Доклады Академии наук СССР 1978. Том 239, No.3, стр. 519–522

УДК 518.12 МАТЕМАТИКА

А.И. Голиков, Ю.Г. ЕВТУШЕНКО

ОБ ОДНОМ КЛАССЕ МЕТОДОВ РЕШЕНИЯ ЗАДАЧ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

(Представлено академиком А.А. Дородницыным 2 XII 1977)
Пересмотрено 18 II 2004

1. Рассмотрим задачу отыскания

$$\min_{x\in X} f(x), \qquad X = \{x \in E_n : g(x) = 0\}; \tag{1}$$

здесь $x \in E_n$, E_i есть i-мерное евклидово пространство; дифференцируемые функции $f(x)$, $g(x)$ осуществляют соответственно отображения $f : E_n \to E_1$, $g : E_n \to E_m$.

Всюду ниже предполагаем, что существуют вектор $p_* \in E_m$ и решение задачи (1) x_* такие, что (x_*, p_*) образуют точку Куна–Таккера, т.е.

$$f_x(x_*) + g_x(x_*)p_* = 0, \qquad x_* \in X. \tag{2}$$

Введем функцию двух скалярных аргументов $\varphi(g^i, p^i)$ и вектор $\varphi(g, p) = [\varphi(g^1, p^1), \ldots, \ldots, \varphi(g^m, p^m)]$. Символы φ_g, φ_p, φ_{gg}, φ_{pp}, φ_{gp} обозначают матрицы первых и вторых производных размерами соответственно $m \times 1$, $m \times 1$, m^2, m^2, m^2. Матрицы вторых производных диагональные. Конкретный элемент матриц получается, если указываются его скалярные аргументы, например, $\varphi_g(g^i, p^i)$. Составим обобщенную функцию Лагранжа

$$H(x, p) = f(x) + \sum_{i=1}^{m} \varphi(g^i(x), p^i). \tag{3}$$

На функцию φ наложим следующее требование:

$\mathbf{A}_1$. Функция $\varphi(g^i, p^i)$ непрерывно дифференцируема, для любых действительных p^i имеет место $\varphi_g(0, p^i) = p^i$, и если $g^i \neq 0$, то $\varphi_g(g^i, p^i) \neq p^i$.

Считая, что задача безусловной минимизации

$$H(x(p), p) = \min_{x\in E_n} H(x, p) \tag{4}$$

имеет решение $x = x(p)$, перейдем к задаче об отыскании корней системы

$$\Phi(p) = \varphi_g(g(x(p)), p) - p = 0. \tag{5}$$

Если p_* — решение (5), $x_* = x(p_*)$, выполнено условие $\mathbf{A}_1$, тогда (x_*, p_*) есть точка Куна–Таккера. Это обстоятельство позволяет использовать широкий класс методов отыскания корней систем уравнений для решения (5) и тем самым решать задачу (1). Предполагая, что $\Phi(p)$ дифференцируема, получим формулу для ее производной:

$$\Phi_p(p) = [I - \varphi_{gg}(g(x(p)), p)M(x(p), p)]\varphi_{gp}(g(x(p)), p) - I;$$

здесь

$$\begin{aligned} M(x,p) &= g_x^\top(x) H_{xx}^{-1}(x,p) g_x(x), \\ H_{xx}(x,p) &= L(x,p) + g_x(x)\varphi_{gg}(g(x),p) g_x^\top(x), \\ L(x,p) &= f_{xx}(x) + \sum_{i=1}^m g_{xx}^i(x)\varphi_g(g^i(x),p^i). \end{aligned}$$

I — единичная матрица m^2, индекс $\top$ означает транспонирование матриц. Обозначим $x_s = x(p_s)$. Приведем три метода решения (1):

$$p_{s+1} = \varphi_g(g(x_s),p_s); \tag{6}$$

$$p_{s+1} = p_s - \Phi_p^{-1}(p_s)[\varphi_g(g(x_s),p_s) - p_s]; \tag{7}$$

$$p_{s+1} = \varphi_g(g(x_s),p_s) + Q(x_s,p_s)g(x_s), \quad Q(x,p) = M^{-1}(x,p) - \varphi_{gg}(g(x),p); \tag{8}$$

здесь (6) — по существу, метод простой итерации, (7) — аналог метода Ньютона.

Укажем примеры простейших функций, удовлетворяющих $\mathbf{A}_1$:

$$\begin{aligned} \varphi^1(g^i,p^i) &= g^i p^i + (g^i)^2/2, \qquad \varphi^2(g^i,p^i) = g^i(p^i - 1) + \exp(g^i), \\ \varphi^3(g^i,p^i) &= g^i p^i + \frac{1}{2\pi}\left[2g^i \operatorname{arctg} g^i - \ln[1+(g^i)^2]\right] \exp(-(p^i)^2). \end{aligned}$$

При численных расчетах целесообразно брать в качестве $\varphi(g^i,p^i)$ функции $\varphi^k(\tau g^i,p^i)/\tau$, где $k = 1,2,3$, τ — положительный параметр. Взяв, например, первую из них, получим, что (7), (8) совпадают, (6) и (7) приводят к схемам

$$p_{s+1} = p_s + \tau g(x_s), \qquad p_{s+1} = p_s + M^{-1}(x_s,p_s)g(x_s).$$

Первый метод хорошо известен (см., например, [1]).

Пусть существует вектор p_*, являющийся решением (5). Введем уравнение

$$|\Phi_p(p_*) + I - \lambda I| = 0. \tag{9}$$

Сходимость (6) и (7) следует из общеизвестных теорем о сходимости. Переформулируем их применительно к (5). Через G обозначим некоторую окрестность p_*.

Теорема 1. *Пусть существуют непрерывно дифференцируемые функции $x(p)$ и $\Phi(p)$, все λ по модулю меньше единицы; тогда (6) локально сходится к p_*.*

Теорема 2. *Пусть $H(x,p)$ дважды непрерывно дифференцируема в окрестности точки Куна–Таккера (x_*,p_*), матрица $H_{xx}(x_*,p_*)$ невырожденна, на G существует непрерывное решение $x(p)$ задачи (4). Если $\Phi_p(p_*)$ невырожденна и на G матрица $\Phi_p(p)$ удовлетворяет условию Липшица, то (7) локально сходится к p_* с квадратичной скоростью. Если $M(x_*,p_*)$ невырожденна, матрица $Q(x(p),p)$ удовлетворяет на G условию Липшица, то (8) локально сходится к p_* с квадратичной скоростью.*

В [2] дана другая модификация метода Ньютона, пригодная для решения (1) и (10), не требующая вспомогательной минимизации H по x. Возможны различные модификации предлагаемых методов, например, аналогом метода Зейделя является процесс

$$p_s^i = \varphi_g(g^i(x(\bar p_{si})), p_{s-1}^i), \quad \bar p_{si} = [p_s^1,\ldots,p_s^{i-1},p_{s-1}^i,p_{s-1}^{i+1},\ldots,p_{s-1}^m], \qquad i \in [1,\ldots,m].$$

$\boxed{\mathbf{2.}}$ Метод (6) можно использовать в случае задач с ограничениями типа неравенств:

$$\min_{x \in X} f(x), \qquad X = \{x \in E_n : g(x) \le 0\}. \tag{10}$$

Предполагаем, что здесь также существует точка Куна–Таккера (x_*, p_*), т.е. в ней имеет место (2) и $p_*^i \ge 0$, $p_*^i g^i(x_*) = 0$ для $i \in [1, \dots, m]$. Функцию Лагранжа строим в виде (3). Через $P(g^i)$ обозначим множество неотрицательных решений уравнения $p^i = \varphi_g(g^i, p^i)$.

На функцию φ вместо $\mathbf{A}_1$ наложим следующее условие:

$\mathbf{A}_2$. $P(g^i) = \emptyset$ при $g^i > 0$; $P(g^i) = 0$ при $g^i < 0$; если $a \ge 0$, то $a \in P(0)$; $\varphi_g(g^i, p^i) \ge 0$ для любых $g^i, p^i \ge 0$.

Для каждого p из (4) находим $x = x(p)$, определив таким образом систему $p = \varphi_g(g(x(p)), p)$. Если p_* — ее решение, $x_* = x(p_*)$, $p_* \ge 0$, то согласно $\mathbf{A}_2$ (x_*, p_*) образует точку Куна–Таккера. Используем для решения (10) схему (6), взяв $p_0 \ge 0$, тогда все $p_s \ge 0$, предельные точки последовательности $(x(p_s), p_s)$ образуют точку Куна–Таккера.

Условию $\mathbf{A}_2$ удовлетворяют, например, функции

$$\begin{aligned}
\varphi^4(g^i, p^i) &= \psi(g^i_+) + p^i e^{g^i}, \\
\varphi^5(g^i, p^i) &= \psi(g^i_+) + p^i \begin{cases} 1 + h g^i + \dfrac{h(h+1)}{2!}(g^i)^2 + \dfrac{h(h+1)(h+2)}{3!}(g^i)^3, & \text{if } g^i \ge 0, \\ \dfrac{1}{(1-g^i)^h}, & \text{if } g^i \le 0; \end{cases}
\end{aligned}$$

здесь $0 < h$, $g^i_+ = \max[0, g^i]$, $\psi(z)$ — некоторая достаточно гладкая функция такая, что $\psi(0) = \psi'(0) = 0$, при $z > 0$ $\psi(z) > 0$, $\psi'(z) > 0$, $\psi''(z) > 0$ (например, $\psi(z) = z^4$).

$\boxed{\mathbf{3.}}$ Количество целых чисел, принадлежащих индексному множеству $B = \{j : g^j(x_*) = 0,\ 1 \le j \le m\}$, обозначим через q. Введем матрицы $\bar{\varphi}_{gg}$, $\bar{\varphi}_{pg}$, $\bar{g}_x$, которые совпадают соответственно с φ_{gg}, φ_{pg}, g_x в задаче (1). В случае (10) в формулах для $\bar{\varphi}_{gg}$, $\bar{\varphi}_{pg}$, $\bar{g}_x$ сохранены только те g^j и их производные, у которых $j \in B$, т.е. для задачи (10) размеры $\bar{\varphi}_{gg}$, $\bar{\varphi}_{pg}$, $\bar{g}_x$ суть соответственно q^2, q^2, $n \times q$.

Приведем два дополнительных условия:

$\mathbf{A}_3$. Функция φ такова, что $\bar{\varphi}_{gg}(g(x_*), p_*)$ положительно определена, матрица $\bar{\varphi}_{pg}(g(x_*), p_*)$ — единичная, в случае (10) для всех $j \notin B$ выполнены условия $\bar{\varphi}_{gg}(g^j(x_*), p_*^i) = 0$, $0 < \varphi_{pg}(g^j(x_*), p_*^j) < 1$.

$\mathbf{A}_4$. Функция $H(x, p)$ дважды непрерывно дифференцируема в окрестности точки Куна–Таккера (x_*, p_*), столбцы матрицы $\bar{g}_x(x_*)$ линейно независимы, $x^\top L(x_*, p_*) x > 0$ для любого ненулевого x такого, что $x^\top \bar{g}_x(x_*) = 0$.

Пусть μ обозначает совокупность корней уравнения

$$|\bar{\varphi}_{gg}(g(x_*), p_*) \bar{g}_x^\top(x_*) L^{-1}(x_*, p_*) \bar{g}_x(x_*) - \mu I| = 0.$$

Здесь в случае задачи (10) единичная матрица I имеет размер q^2. При выполнении $\mathbf{A}_3$ корни приведенного уравнения вещественны, и если они не все положительны, то α обозначает наибольший отрицательный корень.

Теорема 3. *Пусть существует точка Куна–Таккера, в которой матрица L невырожденна, имеют место $\mathbf{A}_3$, $\mathbf{A}_4$, в случае задачи* (1) *выполнено условие $\mathbf{A}_1$, в случае* (10)

— $\mathbf{A}_2$; *в качестве* $\varphi(g,p)$ *взята функция* $\varphi(\tau g,p)/\tau$. *Тогда для любого* $\tau > \bar{\tau}$ *выполнены условия теоремы* 1, *причем если все* $\mu > 0$, *то* $\bar{\tau} = 0$, *и* $\bar{\tau} = -2/\alpha$ *в противном случае.*

4. В случае, если (1) и (10) — задачи выпуклого программирования, вместо (4) можно ввести иную вспомогательную задачу

$$\min_{x \in E_n} \Gamma(g(x), p, \mu, f(x)), \tag{11}$$

где

$$\Gamma(g(x), p, \mu, f(x)) = \gamma(f(x) - \mu) + \sum_{i=1}^{m} \varphi(g^i(x), p^i)$$

выпукла по x, $\gamma(q)$ — непрерывно дифференцируемая функция скалярного аргумента, удовлетворяющая условию

$\mathbf{A}_5$. Для всех $q \neq 0$, $\gamma(q) > 0$, $\gamma'(q) \neq 0$ и $\gamma(0) = \gamma'(0) = 0$.

Пусть на s-м шаге итеративного процесса известны p_s, μ_s, из (11) найдено $x_s = x(p_s, \mu_s)$. Построим метод, в котором

$$p_{s+1} = \varphi_g(g(x_s), p_s) \frac{\gamma'(\bar{f} - \bar{\mu})}{\gamma'(f(x_s) - \mu_s)}; \tag{12}$$

здесь в качестве $\bar{f}$ и $\bar{\mu}$ можно взять любые числа из интервалов

$$\mu_s \leq \bar{\mu} < \bar{f} \leq f(x_s) + \sum_{i=1}^{m} p_s^i g^i(x_s) = F_s.$$

Укажем несколько простейших вариантов выбора $\bar{f}$, $\bar{\mu}$, μ_{s+1}:

$$\begin{array}{lll} \bar{f} = F_s, & \bar{\mu} = \mu_s, & \mu_{s+1} = \mu_s; \\ \bar{f} = F_s, & \bar{\mu} = f(x_s), & \mu_{s+1} = f(x_s); \\ \bar{f} = f(x_s), & \bar{\mu} = \mu_s, & \mu_{s+1} = \mu_s. \end{array}$$

Для приведенных алгоритмов важно, чтобы выполнялось условие $\Gamma(g(x_*), p_s, \mu_s, f(x)) >$ > 0. Если начальные значения μ_0, p_0 выбраны удовлетворяющими этим условиям, то в случае задач выпуклого программирования ($f(x)$ выпуклая, в (1) $g(x)$ линейная, в (10) выпуклая) это свойство автоматически сохраняется и на последующих итерациях. Начальные значения p_0, μ_0 можно найти, используя, например, метод внешних штрафных функций. При этом должно быть $\mu_0 < f(x_*)$. Это условие будет выполнено, если положить $\mu_0 \in [f(x_0), F_0]$, где x_0 — точка минимума внешней штрафной функции.

Если для решения (1) в (11) положить

$$\Gamma(g(x), p, \mu, f(x)) = (f(x) - \mu)^2 + \sum_{i=1}^{m} \varphi^1(\tau g^i(x), p^i)/\tau, \tag{13}$$

то метод (12) переходит в следующий:

$$p_{s+1} = (p_s + \tau g(x_s)) \frac{\bar{f} - \bar{\mu}}{f(x_s) - \mu_s}. \tag{14}$$

В формулах (13), (14) можно положить $p = 0$, $p_s \equiv 0$ соответственно; значение μ_{s+1} брать любым из отрезка $[f(x_s), F_s]$. В частности, если $\mu_{s+1} = f(x_s)$, то приходим к методу

[3], если $\mu_{s+1} = \mu_s + (\Gamma(g(x_s), 0, \mu_s, f(x_s)))^{1/2} = R_s$, то получаем метод [4]. Сходимость будет быстрее, если брать $\mu_{s+1} = F_s$, так как $F_s \geq R_s$. В случае задачи линейного программирования последний алгоритм (в котором $\mu_{s+1} = F_s$) сходится за конечное число шагов.

5. Опыт численного решения задач методами (6), типа Зейделя, алгоритмами из п. 4 свидетельствует о их достаточно высокой эффективности. Алгоритмы п. 4, использующие на каждой итерации значения обычной функции Лагранжа (например, $\bar{f} = F_s$, либо $\mu_{s+1} = R_s$), требуют большей точности решения вспомогательной задачи (11), чем при решении задачи безусловной минимизации (4) в методах п.п. 1,2. Можно показать, что на каждом s-м шаге решение (4) можно прекращать как только найдена точка x_s такая, что $\|H_x(x_s, p_s)\| \leq e_s$, где $e_s \to 0$ при $s \to \infty$.

ЛИТЕРАТУРА

1. *M.J.D. Powell*, in: Optimization, London – N.Y., 1969, p. 283.

2. *Ю.Г. Евтушенко*, ДАН, т. 221, No.5, 1016 (1975).

3. *Б.С. Разумихин*, Автоматика и телемеханика, No.6 (1972).

4. *D. Morrison*, SIAM J. Numer. Anal., v.5, No.1 (1968).