第3卷 第3期
1983年7月

数学研究与评论
JOURNAL OF MATHEMATICAL
RESEARCH AND EXPOSITION

Vol.3 No.3
July, 1983

带有Polak程序的抽象算法模型*

堵丁柱 宋天泰

(中国科学院应用数学所)

§1 引 言

1969 年，Polak[1] 提出了一个对 Rosen 梯度投影法收敛性改进了的新算法。这个改进，关键的一点是引入了所谓 ε 程序，其实质就是使迭代方向的选取依赖于"$g(x)^Td>\max\limits_{j\in J}\{b_j-a_j^Tx\}$?"的判断(详见文内)。这一程序在 1979 年越民义、韓継业[2]首次解决既约梯度法收敛问题时起了作用。自此以后，含 Polak 程序的算法接踵而出[3~8]。在这些文章中，收敛性的证明各异且冗长。本文利用点到集映象的概念，建立了两个针对线性约束极值问题的抽象算法模型，并给出了一组统一的收敛条件。由于这组条件简单且比较易于验证，因此，利用它们会统一而且比较容易地得到[2～5]的收敛结果。此外，它们也将为新算法的设计和改进提供方便。

由于本文是在导师越民义教授和韩继业副教授指导下完成，并吸收了桂湘云副教授的宝贵意见，于此致谢。

§2 记 号

考虑问题：
$$\min_{j\in R} f(x) \tag{2.1}$$

其中 $R=\{x\,|\,a_j^Tx=b_j,\ j=1,\cdots,m;\ a_j^Tx\leqslant b_j,\ j=m+1,\cdots,m+l\}$,

$$x=(x_1,\cdots,x_n)^T\in \mathbb{E}^n,\ a_j\in \mathbb{E}^n,\ b_j\in \mathbb{E}^1,$$

$f(x)$是R上一阶连续可微函数。

记 $L_1=\{1,\cdots,m\}$, $L_2=\{m+1,\cdots,m+l\}$, $L=L_1\cup L_2$,

$J_\varepsilon(x)=\{j\,|\,b_j-a_j^Tx\leqslant\varepsilon, j\in L\}$,

$|J|$为J中元素个数，

$A=(a_1,\cdots,a_{m+l})$, $A_J=(a_j, j\in J)$,

$g(x)=-\nabla f(x)^T$, $\nabla f(x)$为$f(x)$的梯度行向量，

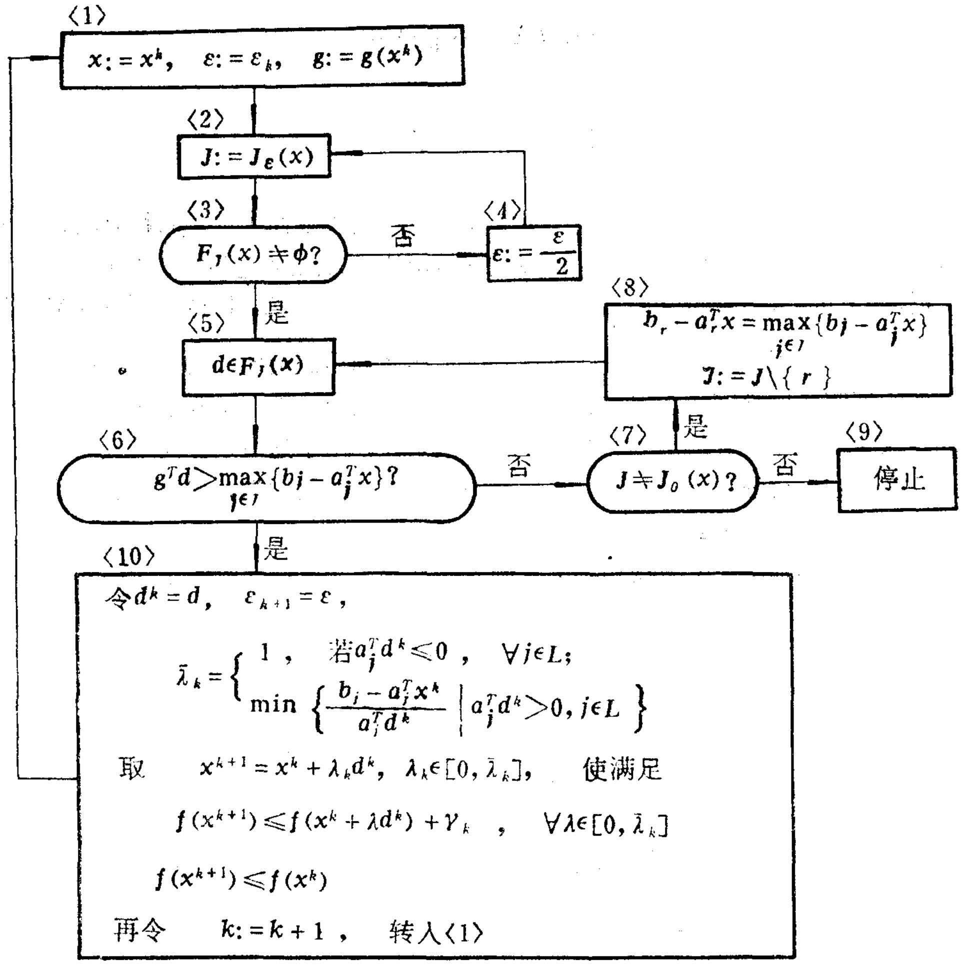
对$X\subseteq \mathbb{E}^n$, $\mathscr{P}(X)$是X之所有子集形成的集。

* 1981年9月18日收到。

§3 模 型

设$\{F_J|J\subseteq L\}$是一族从R到$\mathscr{P}(E^n)$的点到集映象，我们给出如下两个抽象算法。

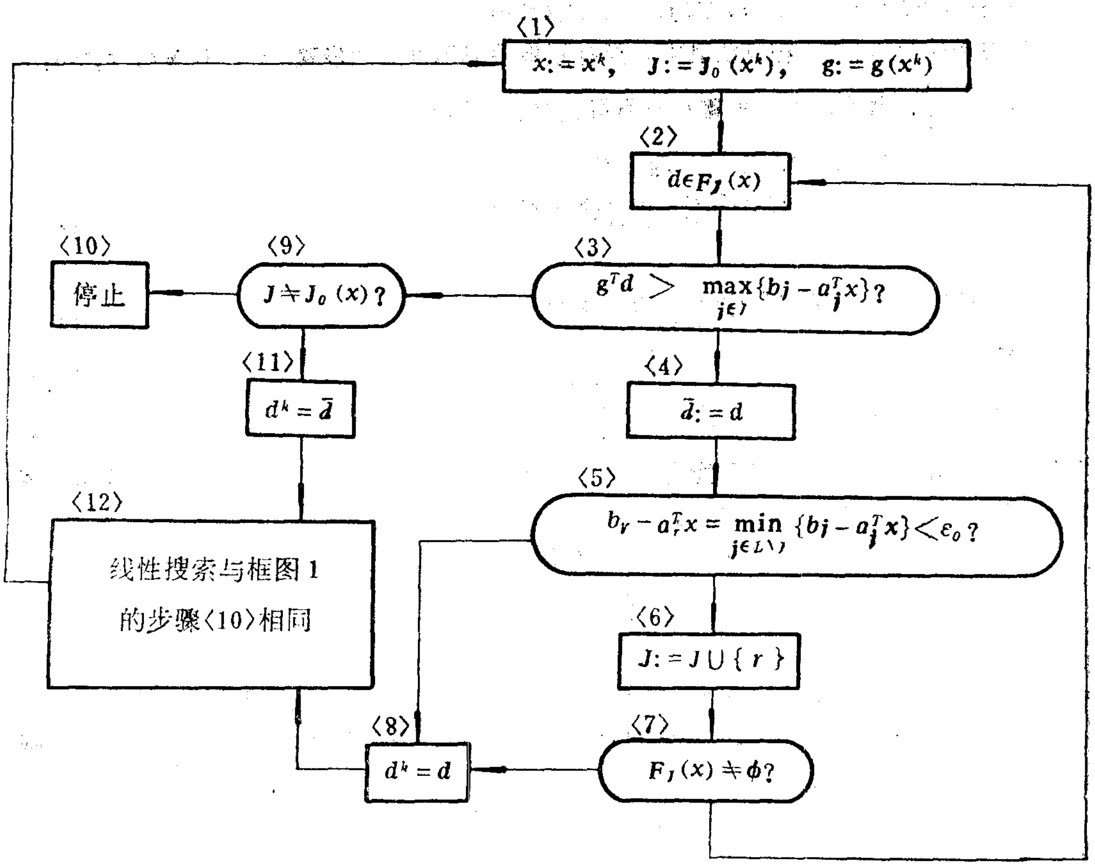
算法A 取$x^0\in R$，$\{\gamma_k\}$为一趋于零正数列，$\varepsilon_0>0$，令$k:=0$，代入框图1运算。



框图 1

算法B 取$x^0\in R$，$\{\gamma_k\}$为一趋于零正数列，$\varepsilon_0>0$，令$k:=0$，代入框图2运算。

注 算法B中引入ε_0是为减少运算次数，后面将会看到它对收敛性无影响。



框图 2

§4　收敛条件

定义　设X，$Y\in \mathbb{E}^n$. 点到集映象$F:X\to\mathscr{P}(Y)$称为在X的子集Ω上是闭的，指$\forall x\in\Omega$，$\forall\{x^k\}\subseteq X$，$\forall\{y^k\}\subseteq Y$，有

$$\left.\begin{array}{l} x^k\to x \\ y^k\to y \\ y^k\in F(x^k),\ \forall k \end{array}\right\}\Rightarrow y\in F(x)$$

定理　设$M\subset R$. 若$\{F_J\,|\,J\subseteq L\}$满足下面的四个条件，那么算法A和B将经有限步终止于M中点或者产生无限点列$\{x^k\}$，其极限点都在M中。

(1)　$\forall J\subseteq L$，F_J在$R\setminus M$上是闭的，且把R中紧集映入$\mathbb{E}$中一紧集之中。

(2)　$\exists\delta_1>0$使对$\forall J\subseteq L$有

$$d\in F_J(x)\Rightarrow\begin{cases} a_j^T d=0, & j\in J\cap L_1 \\ a_j^T d\leqslant 0, & j\in J\cap L_2\cap J_{\delta_1}(x) \end{cases}$$

(3)　$\forall x\in R\setminus M$，$\forall J\subseteq J_0(x)$：$d\in F_J(x)\Rightarrow g(x)^T d>0$.

(4) $\exists\delta_2>0:\forall x\in R,\ \forall J\subseteq J_{\delta_2}(x),\ F_J(x)\neq\emptyset$.

证 由条件(4)可知，在算法A中，经有限步后必不再到达步骤〈4〉，且$\forall k$，$\varepsilon_k\geqslant\delta_3=\min\left\{\varepsilon_0,\frac{\delta_2}{2}\right\}$. 因此，到达步骤〈10〉所得之$d^k$必将满足

a) $d^k\in F_{J^k}(x^k)$，$J^k\supseteq J_{\varepsilon_k}(x^k)\supseteq J_{\delta_3}(x^k)$

$$g(x^k)^Td^k>\max_{j\in J^k}\{b_j-a_j^Tx^k\}\tag{4.1}$$

或者

b) $d^k\in F_{J^k}(x^k)$，$$g(x^k)^Td^k>\max_{j\in T^k}\{b_j-a_j^Tx^k\}\tag{4.2}$$

并且$\exists\tilde d^k\in F_{J^k\cup\{r^k\}}(x^k):$ $$g(x^k)^T\tilde d^k\leqslant\min_{j\in L\setminus J^k}\{b_j-a_j^Tx^k\}\tag{4.3}$$

其中$$\max_{j\in J^k}\{b_j-a_j^Tx^k\}\leqslant b_{r_k}-a_{r_k}^Tx^k=\min_{j\in L\setminus J^k}\{b_j-a_j^Tx^k\}\tag{4.4}$$

容易推证，算法B在进入〈12〉时所得之d^k必也满足上述两条件之一。这样，无论算法A或B均可以此两条件为基础进行讨论。

若算法在点x^k处停止，必$\exists d\in F_{J_0(x^k)}(x^k)$使

$$g(x^k)^Td\leqslant\max_{j\in J_0(x^k)}\{b_j-a_j^Tx^k\}=0$$

由条件(3)知$x^k\in M$.

若算法运算无限次，易知循环只能发生在算法A中〈1〉和〈10〉之间或算法B的〈1〉和〈12〉之间，即产生一无限点列$\{x^k\}$。设x^*是$\{x^k\}$的一极限点，依据条件(1)和J^k、r^k的取法有限性，容易找到子列$\{x^k\}_{\mathscr{K}}\to x^*$且满足

$$\{d^k\}_{\mathscr{K}}\to d^*,\quad J^k=J,\quad \forall k\in\mathscr{K}$$

以及如下之①或②成立。

① $\forall k\in\mathscr{K}$，d^k满足a)；

② $\forall k\in\mathscr{K}$，d^k满足b)且有

$\{\tilde d^k\}_{\mathscr{K}}\to\tilde d$，$r^k=r$，$\forall k\in\mathscr{K}$.

我们下面分两种情形证明$x^*\in M$.

1° $\inf\limits_{k\in\mathscr{K}}\bar\lambda_k=\theta>0$.

首先注意算法是下降的，故对$\forall k$，选$k',k''\in\mathscr{K}$使$k'\leqslant k\leqslant k''$则有$f(x^{k'})\geqslant f(x^k)\geqslant f(x^{k''})$，令$k'$、$k$、$k''$同时趋于$\infty$，得$\lim\limits_{k\to\infty}f(x^k)=f(x^*)$.

其次，对$\forall k\in\mathscr{K}$，$\lambda\in(0,\theta)$，有

$$\begin{aligned}f(x^{k+1})-f(x^k)&\leqslant f(x^k+\lambda d^k)-f(x^k)+\gamma_k\\&=-\lambda g(x^k+\mu_k(\lambda)d^k)^Td^k+\gamma_k\end{aligned}\tag{4.5}$$

其中$0\leqslant\mu_k(\lambda)\leqslant\lambda$. 对固定的$\lambda$，我们选$\{\mu_k(\lambda)\}_{\mathscr{K}}$的收敛子列$\{\mu_k(\lambda)\}_{\mathscr{K}'}\to\mu(\lambda)$并在(4.5)中令$k\xrightarrow{\mathscr{K}'}\infty$，得

$$0=f(x^*)-f(x^*)\leqslant-\lambda g(x^*+\mu(\lambda)d^*)^Td^*$$

即 $$g(x^*+\mu(\lambda)d^*)^Td^*\leqslant0$$

其中$0\leqslant\mu(\lambda)\leqslant\lambda$. 再令$\lambda\to0$，得$g(x^*)^Td^*\leqslant0$. 于是在(4.1)(4.2)中令$k\xrightarrow{\mathscr{K}}\infty$，得

$$0 \geqslant g(x^*)^T d^* \geqslant \max_{j\in J}\{b_j - a_j^T x^*\} \tag{4.6}$$

由此知$J\subseteq J_0(x^*)$。

假若$x^*\notin M$，易知$x^*\in R\setminus M$。依F_J的闭性，得$d^*\in F_J(x^*)$.又依条件(3)，得$g(x^*)^T d^*>0$，这和(4.6)矛盾。因此$x^*\in M$。

2° $\inf\limits_{k\in\mathscr{K}}\bar\lambda_k=0$

不妨设$\{\bar\lambda_k\}_{\mathscr{K}}\to 0$，因否则可考虑其子列，同理也不妨设$\forall k\in\mathscr{K}$，有

$$\bar\lambda_k=\frac{b_{j_0}-a_{j_0}^T x^k}{a_{j_0}^T d^k},\quad a_{j_0}^T d^k>0. \tag{4.7}$$

故$b_{j_0}-a_{j_0}^T x^k\to 0\ (k\xrightarrow{\mathscr{K}}\infty)$。由此并且注意(4.7)和条件(2)可推出$j_0\notin J$且$\{x^k\}_{\mathscr{K}}$必须满足②。由(4.3)、(4.4)得

$$g(x^k)^T\tilde d^k\leqslant b_{j_0}-a_{j_0}^T x^k$$

$$\max_{j\in J\cup\{r\}}\{b_j-a_j^T x^k\}\leqslant b_{j_0}-a_{j_0}^T x^k$$

令$k\xrightarrow{\mathscr{K}}\infty$，得

$$g(x^*)^T d\leqslant 0=\max_{j\in J\cup\{r\}}\{b_j-a_j^T x^k\}$$

与情况1°最后一段同理可证$x^*\in M$。 ∎

§5　应用举例

本节中，我们令M为问题(2.1)的全部k-T点组成之集，举出一些满足条件(1～4)的$\{F_J|J\subseteq L\}$的例子，把它们代入§3中模型将形成[2～5]或新的算法。

例1（梯度投影法[4][5]）对$J\subseteq L$：$\mathrm{rank}A_J=|J|$，记　$P_J=E-A_J(A_J^TA_J)^{-1}A_J^T$（$E$为单位阵）

$$u=(u_j,j\in J)^T=-(A_J^TA_J)^{-1}A_J^Tg(x)$$

定义从R到$\mathscr{P}(E^n)$的点到集映象族$\{F_J|J\subseteq L\}$为：

a) 若$\mathrm{rank}A_J\neq|J|$，则$F_J(x)=\emptyset$，$\forall x\in R$。

b) 若$\mathrm{rank}A_J=|J|$，则

$$F_J(x)=\begin{cases}\{P_Jg(x)\}, & \text{若}\max\limits_{j\in J\cap L_2}u_j\leqslant 0,\\ \{P_{J\setminus\{s\}}g(x)\,|\,u_s=\max\limits_{j\in J\cap L_2}u_j\}, & \text{若}\max\limits_{j\in J\cap L_2}u_j>0.\end{cases}$$

在R满足正则性假设($\forall x\in R$：$\mathrm{rank}A_{J_0(x)}=|J_0(x)|$)时，可证明$\{F_J\subseteq L\}$满足条件(1～4)。把它代入算法A和B将分别得到章祥荪[4]和宋天泰[5]的收敛算法。

例2（可行方向法[3]）设$F_J(x)$为如下线性规划的最优解集：

(LP)　　$$\max_{d\in\Omega_J} g(x)^T d$$

其中　Ω_J：　$a_j^Td=0$，$j\in J\cap L_1$；　$a_j^Td\leqslant 0$，$j\in J\cap L_2$；　$|d_i|\leqslant 1$，$i=1,\cdots,n$。

可证$\{F_J|J\subseteq L\}$满足(1～4)，代入算法B得桂湘云、赖炎连[3]中算法。

例3（既约梯度法[2]）取$l=n$，$b_j-a_j^Tx=x_{j-m}$，$\forall j\in L_2$，即考虑如下形式之线性约束集：

$$R=\{x \mid Bx=b, x\geqslant 0\}$$

其中 $B=A_{L,1}^{T}$，$b=(b_1,\cdots,b_m)^T$。引入记号：

$I\subseteq\{1,2,\cdots,n\}$，$\bar{I}=\{1,2,\cdots,n\}\setminus I$。

$m+I=\{m+i \mid i\in I\}$。

B^I为列指标在 I 中之列组成的 B 之子矩阵。

若$(B^I)^{-1}$存在，称 I 为一组基。

$T(I)=(B^I)^{-1}B$。

$T_r(I)$，$r\in I$表示$T(I)$中列指标为 r 的列中等于1的元素所在的行。

$T_r^i(I)$表示$T_r(I)$中列指标为 i 的元素。

$T^{\bar{I}}(I)=(B^I)^{-1}B^{\bar{I}}$，$t(I)=(B^I)^{-1}b$。

$\nabla_I f(x)=\left(\dfrac{\partial f(x)}{\partial x_i}, i\in I\right)$，$\nabla_{\bar{I}}f(x)=\left(\dfrac{\partial f(x)}{\partial x_i}, i\in \bar{I}\right)$。

$x_I=(x_i, i\in I)^T$，$x_{\bar{I}}=(x_i, i\in\bar{I})^T$。

$\bar{f}(x_{\bar{I}})=f(t(I)-T^{\bar{I}}(I)x_{\bar{I}}, x_{\bar{I}})$。

$\nabla\bar{f}(x_{\bar{I}})=\nabla_{\bar{I}}f(x)-\nabla_I f(x)T^{\bar{I}}(I)$。

我们定义$\{F_J \mid J\subseteq L\}$为：$d\in F_J(x)$当且仅当有满足

$$x_r=\min_{i\in I}x_i\geqslant\frac{1}{2}x_s=\frac{1}{2}\max\{x_i \mid i\in\bar{I}, T_r^i(I)\neq 0\} \tag{5.1}$$

的基I，使

$$d=\begin{pmatrix}-T^{\bar{I}}(I)d_{\bar{I}}\\ d_{\bar{I}}\end{pmatrix}$$

其中$d_{\bar{I}}=(d_i, i\in\bar{I})^T$为：

$$d_i=\begin{cases}0, & \text{若}u_i\geqslant 0\text{且}m+i\in J,\\ -u_i, & \text{其它情况,}\end{cases}$$

这里 $(u_i, i\in\bar{I})=\nabla\bar{f}(x_{\bar{I}})$。

若R是非退化的，则上面定义之$\{F_J \mid J\subseteq L\}$满足条件(1～4)。将它代入算法$B$得越民义、韩继业[2]的结果，只要注意为选出满足(5.1)的基采用的转轴运算应是越民义教授早期采用的或在此基础上修改的[9]。

例4 (MoCormick 既约梯度法)这里利用[10]中方法为主体，利用§3模型构造一新算法。

设A_J^I表示A的行指标在 I 中，列指标在 J 中的子矩阵。对$x\in R$，$J\subseteq L$定义

a) 若 $\operatorname{rank}A_J\neq|J|$，则$F_J(x)=\emptyset$。

b) 若$\operatorname{rank}A_J=|J|$，则$d\in F_J(x)$当且仅当存在使$(A_J^I)^{-1}$存在的$I\subseteq\{1,\cdots,n\}$使

$$d=\left\{\begin{pmatrix}(A_J^I)^{-1}\\0\end{pmatrix}[\mathrm{diag}(\gamma)]\begin{pmatrix}(A_J^I)^{-1}\\0\end{pmatrix}^T+\begin{pmatrix}-(A_J^I)^{-1}A_J^{\bar I}\\E\end{pmatrix}\begin{pmatrix}-(A_J^I)^{-1}A_J^{\bar I}\\E\end{pmatrix}^T\right\}g(x)$$

其中对角方阵$\mathrm{diag}(\gamma)=\mathrm{diag}(\gamma_i, i\in J)$定义为

$$\gamma_j=\begin{cases}0, & 若 j\in L_1 或 j\in L_2 且 u_j\leqslant 0;\\ 1, & 若 j\in L_2 且 u_j>0,\end{cases}$$

这里 $(u_j, j\in J)=\nabla_I f(x)(A_J^I)^{-1}$.

当R为正则的时，上述$\{F_J|J\subseteq L\}$满足(1～4)代入算法A和B将得到新算法.

参 考 文 献

[1] Polak, E., On the Convergence of Optimization Algorithms, *Revue Française d'Informatique et de Recherche Operationnelle*, 3:16(1969), 17-34.

[2] 越民义、韩继业，一个新的既约梯度法及其收敛性，中国科学，4(1979) 345—356.

[3] 桂湘云、赖炎连，约束极值的一个可行方向法，数学学报，23:2(1980).

[4] 章祥荪，改进的Rosen-Polak方法，应用数学学报，2:3(1979) 257—267.

[5] 宋天泰，梯度投影法及其收敛性，中国科学院应用数学所硕士毕业论文.

[6] Kwei Hsiang-Yuin, Wu Fang and Lai Yan-Nian, Extension of a Variable Metric Algorithm to a Linearly Constrained Optimization Problem, K. B. Haley, ed., OR'78, North-Holland Publishing Company (1979), 955-974.

[7] 章祥荪，关于非线性约束条件下的Polak算法的一些讨论，应用数学学报，4:1(1981) 1—13.

[8] 赖炎连，非线性约束凸规划的一个解法及其收敛性，应用数学学报，3:4(1980) 322—331.

[9] 堵丁柱、孙捷、宋天泰，在既约梯度法中有限性转轴的简化，已投应用数学学报.

[10] MoComick, G. P., Anti-zigzagging by Bending, *Monagement Science*, Vol. 15, №.5 (January 1969).

Algorithm Models with Polak's Procedure

Du Dingzhu (堵丁柱) *Song Tiantai* (宋天泰)

(Institute of Applied Mathematics, Academia Sinica)

Abstract

In this paper, we give two algorithm models with Polak's procedure for linearly constrained optimization problem and a group of sufficiant conditions on their global convergence. Using them, it is easily seen that the convergence of algorithms in [2～5] can be proved.