Доклады Академии наук СССР
1973. Том 212, № 5

УДК 518.12 МАТЕМАТИКА

Р. А. ПОЛЯК

К УСКОРЕНИЮ СХОДИМОСТИ МЕТОДОВ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

(Представлено академиком Л. В. Канторовичем 13 II 1973)

В настоящей работе строится общий метод для решения задачи выпуклого программирования, синтезирующей методы условной оптимизации с процессами, базирующимися на методах безусловной оптимизации. Быстрота сходимости различных конкретизаций построенного общего метода определяется скоростью, с которой сходятся соответствующие методы безусловной оптимизации.

1. Пусть $z\in E^N$, $f_i(z)$, $i=0:M$,— выпуклые функции, а $\Omega=\{z\colon f_i(z)\leqslant 0,\ i=1:M\}$ — ограниченное множество. Рассмотрим задачу отыскания

$$z^*=\arg\min\{f_0(z)\,|\,z\in\Omega\}. \qquad (1)$$

Будем считать, что $\{i\colon f_i(z^*)=0\}=\{1,\ldots,m\}$, $m<N$. Рассмотрим систему нелинейных уравнений $f_i(z)=0$, $i=1:m$, или в векторной форме $f(z)\equiv f(x,\ y)=0$, $x\in E^n$, $y\in E^m$, $m+n=N$. Пусть $\partial f_0/\partial z=f'_{0z}=(f'_{0x};\ f'_{0y})=(f'_{0x_1},\ldots,f'_{0x_n};\ f'_{0y_1},\ldots,f'_{0y_m})$; $\partial f/\partial z=f'_z=(f'_x;\ f'_y)=(f'_{ix_1},\ldots,f'_{ix_n};\ f'_{iy_1},\ldots,f'_{iy_m})_{i=1}^m$; $f''_{0zz}=\partial f'_{0z}/\partial z$, причем $f'_z(z_0)$ означает, что элементы матрицы f'_z вычислены в точке z_0.

Пусть $\rho>0$ достаточно мало и $W(z^*,\rho)=\{z\colon |z_j-z_j^*|\leqslant\rho,\ j=1:N\}$.

Сформулируем некоторые условия, определяющие характер задачи (1).

A) $\det(f_y'(z^*))\neq 0$;

B) множество Ω удовлетворяет условию Слейтера;

C) функции $f_i(z)$, $i=0:m$, непрерывно дифференцируемы;

D) частные производные функции $f_i(z)$, $i=0:m$, удовлетворяют условию Липшица в прямоугольнике $W(z^*,\rho)$;

E) функции $f_i(z)$, $i=0:m$, дважды непрерывно дифференцируемы на $W(z^*,\rho)$;

F) вторые частные производные $f_i(z)$, $i=0:m$, удовлетворяют на $W(z^*,\rho)$ условию Липшица.

Если выполнено условие A) и на $W(z^*,\rho)$ для функций $f_i(z)$, $i=0:m$, выполнено условие C), тогда (см., например, [1]) существует такое $\rho'>0$, что в прямоугольнике $V(x^*,\rho')=\{x\colon |x_j-x_j^*|\leqslant\rho',\ j=1:n\}$ система $f(x,y)=0$ определяет $y_1=y_1(x),\ldots,\ y_m=y_m(x)$ как однозначные непрерывно дифференцируемые функции вектора x, причем $y^*=(y_1^*,\ldots,y_m^*)=(y_1(x^*),\ldots,y_m(x^*))$.

Характер функции $\varphi_0(x)=f_0(x,\ y(x))$, определенной на $V(x^*;\ \rho')$, устанавливает

Теорема 1. *Пусть выполнено условие* A), *тогда:*

1) *если имеют место условия* B) *и* E), *то матрица Гессе $H_0(x)$ функции $\varphi_0(x)$ непрерывна и $H_0(x^*)$ неотрицательно определена;*

2) *если в дополнение к условиям п. 1 функция $f_0(z)$ строго либо сильно выпукла, то существует такое $0<\delta\leqslant\rho'$, что в прямоугольнике $V(x^*;\ \delta)$ функция $\varphi_0(x)$ строго либо сильно выпукла, причем минимальные собственные значения μ_1 и μ_2 матриц $f_{0zz}''(z^*)$ и $\varphi_{0xx}''(x^*)$ связаны соотношением $\mu_1\leqslant\mu_2$;*

3) *если на $W(z^*;\rho)$ выполнено условие* D, *то на $V(x^*;\rho')$ вектор $\varphi_{0x}'=$*

$=f_{0x}'+f_{0y}'\cdot y_x'=f_{0x}'-f_{0y}'(f_y')^{-1}f_x'$ * *удовлетворяет условию Липшица, а при выполнении на* $W(z^*;\ \rho)$ *одного из условий* E) *либо* F *матрица Гессе* **

$$H_0(x)=\varphi''_{0xx}=f''_{0xx}+f''_{0xy}\cdot y_x'+(y_x')^T(f''_{0yx}+f''_{0yy}y_x')+$$

$$+\sum f'_{0y_l}y''_{lxx}=\overline{H}_0+\sum f'_{0y_l}y''_{lxx}$$

соответственно непрерывна либо удовлетворяет условию Липшица на $V(x^*;\ \delta)$.

2. Пусть $z^0=(x^0;\ y^0=y(x^0))\in W(z^*;\ \rho)$, а $Q(x^0)=\{x:\ \varphi_0(x)\leqslant\varphi_0(x^0)\}\subset\subset V(x^*;\ \delta)$.

Решение задачи (1) можно получить, отыскав $x^*=\arg\min\{\varphi_0(x)\,|\,x\in\in Q(x^0)\}$ и $y^*=y(x^*)$.

В свою очередь, для отыскания x^*, учитывая теорему 1, достаточно использовать сходящиеся релаксационные процессы безусловной оптимизации, взяв за исходное приближение произвольный вектор $x\in Q(x^0)$.

Оператор $\mathbf{R}$: $\Omega\to\Omega$ назовем оператором релаксации, если $f_0(\mathbf{R}z)\leqslant f_0(z)$. При наличии вектор-функции $y(x)$ на основе методов безусловной оптимизации (см., например, ([2-6])) подобно тому, как это сделано в ([7]), могут быть построены операторы релаксации со следующими свойствами:

1°) $\|\mathbf{R}z^k-z^*\|\leqslant q\|z^k-z^*\|,\quad q<1;$

2°) $\|\mathbf{R}z^k-z^*\|\leqslant q_k\|z^k-z^*\|,\quad q_k\to 0;$

3°) $\|\mathbf{R}z^k-z^*\|\leqslant q\|z^k-z^*\|^2.$

Наличие нелинейных ограничений делает иногда невозможным явное задание $y(x)$. Между тем на основе указанных операторов удается сконструировать операторы со свойствами 1°)–3°), для реализации которых не требуется явного задания $y(x)$.

Проиллюстрируем это на примере оператора сопряженных градиентов, построенного на основе метода ([5]), квадратичная сходимость которого установлена в ([8], [9]).

3. Пусть выбрано достаточно малое $\delta>0$ и такой вектор $z\in\Omega(\delta)=\{z:\ f_i(z)\leqslant\delta,\ i=1:M\}$, что множество $J(z;\ \delta)=\{i:\ |f_i(z)|\leqslant\delta\}=\{1,\dots,m\}$. Рассмотрим вектор-функцию $f(z;\ \delta)=f(x,\ y;\ \delta)=\{f_i(z):\ i\in J(z;\ \delta)\}$ и пусть $y=(y_1,\dots,y_m)$.

Построим матрицу $f_z'(z;\ \delta)=(f_x'(z;\ \delta);\ f_y'(z;\ \delta))$, строками которой являются векторы $f_{iz}'(z)=(f_{ix}'(z);\ f_{iy}'(z))$, $i\in J(z;\ \delta)$, и предположим, что $\Delta(z;\ \delta)=|\det(f_y'(z;\ \delta))|\geqslant\delta$. Пусть $\tilde z^0=(\tilde x^0,\ \tilde y^0)$, $\tilde y^0\in E^m$. Полагаем $y^0=\tilde y^0-(f_y'(\tilde z^0;\ \delta))^{-1}f(\tilde z^0;\ \delta)$, $z^0=(\tilde x^0,\ y^0)$.

Рассмотрим последовательность $\{z^s\}_{s=1}^n$, полученную следующим образом:

$$h^0=-g^0=-[f_{0x}'(z^0)-f_{0y}'(z^0)[f_y'(z^0;\ \delta)]^{-1}f_x'(z^0;\ \delta)],$$

$$z^{s+1}=(x^{s+1};\ y^{s+1}),\quad x^{s+1}=x^s+t_sh^s,\quad y^{s+1}=(y_1^{s+1},\dots,y_m^{s+1}),$$

$$y_l^s(t_s)=y_l^{s+1}=y_l^s+t_s\frac{\partial y_l^s}{\partial x}(h^s)^T+{}^1/_2t_s^2\frac{\partial^2y_l^s}{\partial x^2}(h^s)^T,\quad l=1:m,$$

$$\partial y^s/\partial x=(\partial y_l^s/\partial x)_{l=1}^m=y_x'(x^s;\ \delta)=-(f_y'(z^s;\ \delta))^{-1}f_x'(z^s;\ \delta),$$

$$h^{s+1}=-g^{s+1}+\|g^{s+1}\|^2\|g^s\|^{-2}h^s,$$

$$g^{s+1}=f_{0x}'(z^{s+1})+f_{0y}'(z^{s+1})y_x'(z^s;\ \delta)+t_s\left(\sum f'_{0y_l}(z^{s+1})\frac{\partial^2y_l^s}{\partial x^2}\right)(h^s)^T,$$

$$t_s=\arg\min\{\varphi_{0s}(t)\,|\,t\geqslant 0\},\quad \varphi_{0s}(t)=f_0(x^s+th^s;\ y_1^s(t),\dots,y_m^s(t)).$$

* Для упрощения записи в формулах для φ'_{0x} и φ''_{0xx} опущено указание о том, что элементы всех входящих в соответствующие выражения матриц вычислены в точке $z=(x,y(x))$.

** Здесь и в дальнейшем суммирование производится от 1 до m.

Матрицы $\partial^2 y_l^s/\partial x^2$, $l=1:m$, определяются из системы матричных уравнений

$$\sum f'_{iy_l}(z^s)\frac{\partial^2 y_l}{\partial x^2}=\overline{H}_i(z^s),\quad i=1:m,$$

где $\overline{H}_i(z^s)$ — матрица типа $\overline{H}_0$, соответствующая функции $f_i(z)$ и вычисленная в точке z^s, причем $\Delta(z^s;\ \delta)\geqslant\delta$, $0\leqslant s\leqslant n$. Оператор $\mathbf{R}$ сопряженных градиентов определим по формуле $\mathbf{R}\bar{z}^0=z^n=\bar{z}^1$.

Если для какого-либо $0\leqslant s\leqslant n$ окажется $\Delta(z^s;\ \delta)<\delta$, то будем говорить, что оператор $\mathbf{R}$ для приближения $z=z^0$ реализовать нельзя.

Возможность реализации оператора $\mathbf{R}$ и характер сходимости последовательности $\{\bar{z}^k=\mathbf{R}\bar{z}^{k-1}\}$ устанавливает

Теорема 2. *Пусть выполнены условия* A), B), F) *и функция* $f_0(z)$ *сильно выпукла; тогда существует такое* $\delta>0$, *что для любого вектора* $z\in W(z^*;\ \delta)\subset W(z^*;\ \rho)$ *оператор* $\mathbf{R}$ *реализуем и последовательность* $\{\bar{z}^k=\mathbf{R}\bar{z}^{k-1}\}_{k=1}^{\infty}$ *сходится к* z^* *квадратично, т. е.* $\mathbf{R}$ *обладает свойством* 3^0).

Локальный характер сходимости методов, основанных на использовании операторов типа $\mathbf{R}$, и трудности, связанные с априорным заданием параметра δ, удается преодолеть с помощью приводимого ниже общего метода, синтезирующего сходящиеся процессы выпуклого программирования с методами, использующими операторы типа $\mathbf{R}$.

4. На Ω определим неотрицательную функцию $\mu(z)$, удовлетворяющую условию

$$\mu(z)\geqslant\gamma\|z-z^*\|^\alpha,\quad \gamma>0,\quad \alpha>0. \tag{2}$$

Рассмотрим оператор $\mathbf{C}$: $\Omega\to\Omega$ и будем считать, что последовательность $\{z^k=\mathbf{C}z^{k-1}\}_{k=1}^{\infty}$ содержит сходящуюся к z^* подпоследовательность $\{z^{k_i}\}_{i=1}^{\infty}$, т. е. имеет место

$$\lim\mu(z^{k_i})=0. \tag{3}$$

При выполнении условий B) и C), согласно теореме Куна — Таккера, градиент $f_{0z}'(z^*)$ допускает разложение

$$f_{0z}'(z^*)=\sum u_i^* f_{iz}'(z^*),$$

причем для множителей Лагранжа $u^*=(u_1^*,\ldots,u_m^*)$ имеет место (см., например, ([10])) представление

$$u^*=[f_z'(z^*)(f_z'(z^*))^T)^{-1}f_z'(z^*)(f_{0z}'(z^*))^T\leqslant0. \tag{4}$$

Заметим, что существование $[f_z'(z^*)(f_z'(z^*))^T]^{-1}$ вытекает из условия A).

На множестве тех δ и $z\in\Omega(\delta)$, для которых существует $[f_z'(z;\ \delta)\times\times(f_z'(z;\ \delta))^T]^{-1}$, определим вектор-функцию $u(z;\ \delta)$, заменив в формуле (4) матрицу $f_z'(z^*)$ матрицей $f_z'(z;\ \delta)$, а вектор z^* — вектором z.

На этом же множестве определим вектор-функцию $\varphi_{0x}'(z;\ \delta)=f_{0x}'(z)-f_{0y}'(f_y'(z;\ \delta))^{-1}f_x'(z;\ \delta)$. Заметим, что при достаточно малом $\delta>0$ в силу условий A) и C) матрицы $[f_z'(z;\ \delta)(f_z'(z;\ \delta))^T]^{-1}$ и $[f_y'(z;\ \delta)]^{-1}$ существуют для всех $z\in W(z^*,\ \delta)\subset W(z^*;\ \rho)$. Обозначим $\psi(z)=\max\{f_i(z)\mid i=1:M\}$, $\bar{u}(z;\ \delta)=\max\{u_i(z;\ \delta)\mid,\ \ i\in J(z;\ \delta)\}$. Рассмотрим функцию $\nu=(z;\ \delta)=\max\{\|\varphi_{0x}'(z;\ \delta)\|,\ \bar{u}(z;\ \delta),\ \psi(z)\}$.

Утверждение. *Пусть выполнены условия* A)—C), *тогда*

1) *для того чтобы вектор* z *был решением задачи* (1), *необходимо и достаточно, чтобы*

$$\nu(z;\ 0)=\nu(z)=0,$$

2) *если наряду с условиями* A)—C) *имеет место условие* D, *то существует* $L>0$ *такое, что*

$$\nu(z;\ \delta)=\nu(z;\ \delta)-\nu(z^*;\ \delta)\leqslant L\|z-z^*\|\quad \forall(z;\ \delta),\ J(z;\ \delta)=\{1,\ldots,m\}.$$

Монотонно убывающую последовательность $\{\delta_k\}_{k=0}^{\infty}$ назовем управляющей, если $\lim\delta_k=0$. Будем говорить, что управляющая последовательность согласована с оператором $\mathbf{R}$, если для любой сходящейся к z^* последовательности $\{\tilde{z}^k=\mathbf{R}\tilde{z}^{k-1}\}_{k=1}^{\infty}$ имеет место $\lim \nu(z^k;\ \delta_k)\delta_k^{-1}=0$.

5. Выберем операторы $\mathbf{C}$, $\mathbf{R}$ и управляющую последовательность $\{\delta_i\}_{i=0}^{\infty}$. В качестве исходного приближения берем вектор $z^0\in\Omega$ и полагаем $i(0)=0$. Допустим, что уже проделано k шагов, найден вектор z^k и определено $i(k)$.

Проверяем неравенство $\mu(z^k)>\delta_{i(k)}$. Если это неравенство выполнено, то полагаем $z^{k+1}=\mathbf{C}z^k$, $i(k+1)=i(k)$ и переходим к очередному $k+1$ шагу.

Если $\mu(z^k)\leqslant\delta_{i(k)}$ и $\Delta(z^k;\ \delta_{i(k)})\leqslant\delta_{i(k)}$, то полагаем $z^{k+1}=\mathbf{C}z^k$, $i(k+1)=$ $=i(k)+1$ и переходим к очередному $k+1$ шагу.

В случае $\mu(z^k)\leqslant\delta_{i(k)}$, $\Delta(z^k;\ \delta_{i(k)})>\delta_{i(k)}$ полагаем $\tilde{z}^1=z^k$, $\tilde{\delta}_k=$ $=\min\{[\mu(\tilde{z}_1)]^\beta,\ \delta_i\}$, $\beta\leqslant(1-\varepsilon)\alpha^{-1}$, где $1>\varepsilon>0$ достаточно мало. Проверяем неравенство $\nu(\tilde{z}^1;\ \tilde{\delta}_1)\leqslant\delta_1$. Если это неравенство выполнено, то отыскиваем $\tilde{z}^2=\mathbf{R}\tilde{z}^1$, иначе полагаем $z^{k+1}=\mathbf{C}z^k$ и переходим к очередному $k+1$ шагу.

Пусть уже найдено $\tilde{z}^l$, тогда проверяем неравенство $\nu(\tilde{z}^l;\ \tilde{\delta}_l)\leqslant\delta_l$. Если это неравенство выполнено, то полагаем $\tilde{z}^{l+1}=\mathbf{R}\tilde{z}^l$, иначе полагаем $z^{k+1}=$ $=\mathbf{C}z^k$, $i(k+1)=i(k)+l+1$ и переходим к очередному $k+1$ шагу. Характер сходимости описанного общего метода устанавливаст

Т е о р е м а 3. *Пусть функция* $\mu(z)$ *и оператор* $\mathbf{C}$ *удовлетворяют соответственно условиям* (2) *и* (3), *а управляющая последовательность* $\{\delta_k\}_{=}^{k\infty}$ *согласована с оператором* $\mathbf{R}$, *тогда, начиная с некоторого момента, все приближения будут определяться лишь с помощью оператора* $\mathbf{R}$.

С л е д с т в и е. *Если оператор* $\mathbf{R}$ *удовлетворяет одному из свойств* $1^\circ-3^\circ$), *то последовательность приближений, порожденная описанным процессом, сходится к* z^* *соответственно линейно, сверхлинейно, квадратично.*

Различные операторы $\mathbf{C}$, обладающие свойством (3), и управляющие последовательности, согласованные с операторами, удовлетворяющими $1^\circ)-3^\circ$), указаны в ([7]). С операторами $\mathbf{C}$, обладающими свойством (3), связаны обычно функции $\mu(z)$, удовлетворяющие (2).

Украинский филиал Научно-исследовательского института планирования и нормативов
Киев

Поступило
13 II 1973

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] Г. М. Ф и х т е н г о л ь ц, Курс дифференциального и интегрального исчисления 1. «Наука», 1969. [2] Л. В. К а н т о р о в и ч, ДАН, 56, № 3 (1947). [3] Б. Т. П о л я к, Журн. вычислит. матем. и матем. физ., 3, № 14 (1963). [4] Ю. И. Л ю б и ч, Г. Д. М а й с т р о в с к и й, УМН, 25, в. 1 (1970). [5] R. F l e t c h e r, R. R e e v e s, Comput. J., 7, № 2 (1964). [6] H. H u a n g, J. Optim. Theory Appl., 5, № 6 (1970). [7] Л. А. К и р и е в с к и й, Р. А. П о л я к, ДАН, 209, № 3 (1973). [8] Г. Д. М а й с т р о в с к и й, Вычислит. матем. и вычислит. техн., Физ.-техн. инст. низких температур АН УССР, в. II, 1971. [9] С. А. С м о л я к, Тр. третьей зимней школы, в. III, 1970. [10] J. B. R o z e n, J. Soc. Industr. Appl. Math., 8, № 1 (1960).