О НАПРАВЛЕНИИ НАИСКОРЕЙШЕГО СПУСКА*

В. Н. Малозёмов Г. Ш. Тамасян

v.malozemov@spbu.ru g.tamasyan@spbu.ru

18 апреля 2019 г.

Памяти 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, как при отсутствии ограничений, так и при их наличии. В развитие данной темы В. Ф. Демьянов внёс значительный вклад.

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

$$\frac{\partial f(x_0)}{\partial g} = \langle f'(x_0), g\rangle.$$

*Семинар по конструктивному негладкому анализу и недифференцируемой оптимизации «CNSA & NDO»: `http://www.apmath.spbu.ru/cnsa/`

Вектор $g_0 \in S$, на котором

$$\frac{\partial f(x_0)}{\partial g_0} = \min_{g\in S}\langle f'(x_0), g\rangle, \tag{2}$$

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

$$-\langle f'(x_0), g\rangle = \langle f'(x_0), -g\rangle \leqslant \|f'(x_0)\|, \tag{3}$$

причём 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 (3)

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

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°. Возьмём симметричную положительно определённую матрицу D порядка n. Введём обобщённые скалярное произведение и норму:

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

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

$$\langle f'(x_0), h\rangle \longrightarrow \min \qquad \text{при ограничении} \qquad \langle Dh, h\rangle = 1. \tag{4}$$

Это решение можно получить с помощью обобщённого неравенства Коши—Буняковского

$$\langle x, y\rangle_D \leqslant \|x\|_D\|y\|_D, \tag{5}$$

которое при ненулевых x, y обращается в равенство только тогда, когда $y = \alpha x$ при некотором $\alpha > 0$ (см., например, [1, с. 33]). Согласно (5), при $\|h\|_D = 1$ имеем

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

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 (4) является вектор

$$h_D = -\frac{D^{-1} f'(x_0)}{\sqrt{\langle D^{-1} f'(x_0), f'(x_0)\rangle}}. \tag{6}$$

По этой формуле находится направление D-наискорейшего спуска.

3°. Наряду с задачей (4) рассмотрим задачу минимизации выпуклой квадратичной функции

$$Q(h) := \langle f'(x_0), h\rangle + \tfrac{1}{2}\langle D\,h, h\rangle \longrightarrow \min_{h\in\mathbb{R}^n}. \tag{7}$$

Для неё критерий оптимальности имеет вид $Q'(h) = \mathbf{0}$ или $f'(x_0) + D\,h = \mathbf{0}$. Значит, единственным решением задачи (7) 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 (6) только нормировкой, $h_D = h_0/\|h_0\|_D$.

Получили следующий важный факт: *при* $f'(x_0) \neq \mathbf{0}$ (то есть когда точка x_0 не является стационарной) *решения задач* (4) *и* (7) *различаются лишь нормировкой.*

З а м е ч а н и е. Пусть функция $f(x)$ дважды дифференцируема в точке x_0 и матрица $D = f''(x_0)$ положительно определена. Тогда вектор

$$h_0 = -\,D^{-1}\, f'(x_0) = -\,\big(f''(x_0)\big)^{-1}\, f'(x_0)$$

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:

$$f(x) \longrightarrow \inf, \qquad Ax = b, \tag{8}$$

где A — $(m \times n)$-матрица с линейно независимыми строками. Множество планов задачи (8) 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

$$\begin{aligned} &\langle f'(x_0), h\rangle \longrightarrow \min,\\ &Ah = \mathbf{0}, \quad \|h\| = 1. \end{aligned} \tag{9}$$

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 (8). Она удовлетворяет необходимому условию минимума $f'(x_0) = A^T u$ при $u = (A^T A)^{-1} A f'(x_0)$.

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 (9).

Доказательство. Так как $AP = \mathbb{O}$, то h_0 — план задачи (9). Обозначим $c = f'(x_0)$ и возьмём план h задачи (9), 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, $Ph = h$ 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 (9) рассмотрим задачу минимизации выпуклой квадратичной функции

$$Q(\lambda) := \tfrac{1}{2} \left\| A^T\lambda - f'(x_0) \right\|^2 \longrightarrow \min_{\lambda \in \mathbb{R}^m} . \tag{10}$$

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 (10) 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

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

Таким образом, решение задачи (10) 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 — конченое индексное множество. Для неё 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 точки x_0 по любому направлению $g \in S$ существует. При этом*

$$\frac{\partial\varphi(x_0)}{\partial g} = \max_{i \in M(x_0)} \langle f_i'(x_0), g \rangle. \tag{11}$$

Доказательство этой теоремы опирается на следующий более общий результат.

ТЕОРЕМА 3. *В условиях теоремы 2 справедливо разложение*

$$\varphi(x_0 + h) = \varphi(x_0) + \max_{i \in M(x_0)} \langle f_i'(x_0), h \rangle + o(\|h\|), \tag{12}$$

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

$$\varphi(x_0 + h) = \max_{i \in M(x_0)} f_i(x_0 + h). \tag{13}$$

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

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

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 (13) и отражает этот факт.

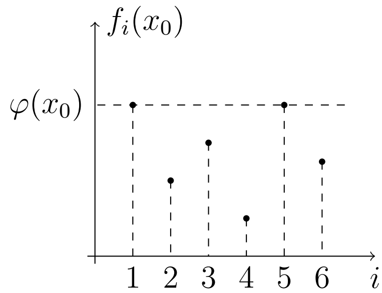

Рис. 1. График $f_i(x_0)$ как функции от i

Согласно (13) 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 $\varepsilon > 0$ найдётся 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 (11) 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:

$$\chi_0(g) := \max_{v \in C(x_0)} \langle v, g \rangle \longrightarrow \min_{g \in S}. \tag{14}$$

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 $\mathbf{0} \in C(x_0)$ (см. рис. 2).

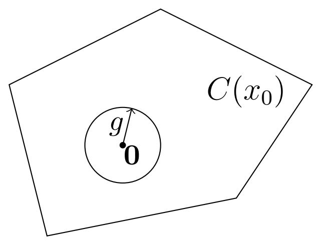

Рис. 2. Пояснение к условию стационарности

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.*

Доказательство. 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 перейдём к пределу при $t \longrightarrow +0$. Получим

$$\langle v, v_*\rangle \geqslant \langle v_*, v_*\rangle \qquad \forall\, v \in C(x_0). \tag{15}$$

Теперь легко доказать требуемое неравенство

$$\chi_0(g_0) \leqslant \chi_0(g) \qquad \forall\, g \in S. \tag{16}$$

Действительно, согласно (15) и определению вектора g_0 имеем

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

причём при $v = v_*$ неравенство выполняется как равенств. Значит,

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

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

$$\langle v_*, -g\rangle \leqslant \|v_*\| \tag{18}$$

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

На основании (17) и (18) при всех $g \in S$ получаем

$$\chi_0(g_0) = -\|v_*\| \leqslant \langle v_*, g\rangle \leqslant \max_{v \in C(x_0)} \langle v, g\rangle = \chi_0(g), \tag{19}$$

что соответствует (16).

Установлено, что вектор g_0 является решением задачи (14). Так как неравенство (18), используемое в (19), при $g \neq g_0$ выполняется как строгое, то g_0 — единственное решение задачи (14).

Теорема доказана. $\square$

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:

$$\chi_0(g) := \max_{v\in C(x_0)} \langle v, g\rangle \longrightarrow \min, \qquad Ag = \mathbf{0}, \quad \|g\| = 1. \tag{20}$$

Введём множества $H = \{h = A^T\lambda \mid \lambda \in \mathbb{R}^m\}$ 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* (20).

Доказательство. 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

$$\langle p_0, h\rangle = 0 \qquad \forall\, h \in H. \tag{21}$$

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 (21).

Учитывая, что $h = A^T\lambda$, на основании (21) 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 (20).

Вычислим $\chi_0(g_0)$. Возьмём вектор $v \in C(x_0)$. Тогда $-v \in P(x_0)$. По неравенству, аналогичному (15), $\langle p, p_0\rangle \geqslant \langle p_0, p_0\rangle$ 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 (21),

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

Теперь при всех $v \in C(x_0)$ имеем

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

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

$$\chi_0(g_0) := \max_{v \in C(x_0)} \langle v, g_0 \rangle = -\|p_0\|. \tag{22}$$

Осталось проверить оптимальность g_0. Возьмём план g задачи (20), 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 $\alpha > 0$, то есть когда $g = g_0$. По условию $g \neq g_0$, поэтому $-\|p_0\| < -\langle p_0, g \rangle$.

С учётом (22) запишем

$$\chi_0(g_0) = -\|p_0\| < -\langle p_0, g \rangle = \langle v_0, g \rangle - \langle h_0, g \rangle. \tag{23}$$

Пусть $h_0 = A^T \lambda_0$. Вектор g как план задачи (20) удовлетворяет условию $Ag = \mathbf{0}$. Поэтому

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

Теперь из (23) следует, 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 её приращение допускает представление

$$f(x_0 + h) - f(x_0) = \max_{v \in \underline{\partial} f(x_0)} \langle v, h \rangle + \min_{w \in \overline{\partial} f(x_0)} \langle w, h \rangle + o(\|h\|), \tag{24}$$

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

Положительно однородный по h функционал

$$\ell(x_0, h) = \max_{v \in \underline{\partial} f(x_0)} \langle v, h \rangle + \min_{w \in \overline{\partial} f(x_0)} \langle w, h \rangle \tag{25}$$

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* множеством соответственно.

Из (24) следует, что у квазидифференцируемой в точке x_0 функции $f(x)$ существуют производные по всем направлениям и что

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

Направлением наискорейшего спуска является решение экстремальной задачи

$$\chi_0(g) := \ell(x_0, g) \longrightarrow \min_{g \in S}. \tag{26}$$

В силу непрерывности $\ell(x_0, h)$ по h решение задачи (26) существует.

Преобразуем выражение для $\ell(x_0, g)$. Согласно (25) имеем

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

Точка x_0, в которой производные по всем направлениям неотрицательны,

$$\ell(x_0, g) \geqslant 0 \qquad \forall g \in S, \tag{27}$$

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

ТЕОРЕМА 6 (Л. Н. Полякова)**.** *Условие* inf-*стационарности* (27) *равносильно вложению*

$$-\overline{\partial} f(x_0) \subset \underline{\partial} f(x_0). \tag{28}$$

Доказательство. Зафиксируем $w \in \overline{\partial} f(x_0)$. Согласно (27) 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 (28). $\square$

Отметим также, что условие (28) эквивалентно соотношению

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

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

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

Покажем, что

$$\min_{g\in S}\chi_0(g)=-a. \tag{29}$$

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

$$\min_{g\in S}\chi_0(g)=\min_{g\in S}\ell(x_0,g)\geqslant-a. \tag{30}$$

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

$$\min_{v\in\underline{\partial}f(x_0)}\|v+w_0\|=a. \tag{31}$$

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

$$\mathbf{0}\notin\underline{\partial}f(x_0)+w_0. \tag{32}$$

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 (32) справедливо равенство

$$\min_{g\in S}\ \max_{v\in\underline{\partial}f(x_0)}\langle v+w_0,g\rangle=-\min_{v\in\underline{\partial}f(x_0)}\|v+w_0\|=-a \tag{33}$$

(мы воспользовались формулой (31)). Значит,

$$\min_{g\in S}\chi_0(g)\leqslant-a. \tag{34}$$

Объединив неравенства (30) и (34), придём к равенству (29).

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 (33) минимум по $g \in S$ достигается на векторе

$$g_0 = -\frac{v_0 + w_0}{\|v_0 + w_0\|}. \tag{35}$$

Таким образом,

$$\max_{v \in \underline{\partial} f(x_0)} \langle v + w_0, g_0 \rangle = -a. \tag{36}$$

Покажем, что g_0 — направление наискорейшего спуска, то есть, что

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

Согласно (36) имеем

$$\begin{aligned} \chi_0(g_0) = \min_{w \in \overline{\partial} f(x_0)} \max_{v \in \underline{\partial} f(x_0)} \langle v + w, g_0 \rangle \leqslant \\ \leqslant \max_{v \in \underline{\partial} f(x_0)} \langle v + w_0, g_0 \rangle = -a. \end{aligned} \tag{37}$$

На основании (29) и (37) заключаем, что $\chi_0(g_0) = -a$.

Подведём 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* (35)*, в которой w_0 — решение экстремальной задачи*

$$\min_{v \in \underline{\partial} f(x_0)} \|v + w\| \longrightarrow \max_{w \in \overline{\partial} f(x_0)} \tag{38}$$

и v_0 — решение экстремальной задачи

$$\|v + w_0\| \longrightarrow \min_{v \in \underline{\partial} f(x_0)} . \tag{39}$$

11°. Приведём характерный пример квазидифференцируемой функции двух переменных.

ПРИМЕР. Рассмотрим функцию

$$f(x) = |x_1| - |x_2|. \tag{40}$$

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

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

Так как

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

то

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

Получили, что функция $f(x)$ вида (40) квазидифференцируема в точке $x_0 = \mathbf{0}$. При этом (см. рис. 3)

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

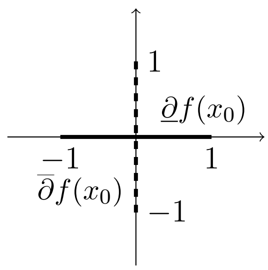

Рис. 3. Суб- и супердифференциальные множества

Условие (28) в данном случае не выполняется. Значит, точка $x_0 = \mathbf{0}$ не является inf-стационарной.

Задача (38) имеет два решения $w_0 = -e_2$ и $w_0 = e_2$. Решением задачи (39) 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 (40) существуют два направления наискорейшего спуска $g_0 = e_2$ и $g_0 = -e_2$.

З а м е ч а н и е. Из геометрических соображений очевидно, что у функции $f(x)$ вида (40) в точке $x_0 = \mathbf{0}$ существуют два направления наискорейшего подъёма $g_1 = e_1$ и $g_1 = -e_1$.

Как известно, в гладком случае направления наискорейшего спуска и наискорейшего подъёма 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. Малозёмов В. Н. *Линейная алгебра без определителей. Квадратичная функция.* СПб.: Изд-во СПбГУ, 1997. 80 с.

2. Малозёмов В. Н. *Метод Ньютона—Рафсона для безусловной минимизации* // Семинар “CNSA&NDO”. Избранные доклады. 14 февраля 2019 г. (`http://apmath.spbu.ru/cnsa/reps19.shtml#0214`)

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 ЛГУ, 1974. 112 с.