \big\{ f_i(x_0 + h) - \varphi(x_0) \big\} = \max_{i \in M(x_0)} \big\{ f_i(x_0 + h) - f_i(x_0) \big\} =$$
$$= \max_{i \in M(x_0)} \big\{ \langle f_i'(x_0), h \rangle + \langle f_i'(x_0 + \theta_i h) - f_i'(x_0), h \rangle \big\} =$$
$$= \max_{i \in M(x_0)} \langle f_i'(x_0), h \rangle + o(\|h\|).$$

Здесь

$$\big| o(\|h\|) \big| \leqslant \max_{i \in M(x_0)} \| f_i'(x_0 + \theta_i h) - f_i'(x_0) \| \cdot \|h\|.$$

Ясно, что по $\varepsilon > 0$ найдется такое $\delta > 0$, что при $\|h\| < \delta$ будет выполняться неравенство

$$\max_{i \in M(x_0)} \| f_i'(x_0 + \theta_i h) - f_i'(x_0) \| < \varepsilon.$$

А это означает, что $o(\|h\|)/\|h\| \longrightarrow 0$ при $h \longrightarrow \mathbf{0}$.

Теорема 3 доказана.

Теорема 2 является очевидным следствием теоремы 3.

7. Обозначим через $C(x_0)$ выпуклую оболочку градиентов $v_i = f_i(x_0)$, $i \in M(x_0)$. Нетрудно проверить, что

$$\max_{i\in M(x_0)} \langle f_i'(x_0), g\rangle = \max_{v\in C(x_0)} \langle v, g\rangle.$$

Приняв это во внимание, перепишем формулу (10) в виде

$$\frac{\partial\varphi(x_0)}{\partial g} = \max_{v\in C(x_0)} \langle v, g\rangle.$$

Направление наискорейшего спуска g_0 функции $\varphi(x)$ в точке x_0 определяется как решение следующей экстремальной задачи:

$$\chi_0(g) := \max_{v\in C(x_0)} \langle v, g\rangle \longrightarrow \min_{g\in S}. \tag{12}$$

Если $\chi_0(g) \geqslant 0$ при всех $g \in S$, то x_0 называется стационарной точкой функции $\varphi(x)$ (в ней производные функции $\varphi(x)$ по всем направлениям неотрицательны). Условие стационарности равносильно включению $\mathbf{0} \in C(x_0)$ (рис. 2).

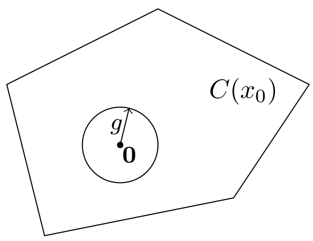

Рис. 2. Пояснение к условию стационарности

Нас интересуют нестационарные точки x_0, в которых $\mathbf{0} \notin C(x_0)$.

Теорема 4 (В. Ф. Демьянов). *Пусть $\mathbf{0} \notin C(x_0)$. Тогда направлением наискорейшего спуска функции $\varphi(x)$ в точке x_0 является вектор $g_0 = -v_*/\|v_*\|$, где v_* — точка многогранника $C(x_0)$, ближайшая к началу координат.*

Д о к а з а т е л ь с т в о. Точка v_* является точкой минимума функции $\|v\|^2$ на $C(x_0)$. Это значит, что $v_* \in C(x_0)$ и $\|v\|^2 \geqslant \|v_*\|^2$ при всех $v \in C(x_0)$. Так как $\mathbf{0} \notin C(x_0)$, то $v_* \neq \mathbf{0}$.

Зафиксируем $v \in C(x_0)$. В силу выпуклости множества $C(x_0)$, точка $v(t) = v_* + t(v - v_*)$ при $t \in (0,1)$ принадлежит $C(x_0)$. Значит, $\|v(t)\|^2 \geqslant \|v_*\|^2$ или $\|v_* + t(v - v_*)\|^2 \geqslant \|v_*\|^2$. Отсюда следует, что

$$2t\langle v_*, v - v_*\rangle + t^2\|v - v_*\|^2 \geqslant 0 \qquad \forall\, t \in (0,1).$$

Поделим на $2t$ и перейдем к пределу при $t \longrightarrow +0$. Получим

$$\langle v, v_*\rangle \geqslant \langle v_*, v_*\rangle \qquad \forall\, v \in C(x_0). \tag{13}$$

Теперь легко доказать требуемое неравенство

$$\chi_0(g_0) \leqslant \chi_0(g) \qquad \forall\, g \in S. \tag{14}$$

Действительно, согласно (13) и определению вектора g_0, имеем

$$\langle v, g_0\rangle = -\tfrac{1}{\|v_*\|}\langle v, v_*\rangle \leqslant -\tfrac{1}{\|v_*\|}\langle v_*, v_*\rangle = -\|v_*\|,$$

причем при $v = v_*$ неравенство выполняется как равенство. Значит,

$$\chi_0(g_0) := \max_{v\in C(x_0)}\langle v, g_0\rangle = -\|v_*\|. \tag{15}$$

Далее, при всех $g \in S$ выполняется неравенство

$$\langle v_*, -g\rangle \leqslant \|v_*\| \tag{16}$$

и только при $-g = \alpha v_*$ с некоторым $\alpha > 0$, т. е. только при $g = g_0$ это неравенство выполняется как равенство.

На основании (15) и (16) при всех $g \in S$ получаем

$$\chi_0(g_0) = -\|v_*\| \leqslant \langle v_*, g\rangle \leqslant \max_{v\in C(x_0)}\langle v, g\rangle = \chi_0(g), \tag{17}$$

что соответствует (14).

Установлено, что вектор g_0 является решением задачи (12). Так как неравенство (16), используемое в (17), при $g \neq g_0$ выполняется как строгое, то g_0 — единственное решение задачи (12).

Теорема доказана. □

8. Рассмотрим дискретную минимаксную задачу при наличии линейных ограничений-равенств:

$$\varphi(x) := \max_{i\in M} f_i(x) \longrightarrow \inf, \qquad Ax = b.$$

Пусть x_0 — план этой задачи. Направлением наискорейшего спуска функции $\varphi(x)$ в точке x_0 при ограничениях $Ax = b$ называется решение следующей экстремальной задачи:

$$\chi_0(g) := \max_{v\in C(x_0)}\langle v, g\rangle \longrightarrow \min, \qquad Ag = \mathbf{0}, \quad \|g\| = 1. \tag{18}$$

Введем множества $H = \left\{h = A^T\lambda \mid \lambda \in \mathbb{R}^m\right\}$ и

$$P(x_0) = H - C(x_0).$$

Множество $P(x_0)$ выпуклое и замкнутое (замкнутость следует из компактности $C(x_0)$). Значит, существует точка $p_0 \in P(x_0)$, ближайшая к началу координат. При $p_0 = \mathbf{0}$ точка x_0 будет стационарной.

Теорема 5. *Пусть $p_0 \neq \mathbf{0}$. Тогда вектор $g_0 = p_0/\|p_0\|$ является единственным решением задачи* (18).

Д о к а з а т е л ь с т в о. По определению множества $P(x_0)$ имеем $p_0 = h_0 - v_0$, где $h_0 \in H$ и $v_0 \in C(x_0)$. Покажем, что

$$\langle p_0, h\rangle = 0 \qquad \forall\, h \in H. \tag{19}$$

Так как вектор αh принадлежит H при любом $\alpha \in \mathbb{R}$, то $\|p_0 - \alpha h\|^2 \geqslant \|p_0\|^2$. Это равносильно неравенству

$$-2\alpha\langle p_0, h\rangle + \alpha^2\|h\|^2 \geqslant 0 \qquad \forall\, \alpha \in \mathbb{R}.$$

Отсюда очевидным образом следует (19).

Учитывая, что $h = A^T\lambda$, на основании (19) получаем

$$0 = \langle p_0, h\rangle = \langle p_0, A^T\lambda\rangle = \langle Ap_0, \lambda\rangle \qquad \forall\, \lambda \in \mathbb{R}^m.$$

Это гарантирует равенство $Ap_0 = \mathbf{0}$. Теперь ясно, что вектор $g_0 = p_0/\|p_0\|$ является планом задачи (18).

Вычислим $\chi_0(g_0)$. Возьмем вектор $v \in C(x_0)$. Тогда $-v \in P(x_0)$. По неравенству, аналогичному (13), $\langle p, p_0\rangle \geqslant \langle p_0, p_0\rangle$ для всех $p \in P(x_0)$, поэтому $\langle -v, p_0\rangle \geqslant \langle p_0, p_0\rangle$ или $\langle v, p_0\rangle \leqslant -\|p_0\|^2$. При $v = v_0$ это неравенство выполняется как равенство. Действительно, согласно (19),

$$\langle v_0, p_0\rangle = -\langle h_0 - v_0, p_0\rangle = -\|p_0\|^2.$$

Теперь при всех $v \in C(x_0)$ имеем

$$\langle v, g_0\rangle = \frac{1}{\|p_0\|}\langle v, p_0\rangle \leqslant -\|p_0\|,$$

причем при $v = v_0$ неравенство выполняется как равенство. Значит,

$$\chi_0(g_0) := \max_{v\in C(x_0)} \langle v, g_0\rangle = -\|p_0\|. \tag{20}$$

Осталось проверить оптимальность g_0. Возьмем план g задачи (18), отличный от g_0. Имеем $\langle p_0, g\rangle \leqslant \|p_0\|$ или $-\|p_0\| \leqslant -\langle p_0, g\rangle$. Равенство в последнем неравенстве выполняется только тогда, когда $g = \alpha p_0$ при некотором $\alpha > 0$, т. е. когда $g = g_0$. По условию $g \neq g_0$, поэтому $-\|p_0\| < -\langle p_0, g\rangle$.

С учетом (20) запишем

$$\chi_0(g_0) = -\|p_0\| < -\langle p_0, g\rangle = \langle v_0, g\rangle - \langle h_0, g\rangle. \tag{21}$$

Пусть $h_0 = A^T\lambda_0$. Вектор g как план задачи (18) удовлетворяет условию $Ag = \mathbf{0}$. Следовательно,

$$\langle h_0, g\rangle = \langle A^T\lambda_0, g\rangle = \langle \lambda_0, Ag\rangle = 0.$$

Теперь из (21) вытекает, что

$$\chi_0(g_0) < \langle v_0, g\rangle \leqslant \max_{v\in C(x_0)} \langle v, g\rangle = \chi_0(g).$$

Теорема доказана. □

9. В. Ф. Демьянов и А. М. Рубинов ввели общее понятие квазидифференцируемой функции [3, 4].

Определение. Функция n переменных $f(x)$ называется *квазидифференцируемой* в точке x_0, если ее приращение допускает представление

$$f(x_0 + h) - f(x_0) = \max_{v\in\underline{\partial} f(x_0)} \langle v, h\rangle + \min_{w\in\overline{\partial} f(x_0)} \langle w, h\rangle + o(\|h\|), \tag{22}$$

где $\underline{\partial} f(x_0)$ и $\overline{\partial} f(x_0)$ — выпуклые компакты.

Положительно однородный по h функционал

$$\ell(x_0, h) = \max_{v\in\underline{\partial} f(x_0)} \langle v, h\rangle + \min_{w\in\overline{\partial} f(x_0)} \langle w, h\rangle \tag{23}$$

называется *квазидифференциалом* функции $f(x)$ в точке x_0. Очевидно, что он непрерывен по h. Множества $\underline{\partial} f(x_0)$ и $\overline{\partial} f(x_0)$, определяющие $\ell(x_0,h)$, называются *субдифференциальным* и *супердифференциальным* множествами соответственно.

Из (22) следует, что у квазидифференцируемой в точке x_0 функции $f(x)$ существуют производные по всем направлениям и что справедлива формула

$$\frac{\partial f(x_0)}{\partial g} = \ell(x_0, g) \qquad \forall g \in S.$$

Направлением наискорейшего спуска является решение экстремальной задачи

$$\chi_0(g) := \ell(x_0, g) \longrightarrow \min_{g\in S}. \tag{24}$$

В силу непрерывности $\ell(x_0,h)$ по h, решение задачи (24) существует.

Преобразуем выражение для $\ell(x_0,g)$. Согласно (23), имеем

$$\ell(x_0,g) = \min_{w\in\overline{\partial} f(x_0)} \bigl\{ \langle w, g\rangle + \max_{v\in\underline{\partial} f(x_0)} \langle v, g\rangle \bigr\} =$$
$$= \min_{w\in\overline{\partial} f(x_0)} \max_{v\in\underline{\partial} f(x_0)} \langle v+w, g\rangle.$$

Точка x_0, в которой производные по всем направлениям неотрицательны,

$$\ell(x_0,g) \geqslant 0 \qquad \forall g \in S, \tag{25}$$

называется inf-*стационарной* точкой квазидифференцируемой функции $f(x)$.

Теорема 6 (Л. Н. Полякова). *Условие* inf-*стационарности* (25) *равносильно вложению*

$$-\overline{\partial} f(x_0) \subset \underline{\partial} f(x_0). \tag{26}$$

Д о к а з а т е л ь с т в о. Зафиксируем $w \in \overline{\partial} f(x_0)$. Согласно (25), имеем

$$\max_{v\in\underline{\partial} f(x_0)} \langle v+w, g\rangle \geqslant 0 \qquad \forall g \in S.$$

Это означает, что $\mathbf{0} \in \underline{\partial} f(x_0) + w$ или

$$-w \in \underline{\partial} f(x_0) \qquad \forall w \in \overline{\partial} f(x_0).$$

Полученное включение равносильно вложению (26). ☐

Отметим также, что условие (26) эквивалентно соотношению

$$\max_{w\in\overline{\partial} f(x_0)} \min_{v\in\underline{\partial} f(x_0)} \|v+w\| = 0.$$

10. Возьмем точку x_0, не являющуюся inf-*стационарной* для квазидифференцируемой функции $f(x)$. Для нее

$$\min_{g\in S} \chi_0(g) < 0 \qquad \text{и} \qquad a := \max_{w\in\overline{\partial} f(x_0)} \min_{v\in\underline{\partial} f(x_0)} \|v+w\| > 0.$$

Покажем, что

$$\min_{g\in S} \chi_0(g) = -a. \tag{27}$$

Имеем $\langle v+w, g\rangle \geqslant -\|v+w\|$, поэтому

$$\max_{v\in\underline{\partial} f(x_0)} \langle v+w, g\rangle \geqslant -\min_{v\in\underline{\partial} f(x_0)} \|v+w\|$$

и

$$\min_{w\in\overline{\partial} f(x_0)} \max_{v\in\underline{\partial} f(x_0)} \langle v+w, g\rangle \geqslant -\max_{w\in\overline{\partial} f(x_0)} \min_{v\in\underline{\partial} f(x_0)} \|v+w\|.$$

Отсюда следует, что

$$\min_{g\in S} \chi_0(g) = \min_{g\in S} \ell(x_0, g) \geqslant -a. \tag{28}$$

Теперь возьмем точку $w_0 \in \overline{\partial} f(x_0)$, на которой

$$\min_{v\in\underline{\partial} f(x_0)} \|v+w_0\| = a. \tag{29}$$

(Единственность w_0 не гарантируется.) Так как $a > 0$, то

$$\mathbf{0} \notin \underline{\partial} f(x_0) + w_0. \tag{30}$$

Запишем

$$\begin{aligned}\min_{g\in S} \chi_0(g) &= \min_{g\in S} \min_{w\in\overline{\partial} f(x_0)} \max_{v\in\underline{\partial} f(x_0)} \langle v+w, g\rangle = \\ &= \min_{w\in\overline{\partial} f(x_0)} \min_{g\in S} \max_{v\in\underline{\partial} f(x_0)} \langle v+w, g\rangle \leqslant \\ &\leqslant \min_{g\in S} \max_{v\in\underline{\partial} f(x_0)} \langle v+w_0, g\rangle.\end{aligned}$$

Согласно теореме 4, при выполнении условия (30) справедливо равенство

$$\min_{g\in S} \max_{v\in\underline{\partial} f(x_0)} \langle v+w_0, g\rangle = -\min_{v\in\underline{\partial} f(x_0)} \|v+w_0\| = -a \tag{31}$$

(мы воспользовались формулой (29)). Значит,

$$\min_{g\in S} \chi_0(g) \leqslant -a. \tag{32}$$

Объединив неравенства (28) и (32), придем к равенству (27).

Обозначим через v_0 решение экстремальной задачи

$$\|v+w_0\| \longrightarrow \min_{v\in\underline{\partial} f(x_0)}.$$

Как отмечалось при доказательстве теоремы 4, в формуле (31) минимум по $g \in S$ достигается на векторе

$$g_0 = -\frac{v_0+w_0}{\|v_0+w_0\|}. \tag{33}$$

Таким образом,

$$\max_{v\in\underline{\partial} f(x_0)} \langle v+w_0, g_0\rangle = -a. \tag{34}$$

Покажем, что g_0 — направление наискорейшего спуска, т. е. что

$$\chi_0(g_0) = -a.$$

Согласно (34), имеем

$$\chi_0(g_0) = \min_{w \in \overline{\partial} f(x_0)} \max_{v \in \underline{\partial} f(x_0)} \langle v + w, g_0 \rangle \leqslant \max_{v \in \underline{\partial} f(x_0)} \langle v + w_0, g_0 \rangle = -a. \tag{35}$$

На основании (27) и (35) заключаем, что $\chi_0(g_0) = -a$.

Подведем итог.

Теорема 7. *Пусть точка x_0 не является* inf*-стационарной для квазидифференцируемой функции $f(x)$. Тогда направление наискорейшего спуска функции $f(x)$ из точки x_0 определяется формулой (33), в которой w_0 — решение экстремальной задачи*

$$\min_{v \in \underline{\partial} f(x_0)} \|v + w\| \longrightarrow \max_{w \in \overline{\partial} f(x_0)} \tag{36}$$

и v_0 — решение экстремальной задачи

$$\|v + w_0\| \longrightarrow \min_{v \in \underline{\partial} f(x_0)}. \tag{37}$$

11. Приведем характерный пример квазидифференцируемой функции двух переменных.

Пример. Рассмотрим функцию

$$f(x) = |x_1| - |x_2|. \tag{38}$$

Возьмем точку $x_0 = \mathbf{0}$. Имеем

$$f(x_0 + h) - f(x_0) = f(h) = |h_1| - |h_2|.$$

Так как

$$|h_1| = \max\{-h_1, h_1\} = \max_{v \in \{-e_1, e_1\}} \langle v, h \rangle, \quad -|h_2| = \min\{-h_2, h_2\} = \min_{w \in \{-e_2, e_2\}} \langle w, h \rangle,$$

то

$$f(x_0 + h) - f(x_0) = \max_{v \in \{-e_1, e_1\}} \langle v, h \rangle + \min_{w \in \{-e_2, e_2\}} \langle w, h \rangle.$$

Получим, что функция $f(x)$ вида (38) квазидифференцируема в точке $x_0 = \mathbf{0}$. При этом (рис. 3)

$$\underline{\partial} f(x_0) = \{-e_1, e_1\}, \qquad \overline{\partial} f(x_0) = \{-e_2, e_2\}.$$

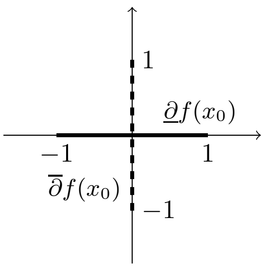

Рис. 3. Суб- и супердифференциальные множества

Условие (26) в данном случае не выполняется. Значит, точка $x_0 = \mathbf{0}$ не является inf-стационарной.

Задача (36) имеет два решения: $w_0 = -e_2$ и $w_0 = e_2$. Решением задачи (37) как при $w_0 = -e_2$, так и при $w_0 = e_2$ является точка $v_0 = \mathbf{0}$. Таким образом, у функции $f(x)$ вида (38) существуют два направления наискорейшего спуска: $g_0 = e_2$ и $g_0 = -e_2$.

Замечание 2. Из геометрических соображений очевидно, что у функции $f(x)$ вида (38) в точке $x_0 = \mathbf{0}$ существуют два направления наискорейшего подъема: $g_1 = e_1$ и $g_1 = -e_1$.

Как известно, в гладком случае направления наискорейшего спуска и наискорейшего подъема противоположны. Для квазидифференцируемых функций это свойство, вообще говоря, не имеет места.

12. Вопросам дифференцируемости по направлениям негладких функций В. Ф. Демьянов посвятил книгу [5].

Литература

1. *Малоземов В. Н.* Линейная алгебра без определителей. Квадратичная функция. СПб.: Изд-во С.-Петерб. ун-та, 1997. 80 с.

2. *Малоземов В. Н.* Метод Ньютона—Рафсона для безусловной минимизации // Семинар «CNSA&NDO». Избр. докл. 14 февраля 2019 г. URL: http://apmath.spbu.ru/cnsa/reps19.shtml#0214 (дата обращения: 13.10.2019 г.)

3. *Демьянов В. Ф., Рубинов А. М.* О квазидифференцируемых функционалах // Докл. АН СССР. 1980. Т. 250. № 1. С. 21–25.

4. *Демьянов В. Ф., Рубинов А. М.* Основы негладкого анализа и квазидифференциальное исчисление. М.: Наука, 1990. 432 с.

5. *Демьянов В. Ф.* Минимакс: дифференцируемость по направлениям. Л.: Изд-во Ленингр. ун-та, 1974. 112 с.

Статья поступила в редакцию 1 ноября 2019 г.
Статья принята к печати 7 ноября 2019 г.

Контактная информация:

Малозёмов Василий Николаевич — д-р физ.-мат. наук, проф.; v.malozemov@spbu.ru

Тамасян Григорий Шаликович — канд. физ.-мат. наук, доц.; g.tamasyan@spbu.ru

On the direction of the steepest descent

V. N. Malozemov, G. Sh. Tamasyan

St. Petersburg State University, 7–9, Universitetskaya nab.,
St. Petersburg, 199034, Russian Federation

For citation: Malozemov V. N., Tamasyan G. Sh. On the direction of the steepest descent. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2019, vol. 15, iss. 4, pp. 489–501. https://doi.org/10.21638/11702/spbu10.2019.406 (In Russian)

The article is dedicated of memory of Professor V. F. Demyanov (1938—2014). The main scientific interests of V. F. Demyanov lay in the field of numerical optimization methods, where the notion of the direction of the steepest descent plays an important role. This notion is introduced for both smooth and nonsmooth functions, both in constrained and unconstrained cases. This paper provides a detailed analysis of methods for constructing the direction of steepest descent. In all cases, it comes down to solving the quadratic programming problem. Particular attention is paid to nonsmooth functions, in the study of which V. F. Demyanov made a significant contribution. An example of a function which is quasidifferentiable at a point is given. This function has two directions of the steepest descent and two directions of the steepest ascent.

Keywords: steepest descent direction, nonsmooth analysis, quasidifferential.

References

1. Malozemov V. N. *Linejnaya algebra bez opredelitelej. Kvadratichnaya funkciya* [*Linear algebra without determinants. Quadratic function*]. St. Petersburg, Saint Petersburg University Publ., 1997, 80 p. (In Russian)

2. Malozemov V. N. Metod N'yutona—Rafsona dlya bezuslovnoj minimizacii [Newton—Raphson method for unconditional minimization]. *Seminar "CNSA&NDO". Selected Papers.* Febr. 14, 2019. Available at: http://apmath.spbu.ru/cnsa/reps19.shtml#0214 (accessed: October 13, 2019). (In Russian)

3. Demyanov V. F., Rubinov A. M. O kvazidifferenciruemyh funkcionalah [On quasidifferentiable functionals]. *Papers of Academy of Science the USSR*, 1980, vol. 250, no. 1, pp. 21–25. (In Russian)

4. Demyanov V. F., Rubinov A. V. *Osnovy negladkogo analiza i kvazidifferencial'noe ischislenie* [*Elements of nonsmooth analysis and quasidifferential calculus*]. Moscow, Nauka Publ., 1990, 432 p. (In Russian)

5. Demyanov V. F. *Minimaks: Differentsiruemost' po napravleniyam* [*Minimax: Directional differentiability*]. Leningrad, Leningrad University Publ., 1974, 112 p. (In Russian)

Received: November 01, 2019.
Accepted: November 07, 2019.

Author's information:

Vassili N. Malozemov — Dr. Sci. in Physics and Mathematics, Professor; v.malozemov@spbu.ru

Grigoriy Sh. Tamasyan — PhD in Physics and Mathematics, Associate Professor; g.tamasyan@spbu.ru