МЕТОД КОДИФФЕРЕНЦИАЛЬНОГО СПУСКА*

М. В. Долгополик
maxim.dolgopolik@gmail.com

28 февраля 2019 г.

Аннотация. Кодифференцируемые функции были введены В.Ф. Демьяновым в работах [1–3]. В данном докладе обсуждается метод минимизации таких функций — метод кодифференциального спуска. Для полноты изложения приводятся некоторые базовые определения и результаты из кодифференциального исчисления, а также ряд примеров вычисления кодифференциалов и усечённых кодифференциалов негладких функций. По поводу более подробного изложения кодифференциального исчисления, метода кодифференциального спуска и его различных модификаций см. [4–9].

1°.Кодифференцируемые функции. Пусть функция f определена на $\mathbb{R}^d$ и принимает вещественные значения. Напомним, что f называется *кодифференцируемой* в точке x, если существует пара выпуклых компактных множеств $\underline{d}f(x)$, $\overline{d}f(x) \subset \mathbb{R} \times \mathbb{R}^d$ таких, что функция f допускает разложение в точке x вида

$$f(x+h) - f(x) = \max_{(a,v)\in\underline{d}f(x)} (a + \langle v, h\rangle) + \min_{(b,w)\in\overline{d}f(x)} (b + \langle w, h\rangle) + o(\|h\|) \qquad (1)$$

и справедливо равенство

$$\max_{(a,v)\in\underline{d}f(x)} a = \min_{(b,w)\in\overline{d}f(x)} b = 0. \qquad (2)$$

Пара множеств $Df(x) = [\underline{d}f(x), \overline{d}f(x)]$ называется кодифференциалом функции f в точке x, множество $\underline{d}f(x)$ называется гиподифференциалом функции f в точке x, а множество $\overline{d}f(x)$ называется гипердифференциалом данной функции в точке x. Элементами гипо- и гипер-дифференциалов являются пары (a, v), в которых a — вещественное число, v — вектор из $\mathbb{R}^d$. При теоретическом анализе кодифференцируемых функций удобнее использовать данную

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

форму записи. Однако, при вычислении кодифференциалов конкретных функций удобнее считать, что гипо- и гипер-дифференциалы — это выпуклые компактные подмножества пространства $\mathbb{R}^{d+1}$, и записывать их элементы в виде вектор-столбцов $\binom{a}{v}$. Везде далее мы будем использовать подобную двойную систему обозначений.

Заметим, что функция $h \mapsto \max\limits_{(a,v)\in \underline{d}f(x)} (a + \langle v, h\rangle)$ является выпуклой, в то время как функция $h \mapsto \min\limits_{(b,w)\in \overline{d}f(x)} (b + \langle w, h\rangle)$ является вогнутой. Таким образом, приращение функции f в (1) аппроксимируется с точностью до первого порядка малости с помощью суммы выпуклой и вогнутой функций или, что тоже самое, с помощью разности выпуклых функций специального вида. Можно показать, что любая функция, приращение которой можно аппроксимировать с помощью разности произвольных выпуклых функций, является кодифференцируемой (см. [10]). Иными словами, функция f кодифференцируема в точке x тогда и только тогда, когда её приращение в этой точке можно аппроксимировать с помощью разности выпуклых функций (d.c. функции). Отсюда, в частности, следует, что любая d.c. функция является кодифференцируемой.

Кодифференциал $Df(x)$ функции f в точке x, очевидно, определяется не единственным образом. Например, нетрудно проверить, что для любого выпуклого компактного множества $C \subset \mathbb{R} \times \mathbb{R}^d$ пара $[\underline{d}f(x) + C, \overline{d}f(x) - C]$ также является кодифференциалом функции f в точке x.

ОПРЕДЕЛЕНИЕ 1. Функция f называется *непрерывно кодифференцируемой в точке* x, если она кодифференцируема в каждой точке некоторой окрестности U точки x, и существует кодифференциальное отображение $U \ni y \mapsto Df(y)$ такое, что соответствующие многозначные отображения $y \mapsto \underline{d}f(y)$ и $y \mapsto \overline{d}f(y)$ непрерывны в метрике Хаусдорфа в точке x.

Функция f называется *непрерывно кодифференцируемой*, если она кодифференцируема в каждой точке $x \in \mathbb{R}^d$, и существует кодифференциальное отображение $x \mapsto Df(x)$ определённое на всём пространстве $\mathbb{R}^d$ и такое, что соответствующие многозначные отображения $x \mapsto \underline{d}f(x)$ и $x \mapsto \overline{d}f(x)$ непрерывны в метрике Хаусдорфа на $\mathbb{R}^d$.

Можно показать, что класс непрерывно кодифференцируемых функций содержит в себе класс непрерывно дифференцируемых функций и замкнут относительно всех алгебраических операций (сложение, умножение, деление), взятия поточного минимума и максимума конечного числа функций, а также композиции с непрерывно дифференцируемыми функциями [4, 9]. Более того, справедливы следующие формулы исчисления непрерывно кодифференцируемых функций:

1) $D(f + c)(x) = Df(x)$ для любого $c \in \mathbb{R}$;

2) $D(f+g)(x) = Df(x) + Dg(x)$;

3) $D(\lambda f)(x) = \lambda Df(x)$;

4) $D(fg)(x) = g(x)Df(x) + f(x)Dg(x)$;

5) $D\left(\dfrac{1}{f}\right)(x) = -\dfrac{1}{f(x)^2}Df(x)$, если $f(x) \neq 0$;

6) Если $f(x) = \max\limits_{1\leqslant i\leqslant m} f_i(x)$, то
$$Df(x) = \left[\operatorname{co}\left\{(f_i(x) - f(x), \mathbf{0}_d) + \underline{d}f_i(x) - \sum_{j\neq i}\overline{d}f_j(x) \;\Big|\; 1 \leqslant i \leqslant m\right\}, \sum_{i=1}^{m}\overline{d}f_i(x)\right],$$
где функции $f_i\colon \mathbb{R}^d \to \mathbb{R}$ непрерывно кодифференцируемы, $\mathbf{0}_d \in \mathbb{R}^d$ — нулевой вектор.

7) Если $f(x) = \min\limits_{1\leqslant i\leqslant m} f_i(x)$, то
$$Df(x) = \left[\sum_{i=1}^{m}\underline{d}f_i(x), \operatorname{co}\left\{(f_i(x) - f(x), \mathbf{0}_d) + \overline{d}f_i(x) - \sum_{j\neq i}\underline{d}f_j(x) \;\Bigg|\; 1 \leqslant i \leqslant m\right\}\right].$$

8) Если функция $F\colon \mathbb{R}^n \to \mathbb{R}$, $F = F(y)$ непрерывно дифференцируема и $g(x) = F(f_1(x), \ldots, f_n(x))$, где функции $f_i\colon \mathbb{R}^d \to \mathbb{R}$ непрерывно кодифференцируемы, то
$$Dg(x) = \sum_{i=1}^{n}\frac{\partial F(f_1(x), \ldots, f_n(x))}{\partial y_i}Df_i(x).$$

Здесь и далее мы используем следующие операции сложения и умножения на число для пар выпуклых компактов:
$$[A_1, B_1] + [A_2, B_2] = [A_1 + A_2, B_1 + B_2],$$
$$\lambda[A, B] = \begin{cases}[\lambda A, \lambda B], & \text{если } \lambda \geqslant 0,\\ [\lambda B, \lambda A], & \text{если } \lambda < 0.\end{cases}$$

В случае, когда функция f непрерывно дифференцируема, можно положить
$$Df(x) = \big[\{(0, \nabla f(x))\}, \{(0, \mathbf{0}_d)\}\big] \text{ или } Df(x) = \big[\{(0, \mathbf{0}_d)\}, \{(0, \nabla f(x))\}\big]. \qquad (3)$$

Приведём несколько примеров вычисления кодифференциалов негладких функций.

ПРИМЕР 1. Пусть $d = 2$ и $f(x) = |x_1| - |x_2|$. Для любого h будет

$$\begin{aligned} f(x+h) - f(x) &= |x_1 + h_1| - |x_2 + h_2| - |x_1| + |x_2| = \\ &= \max\{x_1 + h_1, -x_1 - h_1\} + \min\{x_2 + h_2, -x_2 - h_2\} - |x_1| + |x_2| = \\ &= \max\{x_1 - |x_1| + h_1, -x_1 - |x_1| - h_1\} + \\ &+ \min\{x_2 + |x_2| + h_2, -x_2 + |x_2| - h_2\}. \end{aligned}$$

Положим

$$\underline{d}f(x) = \operatorname{co}\left\{ \begin{pmatrix} x_1 - |x_1| \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -x_1 - |x_1| \\ -1 \\ 0 \end{pmatrix} \right\}, \tag{4}$$

$$\overline{d}f(x) = \operatorname{co}\left\{ \begin{pmatrix} x_2 + |x_2| \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -x_2 + |x_2| \\ 0 \\ -1 \end{pmatrix} \right\}. \tag{5}$$

Тогда, как нетрудно заметить,

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

для всех x и h, и

$$\max_{(a,v)\in\underline{d}f(x)} a = \min_{(b,w)\in\overline{d}f(x)} b = 0.$$

Значит, функция f кодифференцируема в каждой точке $x \in \mathbb{R}^2$, а пара $Df(x) = [\underline{d}f(x), \overline{d}f(x)]$, определённая в (4), (5), является кодифференциалом функции f в этой точке. Многозначные отображения $x \mapsto \underline{d}f(x)$ и $x \mapsto \overline{d}f(x)$, очевидно, являются непрерывными в метрике Хаусдорфа. Значит, функция f непрерывно кодифференцируема.

ПРИМЕР 2. Пусть $f(x) = \max\limits_{1\leqslant i\leqslant m} f_i(x)$, где функции f_i непрерывно дифференцируемы. Положим

$$Df_i(x) = \left[\left\{ \begin{pmatrix} 0 \\ \nabla f_i(x) \end{pmatrix} \right\}, \{\mathbf{0}_{d+1}\} \right].$$

Воспользовавшись правилом 6 вычисления кодифференциала функции максимума, получим, что функция f непрерывно кодифференцируема и

$$\underline{d}f(x) = \operatorname{co}\left\{ \begin{pmatrix} f_i(x) - f(x) \\ \nabla f_i(x) \end{pmatrix} \,\middle|\, 1 \leqslant i \leqslant m \right\}, \quad \overline{d}f(x) = \{\mathbf{0}_{d+1}\}.$$

Если же $f(x) = \max\limits_{1 \leqslant i \leqslant m} |f_i(x)|$, то

$$\underline{d}f(x) = \mathrm{co}\left\{ \begin{pmatrix} f_i(x) - f(x) \\ \nabla f_i(x) \end{pmatrix}, \begin{pmatrix} -f_i(x) - f(x) \\ -\nabla f_i(x) \end{pmatrix} \;\middle|\; 1 \leqslant i \leqslant m \right\}$$

и $\overline{d}f(x) = \{\mathbf{0}_{d+1}\}$.

ПРИМЕР 3. Рассмотрим две непрерывно кодифференцируемые функции, возникающие в задачах кластеризации данных [11–13]. Пусть в пространстве $\mathbb{R}^d$ задан конечный набор точек $A = \{a_1, \ldots, a_m\}$. Основная задача кластеризации данных заключается в разбиении этого множества на ℓ возможно пересекающихся множеств A_j (то есть $A = \bigcup_{j=1}^{\ell} A_j$) согласно некоторому критерию. Каждое из множеств A_j при этом называется *кластером*. См. [14] по поводу различных постановок задач кластеризации.

Часто с каждым кластером A_j связывают *центр* $x_j \in \mathbb{R}^d$, и определяют сам кластер A_j, как совокупность тех точек a_k, для которых расстояние до центра x_j не превосходит расстояния до центров других кластеров, т.е.

$$A_j = \left\{ a_k \in A \;\middle|\; \|x_j - a_k\| \leqslant \min_{i \neq j} \|x_i - a_k\| \right\},$$

где $\|\cdot\|$ — некоторая норма. Везде далее мы будем считать, что $\|\cdot\|$ — евклидова норма. Величина $R_j = \max\{\|x_j - a_k\| \mid a_k \in A_j\}$ называется *радиусом* кластера. В данной постановке задачи кластеризации исходный набор точек A накрывается системой из ℓ шаров с центрами x_j и радиусами R_j. Каждый кластер A_j определяется по соответствующему ему шару. Сама же задача кластеризации данных заключается в нахождении в некотором смысле оптимальной системы таких шаров или, точнее, центров x_j этих шаров.

Пусть точка a_k принадлежит i-му кластеру, т.е. $\|x_i - a_k\| \leqslant \|x_j - a_k\|$ для всех $j \neq i$. Величину уклонения точки a_k от центра соответствующего ей кластера можно оценить с помощью функции

$$f_k(x) = \min_{1 \leqslant j \leqslant \ell} \|x_j - a_k\|^2, \qquad x = (x_1^T, \ldots, x_\ell^T)^T \in \mathbb{R}^{ld}.$$

Одна из возможных постановок задачи кластеризации данных заключается в минимизации суммы таких уклонений, то есть функции

$$f(x) = \sum_{k=1}^{m} f_k(x).$$

Данная функция является непрерывно кодифференцируемой. Вычислим её кодифференциал. Для функции $f_{jk}(x) = \|x_j - a_k\|^2$ положим

$$Df_{jk}(x) = \left[\{\mathbf{0}_{\ell d+1}\}, \left\{ \begin{pmatrix} 0 \\ \nabla f_{jk}(x) \end{pmatrix} \right\} \right].$$

Тогда воспользовавшись правилом 7 вычисления кодифференциала функции минимума, получим $\underline{d}f_k(x) = \{\mathbf{0}_{\ell d+1}\}$ и

$$\overline{d}f_k(x) = \operatorname{co}\left\{ \begin{pmatrix} \|x_j - a_k\|^2 - f_k(x) \\ v_{jk}(x) \end{pmatrix} \;\middle|\; 1 \leqslant j \leqslant \ell \right\}, \quad v_{jk}(x) = \begin{pmatrix} \mathbf{0}_{(j-1)d} \\ 2(x_j - a_k) \\ \mathbf{0}_{(\ell-j)d} \end{pmatrix}.$$

Теперь используя правило 2 вычисления кодифференциала суммы, имеем $\underline{d}f(x) = \{\mathbf{0}_{\ell d+1}\}$ и

$$\overline{d}f(x) = \operatorname{co}\left\{ \begin{pmatrix} \sum\limits_{k=1}^{m} \|x_{j_k} - a_k\|^2 - f(x) \\ \sum\limits_{k=1}^{m} v_{j_k k}(x) \end{pmatrix} \;\middle|\; (j_1, \ldots, j_m) \in J_1 \times \ldots \times J_m \right\}, \qquad (6)$$

где $J_k := \{1, \ldots, \ell\}$ для всех k. Заметим, что гипердифференциал $\overline{d}f(x)$ представляет из себя выпуклую оболочку ℓ^m точек в пространстве $\mathbb{R}^{ld+1}$.

Также можно рассмотреть задачу минимизации максимального уклонения точек a_k от центров соответствующих им кластеров, то есть задачу минимизации функции

$$g(x) = \max_{1 \leqslant k \leqslant m} f_k(x) = \max_{1 \leqslant k \leqslant m} \min_{1 \leqslant j \leqslant \ell} \|x_j - a_k\|^2$$

(если заменить $\|x_j - a_k\|^2$ на $\|x_j - a_k\|$, то это будет в точности задача минимизации максимального радиуса кластеров, которая в случае $\ell = 1$ представляет из себя задачу Сильвестра [15]). Данная функция также является непрерывно кодифференцируемой. Вычислим её кодифференциал. По правилу 6 вычисления кодифференциала функции максимума гипердифференциал $\overline{d}g(x)$ этой функции совпадает с множеством (6), а гиподифференциал имеет вид

$$\underline{d}g(x) = \operatorname{co}\left\{ C_k(x) \;\middle|\; 1 \leqslant k \leqslant m \right\},$$

где

$$C_k(x) = \left\{ \begin{pmatrix} -\sum\limits_{s \neq k} \|x_{j_s} - a_s\|^2 + \sum\limits_{i=1}^{m} f_i(x) - g(x) \\ -\sum\limits_{s \neq k} v_{j_s s}(x) \end{pmatrix} \;\middle|\; j_s \in J_s,\ s \in \{1, \ldots, k-1, k+1, \ldots, m\} \right\}$$

Гиподифференциал $\underline{d}g(x)$ в данном случае является выпуклой оболочкой $m\ell^{m-1}$ точек.

ПРИМЕР 4. Рассмотрим также простой пример применения правила 8 вычисления кодифференциала сложной функции. Пусть $d = 2$ и

$$f(x) = \Big(\sin\big(|x_1| - |x_2|\big) + 1\Big) \cdot e^{|x_2| - |x_1| + 1}.$$

Вычислим кодифференциал данной функции в точке $x_0 = \mathbf{0}_2$. Положим

$$\begin{gathered} f_1(x) = |x_1| - |x_2|, \quad f_2(x) = |x_2| - |x_1| + 1, \\ f_3(x) = \sin f_1(x) + 1, \quad f_4(x) = e^{f_2(x)}. \end{gathered}$$

Тогда $f(x) = f_3(x) \cdot f_4(x)$. Согласно примеру 1 и правилу 3 будет

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

По правилу 8 вычисления кодифференциала сложной функции имеем

$$\begin{gathered} Df_3(x_0) = \cos(f_1(x_0))Df_1(x_0) = Df_1(x_0) \\ Df_4(x_0) = e^{f_2(x_0)}Df_2(x_0) = \left[\mathrm{co}\left\{\begin{pmatrix}0\\0\\e\end{pmatrix}, \begin{pmatrix}0\\0\\-e\end{pmatrix}\right\}, \mathrm{co}\left\{\begin{pmatrix}0\\e\\0\end{pmatrix}, \begin{pmatrix}0\\-e\\0\end{pmatrix}\right\}\right]. \end{gathered}$$

Наконец, по правилу 4 вычисления кодифференциала произведения будет

$$Df(x_0) = f_3(x_0)Df_4(x_0) + f_4(x_0)Df_3(x_0),$$

откуда

$$\underline{d}f(x_0) = \overline{d}f(x_0) = \mathrm{co}\left\{\begin{pmatrix}0\\e\\e\end{pmatrix}, \begin{pmatrix}0\\e\\-e\end{pmatrix}, \begin{pmatrix}0\\-e\\e\end{pmatrix}, \begin{pmatrix}0\\-e\\-e\end{pmatrix}\right\}.$$

Нетрудно заметить, что для любого $h \in \mathbb{R}^2$ будет

$$\max_{(a,v)\in\underline{d}f(x_0)}(a + \langle v, h\rangle) = e(|h_1| + |h_2|) = -\min_{(b,w)\in\overline{d}f(x_0)}(b + \langle w, h\rangle),$$

и поэтому можно положить $\underline{d}f(x_0) = \overline{d}f(x_0) = \{\mathbf{0}_3\}$. Данный пример показывает, что в некоторых случаях, используя особенности конкретной задачи, можно вычислить гораздо более простой кодифференциал исследуемой функции в заданной точке, чем с помощью прямого применение правил кодифференциального исчисления. С другой стороны, применение этих правил гарантирует сохранение свойства непрерывности кодифференциального отображения, в то время как при использовании каких-либо процедур сокращения кодифференциалов непрерывность может нарушаться.

Нетрудно показать, что любая кодифференцируемая в точке x функция является дифференцируемой по направлениям в этой точке. Более того, стандартное необходимое условие минимума в терминах производной по направлениям ($f'(x,h) \geqslant 0$ для любого h) может быть легко переписано в терминах кодифференциалов [4, 9].

ПРЕДЛОЖЕНИЕ 1. *Пусть f кодифференцируема в точке x_*, являющейся точкой локального минимума данной функции. Пусть также $Df(x_*)$ — произвольный кодифференциал функции f в данной точке. Тогда*

$$(0, \mathbf{0}_d) \in \underline{d}f(x_*) + \{(0, w)\} \quad \forall (0, w) \in \overline{d}f(x_*). \tag{7}$$

Более того, данное условие не зависит от выбора кодифференциала, т.е. если оно выполнено для одного кодифференциала функции f в точке x_*, то оно выполнено и для любого другого кодифференциала функции f в этой точке.

З а м е ч а н и е. Из равенств (2) следует, что для любого $(a, v) \in \underline{d}f(x_*)$ будет $a \leqslant 0$, в то время как для любого $(b, w) \in \overline{d}f(x_*)$ будет $b \geqslant 0$. Более того, найдётся по крайней мере одна пара $(b, w) \in \overline{d}f(x_*)$, для которой $b = 0$, т.е. множество $\{w \mid (0, w) \in \overline{d}f(x_*)\}$ в условии (7) заведомо не пусто.

Из теоремы Крейна–Мильмана и выпуклости множеств $\underline{d}f(x_*)$ и $\overline{d}f(x_*)$ следует, что включение в (7) достаточно проверять только для крайних точек множества $\overline{d}f(x_*)$.

СЛЕДСТВИЕ 1. *Необходимое условие минимума* (7) *выполняется тогда и только тогда, когда*

$$(0, \mathbf{0}_d) \in \underline{d}f(x_*) + \{(0, w)\} \quad \forall (0, w) \in \operatorname{ext} \overline{d}f(x_*), \tag{8}$$

где $\operatorname{ext} \overline{d}f(x_*)$ *— совокупность всех крайних точек множества* $\overline{d}f(x_*)$.

Точку x_*, удовлетворяющую условию (8), будем называть inf-стационарной точкой функции f. Согласно предложению 1 inf-стационарность заданной точки не зависит от выбора кодифференциала.

2°.Метод кодифференциального спуска. Перейдём к описанию метода нахождения inf-стационарных точек функции f — метода кодифференциального спуска. Везде далее мы предполагаем, что функция f непрерывно кодифференцируема, и задано некоторое кодифференциальное отображение Df, определённое на $\mathbb{R}^d$, и такое, что соответствующие многозначные отображения $x \mapsto \underline{d}f(x)$ и $x \mapsto \overline{d}f(x)$ непрерывны в метрике Хаусдорфа.

Зафиксируем параметр $\mu \in (0, +\infty]$, и для любого x введём множество

$$\overline{d}_\mu f(x) = \big\{(b, w) \in \overline{d}f(x) \bigm| b \leqslant \mu\big\},$$

называемое *усечённым гипердифференциалом* функции f в точке x. В.Ф. Демьяновым была предложена следующая теоретическая схема метода кодифференциального спуска (см. [4]):

1) Выберем начальное приближение $x_0 \in \mathbb{R}^d$.

2) Переход от n-го приближения к $(n+1)$-му осуществляется следующим образом ($n \geqslant 0$):

   (a) **Вычисление направлений спуска.** Для каждого $z = (b, w) \in \overline{d}_\mu f(x_n)$ вычислим

   $$(a_n(z), v_n(z)) = \arg\min\big\{a^2 + \|v\|^2 \bigm| (a, v) \in \underline{d}f(x_n) + z\big\}.$$

   (b) **Линейный поиск.** Для каждого $z = (b, w) \in \overline{d}_\mu f(x_n)$ найдём $\alpha_n(z) \geqslant 0$ такое, что

   $$\min_{\alpha \geqslant 0} f(x_n - \alpha v_n(z)) = f(x_n - \alpha_n(z) v_n(z)).$$

   (c) **Выбор среди всех найденных точек.** Найдём $z_n \in \overline{d}_\mu f(x_n)$ такое, что

   $$\min_{z \in \overline{d}_\mu f(x_n)} f\big(x_n - \alpha_n(z) v_n(z)\big) = f\big(x_n - \alpha_n(z_n) v_n(z_n)\big),$$

   и положим $x_{n+1} = x_n - \alpha_n(z_n) v_n(z_n)$.

Прокомментируем данный метод. Пусть уже найдено некоторое x_n. Согласно приведённой выше схеме для каждого элемента $z = (b, w)$ усечённого гипердифференциала $\overline{d}_\mu f(x_n)$ необходимо найти элемент $(a_n(z), v_n(z))$ множества $\underline{d}f(x_n) + z$ с минимальной нормой. Данный шаг непосредственно связан с необходимым условием экстремума из предложения 1. Если это условие выполнено в точке x_n, то для любого $z = (b, w) \in \overline{d}_\mu f(x_n)$ такого, что $b = 0$, будет $a_n(z) = 0$, $v_n(z) = \mathbf{0}_d$. Если же необходимое условие экстремума не выполнено, то найдётся по крайней мере одно $z = (0, w) \in \overline{d}_\mu f(x_n)$, для которого $(a_n(z), v_n(z)) \neq (0, \mathbf{0}_d)$. Можно показать, что в этом случае заведомо $v_n(z) \neq \mathbf{0}_d$, и направление $-v_n(z)$ является направлением спуска (не обязательно наискорейшего) функции f в точке x_n, то есть $f'(x_n, -v_n(z)) < 0$. Исходя из этого, может показаться, что в методе кодифференциального спуска стоит положить $\mu = 0$. Однако, такой выбор не позволяет гарантировать сходимость метода. Поэтому, помимо элементов $z = (b, w)$ усечённого гипердифференциала $\overline{d}_\mu f(x_n)$ для которых $b = 0$, необходимо также использовать элементы, для которых b отлично от нуля. При этом использование усечённого гипердифференциала $\overline{d}_\mu f(x_n)$ вместо $\overline{d} f(x_n)$ позволяет существенно сократить количество вычислений на каждой итерации, поскольку оно позволяет отбросить значительное количество точек из исходного гипердифференциала $\overline{d} f(x_n)$.

На втором шаге необходимо осуществить линейный поиск вдоль всех найденных направлений $-v_n(z)$. Следует особо отметить, что в общем случае линейный поиск осуществляется вдоль сразу *нескольких* направлений, даже если в точке x_n функция f непрерывно дифференцируема, что существенно отличает метод кодифференциального спуска от традиционных методов гладкой и негладкой оптимизации.

Как отмечалось выше, если в точке x_n не выполнено необходимое условие экстремума, то найдётся $z = (0, w) \in \overline{d}_\mu f(x_n)$, для которого направление $-v_n(z)$ будет направлением спуска функции f в точке x_n. Отсюда, в частности, следует, что $f(x_{n+1}) \leqslant f(x_n - \alpha_n(z) v_n(z)) < f(x_n)$, то есть метод кодифференциального спуска является релаксационным. С другой стороны, для некоторых направлений $z \in \overline{d}_\mu f(x_n)$ (в особенности это касается случая $b \neq 0$) направление $-v_n(z)$ вовсе не обязано быть направлением спуска функции f. Эта функция может сначала возрастать, а затем убывать в направлении $-v_n(z)$. Данное свойство позволяет методу кодифференциального спуска в некоторых случаях «выходить» из точек локального минимума с помощью подходящего выбора параметра μ. Численный пример проявления данной особенности метода имеется в работе [5].

Наконец, на третьем шаге, среди всех точек, найденных с помощью линейного поиска, необходимо найти ту точку, в которой функция f принимает наименьшее значение, и выбрать её в качестве следующего приближения. Таких точек, вообще говоря, может быть несколько. Выбор среди них осуществляется произвольным образом.

Если функция f непрерывно дифференцируема и её кодифференциал имеет вид (3), то метод кодифференциального спуска сводится к классическому методу наискорейшего спуска. Однако, в негладком случае прямой связи между наискорейшим спуском и методом кодифференциального спуска, вообще говоря, нет.

В случае функции максимума, т. е. когда $f(x) = \max_{1 \leqslant i \leqslant m} f_i(x)$, где f_i непрерывно дифференцируемы (см. пример 2), метод кодифференциального спуска принимает следующий вид:

1) Выберем начальное приближение $x_0 \in \mathbb{R}^d$.

2) Переход от n-го приближения к $(n+1)$-му осуществляется следующим образом ($n \geqslant 0$):

   (a) Вычислим

$$(a_n, v_n) = \arg\min \Big\{ a^2 + \|v\|^2 \ \Big| \\ (a, v) \in \operatorname{co}\big\{ (f_i(x_n) - f(x_n), \nabla f_i(x_n)) \mid 1 \leqslant i \leqslant m \big\} \Big\}.$$

   (b) Найдём $\alpha_n \geqslant 0$ такое, что

$$\min_{\alpha \geqslant 0} f(x_n - \alpha v_n) = f(x_n - \alpha_n v_n),$$

   и положим $x_{n+1} = x_n - \alpha_n v_n$.

Следует особо подчеркнуть, что приведённая выше схема метода кодифференциального спуска является именно *теоретической*, и её практическая реализация в случае, когда гипердифференциал $\overline{d} f(x_n)$ состоит более чем из одной точки, оказывается затруднительной. Действительно, даже для функции $f(x) = |x_1| - |x_2|$ из примера 1 по определению будет

$$\overline{d}_\mu f(x) = \begin{cases} \operatorname{co}\big\{(0,0,-1)^T, (\min\{2x_2, \mu\}, 0, 1)^T\big\}, & \text{если } x_2 \geqslant 0, \\ \operatorname{co}\big\{(0,0,1)^T, (\min\{-2x_2, \mu\}, 0, -1)^T\big\}, & \text{если } x_2 < 0. \end{cases}$$

Таким образом, если $x_2 \neq 0$, то усечённый гипердифференциал представляет из себя отрезок в пространстве $\mathbb{R}^3$, откуда следует, что в методе кодифференциального спуска для данной функции необходимо будет осуществлять линейный поиск вдоль континуума направлений.

Другой трудностью при реализации исходной схемы метода кодифференциального спуска является тот факт, что гиподифференциал $\underline{d} f(x)$, в отличие от гипердифференциала $\overline{d} f(x)$, не усекается, что существенно увеличивает трудоёмкость метода. Например, в задаче минимизации максимального уклонения из примера 3, в случае, когда изначально задано 10 точек в $\mathbb{R}^3$, и требуется найти 2 кластера (т.е. $m = 10$ и $\ell = 2$), гипердифференциал $\overline{d} g(x)$ будет выпуклой оболочкой 1024 точек в пространстве размерности 7, в то время как гиподифференциал $\underline{d} g(x)$ будет, вообще говоря, выпуклой оболочкой 5120 точек из $\mathbb{R}^7$.

Подобное соотношение между «размерами» гипо- и гипер-дифференциалов оказывается нередким в негладких задачах. Для того чтобы преодолеть эту трудность, в статье [5] было предложено рассматривать *усечённый кодифференциал*. Однако, определение усечённого кодифференциала из этой работы обладает недостатками, схожими с определением множества $\overline{d}_\mu f(x)$.

В [9] была предложена следующая модификация метода кодифференциального спуска, которая является более подходящей для практической реализации, чем исходный метод и метод усечённого кодифференциала из [5]. Зафиксируем параметры $\nu > 0$ и $\mu > 0$. Предположим, что заданы некоторые многозначные отображения $\underline{d}_\nu f$, $\overline{d}_\mu f \colon \mathbb{R}^d \rightrightarrows \mathbb{R} \times \mathbb{R}^d$, удовлетворяющие следующим условиям для всех $x \in \mathbb{R}^d$:

$$\Big\{(a, v) \in \operatorname{ext} \underline{d} f(x) \;\Big|\; a \geqslant -\nu\Big\} \subseteq \underline{d}_\nu f(x) \subseteq \underline{d} f(x), \tag{9}$$

$$\Big\{(b, w) \in \operatorname{ext} \overline{d} f(x) \;\Big|\; b \leqslant \mu\Big\} \subseteq \overline{d}_\mu f(x) \subseteq \overline{d} f(x). \tag{10}$$

Пару множеств $[\underline{d}_\nu f(x), \overline{d}_\mu f(x)]$ будем называть *усечённым кодифференциалом* функции f в точке x. Заметим, что множества $\underline{d}_\nu f(x)$ и $\overline{d}_\mu f(x)$ не обязаны быть выпуклыми или замкнутыми, а отображения $x \mapsto \underline{d}_\nu f(x)$ и $x \mapsto \overline{d}_\mu f(x)$ не обязаны быть (полу-)непрерывными. Однако, воспользовавшись условиями (9), (10), нетрудно проверить, что пара $[\operatorname{cl}\operatorname{co} \underline{d}_\nu f(x), \operatorname{cl}\operatorname{co} \overline{d}_\mu f(x)]$ является кодифференциалом функции f в точке x.

Теоретическая схема модифицированного метода кодифференциального спуска имеет следующий вид:

1) Выберем начальное приближение $x_0 \in \mathbb{R}^d$ и параметр линейного поиска $\alpha_* \in (0, +\infty]$.

2) Переход от n-го приближения к $(n+1)$-му осуществляется следующим образом ($n \geqslant 0$):

   (a) **Вычисление направлений спуска.** Для каждого $z = (b, w) \in \overline{d}_\mu f(x_n)$ вычислим

   $$(a_n(z), v_n(z)) = \arg\min \big\{ a^2 + \|v\|^2 \;\big|\; (a, v) \in \operatorname{cl}\operatorname{co}\{\underline{d}_\nu f(x_n)\} + z \big\}.$$

   (b) **Линейный поиск.** Для каждого $z = (b, w) \in \overline{d}_\mu f(x_n)$ найдём $\alpha_n(z) \geqslant 0$ такое, что

   $$\min_{\alpha \in [0, \alpha_*]} f(x_n - \alpha v_n(z)) = f(x_n - \alpha_n(z) v_n(z)).$$

(c) **Выбор среди всех найденных точек.** Найдём $z_n \in \overline{d}_\mu f(x_n)$ такое, что

$$\min_{z \in \overline{d}_\mu f(x_n)} f\big(x_n - \alpha_n(z) v_n(z)\big) = f\big(x_n - \alpha_n(z_n) v_n(z_n)\big),$$

и положим $x_{n+1} = x_n - \alpha_n(z_n) v_n(z_n)$.

Главными отличиями модифицированного метода кодифференциального спуска от исходного метода является другое определение усечённого гипердифференциала $\overline{d}_\mu f(x_n)$, а также использование усечённого гиподифференциала $\underline{d}_\nu f(x_n)$ для вычисления направлений спуска, вместо гиподифференциала $\underline{d} f(x_n)$. В остальном эти методы совпадают.

Для практического применения модифицированного метода кодифференциального спуска необходимо построить многозначные отображения $\underline{d}_\nu f$ и $\overline{d}_\mu f$, удовлетворяющие условиям (9), (10). Укажем, как это можно сделать.

Во всех численных примерах множества $\underline{d} f(x)$ и $\overline{d} f(x)$ представляют из себя выпуклые оболочки конечного числа точек. Если

$$\overline{d} f(x) = \operatorname{co}\{(b_i, w_i) \in \mathbb{R} \times \mathbb{R}^d \mid 1 \leqslant i \leqslant s\}$$

для некоторых (b_i, w_i), то естественно положить

$$\overline{d}_\mu f(x) = \big\{(b_i, w_i) \mid b_i \leqslant \mu,\ 1 \leqslant i \leqslant s\big\}.$$

Множество $\underline{d}_\nu f(x)$ определяется аналогичным образом. На первый взгляд, следующее определение множества $\overline{d}_\mu f(x)$ может показаться более естественным:

$$\overline{d}_\mu f(x) = \Big\{(b, w) \in \operatorname{ext} \overline{d} f(x) \Bigm| b \leqslant \mu\Big\}.$$

Однако, для того чтобы воспользоваться данным определением на практике, необходимо найти все крайние точки множества $\overline{d} f(x)$, что зачастую является очень трудоёмкой процедурой. Поэтому в (9), (10) данное равенство заменено на включение, что позволяет, при необходимости, избегать процедуры нахождения крайних точек.

Наконец, отметим, что для функции $f(x) = |x_1| - |x_2|$ из примера 1 можно положить

$$\overline{d}_\mu f(x) = \begin{cases} \big\{(0,0,-1)^T, (2x_2, 0, 1)^T\big\}, & \text{если } 0 \leqslant x_2 \leqslant \mu/2, \\ \big\{(0,0,-1)^T\big\}, & \text{если } x_2 > \mu/2, \\ \big\{(0,0,1)^T, (-2x_2, 0, -1)^T\big\}, & \text{если } -\mu/2 \leqslant x_2 < 0, \\ \big\{(0,0,1)^T\big\}, & \text{если } x_2 < -\mu/2. \end{cases}$$

При таком выборе усечённого гипердифференциала на каждой итерации модифицированного метода кодифференциального спуска придётся осуществлять линейный поиск не более чем вдоль двух направлений.

Приведём теорему о сходимости метода. Для этого нам потребуется вспомогательное определение локально равномерной кодифференцируемости. Вследствие неединственности кодифференциала данное определение удобнее дать для зафиксированного ранее непрерывного кодифференциального отображения Df.

ОПРЕДЕЛЕНИЕ 2. Будем говорить, что кодифференциальное отображение Df локально равномерно аппроксимирует функцию f, если для любой точки $x_0 \in \mathbb{R}^d$ найдётся окрестность $U(x_0)$ такая, что в разложении

$$f(x+h) - f(x) = \max_{(a,v)\in \underline{d}f(x)} (a + \langle v, h\rangle) + \min_{(b,w)\in \overline{d}f(x)} (b + \langle w, h\rangle) + o(\|h\|)$$

будет $o(\|h\|)/\|h\| \to 0$ при $\|h\| \to 0$ равномерно по всем $x \in U(x_0)$.

Можно показать (см. [9]), что класс локально равномерно и непрерывно кодифференцируемых функций содержит в себе класс непрерывно дифференцируемых функций и замкнут относительно всех алгебраических операций, операций взятия поточечного максимума и минимума конечного числа функций и композиции с непрерывно дифференцируемыми функциями. Более того, приведённые выше правила кодифференциального исчисления сохраняют свойство локальной равномерной кодифференцируемости.

ТЕОРЕМА 1. *[9] Пусть функция f ограничена снизу, непрерывное кодифференциальное отображение Df локально равномерно аппроксимирует функцию f, а последовательность $\{x_n\}$ построена по модифицированному методу кодифференциального спуска. Тогда любая предельная точка последовательности $\{x_n\}$ является* $\inf$*-стационарной.*

Доказательство этой теоремы опирается на целый ряд вспомогательных результатов о кодифференцируемых функциях, и мы его здесь не приводим.

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

ЛИТЕРАТУРА

1. Demyanov V. F. *Continuous generalized gradients for nonsmooth functions.* In: A. Kurzhanski, K. Neumann, and D. Pallaschke, editors, *Optimization, Parallel Processing and Applications*, pp. 24–27. Springer, Berlin, Heidelberg, 1988.

2. Демьянов В. Ф. *О кодифференцируемых функциях* // Вестн. Ленингр. ун-та, 1988, т. 2, № 8, с. 22–26.

3. Demyanov V. F. *Smoothness of nonsmooth functions*. In: F.H. Clarke, V.F. Demyanov, and F. Giannessi, editors, *Nonsmooth Optimization and Related Topics*, pp. 79–88. Springer, Boston, 1989.

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

5. Demyanov V. F., Bagirov A. M., Rubinov A. M. *A method of truncated codifferential with applications to some problems of cluster analysis* // J. Glob. Optim., 2002, vol. 23, no. 1, pp. 63–80.

6. Bagirov A. M., Nazari Ganjehlou A., Ugon J., Tor A. H. *Truncated codifferential method for nonsmooth convex optimization* // Pacific J. Optim., 2010, vol. 6, no. 3, pp. 483–496.

7. Bagirov A. M., Ugon J. *Codifferential method for minimizing nonsmooth DC functions* // J. Glob. Optim., 2011, vol. 50, no. 1, pp. 3–22.

8. Tor A. H., Bagirov A., Karasözen B. *Aggregate codifferential method for nonsmooth dc optimization* // J. Comput. Appl. Math., 2014, vol. 259, pp. 852–867.

9. Dolgopolik M. V. *A convergence analysis of the method of codifferential descent* // Comput. Optim. Appl., 2018, vol. 71, no. 3, pp. 879–913.

10. Dolgopolik M. V. *Abstract convex approximations of nonsmooth functions* // Optimization, 2015, vol. 64, no. 7, pp. 1439–1469.

11. Demyanov A. V., Demyanov V. F., Malozemov V. N. *Minmaxmin problems revisited* // Optim. Methods Softw., 2002, vol. 17, no. 5, pp. 783–804

12. Demyanov A. *On the solution of min-sum-min problems* // J. Glob. Optim., 2005, vol. 31, no. 3, pp. 437–453.

13. Bagirov A. M., Rubinov A. M., Soukhoroukova N. V., Yearwood J. *Unsupervised and supervised data classification via nonsmooth and global optimization* // Top, 2003, vol. 11, no. 1, pp. 1–75.

14. Jain A. K., Murty M. N., Flynn P. J. *Data clustering: a review* // ACM Computing Surveys, 1999, vol. , no. 3, pp. 264–323.

15. Кольцов М. А. *Решение задачи Сильвестра в MATLAB* [Электронный ресурс] // Семинар «CNSA & NDO» [Офиц. сайт]. Избранные доклады. 26 февраля 2015 г. (`http://www.apmath.spbu.ru/cnsa/reps15.shtml#0226`)