Math-Net.Ru
Общероссийский математический портал

В. Г. Жадан, Модифицированные функции Лагранжа в нелинейном программировании, *Ж. вычисл. матем. и матем. физ.*, 1982, том 22, номер 2, 296–308

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
http://www.mathnet.ru/rus/agreement

Параметры загрузки:
IP: 3.139.86.131
5 ноября 2024 г., 00:24:01

ЖУРНАЛ
ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ

Том 22 Март 1982 Апрель № 2

УДК 519.853

МОДИФИЦИРОВАННЫЕ ФУНКЦИИ ЛАГРАНЖА В НЕЛИНЕЙНОМ ПРОГРАММИРОВАНИИ

ЖАДАН В. Г.

(Москва)

Описываются методы решения общей задачи нелинейного программирования с использованием модифицированных функций Лагранжа. Дано доказательство их сходимости.

§ 1. Введение

За последнее время большое количество работ было посвящено модифицированным и обобщенным функциям Лагранжа. Укажем, например, на [1]—[10]. В настоящей статье вводятся три новые модифицированные функции Лагранжа и на их основе строятся методы решения общей задачи нелинейного программирования. В частности, предлагаются методы, в которых одни переменные, прямые или двойственные, отыскиваются из решения соответствующей вспомогательной задачи на безусловный экстремум, а другие находятся как корни специальным образом подобранной системы уравнений. Впервые такой подход был применен в [6], [7]. Здесь он распространяется на более широкий класс методов.

Рассмотрим общую задачу нелинейного программирования: требуется найти

$$\min_{x\in X} f(x), \tag{1.1}$$

$$X=\{x\in E_n : g^i(x)=0,\ i=1,\ 2,\dots,l,\ g^i(x)\leqslant 0,\ i=l+1,\dots,m\},$$

где E_i есть i-мерное евклидово пространство, $f(x)$, $g^i(x)$, $i=1, 2,\dots, m$,— по крайней мере, дважды непрерывно дифференцируемые функции на E_n. Всюду считаем, что множество ее решений X_* непусто.

Пусть $\xi_1(t)$, $\xi_2(t)$ — дважды непрерывно дифференцируемые функции скалярного аргумента такие, что $\xi_2(0)=\xi_2'(0)=0$ и $|\xi_1'(t)|\geqslant c_1>0$, $\xi_2''(t)\geqslant c_2>0$ для любых t. Составим для задачи (1.1) функцию Лагранжа вида

$$L(z)=f(x)+\langle p(u),\ g(x)\rangle. \tag{1.2}$$

Здесь $z=[x,\ u]\in E_{n+m}$, $u\in E_m$; $g(x)$ и $p(u)$ — вектор-функции с компонентами $g^i(x)$, $i=1, 2,\dots, m$, и $p^i(u)=\xi_1(u^i)$, $i=1, 2,\dots, l$, $p^i(u)=\xi_2(u^i)$, $i=l+1,\dots, m$; $\langle a,\ b\rangle$ — скалярное произведение векторов a и b. Функция (1.2) является простейшим обобщением функции Лагранжа, предложенной в [11], где полагалось $\xi_1(t)=t$, $\xi_2(t)=t^2$.

Введем индексные множества $J(x)=\{1\leqslant i\leqslant m : g^i(x)=0\}$, $J_-(x)=\{l<i\leqslant m : g^i(x)=0\}$ и обозначим через f_x, L_x, L_u вектор-столбцы градиентов функций $f(x)$ и $L(z)$, через g_x, p_u, L_{xx}, L_{uu} — матрицы первых и вторых производных с размерами, соответственно, $m\times n$, $m\times m$, $n\times n$, $m\times m$. Матрицы p_u и L_{uu} диагональные. Символ т означает транспонирование матрицы.

Определение 1. Точка $z_*=[x_*, u_*]$, в которой

(1.3) $$L_x(z_*)=f_x(x_*)+g_x^{\mathrm{T}}(x_*)p(u_*)=0,$$

(1.4) $$L_u(z_*)=p_u(u_*)g(x_*)=0,$$

называется стационарной. Если, кроме того, $x_*\in X$, то она является точкой Куна — Таккера.

Определение 2. В точке x ограничения удовлетворяют условию регулярности, если векторы $g_x^i(x)$, $i\in J(x)$, линейно-независимы.

Определение 3. В стационарной точке $z_*=[x_*, u_*]$ выполнено условие строгой дополняющей нежесткости, если $u_*^i\neq 0$ для всех $i\in J_-(x_*)$.

Определение 4. В точке $z=[x, u]$ выполнено условие A_1, если $\langle y, L_{xx}(z)y\rangle>0$ для любых ненулевых y таких, что $p_u(u)g_x(x)y=0$.

Определение 5. В точке $z=[x, u]$ выполнено условие A_2, если $\langle w, L_{uu}(z)w\rangle<0$ для любых ненулевых w таких, что $p_u(u)w=0$.

Справедливы следующие легко проверяемые утверждения (см. [12]).

Лемма 1. *Пусть в стационарной точке z_* выполнено условие* A_2. *Тогда z_* — точка Куна — Таккера и в ней имеет место условие строгой дополняющей нежесткости.*

Лемма 2. *Пусть в стационарной точке z_* выполнены условия* A_1 *и* A_2. *Тогда x_* является изолированным локальным решением задачи* (1.1).

Введем в рассмотрение дважды непрерывно дифференцируемую на $E_1\times E_1$ функцию двух скалярных аргументов $\varphi(t, \theta)$, на которую наложим следующие требования.

Условие $Б_1$. $\varphi_t(t, \theta)=0$ тогда и только тогда, когда $t=0$.

Условие $Б_2$: $\varphi_{tt}(0, \theta)>0$ для любого $\theta\in E_1$.

Условие $Б_3$. $\varphi_{t\theta}(0, \theta)=0$ для любого $\theta\in E_1$.

Согласно условиям $Б_1$ и $Б_2$, функция $\varphi(t, \theta)$ для любого фиксированного $\theta\in E_1$ имеет в точке $t=0$ минимум по t. Простейшими примерами функций, удовлетворяющих $Б_1$ — $Б_3$, являются

$$\varphi^1(t, \theta)=t^2/2, \quad \varphi^2(t, \theta)=e^t-t,$$

$$\varphi^3(t, \theta)=[t \operatorname{arctg} t-0.5\ln(1+t^2)]e^{-\theta},$$

$$\varphi^4(t, \theta)=-e^{-t^2}\operatorname{ch}\theta.$$

Первые три из них строго выпуклые по t. Ниже функция φ используется при построении модифицированных функций Лагранжа.

Пусть a и b — векторы произвольной, но одинаковой размерности. Тогда через $\varphi_1(a, b)$ будем обозначать вектор-столбец с компонентами $\varphi_t(a^j, b^j)$, через $\varphi_{11}(a, b)$ — диагональную матрицу, j-й диагональный элемент которой равен $\varphi_{tt}(a^j, b^j)$. Символ I будет употребляться для обозначения единичных матриц.

§ 2. Двойственные методы

Составим модифицированную функцию Лагранжа

(2.1) $$G(x,u,\sigma)=L(x,u)+\sigma\sum_{i=1}^{m}\varphi(L_u^i(x,u), u^i),$$

где σ — положительный параметр. По своим свойствам функция (2.1) наиболее близка к обобщенной функции Лагранжа, предложенной в [6], [7]. Основное отличие состоит в том, что по-разному учитываются ограничения типа неравенств. Здесь главное внимание обращается не на требование допустимости $g^i(x)\leqslant 0$, а на условие дополняющей нежесткости (1.4).

Если функция φ такова, что $\varphi_\theta(0,\theta)=0$ для любого $\theta\in E_1$, то стационарные точки функции $L(z)$ являются одновременно стационарными точками функции $G(z, \sigma)$. Однако в общем случае это неверно. Из условий (1.3), (1.4) следует лишь, что $G_x(z, \sigma)=0$.

Рассмотрим вспомогательную задачу безусловной минимизации:

$$G(x(u,\sigma),u,\sigma)=\min_n G(x,u,\sigma). \tag{2.2}$$

Когда ее решение $x(u, \sigma)$ существует, то необходимо, чтобы

$$G_x(x, u, \sigma)=L_x(x, u)+\sigma g_x^{\mathrm{T}}(x)p_u(u)\varphi_1(L_u(x, u), u)=0 \tag{2.3}$$

при $x=x(u, \sigma)$.

Лемма 3. *Пусть в стационарной точке $z_*=[x_*, u_*]$ выполнено условие* A_1. *Тогда можно указать константу $\sigma_*>0$ такую, что при $\sigma>\sigma_*$ для всех u из некоторой окрестности $S(u_*, \sigma)$ точки u_* существует изолированное локальное решение $x(u, \sigma)$ задачи* (2.2), *удовлетворяющее условию $x(u_*, \sigma)=x_*$. Функция $x(u, \sigma)$ непрерывно дифференцируема на $S(u_*, \sigma)$.*

Доказательство. Точка z_* удовлетворяет (2.3). Более того, согласно лемме Финслера (см. [13]), матрица

$$G_{xx}(z_*, \sigma)=L_{xx}(z_*)+\sigma g_x^{\mathrm{T}}(x_*)p_u(u_*)\varphi_{11}(0, u_*)p_u(u_*)g_x(x_*) \tag{2.4}$$

положительно определена при σ достаточно больших, $\sigma>\sigma_*$. Тогда, по теореме о неявной функции, можно указать окрестность $S(u_*, \sigma)$ точки u_* такую, что на $S(u_*, \sigma)$ существует непрерывно дифференцируемая функция $x(u, \sigma)$, определяемая уравнением (2.3) и принимающая при $u=u_*$ значение x_*. Так как всегда можно выбрать $S(u_*, \sigma)$ настолько малой, чтобы при $u\in S(u_*, \sigma)$, $x=x(u, \sigma)$ матрица $G_{xx}(x, u, \sigma)$ оставалась положительно-определенной, то $x(u, \sigma)$ является решением задачи (2.2). Лемма доказана.

Обратимся теперь к задаче отыскания корней системы

$$\varphi_1(L_u(x(u, \sigma), u), u)=0. \tag{2.5}$$

Из (2.3) видно, что, решив (2.5), получим стационарные точки функции $L(z)$.

Применим для решения (2.5) метод простой итерации:

$$u_{s+1}=u_s+\alpha\varphi_1(L_u(x_s, u_s), u_s), \tag{2.6}$$

где $x_s=x(u_s, \sigma)$, α — некоторый шаг.

Теорема 1. *Пусть в стационарной точке $z_*=[x_*, u_*]$ выполнены условия* A_1 *и* A_2, *матрица $L_{xx}(z_*)$ не вырождена, ограничения в x_* удовлетворяют условию регулярности. Тогда можно указать $\bar\sigma>0$ и $\bar\alpha(\sigma)$ такие, что для любого $\sigma>\bar\sigma$ и всех $0<\alpha<\bar\alpha(\sigma)$ процесс* (2.6) *локально сходится к u_* со скоростью геометрической прогрессии.*

Доказательство. Согласно теореме Островского (см., например, [14]), процесс (2.6) локально сходится к u_* со скоростью геометрической прогрессии, если отображение $Q(u)=u+\alpha\varphi_1(L_u(x(u,\sigma),u),u)$ дифференцируемо в u_* и спектральный радиус ρ матрицы $Q_u(u_*)$ удовлетворяет условию

(2.7) $\rho(Q_u(u_*))<1.$

Проверим это неравенство и оценим интервал значений α, для которых оно имеет место. В силу леммы 3, если σ взято достаточно большим, $\sigma>\sigma_*$, то функция $x(u,\sigma)$ определена на некоторой окрестности $S(u_*,\sigma)$ и непрерывно дифференцируема на ней. Подставляя $x(u,\sigma)$ в (2.3) и дифференцируя получившееся тождество по u, находим

$$\frac{dx(u,\sigma)}{du}=-G_{xx}^{-1}(x(u,\sigma),u,\sigma)G_{xu}(x(u,\sigma),u,\sigma).$$

Вычислим теперь матрицу $Q_u(u_*)$:

$$Q_u(u_*)=I+\alpha\varphi_{11}(0,u_*)\left\{L_{uu}(z_*)+L_{ux}(z_*)\frac{dx(u_*,\sigma)}{du}\right\}=$$

$$=I+\alpha\varphi_{11}\{L_{uu}-p_ug_xG_{xx}^{-1}g_x^{\mathrm{T}}p_u[I+\sigma\varphi_{11}L_{uu}]\}.$$

Здесь и ниже для сокращения записи опускаются аргументы матриц, поскольку считается, что все они вычислены при $x=x_*$, $u=u_*$. Введя обозначения $\varkappa=\alpha/\sigma$, $R=p_ug_xG_{xx}^{-1}g_x^{\mathrm{T}}p_u$, приходим к

$$Q_u(u_*)=(1-\varkappa)I+\varkappa(I-\sigma\varphi_{11}R)(I+\sigma\varphi_{11}L_{uu}).$$

Матрица $\varphi_{11}=\varphi_{11}(0,u_*)$ положительно определена. Пусть $\varphi_{11}^{1/2}$ — корень квадратный из нее, $\tilde R=\varphi_{11}^{1/2}R\varphi_{11}^{1/2}$, $\tilde L_{uu}=\varphi_{11}^{1/2}L_{uu}\varphi_{11}^{1/2}$. Справедливы представления

$$I-\sigma\varphi_{11}R=\varphi_{11}^{1/2}(I-\sigma\tilde R)\varphi_{11}^{-1/2},\qquad I+\sigma\varphi_{11}L_{uu}=\varphi_{11}^{1/2}(I+\sigma\tilde L_{uu})\varphi_{11}^{-1/2},$$

используя которые, преобразуем $Q_u(u_*)$ к виду

$$Q_u(u_*)=\varphi_{11}^{1/2}[(1-\varkappa)I+\varkappa W]\varphi_{11}^{-1/2},$$

где $W=(I-\sigma\tilde R)(I+\sigma\tilde L_{uu})$.

С помощью формулы Шермана — Моррисона — Вудбери (см. [14]) найдем выражение для матрицы W. Поскольку $L_{xx}(z_*)$ — неособая матрица, то из (2.4) следует

$$G_{xx}^{-1}=L_{xx}^{-1}-\sigma L_{xx}^{-1}g_x^{\mathrm{T}}p_u\varphi_{11}^{1/2}(I+\sigma\varphi_{11}^{1/2}p_ug_xL_{xx}^{-1}g_x^{\mathrm{T}}p_u\varphi_{11}^{1/2})^{-1}\varphi_{11}^{1/2}p_ug_xL_{xx}^{-1}.$$

Отсюда, если обозначить $Z=\varphi_{11}^{1/2}p_ug_xL_{xx}^{-1}g_x^{\mathrm{T}}p_u\varphi_{11}^{1/2}$, получим

(2.8) $I-\sigma\tilde R=I-\sigma Z+\sigma^2Z(I+\sigma Z)^{-1}Z=(I+\sigma Z)^{-1}.$

Таким образом, $W=(I+\sigma Z)^{-1}(I+\sigma\tilde L_{uu})$.

Пусть μ — собственные значения матрицы $Q_u(u_*)$, ν — собственные значения матрицы W. Последние могут быть найдены из решения характеристического уравнения

(2.9) $|W-\nu I|=0.$

Так как матрицы $\varphi_{11}^{1/2}$, $\varphi_{11}^{-1/2}$ неособые, то, на основании теоремы о произведении определителей, числа μ являются одновременно собственными значениями матрицы $(1-\varkappa)I+\varkappa W$, следовательно, они связаны с числами ν соотношением

$$\mu=1+\varkappa(\nu-1). \tag{2.10}$$

Матрица $I+\sigma Z$ неособая, матрица $I+\sigma\tilde{L}_{uu}$ при σ достаточно большом также будет неособой. Поэтому собственные значения ν не изменятся, если умножить матрицу из (2.9) последовательно на $I+\sigma Z$ и на $(I+\sigma\tilde{L}_{uu})^{-1}$. Получим, что должно выполняться уравнение $|U-\eta I|=0$. Здесь $U=(I+\sigma\tilde{L}_{uu})^{-1}(I+\sigma Z)$, $\eta=1/\nu$.

Предположим для простоты, что множество $J(x_*)$ состоит из первых k индексов, $k\geqslant l$. Тогда матрица U представима в виде

$$U=\left\| \begin{array}{c:c} B_1 & 0 \\ \hdashline 0 & B_2 \end{array} \right\|,$$

где $B_1=I+\sigma Z_J$, Z_J — левая верхняя подматрица матрицы Z размера $k\times k$, B_2 — диагональная матрица размера $m-k$ с диагональными элементами

$$[1+\sigma\xi_2''(0)\varphi_{tt}(0,0)g^{k+1}(x_*)]^{-1},\ldots$$
$$\ldots,[1+\sigma\xi_2''(0)\varphi_{tt}(0,0)g^{m}(x_*)]^{-1}.$$

Согласно известному свойству матриц подобного типа, множество характеристических чисел матрицы U состоит из k характеристических чисел матрицы B_1 и $m-k$ характеристических чисел матрицы B_2. Обозначим их последовательно $\eta^1,\ldots,\eta^k$, $\eta^{k+1},\ldots,\eta^m$. Так как обе матрицы B_1 и B_2 симметричны, то все η^i, $1\leqslant i\leqslant m$, вещественны.

Понятно, что $\eta^i=[1+\sigma\xi_2''(0)\varphi_{tt}(0,0)g^i(x_*)]^{-1}$, $k<i\leqslant m$, поэтому при $\sigma>0$ и

$$\alpha<2\{\xi_2''(0)\varphi_{tt}(0,0)\max_{i\in J(x_*)}[-g^i(x_*)]\}^{-1}$$

все соответствующие $|\mu^i|<1$.

Оценим первые k из собственных значений μ^i. Имеем $\eta^i=1+\sigma\lambda^i$, $i\leqslant 1\leqslant k$, где λ^i — собственные значения матрицы Z_J. Отсюда и из (2.10) получаем

$$\mu^i=1-\frac{\alpha\lambda^i}{1+\sigma\lambda^i},\qquad 1\leqslant i\leqslant k.$$

Если все λ^i положительны, то $|\mu^i|<1$ для любых $\sigma>0$ и $\alpha<2\sigma$. В противном случае это неравенство имеет место, когда, например, $\sigma>-2/\lambda_*$, $\alpha\leqslant\sigma$. Здесь λ_* — минимальное по модулю отрицательное λ^i.

Таким образом, условие (2.7) действительно выполняется. Для этого достаточно взять

$$\bar\sigma=\begin{cases}\sigma_*, & \lambda_{\min}>0,\\ \max\{\sigma_*,-2/\lambda_*\}, & \lambda_{\min}<0,\end{cases} \tag{2.11}$$
$$\bar\alpha(\sigma)=\min\{\sigma,2\{\xi_2''(0)\varphi_{tt}(0,0)\max_{i\in J(x_*)}[-g^i(x_*)]\}^{-1}\}.$$

Теорема доказана.

Обсудим вопрос о величине шага α в (2.6). Для эффективной работы метода желательно, чтобы α был близок к верхней оценке $\overline{\alpha}(\sigma)$. Однако из-за наличия в точке x_* неактивных ограничений оценка может стать низкой. Увеличить значение $\overline{\alpha}(\sigma)$, как видно из (2.11), можно за счет выбора функции φ, а именно, следует воспользоваться такой функцией φ, у которой производная $\varphi_{tt}(0, 0)$ мала. Отметим также, что если в точке x_* все ограничения активные, то α можно полагать в точности равным σ. Этот случай реализуется, например, когда в задаче (1.1) отсутствуют ограничения типа неравенств.

Применяя для нахождения корней системы (2.5) метод Ньютона, приходим к итеративному процессу

$$u_{s+1}=u_s-\Lambda^{-1}(u_s, \sigma)\varphi_1(L_u(x_s, u_s), u_s). \tag{2.12}$$

Здесь

$$\Lambda(u, \sigma)=\frac{d}{du}\varphi_1(L_u(x(u, \sigma), u), u).$$

Теорема 2. *Пусть выполнены предположения теоремы* 1 *и матрица* $\Lambda(u, \sigma)$ *в окрестности точки* u_* *удовлетворяет условию Липшица. Тогда существует* $\overline{\sigma}>0$ *такое, что для всех* $\sigma>\overline{\sigma}$ *процесс* (2.12) *локально сходится к* u_* *с квадратичной скоростью.*

Доказательство. Достаточно показать, что матрица $\Lambda(u_*, \sigma)$ не вырождена. Воспользовавшись обозначениями, введенными при доказательстве предыдущей теоремы, получим

$$\Lambda(u_*, \sigma)=\varphi_{11}(0, u_*)\left[L_{uu}(z_*)+L_{ux}(z_*)\frac{dx(u_*, \sigma)}{du}\right]=$$

$$=(I-\sigma\varphi_{11}R)\varphi_{11}L_{uu}-\varphi_{11}R=\varphi_{11}^{1/2}[(I-\sigma\tilde{R})\tilde{L}_{uu}-\tilde{R}]\varphi_{11}^{-1/2}.$$

Отсюда видно, что $\Lambda(u_*, \sigma)$ будет невырожденной, если не вырождена матрица в квадратных скобках (обозначим ее D). Согласно (2.8),

$$D=(I+\sigma Z)^{-1}\tilde{L}_{uu}-\frac{1}{\sigma}[I-(I+\sigma Z)^{-1}],$$

или, если умножить ее на $I+\sigma Z$,

$$Y=(I+\sigma Z)D=\tilde{L}_{uu}-Z.$$

Считая, что множество $J(x_*)$ состоит из первых k индексов, представим матрицу Y в виде

$$Y=\left\|\begin{array}{c|c} C_1 & 0 \\ \hline 0 & C_2 \end{array}\right\|,$$

где $C_1=-Z_J$, C_2 — диагональная матрица размера $m-k$ с диагональными элементами $\xi_2''(0)\varphi_{tt}(0, 0)g^{k+1}(x_*), \ldots, \xi_2''(0)\varphi_{tt}(0, 0)g^m(x_*)$. В силу сделанных предположений, обе матрицы C_1 и C_2 не вырождены. Поэтому вся матрица Y также не вырождена. Поскольку Y — произведение невырожденной матрицы на D, то D не вырождена. Теорема доказана.

Возможны разнообразные варианты метода (2.12), например, матрицу $\Lambda^{-1}(u_s, \sigma)$ можно пересчитывать не на каждом шаге, а лишь на некоторой выделенной подпоследовательности номеров, и т. д.

§ 3. Прямые методы

Рассмотрим два других метода решения задачи (1.1), близких по идее к (2.6), (2.12), но использующих обратную зависимость двойственных переменных от прямых. По аналогии с двойственными методами назовем их прямыми.

Составим модифицированную функцию Лагранжа

$$H(x,u,\tau)=L(x,u)-\tau\sum_{j=1}^{n}\varphi(L_x^j(x,u),x^j), \tag{3.1}$$

где τ — положительный параметр. Ранее частные случаи функции (3.1) рассматривались в [8], [10].

Подобно (2.1), при переходе от функции $L(z)$ к функции $H(z, \tau)$ из условий стационарности (1.3), (1.4) сохраняется только одно, а именно второе.

Ниже покажем, что при определенных предположениях в окрестности точки $x_*\in X_*$ существует решение $u(x, \tau)$ вспомогательной задачи максимизации

$$H(x,u(x,\tau),\tau)=\max_{u\in E_m} H(x,u,\tau). \tag{3.2}$$

Вообще говоря, оно может быть не единственным. Для конкретного выделенного решения $u(x, \tau)$ положим $\Delta(x)=\{l<i\leqslant m\colon u^i(x, \tau)\neq 0\}$.

Лемма 4. *Пусть в стационарной точке $z_*=[x_*, u_*]$ выполнено условие A_2 и ограничения в x_* удовлетворяют условию регулярности. Тогда при $\tau>0$ для всех x из некоторой окрестности $S(x_*, \tau)$ точки x_* существует изолированное локальное решение $u(x, \tau)$ задачи (3.2) такое, что $u(x_*, \tau)=u_*$. Для этого решения $\Delta(x)\equiv\Delta(x_*)=J_-(x_*)$. Функция $u(x, \tau)$ непрерывно дифференцируема на $S(x_*, \tau)$.*

Доказательство. Рассмотрим систему уравнений

$$H_u(x, u, \tau)=p_u(u)[g(x)-\tau g_x(x)\varphi_1(L_x(x, u), x)]=0. \tag{3.3}$$

Точка z_* удовлетворяет ей. Далее имеем

$$H_{uu}(z_*, \tau)=L_{uu}(z_*)-\tau p_u(u_*)g_x(x_*)\varphi_{11}(0, x_*)g_x^{\mathrm{T}}(x_*)p_u(u_*).$$

Пусть для простоты множество $J(x_*)$ состоит из первых k индексов, $k\geqslant l$. Обозначим через $p_u^J(u)$ левую верхнюю диагональную подматрицу матрицы $p_u(u)$ размера k, через $g_x^J(x)$ — матрицу $g_x(x)$, у которой выброшены последние $m-k$ строк. Тогда матрица $H_{uu}(z_*, \tau)$ представима в виде

$$H_{uu}(z_*,\tau)=\left\|\begin{array}{c:c}-\tau N_1(z_*) & 0\\ \hdashline 0 & N_2(z_*)\end{array}\right\|, \tag{3.4}$$

где $N_1(z_*)=p_u^J(u_*)g_x^J(x_*)\varphi_{11}(0, x_*)[g_x^J(x_*)]^{\mathrm{T}}p_u^J(x_*)$, $N_2(z_*)$ — диагональная матрица размера $m-k$ с диагональными элементами $\xi_2''(0)g^{k+1}(x_*),\ldots$ $\ldots,\xi_2''(0)g^m(x_*)$. Понятно, что $N_2(z_*)$ — отрицательно-определенная матрица.

Введем векторы b, h_i, q_i, положив

$$b=\{[\varphi_{tt}(0, x_*^1)]^{1/2},\ldots,[\varphi_{tt}(0, x_*^n)]^{1/2}\}^{\mathrm{T}},$$

$$h_i=\frac{\partial p^i(u_\circ)}{\partial u^i}g_x^i(x_*),\qquad q_i=(b^1h_i^1,\ldots,b^nh_i^n)^\tau,\qquad i=1,2,\ldots,k.$$

Из условия регулярности ограничений следует линейная независимость векторов q_i, $i=1, 2,\ldots, k$. Действительно, если допустить противное, то найдутся коэффициенты $d_1,\ldots, d_k$, не все равные нулю и такие, что $\Sigma_{i=1}^k d_iq_i=0$. Но это означает, что

$$\varphi_{tt}(0,x_*^j)\sum_{i=1}^k d_ih_i^j=0,\qquad j=1,2,\ldots,n,$$

или

$$\sum_{i=1}^k d_ih_i=0.$$

Последнее равенство невозможно, поскольку векторы $h_1,\ldots,h_k$ коллинеарны векторам $g_x^1(x_*),\ldots,g_x^k(x_*)$. Таким образом, $N_1(z_*)$ — матрица Грамма для линейно-независимых векторов $q_1,\ldots,q_k$ и, следовательно, является положительно-определенной. Вся матрица $H_{uu}(z_*, \tau)$ будет отрицательно-определенной для любого $\tau>0$. Поэтому, по теореме о неявной функции, в некоторой окрестности x_* система (3.3) определяет однозначным образом непрерывно дифференцируемую функцию $\tilde{u}(x, \tau)$ такую, что $\tilde{u}(x_*, \tau)=u_*$.

Наряду с (3.3) рассмотрим ее подсистему, состоящую из первых k уравнений:

(3.5) $$p_u^J(u)[g^J(x)-\tau g_x^J(x)\varphi_1(L_x(x, u^J, \hat{u}^{k+1}(x, \tau),\ldots$$
$$\ldots,\hat{u}^m(x, \tau)), x)]=0.$$

Здесь $g^J(x)$ и u^J — векторы, содержащие только первые k компонент, соответственно, векторов $g(x)$ и u; $\hat{u}^i(x, \tau)$, $i=k+1,\ldots, m$,— функции, тождественно равные нулю. Якобиан этой системы по переменным u^J, вычисленный в точке $[x_*, u_*^J]\in E_{n+k}$, есть определитель матрицы $N_1(z_*)$, умноженный на $-\tau$. Следовательно, он не равняется нулю при $\tau>0$. Применяя снова теорему о неявной функции, приходим к выводу, что система (3.5) определяет однозначным образом непрерывно дифференцируемые функции $\hat{u}^i(x, \tau)$, $i=1, 2,\ldots, k$, такие, что $\hat{u}^i(x_*, \tau)=u_*^i$.

Вектор-функция $\hat{u}(x, \tau)=[\hat{u}^1(x, \tau),\ldots,\hat{u}^m(x, \tau)]$ удовлетворяет всей системе (3.3) и, в силу единственности $\tilde{u}(x, \tau)$, совпадает с ней вблизи x_*. Кроме того, на основании необходимых условий локального экстремума функция $u(x, \tau)$ также может быть найдена при посредстве (3.3). Поэтому для любого $\tau>0$ существует окрестность $S(x_*, \tau)$ точки x_* такая, что $u(x, \tau)\equiv\hat{u}(x, \tau)$ при $x\in S(x_*, \tau)$. Отсюда заключаем, что для этого решения $\Delta(x)\equiv\Delta(x_*)=J_-(x_*)$. Лемма доказана.

Рассмотрим теперь задачу отыскания корней системы

(3.6) $$\varphi_1(L_x(x, u(x, \tau)), x)=0.$$

Для ее решения применим итеративный процесс[1)]

(3.7) $$x_{s+1}=x_s-\beta\varphi_1(L_x(x_s, u_s), x_s),$$

где $u_s=u(x_s, \tau)$, β — некоторый шаг.

[1)] Метод предложен совместно с А. И. Голиковым.

Теорема 3. *Пусть в стационарной точке $z_*=[x_*, u_*]$ выполнены условия A_1, A_2 и ограничения в x_* удовлетворяют условию регулярности. Тогда для любого $\tau>0$ можно указать $\bar\beta(\tau)$ такое, что для всех $0<\beta<\bar\beta(\tau)$ процесс (3.7) локально сходится к x_* со скоростью геометрической прогрессии.*

Доказательство будем проводить с помощью теоремы Островского. Покажем, что отображение $F(x)=x-\beta\varphi_1(L_x(x, u(x, \tau)), x)$ дифференцируемо в x_* и спектральный радиус ρ матрицы $F_x(x_*)$ удовлетворяет условию

$$\rho(F_x(x_*))<1. \tag{3.8}$$

Дифференцируемость отображения $F(x)$ в x_* следует из непрерывной дифференцируемости функции $u(x, \tau)$ в некоторой окрестности $S(x_*, \tau)$, установленной в лемме 4. Если подставить $u(x, \tau)$ в (3.3) и продифференцировать получившееся тождество по x, то получим

$$\frac{du(x,\tau)}{dx}=-H_{uu}^{-1}(x,u(x,\tau),\tau)H_{ux}(x,u(x,\tau),\tau).$$

Из представления (3.4) вытекает

$$H_{uu}^{-1}(z_*,\tau)=-\left[\frac{1}{\tau}M_1(z_*)+M_2(z_*)\right],$$

где

$$M_1(z_*)=\left\|\begin{array}{c:c} N_1^{-1}(z_*) & 0 \\ \hdashline 0 & 0\end{array}\right\|,\qquad M_2(z_*)=\left\|\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & -N_2^{-1}(z_*)\end{array}\right\|.$$

Обе матрицы $M_1(z_*)$ и $M_2(z_*)$ неотрицательно определены. Кроме того, в силу условия (1.4) матрица $M_2(z_*)p_u(u_*)$ нулевая. Поэтому

$$\frac{du(x_*,\tau)}{dx}=\frac{1}{\tau}M_1(z_*)p_u(u_*)g_x(x_*)[I-\tau\varphi_{11}(0,x_*)L_{xx}(z_*)].$$

Вычислим теперь матрицу $F_x(x_*)$:

$$F_x(x_*)=I-\beta\varphi_{11}(0,x_*)\left[L_{xx}(z_*)+g_x^{\mathrm T}(x_*)p_u(u_*)\frac{du(x_*,\tau)}{dx}\right]=$$

$$=I-\beta\varphi_{11}\left[L_{xx}+\frac{1}{\tau}g_x^{\mathrm T}p_uM_1p_ug_x(I-\tau\varphi_{11}L_{xx})\right].$$

Здесь и ниже считается, что матрицы, в которых опущены аргументы, вычислены при $x=x_*$, $u=u_*$. Введя обозначения $\varkappa=\beta/\tau$, $P=g_x^{\mathrm T}p_uM_1p_ug_x$, получим

$$F_x(x_*)=(1-\varkappa)I+\varkappa(I-\varphi_{11}P)(I-\tau\varphi_{11}L_{xx}). \tag{3.9}$$

С помощью разложений

$$I-\varphi_{11}P=\varphi_{11}^{1/2}(I-\tilde P)\varphi_{11}^{-1/2},\qquad I-\tau\varphi_{11}L_{xx}=\varphi_{11}^{1/2}(I-\tau\tilde L_{xx})\varphi_{11}^{-1/2},$$

где $\tilde P=\varphi_{11}^{1/2}P\varphi_{11}^{1/2}$, $\tilde L_{xx}=\varphi_{11}^{1/2}L_{xx}\varphi_{11}^{1/2}$, $\varphi_{11}^{1/2}$ — квадратный корень из матрицы $\varphi_{11}(0, x_*)$, выражение (3.9) приводится к

$$F_x(x_*)=\varphi_{11}^{1/2}[(1-\varkappa)I+\varkappa V]\varphi_{11}^{-1/2}.$$

Здесь $V=(I-\tilde P)(I-\tau\tilde L_{xx})$.

Оценим собственные значения μ матрицы $F_x(x_*)$. Проводя рассуждения, почти дословно повторяющие сделанные при доказательстве теоремы 1, убеждаемся, что μ связаны с собственными значениями ν матрицы V соотношением (2.10). Согласно определению, числа ν удовлетворяют равенству

$$(3.10)\qquad Vy=\nu y,$$

где y — собственный вектор матрицы V.

Займемся изучением (3.10). Будем поступать аналогично тому, как это делалось при доказательстве леммы 5.4 из [15]. Пусть ν^j — произвольное собственное значение из набора ν, y^j — соответствующий ему собственный вектор, $1\leqslant j\leqslant n$. Умножив равенство (3.10) поочередно на $\tilde{P}$ и на $I-\tilde{P}$, получим

$$(3.11)\qquad \nu^j\tilde{P}y^j=0,\qquad (I-\tilde{P})(I-\tau\tilde{L}_{xx})y^j=\nu^j(I-\tilde{P})y^j.$$

Возможны два случая.

1. $\tilde{P}y^j\neq0$. Тогда $\nu^j=0$. Подставляя это значение ν^j в (2.10), получаем $\mu^j=1-\varkappa$. Отсюда заключаем, что $|\mu^j|<1$, если $\varkappa<2$.

2. $\tilde{P}y^j=0$, и, стало быть, (3.11) можно переписать в виде $(I-\tilde{P})\times\times(I-\tau\tilde{L}_{xx})(I-\tilde{P})y^j=\nu^jy^j$, т. е. ν^j является одновременно собственным значением матрицы $T=(I-\tilde{P})(I-\tau\tilde{L}_{xx})(I-\tilde{P})$. Эта матрица симметричная, и, значит, ее собственные значения и собственные векторы вещественны. Имеем

$$(3.12)\qquad \nu^j\langle y^j,\ y^j\rangle=\langle y^j,\ Ty^j\rangle=\langle y^j,\ y^j\rangle-\tau\langle y^j,\ \tilde{L}_{xx}y^j\rangle.$$

Но поскольку равенство $\tilde{P}y^j=0$ возможно тогда и только тогда, когда $p_ug_x\varphi_{11}y^j=0$, то, согласно условию A_1,

$$(3.13)\qquad (\nu^j-1)\langle y^j,\ y^j\rangle=-\tau\langle y^j,\ L_{xx}y^j\rangle<0.$$

Из (2.10) и (3.13) следует, что $\mu^j<1$, когда $\tau>0$, $\beta>0$. Кроме того, на основании (3.12) получаем

$$\nu^j=1-\tau\frac{\langle y^j,\ (I-\tilde{P})\tilde{L}_{xx}(I-\tilde{P})y^j\rangle}{\langle y^j,y^j\rangle}\geqslant1-\tau\omega_*,$$

где ω_* — максимальное собственное значение матрицы $(I-\tilde{P})\tilde{L}_{xx}(I-\tilde{P})$. Поэтому $\mu^j\geqslant1-\varkappa\omega_*>-1$, если $\beta<2/\omega_*$. Суммируя все сказанное, приходим к выводу, что условие (3.8) выполняется для любых $\tau>0$ и $\beta<\bar{\beta}(\tau)=2\min\{\tau,\ 1/\omega_*\}$. Теорема доказана.

Как видно из выражения для $\bar{\beta}(\tau)$, при решении задачи (1.1) методом (3.7) параметр τ надо подобрать таким образом, чтобы выполнялось неравенство $\tau<1/\omega_*$, ибо в этом случае, взяв β близким к $\bar{\beta}(\tau)$, можно добиться, чтобы отношение β/τ, которое фактически определяет скорость перемещения в пространстве x, стало наибольшим. Однако брать τ слишком малым, как свидетельствует опыт численных расчетов, не следует.

Рассмотрим теперь аналог метода (2.12). Применение метода Ньютона к решению уравнения (3.6) приводит к схеме

$$(3.14)\qquad x_{s+1}=x_s-\Gamma^{-1}(x_s,\ \tau)\varphi_1(L_x(x_s,\ u_s),\ x_s),$$

где

$$\Gamma(x, \tau) = \frac{d}{dx} \varphi_1(L_x, (x, u(x, \tau)), x).$$

Теорема 4. *Пусть выполнены предположения теоремы 3 и матрица $\Gamma(x, \tau)$ в окрестности точки x_* удовлетворяет условию Липшица. Тогда существует $\bar{\tau}>0$ такое, что для всех $0<\tau<\bar{\tau}$ процесс (3.14) локально сходится к x_* с квадратичной скоростью.*

Доказательство. Сохраняя обозначения, введенные при доказательстве предыдущей теоремы, вычислим матрицу $\Gamma(x_*, \tau)$:

$$\Gamma(x_*, \tau) = \varphi_{11}(0, x_*) \left[L_{xx}(z_*) + L_{xu}(z_*) \frac{du(x_*, \tau)}{dx} \right] =$$

$$= (I - \varphi_{11} P) \varphi_{11} L_{xx} + \frac{1}{\tau} \varphi_{11} P = \varphi_{11}^{1/2} K(\tau) \varphi_{11}^{-1/2},$$

где $K(\tau) = (I - \tilde{P}) \tilde{L}_{xx} + \tilde{P}/\tau$.

Пусть, как и в доказательстве леммы 4, множество $J(x_*)$ состоит из первых k индексов. Тогда $\tilde{P} = \Omega^{\mathrm{T}}(\Omega\Omega^{\mathrm{T}})^{-1}\Omega$, $\Omega = p_u{}^J g_x{}^J \varphi_{11}^{1/2}$ и, следовательно, $\tilde{P}$ и $I - \tilde{P}$ — матрицы проектирования на линейное подпространство, натянутое на векторы $q_1, \ldots, q_k$ и на его ортогональное дополнение. Отсюда, используя условие A_1 и свойства матриц проектирования, с помощью леммы Финслера (см. [13]) убеждаемся, что симметричная матрица $K(\tau) + K^{\mathrm{T}}(\tau)$ при достаточно малых τ положительно определена. Поэтому сама матрица $K(\tau)$ также положительно определена и, стало быть, $\Gamma(x_*, \tau)$ не вырождена.

Таким образом, выполнены условия, гарантирующие локальную сходимость с квадратичной скоростью метода (3.14). Теорема доказана.

При решении (1.1) предложенными методами основной объем вычислений падает на решение вспомогательных задач (2.2), (3.2). Поэтому если размерность n вектора x меньше числа ограничений m, то выгоднее применять двойственные методы. Напротив, в случае, когда n существенно превышает m, более эффективными оказываются прямые методы. При решении задач, в которых числа n и m незначительно отличаются друг от друга, можно использовать как прямые, так и двойственные методы. Однако если вычисление значений функций $f(x)$ и $g(x)$ трудоемко, то некоторое предпочтение можно отдать прямым методам, поскольку они требуют меньшего количества обращений к вычислению значений функций по сравнению с двойственными.

§ 4. Нахождение седловых точек

Комбинируя функции (2.1) и (3.1), построим модифицированную функцию Лагранжа, которая строго выпукло-вогнута в окрестности точек Куна — Таккера. Для этой цели нам потребуются новые условия.

Условие $\mathrm{Б}_4$. $\varphi_t(t, \theta) = 0$ тогда и только тогда, когда $t=0$ или $\theta=0$.

Условие $\mathrm{Б}_5$. $\varphi_{tt}(0, \theta) > 0$ для любых $\theta \neq 0$, $\varphi_{tt}(0, 0) = 0$.

Условие $\mathrm{Б}_6$. $\varphi_\theta(0, \theta) = 0$, $\varphi_{\theta\theta}(0, \theta) \leq 0$ для любых $\theta \in E_1$.

Для всех рассмотренных выше функций $\varphi^r(t, \theta)$, $r=1, 2, 3, 4$, условие $\mathrm{Б}_6$ выполняется. Приведем примеры функций $\varphi(t, \theta)$, для которых выполняются Б$_3$ — Б$_6$:

$$\varphi^5(t,\theta)=\frac{t^2\theta^2}{2},\qquad \varphi^6(t,\theta)=\frac{t^2(\operatorname{ch}\theta-1)}{2},$$

$$\varphi^7(t,\theta)=\left[t\operatorname{arctg}t-\frac{1}{2}\ln(1+t^2)\right]\frac{\theta^2}{2}.$$

Составим модифицированную функцию Лагранжа:

$$\Phi(x,u,\gamma)=L(x,u)-\gamma\sum_{j=1}^{n}\varphi(L_x^j(x,u),x^j)+\tag{4.1}$$

$$+\frac{1}{\gamma}\sum_{i=1}^{m}\psi(L_u^i(x,u),u^i).$$

Здесь γ — положительный параметр, φ и ψ — произвольные функции, удовлетворяющие, соответственно, условиям Б$_1$ — Б$_3$ и Б$_3$ — Б$_5$, обе функции φ и ψ удовлетворяют Б$_6$.

Определение 6. В стационарной точке $z_*=[x_*, u_*]$ выполнено усиленное условие строгой дополняющей нежесткости, если $u_*^i\neq 0$ для всех $i\in J(x_*)$.

Теорема 5. *Пусть в стационарной точке $z_*=[x_*, u_*]$ выполнены условия А$_1$, А$_2$ и усиленное условие строгой дополняющей нежесткости. Пусть, кроме того, ограничения в x_* удовлетворяют условию регулярности. Тогда существует $\bar\gamma>0$ такое, что для всех $0<\gamma<\bar\gamma$ точка z_* будет строгим локальным седлом функции* (4.1).

Согласно этой теореме, модифицированная функция (4.1) обладает важным свойством, которое позволяет использовать известные методы отыскания седловых точек (см. [16]) для нахождения решения задачи (1.1). Укажем среди них градиентный метод и метод Ньютона:

$$x_{s+1}=x_s-\alpha\Phi_x(x_s,u_s,\gamma),\qquad u_{s+1}=u_s+\alpha\Phi_u(x_s,u_s,\gamma),\tag{4.2}$$

$$z_{s+1}=z_s-\Phi_{zz}^{-1}(z_s,\gamma)\,\Phi_z(z_s,\gamma),\tag{4.3}$$

где Φ_x, Φ_u, Φ_z — вектор-градиенты функции Φ по соответствующим компонентам, Φ_{zz} — матрица вторых производных, α — некоторый шаг.

Теорема 6. *Пусть $f(x)$, $g(x)$, $p(u)$ — трижды непрерывно дифференцируемые функции и выполнены предположения теоремы* 3. *Тогда существует $\bar\gamma>0$ такое, что для всех $0<\gamma<\bar\gamma$ метод* (4.2) *при α достаточно малом локально сходится к z_* со скоростью геометрической прогрессии. Если, кроме того, матрица $\Phi_{zz}(z,\gamma)$ удовлетворяет условию Липшица в окрестности точки z_*, то метод* (4.3) *локально сходится к z_* с квадратичной скоростью.*

По сравнению с методами, предложенными в предыдущих параграфах, методы (4.2), (4.3) обладают определенным преимуществом, так как отпадает необходимость решать трудоемкие вспомогательные задачи (2.2) или (3.2). Однако имеются и недостатки. Во-первых, возрастает число переменных. Во-вторых, процесс (4.2) является, по существу, методом с «малым» шагом, в котором α нельзя брать большим. В третьих, методы (4.2), (4.3) требуют знания производных функций $f(x)$ и $g(x)$ более высокого порядка. Улучшить характеристики сходимости методов (4.2), (4.3) можно, применяя подходы, описанные в [17].

В заключение автор выражает признательность Ю. Г. Евтушенко за внимание к работе и сделанные замечания.

Литература

1. *Hestenes M. R.* Multiplier and gradient methods.– J. Optimizat. Theory and Appl., 1969, v. 4, № 5, p. 303–320.
2. *Поляк Б. Т., Третьяков Н. В.* Метод штрафных оценок для задач на условный экстремум.– Ж. вычисл. матем. и матем. физ., 1973, т. 13, № 1, с. 33–46.
3. *Гольштейн Е. Г., Третьяков Н. В.* Модифицированные функции Лагранжа.– Экономика и матем. методы, 1974, т. 10, № 3, с. 568–591.
4. *Mangasarian O. L.* Unconstrained Lagrangians in nonlinear programming.– SIAM J. Control, 1975, v. 13, № 4, p. 772–791.
5. *Kort B. W., Bertsekas D. P.* Combined primal-dual and penalty methods for convex programming.– SIAM J. Control and Optimization, 1976, v. 14, № 2, p. 268–295.
6. *Голиков А. И., Евтушенко Ю. Г.* Об одном классе методов решения задач нелинейного программирования.– Докл. АН СССР, 1978, т. 239, № 3, с. 519–522.
7. *Евтушенко Ю. Г.* Применение модифицированных функций Лагранжа для решения задач нелинейного программирования.– В кн.: Иссл. операций. Вып. 7. М.: ВЦ АН СССР, 1979, с. 3–24.
8. *Антипин А. С.* Методы математического программирования, основанные на прямой и двойственной модификации функции Лагранжа.– Препринт ВНИИ системных иссл. М., 1979.
9. *Гольштейн Е. Г., Третьяков Н. В.* Слабые модифицированные функции Лагранжа и связанные с ними теоремы двойственности.– Экономика и матем. методы, 1979, т. 15, № 6, с. 1180–1193.
10. *Голиков А. И., Жадан В. Г.* Итеративные методы решения задач нелинейного программирования с использованием модифицированных функций Лагранжа.– Ж. вычисл. матем. и матем. физ., 1980, т. 20, № 4, с. 874–888.
11. *Евтушенко Ю. Г.* Численные методы нелинейного программирования.– Докл. АН СССР, 1975, т. 221, № 5, с. 1016–1019.
12. *Фиакко А., Мак-Кормик Г.* Нелинейное программирование. Методы последовательной безусловной минимизации. М.: Мир, 1972.
13. *Беллман Р.* Введение в теорию матриц. М.: Наука, 1969.
14. *Ортега Дж., Рейнболдт В.* Итерационные методы решения нелинейных систем уравнений со многими неизвестными. М.: Мир, 1975.
15. *Пшеничный Б. Н., Данилин Ю. М.* Численные методы в экстремальных задачах. М.: Наука, 1975.
16. *Демьянов В. Ф., Певный А. Б.* Численные методы разыскания седловых точек.– Ж. вычисл. матем. и матем. физ., 1972, т. 12, № 5, с. 1099–1127.
17. *Данилин Ю. М., Панин В. М.* О некоторых методах поиска седловых точек.– Кибернетика, 1974, № 3, с. 119–124.

Поступила в редакцию 21.V.1980