ТЕХНИЧЕСКАЯ КИБЕРНЕТИКА
No.1 · 1983 · с. 47 – 59

УДК 519.8

ОБ ОДНОМ ПОДХОДЕ К СИСТЕМАТИЗАЦИИ ЧИСЛЕННЫХ МЕТОДОВ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

ЕВТУШЕНКО Ю.Г., ЖАДАН В.Г.

Введение. Начатые в пятидесятых годах работы по созданию методов решения задач нелинейного программирования (НЛП) привели к появлению многочисленных публикаций, давших широкий спектр разнообразных численных методов. В последние годы появились пакеты прикладных программ, в которых были реализованы многие из существующих методов решения задач НЛП. Программная реализация методов, попытки создания управляющих программ привели к необходимости систематизации существующих методов, к выяснению истоков отдельных методов. Возник вопрос: нельзя ли с единой позиции описать большинство из существующих методов НЛП? В данной статье делается попытка дать положительный ответ на этот вопрос. Предлагается многие из существующих методов решения задач НЛП трактовать как методы решения систем нелинейных уравнений, в которых часть переменных определяется из решения вспомогательных задач отыскания безусловного экстремума функций многих переменных. При этом методы весьма условно делятся на три группы: прямые (те, в которых решаются системы уравнений для исходных переменных), непрямые (в которых решаются системы для вспомогательных переменных) и прямые-двойственные (в которых решаются системы для исходных и двойственных переменных).

1. Постановка задачи, основные обозначения. Рассмотрим общую задачу нелинейного программирования об отыскании

$$\min_{x\in X} f(x), \qquad X = \{x \in \mathbb{E}^n : g(x) = 0,\ h(x) \le 0\}. \tag{1.1}$$

Здесь $\mathbb{E}^i$ есть i-мерное евклидово пространство, $x = [x^1, \dots, x^n]$. Функции f, g, h, определяющие задачу НЛП, по меньшей мере всюду непрерывны и осуществляют отображения

$$f : \mathbb{E}^n \to \mathbb{E}^1, \qquad g : \mathbb{E}^n \to \mathbb{E}^e, \qquad h : \mathbb{E}^n \to \mathbb{E}^c.$$

Множество решений задачи (1.1) обозначим X_*. Определим

$$X^0 = \{x \in \mathbb{E}^n : g(x) = 0,\ h(x) < 0\}.$$

Всюду предполагается, что множества X_* и X^0 непусты и замыкание X^0 совпадает с X.

Обозначим через $m = e + c$ общее число ограничений в задаче (1.1), через $\mathbb{E}^c_+$ и $\mathbb{E}^c_-$ — неотрицательный и неположительный ортанты $\mathbb{E}^c$, т.е. совокупности таких векторов $v \in$

$\in \mathbb{E}^c$, у которых все координаты, соответственно, либо неотрицательны, либо неположительны. Обозначим $\tilde{\mathbb{E}}^1 = \mathbb{E}^1 \cup \{+\infty\}$, $\tilde{\mathbb{E}}^1_+ = \mathbb{E}^1_+ \cup \{+\infty\}$. Под $\|a\|$, если не оговорено противное, будем понимать евклидову норму вектора a. Границу, внутренность и замыкание множества D обозначим, соответственно, $\Gamma(D)$, $\operatorname{int} D$, и $\bar{D}$. Символ внизу рядом с обозначением функции указывает на ее производную по этому аргументу.

Всякому вектору p соответствуют векторы p_+ и p_-, у которых i-е компоненты выражаются через p^i по, соответственно, формулам

$$p^i_+ = \max[0, p^i], \qquad p^i_- = \min[0, p^i].$$

Аналогично определяются положительные и отрицательные части функции $\varphi(x)$:

$$\varphi_+(x) = \max[0, \varphi(x)], \qquad \varphi_-(x) = \min[0, \varphi(x)].$$

Непрерывная при всех $x \in \mathbb{E}^n$ функция, обращающаяся в нуль при $x \in X$ и строго большая нуля при $x \notin X$, называется **штрафной** и обозначается $S(x)$.

Введем функцию Лагранжа

$$L(x, u, v) = f(x) + \langle u, g(x)\rangle + \langle v, h(x)\rangle.$$

Здесь $\langle a, b\rangle$ — скалярное произведение векторов a и b, $u \in \mathbb{E}^c$, $v \in \mathbb{E}^c_+$. Будем говорить, что $[x_*, u_*, v_*]$ является седловой точкой функции Лагранжа, если для любых $x \in \mathbb{E}^n$, $u \in \mathbb{E}^e$, $v \in \mathbb{E}^c_+$ имеют место неравенства

$$L(x_*, u, v) \le L(x_*, u_*, v_*) \le L(x, u_*, v_*).$$

Если функции, определяющие задачу, дифференцируемы, выполнено условие регулярности ограничений, то необходимое условие минимума состоит в существовании такой точки $[x_*, u_*, v_*]$, что

$$L_x(x_*, u_*, v_*) = 0, \tag{1.2}$$

$$x_* \in X, \quad v_* \in \mathbb{E}^c_+, \quad \langle v_*, h(x_*)\rangle = 0. \tag{1.3}$$

Точки $[x_*, u_*, v_*]$, удовлетворяющие условиям (1.2), (1.3), будем называть точками Куна–Таккера. Множество точек x_* таких, что x_* вместе с некоторыми u_* и v_* образуют точку Куна–Таккера, обозначим X^1_*.

2. Общая схема непрямых методов. Введем функцию $H(x, y)$ и вектор-функцию $G(x, y)$, зависящие от x и некоторого вектора y из множества M. Размерность G и y пока не конкретизируем. Рассмотрим вспомогательную задачу безусловной минимизации

$$\min_{x \in \mathbb{E}^n} H(x, y). \tag{2.1}$$

Предполагая, что ее решение существует, определим точечно-множественное отображение

$$X(y) = \operatorname{Arg} \min_{x \in \mathbb{E}^n} H(x, y).$$

Определение 2.1. *Точку $z_* = [x_*, y_*] \in \mathbb{E}^n \times M$ назовем особой точкой пары функций $\{H, G\}$, если выполнены условия*

$$x_* \in X(y_*), \qquad G(x_*, y_*) = 0. \tag{2.2}$$

Определение 2.2. *Точку $z_* = [x_*, y_*] \in \mathbb{E}^n \times \bar{M}$ назовем обобщенной особой точкой пары функций $\{H, G\}$, если можно указать такие последовательности $\{x_k\}$, $\{y_k\}$, что $x_k \to x_*$, $y_k \to y_*$, $y_k \in M$ и выполнены условия*

$$x_k \in X(y_k), \qquad \lim_{k\to\infty} G(x_k, y_k) = 0. \tag{2.3}$$

Множество всех особых точек (в том числе и обобщенных) пары функций $\{H, G\}$ обозначим через $Z_*(H, G)$.

Определение 2.3. *Пара $\{H, G\}$ согласована с задачей* (1.1)*, если всякая ее особая точка $[x_*, y_*] \in Z_*(H, G)$ такова, что $x_* \in X_*$.*

Если для (1.1) построена согласованная с ней пара $\{H, G\}$, найдена ее особая точка $[x_*, y_*]$, то задача (1.1) решена, так как $x_* \in X_*$. Этот путь дает возможность использовать для решения задачи (1.1) разнообразные численные методы отыскания решений систем нелинейных уравнений в сочетании с методами безусловной минимизации функций многих переменных. Простейший подход основан на применении зависимости $x(y)$, полученной из решения вспомогательной задачи (2.1). Подставив $x(y)$ во второе условие (2.2), приходим к уравнению

$$G(x(y), y) = 0. \tag{2.4}$$

Предположим, что размерности G и y совпадают. Для решения (2.4) воспользуемся методом простой итерации:

$$y_{k+1} = y_k + \alpha G(x_k, y_k), \qquad x_k \in X(y_k), \tag{2.5}$$

где α — некоторый коэффициент, обеспечивающий сходимость процесса (в ряде случаев его можно полагать равным единице). Если функции $G(x, y)$ и $x(y)$ достаточно гладкие, то применение метода Ньютона приводит к формулам

$$y_{k+1} = y_k - \left[\frac{d}{dt} G(x(y_k), y_k)\right]^{-1} G(x_k, y_k), \qquad x_k = x(y_k) \in X(y_k). \tag{2.6}$$

В задачах оптимизации часто решение исходной задачи заменяется нахождением точек, удовлетворяющих необходимым условиям экстремума. Основываясь на этой идее, построим другой подход к отысканию особых точек. Пусть функция H дифференцируема по x. Для того, чтобы точка $[x_*, y_*]$ была особой, необходимо, чтобы

$$H_x(x_*, y_*) = 0, \qquad G(x_*, y_*) = 0. \tag{2.7}$$

Для отыскания точек, удовлетворяющих (2.7), будем при каждом x определять

$$Y_1(x) = \operatorname{Arg}\min_{y\in M} \|G(x, y)\|,$$

а для решения первого уравнения использовать метод простой итерации

$$x_{k+1} = x_k - \alpha H_x(x_k, y_k), \qquad y_k \in Y_1(x_k). \tag{2.8}$$

Коэффициент α можно сделать переменным, в частности, выбирать из условия минимизации функции $H(x, y_k)$ вдоль направления $-H_x(x_k, y_k)$, либо использовать другие приближенные схемы. Тогда процесс (2.8) превращается, по существу, в метод градиентного спуска.

В том случае, когда функция H не дифференцируема, первое из условий (2.7) заменяется на

$$\inf_{\|s\|=1} \frac{\partial_x H(x_*, y_*)}{\partial s} \geq 0,$$

где через $\partial_x H(x_*, y_*)/\partial s$ обозначена производная функции H по x вдоль направления s. Вместо (2.8) итерации ведутся по схеме

$$x_{k+1} = x_k + \alpha_k s(x_k, y_k), \qquad y_k \in Y_1(x_k). \tag{2.9}$$

Здесь: $s(x_k, y_k)$ — направление убывания функции $H(x, y_k)$ в точке x_k; α — переменный шаг спуска.

Широкий класс методов получается, если положить

$$H(x, y) = f(x) + \xi(x, y). \tag{2.10}$$

Переформулируем для этого случая доказанную в [1] теорему 1.6.4.

Теорема 2.1. *Пусть существует особая точка $[x_*, y_*]$ пары $\{H, G\}$; функции $G(x, y)$ и $\xi(x, y)$ таковы, что из условий $G(x_*, y_*) = 0$, $x \in X$ следует, что $x_* \in X$ и*

$$\xi(x, y_*) \leq \xi(x_*, y_*); \tag{2.11}$$

тогда $x_* \in X_*$.

В [1] доказано более сильное утверждение: $X_* = X \cap X(y_*)$. Обобщением теоремы 2.1 является следующая теорема.

Теорема 2.2. *Пусть $[x_*, y_*]$ — обобщенная особая точка пары функций $\{H, G\}$; функции $G(x, y)$ и $\xi(x, y)$ таковы, что из условия существования последовательностей $x_k \to x_*$, $y_k \to y_*$, удовлетворяющих (2.3), следует, что $x_* \in X$ и для любого $x \in X$*

$$\lim_{k\to\infty} \xi(x, y_k) \leq \lim_{k\to\infty} \xi(x_k, y_k); \tag{2.12}$$

тогда $x_* \in X_*$.

Таким образом, если для всех точек из множества $Z_*(H, G)$ выполнены условия либо теоремы 2.1, либо теоремы 2.2, то пара функций $\{H, G\}$ согласована с задачей (1.1). В тех случаях, когда $X^0 \neq \emptyset$, проверку (2.12) можно проводить только для точек $x \in X^0$.

3. Точные штрафные функции. Положим в (2.10)

$$\xi(x, y) = y\|\Phi(x)\|_q. \tag{3.1}$$

Здесь $y \in \mathbb{E}^1$, $\|\Phi\|_q$ — норма Гельдера вектора Φ:

$$\Phi(x) = [|g^1(x)|, \ldots, |g^e(x)|, h_+^1(x), \ldots, h_+^c(x)], \qquad \|\Phi(x)\|_q = \left[\sum_{i=1}^{m} |\Phi^i(x)|^q\right]^{1/q}.$$

Через $\|\cdot\|_r$, $1/q + 1/r = 1$, обозначим норму, сопряженную к $\|\cdot\|_q$, и положим

$$y_* = \|w_*\|_r, \quad w_* = [u_*, v_*] \in \mathbb{E}^m.$$

Очевидно, что $\|\Phi(x)\|_q$ является штрафной функцией.

Теорема 3.1. *Пусть в задаче* (1.1) *существует седловая точка* $[x_*, u_*, v_*]$ *функции Лагранжа; функция* H *задается* (2.10) *и* (3.1). *Тогда* $X(y) \subseteq X_*$ *при любых* $y > y_*$.

Таким образом, функция H, задаваемая (2.10) и (3.1), обладает замечательным свойством: для нее множество $X(y)$ совпадает с X_*, если только $y > y_*$. Решение исходной задачи сводится к безусловной минимизации. Функция ξ, обеспечивающая это свойство, называется точной штрафной функцией. Впервые на существование таких функций указали И.И. Еремин [2] и У. Зангвилл [3]. Доказательство теоремы 3.1 можно найти в [1]. К сожалению, функция ξ не дифференцируема в точках $x \in \Gamma(X)$, что существенно усложняет численное решение задачи (2.1). Поэтому целесообразно искать другие пути решения исходной задачи (1.1).

4. Методы штрафных функций. В этом пункте предполагается для простоты, что в (1.1) отсутствуют ограничения типа равенства. Пусть $y \in \mathbb{E}^1$, $y > 0$. Функции $\xi(x,y)$ и $G(x,y)$ будем строить в виде

$$\xi(x,y) = yR(h(x)), \tag{4.1}$$

$$G(x,y) = \begin{cases} P(R(h(x)) - R(0), y), & \text{если } R(0) < +\infty, \\ -Q(y) & \text{в противном случае.} \end{cases} \tag{4.2}$$

На функции $R(p)$, $P(t,\theta)$ и $Q(t)$, отображающие, соответственно, $\mathbb{E}^c$ в $\mathbb{E}^1$, $\tilde{\mathbb{E}}^1 \times \mathbb{E}^1_+$ в $\tilde{\mathbb{E}}^1$ и $\mathbb{E}^1_+$ в $\mathbb{E}^1_+$, наложим требования:

$\boxed{A_1.}$ Функция $R(p)$ непрерывна в любой точке p, в которой она конечна; если $\lim\limits_{k\to\infty} p_k = \tilde{p}$ и $R(\tilde{p}) = +\infty$, то $\lim\limits_{k\to\infty} R(p_k) = +\infty$; для любых $p_1 \in \operatorname{int}\mathbb{E}^c_-$ и $p_2 \notin \mathbb{E}^c_-$ имеет место неравенство

$$R(p_1) < R(p_2). \tag{4.3}$$

$\boxed{A_2.}$ Функция $R(p)$ удовлетворяет A_1 и либо $R(0) = +\infty$, либо $R(0) < R(p)$ для всех $p \notin \mathbb{E}^c_-$.

$\boxed{A_3.}$ Функция $P(t,\theta)$ непрерывна на $\mathbb{E}^1 \times \mathbb{E}^1_+$; для любых последовательностей $t_k \to t_*$, $\theta_k \to \theta_*$, где $\theta_k > 0$, $\theta_* \in \tilde{\mathbb{E}}^1_+$, равенство $\lim\limits_{k\to\infty} P(t_k, \theta_k) = 0$ возможно тогда и только тогда, когда или $t_* = 0$, или $t_* < 0$ и $\theta_* = 0$.

$\boxed{A_4.}$ Функция $Q(t)$ непрерывна на $\mathbb{E}^1_+$; для любой последовательности $t_k \to t_*$, где $t_k > 0$, $t_* \in \tilde{\mathbb{E}}^1_+$, равенство $\lim\limits_{k\to\infty} Q(t_k) = 0$ возможно тогда и только тогда, когда $t_* = 0$.

Для определенности будем считать, что $P(t,\theta) \geq 0$, $Q(t) \geq 0$ для всех $t \geq 0$, $\theta \geq 0$.

Приведем достаточные условия оптимальности точки x_*, выраженные посредством функций H и G.

Теорема 4.1. *Пусть функции* H *и* Q *строятся в виде* (4.1), (4.2), *причем функции* R, P *и* Q *удовлетворяют условиям* A_2–A_4. *Тогда если* $[x_*, y_*]$ *является особой точкой пары функций* $\{H, G\}$ *(быть может, в обобщенном смысле), то* $x_* \in X_*$.

Согласно условию A_1 функция $R(p)$ всюду на $\operatorname{int}\mathbb{E}^c_-$ принимает конечные значения. Кроме того, из неравенства (4.3) следует, что она должна быть постоянной всюду на границе ортанта $\mathbb{E}^c_-$. Приведем примеры функций, удовлетворяющих A_1 и A_2. Обозначим

$$\mu_{q,r}(p) \;=\; \left(\sum_{i=1}^{c} |p^i|^q\right)^{r/q}, \qquad -\infty < q < +\infty, \quad q \neq 0, \quad r > 0,$$

$$
\begin{aligned}
\mu_{+\infty,r}(p) &= \max_{1\le i\le c}\{|p^i|^r\},\\
\mu_{-\infty,r}(p) &= \min_{1\le i\le c}\{|p^i|^r\}.
\end{aligned}
$$

При этом считается, что $\mu_{q,r}(p)=0$, когда $q<0$ и $p^i=0$ хотя бы для одного индекса $i\in[1:c]$. Если $r=1$, то будем писать просто $\mu_q(p)$. При $q\ge 1$ функция $\mu_q(p)$ совпадает с гельдеровской нормой $\|p\|_q$ вектора p. Из всевозможных функций, удовлетворяющих A_1 и A_2, выделим следующие три класса.

К л а с с R_1. Для каждой функции из данного класса существует такое число $\gamma_1\in\mathbb{E}^1$, что $R(p)=\gamma^1$ для любых $p\in\mathbb{E}^c_-$ и $R(p)>\gamma_1$, если $p\notin\mathbb{E}^c_-$.

Очевидно, что все функции этого класса удовлетворяют A_2. Простейшими примерами служат

$$
\begin{aligned}
R(p) &= \mu_{q,r}(p_+), \qquad 0<q\le+\infty, \quad r>0,\\
R(p) &= \sum_{i=1}^{c}\lambda(p^i),
\end{aligned}
\tag{4.4}
$$

где $\lambda(t)$ — непрерывная функция скалярного аргумента такая, что $\lambda(t)=0$ при $t\le 0$ и $\lambda(t)>0$ при $t>0$. Для обеих функций $\gamma_1=0$ и обе они являются штрафными.

К л а с с R_2. Для каждой функции из этого класса существует такое число $\gamma_2\in\tilde{\mathbb{E}}^1$, что $R(p)<\gamma_2$ для любых $p\in\operatorname{int}\mathbb{E}^c_-$ и $R(p)=\gamma_2$, если $p\notin\operatorname{int}\mathbb{E}^c_-$. Примерами таких функций являются

$$
R(p) = -\mu_{q,r}(p_-), \qquad -\infty\le q<0, \quad r>0, \tag{4.5}
$$

$$
R(p) = \sum_{i=1}^{c}\nu(p^i), \tag{4.6}
$$

где $\nu(t)$ — непрерывная функция скалярного аргумента такая, что $\nu(t)<+\infty$ при $t<0$ и $\nu(t)\equiv+\infty$ при $t\ge 0$. Здесь в случае (4.5) $\gamma_2=0$, в случае (4.6) — $\gamma_2=+\infty$. Функция (4.5) удовлетворяет условию A_1, функция (4.6) — условию A_2.

К л а с с R_3. К этому классу будем относить те функции $R(p)$, которые не принадлежат ни R_1, ни R_2. Их можно получить, например, суммируя функции из R_1 и R_2. В частности, складывая (4.4) и (4.5), приходим к функции, удовлетворяющей A_2,

$$
R(p)=\mu_{q,r}(p_+)-\mu_{-q,r}(p_-), \qquad 0<q<+\infty, \quad r>0.
$$

В предельном случае, когда $q=+\infty$, $r=1$, она принимает вид

$$
R(p)=\max_{1\le i\le c}\{p^i\}. \tag{4.7}
$$

Из примеров функций $P(t,\theta)$ и $Q(t)$, удовлетворяющих, соответственно, A_3 и A_4, укажем

$$
P(t,\theta)=\begin{cases} t, & t\ge-\theta,\\ -\theta, & t<-\theta,\end{cases} \qquad P(t,\theta)=\begin{cases} t^2/2+t\theta, & t\ge-\theta,\\ -\theta^2/2, & t<-\theta,\end{cases}
$$

$$
Q(t)=t, \quad Q(t)=\sqrt{t}, \quad Q(t)=\ln(1+t).
$$

Отметим, что если при построении ξ используется функция R, принадлежащая классу R_1, то $R(h(x))\ge R(0)$ для всех x; вместо (4.2) тогда целесообразно взять

$$
G(x,y)=Q(R(h(x))-R(0))=S(x). \tag{4.8}
$$

Беря в качестве H и G в (2.5) или (2.6) функции (4.1), (4.2), (4.8), можно получить разные варианты метода штрафных функций. В случае, когда R является функцией типа R_1, эти методы будут методами внешней точки. В частности, полагая в (4.1) и (4.8) $R(p) = \mu_{2,2}(p)$, $Q(t) = t^{1/2}$, из (2.5) приходим к следующей схеме:

$$\begin{gathered} x_k \in \operatorname{Arg} \min_{x \in \mathbb{E}^n} \{ f(x) + y_k \|h_+(x)\|^2 \}, \\ y_{k+1} = y_k + \alpha \|h_+(x_k)\|. \end{gathered} \tag{4.9}$$

Параметр y в данном методе может лишь возрастать в процессе расчетов.

Методы внутренней точки реализуются с помощью функций R из класса R_2, у которых $R(0) = +\infty$. Так, например, подставляя в (4.1), (4.2)

$$R(p) = -\sum_{i=1}^{c} (p_-^i)^{-1}, \qquad Q(t) = t,$$

получим из (2.5)

$$x_k \in \operatorname{Arg} \min_{x \in X^0} \left\{ f(x) - y_k \sum_{i=1}^{c} \frac{1}{h^i(x)} \right\}, \qquad y_{k+1} = y_k(1-\alpha),$$

где $0 < \alpha < 1$. В отличие от метода (4.9) здесь параметр y стремится к нулю.

Использование в качестве R функций из класса R_3 приводит к объединенным методам внутренней и внешней точки. Текущие точки x_k при этом могут находиться как внутри, так и вне допустимой области. Более подробно различные варианты метода штрафных функций описаны в [1, 4, 5].

5. Методы параметризации целевой функции. На базе функций R, введенных в предыдущем пункте, построим другой класс методов решения задачи (1.1). Снова для простоты будем предполагать, что в (1.1) присутствуют только ограничения типа неравенства. Пусть $y \in \mathbb{E}^1$. Обозначим через $h^0(x,y)$ функцию $h^0(x,y) = f(x) - y$, через $\tilde{h}(x,y)$ — $(c+1)$-мерную вектор-функцию с компонентами $h^0(x,y), h^1(x), \ldots, h^c(x)$. Положим

$$H(x,y) = R(\tilde{h}(x,y)), \tag{5.1}$$

$$G(x,y) = \begin{cases} R(\tilde{h}(x,y)) - R(0), & \text{если} R(0) < +\infty, \\ h^0(x,y) & \text{в противном случае,} \end{cases} \tag{5.2}$$

где $R(p)$ — функция от $(c+1)$-мерного вектора, удовлетворяющая условию A_1. Имеют место следующие утверждения.

Теорема 5.1. *Пусть функция R удовлетворяет условию A_2 и $R(0) < +\infty$. Тогда если $[x_*, y_*]$ является особой точкой пары функций $\{H, G\}$, задаваемых* (5.1) *и* (5.2), *причем множество $X(y_*)$ состоит из единственной точки x_*, то $x_* \in X_*$.*

Теорема 5.2. *Пусть функция R удовлетворяет условию A_2 и $R(0) = +\infty$. Тогда если $[x_*, y_*]$ является обобщенной особой точкой пары функций $\{H, G\}$, задаваемых* (5.1) *и* (5.2), *то $x_* \in X_*$.*

Опираясь на приведенные теоремы, можно построить разнообразные непрямые методы решения задачи (1.1). Если функция R принадлежит классу R_1, то из (2.5) и (2.6) приходим к методам параметризации целевой функции (иногда их называют методами внешних центров). Параметр y в них играет роль нижней оценки оптимального значения целевой функции $f_* = \min\limits_{x \in X} f(x)$. В частности, беря в (2.5) в качестве R функцию $\mu_2(p)$ и полагая $\alpha = 1$, получаем хорошо известный метод Моррисона [6]

$$x_k \in \operatorname{Arg} \min_{x \in \mathbb{E}^n} \|\tilde{h}_+(x, y_k)\| = \operatorname{Arg} \min_{x \in \mathbb{E}^n} \|\tilde{h}_+(x, y_k)\|^2, \quad y_{k+1} = y_k + \|\tilde{h}_+(x_k, y_k)\|.$$

Вторая версия метода Моррисона, отличающаяся от приведенной выше лишь правилом изменения параметра y, следует из метода Ньютона (2.6). С учетом условия $\dfrac{\partial \|\tilde{h}_+(x_k, y_k)\|}{\partial x} = 0$, справедливого для всех x_k и y_k таких, что $\|\tilde{h}_+(x_k, y_k)\| > 0$, из (2.6) получаем

$$y_{k+1} = y_k + \frac{\|\tilde{h}_+(x_k, y_k)\|}{f(x_k) - y_k}.$$

Можно показать, что в случае, когда (1.1) является задачей выпуклого программирования, $y_0 < f_*$, то все последующие y_k не превосходят f_*.

Если обратиться к функциям типа R_2, то получим другие способы решения задачи (1.1), известные как методы внутренних центров. В них в отличие от предыдущих методов параметр y служит уже верхней оценкой величины f_*. В частности, полагая R равной $R(p) = -\sum\limits_{i=0}^{c} (p_-^i)^{-1}$, $\alpha = 1$, из (2.5) и (5.2) приходим к схеме

$$x_k \in \operatorname{Arg} \min_{x \in X_k^0} \left\{ \frac{1}{y_k - f(x)} - \sum_{i=1}^{c} \frac{1}{h^i(x)} \right\}, \qquad y_{k+1} = f(x_k),$$

где $X_k^0 = \{x \in X^0 : f(x) < y_k\}$. В данном методе все текущие точки x_k лежат внутри допустимой области, параметр y убывает в процессе расчетов, стремясь к f_*.

Функции типа R_3 дают возможность построить общие методы внутренних и внешних точек. Более детально все варианты методов параметризации целевой функции рассматриваются в [1, 5]. Там же обсуждается их связь с методами штрафных функций.

6. Метод возможных направлений. Возьмем в качестве H и G функции (5.1), (5.2), в качестве R — функцию (4.7); получим

$$H(x, y) = \max_{0 \le i \le c} \{h^i(x, y)\}, \qquad G(x, y) = H(x, y), \tag{6.1}$$

где для единообразия функции $h^i(x)$ обозначены как $h^i(x, y)$, $i \in [1 : c]$. Согласно утверждению теоремы 5.1, если $[x_*, y_*]$ является особой точкой пары (6.1), то $x_* \in X_*$. Функции H и G непрерывны, но не дифференцируемы по x даже при гладких f и h. Применим для отыскания особых точек процесс (2.9), причем поскольку $f(x) \in Y_1(x)$ для любого $x \in \mathbb{E}^n$, то на каждом шаге будем полагать $y_k = f(x_k)$. Известно много способов построения векторов $s(x_k, y_k)$, которым соответствуют разные варианты метода возможных направлений. Если воспользоваться направлением ε-наискорейшего спуска, то приходим к наиболее популярному варианту метода возможных направлений, предложенному Зойтендейком [7] и обобщенному в [8]. В нем для отыскания направления $s(x_k, y_k)$ решается вспомогательная задача

$$\min_{s, \sigma} \sigma, \qquad \langle h_x^i(x_k, y_k), s \rangle \le \sigma, \qquad i \in I_\varepsilon(x_k, y_k), \quad \|s\| \le 1. \tag{6.2}$$

Здесь $\varepsilon > 0$, $I_\varepsilon(x, y) = \{0 \le i \le c : h^i(x, y) \ge H(x, y) - \varepsilon\}$, $\|\cdot\|$ — произвольная норма в $\mathbb{E}^n$ (для того, чтобы (6.2) стала задачей линейного программирования, обычно берут $\|\cdot\|_1$

или $\|\cdot\|_\infty$). Шаг спуска α_k выбирается из приближенного решения задачи минимизации функции $H(x, y_k)$ вдоль направления $s(x_k, y_k)$.

7. Методы модифицированных функций Лагранжа. Положим в (2.10)

$$\xi(x,y) = \sum_{i=1}^{e} \varphi(g^i(x), u^i) + \sum_{j=1}^{c} \psi(h^j(x), v^j), \qquad y = [u, v] \in \mathbb{E}^m. \tag{7.1}$$

Предполагая, что функции φ, ψ, f, g, h дифференцируемы, запишем необходимое условие минимума в задаче (2.1) в виде

$$H_x(x,u,v) = f_x(x) + \sum_{i=1}^{e} \varphi_g(g^i(x), u^i) g_x^i(x) + \sum_{j=1}^{c} \psi_h(h^j(x), v^j) h_x^j(x) = 0.$$

Сравним это условие с (1.2). Оба условия совпадут, если

$$\varphi_g(g^i(x), u^i) = u^i, \quad i \in [1:e], \tag{7.2}$$

$$\psi_h(h^j(x), v^j) = v^j, \quad j \in [1:c]. \tag{7.3}$$

Наложим на функции φ и ψ следующее требование.

$\boxed{A_5.}$ Функции φ и ψ таковы, что система (7.2), (7.3) разрешима тогда и только тогда, когда $x \in X$.

Объединим вектор-функции из (7.2) и (7.3) единым символом, положив

$$G(x,y) = [\varphi_g(g^1(x), u^1) - u^1, \ldots, \varphi_g(g^e(x), u^e) - u^e, \psi_h(h^1(x), v^1) - v^1, \ldots, \psi_h(h^c(x), v^c) - v^c]. \tag{7.4}$$

Тогда система (7.2), (7.3) примет вид $G(x,y) = 0$. Предполагая, что при каждом $y \in \mathbb{E}^m$ из (2.1) можно определить $x \in X(y)$, поставим задачу решения системы (2.4). Теорему 2.1 можно в данном случае переформулировать следующим образом.

Теорема 7.1. *Пусть функции H, G определены формулами* (2.10), (7.1), (7.4), *выполнено условие A_5. Пусть существует особая точка $[x_*, y_*]$ пары $\{H, G\}$, функции φ и ψ таковы, что для любых $x \in X$ имеет место* (2.11). *Тогда $x_* \in X_*$.*

Если в дополнение к условиям теоремы 7.1 потребовать, чтобы в каждой точке $x_* \in X_*$ векторы-градиенты всех ограничений были линейно независимы, то всякая особая точка будет точкой Куна–Таккера.

Для решения системы (2.4) можно применить методы (2.5), (2.6). Теорема 7.1 гарантирует в случае сходимости методов к точке x_* выполнение условия $x_* \in X_*$. Для того, чтобы гарантировать сходимость методов (2.5), (2.6), нужно обеспечить выполнение некоторых достаточных условий, которые приводят к дополнительным ограничениям на выбор функций φ и ψ. Некоторые такие условия даны в [1]. Если взять

$$\varphi(g^i, u^i) = g^i u^u + \frac{\tau}{2}(g^i)^2, \qquad \psi(h^j, v^j) = \frac{1}{2\tau}[(\tau h^j + v^j)_+]^2, \tag{7.5}$$

где $\tau > 0$, то метод простой итерации (2.5) при $\alpha = 1$ приводит к формулам

$$u_{k+1} = u_k^i + \tau g^i(x_k), \qquad v_{k+1} = (v_k^j + \tau h^j(x_k))_+, \qquad i \in [1:e], \quad j \in [1:c].$$

Это наиболее популярный вариант метода, исследованный многими авторами (см., например, [9, 10]). Его недостатком является то, что функция H дифференцируема по x только один раз.

Проверку условий теоремы 7.1 проводить проще, если в дополнение к A_5 вводить предположение о том, что система (7.2), (7.3) разрешима тогда и только тогда, когда $v^j \geq 0$ и $v^j h^j = 0$, $j \in [1:c]$. Более того, в этом случае можно отказаться от выполнения условия (2.11), однако тогда можно лишь утверждать, что $x_* \in X_*^1$. Примером таких функций является

$$\psi(h^j, v^j) = \frac{1}{4\tau}[(v^j + \tau h^j(x))_+^4 - 4\tau h^j (v^j)^3] + v^j \cdot \begin{cases} h^j, & h^j \geq 0, \\ \mathrm{arctg} h^j, & h^j < 0. \end{cases} \tag{7.6}$$

Для нее итерации по v имеют вид

$$v_{k+1}^j = (v_k^j + \tau h^j(x_k))_+^3 - (v_k^j)^3 + v_k^j[1 + (\tau h_-^j(x_k))^2]^{-1}, \qquad j \in [1:c].$$

По сравнению с (7.5) функция (7.6) обладает более высокой гладкостью, что позволяет для решения (2.1) использовать быстросходящиеся методы (например, метод Ньютона).

8. Общая схема прямых методов. Видоизменим описанную в п. 2 последовательность действий. При фиксированном $x \in \mathbb{E}^n$ решим вспомогательную задачу максимизации

$$\max_{y \in M} H(x, y). \tag{8.1}$$

Множество всех ее решений определяет отображение

$$Y(x) = \mathrm{Arg} \max_{y \in M} H(x, y).$$

В данном пункте и ниже особой точкой пары функций $\{H, G\}$ будем называть точку $[x_*, y_*]$, удовлетворяющую условиям

$$y_* \in Y(x_*), \qquad G(x_*, y_*) = 0.$$

Как и прежде, множество всех особых точек $\{H, G\}$ будем обозначать $Z_*(H, G)$. От функций H и G желательно потребовать, чтобы всякая их особая точка $[x_*, y_*] \in Z_*(H, G)$ обладала свойством: $x_* \in X_*$. Построить такие функции в случае общей задачи нелинейного программирования авторам не удалось. Поэтому рассмотренный подход реализуем в несколько упрощенном варианте, а именно: подберем функции H и G таким образом, чтобы все их особые точки $[x_*, y_*]$ были таковы, что $x_* \in X_*^1$. Тогда нахождение точек $x_* \in X_*^1$ сводится к решению системы

$$G(x, y(x)) = 0, \tag{8.2}$$

где $y(x)$ — зависимость, полученная из решения вспомогательной задачи (8.1). Для решения (8.2) можно применять общеизвестные методы. Так, например, метод простой итерации записывается в виде

$$x_{k+1} = x_k - \alpha G(x_k, y_k), \qquad y_k \in Y(x_k). \tag{8.3}$$

В тех случаях, когда множество M имеет достаточно простую структуру, аналогично (2.8) можно строить методы, предназначенные для отыскания точек, удовлетворяющих необходимым условиям экстремума.

Объединим двойственные векторы одним символом $y = [u, v] \in \mathbb{E}^m$. В качестве M возьмем множество $M = \{y \in \mathbb{E}^m : y^i \geq 0,\ i = e+1, \ldots, m\}$. Положим

$$H(x, y) = L(x, y) - \eta(x, y), \tag{8.4}$$

где $\eta(x,y)$ — функция, непрерывно дифференцируемая по y.

Теорема 8.1. *Пусть функции $G(x,y)$ и $\eta(x,y)$ таковы, что во всякой особой точке $z_* = [x_*, y_*]$ пары $\{H, G\}$ выполнены условия*

$$L_x(x_*, y_*) = 0, \qquad \eta_y(x_*, y_*) = 0.$$

Тогда z_* есть точка Куна–Таккера.

Про пару функций $\{H, G\}$ будем говорить, что она согласована с задачей (1.1), если всякая ее особая точка $[x_*, y_*]$ такова, что $x_* \in X_*^1$.

9. Прямые методы модифицированной функции Лагранжа. Введем в рассмотрение непрерывно дифференцируемую функцию $\rho(t)$, отображающую $\mathbb{E}^1$ в $\mathbb{E}^1$, на которую наложим требование

$\boxed{A_6.}$ Функция ρ такова, что $\rho_t(t) = 0$ тогда и только тогда, когда $t = 0$.

В качестве примеров функции $\rho(t)$ укажем

$$\rho_1(t) = t^2/2, \qquad \rho_2(t) = e^t - t, \qquad \rho_3(t) = \operatorname{ch} t.$$

Чтобы удовлетворить условиям теоремы 8.1, функции $\eta(x,y)$ и $G(x,y)$ будем строить в виде

$$\eta(x,y) = \tau \sum_{j=1}^{n} \rho\left(\frac{\partial L(x,y)}{\partial x^j}\right), \qquad G(x,y) = \tau \rho_t(L_x(x,y)). \tag{9.1}$$

Здесь $\rho_t(L_x)$ — вектор с компонентами $\rho_t\left(\dfrac{\partial L(x,y)}{\partial x^j}\right)$, $j = 1, 2, \ldots, n$; $\tau > 0$.

Из теоремы 8.1 следует

Теорема 9.1. *Пусть функции H, G определены формулами* (8.4), (9.1), *выполнено условие A_6. Тогда если $z_* = [x_*, y_*]$ — особая точка пары $\{H, G\}$, то z_* есть точка Куна–Таккера.*

Подставляя функции (8.4), (9.1) в (8.3) и полагая $\alpha = 1$, приходим к процессу

$$x_{k+1} = x_k - \tau \rho_t(L_x(x_k, y_k)), \qquad y_k \in Y(x_k). \tag{9.2}$$

Можно показать (см. [11]), что если $\rho(t)$ — дважды непрерывно дифференцируемая строго выпуклая функция, то в окрестности любой точки $x_* \in X_*$, в которой выполнены стандартные достаточные условия минимума задачи НЛП и условие регулярности ограничений, решение задачи (8.1) всегда существует, метод (9.2) локально сходится к x_* для достаточно малых τ.

Если все функции, определяющие задачу (1.1), трижды непрерывно дифференцируемы, функция $\rho(t)$ дважды непрерывно дифференцируема, то для решения уравнения (8.2) может быть применен метод Ньютона.

10. Метод приведенного градиента. Покажем, что из (9.2) следуют многие известные численные методы. Опустим в задаче (1.1) ограничения типа неравенства, в качестве ρ возьмем функцию ρ_1; тогда задача (8.1) легко решается:

$$y(x) = [g_x^\top(x) g_x(x)]^{-1} \left[\frac{1}{\tau} g(x) - g_x^\top(x) f_x(x)\right].$$

Здесь g_x — матрица размером $n \times m$, символ "$\top$" означает транспонирование. Непрерывный вариант метода (9.2) имеет вид

$$\frac{dx}{dt} = -[f_x(x) + g_x(x)y(x)], \qquad x(0) = x_0. \tag{10.1}$$

Решение этой системы обозначим через $x(x_0, t)$. Продифференцируем f и g в силу системы (10.1); получим

$$\frac{df}{dt} = -\|f_x(x) + g_x(x)y(x)\|^2 + \frac{1}{\tau}\langle y(x), g(x)\rangle, \tag{10.2}$$

$$\frac{dg}{dt} = -\frac{1}{\tau}g. \tag{10.3}$$

В том частном случае, когда x_0 таково, что $g(x_0) = 0$, из (10.2), (10.3),) получаем для всех $t \geq 0$

$$\frac{d}{dt}f(x(x_0, t)) \leq 0, \qquad g(x(x_0, t)) \equiv 0,$$

т.е. допустимое множество X инвариантно относительно системы (10.1), вдоль траекторий минимизируемая функция монотонно убывает. Этот метод был предложен рядом авторов, сошлемся лишь на [12]. Метод (10.1) является его обобщением на тот случай, когда $x_0 \notin X$. Из (10.3) следует, что

$$g(x(x_0, t)) = g(x_0)e^{-t/\tau} \to 0$$

при $t \to +\infty$, т.е. траектории системы (10.3) стремятся к допустимому множеству, если $t \to +\infty$. Функция $f(x(x_0, t))$ вне допустимого множества может изменяться немонотонно, но вблизи X она монотонно уменьшается.

Аналогичным образом можно изучить случай задачи с ограничениями типа неравенства. Однако расчетные формулы при этом будут более громоздкими, поэтому мы их не приводим. Заметим только, что если $x_0 \in X^0$, то обобщение метода (10.1) дано в [1, 13].

11. Методы линеаризации. Предположим, что функция $\rho(t)$, введенная в п. 9, строго выпукла по t. Тогда в силу условия A_6 она достигает своего минимума на $\mathbb{E}^1$ в нуле. Пусть $\rho^*(t)$ обозначает функцию, сопряженную к $\rho(t)$. Составим вспомогательную задачу выпуклого программирования, линеаризовав в (1.1) целевую функцию и ограничения

$$\begin{gathered}
\min_s \langle f_x(x), s\rangle + \tau \sum_{j=1}^{n} \rho^*(-s^j/\tau), \\
g^i(x) + \langle g^i_x(x), s\rangle = 0, \qquad i \in [1:e], \\
h^j(x) + \langle h^j_x(x), s\rangle \leq 0, \qquad j \in [1:c].
\end{gathered} \tag{11.1}$$

Добавление второго члена к линеаризованной целевой функции позволяет избежать бесконечно больших решений.

Из сравнения необходимых и достаточных условий экстремума в задачах (8.1) и (11.1) следует, что вектор $-\tau\rho_t(L_x(x, y(x))$ является решением задачи (11.1). Таким образом, если рассмотреть процесс вида

$$x_{k+1} = x_k + s_k, \tag{11.2}$$

где s_k есть решение задачи (11.1) при $x = x_k$, то для любого начального приближения $x_0 \in \mathbb{E}^n$ последовательность $\{x_k\}$, вырабатываемая (11.2), совпадет с последовательностью, вырабатываемой процессом (9.2). Указанная связь позволяет интерпретировать многие методы линеаризации как прямые методы типа (9.2). В частности, если в (11.1) в качестве ρ

взять функцию ρ_1, то приходим к вспомогательной задаче, близкой к задаче, используемой в методе линеаризации [14].

12. Прямые-двойственные методы. В методах этой группы совершаются одновременные итерации по исходным и вспомогательным переменным. В простейшем случае отыскиваются решения системы 2.7). Метод простой итерации и метод Ньютона приводят к следующим расчетным схемам:

$$x_{k+1} = x_k - \alpha H_x, \quad y_{k+1} = y_k + \alpha G, \tag{12.1}$$

$$\begin{aligned} x_{k+1} &= x_k - H_{xx}^{-1}(H_x + H_{xy}\delta y_k), \\ y_{k+1} - y_k &= \delta y_k = (G_x H_{xx}^{-1} H_{xy} - G_y)^{-1}(G - G_x H_{xx}^{-1} H_x). \end{aligned} \tag{12.2}$$

Здесь функции H, G и их производные вычислены в точке $[x_k, y_k]$..

В [1] доказана локальная сходимость этих методов в предположениях п. 7. Если в (1.1) отсутствуют ограничения типа неравенства, $H = L$, $G = g$, то (12.1) переходит в метод Эрроу–Гурвица [12], в этом же случае метод (12.2) исследовался в [15].

Заключение. Изложенная трактовка численных методов решения задач НЛП как решение систем уравнений не является единственной. Первоначально авторы пытались за основу взять методы отыскания седловых точек, решения минимаксных и максиминных задач. Однако при таком подходе удалось охватить с единых позиций меньшее число методов.

Всюду выше, проводя редукцию (1.1) к решению систем уравнений, мы использовали метод простой итерации и метод Ньютона. Ясно, что для решения систем можно применять многие другие методы, например, квазиньютоновские, метод секущих, аналог метода обобщенного градиента. Используя вместо (2.5) последний из упомянутых методов, получим

$$y_{k+1} = y_k + \alpha_k \omega_k G(x_k, y_k), \qquad x_k \in X(y_k).$$

Здесь α_k, ω_k — некоторые положительные коэффициенты, изменяющиеся по законам

$$\omega_k = (1 + \|G(x_k, y_k)\|)^{-1}, \qquad \sum_{k=1}^{\infty} \alpha_k = \infty, \qquad \sum_{k=1}^{\infty} \alpha_k^2 < \infty.$$

С помощью теоремы 2.5.2 из [1] легко сформулировать условия сходимости метода. Аналогично тривиальным образом можно выписать другие вычислительные схемы подобного типа и дать их обоснование.

ЛИТЕРАТУРА

1. *Евтушенко Ю.Г.* Методы решения экстремальных задач и их применение в системах оптимизации.– М.: Наука, 1982.

2. *Еремин И.И.* Метод “штрафов” в выпуклом программировании.– ДАН СССР, 1967, 173, No. 4.

3. *Zangwill W.* Non-linear programming via penalty function.– Management Sci., 1967, 13, No.5.

4. *Фиакко А., Мак-Кормик Г.* Нелинейное программирование. Методы последовательной безусловной минимизации.– М.: Мир, 1972.

5. *Гроссман К., Каплан А.А.* Нелинейное программирование на основе безусловной минимизации.– Новосибирск: Наука, 1981.

6. *Morrison D.* Optimization by least squars.– SIAM J. Numerical Analysis, 1968, 5, N1.

7. *Зойтендейк Г.* Методы возможных направлений.– М.: Изд-во иностр. лит., 1963.

8. *Polak E., Trahan R., Mayne D.Q.* Combined phase I — phase II methods of feasible directions.– Mathematical programming, 1979, No.17

9. *Гольштейн Е.Г., Третьяков Н.В.* Модифицированные функции Лагранжа.– Экономика и математические методы, 1974, 10, j 3.

10. *Поляк Б.Т., Третьяков Н.В.* Метод штрафных оценок для задач на условный экстремум.– ЖВМ и МФ, 1973, 13, j 1.

11. *Голиков А.И., Жадан В.Г.* Итеративные методы решения задач нелинейного программирования с использованием модифицированных функций Лагранжа.– ЖВМ и МФ, 1980, 20, No.4.

12. *Эрроу К.Дж., Гурвиц Л., Удзава Х.* Исследования по линейному и нелинейному программированию.– М.: Изд-во иностр. лит., 1962.

13. *Евтушенко Ю.Г., Жадан В.Г.* Релаксационный метод решения задач нелинейного программирования.– ЖВМ и МФ, 1977, 17, No.4.

14. *Пшеничный Б.Н., Данилин Ю.М.* Численные методы в экстремальных задачах.– М.: Наука, 1975.

15. *Поляк Б.Т.* Итерационные методы, использующие множители Лагранжа для решения экстремальных задач с ограничениями типа равенства.– ЖВМ и МФ, 1970, 10, No.5.