МЕТОД КОДИФФЕРЕНЦИАЛЬНОГО СПУСКА ДЛЯ КУСОЧНО-АФФИННЫХ ФУНКЦИЙ*

М. В. Долгополик

maxim.dolgopolik@gmail.com

16 мая 2019 г.

Аннотация. В докладе обсуждается модификация метода кодифференциального спуска для минимизации кусочно-аффинных функций. Доказано, что модифицированный метод сходится к точке глобального минимума невыпуклой кусочно-аффинной функции за конечное число шагов. Более подробное изложение результатов данного доклада имеется в статье [1].

1°. Кусочно-аффинные функции. Интуитивно понятно, что кусочно-аффинной называется функция $f\colon \mathbb{R}^d \to \mathbb{R}$, получаемая с помощью склейки конечного числа аффинных «кусочков», то есть аффинных функций вида $\ell(x) = a + \langle v, x \rangle$, $x \in \mathbb{R}^d$, заданных на непересекающихся выпуклых многогранных множествах. Данное интуитивное определение можно формализовать следующим образом (см. [2, 3]).

Напомним, что подмножество Q пространства $\mathbb{R}^d$ называется *полиэдральным*, если оно представимо в виде пересечения конечного числа замкнутых полупространств. *Полиэдральным разбиением* пространства $\mathbb{R}^d$ называется конечный набор $\sigma = \{Q_1, \ldots, Q_n\}$ полиэдральных подмножеств Q_i пространства $\mathbb{R}^d$, удовлетворяющий следующим условиям:

1) каждое из множеств Q_i имеет непустую внутренность,

2) внутренности множеств Q_i попарно не пересекаются ($\operatorname{int} Q_i \cap \operatorname{int} Q_j = \varnothing$, если $i \neq j$),

3) σ является разбиением пространства $\mathbb{R}^d$, т. е. $\mathbb{R}^d = \bigcup_{k=1}^n Q_i$.

Наконец, функция $f\colon \mathbb{R}^d \to \mathbb{R}$ называется *кусочно-аффинной*, если существует такое полиэдральное разбиение $\sigma = \{Q_1, \ldots, Q_n\}$ пространства $\mathbb{R}^d$, что на

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

каждом множестве Q_i функция f является аффинной, то есть для любого i найдутся $a_i \in \mathbb{R}$ и $v_i \in \mathbb{R}^d$ такие, что $f(x) = a_i + \langle v_i, x \rangle$ для всех $x \in Q_i$.

На первый взгляд может показаться, что кусочно-аффинная функция может быть разрывной, и поэтому необходимо дополнительно требовать непрерывности функции f. Однако, непрерывность кусочно-аффинной функции вытекает непосредственно из определения. Покажем это в одномерном случае.

Как нетрудно заметить, при $d = 1$ существует лишь три типа нетривиальных полиэдральных множеств: отрезки $[\alpha, \beta]$ (случай $\alpha = \beta$ не исключается) и бесконечные полуинтервалы вида $(-\infty, \alpha]$ и $[\alpha, +\infty)$. Поэтому любое полиэдральное разбиение вещественной прямой однозначно определяется конечным набором точек $\alpha_1 < \alpha_2 < \ldots < \alpha_{n-1}$ и имеет вид:

$$\sigma = \Big\{(-\infty, \alpha_1], [\alpha_1, \alpha_2], \ldots, [\alpha_{n-2}, \alpha_{n-1}], [\alpha_{n-1}, +\infty)\Big\}.$$

Неравенства $\alpha_1 < \alpha_2 < \ldots < \alpha_{n-1}$ вытекают из предположения о непустоте и попарном непересечении внутренностей множеств $Q_i \in \sigma$. Таким образом, функция $f \colon \mathbb{R} \to \mathbb{R}$ является кусочно-аффинной тогда и только тогда, когда найдутся конечные наборы точек $\alpha_1 < \alpha_2 < \ldots < \alpha_{n-1}$ и чисел $a_i \in \mathbb{R}$, $v_i \in \mathbb{R}$ такие, что

$$f(x) = \begin{cases} a_1 + v_1 x, & \text{если } x \in (-\infty, \alpha_1], \\ a_i + v_i x, & \text{если } x \in [\alpha_{i-1}, \alpha_i],\ i \in \{2, \ldots, n-1\}, \\ a_n + v_n x, & \text{если } x \in [\alpha_{n-1}, +\infty). \end{cases}$$

То есть функция f является кусочно-аффинной тогда и только тогда, когда она является ломаной. Теперь понятно, что непрерывность кусочно-аффинной функции f вытекает непосредственно из определения, поскольку по определению будет $f(\alpha_i) = a_i + v_i \alpha_i = a_{i+1} + v_{i+1} \alpha_i$ для всех i. Схожим образом можно доказать непрерывность кусочно-аффинной функции в общем случае.

При исследовании экстремальных задач удобнее иметь дело с аналитическим представлением кусочно-аффинной функции, чем с её определением через полиэдральное разбиение. В статье [3] (см. также [4]) было доказано, что любая кусочно-аффинная функция допускает аналитическое представление специального вида.

ТЕОРЕМА 1 (Гороховик-Зорько)**.** *Функция $f \colon \mathbb{R}^d \to \mathbb{R}$ является кусочно-аффинной тогда и только тогда, когда найдутся конечные наборы $(a_i, v_i) \in \mathbb{R} \times \mathbb{R}^d$, $i \in I = \{1, \ldots, \ell\}$, и $(b_j, w_j) \in \mathbb{R} \times \mathbb{R}^d$, $j \in J = \{1, \ldots, s\}$, такие, что*

$$f(x) = \max_{i \in I}(a_i + \langle v_i, x \rangle) + \min_{j \in J}(b_j + \langle w_j, x \rangle) \quad \forall x \in \mathbb{R}^d. \tag{1}$$

Заметим также, что множество кусочно-аффинных функций является линейным пространством, содержащим все аффинные функции и замкнутым относительно операций взятия поточечного максимума и минимума конечных семейств функций (см. [3, 4]). Так, например, функция

$$f(x_1, x_2) = \min\big\{ \max\{|x_1|, |x_2|\}, 1 + \max\{2|x_1 - 2|, |x_2 - 2|\}\big\}$$

является кусочно-аффинной.

Методы построения некоторого аналитического представления кусочно-аффинной функции по заданному полиэдральному разбиению изучались в работе [2]. Методы вычисления представления (1) по произвольному аналитическому представлению кусочно-аффинной функции исследовались в [5]. Одномерный случай подробно изучен в докладе [6].

2°.Глобальные кодифференциалы кусочно-аффинных функций. Пусть задана кусочно-аффинная функция $f\colon \mathbb{R}^d \to \mathbb{R}$, и известно её представление вида (1). Положим

$$\underline{f}(x) = \max_{i\in I}(a_i + \langle v_i, x\rangle), \quad \overline{f}(x) = \min_{j\in J}(b_j + \langle w_j, x\rangle). \tag{2}$$

Тогда $f = \underline{f} - (-\overline{f})$ является DC разложением[1] функции f, то есть представлением функции f в виде разности выпуклых функций. Введём многозначные отображения

$$\begin{aligned} \underline{d}f(x) &= \operatorname{co}\big\{(a_i - \underline{f}(x) + \langle v_i, x\rangle, v_i) \in \mathbb{R}^{d+1} \;\big|\; i \in I\big\},\\ \overline{d}f(x) &= \operatorname{co}\big\{(b_j - \overline{f}(x) + \langle w_j, x\rangle, w_j) \in \mathbb{R}^{d+1} \;\big|\; j \in J\big\}. \end{aligned} \tag{3}$$

По определению, для любых $x, \Delta x \in \mathbb{R}^d$ будет

$$f(x + \Delta x) - f(x) = \underline{f}(x + \Delta x) + \overline{f}(x + \Delta x) - (\underline{f}(x) + \overline{f}(x))$$

Воспользовавшись равенствами (2), получим

$$\begin{aligned} &f(x + \Delta x) - f(x) =\\ &\quad = \max_{i\in I}(a_i - \underline{f}(x) + \langle v_i, x\rangle + \langle v_i, \Delta x\rangle) + \min_{j\in J}(b_j - \overline{f}(x) + \langle w_j, x\rangle + \langle w_j, \Delta x\rangle), \end{aligned}$$

откуда с учётом (3) следует, что

$$f(x + \Delta x) - f(x) = \max_{(a,v)\in \underline{d}f(x)}(a + \langle v, \Delta x\rangle) + \min_{(b,w)\in \overline{d}f(x)}(b + \langle w, \Delta x\rangle). \tag{4}$$

[1] от англ. «Difference of Convex», т. е. «разность выпуклых»

Ясно, что множества $\underline{d}f(x)$ и $\overline{d}f(x)$ выпуклы и компактны. Более того, для любого $x \in \mathbb{R}^d$ справедливы равенства

$$\max_{(a,v)\in\underline{d}f(x)} a = \max_{i\in I}(a_i + \langle v_i, x\rangle) - \underline{f}(x) = \underline{f}(x) - \underline{f}(x) = 0, \tag{5}$$

и аналогично $\min\limits_{(b,w)\in\overline{d}f(x)} b = 0$. Таким образом, пара $Df(x) = [\underline{d}f(x), \overline{d}f(x)]$ является кодифференциалом функции f в точке x (см. [7, 8]). С другой стороны, кодифференциал определяется через локальную аппроксимацию исследуемой функции, в то время как равенство (4) справедливо для всех x, $\Delta x \in \mathbb{R}^d$, то есть глобально.

ОПРЕДЕЛЕНИЕ 1. Пара $Df = [\underline{d}f, \overline{d}f]$, определяемая равенствами (3), называется *глобальным кодифференциалом* кусочно-аффинной функции f, соответствующим DC разложению $f = \underline{f} - (-\overline{f})$). Многозначное отображение $\underline{d}f$ называется *глобальным гиподифференциалом* функции f, в то время как многозначное отображение $\overline{d}f$ называется *глобальным гипердифференциалом* функции f.

Глобальный кодифференциал кусочно-аффинной функции определяется неоднозначно, поскольку существует бесконечно много различных DC разложений кусочно-аффинных функций вида (1). Отметим также, что глобальный кодифференциал кусочно-аффинной функции впервые был неявно введён проф. Л.Н. Поляковой в [9].

Для глобальных кодифференциалов справедливо исчисление, совпадающее с обычным кодифференциальным исчислением [7, 8].

ПРЕДЛОЖЕНИЕ 1. *Пусть $f_m \colon \mathbb{R}^d \to \mathbb{R}$, $m \in M = \{1, \ldots, \ell\}$, — кусочно-аффинные функции, а Df_m — их глобальные кодифференциалы. Справедливы следующие утверждения:*

1) *если $f(x) = a + \langle v, x\rangle$, то многозначные отображения*

$$Df(\cdot) \equiv [\{(0, v)\}, \{(0, 0)\}], \quad Df(\cdot) \equiv [\{(0, 0)\}, \{(0, v)\}]$$

являются глобальными кодифференциалами функции f;

2) *если $f = f_1 + c$ для некоторого $c \in \mathbb{R}$, то $Df = Df_1$;*

3) *если $f = \lambda f_1$, то $Df = [\lambda\underline{d}f_1, \lambda\overline{d}f_1]$ в случае $\lambda \geqslant 0$, и $Df = [\lambda\overline{d}f_1, \lambda\underline{d}f_1]$ в случае $\lambda < 0$;*

4) *если $f = \sum_{m=1}^{\ell} f_m$, то $Df = [\sum_{m=1}^{\ell} \underline{d}f_m, \sum_{m=1}^{\ell} \overline{d}f_m]$;*

5) если $f = \max_{m \in M} f_m$, *то*

$$Df(\cdot) = \left[\mathrm{co} \left\{ (f_m(\cdot) - f(\cdot), 0) + \underline{d} f_m(\cdot) - \sum_{k \neq m} \overline{d} f_i(\cdot) \mid m \in M \right\}, \sum_{m=1}^{\ell} \overline{d} f_m(\cdot) \right];$$

6) если $f = \min_{m \in M} f_m$, *то*

$$Df(\cdot) = \left[\sum_{m=1}^{\ell} \underline{d} f_m(\cdot), \mathrm{co} \left\{ (f_m(\cdot) - f(\cdot), 0) + \overline{d} f_m(\cdot) - \sum_{k \neq m} \underline{d} f_i(\cdot) \mid m \in M \right\} \right].$$

Очень громоздкое, но содержательно простое доказательство этого утверждения, основанное на применении формул (3), имеется в [1].

3°. Условия глобальной оптимальности. В терминах глобальных кодифференциалов возможно выписать необходимые и достаточные условия глобальной оптимальности для кусочно-аффинных функций. Впервые условия такого рода были получены в [9]. Здесь мы приведём другие условия глобальной оптимальности из работы [1], тесно связанные с методом кодифференциального спуска [7, 8].

Нам потребуется следующий вспомогательный результат о необходимых и достаточных условиях неотрицательности конечной полиэдральной функции.

ЛЕММА 1. *Пусть функция* $f\colon \mathbb{R}^d \to \mathbb{R}$ *имеет вид*

$$f(x) = \max_{i \in I} (a_i + \langle v_i, x \rangle) \quad \forall x \in \mathbb{R}^d$$

для некоторых $(a_i, v_i) \in \mathbb{R} \times \mathbb{R}^d$, *где* $I = \{1, \ldots, \ell\}$. *Тогда* $f(x) \geqslant 0$ *для всех* $x \in \mathbb{R}^d$ *тогда и только тогда, когда выполняется одно из двух следующих условий:*

1) $0 \in C = \mathrm{co}\{(a_i, v_i) \mid i \in I\}$

2) функция f *ограничена снизу и* $a^* > 0$, *где*

$$\{(a^*, v^*)\} = \arg\min \left\{ \|(a, v)\|^2 \mid (a, v) \in C \right\} \tag{6}$$

и $\|\cdot\|$ *— евклидова норма.*

Д о к а з а т е л ь с т в о. Пусть функция f неотрицательна. Предположим, что $0 \notin C$ и $a^* \leqslant 0$. Из определения (a^*, v^*) и необходимых и достаточный условий минимума выпуклой функции на выпуклом множестве (см., например, [10, Предложение II.2.1]) следует, что

$$a^*(a - a^*) + \langle v^*, v - v^* \rangle \geqslant 0 \quad \forall (a, v) \in C. \tag{7}$$

Если $a^* = 0$, то $v^* \neq 0$ (так как иначе $0 \in C$) и $\langle v, -v^* \rangle \leqslant -\|v^*\|^2$ для всех $(a, v) \in C$. Отсюда для всех $\alpha \geqslant 0$ имеем

$$f(-\alpha v^*) = \max_{i \in I}(a_i + \langle v_i, -\alpha v^* \rangle) \leqslant \max_{i \in I}(a_i - \alpha\|v^*\|^2) = f(0) - \alpha\|v^*\|^2, \tag{8}$$

что противоречит предположению о неотрицательности функции f.

Если же $a^* < 0$, то, поделив (7) на a^*, получим

$$a + \left\langle v, \frac{1}{a^*} v^* \right\rangle \leqslant -\frac{1}{|a^*|}\|(a^*, v^*)\|^2 < 0 \quad \forall (a, v) \in C.$$

Беря максимум по всем $(a, v) \in C$, приходим к неравенству $f([a^*]^{-1} v^*) < 0$, которое вновь противоречит неотрицательности функции f.

Предположим теперь, что выполнено одно из двух условий леммы. Если $0 \in C$, то для любого $x \in \mathbb{R}^d$ будет

$$f(x) = \max_{(a,v) \in C}(a + \langle v, x \rangle) \geqslant 0 + \langle 0, x \rangle = 0,$$

т.е. функция f неотрицательна. Предположим теперь, что $0 \notin C$, функция f ограничена снизу и $a^* > 0$, но найдётся такое $x \in \mathbb{R}^d$, что $f(x) < 0$.

Положим $f^* = \inf_{x \in \mathbb{R}^d} f(x)$. По нашему предположению $-\infty < f^* < 0$. Покажем, что $(f^*, 0) \in C$. Тогда $(1 - \alpha)(f^*, 0) + \alpha(a^*, v^*) \in C$ для любого $\alpha \in [0, 1]$. Полагая $\alpha = |f^*|/(|f^*| + a^*) \in (0, 1)$, получим $(0, \alpha v^*) \in C$, что противоречит определению (a^*, v^*) (см. (6)), поскольку $\|(0, \alpha v^*)\|^2 < \|(a^*, v^*)\|^2$. Значит, функция f неотрицательна.

Для любого $\varepsilon > 0$ существует x_ε такое, что $f(x_\varepsilon) < f^* + \varepsilon$. Из определения ε-субдифференциала выпуклой функции следует, что $0 \in \partial_\varepsilon f(x_\varepsilon)$, где $\partial_\varepsilon f(x_\varepsilon)$ — это ε-субдифференциал функции f в точке x_ε. Согласно [11, Пример XI.3.5.3] справедливо равенство

$$\partial_\varepsilon f(x) = \big\{ v \in \mathbb{R}^d \;\big|\; \exists a \in \mathbb{R} \colon (a, v) \in C,\ a + \langle v, x \rangle \geqslant f(x) - \varepsilon \big\}.$$

Следовательно, для всех $\varepsilon > 0$ существует $a_\varepsilon \geqslant f(x_\varepsilon) - \varepsilon \geqslant f^* - \varepsilon$ такое, что $(a_\varepsilon, 0) \in C$. Заметим, что для любого $(a, 0) \in C$ будет $f(x) \geqslant a$ для всех $x \in \mathbb{R}^d$, откуда вытекает, что $f^* \geqslant a$. Таким образом, $f^* \geqslant a_\varepsilon \geqslant f^* - \varepsilon$. Переходя к пределу при ε стремящемся к нулю, получаем, что $(f^*, 0) \in C$, поскольку множество C замкнуто. $\square$

СЛЕДСТВИЕ 1. *Пусть выполнены все предположения леммы 1, и f ограничена снизу. Тогда $f(x) \geqslant 0$ для всех $x \in \mathbb{R}^d$ тогда и только тогда, когда $a^* \geqslant 0$.*

Доказательство. Если функция f неотрицательна, то по лемме 1 либо $a^* > 0$, либо $0 \in C$. В случае $0 \in C$, по определению будет $(a^*, v^*) = (0, 0)$, т.е. $a^* = 0$.

Докажем обратное утверждение. Предположим, что $a^* \geqslant 0$. Если $a^* > 0$, то f неотрицательна по Лемме 1. Пусть теперь $a^* = 0$. Если $v^* = 0$, то $0 \in C$ и, вновь воспользовавшись леммой 1, получаем, что f неотрицательна. С другой стороны, если $v^* \neq 0$, то, рассуждая как и при доказательстве леммы (см. (8)), получим, что f не ограничена снизу, что противоречит нашему предположению. □

С помощью предыдущего следствия нетрудно получить необходимые и достаточные условия глобальной оптимальности для кусочно-аффинной функции.

ТЕОРЕМА 2. *Пусть f — ограниченная снизу кусочно-аффинная функция вида (1), а Df — её глобальный кодифференциал. Для произвольной точки $x^* \in \mathbb{R}^d$ и всех $j \in J$ положим $z_j = (b_j - \overline{f}(x^*) + \langle w_j, x^* \rangle, w_j) \in \overline{d}f(x^*)$ и*

$$\big\{(a_j^*, v_j^*)\big\} = \arg\min\big\{\|(a, v)\|^2 \;\big|\; (a, v) \in \underline{d}f(x^*) + z_j\big\}.$$

Для того чтобы точка x^* была точкой глобального минимума функции f необходимо и достаточно, чтобы для любого $j \in J$ выполнялось неравенство $a_j^* \geqslant 0$.

Доказательство. Полагая $x = x^*$ и $\Delta x = x - x^*$ в равенстве (4), получим

$$\begin{aligned} f(x) - f(x^*) &= \\ = \max_{(a,v)\in\underline{d}f(x^*)} &(a + \langle v, x - x^* \rangle) + \min_{j\in J}(b_j - \overline{f}(x^*) + \langle w_j, x^* \rangle + \langle w_j, x - x^* \rangle) = \\ &= \min_{j\in J} \max_{(a,v)\in\underline{d}f(x^*)} \Big(a + b_j - \overline{f}(x^*) + \langle w_j, x^* \rangle + \langle v + w_j, x - x^* \rangle\Big) \end{aligned}$$

для всех $x \in \mathbb{R}^d$, откуда следует, что

$$f(x) - f(x^*) = \min_{j\in J} \max_{(a,v)\in\underline{d}f(x^*)+z_j} (a + \langle v, x - x^* \rangle) = \min_{j\in J} g_j(x - x^*), \tag{9}$$

где $g_j(x) = \max\limits_{(a,v)\in\underline{d}f(x^*)+z_j} (a + \langle v, x \rangle)$. Следовательно, точка x^* будет точкой глобального минимума функции f тогда и только тогда, когда для всех $j \in J$

полиэдральные функции g_j неотрицательны. Заметим, что функции g_j ограничены снизу, поскольку $g_j(x) \geqslant f(x) - f(x^*)$ для всех $x \in \mathbb{R}^d$ (см. (9)), и f ограничена снизу. Применяя следствие 1 к функциям g_j, получаем, что x^* является точкой глобального минимума функции f тогда и только тогда, когда $a_j^* \geqslant 0$ для всех $j \in J$. $\square$

4°. **Модифицированный метод кодифференциального спуска.** С помощью необходимых и достаточных условий глобальной оптимальности из теоремы 2 можно предложить модификацию метода кодифференциального спуска [8] для минимизации кусочно-аффинных функций. Для этого зафиксируем произвольную кусочно-аффинную функцию f вида (1), и обозначим через Df её глобальный кодифференциал. Для всех $x \in \mathbb{R}^d$ и $j \in J$ положим

$$z_j(x) = (b_j - \overline{f}(x) + \langle w_j, x \rangle, w_j) \in \overline{d}f(x), \tag{10}$$

$$\big\{(a_j(x), v_j(x))\big\} = \arg\min\big\{\|(a,v)\|^2 \;\big|\; (a,v) \in \underline{d}f(x) + z_j(x)\big\}. \tag{11}$$

Предположим, что точка x не является точкой глобального минимума функции f, и зафиксируем произвольное $j \in J$. Из определения $(a_j^*(x), v_j(x))$ и необходимого и достаточного условия минимума выпуклой функции на выпуклом множестве следует, что

$$a_j(x)(a - a_j(x)) + \langle v_j(x), v - v_j(x) \rangle \geqslant 0 \quad \forall (a,v) \in \underline{d}f(x) + z_j(x).$$

Если $a_j(x) < 0$, то, поделив это неравенство на $a_j(x)$ и взяв максимум по всем $(a,v) \in \underline{d}f(x) + z_j(x)$, получим (см. (9))

$$f\left(x + \frac{1}{a_j(x)} v_j(x)\right) - f(x) \leqslant -\frac{1}{|a_j(x)|}\|(a_j(x), v_j(x))\|^2 < 0. \tag{12}$$

Если $a_j(x) = 0$, но $v_j(x) \neq 0$, то $\langle v, -v_j(x) \rangle \leqslant -\|v_j(x)\|^2$ для всех $(a,v) \in \underline{d}f(x) + z_j(x)$. Отсюда и из (9) следует, что

$$f(x - \alpha v_j(x)) - f(x) \leqslant \max_{(a,v) \in \underline{d}f(x) + z_j(x)} a - \alpha \|v_j(x)\|^2 \quad \forall \alpha \geqslant 0, \tag{13}$$

т.е. функция f не ограничена снизу. Таким образом, если $a_j(x) = 0$ и f ограничена снизу, то $v_j(x) = 0$.

Наконец, если $a_j(x) > 0$, то индекс j можно отбросить. Действительно, по лемме 1 будет

$$\max_{(a,v) \in \underline{d}f(x) + z_j(x)} (a + \langle v, y \rangle) \geqslant 0 \quad \forall y \in \mathbb{R}^d. \tag{14}$$

Согласно (9) имеем

$$f(y) - f(x) = \min_{k \in J} \max_{(a,v) \in \underline{d}f(x) + z_k(x)} (a + \langle v, y - x \rangle) \quad \forall y \in \mathbb{R}^d.$$

Из неравенства (14) следует, что для любого y такого, что $f(y) < f(x)$ минимум в данном равенстве не может достигаться при $k = j$. Поэтому

$$f(y) - f(x) = \min_{k \in J \setminus \{j\}} \max_{(a,v) \in \underline{d}f(x) + z_i(x)} (a + \langle v, y - x \rangle)$$

для всех $y \in \mathbb{R}$, удовлетворяющих неравенству $f(y) < f(x)$. Другими словами, индекс j и соответствующий ему вектор (b_j, w_j) не нужны для вычисления $f(y)$, если $f(y) < f(x)$. Отсюда, в частности, следует, что для любого y, удовлетворяющего неравенству $f(y) < f(x)$, также будет $a_j(y) \geqslant 0$. Докажем этот результат строго.

ЛЕММА 2. *Предположим, что функция f ограничена снизу, и для некоторых $j \in J$ и $x \in \mathbb{R}^d$ выполняется неравенство $a_j(x) \geqslant 0$. Тогда $a_j(y) \geqslant 0$ для всех $y \in \mathbb{R}^d$ удовлетворяющих неравенству $f(y) \leqslant f(x)$.*

Доказательство. Для любых $\Delta y, y \in \mathbb{R}^d$ положим

$$g_j(\Delta y, y) = \max_{(a,v) \in \underline{d}f(y) + z_j(y)} (a + \langle v, \Delta y \rangle).$$

Воспользовавшись равенствами (4) и (3), получим

$$\begin{aligned} f(y + \Delta y) - f(y) &= \max_{(a,v) \in \underline{d}f(y)} (a + \langle v, \Delta y \rangle) + \min_{(b,w) \in \overline{d}f(y)} (b + \langle w, \Delta y \rangle) \leqslant \\ &\leqslant \max_{(a,v) \in \underline{d}f(y)} (a + \langle v, \Delta y \rangle) + b_j - \overline{f}(y) + \langle w_j, y \rangle + \langle w_j, \Delta y \rangle = g_j(\Delta y, y). \end{aligned}$$

для всех $\Delta y, y \in \mathbb{R}^d$. Следовательно, $g_j(\cdot, y) \geqslant \inf_{z \in \mathbb{R}^d} f(z) - f(y)$ для любого $y \in \mathbb{R}^d$, то есть полиэдральная функция $g_j(\cdot, y)$ ограничена снизу. Заметим также, что по определению $\underline{d}f(\cdot)$ и $z_j(\cdot)$ (см. (3) и (10)) будет

$$g_j(\Delta y, y) = \max_{i \in I} \big(a_i + \langle v_i, y \rangle - \underline{f}(y) + \langle v_i, \Delta y \rangle + b_j + \langle w_j, y \rangle - \overline{f}(y) + \langle w_j, \Delta y \rangle\big) \quad (15)$$

для всех $\Delta y, y \in \mathbb{R}^d$.

Из неравенства $a_j(x) \geqslant 0$ и следствия 1 вытекает, что функция $g(\cdot, x)$ неотрицательна. Поэтому, согласно равенству (15) для любого $\Delta x \in \mathbb{R}^d$ существует $i \in I$ такое, что

$$a_i + \langle v_i, x \rangle - \underline{f}(x) + \langle v_i, \Delta x \rangle + b_j + \langle w_j, x \rangle - \overline{f}(x) + \langle w_j, \Delta x \rangle \geqslant 0.$$

Полагая $\Delta x = y - x + \Delta y$ и пользуясь равенством $f(x) = \underline{f}(x) + \overline{f}(x)$ (см. (1) и (2)), получаем, что для всех $\Delta y, y \in \mathbb{R}^d$ найдётся такой индекс $i \in I$, что

$$a_i + \langle v_i, y \rangle + \langle v_i, \Delta y \rangle + b_j + \langle w_j, y \rangle + \langle w_j, \Delta y \rangle \geqslant f(x).$$

Вычитая $f(y) = \underline{f}(y) + \overline{f}(y)$, приходим к неравенству

$$a_i + \langle v_i, y\rangle - \underline{f}(y) + \langle v_i, \Delta y\rangle + b_j + \langle w_j, y\rangle - \overline{f}(y) + \langle w_j, \Delta y\rangle \geqslant f(x) - f(y).$$

Беря максимум по всем $i \in I$, получаем, что $g_j(\Delta y, y) \geqslant f(x) - f(y)$ для всех $\Delta y, y \in \mathbb{R}^d$ (см. (15)). Следовательно, функция $g_j(\cdot, y)$ неотрицательна для любого y, удовлетворяющего неравенству $f(y) \leqslant f(x)$. Отсюда и из следствия 1 вытекает, что $a_j(y) \geqslant 0$ для любого такого y. □

Опираясь на необходимые и достаточные условия глобальной оптимальности из теоремы 2, а также лемму 2, можно предложить следующую модификацию метода кодифференциального спуска для минимизации кусочно-аффинных функций, которую мы называем *методом глобального кодифференциального спуска.*

Алгоритм 1: Метод глобального кодифференциального спуска (МГКС).

Шаг 1. Выберем начальную точку $x_0 \in \mathbb{R}^d$. Положим $n := 0$ и $M_n = J = \{1, \ldots, s\}$.
Шаг 2. Вычислим $\underline{d}f(x_n)$ и $z_j(x_n)$ для всех $j \in M_n$.
Шаг 3. Для каждого $j \in M_n$ найдём оптимальный план $(a_j(x_n), v_j(x_n)) \in \mathbb{R}^{d+1}$ задачи

$$\|(a, v)\|^2 \to \min, \quad (a, v) \in \underline{d}f(x_n) + z_j(x_n).$$

Если $a_j(x_n) \geqslant 0$, то $M_n := M_n \setminus \{j\}$.
Шаг 4. Если $M_n = \varnothing$, то **Стоп**. Иначе, найдём $j(n) \in M_n$ такое, что

$$j(n) \in \operatorname*{arg\,min}_{j \in M_n} f\left(x_n + \frac{1}{a_j(x_n)} v_j(x_n)\right).$$

Положим $x_{n+1} = x_n + [a_{j(n)}(x_n)]^{-1} v_{j(n)}(x_n)$, $M_{n+1} = M_n$, $n := n + 1$, и перейдём на **Шаг 2**.

Пусть последовательность $\{x_n\}$ построена по методу глобального кодифференциального спуска (МГКС). Заметим, что из неравенства (12) следует, что для любого $n \in \mathbb{N}$, либо $f(x_{n+1}) < f(x_n)$, либо $M_n = \varnothing$. Поэтому, если для некоторых $j \in J$ и $n \in \mathbb{N}$ будет $a_j(x_n) \geqslant 0$ (в этом случае МГКС отбрасывает индекс j из индексного множества M_n), то по лемме 2

$$a_j(x_k) \geqslant 0 \quad \forall k \geqslant n. \tag{16}$$

Следовательно, если МГКС заканчивает свою работу на n-й итерации (т.е. на n-й итерации оказалось, что $M_n = \varnothing$), то $a_j(x_n) \geqslant 0$ для всех $j \in J$, откуда с помощью теоремы 2 можно сделать вывод, что x_n является точкой глобального минимума функции f. Ниже мы покажем, что МГКС всегда заканчивает свою работу за конечное число шагов, то есть всегда находит точку глобального минимума невыпуклой кусочно-аффинной функции за конечное число шагов.

Прежде чем перейти к доказательству этого результата, приведём геометрическую интерпретацию каждой итерации метода глобального кодифференциального спуска, на основании которой и будет строиться доказательство. Предположим для простоты, что функция f выпукла, т.е. $\overline{d}f(x) \equiv \{0\}$ (см (2)). Пусть точка x_n была получена на n-й итерации МГКС. Гиподифференциал $\underline{d}f(x_n)$ функции f в точке x_n является выпуклым многогранником в $\mathbb{R}^{d+1}$. Согласно неравенству (5) будет $a \leqslant 0$ для всех $(a, v) \in \underline{d}f(x_n)$ и $\max_{(a,v)\in\underline{d}f(x_n)} a = 0$. Таким образом, множество $\{(a, v) \in \mathbb{R}^{d+1} \mid a = 0\} \cap \underline{d}f(x_n)$ является непустой гранью[2] многогранника $\underline{d}f(x_n)$. Заметим, что согласно неравенству (13) данная грань будет собственной, то есть она не совпадает с $\underline{d}f(x_n)$, поскольку в противном случае функция f была бы не ограничена снизу. Мы называем данную грань *активной гранью* многогранника $\underline{d}f(x_n)$. Нетрудно заметить, что субдифференциал $\partial f(x_n)$ выпуклой функции f в точке x_n состоит из всех таких векторов v, что пара $(0, v)$ принадлежит активной грани гиподифференциала $\underline{d}f(x_n)$.

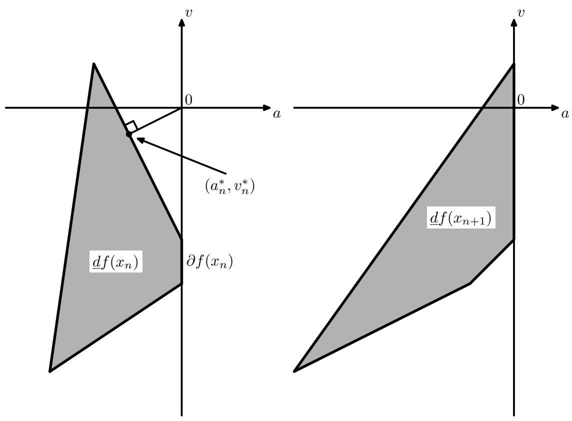

Рис. 1. Преобразование глобального кодифференциала на каждой итерации МГКС: $\underline{d}f(x_n)$ (левый рис.) и $\underline{d}f(x_{n+1})$ (правый рис.). Заметим, что все точки сдвигаются только горизонтально, т.е. только вдоль оси a (см. (3)).

[2]Напомним, что гранью выпуклого многогранника $P \subset \mathbb{R}^d$ называется любое множество вида $G = P \cap \{x \colon \langle c, x\rangle = c_0\}$, где $c \in \mathbb{R}^d$ и $c_0 \in \mathbb{R}$ удовлетворяют неравенству $\langle c, x\rangle \leqslant c_0$ для всех $x \in P$ (см., например, [12, Определение 2.1]).

Точка
$$\{(a_n^*, v_n^*)\} = \arg\min \left\{ \|(a,v)\|^2 \;\middle|\; (a,v) \in \underline{d}f(x_n) \right\} \tag{17}$$
лежит на некоторой грани G многогранника $\underline{d}f(x_n)$, которая не является активной, так как иначе функция f была бы не ограниченной снизу в соответствии с (13). При совершении одной итерации МГКС многогранник $\underline{d}f(x_n)$ преобразуется, согласно формуле (3). Как будет показано в доказательстве ниже, он преобразуется таким образом, что грань G становится активной грань многогранника $\underline{d}f(x_{n+1})$. Таким образом, точка (a_n^*, v_n^*) с наименьшей нормой принадлежит той грани гиподифференциала, которая станет активной на следующей итерации (см. Рис. 1).

Учитывая данное наблюдение, конечную сходимость МГКС в выпуклом случае можно показать следующим образом. Если доказано, что на некоторой итерации пара (a_n^*, v_n^*) с минимальной нормой принадлежит грани $\underline{d}f(x_n)$, пересекающейся с осью $\{(a,0) \in \mathbb{R}^{d+1} \mid a \in \mathbb{R}\}$, то $0 \in \partial f(x_{n+1})$ и, значит, точка x_{n+1} является точкой глобального минимума функции f.

В невыпуклом случае с помощью похожих рассуждений можно доказать, что за конечное число итераций МГКС отбросит, по крайней мере, один индекс j. Повторяя данное рассуждение s раз получим, что за конечное число шагов все индексы будут отброшены, и МГКС прекратит свою работу.

ТЕОРЕМА 3. *Пусть кусочно-аффинная функция f ограничена снизу. Тогда f достигает глобального минимума, и МГКС сходится к точке глобального минимума этой функции за конечное число шагов.*

Доказательство. Пусть конечная или бесконечная последовательность точек $\{x_n\}$ построена по МГКС для функции f. Обозначим $a_n^* = a_{j(n)}(x_n)$ и $v_n^* = v_{j(n)}(x_n)$, где индекс $j(n)$ вычисляется на Шаге 4 МГКС. Заметим, что данное определение пары (a_n^*, v_n^*) совпадает с определением (17) в выпуклом случае (т.е. когда $\overline{f}(x) \equiv 0$), поскольку в этом случае $z_j(x) \equiv 0$ для всех j (см. формулы (2) и (10)).

Из теоремы 2 следует, что если x_n не является точкой глобального минимума функции f, то существует $j \in J$ такое, что $a_j(x_n) < 0$ и
$$\begin{aligned} f(x_{n+1}) \leqslant f\left(x_n + \frac{1}{a_j(x_n)} v_j(x_n)\right) \leqslant f(x_n) - \frac{1}{|a_j(x_n)|} \big\|(a_j(x_n), v_j(x_n))\big\|^2 = \\ = f(x_n) - |a_j(x_n)| - \frac{1}{|a_j(x_n)|}\|v_j(x_n)\|^2 \end{aligned} \tag{18}$$
(см. (12) и Шаг 4 МГКС). Заметим, что
$$-|a_j(x_n)| - \frac{1}{|a_j(x_n)|}\|v_j(x_n)\|^2 \leqslant \begin{cases} -1, & \text{если } |a_j(x_n)| \geqslant 1, \\ -\|v_j(x_n)\|^2, & \text{если } |a_j(x_n)| < 1. \end{cases}$$

Следовательно, если x_n не является точкой глобального минимума функции f, то справедливы следующие неравенства:

$$f(x_{n+1}) - f(x_0) \leqslant -\sum_{k=0}^{n} \left(|a_k^*| + \frac{1}{|a_k^*|} \left\| v_k^* \right\|^2 \right) \leqslant -\sum_{k=0}^{n} \min \left\{ 1, \|v_k^*\|^2 \right\}. \tag{19}$$

Обозначим через $\mathcal{E}$ совокупность всех выпуклых множеств $C \subset \mathbb{R}^d$ таких, что $0 \notin C$ и

$$C = \operatorname{co}\{v_{i_1}, \ldots, v_{i_i}\} + w_j$$

для некоторых $i_1, \ldots, i_i \in I$, $1 \leqslant k \leqslant \ell$, и $j \in J$, где векторы v_i и w_j входят в представление функции f в виде разности выпуклых функций (1). Понятно, что $\mathcal{E}$ — это конечное семейство выпуклых компактных множеств и $\theta = \min_{C \in \mathcal{E}} \min_{v \in C} \|v\|^2 > 0$.

Обозначим $f^* = \inf_{x \in \mathbb{R}^d} f(x) > -\infty$ и $n^* = \lfloor (f(x_0) - f^*) / \min\{\theta, 1\} \rfloor + 1$ (здесь $\lfloor t \rfloor$ — наибольшее целое число не превосходящее $t \in \mathbb{R}$). Из неравенства (19) следует, что найдётся $n \leqslant n^*$ такое, что, либо МГКС прекращает свою работу на шаге n, либо $a_n^* < 0$ и $\|v_n^*\|^2 < \theta$. Как было замечено выше, в первом случае точка x_n является точкой глобального минимума функции f.

Предположим, что x_n не является точкой глобального минимума функции f. По определению, пара (a_n^*, v_n^*) принадлежит выпуклому многограннику $\underline{d}f(x_n) + z_{j(n)}(x_n)$ (см. Шаг 3 МГКС). Как известно, любой выпуклый многогранник является объединением относительных внутренностей своих граней (см., например, [12], с. 61). Поэтому пара (a_n^*, v_n^*) принадлежит относительной внутренности $\operatorname{relint} G$ некоторой грани G многогранника $\underline{d}f(x_n) + z_{j(n)}(x_n)$.

Из определения (a_n^*, v_n^*) и необходимого и достаточного условия минимума выпуклой функции на выпуклом множестве следует, что

$$a_n^* a + \langle v_n^*, v \rangle \geqslant \|(a_n^*, v_n^*)\|^2 \quad \forall (a, v) \in \underline{d}f(x_n) + z_{j(n)}(x_n). \tag{20}$$

Ясно, что это неравенство выполняется как равенство при $(a, v) = (a_n^*, v_n^*)$. Согласно [12, Предложение 2.3] грань выпуклого многогранника сама является выпуклым многогранником. Отсюда, воспользовавшись Леммой 2.9 из [12] о свойствах относительной внутренности выпуклого многогранника и условием $(a_n^*, v_n^*) \in \operatorname{relint} G$, получим

$$a_n^* a + \langle v_n^*, v \rangle = \|(a_n^*, v_n^*)\|^2 \quad \forall (a, v) \in G. \tag{21}$$

Вершины грани G, как выпуклого многогранника, также являются вершинами многогранника $\underline{d}f(x_n) + z_{j(n)}(x_n)$ по Предложению 2.3 из [12]. Поэтому

$$G = \operatorname{co}\{(a_{i_r} + \langle v_{i_r}, x_n \rangle - \underline{f}(x_n), v_{i_r}) \mid 1 \leqslant r \leqslant k\} + z_{j(n)}(x_n)$$

для некоторых $i_1, \ldots, i_k \in I$ и $1 \leqslant k \leqslant \ell$ (см. (3)). Заметим, что из определения θ и условий $(a_n^*, v_n^*) \in G$ и $\|v_n^*\|^2 < \theta$ следует, что $G \cap (\mathbb{R} \times \{0\}) \neq \varnothing$.

Введём выпуклую функцию

$$g_n(x) = \max_{(a,v)\in \underline{d}f(x_{n+1})+z_{j(n)}(x_{n+1})} (a + \langle v, x\rangle). \tag{22}$$

Покажем, что $0 \in \partial g_n(0)$. Действительно, по определению $z_j(x)$ (см. (10)) имеем

$$\begin{aligned} z_{j(n)}(x_{n+1}) &= \big(b_{j(n)} - \overline{f}(x_{n+1}) + \langle w_{j(n)}, x_{n+1}\rangle, w_{j(n)}\big) = \\ &= z_{j(n)}(x_n) + \big(\overline{f}(x_n) - \overline{f}(x_{n+1}) + \langle w_{j(n)}, x_{n+1} - x_n\rangle, 0\big). \end{aligned}$$

Из определения глобального кодифференциала (3) также следует, что

$$\underline{d}f(x_{n+1}) = \Big\{(a + \underline{f}(x_n) - \underline{f}(x_{n+1}) + \langle v, x_{n+1} - x_n\rangle, v) \in \mathbb{R}^{d+1} \;\Big|\; (a, v) \in \underline{d}f(x_n)\Big\}$$

Следовательно, воспользовавшись равенством $f(x) = \underline{f}(x) + \overline{f}(x)$, получим

$$\begin{aligned} &\underline{d}f(x_{n+1}) + z_{j(n)}(x_{n+1}) = \\ &\quad = \Big\{(a + \langle v, x_{n+1} - x_n\rangle - f(x_{n+1}) + f(x_n), v) \;\Big|\; (a, v) \in \underline{d}f(x_n) + z_{j(n)}(x_n)\Big\}. \end{aligned}$$

Поэтому

$$g_n(0) = \max_{(a,v)\in \underline{d}f(x_n)+z_{j(n)}(x_n)} (a + \langle v, x_{n+1} - x_n\rangle) - f(x_{n+1}) + f(x_n).$$

Отсюда, учитывая (20), (21), равенство $x_{n+1} - x_n = [a_n^*]^{-1} v_n^*$ и неравенство $a_n^* < 0$, получаем следующее неравенство:

$$g_n(0) = \frac{1}{a_n^*}\|(a_n^*, v_n^*)\|^2 - f(x_{n+1}) + f(x_n) \geqslant 0 \tag{23}$$

(справедливость последнего неравенства следует из (18)). Более того, максимум в определении $g_n(0)$ достигается в точках вида

$$(a + \langle v, x_{n+1} - x_n\rangle - f(x_{n+1}) + f(x_n), v), \quad (a, v) \in G.$$

Поэтому, согласно формуле для субдифференциала супремума бесконечного семейства выпуклых функций [13, Теорема 4.2.3], справедливо включение

$$\Big\{v \in \mathbb{R}^d \;\Big|\; \exists a \in \mathbb{R}\colon (a, v) \in G\Big\} \subseteq \partial g_n(0),$$

из которого вытекает, что $0 \in \partial g_n(0)$, поскольку $G \cap (\mathbb{R} \times \{0\}) \neq \varnothing$. Таким образом, 0 является точкой глобального минимума выпуклой функции $g_n(x)$.

Следовательно, функция g_n неотрицательна, так как $g_n(0) \geqslant 0$ по неравенству (23). Отсюда, учитывая определение функции g_n (см. (22)) и пользуясь следствием 1, получаем, что $a_{j(n)}(x_{n+1}) \geqslant 0$. Значит, индекс $j(n)$ отбрасывается методом на $(n+1)$-й итерации, и по Лемме 2 для всех $k \geqslant n+1$ справедливо неравенство $a_{j(n)}(x_k) \geqslant 0$.

Таким образом, найдётся $n_1 \leqslant n^* := \lfloor (f(x_0) - f^*)/\min\{\theta, 1\} \rfloor + 1$ такое, что МГКС отбрасывает индекс $j(n_1)$ на $(n_1 + 1)$-й итерации. Из определения n^* и неравенства (19) следует, что найдётся $n_2 \leqslant n_1 + n^* \leqslant 2n^*$ такое, что либо МГКС заканчивает свою работу n_2-й итерации, либо $a_{n_2} < 0$ и выполняется неравенство $\|v_{n_2}\|^2 < \theta$. Повторяя рассуждение, приведённое выше, нетрудно показать, что МГКС отбрасывает индекс $j(n_2)$ на $(n_2 + 1)$-й итерации, и $a_{j(n_2)}(x_k) \geqslant 0$ для всех $k \geqslant n_2 + 1$. Повторяя то же самое рассуждение s раз, получаем что МГКС отбрасывает все индексы $j \in J$ не более чем за sn^* итераций, и, тем самым, заканчивает свою работу за конечное число шагов. Более того, если МГКС заканчивает свою работу на n-й итерации, то как было указано выше (см. (16)), по Лемме 2 будет $a_j(x_n) \geqslant 0$ для всех $j \in J$. Отсюда и из Теоремы 2 следует, что x_n является точкой глобального минимума функции f. Теорема доказана. $\square$

Приведём два простых примера работы метода глобального кодифференциального спуска.

ПРИМЕР 1. Пусть $d = 1$. Рассмотрим ломаную $f(x)$, заданную на Рис. 2. Воспользовавшись результатами доклада [6], нетрудно выписать следующее аналитическое представление данной ломаной:

$$\begin{aligned} f(x) &= -x + 2x_+ - 2(x-2)_+ + 2(x-3)_+ = \\ &= \Big(2x_+ + (2x-6)_+\Big) - \Big(x + (2x-4)_+\Big) = \\ &= \max\{0, 2x, -6+4x\} + \min\{-x, 4-3x\}. \end{aligned}$$

В данном случае $\ell = 3$, $s = 2$ и

$$\begin{pmatrix} a_1 \\ v_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} a_2 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} a_3 \\ v_3 \end{pmatrix} = \begin{pmatrix} -6 \\ 4 \end{pmatrix},$$
$$\begin{pmatrix} b_1 \\ w_1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} b_2 \\ w_2 \end{pmatrix} = \begin{pmatrix} 4 \\ -3 \end{pmatrix}$$

(см. (1)). Кроме того, согласно (2) положим

$$\underline{f}(x) = \max\{0, 2x, -6+4x\}, \quad \overline{f}(x) = \min\{-x, 4-3x\}.$$

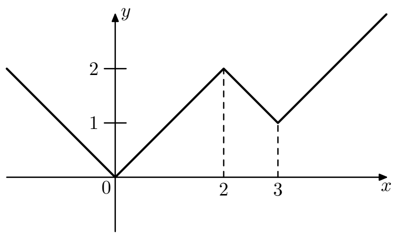

Рис. 2. График ломаной $f(x)$.

Воспользуемся методом глобального кодифференциального спуска для минимизации функции f. В качестве начального приближения выберем точку $x_0 = 3$, которая является точкой локального минимума функции f.
Нулевая итерация. Шаг 1. Положим $n := 0$ и $M_0 = J = \{1, 2\}$.
Шаг 2. Вычислим $\underline{d}f(x_0)$ и $z_j(x_0)$, $j \in M_0$, по формулам (3) и (10) соответственно. Имеем $\underline{f}(x_0) = 6$ и

$$\underline{d}f(x_0) = \mathrm{co}\left\{ \begin{pmatrix} a_i - \underline{f}(x_0) + \langle v_i, x_0\rangle \\ v_i \end{pmatrix} \;\middle|\; i \in I \right\} = \mathrm{co}\left\{ \begin{pmatrix} -6 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \end{pmatrix} \right\}.$$

Также $\overline{f}(x_0) = -5$ и

$$z_1(x_0) = \begin{pmatrix} b_1 - \overline{f}(x_0) + \langle w_1, x_0\rangle \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix},$$
$$z_2(x_0) = \begin{pmatrix} b_2 - \overline{f}(x_0) + \langle w_2, x_0\rangle \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$$

Шаг 3. Пусть $j = 1$. В этом случае

$$\underline{d}f(x_0) + z_1(x_0) = \mathrm{co}\left\{ \begin{pmatrix} -4 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\}.$$

Оптимальным планом задачи

$$\|(a, v)\|^2 \to \min, \quad (a, v) \in \underline{d}f(x_0) + z_1(x_0)$$

является точка $(a_1(x_0); v_1(x_0)) = (-0,1; 0,3)$. Заметим, что $a_1(x_0) < 0$, то есть в точке x_0 не выполнены условия глобального минимума из теоремы 2.

Пусть теперь $j = 2$. В этом случае

$$\underline{d}f(x_0) + z_2(x_0) = \mathrm{co}\left\{ \begin{pmatrix} -6 \\ -3 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Оптимальным планом задачи

$$\|(a, v)\|^2 \to \min, \quad (a, v) \in \underline{d}f(x_0) + z_2(x_0)$$

является точка $(a_2(x_0); v_2(x_0)) = (0; 0)$. Поскольку $a_2(x_0) \geqslant 0$, то индекс $j = 2$ отбрасывается, то есть $M_0 = \{1\}$.

Шаг 4. Так как индексное множество M_0 состоит из единственного элемента $j = 1$, то полагаем $x_1 := x_0 + [a_1(x_0)]^{-1}v_1(x_0) = 0$, $n = 1$ и $M_1 = M_0$.
Первая итерация. Шаг 2. Вычислим $\underline{d}f(x_1)$ и $z_j(x_1)$, $j \in M_1 = \{1\}$. Имеем $\underline{f}(x_1) = 0$ и

$$\underline{d}f(x_1) = \mathrm{co}\left\{ \begin{pmatrix} a_i - \underline{f}(x_1) + \langle v_i, x_1 \rangle \\ v_i \end{pmatrix} \;\middle|\; i \in I \right\} = \mathrm{co}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} -6 \\ 4 \end{pmatrix} \right\},$$

а также $\overline{f}(x_1) = 0$ и

$$z_1(x_1) = \begin{pmatrix} b_1 - \overline{f}(x_1) + \langle w_1, x_1 \rangle \\ w_1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

Шаг 3. Для $j = 1$ имеем

$$\underline{d}f(x_0) + z_1(x_0) = \mathrm{co}\left\{ \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -6 \\ 3 \end{pmatrix} \right\}.$$

Оптимальным планом задачи

$$\|(a, v)\|^2 \to \min, \quad (a, v) \in \underline{d}f(x_1) + z_1(x_1)$$

является точка $(a_1(x_1); v_1(x_1)) = (0; 0)$. Поскольку $a_1(x_1) \geqslant 0$, то индекс $j = 1$ отбрасывается, то есть $M_1 = \varnothing$, и алгоритм заканчивается свою работу.

Таким образом, метод глобального кодифференциального спуска «перепрыгнул» из точки локального минимума $x_0 = 3$ функции f в точку глобального минимума $x_1 = 0$ за одну итерацию.

ПРИМЕР 2. Пусть $d = 2$ и

$$f(x) = \min\big\{ \max\{|x_1|, |x_2|\}, 1 + \max\{2|x_1 - 2|, |x_2 - 2|\}\big\}.$$

В качестве начального приближения выберем точку $x_0 = (2, 2)$. Нетрудно заметить, что x_0 — точка локального минимума функции f, в то время как глобальный минимум достигается в точке $x^* = (0, 0)$.

Вместо того, чтобы находить DC разложение функции f в точке x и затем пользоваться определением глобального кодифференциала (3) для того чтобы вычислить $Df(x_0)$, мы вычислим кодифференциал с помощью предложения 1.

Положим.

$$f_1(x) = \max\{|x_1|, |x_2|\}, \quad f_2(x) = 1 + \max\{2|x_1 - 2|, |x_2 - 2|\}.$$

Тогда $f(x)=\min\{f_1(x),f_2(x)\}$. Воспользовавшись пунктами 1, 2 и 5 предложения 1, получим

$$\underline{d}f_1(x_0)=\operatorname{co}\left\{\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}-4\\-1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix},\begin{pmatrix}-4\\0\\-1\end{pmatrix}\right\},\quad \overline{d}f_1(x_0)=\{0\},$$
$$\underline{d}f_2(x_0)=\operatorname{co}\left\{\begin{pmatrix}0\\2\\0\end{pmatrix},\begin{pmatrix}0\\-2\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix},\begin{pmatrix}0\\0\\-1\end{pmatrix}\right\},\quad \overline{d}f_2(x_0)=\{0\}.$$

Теперь применяя п. 6 предложения 1, получаем, что $\underline{d}f(x_0)=\underline{d}f_1(x_0)+\underline{d}f_2(x_0)$, т. е.

$$\underline{d}f(x_0)=\operatorname{co}\left\{\begin{pmatrix}0\\3\\0\end{pmatrix},\begin{pmatrix}-4\\1\\0\end{pmatrix},\begin{pmatrix}0\\2\\1\end{pmatrix},\begin{pmatrix}-4\\2\\-1\end{pmatrix},\begin{pmatrix}0\\-1\\0\end{pmatrix},\begin{pmatrix}-4\\-3\\0\end{pmatrix},\begin{pmatrix}0\\-2\\1\end{pmatrix},\begin{pmatrix}-4\\-2\\-1\end{pmatrix},\right.$$
$$\left.\begin{pmatrix}0\\1\\1\end{pmatrix},\begin{pmatrix}-4\\-1\\1\end{pmatrix},\begin{pmatrix}0\\0\\2\end{pmatrix},\begin{pmatrix}-4\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\-1\end{pmatrix},\begin{pmatrix}-4\\-1\\-1\end{pmatrix},\begin{pmatrix}0\\0\\0\end{pmatrix},\begin{pmatrix}-4\\0\\-2\end{pmatrix}\right\}.$$

Также имеем

$$\overline{d}f(x_0)=\operatorname{co}\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}+\overline{d}f_1(x_0)-\underline{d}f_2(x_0),\overline{d}f_2(x_0)-\underline{d}f_1(x_0)\right\}=$$
$$=\operatorname{co}\left\{\begin{pmatrix}1\\2\\0\end{pmatrix},\begin{pmatrix}1\\-2\\0\end{pmatrix},\begin{pmatrix}1\\0\\1\end{pmatrix},\begin{pmatrix}1\\0\\-1\end{pmatrix},\begin{pmatrix}0\\-1\\0\end{pmatrix},\begin{pmatrix}4\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\-1\end{pmatrix},\begin{pmatrix}4\\0\\1\end{pmatrix}\right\}.$$

Воспользуемся методом глобального кодифференциального спуска (детали опущены для краткости). Оптимальным планом задачи

$$\|(a,v)\|^2\to\min,\quad (a,v)\in\underline{d}f(x_0)+z_i(x_0)$$

(Шаг 3 МГКС) для $z_1(x_0)=(1,2,0)^T\in\overline{d}f(x_0)$ является вектор

$$(a_1(x_0),v_1(x_0))\approx(-0.1111,0.2222,0.2222).$$

Поскольку $a_1(x_0)<0$, то по теореме 2 можно заключить, что x_0 не является точкой глобального минимума функции f. Более того, положив согласно МГКС $x_1=x_0+[a_1(x_0)]^{-1}v_1(x_0)$, получим $x_1=(0,0)=x^*$, т.е. метод глобального кодифференциального спуска вновь «перепрыгнул» из точки локального минимума x_0 в точку глобального минимума x^* всего за один шаг.

В заключение отметим, что, рассуждая как и при доказательстве теоремы 3, можно показать, что в случае $\mu=\nu=+\infty$ метод кодифференциального спуска [8] также сходится к точке глобального минимума кусочно-аффинной функции за конечное число шагов. Доказательство этого результата имеется в работе [1].

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 18-31-00014.

ЛИТЕРАТУРА

1. M.V. Dolgopolik. *The method of codifferential descent for convex and global piecewise affine optimization.* // Optimization Methods & Software, 2019. doi: 10.1080/10556788.2019.1571590. URL: https://arxiv.org/pdf/1807.05538.pdf

2. A. Kripfgang, R. Schulze. *Piecewise affine functions as a difference of two convex functions.* // Optimization, 1987, vol. 18, pp. 23–29.

3. V. V. Gorokhovik, O. I. Zorko. *Piecewise affine functions and polyhedral sets* // Optimization, 1994, vol. 31, pp. 209–221.

4. V.V. Gorokhovik, *Geometrical and analytical characteristic properties of piecewise affine mappings* // arXiv preprint: 1111.1389, 2011. URL: https://arxiv.org/pdf/1111.1389.pdf

5. Т.А. Ангелов. *Представление кусочно-аффинных функций в виде разности полиэдральных* // Вестник С.-Петербург. ун-та. Сер. 10. Прикл. матем. Информ. Проц. упр., 2016, № 1, 4–18. URL: http://cyberleninka.ru/article /v/predstavlenie-kusochno-affinnyh-funktsiy-v-vide-raznosti-poliedralnyh

6. В.Н. Малозёмов, Г.Ш. Тамасян. *Представления непрерывных кусочно-аффинных функций* // Семинар «CNSA & NDO». Избранные доклады. 10 сентября 2019 г. (`http://apmath.spbu.ru/cnsa/reps19.shtml#0910`)

7. В.Ф. Демьянов, А.М. Рубинов. Основы негладкого анализа и квазидифференциальное исчисление. М.: Наука, 1990.

8. М.В. Долгополик. *Метод кодифференциального спуска* // Семинар «CNSA & NDO». Избранные доклады. 28 февраял 2019 г. (`http://apmath.spbu.ru/cnsa/reps19.shtml#0228`)

9. L. N. Polyakova. *On global unconstrained minimization of the difference of polyhedral functions* // J. Glob. Optim., 2011, vol. 50, no. 2, pp. 179–195.

10. И. Экланд, Р. Темам. Выпуклый анализ и вариационные проблемы. М.: Мир, 1979.

11. J.-B. Hiriart-Urruty, C. Lemaréchal. Convex Analysis and Minimization Algorithms II. Advanced Theory and Bundle Methods. Berlin, Heidelberg: Springer-Verlag, 1993.

12. G. M. Zeigler. Lectures on Polytopes. New York: Springer Verlag, 1995.

13. А.Д. Иоффе, В.М. Тихомиров. Теория экстремальных задач. М.: Наука, 1974.