ОПТИМИЗАЦИЯ В ПОЛИЭДРАЛЬНЫХ НОРМАХ*

В. Н. Малозёмов
v.malozemov@spbu.ru

23 марта 2023 г.

В начале 1980-х годов мы с Б. П. Дербеневой занимались оптимизационными задачами, в которых использовались полиэдральные нормы. На эту деятельность нас вдохновили публикации Андерсона и Осборна [1, 2]. В 1981 и 1982 годах мы опубликовали две свои работы на данную тему [3, 4]. В 1983 году Б. П. Дербенева защитила кандидатскую диссертацию, посвященную оптимизационной задаче, целевая функция в которой зависит от несимметричной ℓ^1-нормы [5].

В подготовленном мной докладе приводятся начальные сведения о полиэдральных нормах и их применении в оптимизации.

1°. Пусть $c_1, c_2, \ldots, c_p$ — векторы в $\mathbb{R}^m$, такие, что множество

$$V = \big\{y \in \mathbb{R}^m \mid \langle c_k, y\rangle \leqslant 1 \ \text{ для всех } \ k \in 1:p\big\} \tag{1}$$

ограничено. Введем функцию Минковского

$$\mathcal{D}(y) = \inf\{\lambda > 0 \mid y/\lambda \in V\}.$$

Учитывая определение V, будем называть функцию $\mathcal{D}(y)$ *полиэдральной нормой* в $\mathbb{R}^m$.

Рассмотрим в $\mathbb{R}^n$ вектор-функцию $F(x) = \big(f_1(x), f_2(x), \ldots, f_m(x)\big)$ и поставим экстремальную задачу

$$\mathcal{D}\big(F(x)\big) \to \inf_{x \in \mathbb{R}^n}. \tag{2}$$

Покажем, что задача (2) сводится к минимаксной задаче.

ЛЕММА 1. *Справедливо равенство*

$$\mathcal{D}(y) = \max_{k \in 1:p} \langle c_k, y\rangle. \tag{3}$$

*Семинар по оптимизации, машинному обучению и искусственному интеллекту «O&ML» `http://oml.cmlaboratory.com/`

Д о к а з а т е л ь с т в о. Если $y = \mathbf{0}$, то утверждение тривиально. Допустим, что $y \neq \mathbf{0}$. Тогда $\langle c_k, y\rangle > 0$ хотя бы при одном k. Действительно, иначе $\langle c_k, y\rangle \leqslant 0$ и $\langle c_k, ty\rangle \leqslant 0 \leqslant 1$ при всех $k \in 1:p$ и всех $t > 0$. Но это противоречит ограниченности множества V. Теперь в силу (1) имеем

$$\mathcal{D}(y) = \inf\bigl\{\lambda > 0 \mid \langle c_k, y\rangle \leqslant \lambda \ \ \forall k \in 1:p\bigr\} = \max_{k\in 1:p}\langle c_k, y\rangle.$$

Лемма доказана. □

Для справедливости равенства (3) существенна ограниченность множества V. Сформулируем критерий ограниченности V в терминах c_k. Обозначим через V^* выпуклую оболочку векторов $c_1, c_2, \ldots, c_p$.

ЛЕММА 2. *Для того чтобы множество V было ограниченным, необходимо и достаточно, чтобы выполнялось условие $\mathbf{0} \in \operatorname{int} V^*$.*

Д о к а з а т е л ь с т в о. Н е о б х о д и м о с т ь. Допустим противное. По теореме отделимости найдется вектор $g \neq \mathbf{0}$, такой, что $\langle y, g\rangle \leqslant 0$ при всех $y \in V^*$. В частности, $\langle c_k, g\rangle \leqslant 0$ при всех $k \in 1:p$. Отсюда следует, что $tg \in V$ при всех $t > 0$. Но это противоречит ограниченности множества V.
Д о с т а т о ч н о с т ь. Зафиксируем y из V. Имеем $\langle c_k, y\rangle \leqslant 1$ при всех $k \in 1:p$. Отсюда следует, что $\langle z, y\rangle \leqslant 1$ при всех $z \in V^*$. Обозначим $B_\delta = \bigl\{z \in \mathbb{R}^m \mid \|z\| \leqslant \delta\bigr\}$, где $\|z\|$ — евклидова норма вектора z. По условию теоремы, $B_\delta \subset V^*$ при некотором $\delta > 0$. Значит, $\langle z, y\rangle \leqslant 1$ при всех $z \in B_\delta$. Подставив в это неравенство $z = \delta y/\|y\|$, получим $\|y\| \leqslant 1/\delta$, что гарантирует ограниченность множества V.

Лемма доказана. □

З а м е ч а н и е. На самом деле, из доказательства леммы 1 следует, что для ее справедливости достаточно, чтобы для любого ненулевого вектора $y \in \mathbb{R}^m$ при некотором $k \in 1:p$ выполнялось неравенство $\langle c_k, y\rangle > 0$.

2°. На основании леммы 1 получаем

$$\mathcal{D}\bigl(F(x)\bigr) = \max_{k\in 1:p}\bigl\langle c_k, F(x)\bigr\rangle.$$

Пусть $c_k = (c_{k1}, c_{k2}, \ldots, c_{km})$. Тогда задача (2) примет вид

$$\varphi(x) := \max_{k\in 1:p}\sum_{i=1}^{m} c_{ki} f_i(x) \to \inf_{x\in\mathbb{R}^n}. \tag{4}$$

Задача (4) — это минимаксная задача. Для нее справедливы все результаты минимаксной теории [6]. Одним из принципиальных результатов является дифференцируемость функции максимума $\varphi(x)$ по всем возможным направлениям в любой точке $x_0 \in \mathbb{R}^n$, в которой $\varphi(x_0) > 0$, при условии, что все функции $f_i(x)$, $i \in 1:m$, непрерывно дифференцируемы в окрестности этой точки. Чтобы записать формулу для производной по направлению, введем индексное множество

$$R(x_0) = \Big\{k \in 1:p \mid \sum_{i=1}^{m} c_{ki} f_i(x_0) = \varphi(x_0)\Big\}.$$

Для любого вектора $g \in \mathbb{R}^n$ с единичной евклидовой нормой имеем [6, с. 71]

$$\frac{\partial\varphi(x_0)}{\partial g} = \max_{k \in R(x_0)} \Big\{\sum_{i=1}^{m} c_{ki} \big\langle f_i'(x_0), g\big\rangle\Big\}. \tag{5}$$

3°. Приведем три примера полиэдральных норм. Они будут различаться определяющим набором нормалей $c_1, c_2, \ldots, c_p$.

ПРИМЕР 1. Пусть $p = 2m$ и система $\{c_k\}$ состоит из векторов

$$e_1, -e_1, e_2, -e_2, \ldots, e_m, -e_m,$$

где e_k — k-й орт в пространстве $\mathbb{R}^m$. Покажем, что в этом случае при $y = (y_1, y_2, \ldots, y_m)$ справедлива формула

$$\mathcal{D}(y) = \max_{i \in 1:m} |y_i|. \tag{6}$$

При $y = \mathbf{0}$ равенство (6) тривиально, поэтому считаем, что $y \neq \mathbf{0}$. Пусть $y_{i_0} \neq 0$. Возьмем $c_k = e_{i_0} \operatorname{sign} y_{i_0}$. Имеем

$$\langle c_k, y\rangle = |y_{i_0}| > 0.$$

По замечанию к лемме 2 и лемме 1 верна формула (3). Значит, проверка равенства (6) сводится к проверке равенства

$$\max_{k \in 1:p} \langle c_k, y\rangle = \max_{i \in 1:m} |y_i|. \tag{7}$$

В силу определения нормалей c_k формулу (7) можно переписать в виде

$$\max\{y_1, -y_1, y_2, -y_2, \ldots, y_m, -y_m\} = \max_{i \in 1:m} |y_i|.$$

Справедливость последнего равенства очевидна.

Полиэдральная норма (6) называется *чебышевской*.

ПРИМЕР 2. Пусть $p = 2^m$ и все компоненты всех нормалей c_k равны $+1$ или -1. Покажем, что в этом случае

$$\mathcal{D}(y) = \sum_{i=1}^{m} |y_i|. \tag{8}$$

При $y = \mathbf{0}$ равенство (8) тривиально. Зафиксируем вектор $y \neq \mathbf{0}$. Возьмем нормаль c_{k_0} с компонентами

$$c_{k_0 i} = \begin{cases} 1, & \text{если } y_i \geqslant 0, \\ -1, & \text{если } y_i < 0. \end{cases}$$

Запишем

$$\langle c_{k_0}, y \rangle = \sum_{i=1}^{m} c_{k_0 i} y_i = \sum_{i=1}^{m} |y_i|. \tag{9}$$

В частности, $\langle c_{k_0}, y \rangle > 0$. По замечанию к лемме 2 и лемме 1 верна формула (3). Значит, проверка равенства (8) сводится к проверке равенства

$$\max_{k \in 1:p} \langle c_k, y \rangle = \sum_{i=1}^{m} |y_i|. \tag{10}$$

В силу определения нормалей c_k при всех $k \in 1:p$ имеем

$$\langle c_k, y \rangle \leqslant \sum_{i=1}^{m} |y_i|. \tag{11}$$

Равенство (10) является следствием соотношений (11) и (9).

Полиэдральная норма (8) называется *ℓ^1-нормой.*

ПРИМЕР 3. Пусть $p = 2^m$ и $\lambda \in (0,1)$ — некоторый параметр. Будем считать, что все компоненты всех нормалей c_k равны λ или $-(1-\lambda)$. Покажем, что в этом случае

$$\mathcal{D}(y) = \sum_{i=1}^{m} |y_i|_\lambda, \tag{12}$$

где

$$|y_i|_\lambda = \max\{\lambda y_i, -(1-\lambda) y_i\} = \begin{cases} \lambda y_i, & \text{если } y_i \geqslant 0, \\ -(1-\lambda) y_i, & \text{если } y_i < 0. \end{cases}$$

Введем обозначение

$$\operatorname{sign}_\lambda(y_i) = \begin{cases} \lambda, & \text{если } y_i \geqslant 0, \\ -(1-\lambda), & \text{если } y_i < 0. \end{cases}$$

Нетрудно проверить, что
$$|y_i|_\lambda = y_i \operatorname{sign}_\lambda(y_i).$$

При $y = \mathbf{0}$ равенство (12) тривиально. Зафиксируем вектор $y \neq \mathbf{0}$. Возьмем нормаль c_{k_0} с компонентами $c_{k_0 i} = \operatorname{sign}_\lambda(y_i)$. Запишем

$$\langle c_{k_0}, y\rangle = \sum_{i=1}^m y_i \operatorname{sign}_\lambda(y_i) = \sum_{i=1}^m |y_i|_\lambda. \tag{13}$$

В частности, $\langle c_{k_0}, y\rangle > 0$. По замечанию к лемме 2 и лемме 1 верна формула (3). Значит, проверка равенства (12) сводится к проверке равенства

$$\max_{k\in 1:p}\langle c_k, y\rangle = \sum_{i=1}^m |y_i|_\lambda. \tag{14}$$

В силу определения нормалей c_k при всех $k \in 1:p$ имеем

$$\langle c_k, y\rangle = \sum_{i=1}^m c_{ki} y_i \leqslant \sum_{i=1}^m \max\bigl\{\lambda y_i, -(1-\lambda)y_i\bigr\} = \sum_{i=1}^m |y_i|_\lambda. \tag{15}$$

Равенство (14) является следствием соотношений (15) и (13).

Полиэдральная норма (12) называется *несимметричной* ℓ^1 *нормой*. Для такой нормы при λ, близком к единице, более значимыми являются положительные компоненты вектора y.

4°. Вернемся к формуле (5) для производной по направлению и приведем ее детальный анализ в случае ℓ^1-нормы. Запишем начальную информацию:

$$\varphi(x) = \sum_{i=1}^m \bigl|f_i(x)\bigr|,$$
$$\frac{\partial\varphi(x_0)}{\partial g} = \max_{k\in R(x_0)}\Bigl\{\sum_{i=1}^m c_{ki}\bigl\langle f_i'(x_0), g\bigr\rangle\Bigr\}. \tag{16}$$

Здесь все коэффициенты c_{ki} равны $+1$ или -1 и

$$R(x_0) = \Bigl\{k \in 1:2^m \mid \sum_{i=1}^m c_{ki} f_i(x_0) = \sum_{i=1}^m \bigl|f_i(x_0)\bigr|\Bigr\}.$$

Отметим, что при всех k, i справедливо неравенство

$$c_{ki} f_i(x_0) \leqslant \bigl|f_i(x_0)\bigr|. \tag{17}$$

Отсюда следует, что k принадлежит $R(x_0)$ тогда и только тогда, когда при всех $i \in 1:m$ неравенство (17) выполняется как равенство. Введем множества

$$I(x_0) = \bigl\{i \in 1:m \mid f_i(x_0) \neq 0\bigr\}, \quad I_0(x_0) = \bigl\{i \in 1:m \mid f_i(x_0) = 0\bigr\}. \tag{18}$$

С их помощью критерий принадлежности индекса k множеству $R(x_0)$ можно переформулировать так: для справедливости включения $k \in R(x_0)$ необходимо и достаточно, чтобы компоненты нормали c_k имели вид

$$c_{ki} = \begin{cases} \operatorname{sign} f_i(x_0) & \text{при } i \in I(x_0), \\ \pm 1 & \text{при } i \in I_0(x_0). \end{cases} \tag{19}$$

В дальнейшем будем использовать обозначение $\xi_i = \operatorname{sign} f_i(x_0)$.

Обратимся к формуле (16). Возьмем $k \in R(x_0)$ и запишем

$$\sum_{i=1}^{m} c_{ki}\bigl\langle f_i'(x_0), g\bigr\rangle = \sum_{i \in I(x_0)} c_{ki}\bigl\langle f_i'(x_0), g\bigr\rangle + \sum_{i \in I_0(x_0)} c_{ki}\bigl\langle f_i'(x_0), g\bigr\rangle \leqslant$$
$$\leqslant \sum_{i \in I(x_0)} \xi_i\bigl\langle f_i'(x_0), g\bigr\rangle + \sum_{i \in I_0(x_0)} \Bigl|\bigl\langle f_i'(x_0), g\bigr\rangle\Bigr|. \tag{20}$$

На индексе k_0 из $R(x_0)$, которому соответствует нормаль c_{k_0} с компонентами

$$c_{k_0 i} = \begin{cases} \xi_i & \text{при } i \in I(x_0), \\ \operatorname{sign}\bigl\langle f_i'(x_0), g\bigr\rangle & \text{при } i \in I_0(x_0) \text{ и } \bigl\langle f_i'(x_0), g\bigr\rangle \neq 0, \\ 1 & \text{при } i \in I_0(x_0) \text{ и } \bigl\langle f_i'(x_0), g\bigr\rangle = 0 \end{cases}$$

предыдущее неравенство выполняется как равенство. Значит,

$$\max_{k \in R(x_0)} \Bigl\{\sum_{i=1}^{m} c_{ki}\bigl\langle f_i'(x_0), g\bigr\rangle\Bigr\} = \sum_{i \in I(x_0)} \xi_i\bigl\langle f_i'(x_0), g\bigr\rangle + \sum_{i \in I_0(x_0)} \Bigl|\bigl\langle f_i'(x_0), g\bigr\rangle\Bigr|.$$

Приходим к следующему результату.

ТЕОРЕМА 1. *Пусть функции $f_1(x), f_2(x), \ldots, f_m(x)$ непрерывно дифференцируемы в некоторой окрестности точки x_0 и функция*

$$\varphi(x) = \sum_{i=1}^{m} \bigl|f_i(x)\bigr|$$

положительны при $x = x_0$. Тогда для производной функции $\varphi(x)$ в точке $x = x_0$ по любому направлению g с евклидовой нормой, равной единице, справедлива формула

$$\frac{\partial \varphi(x_0)}{\partial g} = \sum_{i \in I(x_0)} \xi_i\bigl\langle f_i'(x_0), g\bigr\rangle + \sum_{i \in I_0(x_0)} \Bigl|\bigl\langle f_i'(x_0), g\bigr\rangle\Bigr|,$$

где индексные множества $I(x_0)$ и $I_0(x_0)$ определены записью (18) и $\xi_i = \operatorname{sign} f_i(x_0)$.

5°. В общем случае формулу (5) для производной по направлению можно переписать так:

$$\frac{\partial\varphi(x_0)}{\partial g} = \max_{k\in R(x_0)} \langle Z_k, g\rangle, \tag{21}$$

где

$$Z_k = \sum_{i=1}^{m} c_{ki} f_i'(x_0).$$

Если x_0 — точка локального минимума в задаче (4), то необходимо

$$\min_{g\in S} \frac{\partial\varphi(x_0)}{\partial g} \geqslant 0. \tag{22}$$

Здесь $S = \big\{g \in \mathbb{R}^n \mid \|g\| = 1\big\}$. С учетом (21) необходимое условие локального минимума (22) принимает вид

$$\min_{g\in S} \max_{k\in R(x_0)} \langle Z_k, g\rangle \geqslant 0. \tag{23}$$

Обозначим через $C(x_0)$ выпуклую оболочку векторов Z_k, $k \in R(x_0)$. Как показано в [6, с. 77], неравенство (23) равносильно включению

$$\mathbf{0} \in C(x_0). \tag{24}$$

Выясним, как выглядит необходимое условие локального минимума (24) в случае ℓ^1-нормы.

Введем обозначения

$$z(x_0) = \sum_{i\in I(x_0)} \xi_i f_i'(x_0),$$

$$P(x_0) = \Big\{z = \sum_{i\in I_0(x_0)} \beta_i f_i'(x_0) \mid \beta_i \in [-1,1] \ \ \forall i \in I_0(x_0)\Big\}.$$

ТЕОРЕМА 2. *При ℓ^1-норме необходимое условие локального минимума* (22) *эквивалентно включению*

$$z(x_0) \in P(x_0).$$

Доказательство. Напомним описание множества $R(x_0)$ при ℓ^1-норме. Индекс k принадлежит $R(x_0)$ тогда и только тогда, когда ему соответствует нормаль c_k с компонентами

$$c_{ki} = \begin{cases} \operatorname{sign} f_i(x_0) & \text{при } i \in I(x_0), \\ \pm 1 & \text{при } i \in I_0(x_0). \end{cases}$$

Понятно, что $\big|R(x_0)\big| = 2^{|I_0(x_0)|}$. □

При $k \in R(x_0)$ имеем

$$Z_k := \sum_{i=1}^{m} c_{ki} f_i'(x_0) = \sum_{i \in I(x_0)} \xi_i f_i'(x_0) + \sum_{i \in I_0(x_0)} c_{ki} f_i'(x_0). \tag{25}$$

Обозначим

$$z_k = \sum_{i \in I_0(x_0)} c_{ki} f_i'(x_0).$$

Равенство (25) примет компактный вид

$$Z_k = z(x_0) + z_k, \quad k \in R(x_0). \tag{26}$$

Напомним, что через $C(x_0)$ обозначена выпуклая оболочка векторов Z_k, $k \in R(x_0)$. Введем еще выпуклую оболочку $c(x_0)$ векторов z_k, $k \in R(x_0)$. Из (26) следует, что

$$C(x_0) = z(x_0) + c(x_0). \tag{27}$$

Покажем, что $c(x_0) = P(x_0)$.

Возьмем точку z из $c(x_0)$. Запишем

$$z = \sum_{k \in R(x_0)} \alpha_k z_k = \sum_{k \in R(x_0)} \alpha_k \sum_{i \in I_0(x_0)} c_{ki} f_i'(x_0) =$$
$$= \sum_{i \in I_0(x_0)} f_i'(x_0) \sum_{k \in R(x_0)} \alpha_k c_{ki}.$$

Здесь коэффициенты α_k неотрицательны и в сумме равны единице. Обозначим

$$\beta_i = \sum_{k \in R(x_0)} \alpha_k c_{ki}, \quad i \in I_0(x_0).$$

Получим

$$z = \sum_{i \in I_0(x_0)} \beta_i f_i'(x_0). \tag{28}$$

Напомним, что c_{ki} при $k \in R(x_0)$ и $i \in I_0(x_0)$ равны $+1$ или -1, поэтому $\beta_i \in [-1, 1]$ при всех $i \in I_0(x_0)$. Равенство (28) гарантирует, что $z \in P(x_0)$. Установлено включение $c(x_0) \subset P(x_0)$.

Для проверки обратного включения $P(x_0) \subset c(x_0)$ достаточно выяснить, что система линейных уравнений

$$\sum_{k \in R(x_0)} c_{ki} \alpha_k = \beta_i, \;\; i \in I_0(x_0),$$
$$\sum_{k \in R(x_0)} \alpha_k = 1$$

при $\beta_i \in [-1, 1]$ имеет неотрицательное решение. Здесь мы воспользуемся известным результатом [7, с. 24]: *система $Ax = b$ с произвольной матрицей A имеет неотрицательное решение $x \geqslant \mathbf{0}$ тогда и только тогда, когда для любого u, удовлетворяющего условию $A^T u \geqslant \mathbf{0}$, выполняется неравенство* $\langle b, u \rangle \geqslant 0$. В нашем случае нужно доказать, что любое решение системы неравенств

$$\sum_{i \in I_0(x_0)} c_{ki} v_i + u \geqslant 0, \quad k \in R(x_0), \tag{29}$$

удовлетворяет условию

$$\sum_{i \in I_0(x_0)} \beta_i v_i + u \geqslant 0. \tag{30}$$

Возьмем какое-нибудь решение системы (29). В частности, будет выполняться неравенство

$$\sum_{i \in I_0(x_0)} c_{k_0 i} v_i + u \geqslant 0 \tag{31}$$

на индексе $k_0 \in R(x_0)$, которому соответствует нормаль c_{k_0} с компонентами

$$c_{k_0 i} = \begin{cases} \xi_i & \text{при } i \in I(x_0), \\ -\operatorname{sign} v_i & \text{при } i \in I_0(x_0) \text{ и } v_i \neq 0, \\ -1 & \text{при } i \in I_0(x_0) \text{ и } v_i = 0. \end{cases}$$

Отметим, что

$$c_{k_0 i} v_i = -|v_i| \leqslant \beta_i v_i \quad \text{при} \quad i \in I_0(x_0).$$

На основании (31) получим

$$\sum_{i \in I_0(x_0)} \beta_i v_i + u \geqslant \sum_{i \in I_0(x_0)} c_{k_0 i} v_i + u \geqslant 0.$$

Неравенство (30) доказано.

Как следствие, справедливо включение $P(x_0) \subset c(x_0)$, которое вместе с доказанным ранее обратным включением приводит к равенству $c(x_0) = P(x_0)$.

Перепишем формулу (27) в виде

$$C(x_0) = z(x_0) + P(x_0).$$

Очевидно, что $P(x) = -P(x)$, поэтому

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

Остается отметить, что необходимое условие локального минимума $\mathbf{0} \in C(x_0)$ эквивалентно включению

$$z(x_0) \in P(x_0).$$

Теорема доказана. $\square$

ЛИТЕРАТУРА

1. Anderson D. H., Osborne M. R. *Discrete, linear approximation in polyhedral norms* // Numer. Math., 1976. Vol. 26. N. 2. P. 179–189.

2. Anderson D. H., Osborne M. R. *Discrete, non-linear approximation in polyhedral norms* // Numer. Math., 1977. Vol. 28. N. 2. P. 143–156.

3. Дербенева Б. П., Малозёмов В. Н. *Оптимизация в полиэдральных нормах* // Вестник ЛГУ. Сер. 1. 1981. Вып. 4 (№ 19). С. 15–19.

4. Дербенева Б. П., Малозёмов В. Н. *Градиентный метод для дискретной аппроксимации в норме* ℓ^1 // Вестник ЛГУ. Сер. 1. 1982. Вып. 3 (№ 13). С. 19–23.

5. Дербенева Б. П. *Оптимизация в полиэдральной норме* / Автореферат канд. дисс. ЛГУ. 1983.

6. Демьянов В. Ф., Малозёмов В. Н. *Введение в минимакс.* М.: Наука, 1972. 368 с.

7. Гавурин М. К., Малозёмов В. Н. *Экстремальные задачи с линейными ограничениями.* Л.: Изд-во ЛГУ, 1984. 176 с.