ОСНОВЫ ТЕОРИИ ТОЧНЫХ ШТРАФНЫХ ФУНКЦИЙ*

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

24 декабря 2015 г.

Аннотация. В докладе обсуждаются базовые результаты теории точных штрафных функций для задачи математического программирования. В частности, приводятся определения локальной и глобальной точности штрафной функции. Доказывается простое достаточное условие для локальной точности штрафной функции в терминах выполнения условия регулярности ограничений, и обсуждается связь между глобальной точностью штрафной функции и её локальной точностью в оптимальных планах исходной задачи. Также рассматривается один из возможных подходов к построению *гладких* точных штрафных функций.

1°. Локальная точность штрафной функции. В данном докладе мы будем обсуждать теорию точных штрафных функций для задачи математического программирования вида

$$f(x) \to \min, \tag{1}$$
$$g_i(x) \leqslant 0, \quad i \in I := 1 \colon m, \tag{2}$$
$$h_j(x) = 0, \quad j \in J := 1 \colon \ell. \tag{3}$$

Здесь и далее мы предполагаем, что функции f, g_i, $i \in I$, и h_j, $j \in J$, определены и непрерывны на $\mathbb{R}^n$, и существует оптимальный план задачи (1)–(3). Обозначим через Ω множество планов данной задачи.

Один из возможных подходов к решению задачи (1)–(3) заключается в построении некоторой вспомогательной функции $F(x, \lambda)$ (здесь λ — параметр), множество точек глобального минимума на $\mathbb{R}^n$ которой совпадает с множеством оптимальных планов задачи (1)–(3) при некоторых значениях параметра λ. Функции $F(x, \lambda)$ такого типа называют *точными отделяющими функциями.* Данное название связно с тем, что любая теорема существования подоб-

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

ного рода функций является, по существу, обобщением теоремы об отделимости для выпуклых множеств. Более подробно по поводу отделяющих функций см. [1].

Если функция $F(x, \lambda)$ построена, то, решая задачу минимизации этой функции на всём пространстве $\mathbb{R}^n$, мы сможем найти решения исходной задачи. Таким образом, данный подход позволяет в некоторых случаях свести (по крайней мере, с теоретической точки зрения) задачу условной оптимизации (1)–(3) к задаче безусловной минимизации вспомогательной функции $F(x, \lambda)$.

В рамках теории точных штрафных функций изучаются функции $F(x, \lambda)$ вида

$$F(x, \lambda) = f(x) + \lambda\varphi(x),$$

где $\lambda \geqslant 0$, а φ — неотрицательная функция, определённая на $\mathbb{R}^n$, и такая, что

$$\varphi(x) = 0 \iff x \in \Omega,$$

то есть $\varphi(x) = 0$ тогда и только тогда, когда $g_i(x) \leqslant 0$, $i \in I$, и $h_j(x) = 0$, $j \in J$. Функция $F(x, \lambda)$ при этом называется *штрафной функцией*, множитель λ называется *штрафным параметром*, а функция φ называется *штрафным членом*.

Из определения штрафной функции следует, что $F(x, 0) = f(x)$, а для любого $\lambda > 0$ функция $F(x, \lambda)$ совпадает с целевой функцией $f(x)$, если вектор x является планом задачи (1)–(3), и $F(x, \lambda) > f(x)$ в противном случае. Кроме того, функция $F(x, \lambda)$ не убывает по λ и $F(x, \lambda) \to +\infty$ при $\lambda \to +\infty$ для любого $x \notin \Omega$.

Прежде чем перейти к основным определениям и результатам теории точных штрафных функций, заметим, что для того чтобы точка глобального минимума штрафной функции $F(x, \lambda) = f(x) + \lambda\varphi(x)$ была оптимальным планом задачи (1)–(3), в общем случае необходимо, чтобы функция φ была негладкой.

Действительно, предположим, что функции f и φ непрерывно дифференцируемы, и для некоторого $\lambda \geqslant 0$ оптимальный план x^* задачи (1)–(3) является точкой глобального минимума штрафной функции $F(x, \lambda)$. Тогда по необходимому условию минимума для дифференцируемой функции имеем

$$\frac{\partial F}{\partial x}(x^*, \lambda) = f'(x^*) + \lambda\varphi'(x^*) = 0.$$

Так как функция φ неотрицательна и x^* — план задачи (1)–(3), т. е. $\varphi(x^*) = 0$, то x^* — точка глобального минимума функции φ. Поэтому $\varphi'(x^*) = 0$ и

$$\frac{\partial F}{\partial x}(x^*, \lambda) = f'(x^*) = 0.$$

Однако, в общем случае равенство $f'(x^*) = 0$ не выполнено, поскольку x^* — оптимальный план задачи *условной* оптимизации (1)–(3), и нас интересует случай, когда этот оптимальный план лежит на границе множества планов Ω, а не принадлежит его внутренности.

Учитывая предыдущее замечание, мы определим

$$\varphi(x) = \sum_{j=1}^{\ell} |h_j(x)| + \sum_{i=1}^{m} [g_i(x)]_+$$

и

$$F_\lambda(x) = f(x) + \lambda\varphi(x).$$

Штрафную функцию F_λ часто называют ℓ_1 штрафной функцией.

ОПРЕДЕЛЕНИЕ 1. Пусть x^* — точка локального минимума в задаче (1)–(3). Штрафная функция F_λ называется *точной* в точке x^*, если существует $\lambda_0 \geqslant 0$ такое, что x^* является точкой локального минимума функции F_{λ_0}. Точная нижняя грань всех таких λ_0 обозначается $\lambda^*(x^*)$ и называется *наименьшим точным штрафным параметром* функции F_λ в точке x^*.

Предположим, что функция F_λ точна в точке локального минимума x^* в задаче (1)–(3). Поскольку функция F_λ не убывает по λ, то для любого $\lambda > \lambda^*(x^*)$, план x^* будет точкой локального минимума штрафной функции F_λ. Кроме того, из определения $\lambda^*(x^*)$ следует, что для любого $0 \leqslant \lambda < \lambda^*(x^*)$ план x^* не является точкой локального минимума функции F_λ.

Существует простое достаточное условие для точности штрафной функции в точке локального минимума в задаче (1)–(3), выражаемое в терминах условия регулярности ограничений данной задачи. Напомним (см. [2]), что ограничения в задаче (1)–(3) называются *регулярными* в точке $x \in \Omega$, если градиенты активных ограничений $g_i'(x)$, $i \in I(x)$, и $h_j'(x)$, $j \in J$, линейно независимы. Здесь $I(x) = \{i \in I \mid g_i(x) = 0\}$ — множество индексов активных ограничений неравенств.

Для любых $x \in \mathbb{R}^n$ и $r > 0$ обозначим

$$B(x,r) = \big\{y \in \mathbb{R}^n \mid \|x-y\| \leqslant r\big\}, \quad d(x,\Omega) = \inf_{y\in\Omega} \|x-y\|,$$

где $\|\cdot\|$ — евклидова норма. Напомним также, что величина

$$\varphi^{\downarrow}(x) = \liminf_{y\to x} \frac{\varphi(y)-\varphi(x)}{\|x-y\|}$$

называется *скоростью наискорейшего спуска* функции φ в точке x.

ТЕОРЕМА 1. *Пусть x^* — точка локального минимума в задаче* (1)–(3). *Предположим также, что выполнены следующие условия:*

1) *функция f удовлетворяет условию Липшица в некоторой окрестности точки x^*;*

2) *функции g_i, $i \in I(x)$, и h_j, $j \in J$, дифференцируемы в некоторой окрестности точки x^* и их градиенты непрерывны в данной точке;*

3) *ограничения в задаче* (1)–(3) *регулярны в точке x^*.*

Тогда штрафная функция F_λ является локально точной в точке x^*.

Доказательство. Покажем, что при сделанных предположениях существуют $a > 0$ и $r > 0$ такие, что

$$\varphi^{\downarrow}(x) \leqslant -a \quad \forall x \in B(x_0, r) \setminus \Omega.$$

Тогда, воспользовавшись теоремой 4 из доклада [3], получим, что существует $\delta > 0$ такое, что

$$\varphi(x) \geqslant ad(x, \Omega) \quad \forall x \in B(x_0, \delta).$$

Отсюда и из [3], теорема 2, вытекает, что штрафная функция F_λ является точной в точке x^*.

Для любого $i \in I(x^*)$ обозначим через M_i линейное подпространство натянутое на векторы

$$g'_k(x^*), \quad k \in I(x^*) \setminus \{i\}, \quad h'_s(x^*), \quad s \in J,$$

а для любого $j \in J$ обозначим через N_j линейное подпространство натянутое на векторы

$$g'_k(x^*), \quad k \in I(x^*), \quad h'_s(x^*), \quad s \in J \setminus \{j\}.$$

Поскольку в точке x^* ограничения в задаче (1)–(3) регулярны, то

$$g'_i(x^*) \notin M_i \quad \forall i \in I(x^*), \quad h'_j(x^*) \notin N_j \quad \forall j \in J.$$

Поэтому для любого $i \in I(x^*)$ найдётся ненулевой вектор v_i из ортогонального дополнения к подпространству M_i такой, что $\langle g'_i(x^*), v_i \rangle = -1$. Аналогично, для любого $j \in J$ найдётся ненулевой вектор w_j из ортогонального дополнения к подпространству N_j такой, что $\langle h'_j(x^*), w_j \rangle = -1$. Таким образом, для любого $i \in I(x^*)$ будет

$$\begin{aligned}
&\langle g'_i(x^*), v_i \rangle = -1, \\
&\langle g'_k(x^*), v_i \rangle = 0 \qquad \forall k \in I(x^*) \setminus \{i\}, \\
&\langle h'_s(x^*), v_i \rangle = 0 \qquad \forall s \in J,
\end{aligned}$$

а для любого $j \in J$ будет

$$\begin{aligned}
&\langle g_k'(x^*), w_j\rangle = 0 \qquad \forall k \in I(x^*),\\
&\langle h_j'(x^*), w_j\rangle = -1,\\
&\langle h_s'(x^*), w_j\rangle = 0 \qquad \forall s \in J \setminus \{j\}.
\end{aligned}$$

Положим $\varepsilon = \frac{1}{2(m+\ell)}$. Так как градиенты ограничений, активных в точке x^*, непрерывны в данной точке, то существует $r > 0$ такое, что для всех $x \in \in B(x^*, r)$ и для любого $i \in I(x^*)$ будет

$$\begin{aligned}
&\langle g_i'(x), v_i\rangle < -1 + \varepsilon,\\
&|\langle g_k'(x), v_i\rangle| < \varepsilon \qquad \forall k \in I(x^*) \setminus \{i\},\\
&|\langle h_s'(x), v_i\rangle| < \varepsilon \qquad \forall s \in J,
\end{aligned}$$

а для любого $j \in J$ будет

$$\begin{aligned}
&|\langle g_k'(x), w_j\rangle| < \varepsilon \qquad \forall k \in I(x^*),\\
&\langle h_j'(x), w_j\rangle < -1 + \varepsilon,\\
&|\langle h_s'(x), w_j\rangle| < \varepsilon \qquad \forall s \in J \setminus \{j\}.
\end{aligned}$$

Кроме того, выбирая $r > 0$ достаточно малым, можно считать, что $g_i(x) < 0$ для всех $x \in B(x^*, r)$ и $i \in I \setminus I(x^*)$.

Зафиксируем произвольное $x \in B(x^*, r) \setminus \Omega$. Тогда, либо $g_i(x) > 0$ для некоторого $i \in I(x^*)$, либо $h_j(x) \neq 0$ для некоторого $j \in J$. Рассмотрим сначала первый случай. Из определения вектора v_i и того факта, что $g_i(x) > 0$, следует, что производная функции $[g_i(x)]_+$ по направлению v_i в точке x удовлетворяет неравенству

$$[g_i]_+'(x, v_i) = \langle g_i'(x), v_i\rangle < -1 + \varepsilon.$$

При этом

$$|[g_i]_+'(x, v_i)| \leqslant |\langle g_k'(x), v_i\rangle| < \varepsilon, \quad \forall k \in I(x^*) \setminus \{i\}$$

и

$$\Big||h_s|'(x, v_i)\Big| \leqslant |\langle h_s'(x), v_i\rangle| < \varepsilon \quad \forall j \in J.$$

Поэтому

$$\begin{aligned}
\varphi'(x, v_i) = \sum_{s=1}^{\ell} |h_s|'(x, v_i) + \sum_{k=1}^{m} [g_i]_+'(x, v_i) <\\
< \sum_{s=1}^{\ell} \varepsilon + \sum_{k \in I(x^*)\setminus\{i\}} \varepsilon + (-1 + \varepsilon) = -1 + (m+\ell)\varepsilon = -\frac{1}{2}.
\end{aligned}$$

Отсюда, учитывая пункт 2 предложения 3 из [3], получаем, что

$$\varphi^{\downarrow}(x) \leqslant \varphi'\left(x, \frac{v_i}{\|v_i\|}\right) \leqslant -\frac{1}{2\|v_i\|}.$$

Рассуждая аналогичным образом, нетрудно проверить, что в случае, когда $h_j(x) \neq 0$ для некоторого $j \in J$, справедливы неравенства

$$\varphi^{\downarrow}(x) \leqslant \varphi'\left(x, \operatorname{sign}(h_j(x))\frac{w_i}{\|w_i\|}\right) \leqslant -\frac{1}{2\|w_j\|}.$$

Поэтому

$$\varphi^{\downarrow}(x) \leqslant -a \quad \forall x \in B(x^*, r) \setminus \Omega,$$

где

$$a = \frac{1}{2} \min_{i \in I(x^*), j \in J} \left\{ \frac{1}{\|v_i\|}, \frac{1}{\|w_j\|} \right\},$$

что и требовалось доказать. □

З а м е ч а н и е 1. Другое доказательство теоремы 1 в предположении непрерывной дифференцируемости функций f, g_i и h_j в некоторой окрестности рассматриваемой точки имеется в [4].

2°. Глобальная точность штрафной функции. Перейдём теперь к изучению глобальной точности штрафной функции F_λ.

ОПРЕДЕЛЕНИЕ 2. Штрафная функция F_λ называется *точной*, если существует $\lambda_0 \geqslant 0$ такое, что штрафная функция F_{λ_0} достигает глобального минимума, и множество точек глобального минимума этой функции совпадает с множеством оптимальных планов задачи (1)–(3). Точная нижняя грань всех таких λ_0 обозначается через λ^* и называется *наименьшим точным штрафным параметром* штрафной функции F_λ.

Предположим, что штрафная функция F_λ является точной. Так как функция $F_\lambda(x)$ строго возрастает по λ при каждом фиксированном $x \notin \Omega$, то для любого $\lambda > \lambda^*$ штрафная функция F_λ достигает глобального минимума и множество точек глобального минимума функции F_λ совпадает с множеством оптимальных планов задачи (1)–(3). Таким образом, если штрафная функция F_λ является точной, то для любого $\lambda > \lambda^*$ задача (1)–(3) эквивалентна (в смысле совпадения множества оптимальных планов) задаче безусловной минимизации штрафной функции F_λ.

З а м е ч а н и е 2. Отметим, что при доказательстве теорем удобнее пользоваться различными переформулировками определения глобальной точности штрафной функции, чем непосредственно самим определением. Например, нетрудно проверить, что штрафная функция является точной тогда и только тогда, когда выполнено одно из двух следующих условий:

1) существует $\lambda_0 \geqslant 0$ и оптимальный план x^* задачи (1)–(3), который является точкой глобального минимума функции F_{λ_0};

2) существует $\lambda_0 \geqslant 0$ такое, что
$$\inf_{x \in \mathbb{R}^n} F_{\lambda_0}(x) = f^* := \min_{x \in \Omega} f(x).$$

Прокомментируем эти условия. Точность штрафной функции формулируется, как условие совпадения двух множеств: множества оптимальных планов задачи (1)–(3) и множества точек глобального минимума функции F_λ на $\mathbb{R}^n$. Условие 1 означает, что вместо совпадения этих множеств, достаточно проверить, что они пересекаются. В свою очередь, условие 2 позволяет вместо условия совпадения множеств, проверять справедливость оценки $F_\lambda(x) \geqslant f^*$ для всех $x \in \mathbb{R}^n$.

Мы начнём изучение достаточных условий для глобальной точности штрафной функции с выпуклого случая. Предположим, что задача (1)–(3) является задачей выпуклого программирования. То есть предположим, что функции f и g_i, $i \in I$, выпуклы, а функции h_j, $j \in J$, аффинны. В этом случае штрафная функция F_λ также будет выпуклой.

Поскольку любая точка локального минимума в задаче выпуклого программирования является оптимальным планом данной задачи, то справедливо следующее утверждение.

ПРЕДЛОЖЕНИЕ 1. *Предположим, что задача* (1)–(3) *является задачей выпуклого программирования. Для того чтобы штрафная функция F_λ была точной, необходимо и достаточно, чтобы F_λ была локально точной в одном из* (*или, что эквивалентно, в каждом из*) *оптимальных планов задачи* (1)–(3).

Из предыдущего предложения вытекает, что если в задаче выпуклого программирования вида (1)–(3) ограничения регулярны хотя бы в одном из оптимальных планов этой задачи, то штрафная функция F_λ будет точной. Однако, существует более простое достаточное условие точности штрафной функции в выпуклом случае.

ПРЕДЛОЖЕНИЕ 2. *Предположим, что задача* (1)–(3) *является задачей выпуклого программирования, и пусть в этой задаче выполнено условие Слейтера. Тогда штрафная функция* F_λ *является точной.*

Доказательство. Пусть x^* — оптимальный план задачи (1)–(3). Поскольку в этой задаче выполнено условие Слейтера, то найдутся множители Лагранжа $\lambda_i \geqslant 0$, $i \in I$ и $\mu_j \in \mathbb{R}$, $j \in J$, такие, что

$$0 \in \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*) + \sum_{j=1}^{\ell} \mu_j h_j'(x^*)$$

и $\lambda_i g_i(x^*) = 0$ для всех $i \in I$, где «∂» обозначает субдифференциал в смысле выпуклого анализа (см., например, [5], теорема 1.1.5). Положим

$$\lambda = \max_{i \in I, j \in J} \big\{ \lambda_i, |\mu_j| \big\}.$$

Тогда, как нетрудно проверить, справедливо включение $0 \in \partial F_\lambda(x^*)$, из которого, в силу выпуклости функции F_λ, вытекает, что x^* является точкой глобального минимума функции F_λ. Следовательно, штрафная функция F_λ является точной по замечанию 2. $\square$

Приведём пример применения предыдущего предложения.

ПРИМЕР 1. Рассмотрим задачу поиска расстояния между эллипсоидами E_1 и E_2, заданными соотношениями

$$E_i = \big\{ x \in \mathbb{R}^n \mid \langle x, A_i x \rangle + \langle b_i, x \rangle \leqslant c_i \big\}, \quad i \in \{1, 2\}.$$

Мы предполагаем, что матрицы A_i являются положительно определёнными, и эллипсоиды E_i не вырождаются в точку, т. е.

$$c_i > \min_{x \in \mathbb{R}^n} \big(\langle x, A_i x \rangle + \langle b_i, x \rangle \big), \quad i \in \{1, 2\}. \tag{4}$$

Задачу поиска расстояния между эллипсоидами можно сформулировать в виде следующей задачи математического программирования:

$$\|x_1 - x_2\| \to \min_{x_1, x_2}, \tag{5}$$

$$\langle x_1, A_1 x_1 \rangle + \langle b_1, x_1 \rangle \leqslant c_1, \tag{6}$$

$$\langle x_2, A_2 x_2 \rangle + \langle b_2, x_2 \rangle \leqslant c_2. \tag{7}$$

Поскольку матрицы A_i являются положительно определёнными, и для любых $\alpha \in [0, 1]$ и $x_1, x_2, y_1, y_2 \in \mathbb{R}^n$ справедливо неравенство

$$\|\alpha x_1 + (1 - \alpha) y_1 - (\alpha x_2 + (1 - \alpha) y_2)\| \leqslant \alpha \|x_1 - x_2\| + (1 - \alpha) \|y_1 - y_2\|,$$

то задача (5)–(7) является задачей выпуклого программирования. Более того, условие (4) гарантирует, что в этой задаче выполнено условие Слейтера. Поэтому штрафная функция

$$F_\lambda(x_1, x_2) = \|x_1 - x_2\| + \lambda\big[\langle x_1, A_1 x_1\rangle + \langle b_1, x_1\rangle - c_1\big]_+ + \\ + \lambda\big[\langle x_2, A_2 x_2\rangle + \langle b_2, x_2\rangle - c_2\big]_+.$$

является точной штрафной функцией для задачи (5)–(7).

Перейдём теперь к рассмотрению невыпуклого случая. Для получения достаточных условий для точности штрафной функции в невыпуклом случае, нам потребуется вспомогательное утверждение, справедливое для всех штрафных функций.

ПРЕДЛОЖЕНИЕ 3. *Предположим, то штрафная функция F_λ достигает глобального минимума для всех λ больше некоторого $\lambda_0 \geqslant 0$. Тогда для любой неограниченно возрастающей последовательности $\{\lambda_k\} \subset (\lambda_0, +\infty)$ и для любых $x_{\lambda_k} \in \operatorname*{arg\,min}_{x \in \mathbb{R}^n} F_{\lambda_k}(x)$ будет*

$$\varphi(x_{\lambda_k}) \to 0 \quad \textit{при} \quad k \to \infty.$$

По поводу доказательства данного предложения см., например, [3], предложение 2.

Предложение 3 гарантирует, что точки глобального минимума функции F_λ «приближаются» к множеству допустимых планов Ω при неограниченном возрастании штрафного параметра. Однако, если множество Ω является неограниченным или $\varphi(x) \to 0$ при $\|x\| \to \infty$, то последовательность $\{x_{\lambda_k}\}$ точек глобального минимума функции F_{λ_k} может оказаться неограниченной. Для того чтобы избежать такого патологического поведения точек глобального минимума, мы введём дополнительное ограничение на штрафную функцию.

ОПРЕДЕЛЕНИЕ 3. Штрафная функция F_λ называется *невырожденной*, если существуют $\lambda_0 \geqslant 0$ и $R > 0$ такие, что для любого $\lambda \geqslant \lambda_0$ штрафная функция F_λ достигает глобального минимума и найдётся $x_\lambda \in \operatorname*{arg\,min}_{x \in \mathbb{R}^n} F_\lambda(x)$ такое, что $\|x_\lambda\| \leqslant R$.

Оказывается, что можно сформулировать необходимое и достаточное условие для точности штрафной функции в терминах условия невырожденности [6].

ТЕОРЕМА 2. *Штрафная функция F_λ является точной тогда и только тогда, когда она является невырожденной и локально точной в каждом из оптимальных планов задачи* (1)–(3).

Доказательство. Необходимость. Если функция F_λ является точной, то для любого $\lambda > \lambda^*$ любой оптимальный план задачи (1)–(3) является точкой глобального (и, следовательно, локального) минимума функции F_λ. Поэтому функция F_λ является точной в каждом оптимальном плане задачи (1)–(3).

Зафиксируем произвольный оптимальный план x^* задачи (1)–(3). Тогда для любых $\lambda \geqslant \lambda_0 > \lambda^*$ будет $x^* \in \arg\min\limits_{x \in \mathbb{R}^n} F_\lambda(x)$, а это и означает, что штрафная функция F_λ является невырожденной.

Достаточность. Поскольку штрафная функция F_λ является невырожденной, то существует $\lambda_0 \geqslant 0$ и $R > 0$ такое, что для любого $\lambda \geqslant \lambda_0$ найдётся $x_\lambda \in \arg\min\limits_{x \in \mathbb{R}^n} F_\lambda(x)$, удовлетворяющее неравенству $\|x_\lambda\| \leqslant R$. Выберем произвольную неограниченно возрастающую последовательность $\{\lambda_k\} \subset [\lambda_0, +\infty)$, и соответствующую её последовательность $x_{\lambda_k} \in \arg\min\limits_{x \in \mathbb{R}^n} F_{\lambda_k}(x)$, $k \in \mathbb{N}$, удовлетворяющую неравенству $\|x_{\lambda_k}\| \leqslant R$.

Так как последовательность $\{x_{\lambda_k}\}$ ограничена, то из неё можно выбрать сходящуюся подпоследовательность. Не ограничивая общности, можно считать, что сама последовательность $\{x_{\lambda_k}\}$ сходится. Обозначим через x^* её предельную точку. Покажем, что x^* является оптимальным планом задачи (1)–(3).

Действительно, из предложения 3 и непрерывности функции φ (напомним, что мы предполагаем, что все функции f, g_i и h_j непрерывны) следует, что $\varphi(x^*) = 0$, то есть x^* — план задачи (1)–(3). Поскольку x_{λ_k} — точка глобального минимума функции F_{λ_k}, то

$$f(x_{\lambda_k}) \leqslant f^* := \min_{x \in \Omega} f(x).$$

Отсюда, переходя к пределу при $k \to \infty$, получаем, что $f(x^*) = f^*$, то есть x^* — оптимальный план задачи (1)–(3).

Поскольку штрафная функция F_λ точна в оптимальных планах задачи (1)–(3), то для любого $\lambda > \lambda^*(x^*)$ существует $r > 0$ такое, что

$$F_\lambda(x) \geqslant F_\lambda(x^*) = f(x^*) \quad \forall x \in B(x^*, r).$$

Зафиксируем произвольное $\lambda_0 > \lambda^*(x^*)$. Учитывая, что штрафная функция не убывает по λ, получаем, что для любого $\mu \geqslant \lambda_0$ будет

$$F_\mu(x) \geqslant F_\lambda(x) \geqslant f(x^*) = F_\mu(x^*) \quad \forall x \in B(x^*, r).$$

Так как последовательность $\{\lambda_k\}$ неограниченно возрастает, а соответствующая ей последовательность $\{x_{\lambda_k}\}$ сходится к x^*, то для всех достаточно больших $k \in \mathbb{N}$ будет $\lambda_k \geqslant \lambda_0$ и $x_{\lambda_k} \in B(x^*, r)$. Следовательно, для любого достаточно большого k справедливо неравенство

$$F_{\lambda_k}(x_{\lambda_k}) \geqslant F_{\lambda_k}(x^*).$$

Так как x_{λ_k} — это точка глобального минимума функции F_{λ_k}, то из предыдущего неравенства следует, что оптимальный план x^* также является точкой глобального минимума функции F_{λ_k}. Отсюда, воспользовавшись замечанием 2, получаем, что штрафная функция F_λ является точной. $\square$

СЛЕДСТВИЕ 1. *Пусть функция f является локально липшицевой, а функции g_i, $i \in I$, и h_j, $j \in J$, непрерывно дифференцируемы на $\mathbb{R}^n$. Предположим, что ограничения в задаче* (1)–(3) *регулярны в каждом из оптимальных планов этой задачи и существует $\mu \geqslant 0$ такое, что множество*

$$\big\{x \in \mathbb{R}^n \mid F_\mu(x) < f^*\big\}$$

ограниченно, где $f^* = \min\limits_{x \in \Omega} f(x)$. Тогда штрафная функция F_λ является точной.

Приведём пример применения предыдущего следствия.

ПРИМЕР 2. Рассмотрим задачу поиска расстояния между квадриками K_1 и K_2, заданными соотношениями

$$K_i = \big\{x \in \mathbb{R}^n \mid \langle x, A_i x\rangle + \langle b_i, x\rangle = c_i\big\}, \quad i \in \{1, 2\}.$$

Мы предполагаем, что матрицы A_i являются положительно определёнными, и квадрики K_i не вырождаются в точку, т. е.

$$c_i > \min_{x \in \mathbb{R}^n} \big(\langle x, A_i x\rangle + \langle b_i, x\rangle\big), \quad i \in \{1, 2\}. \tag{8}$$

Задачу поиска расстояния между квадриками можно сформулировать в виде следующей задачи математического программирования:

$$\|x_1 - x_2\| \to \min_{x_1, x_2}, \tag{9}$$

$$h_1(x_1, x_2) = \langle x_1, A_1 x_1\rangle + \langle b_1, x_1\rangle - c_1 = 0, \tag{10}$$

$$h_2(x_2, x_2) = \langle x_2, A_2 x_2\rangle + \langle b_2, x_2\rangle - c_2 = 0. \tag{11}$$

Проверим, что в данной задаче выполнены условия следствия 1.

Действительно, градиенты ограничений h_1 и h_2 имеют вид

$$h_1'(x_1, x_2) = \big(2A_1 x_1 + b_1, 0_n\big), \quad h_2'(x_1, x_2) = \big(0_n, 2A_2 x_2 + b_2\big),$$

где $0_n \in \mathbb{R}^n$ — нулевой вектор. Из условия (8) и положительной определённости матриц A_i вытекает, что $h_i'(x_1, x_2) \neq 0$, $i \in \{1, 2\}$ в любом плане задачи (9)–(11). Поэтому ограничения в задаче (9)–(11) регулярны во всех планах этой задачи.

Поскольку матрицы A_i являются положительно определёнными, то существует $\mu > 0$ такое, что

$$\begin{aligned}\big|\langle x_1, A_1 x_1\rangle + \langle b_1, x_1\rangle - c_1\big| + \big|\langle x_2, A_2 x_2\rangle + \langle b_2, x_2\rangle - c_2\big| \geqslant \\ \geqslant \mu\big(\|x_1\|^2 + \|x_2\|^2\big) - \|b_1\|\|x_1\| - \|b_2\|\|x_2\| - |c_1| - |c_2|.\end{aligned}$$

Поэтому для любого $\lambda > 0$ будет $F_\lambda(x_1, x_2) \to \infty$ при $\|x_1\| + \|x_2\| \to \infty$, т. е. для любых $\lambda > 0$ и $s \in \mathbb{R}$ множество

$$\big\{(x_1, x_2) \in \mathbb{R}^{2n} \mid F_\lambda(x_1, x_2) < s\big\}$$

ограниченно. Отсюда, воспользовавшись следствием 1, получаем, что штрафная функция

$$\begin{aligned}F_\lambda(x_1, x_2) = \|x_1 - x_2\| + \lambda\big|\langle x_1, A_1 x_1\rangle + \langle b_1, x_1\rangle - c_1\big| + \\ + \lambda\big|\langle x_2, A_2 x_2\rangle + \langle b_2, x_2\rangle - c_2\big|.\end{aligned}$$

является точной штрафной функцией для задачи (9)–(11).

3°. Гладкие точные штрафные функции. В начале доклада было показано, что для того чтобы штрафная функция для задачи (1)–(3) вида $f(x) + \lambda\varphi(x)$ была точной, в общем случае необходимо, чтобы штрафной член $\varphi(x)$ был негладким. Однако, существует несколько различных подходов к построению *гладких* точных штрафных функций. Мы обсудим один из таких подходов, предложенный в [7] (см. также [8, 9]).

Для упрощения изложения теории гладких точных штрафных функций, мы будем предполагать, что в задаче (1)–(3) ограничения неравенства отсутствуют, т. е. эта задача имеет вид

$$f(x) \to \min, \tag{12}$$
$$h_j(x) = 0, \quad j \in J := \{1, \ldots, \ell\}. \tag{13}$$

Формально, данная задача эквивалентна задаче безусловной минимизации функции

$$\widehat{f}(x) = \begin{cases} f(x), & \text{если } x \in \Omega, \\ +\infty, & \text{если } x \notin \Omega. \end{cases}$$

Однако, с практической точки зрения задача минимизации функции $\widehat{f}$ ничем не отличается от задачи (12)–(13). Поэтому мы рассмотрим аппроксимацию функции $\widehat{f}$ вида

$$f(x,\varepsilon) = f(x) + \frac{1}{\varepsilon}\sum_{j=1}^{\ell} h_j^2(x), \quad \varepsilon > 0. \tag{14}$$

Заметим, что функция $f(x,\varepsilon)$ является классической штрафной функцией для задачи (12)–(13), и $f(x,\varepsilon) \to \widehat{f}(x)$ при $\varepsilon \to +0$.

Основная идея построения гладкой точной штрафной функции заключается в рассмотрении величины $\varepsilon > 0$ в (14), как дополнительной переменной. В этом случае мы можем рассматривать неравенство $\varepsilon > 0$, как дополнительное ограничение, и ввести штрафную функцию

$$F_\lambda(x,\varepsilon) = f(x,\varepsilon) + \lambda\varepsilon, \quad \lambda \geqslant 0,$$

явно учитывающую ограничение $\varepsilon > 0$. Доопределяя штрафную функцию $F_\lambda(x,\varepsilon)$ при $\varepsilon = 0$ с помощью функции $\widehat{f}$, мы приходим к определению штрафной функции из [7].

Дадим формальное определение. Зафиксируем произвольный вектор $w \in \in \mathbb{R}^\ell$ и положим

$$\Delta(x,\varepsilon) = \sum_{j=1}^{\ell}(h_j(x) - \varepsilon w)^2.$$

Тогда для любого $\lambda \geqslant 0$ определим *расширенную штрафную функцию* для задачи (12)–(13) следующим образом:

$$F_\lambda(x,\varepsilon) = \begin{cases} f(x), & \text{если } x \in \Omega,\ \varepsilon = 0, \\ +\infty, & \text{если } x \notin \Omega,\ \varepsilon = 0, \\ f(x) + \varepsilon^{-1}\Delta(x,\varepsilon) + \lambda\varepsilon, & \text{если } \varepsilon > 0. \end{cases}$$

Вектор w введён в определение $\Delta(x,\varepsilon)$ для того, чтобы штрафная функция $F_\lambda(x,\varepsilon)$ была похожа на модифицированную функцию Лагранжа для задачи (12)–(13) при $\varepsilon > 0$ (более подробно см. [7]).

Замечание 3. Легко видеть, что если функции f и h_j, $j \in J$, непрерывно дифференцируемы на $\mathbb{R}^n$, то функция $F_\lambda(x,\varepsilon)$ будет непрерывно дифференцируема на $\mathbb{R}^n \times (0,+\infty)$. Отметим также, что при практическом применении штрафной функции $F_\lambda(x,\varepsilon)$, её рассматривают только при положительных значениях ε, и решают задачу минимизации функции

$$F_\lambda(x,\varepsilon) = f(x) + \frac{1}{\varepsilon}\sum_{j=1}^{\ell}(h_j(x) - \varepsilon w_j)^2 + \lambda\varepsilon$$

по переменным (x,ε) на множестве $\mathbb{R}^n \times (0,+\infty)$.

Определение точности штрафной функции естественным образом переносится на случай расширенной штрафной функции $F_\lambda(x, \varepsilon)$.

ОПРЕДЕЛЕНИЕ 4. Штрафная функция $F_\lambda(x, \varepsilon)$ называется *точной*, если существует λ_0 такое, что:

1) функция $F_{\lambda_0}(x, \varepsilon)$ достигает глобального минимума на множестве $\mathbb{R}^n \times$ $\times [0, +\infty)$;

2) вектор (x^*, ε^*) является точкой глобального минимума функции $F_{\lambda_0}(x, \varepsilon)$ на множестве $\mathbb{R}^n \times [0, +\infty)$ тогда и только тогда, когда $\varepsilon^* = 0$ и x^* является оптимальным планом задачи (12)–(13).

Точная нижняя грань всех таких λ_0 обозначается λ^*_{ext} и называется *наименьшим точным штрафным параметром* расширенной штрафной функции $F_\lambda(x, \varepsilon)$.

Таким образом, если штрафная функция $F_\lambda(x, \varepsilon)$ является точной, то для любого $\lambda > \lambda^*_{ext}$ оптимальные планы задача (12)–(13) однозначно восстанавливаются по оптимальным планам задачи

$$F_\lambda(x, \varepsilon) \to \min_{x, \varepsilon},$$
$$\varepsilon \geqslant 0.$$

Кроме того, оказывается, что штрафная функция $F_\lambda(x, \varepsilon)$ является точной тогда и только тогда, когда ℓ_1 штрафная функция F_λ является точной (см. теорему 3 ниже). Таким образом, с теоретической точки зрения, подход, основанный на применении расширенной точной штрафной функции $F_\lambda(x, \varepsilon)$, равносилен подходу, основанному на применении негладких точных штрафных функций.

Прежде чем сформулировать теорему о необходимых и достаточных условиях для точности штрафной функции $F_\lambda(x, \varepsilon)$ заметим, что ℓ_1 штрафная функция

$$F_\lambda(x) = f(x) + \lambda \sum_{j=1}^{\ell} |h_j(x)|$$

является точной тогда и только тогда, когда ℓ_2 штрафная функция

$$G_\sigma(x) = f(x) + \sigma \sqrt{\sum_{j=1}^{\ell} h_j^2(x)}, \quad \sigma \geqslant 0$$

является точной, и справедливо следующее соотношение между наименьшими точным штрафными параметрами этих функций:

$$\frac{1}{\sqrt{\ell}} \sigma^* \leqslant \lambda^* \leqslant \sigma^*.$$

Справедливость данных утверждений вытекает их хорошо известных неравенств для ℓ_1 и евклидовой норм.

ТЕОРЕМА 3. *Расширенная штрафная функция $F_\lambda(x,\varepsilon)$ является точной тогда и только тогда, когда ℓ_2 штрафная функция G_σ является точной. Кроме того, наименьший точный штрафной параметр λ^*_{ext} функции $F_\lambda(x,\varepsilon)$ удовлетворяет неравенствам*

$$\frac{(\sigma^*)^2}{4} - \|w\|\sigma^* \leqslant \lambda^*_{ext} \leqslant \frac{(\sigma^*)^2}{4} + \|w\|\sigma^*, \tag{15}$$

где σ^* — наименьший точный штрафной параметр ℓ_2 штрафной функции G_σ.

Доказательство. Зафиксируем произвольные $\lambda > 0$ и $x \in \mathbb{R}^n$, и найдём верхнюю и нижнюю оценки величины

$$\inf_{\varepsilon \geqslant 0} F_\lambda(x,\varepsilon). \tag{16}$$

Если $x \in \Omega$, то

$$F_\lambda(x,\varepsilon) = f(x) + (\lambda + \|w\|^2)\varepsilon.$$

Поэтому для любого $x \in \Omega$ будет

$$\min_{\varepsilon \geqslant 0} F_\lambda(x,\varepsilon) = f(x).$$

Предположим теперь, что $x \notin \Omega$. Поскольку в этом случае $F_\lambda(x,0) = +\infty$, то инфимум в (16) можно брать по всем $\varepsilon \in (0,+\infty)$.

Заметим, что для любого $\varepsilon > 0$ справедливо равенство

$$F_\lambda(x,\varepsilon) = f(x) + \frac{1}{\varepsilon}\|h(x)\|^2 - 2\langle h(x), w\rangle + (\lambda + \|w\|^2)\varepsilon, \tag{17}$$

где $h(x) = (h_1(x), \ldots, h_\ell(x))$. Воспользовавшись неравенством Коши-Буняковского, получим, что для любого $\varepsilon > 0$ будет

$$F_\lambda(x,\varepsilon) \geqslant f(x) + \frac{1}{\varepsilon}\|h(x)\| - 2\|w\|\|h(x)\| + (\lambda + \|w\|^2)\varepsilon.$$

Для любого $\varepsilon > 0$ определим функцию

$$\omega(\varepsilon) = \frac{1}{\varepsilon}\|h(x)\| - 2\|w\|\|h(x)\| + (\lambda + \|w\|^2)\varepsilon.$$

Нетрудно видеть, что функция ω непрерывна на $(0,+\infty)$, и $\omega(\varepsilon) \to +\infty$, если $\varepsilon \to +\infty$ или $\varepsilon \to +0$. Поэтому функция ω достигает глобального минимума на множестве $(0,+\infty)$. Решая уравнение

$$\omega'(\varepsilon) = -\frac{1}{\varepsilon^2}\|h(x)\|^2 + \lambda + \|w\|^2 = 0,$$

получаем, что единственной точкой глобального минимума функции ω на множестве $(0, +\infty)$ является точка

$$\varepsilon^* = \frac{\|h(x)\|}{\sqrt{\lambda + \|w\|^2}}$$

(отметим, что поскольку $x \notin \Omega$, то $\|h(x)\| \neq 0$ и $\varepsilon^* > 0$). Отсюда получаем, что

$$\inf_{\varepsilon>0} F_\lambda(x,\varepsilon) \geqslant \min_{\varepsilon>0}(f(x) + \omega(\varepsilon)) = f(x) + \omega(\varepsilon^*) =$$
$$= f(x) + 2\big(\sqrt{\lambda + \|w\|^2} - \|w\|\big)\|h(x)\|. \quad (18)$$

Снова воспользовавшись неравенство Коши-Буняковского в (17), имеем

$$F_\lambda(x,\varepsilon) \leqslant f(x) + \frac{1}{\varepsilon}\|h(x)\| + 2\|w\|\|h(x)\| + (\lambda + \|w\|^2)\varepsilon.$$

Находя минимум правой части последнего неравенства по всем $\varepsilon > 0$, получим, что

$$\inf_{\varepsilon>0} F_\lambda(x,\varepsilon) \leqslant f(x) + 2\big(\sqrt{\lambda + \|w\|^2} + \|w\|\big)\|h(x)\|. \quad (19)$$

Из неравенств (18) и (19) следует, что

$$G_{\sigma_1}(x) \leqslant \inf_{\varepsilon\geqslant 0} F_\lambda(x,\varepsilon) \leqslant G_{\sigma_2}(x) \quad \forall x \in \mathbb{R}^n,$$

где $\sigma_1 = 2(\sqrt{\lambda + \|w\|^2} - \|w\|)$ и $\sigma_2 = 2(\sqrt{\lambda + \|w\|^2} + \|w\|)$. Отсюда легко получить, что штрафная функция $F_\lambda(x,\varepsilon)$ является точной тогда и только тогда, когда ℓ_2 штрафная функция G_σ является точной и справедливы неравенства (15). □

Замечание 4. Отметим, что можно легко модифицировать штрафную функцию $F_\lambda(x,\varepsilon)$, чтобы существенно уменьшить её точный штрафной параметр. А именно, для любого $\varepsilon > 0$ положим

$$F_\lambda(x,\varepsilon) = f(x) + \lambda\Big(\frac{1}{\varepsilon}\sum_{j=1}^{\ell}(h_j(x) - \varepsilon w_j)^2 + \varepsilon\Big).$$

Тогда рассуждая как и при доказательстве предыдущей теоремы, нетрудно показать, что наименьший точный штрафной параметр данной функции удовлетворяет неравенствам

$$\frac{\sigma^*}{2\big(\sqrt{1 + \|w\|^2} + \|w\|\big)} \leqslant \lambda^*_{ext} \leqslant \frac{\sigma^*}{2\big(\sqrt{1 + \|w\|^2} - \|w\|\big)},$$

где σ^* — наименьший точный штрафной параметр ℓ_2 штрафной функции G_σ.

ЛИТЕРАТУРА

1. Giannessi F. *Constrained Optimization and Image Space Analysis.* New York, Springer, 2005.

2. Малозёмов В.Н. *Теорема Куна–Таккера в дифференциальной форме* [Электронный ресурс] // Семинар «DHA & CAGD» [Офиц. сайт]. Избранные доклады. 27.02.2010. URL: `http://dha.spb.ru/PDF/KuhnTucker.pdf` (дата обращения: 27.12.2015).

3. Долгополик М.В. *Точные штрафные функции в негладкой оптимизации* [Электронный ресурс] // Семинар «CNSA & NDO» [Офиц. сайт]. Избранные доклады. 8.05.2014. URL: `http://www.apmath.spbu.ru/cnsa/pdf/DolgopolikM_ExactPenalty.pdf` (дата обращения: 27.12.2015).

4. Han S.P., Mangasarian O.L. *Exact penalty functions in nonlinear programming* // Mathematical Programming, 1979. vol. 17. pp. 251–269.

5. Иоффе А.Д., Тихомиров В.М. *Теория экстремальных задач.* М.: Наука, 1974. 479 с.

6. Dolgopolik M.V. *A unifying theory of exactness of linear penalty functions* // Optimization, 2015. doi: 10.1080/02331934.2015.1122005.

7. Huyer W., Neumaier A. *A new exact penalty function* // SIAM J. Optim., 2003. vol. 13. pp. 1141–1158.

8. Wang C., Ma C., Zhou J. *A new class of exact penalty functions and penalty algorithms* // J. Glob. Optim., 2014. vol. 58. pp. 51–73.

9. Dolgopolik M.V. *Smooth exact penalty functions II: a reduction to standard exact penalty functions* // Optimization Letters, 2015. doi: 10.1007/s11590-015-0961-9.