Доклады Академии наук СССР
1973. Том 209, № 5

МАТЕМАТИКА

Л. А. КИРИЕВСКИЙ, Р. А. ПОЛЯК

ОБ УСКОРЕНИИ СХОДИМОСТИ МЕТОДОВ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

(Представлено академиком Л. В. Канторовичем 28 VI 1972)

В работе для задач выпуклого программирования при линейных ограничениях построено три общих сходящихся метода. Их отличительной особенностью является то, что, начиная с некоторого момента, все приближения оказываются в линейном многообразии, содержащем искомый экстремум. С этого момента различные конкретизации построенных процессов автоматически превращаются в методы отыскания экстремума функции при отсутствии ограничений, причем в пространстве, размерность которого, вообще говоря, ниже размерности исходного пространства.

Это достигается благодаря введению некоторого параметра, который не только удерживает вычислительный процесс на многообразии, содержащем экстремум, но и устраняет, в случае необходимости, возможность «зигзагообразного» движения в окрестности точки, не являющейся решением.

1. Пусть в $\mathbf{R}^n$ задана вогнутая функция $f(x)$ и многогранник $\Omega = \{x: (a^j, x) \geqslant b_j, j \in J = \{1, \ldots, r\}$.

Рассмотрим задачу отыскания

$$f(x^*) = \max \{f(x): x \in \Omega\}. \tag{1}$$

На Ω определим неотрицательную функцию $\mu(x)$, являющуюся «мерой неоптимальности» вектора x, т. е. такую, что из $\mu(x) = 0$ следует $x = x^*$.

Для $x \in \Omega$ определим $J(x) = \{j: (a^j, x) = b_j\}$ и грань $\Gamma(x) = \{y: (a^j, y) = b_j, j \in J(x); y \in \Omega\}$.

Пусть $Q(\Gamma)$ — линейное многообразие минимальной размерности, содержащее Γ. Для простоты рассуждений предположим, что для каждого $\Gamma \subset \Omega$ существует и единственная точка $x_\Gamma{}^* = \arg\max \{f(x) \,|\, x \in Q(\Gamma)\}$.

Для $\delta \geqslant 0$ и $x \in \Omega$ рассмотрим неотрицательную функцию $\mu(x, \delta)$, удовлетворяющую следующему условию:

$$\varliminf_{x^k \to x,\ \delta_k \to 0} \mu(x^k, \delta_k) \geqslant \mu(x, 0) = \mu(x), \tag{2}$$

а также оператор $C: \Omega \to \Omega$, удовлетворяющий условию: для $\forall \delta > 0$ найдется $\sigma(\delta) > 0$ такое, что при $\mu(x, \delta) > \delta$ имеет место

$$f(Cx) - f(x) \geqslant \sigma(\delta) \qquad \forall x \in \Omega. \tag{3}$$

Назовем R: $\Omega \to \Omega$ оператором релаксации, если $\Gamma(Rx) \subset \Gamma(x)$ и $f(Rx) \geqslant f(x)$. При различных предположениях относительно $f(x)$ операторы R могут обладать некоторыми из следующих свойств:

S) Если $\Gamma(Rx^k) = \Gamma(x^k) = \Gamma$, $k = 1, 2, \ldots$, $f(x^{k+1}) \leqslant f(Rx^k)$, то $f(x^k) \to f(x^*)$.

Пусть $x^{k+1} = Rx^k$, $\Gamma(x^k) = \Gamma$, $k = 1, 2, \ldots$, $x^k \to x^*$. Тогда

LS) $f(x^*) - f(x^{k+1}) \leqslant q[f(x^*) - f(x^k)]$, $q < 1$.

SLS) $f(x^*) - f(x^{k+1}) \leqslant q_k [f(x^*) - f(x^k)]$, $\lim\limits_{k\to\infty} q_k = 0$.

KS) $f(x^*) - f(x^{k+1}) \leqslant q[f(x^*) - f(x^k)]^2$.

2. Приведем три общих метода решения задачи (1).

Метод CR1 с постоянным параметром заключается в следующем. Пусть параметр δ удовлетворяет условию

$$0 < \delta < \min\{\mu(x_\Gamma^*, \delta):\ x_\Gamma^* \in \Omega,\ x_\Gamma^* \neq x^*\}. \tag{4}$$

В качестве исходного приближения выбираем $x^1 \in \Omega$ и полагаем $\theta_1 = 0$. Пусть уже построено приближение x^k и найдено θ_k.

В случае $\theta_k = 0$ полагаем $\theta_{k+1} = 1$ и проверяем неравенство $\mu(x^k, \delta) > \delta$. Если это неравенство выполнено, то $x^{k+1} = Cx^k$, в противном случае $x^{k+1} = x^k$.

В случае $\theta_k = 1$ полагаем $x^{k+1} = Rx^k$, если $\Gamma(x^{k+1}) = \Gamma(x^k)$, то $\theta_{k+1} = 0$, иначе $\theta_{k+1} = 1$.

Выбор параметра δ из условия (4) представляет определенные трудности. В приведенных ниже методах отсутствует априорное требование (4) на значение параметра δ.

Последовательность $\{\delta_i\}$ будем называть *управляющей*, если $\delta_i > 0$, $\delta_i \to 0$.

Метод CR2 заключается в следующем. Выберем произвольную управляющую последовательность $\{\delta_i\}$. В качестве исходного приближения выбираем $x^1 \in \Omega$ и полагаем $\theta_1 = 0$, $i(1) = 1$.

Пусть уже построено приближение x^k и найдены θ_k и $i(k)$.

В случае $\theta_k = 0$ полагаем $\theta_{k+1} = 1$ и проверяем неравенство $\mu(x^k, \delta_{i(k)}) > \delta_{i(k)}$. Если оно выполнено, то полагаем $x^{k+1} = Cx^k$ и $i(k+1) = i(k)$. В противном случае $x^{k+1} = x^k$, $i(k+1) = i(k) + 1$.

При $\theta_k = 1$ полагаем $i(k+1) = i(k)$, $x^{k+1} = Rx^k$ и, если $\Gamma(x^{k+1}) = \Gamma(x^k)$, то $\theta_{k+1} = 0$, иначе $\theta_{k+1} = 1$.

Иногда в методе CR2 управляющую последовательность удается подчинить дополнительному условию

$$\varlimsup_{\substack{\theta_k = 0,\\ k\to\infty}} (f(x^*) - f(x^k))\,[\sigma(\delta_{i(k)})]^{-1} < 1. \tag{5}$$

На $Q(\Gamma)$ определим неотрицательную функцию $\nu(x)$, являющуюся «мерой неоптимальности на $Q(\Gamma)$» вектора x, т. е. такую, что из $\nu(x) = 0$ следует $x = x_\Gamma^*$. Пусть $\mu(x)$ и $\nu(x)$ удовлетворяет следующим условиям:

если $x^k \in \Gamma$, $x^k \to x_\Gamma^*$, то $\nu(x^k) \to 0$;

если $x^k \to x^*$, то

$$\mu(x^k)\,[\nu(x^k)]^{-1} \to 0. \tag{6}$$

Метод CR3 со стабилизирующимся параметром получим, видоизменив следующим образом метод CR2: в случае $\theta_k = 0$ при $\mu(x^k, \delta_{i(k)}) \leqslant \delta_{i(k)}$ полагаем $x^{k+1} = x^k$ и проверяем неравенство $\mu(x^k, \delta_{i(k)}) > \nu(x^k)$; если оно выполнено, то $i(k+1) = i(k) + 1$, иначе $i(k+1) = i(k)$.

3. Характер сходимости методов CR1 — CR3 устанавливает следующая

Теорема. *Пусть* $\mu(x, \delta)$ *удовлетворяет* (2), *оператор* C *удовлетворяет* (3), *а оператор* R *обладает свойством* S. *Тогда:*

1) *если в* CR1 *параметр* δ *выбран из условия* (4), *а в* CR2, CR3 *значение параметра определяют элементы произвольной управляющей последовательности* $\{\delta_i\}$, *то методы* CR1 — CR3 *порождают последовательность* $\{x^k\}$, *каждая предельная точка которой является решением задачи* (1), *а* $f(x^k) \to f(x^*)$;

2) *если в* CR1 *параметр* δ *выбран из условия* (4), *в* CR2 *последовательность* $\{\delta_i\}$ *удовлетворяет* (5), *а в* CR3 *функция* $\nu(x)$ *удовлетворяет* (6), *то методы* CR1 — CR3 *порождают такую последовательность* $\{x^k\}$, *что, начиная с некоторого* k_0, *имеет место:*

а) $x^k \in \Gamma$, Γ: $x_\Gamma^* = x^*$, б) *при* $\theta_k = 0$ *приближение* $x^{k+1} = x^k$, *т. е., начиная с* k, *новые приближения получаются лишь применением оператора* R;

в) *в методе* CR3 *дополнительно имеет место* $i(k) = i(k_0)$, $k \geqslant k_0$;

3) *если оператор* R *обладает одним из свойств* LS — KS, *то при выполнении условий предыдущего пункта теоремы каждый из методов* CR1 — CR3 *порождает последовательность* $\{x^k\}$, *сходящуюся к* x^* *соответственно линейно, сверхлинейно, квадратично.* *

Замечание. Если x^* — вершина Ω, то решение задачи (1) может быть найдено незначительной модификацией методов CR1 — CR3 за конечное число шагов.

4. Рассмотрим три примера функций $\mu(x, \delta)$, $\nu(x)$ и операторов C.

1^0) Пусть δ удовлетворяет (4) и такое, что для $\forall x \in \Omega$ существует $y \in \Omega$, для которого $J(y) = J(x, \delta) = \{j\colon b_j \leqslant (a^j, x) < b_j + \delta\}$. Рассмотрим $x \in \partial\Omega$ и пусть $J(x, \delta) = \{1, \ldots, l\}$. Для простоты будем считать, что $b_j = 0$, $j \in J(x, \delta)$, и векторы-столбцы a^j, $j \in J(x, \delta)$ линейно независимы.

Рассмотрим векторы $u(x, \delta) = (A_l^T A_l)^{-1} A_l^T \nabla f(x)$, $z(x, \delta) = \nabla f(x) - A_l u(x, \delta)$, $z_{-1}(x, \delta) = \nabla f(x) - A_{l-1}(A_{l-1}^T A_{l-1})^{-1} A_{l-1}^T \nabla f(x)$, где $A_l = (a^1, \ldots, a^l)$, а A_{l-1} получается из A_l удалением столбца $a^{\bar{l}}$: $u_{\bar{l}}(x, \delta) = \max\limits_{1 \leqslant j < l} u_j(x, \delta)$.

Пусть $\mu 1(x, \delta) = \max\{\|z(x, \delta)\|, u_{\bar{l}}(x, \delta)\}$, а оператор проекции градиента $C1$ определим равенством

$$C1(x) = x + t(x)\bar{z}(x, \delta),$$

где

$$\bar{z}(x, \delta) = \begin{cases} z(x, \delta), & \text{если } \mu 1(x, \delta) = \|z(x, \delta)\|, \\ z_{-1}(x, \delta), & \text{если } \mu 1(x, \delta) = u_{\bar{l}}(x, \delta), \end{cases}$$

$$t(x) = \arg\max\{f(x + t\bar{z}(x, \delta)) \,|\, 0 \leqslant t \leqslant t'(x)\},$$

$$t'(x) = \max\{t \,|\, x + t\bar{z}(x, \delta) \in \Omega\}.$$

Здесь и в дальнейшем $\nu(x) = \|z(x, 0)\|^{1/2}$.

2^0) Рассмотрим $\mu 2(x) = \max\{(\nabla f(x), z) \,|\, A_l^T z \geqslant 0, \|z\| \leqslant 1\} = (\nabla f(x), \bar{z}(x, \delta))$, $\|\cdot\|$ — некоторая норма вектора в $\mathbf{R}^N$.

Оператор наискорейшего спуска $C2$ определим аналогично $C1$.

3^0) Пусть $\mu 3(x) = \mu 3(x, \delta) = \max\{(\nabla f(x), y - x) \,|\, y \in \Omega\} = (\nabla f(x), \bar{z}(x, \delta)) = (\nabla f(x), \bar{z}(x))$. Оператор условного градиента $C3$ определяется аналогично $C1$.

Функции $\mu 1$, $\mu 2$, $\mu 3$ удовлетворяют (2), а гладкости $f(x)$ достаточно для того, чтобы операторы $C1$ — $C3$ удовлетворяли (3).

5. Приведем примеры операторов релаксации. Пусть $\varphi(\bar{x})$, $\bar{x} \in \mathbf{R}^n$, — сужение $f(x)$ относительно многообразия $Q(\Gamma)$, а $g(\bar{x}) = \nabla\varphi(\bar{x})$.

1^{00}) Рассмотрим последовательность $\{\bar{x}^s\}_0^n$, порождаемую процессом сопряженных градиентов ([1]):

$$\bar{x}^0 = \bar{x}, \quad p^0 = g(\bar{x}^0); \quad \bar{x}^{s+1} = \bar{x}^s + t_s p_s,$$

где $t_s = \arg\max\{\varphi(\bar{x}^s + tp_s) \,|\, 0 < t_s \leqslant t_s'\}$, а $t_s' = \max\{t \,|\, \bar{x}^s + tp^s \in \Omega\}$, $p^{s+1} = g(\bar{x}^{s+1}) + \|g(\bar{x}^{s+1})\|^2 \cdot \|g(\bar{x}^s)\|^{-2} p^s$.

Оператор сопряженных градиентов $R1$ определим по формуле

$$R1(\bar{x}) = \bar{x}^{\bar{s}+1}, \quad \bar{s} = \min\{n - 1, \min\{s \,|\, t_s = t_s', 0 \leqslant s < n\}\}.$$

2^{00}) Рассмотрим, следуя ([2]), последовательность $\{\bar{x}^s\}^n$ типа (8), в которой $p^s = H_s^T g(\bar{x}^s)$, $H_s = E + g(\bar{x}^s)[p^{s-1}]^T[(p^{s-1}, g(\bar{x}^{s-1}))]^{-1}$, где $H_0 = E$ — единичная матрица. Оператор $R2$ определим для указанной последовательности аналогично $R1$.

---

* Утверждение 1) для метода CR2 имеет место при произвольном операторе релаксации, например при RO: $RO(x) = x$.

3^{00}) Оператор градиентной релаксации $R3$ определим по формуле

$$R3(\bar{x}) = \bar{x} + t(\bar{x})g(\bar{x}),$$

где $t(\bar{x})$ — некоторый параметр.

Если для гессиана $H(\bar{x})$ функции $\varphi(\bar{x})$ выполнены условия

$$m(y, y) \leqslant -(H(\bar{x}^*)y, y) \leqslant M(y, y) \quad \forall y \in R^n, \\ \|H(\bar{x}^1) - H(\bar{x}^2)\| \leqslant L\|\bar{x}^1 - \bar{x}^2\|, \tag{7}$$

где $\bar{x}^1$ и $\bar{x}^2$ принадлежат достаточно малой окрестности $\bar{x}^*$, то, как установлено в ([3]) (см. также ([4], [5])), оператор $R1$ обладает свойством KS.

В случае непрерывного гессиана, удовлетворяющего лишь первому из условий (7), оператор $R2$ обладает (см., например, ([6])) свойством SLS.

В ([7–9]) для $\varphi(\bar{x})$ и параметра $t(\bar{x})$ найдены условия, при которых оператор $R3$ обладает свойствами S и LS. В связи с этим заметим, что конструкция процессов CR1 — CR3 позволяет сохранить для задачи (1) все факты общей теории релаксации ([8], [9]).

Заметим далее, что управляющие последовательности $\{\delta_k = k^{-1}\}$, $\{\delta_k = -q^{k/2}, q < 1\}$ и $\{\delta_k = k^{-k/2}\}$ удовлетворяют условию (5), если оператор R обладает соответственно свойствами LS — KS, а $g(\bar{x})$ удовлетворяет условию Липшица.

В заключение отметим, что метод условного градиента ([10], [11]) является конкретизацией CR1 с $\delta = 0$, $C = C3$ и $R = RO$. Методы наискорейшего спуска ([12], [13]) являются реализациями CR2 с $C = C2$ и $R = RO$, а метод проекции градиента ([14]) конкретизирует CR1 с $\delta = 0$, оператором $C = C1$ и $R = RO$, причем в ([14]) установлено лишь, что $\lim\limits_{k\to\infty} \mu(x^k)t(x^k) = 0$, из чего, вообще говоря, не следует сходимость метода. Отмеченного недостатка лишен базирующийся на методе ([14]) алгоритм ([15]), в котором, как и в методах ([12], [13]), для устранения «зигзагообразного» движения введен «сглаживающий» параметр.

Государственный институт по проектированию «Киевпроект»

Украинский филиал научно-исследовательского института планирования и нормативов
Киев

Поступило
22 VI 1972

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] R. Fletcher, R. Reeves, Comput. J., 7, № 2 (1964). [2] H. Huang, J. Opt. Theory Appl., 5, № 6 (1970). [3] Г. Д. Майстровский, Вычислит. матем. и вычислит. техн., в. (1971). [4] С. А. Смоляк, Тр. третьей зимн. школы по матем. програм., в. 3, М., 1970. [5] Б. Т. Поляк, Тр. второй зимн. школы по матем. програм., в. 1, М., 1969. [6] Ю. М. Данилин, Кибернетика, № 5 (1971). [7] Б. Т. Поляк, Журн. вычислит. матем. и матем. физ., 3, № 4 (1963). [8] И. М. Глазман, Тр. первой зимн. школы по матем. програм., в. 1, М., 1969. [9] Ю. И. Любич, Г. Д. Майстровский, УМН, 25, в. 1 (151) (1970). [10] M. Frank, Ph. Wolfe, Nav. Res. Log. Quart., 3 (1956). [11] В. Ф. Демьянов, А. М. Рубинов, Вестн. Ленингр. унив., № 19 (1964). [12] Г. Зойтендейк, Методы возможных направлений, ИЛ, 1963. [13] С. И. Зуховицкий, Р. А. Поляк, М. Е. Примак, ДАН, 153, № 5 (1963). [14] J. B. Rozen, J. Soc. Intust. Appl. Math., 8, № 1 (1960). [15] Б. Н. Пшеничный, И. Ф. Ганжела, Кибернетика, № 3 (1970).