ЭКОНОМИКА 1973
И МАТЕМАТИЧЕСКИЕ МЕТОДЫ том IX, вып. 3

МЕТОД ШТРАФНЫХ ОЦЕНОК ДЛЯ ЗАДАЧ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

Н. В. ТРЕТЬЯКОВ

(Москва)

1. В [1—3] применительно к различным классам задач математического программирования рассматривался один общий метод решения, основанный на использовании модифицированной функции Лагранжа вместо обычной и применении к ней алгоритма Удзавы [4, гл. 10, § 4, 5] для отыскания седловых точек. Модификация функции Лагранжа строится путем дополнительного введения в нее штрафа за нарушение ограничений задачи; цель такой модификации — сохранив неизменным множество седловых точек, улучшить некоторые свойства самой функции и обеспечить этим сходимость алгоритма для более широкого класса задач и с большей скоростью. Из соображений, связанных с экономической интерпретацией метода (см. [1, 2]), естественно называть его «методом штрафных оценок». Этот же метод можно интерпретировать иначе, именно как модификацию метода штрафных функций [5], свободную от необходимости неограниченно увеличивать штрафной коэффициент (что, как известно, приводит к серьезным вычислительным трудностям), за счет итерационного подбора надлежащего сдвига штрафа [1, 3]. В [1] метод изучался для нелинейных задач на экстремум с ограничениями в форме равенств, в [2] — для задач линейного программирования, в [3] — для невыпуклых задач с ограничениями в форме неравенств. В настоящей статье дается обоснование метода штрафных оценок и исследуется его сходимость для задач выпуклого программирования. Устанавливается, что при естественных для выпуклого программирования предположениях метод позволяет построить две итерационные последовательности — исходных и двойственных переменных, которые сходятся к множествам решений соответственно прямой и двойственной задач. Показывается, что при условиях регулярности решения исходной задачи метод сходится в окрестности седловой точки не медленнее убывающей геометрической прогрессии, причем знаменатель прогрессии может быть сделан сколь угодно малым за счет выбора достаточно большого значения штрафного коэффициента.

2. Рассматривается задача

$$f(x) \to \min,$$
$$g(x) \leqslant 0, \tag{1}$$
$$x \in Q,$$

где $x \in R^n$, $g(x) = (g_1(x), \ldots, g_m(x)) \in R^m$, $f(x)$, $g_i(x)$ — выпуклые непрерывные функции; Q — выпуклое замкнутое множество в R^n.

Составим для (1) модифицированную функцию Лагранжа

$$M(x, y, K) = f(x) + \frac{1}{2K}\|[y + Kg(x)]_+\|^2 - \frac{1}{2K}\|y\|^2. \tag{2}$$

Здесь $K \in R^1$, $K > 0$, $y \in R^m$, $t_+ = \max\{0, t\}$. Отметим простейшие свойства функции $M(x, y, K)$.

а) $M(x, y, K)$ выпукла и непрерывна по x при любых $y \in R^m$, $K > 0$, вогнута и непрерывно дифференцируема по y при любых $x \in R^n$, $K > 0$. Действительно, $f(x)$ выпукла, а второе слагаемое в правой части (2) выпукло по x как сумма квадратов выпуклых неотрицательных функций. Вогнутость $M(x, y, K)$ по y видна из представления

$$M(x, y, K) = f(x) + \sum_{i=1}^{m} q_i(x, y_i, K),$$

$$q_i(x, y_i, K) = \begin{cases} -\dfrac{1}{2K} y_i^2, & y_i \leqslant -Kg_i(x), \\ y_i g_i(x) + \dfrac{K}{2} g_i^2(x), & y_i > -Kg_i(x). \end{cases} \tag{3}$$

б) При $y = 0$ функция $M(x, y, K)$ совпадает со штрафной функцией [5] $P(x, K) = M(x, 0, K) = f(x) + (K/2)\|[g(x)]_+\|^2$, а при $K \to +0$ и любых $x \in R^n$, $y \in R^m$ существует предел

$$\lim_{K \to +0} M(x, y, K) = \begin{cases} L(x, y), & \text{если } y \geqslant 0, \\ -\infty, & \text{в противном случае,} \end{cases}$$

где $L(x, y) = f(x) + (y, g(x))$ — функция Лагранжа задачи (1). Справедливость последнего утверждения видна, например, из (3).

в) Имеют место неравенства

$$L(x, y) \leqslant M(x, y, K) \quad \text{при } y \geqslant 0,\ K > 0, \tag{4}$$

$$M(x, y, K) \leqslant f(x) \quad \text{при } y \geqslant 0,\ g(x) \leqslant 0,\ K > 0. \tag{5}$$

Чтобы убедиться в этом, заметим, что (2) может быть записано в виде

$$M(x, y, K) = f(x) + \frac{1}{2K}\|[y + Kg(x)]_+ - y\|^2 + \frac{1}{K}(y, [y + Kg(x)]_+ - y). \tag{6}$$

Отсюда $M(x, y, K) \geqslant f(x) + (1/K)(y, [y + Kg(x)]_+ - y)$, и, учитывая неравенство

$$[y + Kg(x)]_+ - y \geqslant Kg(x), \tag{7}$$

получаем (4). Неравенство (5) сразу следует из (2), если заметить, что

$$0 \leqslant [y + Kg(x)]_+ \leqslant y \tag{8}$$

при $y \geqslant 0$, $g(x) \leqslant 0$, $K > 0$.

Чтобы установить дальнейшие свойства $M(x, y, K)$, нам понадобится одна лемма из теории выпуклых множеств и функций в конечномерных пространствах.

Лемма 1. *Пусть выпуклое множество $D = Q \cap \{x \in R^n : h(x) \leqslant 0\}$ непусто и ограничено (здесь $h(x) = (h_1(x), \ldots, h_l(x))$ — выпуклая непрерывная вектор-функция, Q — выпуклое замкнутое подмножество R^n). Тогда при всяком $z \in R^l$ множество $D_z = Q \cap \{x \in R^n : h(x) \leqslant z\}$ ограничено (в частности, D_z может быть пустым).*

Формулировку и доказательство леммы 1 для случая $Q = R^n$ можно найти в [5, гл. 6]; случай $Q \neq R^n$ сводится к предыдущему случаю, например, заменой $Q = \{x \in R^n : \rho(x, Q) \leqslant 0\}$, где $\rho(x, Q)$ означает расстояние от точки x до множества Q.

Обозначим через X^* множество решений задачи (1), т. е. $X^* = \{x^* \in Q : g(x^*) \leqslant 0,\ f(x^*) \leqslant f(x)$ для любого $x \in Q,\ g(x) \leqslant 0\}$. Рассмотрим двойственную задачу

$$\varphi(y) \to \max, \qquad y \geqslant 0, \tag{9}$$

где $\varphi(y) = \inf\limits_{x \in Q} L(x, y)$, и обозначим через Y^* множество ее решений, т. е. $Y^* = \{y^* \geqslant 0 : \varphi(y^*) \geqslant \varphi(y)$ для любого $y \geqslant 0\}$. Всюду в дальнейшем будем предполагать выполненными следующие два условия.

I. $X^* \neq \phi,\ Y^* \neq \phi$.

II. X^* ограничено.

Из условий I, II следует [6], что $X^* \times Y^*$ — множество седловых точек функции Лагранжа, т. е.

$$L(x^*, y) \leqslant L(x^*, y^*) \leqslant L(x, y^*) \tag{10}$$

при всех $x^* \in X^*$, $x \in Q$, $y^* \in Y^*$, $y \geqslant 0$. При этом $L(x^*, y^*) = f(x^*)$ и из (4), (5) получаем, что $M(x^*, y^*, K) = L(x^*, y^*) = f(x^*)$ при любых $x^* \in X^*$, $y^* \in Y^*$, $K > 0$.

Лемма 2. *При всех $y \geqslant 0$, $K > 0$ функция $M(x, y, K)$ достигает минимума по $x \in Q$ на непустом ограниченном множестве $\tilde{X}(y, K)$, причем $\tilde{X}(y^*, K) = X^*$ для всякого $y^* \in Y^*$. Для любых $\tilde{x} \in \tilde{X}(y, K)$, $y' \geqslant 0$, $\tilde{x}' \in \tilde{X}(y', K)$ имеют место неравенства*

$$\|[g(\tilde{x})]_+\| \leqslant \|M_y(\tilde{x}, y, K)\| \leqslant \frac{1}{K}\rho(y, Y^*) = \frac{1}{K}\min_{y^* \in Y^*}\|y - y^*\|, \tag{11}$$

$$\|M_y(\tilde{x}, y, K) - M_y(\tilde{x}', y', K)\| \leqslant \frac{1}{K}\|y - y'\|, \tag{12}$$

где $M_y(\tilde{x}, y, K) = [y/K + g(\tilde{x})]_+ - y/K$ — градиент непрерывно дифференцируемой функции $M_{\{\tilde{x}, K\}}(y) = M(\tilde{x}, y, K)$ в точке y.

Доказательство. Пусть $x^* \in X^*$, $y^* \in Y^*$, $y \geqslant 0$, $K > 0$. Достаточно рассмотреть функцию $M(x, y, K)$ на непустом (содержащем x^*) замкнутом множестве $\tilde{Q} = \{x \in Q : M(x, y, K) \leqslant M(x^*, y, K)\}$. При $x \in \tilde{Q}$ имеем, согласно (5) и (10), $M(x, y, K) \leqslant M(x^*, y, K) \leqslant f(x^*) = L(x^*, y^*) \leqslant L(x, y^*) = f(x) + (y^*, g(x))$. Отсюда в силу неравенства (7) следует, что $M(x, y, K) \leqslant f(x) + (1/K)(y^*, [y + Kg(x)]_+ - y)$. Используя (6) для $M(x, y, K)$, получаем из предыдущего неравенства $\frac{1}{2}\|[y + Kg(x)]_+ - y\|^2 + (y, [y + Kg(x)]_+ - y) \leqslant (y^*, [y + Kg(x)]_+ - y)$. Следовательно, $\|[y + Kg(x)]_+ - y\|^2 \leqslant 2(y^* - y, [y + Kg(x)]_+ - y) \leqslant 2\|y - y^*\| \cdot \|[y + Kg(x)]_+ - y\|$, откуда

$$\|[y + Kg(x)]_+ - y\| \leqslant 2\|y - y^*\|. \tag{13}$$

Нетрудно проверить, что при $y \geqslant 0$ наряду с (7) имеет место равенство $[Kg(x)]_+ = \{[y + Kg(x)]_+ - y\}_+$, из которого следует

$$\|[Kg(x)]_+\| = \|\{[y + Kg(x)]_+ - y\}_+\| \leqslant \|[y + Kg(x)]_+ - y\|. \tag{14}$$

Из (13) и (14) вытекает

$$\|[g(x)]_+\| \leqslant \frac{2}{K}\|y - y^*\|. \tag{15}$$

Неравенства (13) и (15) позволяют установить ограниченность множества $\tilde{Q}$. Действительно, поскольку из $x \in \tilde{Q}$ следует $M(x, y, K) \leqslant M(x^*, y, K) \leqslant f(x^*)$, а отсюда $f(x) \leqslant f(x^*) - (1/2K)\|[y + Kg(x)]_+ - y\|^2 - (1/K)(y, [y + Kg(x)]_+ - y) \leqslant f(x^*) - (1/K)(y, [y + Kg(x)]_+ -$

$-y)$, то используя (13), приходим к неравенству $f(x)\leqslant f(x^*)+(2/K)\|y\|\|y-y^*\|$ для всех $x\in\tilde{Q}$. Таким образом, $\tilde{Q}\subseteq\{x\in Q:\|[g(x)]_+\|\leqslant[2/K)\|y-y^*\|, f(x)\leqslant f(x^*)+(2/K)\|y\|\|y-y^*\|\}$ и по лемме 1 множество $\tilde{Q}$ ограничено, поскольку ограничено множество $X^*=\{x\in Q:\|[g(x)]_+\|\leqslant 0, f(x)\leqslant f(x^*)\}$. Из компактности $\tilde{Q}$ следует, что непрерывная по x функция $M(x,y,K)$ достигает своего минимума на $\tilde{Q}$, а значит, и на Q. При этом множество ее точек минимума $\tilde{X}(y,K)\subseteq\tilde{Q}$, следовательно, $\tilde{X}(y,K)$ ограничено.

Докажем теперь оценку (12). Пусть $\tilde{x}\in\tilde{X}(y,K)$, $\tilde{x}'\in\tilde{X}(y',K)$, $y\geqslant 0$, $y'\geqslant 0$. Необходимое и достаточное условие минимума по $x\in Q$ функции $M(x,y,K)$ в точке $\tilde{x}$ имеет вид [7] $(M_x(\tilde{x},y,K),\tilde{x}-x)\leqslant 0$ для всех $x\in Q$. Здесь $M_x(\tilde{x},y,K)$ означает некоторый опорный функционал к выпуклой непрерывной функции $M_{(y,K)}(x)=M(x,y,K)$ в точке $\tilde{x}\in Q$. Как следует из известных теорем о представлении опорного к сумме и к максимуму двух выпуклых непрерывных функционалов [7, § 1], а также к суперпозиции монотонного выпуклого функционала и выпуклого оператора [6, гл 4], для всякого $M_x(\tilde{x},y,K)$ найдутся такие опорные в точке $\tilde{x}$ функционалы $f'(\tilde{x})$, $g_i'(\tilde{x})$, $i=1,\ldots,m$, к функциям $f(x)$ и $g_i(x)$ соответственно, что $M_x(\tilde{x},y,K)=f'(\tilde{x})+[g'(\tilde{x})]^*[y+Kg(\tilde{x})]_+$, где $g'(\tilde{x})$ — матрица $m\times n$, строками которой служат векторы $g_i'(\tilde{x})$, а запись $[\cdot]^*$ означает транспонирование. В частности, при $x=\tilde{x}'$ условие минимума дает:

$$0\geqslant(M_x(\tilde{x},y,K),\tilde{x}-\tilde{x}')=(f'(\tilde{x})+[g'(\tilde{x})]^*[y+Kg(\tilde{x})]_+,\tilde{x}-\tilde{x}').$$

Аналогично, из условия $\tilde{x}'\in\tilde{X}(y',K)$

$$0\geqslant(M_x(\tilde{x}',y',K),\tilde{x}'-\tilde{x})=(f'(\tilde{x}')+[g'(\tilde{x}')]^*[y'+Kg(\tilde{x}')]_+,\tilde{x}'-\tilde{x}).$$

Складывая последние два неравенства и учитывая, что из выпуклости $f(x)$ следует $(f'(\tilde{x})-f'(\tilde{x}'),\tilde{x}-\tilde{x}')\geqslant 0$, получаем

$$([g'(\tilde{x})]^*[y+Kg(\tilde{x})]_+-[g'(\tilde{x}')]^*[y'+Kg(\tilde{x}')]_+,\tilde{x}-\tilde{x}')\leqslant 0.$$

Отсюда $([y+Kg(\tilde{x})]_+,\ g'(\tilde{x})(\tilde{x}-\tilde{x}'))-([y'+Kg(\tilde{x}')]_+,\ g'(\tilde{x}')\cdot(\tilde{x}-\tilde{x}'))\leqslant 0$. В силу выпуклости $g_i(x)$, $i=1,\ldots,m$, имеем $g'(\tilde{x})(\tilde{x}-\tilde{x}')\geqslant g(\tilde{x}')-g(\tilde{x}')$, $g'(\tilde{x}')(\tilde{x}'-\tilde{x})\geqslant g(\tilde{x}')-g(\tilde{x})$, поэтому $([y+Kg(\tilde{x})]_+-[y'+Kg(\tilde{x}')]_+,g(\tilde{x})-g(\tilde{x}'))\leqslant 0$ или, что то же самое,

$$([y+Kg(\tilde{x})]_+-[y'+Kg(\tilde{x}')]_+,\ \{[y+Kg(\tilde{x})]-y\}-$$
$$-\{[y'+Kg(\tilde{x}')]-y'\})\leqslant 0.$$

Из последнего неравенства, учитывая очевидные соотношения типа $([y+Kg(\tilde{x})]_+,\ y+Kg(\tilde{x}))=([y+Kg(\tilde{x})]_+,\ [y+Kg(\tilde{x})]_+)$, получаем

$$([y+Kg(\tilde{x})]_+-[y'+Kg(\tilde{x}')]_+,\ \{[y+Kg(\tilde{x})]_+-y\}-$$
$$-\{[y'+Kg(\tilde{x}')]_+-y'\})\leqslant 0.$$

Это приводит к

$$\|\{[y+Kg(\tilde{x})]_+-y\}-\{[y'+Kg(\tilde{x}')]_+-y'\}\|^2\leqslant$$
$$\leqslant(y'-y,\ \{[y+Kg(\tilde{x})]_+-y\}-\{[y'+Kg(\tilde{x}')]_+-y'\}).\qquad(16)$$

Оценив скалярное произведение в правой части (16) произведением норм сомножителей, получаем $\|\{[y+Kg(\tilde{x})]_+-y\}-\{[y'+Kg(\tilde{x}')]_+-y'\}\|\leqslant\|y-y'\|$, откуда следует (12).

Покажем, что при $y^*\in Y^*$ имеет место равенство $\tilde{X}(y^*,K)=X^*$. Прежде всего, включение $X^*\subseteq\tilde{X}(y^*,K)$ сразу следует из соотношений $M(x,y^*,K)\geqslant L(x,y^*)\geqslant L(x^*,y^*)=M(x^*,y^*,K)$. Для доказательства обратного включения предположим, что $\tilde{x}\in\tilde{X}(y^*,K)$. Тогда в точке $\tilde{x}$ выполнено условие минимума $M(x,y^*,K)$ по $x\in Q:(f'(\tilde{x})+[g'(\tilde{x})]^*[y^*+Kg(\tilde{x})]_+,\tilde{x}-x)\leqslant$

$\leqslant 0$, $x \in Q$. Для произвольного $x^* \in X^*$ по доказанному $x^* \in \tilde{X}(y^*, K)$ и из (12) следует, что $[y^*/K + g(\tilde{x})]_+ - y^*/K = [y^*/K + g(x^*)]_+ - y^*/K$. При этом правая часть последнего равенства равна нулю, поскольку $y^* \geqslant 0$, $g(x^*) \leqslant 0$, $(y^*, g(x^*)) = 0$, следовательно, $[y^* + Kg(\tilde{x})]_+ = y^*$. Отсюда получаем, что в точке $\tilde{x}$ выполнены условия $g(\tilde{x}) \leqslant 0$, $(y^*, g(\tilde{x})) = 0$ и $(f'(\tilde{x}) + [g'(\tilde{x})]^* y^*, \tilde{x} - x) \leqslant 0$, $x \in Q$, которые означают, что $\tilde{x} \in X^*$. Наконец, полагая в (12) $y' = y^* \in Y^*$, $\tilde{x}' = x^* \in X^*$ и учитывая (14), получаем (11). Лемма доказана.

С л е д с т в и е. $M_y(\tilde{x}, y, K) = [y/K + g(\tilde{x})]_+ - y/K$ принимает одно и то же значение при всех $\tilde{x} \in \tilde{X}(y, K)$ и непрерывно зависит от y.

Справедливость этого утверждения вытекает непосредственно из (12).

Отметим, что если заменить условие II более слабым требованием (10), то все утверждения леммы 2, кроме непустоты $\tilde{X}(y, K)$, останутся в силе, и их доказательство не изменится. При этом вместо непустоты $\tilde{X}(y, K)$, т. е. вместо существования точек минимума $M(x, y, K)$, можно показать лишь ограниченность снизу функции $M(x, y, K)$. Простой пример задачи минимизации $f(x) \equiv C = \text{const}$, $x \in R^1$ при ограничении $e^{-x} - 1 \leqslant 0$ показывает, что если условие II не выполнено, то минимум $M(x, y, K)$ может не достигаться. Действительно, здесь $X^* = [0, +\infty)$ — неограничено, $V^* = \{0\}$ — непусто и ограничено, следовательно, выполняются условия I и (10). При $0 \leqslant y < K$ функция $M(x, y, K)$ достигает минимума на множестве $\tilde{X}(y, K) = [-\ln(1 - y/K), +\infty)$, однако при $y \geqslant K$ функция $M(x, y, K)$, оставаясь ограниченной снизу, не достигает своей нижней грани, $\tilde{X}(y, K) = \varnothing$.

Перейдем к исследованию свойств двойственной функции $\psi(y, K) = \min\limits_{x \in Q} M(x, y, K)$, которая, согласно лемме 2, определена при всех $y \geqslant 0$, $K > 0$.

Лемма 3. *При всех $y \geqslant 0$, $K > 0$ функция $\psi(y, K) = \min\limits_{x \in Q} M(x, y, K)$ вогнута и непрерывно дифференцируема по y, а ее градиент $\psi_y(y, K) = M_y(\tilde{x}, y, K) = [y/K + g(\tilde{x})]_+ - y/K$, $\tilde{x} \in \tilde{X}(y, K)$ удовлетворяет по y условию Липшица с постоянной $1/K$. Множество точек максимума $\psi(y, K)$ по $y \geqslant 0$ совпадает с Y^*.*

Д о к а з а т е л ь с т в о. Вогнутость $\psi(y, K)$ по y легко следует из ее определения и вогнутости по y функции $M(x, y, K)$. Покажем, что $\psi(y, K)$ дифференцируема. Пусть $y \geqslant 0$, $y + \Delta y \geqslant 0$, $\tilde{x} \in \tilde{X}(y, K)$, $\tilde{x} + \Delta\tilde{x} \in \tilde{X}(y + \Delta y, K)$. Имеем, согласно определению $\psi(y, K)$

$$M(\tilde{x} + \Delta\tilde{x}, y + \Delta y, K) = \psi(y + \Delta y, K) \leqslant M(\tilde{x}, y + \Delta y, K),$$

$$M(\tilde{x} + \Delta\tilde{x}, y, K) \geqslant \psi(y, K) = M(\tilde{x}, y, K).$$

Вычитая почленно из верхних неравенств нижние, получаем

$$M(\tilde{x} + \Delta\tilde{x}, y + \Delta y, K) - M(\tilde{x} + \Delta\tilde{x}, y, K) \leqslant \psi(y + \Delta y, K) - \psi(y, K) \leqslant \\ \leqslant M(\tilde{x}, y + \Delta y, K) - M(\tilde{x}, y, K). \tag{17}$$

В силу вогнутости $M(x, y, K)$ по y имеем

$$M(\tilde{x}, y + \Delta y, K) - M(\tilde{x}, y, K) \leqslant (M_y(\tilde{x}, y, K), \Delta y). \tag{18}$$

Аналогичным образом, $M(\tilde{x} + \Delta\tilde{x}, y + \Delta y, K) - M(\tilde{x} + \Delta\tilde{x}, y, K) \geqslant (M_y(\tilde{x} + \Delta\tilde{x}, y + \Delta y, K), \Delta y) \geqslant (M_y(\tilde{x}, y, K), \Delta y) - \|M_y(\tilde{x} + \Delta\tilde{x}, y + \Delta y, K) - M_y(\tilde{x}, y, K)\| \|\Delta y\|$.

Согласно (12), $\|M_y(\tilde{x} + \Delta\tilde{x}, y + \Delta y, K) - M_y(\tilde{x}, y, K)\| \leqslant (1/K)\|\Delta y\|$, следовательно,

$$M(\tilde{x}+\Delta\tilde{x}, y+\Delta y, K) - M(\tilde{x}+\Delta\tilde{x}, y, K) \geqslant (M_y(\tilde{x}, y, K), \Delta y) - \frac{1}{K}\|\Delta y\|^2. \tag{19}$$

Поскольку $M_y(\tilde{x}, y, K) = [y/K + g(\tilde{x})]_+ - y/K$ принимает одно и то же значение для всех $\tilde{x} \in \tilde{X}(y, K)$, то из (17)–(19) следует, что $\psi(y, K)$ дифференцируема по y, а ее градиент равен $\psi_y(y, K) = M_y(\tilde{x}, y, K)$, $\tilde{x} \in \tilde{X}(y, K)$. Неравенство (12) показывает также, что $\psi_y(y, K)$ удовлетворяет условию Липшица с постоянной $1/K$. Наконец, завершая доказательство леммы, заметим, что $\psi(y, K)$ достигает максимума по $y \geqslant 0$ в тех и только тех точках, для которых

$$\begin{gathered} \psi_y(y, K) = \left[\frac{1}{K}y + g(\tilde{x})\right]_+ - \frac{1}{K}y \leqslant 0, \\ (\psi_y(y, K), y) = \left(\left[\frac{1}{K}y + g(\tilde{x})\right]_+ - \frac{1}{K}y, y\right) = 0. \end{gathered} \tag{20}$$

При $y = y^* \in Y^*$ это имеет место, поскольку $\tilde{X}(y^*, K) = X^*$, $g(x^*) \leqslant 0$, $(y^*, g(x^*)) = 0$ и, следовательно, $[y^*/K + g(x^*)]_+ - y^*/K = 0$. Обратно, из (20) легко вывести, что $[y/K + g(\tilde{x})]_+ - y/K = 0$, поэтому при любом $\tilde{x} \in \tilde{X}(y, K)$ имеем $g(\tilde{x}) \leqslant 0$, $(y, g(\tilde{x})) = 0$ и, кроме того, условие минимума $M(x, y, K)$ в точке $\tilde{x}$ принимает вид $(f'(\tilde{x}) + [g'(\tilde{x})]^* y, \tilde{x} - x) \leqslant 0$, $x \in Q$. Но это означает [6, 7], что $\tilde{x} \in X^*$, $y \in Y^*$.

Из леммы 3 вытекает утверждение, имеющее форму теоремы двойственности.

Следствие. При $K > 0$ функция $M(x, y, K)$ имеет седловые точки на множестве $x \in Q$, $y \geqslant 0$ те же, что и функция $L(x, y)$ (т. е. $X^* \times Y^*$). Для всякого допустимого x ($x \in Q$, $g(x) \leqslant 0$) и любого $y \geqslant 0$ справедливы соотношения

$$\max_{y \geqslant 0} M(x, y, K) = f(x) \geqslant f(x^*) = \psi(y^*, K) \geqslant \psi(y, K) = \min_{x \in Q} M(x, y, K),$$
$$x^* \in X^*, \quad y^* \in Y^*,$$

в которых неравенства переходят в равенства лишь при $x \in X^*$, $y \in Y^*$.

Подчеркнем, что обычная функция Лагранжа $L(x, y)$ не обладает свойствами, установленными леммами 2 и 3 для модифицированной функции $M(x, y, K)$: $L_y(\tilde{x}, y) = g(\tilde{x})$, не определяется однозначно для всех точек минимума $\tilde{x}$ функции $L(x, y)$, двойственная функция $\varphi(y) = \inf_{x \in Q} L(x, y)$ недифференцируема и, кроме того, точки минимума $L(x, y^*)$ по $x \in Q$ не исчерпываются множеством X^*.

3. Перейдем непосредственно к методу штрафных оценок для решения задачи (1). В этом методе строятся итерационные последовательности $x^n \in Q$, $y^n \geqslant 0$, $K_n > 0$, причем K_n могут выбираться произвольно, с единственным условием $\inf_n K_n \geqslant K > 0$, x^n ищется по y^n и K_n путем минимизации $M(x, y^n, K_n)$ по $x \in Q$, y^{n+1} пересчитывается по x^n, y^n, K_n. Формулы метода имеют вид

$$\begin{cases} K_n \geqslant K > 0, \\ M(x^n, y^n, K_n) = \min\limits_{x \in Q} M(x, y^n, K_n), \qquad x^n \in Q, \\ y^{n+1} = [y^n + K_n g(x^n)]_+. \end{cases} \tag{21}$$

В соответствии с леммой 2 итерационные последовательности x^n и y^n могут быть построены. Кроме того, из леммы 3 следует, что при каждом n формулы (21) представляют собой запись одной итерации градиентного метода максимизации вогнутой дифференцируемой функции $\psi(y, K_n)$ с длиной шага K_n, поскольку $y^{n+1}=[y^n+K_n g(x^n)]_+=y^n+K_n\psi_y(y^n, K_n)$.

Следующее утверждение гарантирует сходимость метода (21) при любом начальном приближении y^0.

Теорема 1. *При любых $K_n\geqslant K>0$, $y^0\geqslant 0$ имеет место сходимость $\rho(x^n, X^*)=\min\limits_{x^*\in X^*}\|x^n-x^*\|\to 0$, $\rho(y^n, Y^*)=\min\limits_{y^*\in Y^*}\|y^n-y^*\|\to 0$, когда $n\to\infty$. При этом обе последовательности x^n и y^n ограничены. Для любого $n\geqslant 0$ справедлива оценка*

$$M(x^n, y^n, K_n)\leqslant f(x^*),\quad x^*\in X^*. \tag{22}$$

Доказательство. Зафиксируем произвольные $x^*\in X^*$, $y^*\in Y^*$. Воспользуемся неравенством (16), полученным при доказательстве леммы 2, положив в нем $y=y^n$, $\tilde{x}=x^n$, $[y+Kg(\tilde{x})]_+=y^{n+1}$, $y'=y^*$, $\tilde{x}'=x^*$, $K=K_n$. Получим $\|y^{n+1}-y^n\|^2\leqslant(y^*-y^n,\ y^{n+1}-y^n)=-(1/2)\|y^{n+1}-y^*\|^2+(1/2)\|y^n-y^*\|^2+(1/2)\|y^{n+1}-y^n\|^2$. Отсюда $\|y^{n+1}-y^*\|^2\leqslant\|y^n+y^*\|^2-\|y^{n+1}-y^n\|^2$. Суммируя подобные неравенства от $n=0$ до $n=N$, находим

$$\|y^{N+1}-y^*\|^2\leqslant\|y^0-y^*\|^2-\sum_{n=0}^{N}\|y^{n+1}-y^n\|^2. \tag{23}$$

Из (23) следует, прежде всего, что $\sum\limits_{n=0}^{\infty}\|y^{n+1}-y^n\|^2\leqslant\|y^0-y^*\|^2$, поэтому $\|y^{n+1}-y^n\|\to 0$ и $\left\|\left[\frac{1}{K_n}y^n+g(x^n)\right]_+-\frac{1}{K_n}y^n\right\|=\frac{1}{K_n}\|y^{n+1}-y^n\|\to 0$, так как $K_n\geqslant K>0$. В силу (14) получаем, что $\|[g(x^n)]_+\|\to 0$. Кроме того, из (23) следует, что $\|y^{N+1}-y_*\|\leqslant\|y^0-y^*\|$, т. е. $\{y^n\}$ ограничена. Для любого n имеем $M(x^n, y^n, K_n)\leqslant M(x^*, y^n, K_n)\leqslant f(x^*)$, откуда помимо оценки (22) получаем

$$f(x^n)\leqslant f(x^*)-\frac{1}{2K_n}\|[y^n+K_n g(x_n)]_+-y^n\|^2-\frac{1}{K_n}(y^n,[y^n+K_n g(x^n)]_+-y^n)\leqslant$$
$$\leqslant f(x^*)+\|y^n\|\frac{1}{K_n}\|y^{n+1}-y^n\|. \tag{24}$$

С другой стороны,

$$f(x^*)=L(x^*,y^*)\leqslant L(x^n,y^*)=f(x^n)+(y^*,g(x^n))\leqslant f(x^n)+$$
$$+\left(y^*,\left[\frac{1}{K_n}y^n+g(x^n)\right]_+-\frac{1}{K_n}y^n\right)\leqslant f(x^n)+\|y^*\|\frac{1}{K_n}\|y^{n+1}-y^n\|. \tag{25}$$

Ввиду ограниченности $\|y^n\|$ из (24) и (25) следует, что $f(x^n)\to f(x^*)$.

Из доказанного вытекает, что $\rho(x^n, X^*)\to 0$ при $n\to\infty$. Действительно, точки x^n содержатся, начиная с некоторого номера, в множестве $\{x\in Q: f(x)-f(x^*)\leqslant 1,\ \|[g(x)]_+\|\leqslant 1\}$, ограниченном по лемме 1 в силу непустоты и ограниченности множества $\{x\in Q: f(x)-f(x^*)\leqslant 0,\ \|[g(x)]_+\|\leqslant 0\}=X^*$. Следовательно, $\{x^n\}$ ограничена. Поэтому, если для некоторой подпоследовательности $\{x^{n_k}\}\subseteq\{x^n\}$ имеет место $\rho(x^{n_k}, X^*)\geqslant\varepsilon>0$, то найдется предельная точка $\bar{x}$ этой подпоследовательности, $\rho(\bar{x}, X^*)\geqslant\varepsilon$. Но это невозможно, так как для любой предельной точки $\bar{x}$ последовательности $\{x^n\}$ имеем $x\in Q$, $f(\bar{x})=\lim\limits_{n\to\infty}f(x^*)=f(x^*)$, $[g(\bar{x})]_+=\lim\limits_{n\to\infty}[g(x^n)]_+=0$, т. е. $\bar{x}\in X^*$.

Покажем, что $\rho(y^n, Y^*) \to 0$. Прежде всего, $y^n \geqslant 0$ и $\{y^n\}$ ограничена. Кроме того, $\psi(y^n, K_n) = M(x^n, y^n, K_n) = f(x^n) + \frac{1}{2K_n} \| [y^n + K_n g(x^n)]_+ - y^n \|^2 + \frac{1}{K_n} (y^n, [y^n + K_n g(x^n)]_+ - y^n) \to f(x^*) = \psi(y^*, K_n)$. Отсюда, аналогично предыдущему, следует $\rho(y^n, Y^*) \to 0$. Теорема доказана.

Заметим, что если ослабить сделанные в п. 2 предположения, заменив условие II условиями (10), и предположить дополнительно существование точек минимума $x^n \in Q$ функции $M(x, y^n, K_n)$ для всех y^n, K_n, то, повторяя проведенные выше рассуждения, можно показать, что $[g(x^n)]_+ \to 0$, $f(x^n) \to f(x^*)$, $\rho(y^n, Y^*) \to 0$. Отметим также, что приведенные результаты охватывают случай, когда некоторые из ограничений задачи (1) суть линейные равенства, записанные в виде пары взаимно противоположных неравенств.

Теорема 1 была доказана примерно по той же схеме, что и теорема о сходимости метода для линейного программирования [2, теорема 1]. Однако, если для задачи линейного программирования метод, как установлено в [2], является конечным, т. е. приводит к точному решению за конечное число итераций, то для более общих задач выпуклого программирования это не имеет места. Простейшие примеры показывают, что метод будет бесконечным даже для наиболее близкого к линейному случаю — задачи квадратичного программирования. Так, для задачи минимизации $f(x) = x^2$, $x \in R^1$, при ограничении $x \leqslant 0$ метод определяет две бесконечные итерационные последовательности

$$x^n = \frac{-1}{(1+K_n)\dots(1+K_0)} y^0 \quad \text{и} \quad y^n = \frac{1}{(1+K_{n-1})\dots(1+K_0)} y^0,$$

сходящиеся соответственно к $x^* = 0$ и $y^* = 0$, при этом $x^n \neq x^*$, $y^n \neq y^*$ для любых $n \geqslant 0$.

4. В дальнейшем нам потребуются две нижеследующие леммы, которые можно рассматривать как модификации лемм 1 и 3 из [1]. Для сопоставления с [1], а также учитывая некоторое самостоятельное значение этих утверждений, докажем их для гильбертовых пространств, хотя для целей настоящей статьи они потребуются лишь в конечномерном случае, где доказательства можно провести проще.

Лемма 4. *Пусть A — неотрицательный ограниченный линейный оператор в гильбертовом пространстве H,* т. е. $(Ax, x) \geqslant 0$ *для всех* $x \in H$. *Пусть, далее, C — невырожденный ограниченный линейный оператор из H в гильбертово пространство H_1 (т. е. $CH = H_1$). Пусть, кроме того, $(Ax, x) \geqslant m\|x\|^2$, $m > 0$, при $x \in N_C = \{x \in H : Cx = 0\}$. Тогда для всякого $K > 0$ найдется $\gamma > 0$, такое, что*

$$((A + KC^*C)x, x) \geqslant \gamma \|x\|^2 \tag{26}$$

для всех $x \in H$.

Доказательство. Невырожденность оператора C эквивалентна условию $\|C^*y\| \geqslant \mu\|y\|$, $\mu > 0$, для всех $y \in H_1$ или $\|(CC^*)^{-1}\| \leqslant 1/\mu^2$ (C^* — оператор из H_1 в H, сопряженный к C). Зафиксируем произвольное $K > 0$. Для любого $x \in H$ справедливо разложение $x = x_1 + x_2$, где $x_1 \in N_C$, $x_2 \perp N_C$. При этом $x_2 = C^*(CC^*)^{-1}Cx$, откуда

$$\|x_2\| \leqslant \frac{\|C\|}{\mu^2}\|Cx\| = \frac{\|C\|}{\mu^2}(C^*Cx, x)^{1/2} \leqslant \frac{\|C\|}{\mu^2 K^{1/2}}((A + KC^*C)x, x)^{1/2}.$$

Используя это, получаем

$((A+KC^*C)x, x) \geqslant (Ax, x) = (Ax_1, x_1) + 2(Ax_1, x_2) + (Ax_2, x_2) \geqslant m\|x_1\|^2 - 2\|A\|\|x_1\|\|x_2\| \geqslant m\|x\|^2 - 2\|A\|\|x\|\|x_2\| - m\|x_2\|^2 \geqslant m\|x\|^2 - \dfrac{2\|A\|\|C\|}{\mu^2K^{1/2}}\|x\|((A + KC^*C)x, x)^{1/2} - \dfrac{m\|C\|^2}{\mu^4K}((A+KC^*C)x, x)$. Отсюда при $\dfrac{m}{2}\|x\| - \dfrac{2\|A\|\|C\|}{\mu^2K^{1/2}} \times ((A+KC^*C)x, x)^{1/2} \geqslant 0$ получаем $((A+KC^*C)x, x) \geqslant \dfrac{m}{2}\|x\|^2 - \dfrac{m\|C\|^2}{\mu^4K} \times ((A+KC^*C)x, x)$, т. е. $((A+KC^*C)x, x) \geqslant \dfrac{m}{2}\left(1+\dfrac{m\|C\|^2}{\mu^4K}\right)^{-1}\|x\|^2$. В противном случае $\dfrac{m}{2}\|x\| - \dfrac{2\|A\|\|C\|}{\mu^2K^{1/2}}((A+KC^*C)x, x)^{1/2} \leqslant 0$, но тогда $((A + KC^*C)x, x) \geqslant \left(\dfrac{m\mu^2K^{1/2}}{4\|A\|\|C\|}\right)^2\|x\|^2$. Итак, всегда $((A+KC^*C)x, x) \geqslant \gamma\|x\|^2$ с $\gamma = \min\left\{\dfrac{m}{2}\left(1+\dfrac{m\|C\|^2}{\mu^4K}\right)^{-1}, \left(\dfrac{m\mu^2K^{1/2}}{4\|A\|\|C\|}\right)^2\right\}$, что и требовалось доказать.

Лемма 5. *В условиях леммы 4 для всякого $K > 0$ найдется α, $0 < \alpha \leqslant 1$, такое что для всех $y \in H_1$*

$$\alpha\|y\|^2 \leqslant (KC(A+KC^*C)^{-1}C^*y, y) \leqslant \|y\|^2. \tag{27}$$

Если $0 < K_1 \leqslant K_2$, то

$$(K_1C(A+K_1C^*C)^{-1}C^*y, y) \leqslant (K_2C(A+K_2C^*C)^{-1}C^*y, y). \tag{28}$$

Доказательство. В силу леммы 4 существует такое $\gamma > 0$, что $\gamma\|x\|^2 \leqslant ((A+KC^*C)x, x) \leqslant (\|A\| + K\|C\|^2)\|x\|^2$, поэтому существует линейный ограниченный оператор $(A+KC^*C)^{-1}$ и $((A+KC^*C)^{-1}x, x) \geqslant \dfrac{1}{\|A\|+K\|C\|^2}\|x\|^2$. Получаем $(KC(A+KC^*C)^{-1}C^*y, y) = K((A+KC^*C)^{-1} \times C^*y, C^*y) \geqslant \dfrac{K}{\|A\|+K\|C\|^2}\|C^*y\|^2 \geqslant \dfrac{K\mu^2}{\|A\|+K\|C\|^2}\|y\|^2$, чем доказано левое неравенство (27). С другой стороны, поскольку $(Ax, x) \geqslant 0$ для всех $x \in H$, имеем

$$K\|C(A+KC^*C)^{-1}C^*y\|^2 = (KC^*C(A+KC^*C)^{-1}C^*y, (A+KC^*C)^{-1}C^*y) \leqslant ((A+KC^*C)(A+KC^*C)^{-1}C^*y, (A+KC^*C)^{-1}C^*y) = (C(A+KC^*C)^{-1}C^*y, y) \leqslant \|C(A+KC^*C)^{-1}C^*y\|\|y\|.$$

Отсюда $\|KC(A+KC^*C)^{-1}C^*y\| \leqslant \|y\|$, следовательно, справедливо правое неравенство (27). Наконец, так как $(((1/K_1)A + C^*C)x, x) \geqslant (((1/K_2)A+C^*C)x, x)$ при $0 < K_1 \leqslant K_2$ и произвольном $x \in H$, то $(K_1(A + K_1C^*C)^{-1}x, x) = (((1/K_1)A+C^*C)^{-1}x, x) \leqslant (((1/K_2)A+C^*C)^{-1}x, x) = (K_2(A+K_2C^*C)^{-1}x, x)$, откуда, полагая $x = C^*y$, получаем (28).

Отличие доказанных лемм от соответствующих утверждений [1] состоит в дополнительном требовании неотрицательности A ($(Ax, x) \geqslant 0$, $x \in H$), которое обеспечивает сильную положительность оператора $A + KC^*C$ (неравенство (26)) при любых $K > 0$. Если же не требовать неотрицательности A на всем H, предположив лишь его сильную положительность на подпространстве $Cx = 0$, то (26) будет иметь место только при достаточно больших $K > 0$ [1, лемма 1]. При этом за счет выбора достаточно большого K оператор $KC(A+KC^*C)^{-1}C^*$ может быть сделан сколь угодно близким по норме к единичному [1, лемма 3].

5. Для сходимости метода (21) достаточно, как показывает теорема 1, выполнения условий I и II. При более сильных предположениях можно дать оценку скорости сходимости итерационных последовательностей x^n, y^n. Будем предполагать, что в задаче (1) $Q = R^n$ и, кроме того, существует $x^* \in X^*$, для которого выполнены следующие условия.

III. В некоторой окрестности x^* функции $f(x)$, $g_i(x)$, $i = 1, \ldots, m$, дважды непрерывно дифференцируемы.

IV. Векторы $g_i'(x^*)$ при $i \in I^* = \{i: 1 \leqslant i \leqslant m,\ g_i(x^*) = 0\}$ линейно независимы.

Из III, IV вытекает существование $y^* \in Y^*$ и правило множителей Лагранжа: $L_x(x^*, y^*) = 0$, $(y^*, g(x^*)) = 0$, где $L_x(x^*, y^*)$ — градиент по x функции $L(x, y^*)$, вычисленный в точке x^*.

V. Если $(g_i'(x^*), x) = 0$ для всех $i \in I^*$, то $(L_{xx}(x^*, y^*)x, x) \geqslant m\|x\|^2$, $m > 0$ ($L_{xx}(x^*, y^*)$ — матрица вторых частных производных функции $L(x, y^*)$ в точке x^*).

VI. Выполнены условия строгой дополняющей нежесткости, т. е. $y_i^* > 0$ при $i \in I^*$.

Предположения III — VI суть достаточные условия того, что x^* — единственная и невырожденная точка минимума $f(x)$ при ограничениях $g(x) \leqslant 0$ [5, гл. II]. Из III — VI следует, таким образом, выполнение условий I и II. При этом Y^* также состоит из единственной точки y^*.

Лемма 6. *Пусть задача* (1) *с* $Q = R^n$ *имеет решение* x^*, *для которого выполнены условия III — VI. Тогда для всякого* $\varepsilon > 0$ *найдется такое* $\delta > 0$, *что при* $K_n \geqslant K > 0$, $\dfrac{\|y^0 - y^*\|}{K} < \delta$ *будет* $\|x^n - x^*\| < \varepsilon$, $\|y^{n+1} - y^*\| < \varepsilon$ *для всех* $n \geqslant 0$, *а метод* (21) *принимает вид*

$$f'(x^n) + \sum_{i \in I^*} [y_i^n + K_n g_i(x^n)] g_i'(x^n) = 0, \tag{29}$$

$$y_i^{n+1} = \begin{cases} y_i^n + K_n g_i(x^n), & i \in I^*, \\ 0, & i \overline{\in} I^*, \end{cases} \tag{30}$$

$$n = 0, 1, \ldots .$$

Доказательство. В силу условия VI имеем $\min\{y_i^*, i \in I^*\} = \Delta_1 > 0$. Кроме того, $\min\{-g_i(x^*), i \overline{\in} I^*\} = \Delta_2 > 0$. Положим $\Delta = \min\{\Delta_1, \Delta_2\}$. Заметим, что при всех n точка x^n является решением задачи

$$f(x) \to \min,$$

$$g(x) \leqslant \left[\frac{1}{K_n} y^n + g(x^n)\right]_+ - \frac{1}{K_n} y^n, \tag{31}$$

а y^{n+1} — решением задачи, двойственной к (31). Действительно, x^n удовлетворяет ограничениям (31). Если $\bar{x}$ также удовлетворяет этим ограничениям, то $y^n / K_n + g(\bar{x}) \leqslant [y^n / K^n + g(x^n)]_+$ и потому $\|[y^n / K_n + g(\bar{x})]_+\| \leqslant \|[y^n / K_n + g(x^n)]_+\|$. Если допустить, что при этом $f(\bar{x}) < f(x^n)$, то, как видно из (2), $M(\bar{x}, y^n, K_n) < M(x^n, y^n, K_n)$, что противоречит (21). Поскольку x^n — точка минимума $M(x, y^n, K_n)$, то

$$\begin{aligned} 0 = M_x(x^n, y^n, K_n) &= f'(x^n) + [g'(x^n)]^* [y^n + K_n g(x^n)]_+ = \\ &= f'(x^n) + [g'(x^n)]^* y^{n+1}, \end{aligned} \tag{32}$$

причем $(y^{n+1}, g(x^n) - [y^n / K_n + g(x^n)]_+ + y^n / K_n) = ([y^n / K_n + g(x^n)]_+, y^n / K_n + g(x^n) - [y^n / K_n + g(x^n)]_+) = 0$. Но это означает, что y^{n+1} — множитель Лагранжа для задачи (31).

Из неравенства (23), полученного при доказательстве теоремы 1, следует, что $\|y^n - y^*\| \leqslant \|y^0 - y^*\|$, $n \geqslant 0$. Отсюда и из (11) получаем $\|[y^n/K_n + g(x^n)]_+ - y^n/K_n\| \leqslant (1/K_n)\|y^n - y^*\| \leqslant (1/K_n)\|y^0 - y^*\| \leqslant$ $\leqslant (1/K)\| \, \|y^0 - y^*\|$, поэтому правая часть ограничений (31) может быть за счет малости $\dfrac{\|y_0 - y^*\|}{K}$ сделана сколь угодно малой по норме при всех $n \geqslant 0$. При соблюдении условий III–VI можно воспользоваться теоремой об устойчивости решения задачи математического программирования [5, стр. 52]. Получаем, что при достаточно малых $\dfrac{\|y^0 - y^*\|}{K}$ точки x^n и y^{n+1} для всех $n \geqslant 0$ сколь угодно близки к x^* и y^* соответственно. В частности, поскольку $y_i^* \geqslant \Delta$ при $i \in I^*$ и $g_i(x^*) \leqslant -\Delta$ при $i \overline{\in} I^*$, то с учетом непрерывности $g_i(x)$ можно утверждать существование такого δ, $0 < \delta < \Delta/2$, что при $\dfrac{\|y^0 - y^*\|}{K} < \delta$ будет $y_i^{n+1} > \dfrac{\Delta}{2}$, $i \in I^*$, и $g_i(x^n) < -\Delta/2$, $i \overline{\in} I^*$, для всех n. Получаем

$$[y_i^n + K_n g_i(x^n)]_+ = y_i^{n+1} > \frac{\Delta}{2} > 0, \quad i \in I^*, \quad n \geqslant 0. \tag{33}$$

Кроме того, поскольку $y_i^n - y_i^* \leqslant \|y^n - y^*\| \leqslant \|y^0 - y^*\|$ и $y_i^* = 0$ при $i \overline{\in} I^*$, будем иметь

$$y_i^n + K_n g_i(x^n) \leqslant \|y^0 - y^*\| - K\frac{\Delta}{2} < K\left(\delta - \frac{\Delta}{2}\right) < 0, \quad i \overline{\in} I^*, \quad n \geqslant 0. \tag{34}$$

Из (33) и (34) вытекает (30), а из (30) и (32) следует (29). Лемма доказана.

Условие (29) означает, что x^n — стационарная точка функции $M^*(x, y^n, K_n) = f(x) + \sum\limits_{i \in I^*} \left[y_i^n g_i(x) + \dfrac{K_n}{2} g_i^2(x) \right]$. Иначе говоря, при малых $\dfrac{\|y^0 - y^*\|}{K}$ метод (21) для задачи (1) совпадает с методом штрафных оценок [1] для следующей задачи с ограничениями в форме равенств

$$\begin{gathered} f(x) \to \min, \\ g_i(x) = 0, \; i \in I^*. \end{gathered} \tag{35}$$

В действительности, как нетрудно доказать, $M^*(x, y^n, K_n)$ строго выпукла в окрестности x^n, так что x^n — локально единственная точка минимума $M^*(x, y^n, K_n)$, а не просто стационарная точка. Более того, хотя $M^*(x, y^n, K_n)$ не является выпуклой всюду, x^n — точка ее глобального минимума.

Теорема 2. *Пусть задача* (1) *с* $Q = R^n$ *имеет решение* x^*, *для которого выполнены условия* III–VI. *Тогда для любых* $K_n \geqslant K > 0$ *найдется* $\rho > 0$ *такое, что при* $\|y^0 - y^*\| < \rho$ *итерационные последовательности* (21) *для всех* $n \geqslant 0$ *удовлетворяют неравенствам*

$$\begin{gathered} \|x^n - x^*\| \leqslant c\|y^n - y^*\|, \; c > 0, \\ \|y^{n+1} - y^*\| \leqslant q\|y^n - y^*\|, \; 0 < q < 1. \end{gathered} \tag{36}$$

Доказательство. Можем считать, что $\sup_n K_n < +\infty$, так как противоположный случай охватывается нижеследующей теоремой 3. В силу леммы 6 для любых $K_n \geqslant K > 0$ найдутся столь малые значения $\|y^0 - y^*\|$, что будет выполняться (29). При этом x^n будет близок к x^*, поэтому, используя гладкость $f(x)$, $g_i(x)$ (условие III), можно при $\sup_n K_n < +\infty$ записать (29) в виде

$$f'(x^*)+f''(x^*)(x^n-x^*)+\sum_{i\in I^*}[y_i^n+K_n(g_i'(x^*),x^n-x^*)]\times$$
$$\times[g_i'(x^*)+g_i''(x^*)(x^n-x^*)]+o(\|x^n-x^*\|)=0. \quad (37)$$

Положим $A = L_{xx}(x^*, y^*) = f''(x^*) + \sum_{i\in I^*} y_i^* g_i''(x^*)$. Обозначим через m^* число индексов в I^*, $m^* \leqslant m$; через C — матрицу $m^* \times n$, которая получается из $g'(x^*)$ вычеркиванием строк с номерами $i \overline{\in} I^*$; через $\tilde{y}$ — вектор длины m^*, который получается из $y = (y_1, \ldots, y_m)$ вычеркиванием компонент с номерами $i \overline{\in} I^*$. Тогда, учитывая, что $f'(x^*) + \sum_{i\in I^*} y_i^* g_i'(x^*) = 0$, получим из (37)

$$(A + K_n C^* C)(x^n - x^*) + C^*(\tilde{y}^n - \tilde{y}^*) + o(\|x^n - x^*\| + \|\tilde{y}^n - \tilde{y}^*\|) = 0. \quad (38)$$

Аналогично из (30) находим

$$\tilde{y}^{n+1} - \tilde{y}^* = \tilde{y}^n - \tilde{y}^* + K_n C(x^n - x^*) + o(\|x^n - x^*\|). \quad (39)$$

Из предположений IV, V и выпуклости $f(x)$, $g_i(x)$ вытекает, что матрицы A и C удовлетворяют условиям леммы 4 ($H = R^n$, $H_1 = R^{m^*}$). Согласно лемме 4, для данного K, $K_n \geqslant K > 0$ найдется такое $\gamma > 0$, что $((A + K_n C^* C)x, x) \geqslant ((A + KC^*C)x, x) \geqslant \gamma(x, x)$. Отсюда следует, что существуют $(A + K_n C^* C)^{-1}$, причем $\|(A + K_n C^* C)^{-1}\| \leqslant \frac{1}{\gamma}$ для всех n. Поэтому (38) можно переписать в виде

$$x^n - x^* = -(A + K_n C^* C)^{-1} C^* (\tilde{y}^n - \tilde{y}^*) + o(\|x^n - x^*\| + \|\tilde{y}^n - \tilde{y}^*\|), \quad (40)$$

откуда, в частности, следует, что при достаточно малых $\|\tilde{y}^n - \tilde{y}^*\|$ имеет место

$$\|x^n - x^*\| \leqslant \frac{2\|C\|}{\gamma} \|\tilde{y}^n - \tilde{y}^*\|. \quad (41)$$

Подставляя (40) в (39) и учитывая, что ввиду (41) можно заменить $o(\|x^n - x^*\| + \|\tilde{y}^n - \tilde{y}^*\|)$ на $o(\|\tilde{y}^n - \tilde{y}^*\|)$, получаем

$$\tilde{y}^{n+1} - \tilde{y}^* = \tilde{y}^n - \tilde{y}^* - K_n C(A + K_n C^* C)^{-1} C^* (\tilde{y}^n - \tilde{y}^*) + o(\|\tilde{y}^n - \tilde{y}^*\|) =$$
$$= (I - K_n C(A + K_n C^* C)^{-1} C^*)(\tilde{y}^n - \tilde{y}^*) + o(\|\tilde{y}^n - \tilde{y}^*\|), \quad (42)$$

где I — единичная матрица порядка m^*. В силу леммы 5 найдется такое α, $0 < \alpha \leqslant 1$, что

$$\alpha(\tilde{y}, \tilde{y}) \leqslant (KC(A + KC^*C)^{-1} C^* \tilde{y}, \tilde{y}) \leqslant (K_n C(A +$$
$$+ K_n C^* C)^{-1} C^* \tilde{y}, \tilde{y}) \leqslant (\tilde{y}, \tilde{y})$$

для всех $\tilde{y} \in R^{m^*}$. Поэтому

$$0 \leqslant ([I - K_n C(A + K_n C^* C)^{-1} C^*] \tilde{y}, \tilde{y}) \leqslant (1 - \alpha)(\tilde{y}, \tilde{y}). \quad (43)$$

Известно [8, гл. 6], что при условии (43) итерационная последовательность, удовлетворяющая (42), сходится к нулю при малых $\|\tilde{y}^n - \tilde{y}^*\|$

не медленнее геометрической прогрессии, т. е.

$$\|\tilde{y}^{n+1}-\tilde{y}^*\| \leqslant q\|\tilde{y}^n-\tilde{y}^*\|, \quad 0<q<1. \tag{44}$$

Но ввиду неравенств $\|\tilde{y}^n-\tilde{y}^*\| \leqslant \|y^n-y^*\| \leqslant \|y^0-y^*\|$ можно сделать $\|\tilde{y}^n-\tilde{y}^*\|$ сколь угодно малыми за счет малости $\|y^0-y^*\|$. Поэтому из (41) и (44) следует утверждение теоремы.

Оценки (36) показывают, что при любых $K_n \geqslant K > 0$ достаточно хорошее начальное приближение по двойственным переменным обеспечивает линейную сходимость обеих итерационных последовательностей, именно: $\|x^n-x^*\| \leqslant cq^n\|y^0-y^*\|$, $\|y^{n+1}-y^*\| \leqslant q^{n+1}\|y^0-y^*\|$ при $n \geqslant 0$. Поскольку по теореме 1 при любом $y^0 \geqslant 0$ имеет место сходимость $y^n \to y^*$, то, комбинируя теоремы 1 и 2, получаем, что при произвольном начальном приближении по двойственным переменным сходимость станет линейной после надлежащего числа итераций.

Следствие. При условиях теоремы 2, для любых $K_n \geqslant K > 0$, $y^0 \geqslant 0$ найдется такое N, что при $n \geqslant N$ итерационные последовательности $\{x^n\}$ и $\{y^n\}$ удовлетворяют неравенствам (36).

Линейную сходимость можно было установить и несколько иным путем. Можно показать, что в достаточно малой окрестности y^* двойственная функция $\psi(y, K) = \min\limits_{x \in R^n} M(x, y, K)$ строго вогнута и дважды непрерывно дифференцируема по y при любых $K > 0$. Рассматривая метод (21) при каждом n как итерацию градиентного метода максимизации $\psi(y, K_n)$ с длиной шага K_n, можно получить (36) из теоремы о сходимости градиентного метода [9, теорема 6] (соблюдение условий этой теоремы обеспечивается леммами 4 и 5).

6. Остановимся теперь на сравнении метода штрафных оценок с методом Удзавы [4, гл. 10, § 4, 5], который для задачи (1) с $Q = R^n$ имеет вид

$$L(x^n, y^n) = \min_{x \in R^n} L(x, y^n),$$

$$y^{n+1} = [y^n + \alpha L_y(x^n, y^n)]_+, \tag{45}$$

где $L(x, y)$ — функция Лагранжа задачи (1), $\alpha > 0$ — постоянная длина шага. Очевидно, если заменить в (45) обычную функцию Лагранжа $L(x, y)$ на модифицированную $-M(x, y, K)$, а также α на K_n, то придем к методу (21).

В [4, гл. 10, § 4] получен результат о глобальной сходимости метода (45) в предположениях, существенно более жестких, чем в теореме 1 настоящей работы: если $f(x)$ строго выпукла и соблюдается условие Слейтера, то метод (45) при любом $y^0 \geqslant 0$ приводит в произвольную окрестность множества Y^*, когда α выбрано достаточно малым. Отказ от условия Слейтера, гарантирующего компактность Y^*, ведет, вообще говоря, к расходимости (45), тогда как метод (21) сходится и без условий такого типа (см. теорему 1). Это ослабление условий сходимости достигается за счет модификации функции Лагранжа — для применения алгоритма типа Удзавы функция $M(x, y, K)$ предпочтительнее по своим свойствам, чем $L(x, y)$ (см. леммы 2 и 3). Отметим, что для расширения области приложимости других алгоритмов, использующих функцию Лагранжа, могут оказаться целесообразными другие модификации $L(x, y)$ (см. [10, 11]).

В [12, теорема 2] для задачи с ограничениями в форме равенств и при условиях регулярности типа III–VI доказано, что для достаточно малых $\alpha > 0$ метод (45) сходится в окрестности y^* со скоростью геометрической прогрессии. Аналогом этого результата для метода (21) является теорема 2 настоящей работы. Как показывает сравнение условий обеих теорем, в методе штрафных оценок линейная сходимость имеет место при более слабых предположениях о функциях: вместо положительности $L_{xx}(x^*, y^*)$ на всем пространстве (т. е. строгой выпуклости $L(x, y^*)$ вблизи x^*) достаточно соблюдения V. Подчеркнем также, что требуемая для сходимости (45) длина шага α априорно обычно неизвестна, и не вполне ясно, как ее подбирать при практическом применении метода. В методе штрафных оценок (21) этой проблемы, очевидно, не возникает.

Как в (45), так и в (21) при малых K_n сходимость может оказаться медленной (прогрессия со знаменателем, близким к единице). Однако в отличие от (45) в методе (21) сходимость итерационной последовательности можно ускорить. Следующая теорема показывает, что за счет увеличения K_n знаменатель прогрессии, мажорирующей расстояние до точки минимума, может быть сделан как угодно малым. Эта теорема опирается на результаты [1].

Теорема 3. *Пусть задача* (1) *с* $Q=R^n$ *имеет решение* x^*, *для которого выполнены условия III — VI и, кроме того,* $f''(x)$, $g_i''(x)$, $i \in I^*$ *удовлетворяют условию Липшица в некоторой окрестности* x^*. *Тогда для любого* $y^0 \geqslant 0$ *найдется столь большое* $K=K(\|y^0-y^*\|)>0$, *что при* $K_n \geqslant K$ *итерационные последовательности* (21) *для всех* $n \geqslant 0$ *удовлетворяют неравенствам*

$$\|x^n-x^*\| \leqslant \frac{c}{K_n}\|y^n-y^*\|,$$
$$\|y^{n+1}-y^*\| \leqslant \frac{c}{K_n}\|y^n-y^*\|, \tag{46}$$

где постоянная $c>0$ не зависит от K_n.

Доказательство. Вновь, как и при доказательстве теоремы 2, воспользуемся леммой 6, однако на этот раз ее утверждения будут иметь место не за счет малости $\|y^0-y^*\|$, а за счет больших K_n. По лемме 6 для произвольного $y^0 \geqslant 0$ найдется такое K, зависящее от $\|y^0-y^*\|$, что метод (21) примет вид (29), (30), а x^n, y^n будут близки к x^*, y^*. При соблюдении условий теоремы можно воспользоваться результатами [1], применяя их к методу (29), (30) для задачи на условный экстремум (35). Доказываемое утверждение следует непосредственно из теоремы 1 работы [1].

Из (46) имеем $\|x^n-x^*\| \leqslant (c/K_n)\dots(c/K_0)\|y^0-y^*\|$, $\|y^{n+1}-y^*\| \leqslant$ $\leqslant (c/K_n)\dots(c/K_0)\|y^0-y^*\|$. Таким образом, увеличивая K_n, можно добиться более быстрой сходимости метода. В частности, при $K_n \equiv K$ метод сходится не медленнее прогрессии, знаменатель которой c/K может быть сделан сколь угодно малым за счет выбора K. В связи с этим подчеркнем важное различие между методом (21) и методом штрафных функций [5], в котором для решения задачи (1) последовательно ищут точки минимума функций $P(x, K_n)=M(x, 0, K_n)=f(x)+(K_n/2)\|[g(x)]_+\|^2$ при $K_n \to +\infty$, т. е. строят итерационную последовательность x^n по формулам

$$K_n \to +\infty,$$
$$P(x^n, K_n)=\min_x P(x, K_n). \tag{47}$$

Известно, что численная реализация (47) сопряжена с серьезными трудностями. Это связано с тем, что пропорционально росту K_n ухудшается обусловленность функции $P(x, K_n)$ и, как следствие, возрастает трудоемкость отыскания ее минимума, поскольку большинство методов безусловной минимизации так или иначе чувствительны к обусловленности. Иначе говоря, в методе штрафных функций на каждой итерации приходится решать все более трудную в вычислительном отношении задачу. По этой причине метод (47) часто не позволяет найти решение исходной задачи с нужной точностью. В методе штрафных оценок, благодаря использованию приближений y^n для множителей Лагранжа, не требуется неограниченно увеличивать K_n; достаточно, подобрав в ходе итерационного процесса некоторое K, обеспечивающее приемлемую скорость сходимости, в дальнейшем зафиксировать $K_n \equiv K$. (Заметим, что если взять $K_n \to +\infty$, то метод, как видно из (46), будет сходиться сверхлинейно, т. е. быстрее любой убывающей геометрической прогрессии.) Вопросы численной реализации метода штрафных оценок обсуждаются в [1].

В заключение отметим результаты о методе (21), полученные в [3] для невыпуклой задачи с ограничениями в форме неравенств. Разумеется, для невыпуклой задачи удается получить лишь результаты о локальной сходимости к изолированной точке локального минимума. Вместо выпуклости функций роль основного требования играет компактность допустимой области $\{x: g(x) \leqslant 0\}$. Сходимость в этом случае имеет место не при любых, а лишь при достаточно больших значениях K_n, и для того, чтобы K_n могли оставаться ограниченными сверху (например, чтобы иметь возможность взять $K_n \equiv K$), требуются довольно жесткие предположения регулярности типа условий III—VI. При этих предположениях в [3] получен результат, близкий к теореме 3, — локальная сходимость со скоростью прогрессии, знаменатель которой стремится к нулю с ростом K. В [3], кроме того, описана вычислительная схема метода, в которой K_n подбираются в процессе счета так, чтобы обеспечить заданную скорость убывания невязок ограничений.

Автор выражает глубокую благодарность Б. Т. Поляку за постоянное внимание к работе, а также Е. Г. Гольштейну за полезное обсуждение результатов.

ЛИТЕРАТУРА

1. Б. Т. Поляк, Н. В. Третьяков. Метод штрафных оценок для задач на условный экстремум. Ж. вычисл. матем. и матем. физ., 1973, т. 13, № 1.
2. Б. Т. Поляк, Н. В. Третьяков. Об одном итерационном методе линейного программирования и его экономической интерпретации. Экономика и матем. методы, 1972, т. VIII, вып. 5.
3. A. P. Wierzbicky. A Penalty Function Shifting Method in Constrained Static Optimization and its Convergence Properties. Arch. Autom. i Telemech., 1971, v. 16, N 4.
4. К. Дж. Эрроу, Л. Гурвиц, Х. Удзава. Исследования по линейному и нелинейному программированию. М., Изд-во иностр. лит., 1962.
5. А. Фиакко, Г. Мак-Кормик. Нелинейное программирование. Методы последовательной безусловной минимизации. М., «Мир», 1972.
6. Е. Г. Гольштейн. Теория двойственности в математическом программировании и ее приложения. М., «Наука», 1971.
7. Б. Н. Пшеничный. Необходимые условия экстремума. М., «Наука», 1969.
8. Р. Беллман. Теория устойчивости решений дифференциальных уравнений. М., Изд-во иностр. лит., 1954.
9. Б. Т. Поляк. Градиентные методы минимизации функционалов. Ж. вычисл. матем. и матем. физ., 1963, т. 3, № 4.
10. Б. И. Коробочкин. Игровой метод решения обобщенной задачи выпуклого программирования. В сб. Первая Всесоюзная конференция по оптимальному планированию и управлению народным хозяйством. Тезисы докл., М., 1971 (ЦЭМИ АН СССР).
11. Е. Г. Гольштейн. Обобщенный градиентный метод отыскания седловых точек. Экономика и матем. методы, 1972, т. VIII, вып. 4.
12. Б. Т. Поляк. Итерационные методы, использующие множители Лагранжа для решения экстремальных задач с ограничениями типа равенств. Ж. вычисл. матем. и матем. физ., 1970, т. 10, № 5.

Поступила в редакцию
13 XI 1972