维普资讯 http://www.cqvip.com

第32卷第4期 上海师范大学学报(自然科学版) Vol. 32, No. 4
2003年12月 Journal of Shanghai Normal University(Natural Sciences) 2003, Dec.

非单调线搜索解线性约束优化的信赖域内点算法

钱纯青

(上海师范大学 数理信息学院，上海 200234)

摘　要：提供了求解线性约束的非线性优化问题的非单调信赖域内点算法. 在合理的条件下，证明了算法的整体收敛性，并且在最优解局部范围内获得单位步长的可接受性，从而保证了局部超线性收敛速率.

关键词：信赖域；非单调线技术；内点法

中图分类号：O221.2　　**文献标识码**：A　　**文章编号**：1000-5137(2003)04-0023-05

本文研究线性约束下非线性优化问题

$$\text{(P)} \qquad \min\ f(x) \quad \text{s.t.} \quad Ax = b,\ x \geqslant 0,$$

其中 f 是 $\mathscr{R}^n$ 到 $\mathscr{R}$ 的光滑函数，但未必为凸函数；A 是 $p \times n$ 矩阵；$b \in \mathscr{R}^p$. 信赖域和线搜索是获得整体收敛性的两种最基本策略，信赖域算法解决非线性目标函数的早期工作可见 Dikin 和 Zorkalcev[3]，采用内点算法解决约束凸优化问题后来受到了广泛的关注，相关工作有 Den Hertog, Roos, Telaky[4]，McCormick[6] 等. Gonzaga[5] 研究了椭球形的信赖域，此算法产生的点列都在可行域的内部. 更进一步，Bonnans 和 Pola [2] 采用了此种椭球信赖域研究了线性约束优化问题的内点算法，得到了整体收敛性和局部单位 步长可接受性的良好结果. 但这些工作都是建立在每一步迭代目标函数单调下降的基础上. 最近 Grippo 等[1] 提出了非单调技术解决复杂度较高的无约束非线性优化问题. 非单调思想启发我们采用非单调线搜索技巧来研究线性约束优化问题的信赖域内点算法.

本文第二节首先描述与分析了非单调信赖域内点算法，然后在第三节论述了算法的整体收敛性，得到了算法的收敛点满足问题(P)的一阶优化条件，第四节则讨论了算法的局部收敛性质，在合理的条件下，得到了最优解局部范围内单位步长的可接受性.

1　算法

首先引进可行集合与严格内点可行集：

$$F = \{x \in \mathscr{R}^n; Ax = b, x \geqslant 0\}, \qquad \mathring{F} = \{x \in \mathscr{R}^n; Ax = b, x > 0\},$$

在本文中，将假设 F 有界，且 $\mathring{F}$ 非空，$g^k = g(x^k)$ 表示 $f(x)$ 在 x^k 处的梯度 $\nabla f(x^k)$.

在算法的每一次迭代中将用到两个矩阵，一个是 $X_k = \operatorname{diag}\{x_1^k, \cdots, x_n^k\}$，其中 x_i^k 是向量 x^k 的第 i 个分量，另一个是函数 $f(x)$ 的 Hessian 矩阵的对称近似阵 B_k. 假定 B_k 是半正定矩阵.

算法 1.1

步 1：给出 $x^0 \in \mathring{F}$, $\delta \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1)$，整数 $M \geqslant 0$；给定 $m(0) = 0$，$k \leftarrow 0$.

收稿日期：2002-02-08

作者简介：钱纯青(1965-)，男，上海师范大学数理信息学院讲师.

维普资讯 http://www.cqvip.com

步 2:取定一 $n\times n$ 对称矩阵 B_k. 计算 $\delta_k\in(\delta,1/\delta)$,使 d^k 满足

(SP) $\quad \min\ \varphi_k(d)=(g^k)^Td+\frac{1}{2}d^TB_kd$, s. t. $\quad Ad=0,\ d^TX_k^{-2}d\leqslant\delta_k^2$.

且满足 $x^k+d^k>0$,这里称 δ_k 是信赖域半径.

步 3: 若 $\varphi_k(0)=\varphi_k(d^k)$,则停止,即 x^k 为最优解.

步 4:使用线搜索,计算 $\rho^k=\beta^{l_k}$, l_k 是使下式成立的最小非负整数,

$$f(x^k+\beta^{l_k}d^k)\leqslant f(x^{l(k)})+\gamma\beta^{l_k}[(g^k)^Td^k+\frac{1}{2}(d^k)^TB_kd^k]. \tag{1}$$

其中 $f(x^{l(k)})=\max\limits_{0\leqslant j\leqslant m(k)}f(x^{k-j})$, 令 $x^{k+1}=x^k+\rho^kd^k$.

步 5:令 $m(k+1)\leqslant\min\{m(k)+1,M\}$;置 $k\leftarrow k+1$. 转步 2.

当 $M=0$ 时,非单调算法可看成通常的单调算法,从而非单调算法可视为通常单调算法的推广.

如果算法停止于第 k 步迭代,那么 x^k 满足(P)的一阶必要条件(见[7]).

引理 1.2 若 d^k 是(SP)的解,则必存在 $\lambda^{k+1}\in\mathscr{R}^p$, $v_k\geqslant0$,使得

$$g^k+B_kd^k+A^T\lambda^{k+1}+2v_kX_k^{-2}d^k=0,\quad Ad^k=0, \tag{2}$$

$$v_k\geqslant0,\ (d^k)^TX_k^{-2}d^k\leqslant\delta_k^2,\ v_k[(d^k)^TX_k^{-2}d^k-\delta_k^2]=0. \tag{3}$$

在算法 1.1 的步 3 中,若 $\varphi_k(0)=\varphi_k(d^k)$,则算法停止,此时

$$\varphi_k(0)-\varphi_k(d^k)=-(g^k)^Td^k-\frac{1}{2}(d^k)^TB_kd^k=(\lambda^{k+1})^TAd^k+2v_k(d^k)^TX_k^{-2}d^k+\frac{1}{2}(d^k)^TB_kd^k.$$

由(2)~(3)可得

$$\varphi_k(0)-\varphi_k(d^k)=2v_k\delta_k^2+\frac{1}{2}(d^k)^TB_kd^k=0. \tag{4}$$

由于 $\delta_k\in(\delta,1/\delta)$, B_k 为半正定矩阵,故有 $v_k\delta_k^2=0$, $(d^k)^TB_kd^k=0$. 因为 $(d^k)^TB_kd^k=(d^k)^TB_k^{\frac{1}{2}}B_k^{\frac{1}{2}}d^k=(B_k^{\frac{1}{2}}d^k)^TB_k^{\frac{1}{2}}d^k=0$. 这意味着 $B_k^{\frac{1}{2}}d^k=0$,又 $\delta_k>0$,因此 $v_k=0$, $B_kd^k=0$. 由此从(2)可见

$$g^k+A^T\lambda^{k+1}=0,\ Ax^k=b,\ x^k>0. \tag{5}$$

所以,终止步意味着 x^k 满足(P)的一阶必要性条件.

2 算法的整体收敛性

在本节中,将证明所提供算法的整体收敛性. 为了论述的方便,引入有效约束集合

$$I(x)=\{i\in\{1,\cdots,n\};x_i=0\};\ \forall x\in F.$$

对于任意 $I\subset\{1,\cdots,n\}$,其相应的优化问题是

$(P)_I \qquad \min\ f(x);\ Ax=b;\ x_I=0.$

它的一阶必要性优化条件为

$$(Q)_I\qquad\begin{cases}\nabla f(x)+A^T\lambda-\mu=0,\\ Ax=b,\\ x_I=0,\mu_i=0,\ i\in I\end{cases}$$

在本节中,将作如下假设

(A1): 对于任意 $I\subset\{1,\cdots,n\}$, $(Q)_I$ 没有非孤立解.

(A2): 存在 $\alpha>0$, $(d^k)^T(B_k+2v_kX_k^{-2})d^k\geqslant\alpha\|d^k\|^2$.

(A3): 对于任何解 $\bar{x}$, 以及相应的 Lagrange 乘子 $\bar{\lambda}$. 若 $(A^T\bar{\lambda})_i=0$, $\forall i\notin I(\bar{x})$, 则 $\bar{\lambda}=0$.

事实上, (A2)可重新表示为存在 $\alpha>0$,有

维普资讯 http://www.cqvip.com

$$\varphi_k(0)-\varphi_k(d^k)=-(g^k)^Td^k-\frac{1}{2}(d^k)^TB_kd^k=\frac{1}{2}(d^k)^T(B_k+2v_kX_k^{-2})d^k\geqslant\frac{\alpha}{2}\|d^k\|^2.$$

从(5)式中可直接推得. 若 B_k 一致正定，即 $\exists\ \alpha>0$; $(d^k)^TB_kd^k\geqslant\alpha\|d^k\|^2$. 从而(A2)可得到满足.

引理 2.1 设 $\{x^k\}$ 是由算法 2.1 产生的序列,则有 $\lim\limits_{k\to\infty}[\varphi_k(0)-\varphi_k(d^k)]=0$.

证明 据 $l(k)$ 的定义，$l(k)$ 是一整数且满足 $k-m(k)\leqslant l(k)\leqslant k$, $f(x^{l(k)})=\max\limits_{0\leqslant j\leqslant m(k)}[f(x^{k-j})]$. 考虑到 $m(k+1)\leqslant m(k)+1$, 有

$$f(x^{l(k+1)})=\max_{0\leqslant j\leqslant m(k+1)}[f(x^{k+1-j})]\leqslant\max_{0\leqslant j\leqslant m(k)+1}[f(x^{k+1-j})]=\max\{f(x^{l(k)}),\ f(x^{k+1})\}=f(x^{l(k)}).$$

这意味着序列 $\{f(x^{l(k)})\}$ 单调非增.

进一步,当 $k>M$,有(1)式知

$$\begin{aligned}f(x^{l(k)})&=f(x^{l(k)-1}+\rho^{l(k)-1}d^{l(k)-1})\\&\leqslant\max_{0\leqslant j\leqslant m(l(k)-1)}[f(x^{l(k)-1-j})]+\gamma\rho^{l(k)-1}[(g^{l(k)-1})^Td^{l(k)-1}+\frac{1}{2}(d^{l(k)-1})^TB_{l(k)-1}d^{l(k)-1}]\\&=f(x^{l(l(k)-1)})+\gamma\rho^{l(k)-1}[(g^{l(k)-1})^Td^{l(k)-1}+\frac{1}{2}(d^{l(k)-1})^TB_{l(k)-1}d^{l(k)-1}].\end{aligned}$$

故当 $k\to\infty$ 时,非增序列 $\{f(x^{l(k)})\}$ 必有极限,所以上式可得

$$\lim_{k\to\infty}\rho^{l(k)-1}[(g^{l(k)-1})^Td^{l(k)-1}+\frac{1}{2}(d^{l(k)-1})^TB_{l(k)-1}d^{l(k)-1}]=0.$$

由(A2)得 $\rho^{l(k)-1}[(g^{l(k)-1})^Td^{l(k)-1}+\frac{1}{2}(d^{l(k)-1})^TB_{l(k)-1}d^{l(k)-1}]\leqslant-\frac{\alpha}{2}\rho^{l(k)-1}\|d^{l(k)-1}\|^2\leqslant0$. 这意味着

$$\lim_{k\to\infty}\rho^{l(k)-1}\|d^{l(k)-1}\|=0.\tag{6}$$

令 $\hat{l}(k)\overset{\text{def}}{=}l(k+M+2)$. 现在通过数学归纳法证明,对任给的 $j\geqslant1$

$$\lim_{k\to\infty}\rho^{\hat{l}(k)-j}\|d^{\hat{l}(k)-j}\|=0.\tag{7}$$

和

$$\lim_{k\to\infty}f(x^{\hat{l}(k)-j})=\lim_{k\to\infty}f(x^{l(k)}).\tag{8}$$

不失一般性,下文中假设迭代指标 k 足够大,以避免负的下标的出现,既 $k\geqslant j-1$. 若 $j=1$,由于 $\{\hat{l}(k)\}\subset\{l(k)\}$,从(6)既可推得(7),从而 $\|x^{\hat{l}(k)}-x^{\hat{l}(k)-1}\|\to0$. 故当 $j=1$ 时,(8)式成立. 由于 $f(x)$ 在 F 上的一致连续性,现假定对任给的 j ,(7)和(8)式成立,由(1)式知

$$f(x^{\hat{l}(k)-j})\leqslant f(x^{l(\hat{l}(k)-j-1)})+\gamma\rho^{\hat{l}(k)-j-1}[(g^{\hat{l}(k)-j-1})^Td^{\hat{l}(k)-j-1}+\frac{1}{2}(d^{\hat{l}(k)-j-1})^TB_{\hat{l}(k)-j-1}d^{\hat{l}(k)-j-1}].$$

令 $k\to\infty$,由(8)式可知

$$\lim_{k\to\infty}\rho^{\hat{l}(k)-j-1}[(g^{\hat{l}(k)-j-1})^Td^{\hat{l}(k)-j-1}+\frac{1}{2}(d^{\hat{l}(k)-j-1})^TB_{\hat{l}(k)-j-1}d^{\hat{l}(k)-j-1}]=0.$$

类似于(6)的推导,可得

$$\lim_{k\to\infty}\rho^{\hat{l}(k)-j-1}\|d^{\hat{l}(k)-j-1}\|=0.$$

进一步,这意味着

$$\|x^{\hat{l}(k)-j-1}-x^{\hat{l}(k)-j-1}\|\to0.$$

故由 $f(x)$ 在 F 上的一致连续性,

$$\lim_{k\to\infty}f(x^{\hat{l}(k)-j-1})=\lim_{k\to\infty}f(x^{\hat{l}(k)-j-1})=\lim_{k\to\infty}f(x^{l(k)}).$$

从而证明了对 $j\geqslant1$, (7)和(8)式都成立.

对任意的 k ,有

$$x^{k+1}=x^{\hat{l}(k)}-\sum_{j=1}^{\hat{l}(k)-j-1}\rho^{\hat{l}(k)-j}d^{\hat{l}(k)-j}.\tag{9}$$

维普资讯 http://www.cqvip.com

因为 $\hat{l}(k)-k-1=l(k+M+2)-k-1\leqslant M+1$. 所以通过(7),(9)式可得 $\lim\limits_{k\to\infty}\|x^{k+1}-x^{\hat{l}(k)}\|=0$. 由于 $f(x^{l(k)})$ 的极限值唯一,由 f 的一致收敛性

$$\lim_{k\to\infty}f(x^k)=\lim_{k\to\infty}f(x^{\hat{l}(k)})=\lim_{k\to\infty}f(x^{l(k)}). \tag{10}$$

另一方面,当 ρ 足够小时,$f(x^k+\rho d^k)=f(x^k)+\rho(g^k)^{\mathrm{T}}d^k+o(\rho)$. 故必存在一常数 $c>0$,使得 $f(x^k+\rho d^k)\leqslant f(x^k)+\rho(g^k)^{\mathrm{T}}d^k+c\rho^2$. 由于 φ_k 是凸函数,有 $\varphi_k(d^k)-\varphi_k(0)\geqslant\nabla\varphi_k(0)^{\mathrm{T}}d^k=(g^k)^{\mathrm{T}}d^k$. 所以存在正常数 c

$$f(x^k+\rho d^k)\leqslant f(x^k)+\rho[\varphi_k(d^k)-\varphi_k(0)]+c\rho^2\leqslant f(x^{l(k)})+\rho[\varphi_k(d^k)-\varphi_k(0)]+c\rho^2.$$

显然当下式成立时,线搜索能够得到满足,$\rho\leqslant\hat{\rho}_k=\min\{1,\frac{1-\gamma}{c}[\varphi_k(d^k)-\varphi_k(0)]$. 这意味着 $\rho^k\geqslant\beta\hat{\rho}_k$,所以算法 1.1 中的线搜索满足

$$f(x^{k+1})\leqslant f(x^{l(k)})-\gamma\beta\frac{1-\gamma}{c}(\varphi_k(0)-\varphi_k(d^k))^2.$$

于是对(11)两边取极限,再利用(10)式,有 $\lim\limits_{k\to\infty}[\varphi_k(0)-\varphi_k(d^k)]=0$. □

引理 2.2 设 $\{x^k\}$ 是由算法 2.1 产生的序列,若 $\{B_k\}$ 有界,则 $\lim\limits_{k\to\infty}X_k[g^k+A^{\mathrm{T}}\lambda^{k+1}]=0$.

证明 由(10)和(11)式,可得

$$\varphi_k(0)-\varphi_k(d^k)=v_k\delta_k^2+\frac{1}{2}(d^k)^{\mathrm{T}}B_kd^k\to 0.$$

由于等式右边两个式子均不小于零,且 δ_k 有界,故此可得,当 $k\to\infty$ 时,$v_k\to 0$,$(d^k)^{\mathrm{T}}B_kd^k\to 0$,进而可得 $B_k^{\frac{1}{2}}d^k\to 0$. 余下部分内容的证明可参见[2]引理 2.3. □

定理 2.3 设 $\{x^k\}$ 是由算法 2.1 产生的序列,假设 $\{B_k\}$ 有界,则

(i) $\{x^k\}$ 的任一极限点都是 $(Q)_{I(\bar{x})}$ 的解;

(ii)若(A1)与(A2) 成立. 则 $\{x^k\}$ 收敛,若(A3)也成立,那么 $\bar{x}$ 满足(P)的一阶优化条件,即

$$(Q)_I\qquad\begin{cases}\nabla f(\bar{x})+A^{\mathrm{T}}\bar{\lambda}-\bar{\mu}=0, & A\bar{x}=b,\\ \bar{x}\geqslant 0,\bar{\mu}\geqslant 0,\bar{x}^{\mathrm{T}}\bar{\mu}=0.\end{cases}$$

证明 由(A2)及引理 2.2 可推得 $\lim\limits_{k\to\infty}\|d^k\|=0$,故以下的证明可参见[2] 定理 2.2. □

3 算法的局部收敛速率

为了使获得局部收敛性,在此不妨假设矩阵 A 行满秩,且算法 1.1 不是有限步终止,即算法 1.1 所得序列 $\{x^k\}$ 是一个无穷点列. 同时,假设下式成立.

$$\text{(A4)}\qquad\lim_{k\to\infty}\frac{\|P(B_k-\nabla^2f(x^k))d^k\|}{\|d^k\|}=0.$$

其中 $P=I-A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}A$ 为 A 的零空间的投影矩阵,事实上,(A4)是线性约束问题的超线性收敛的充分必要条件.

定理 3.1 假设(A1)~(A4)成立,$\{x^k\}$ 是算法 2.1 所得序列,若 x^k 充分接近于(P)的一个局部极小值点 $\bar{x}$ 时,有 $x^k\to\bar{x}$,且当 k 充分大时,算法中(1)式的 $\rho^k\equiv 1$.

证明 由定理 2.3 直接可得 $x^k\to\bar{x}$.

$$\begin{aligned}f(x^k+d^k)&=f(x^k)+(g^k)^{\mathrm{T}}d^k+\frac{1}{2}(d^k)^{\mathrm{T}}\nabla^2f(x^k)d^k+o(\|d^k\|^2)\\&\leqslant f(x^{l(k)})+\gamma[(g^k)^{\mathrm{T}}d^k+\frac{1}{2}(d^k)^{\mathrm{T}}B_kd^k]+(1-\gamma)[(g^k)^{\mathrm{T}}d^k+\frac{1}{2}(d^k)^{\mathrm{T}}B_kd^k]\end{aligned}$$

维普资讯 http://www.cqvip.com

$$+\frac{1}{2}(d^k)^T[\nabla^2 f(x^k)-B_k]d^k+o(\|d^k\|^2).$$

此时，由(A2)

$$(g^k)^T d^k+\frac{1}{2}(d^k)^T B_k d^k \leqslant -\frac{\alpha}{2}\|d^k\|^2. \tag{12}$$

否则，$(g^k)^T d^k+\frac{1}{2}(d^k)^T B_k d^k=0$. 则算法1.1在第$k$步终止.再由(A4)知,因为$Pd^k=d^k$,有

$$(d^k)^T B_k d^k=(d^k)^T \nabla^2 f(x^k) d^k+o(\|d^k\|^2). \tag{13}$$

从而当k足够大时,利用(12)~(13)

$$(1-\gamma)[(g^k)^T d^k+\frac{1}{2}(d^k)^T B_k d^k]+\frac{1}{2}(d^k)^T[\nabla^2 f(x^k)-B_k]d^k+o(\|d^k\|^2)<0.$$

所以，$f(x^k+d^k)\leqslant f(x^k)+\gamma[(g^k)^T d^k+\frac{1}{2}(d^k)^T B_k d^k]$. 这就意味着当$k$足够大时，$\rho^k\equiv 1$. □

定理4.1的成立意味着当k充分大时,步长$\rho^k=1$,那么收敛速率取决于d^k.若d^k是拟牛顿步,由假设条件(A4)，则算法(1.1)所得序列$\{x^k\}$超线性收敛于$\bar{x}$.

参考文献:

[1] GRIPPO L, LAMPARIELLO F, LUCIDI S. A Nonmonotone Line Search Technique for Newton's Methods[J]. SIAM J Numer Anal, 1986,23:707-716.

[2] BONNANS F J, POLA C. A Trust Region Interior Algorithm for Linearly Constrained Optimization[J]. SIAM J. Optim, 1997, 7(3):717-731.

[3] DINKIN I I, ZORKALCEV V I. Iterative Solution of Mathematical Programming Problems (Algorithms of Interior Point Methods)[M]. Nauka Publishers,Novosibirsk,1980(in Russian).

[4] HERTOG J E DEN, ROOS G, TERLAKY T. On the Classical Logarithmic Barrier Function Method for a Class of Smooth Convex Programming Problems. J Optim Theory Appl, 1992,73:1-25.

[5] GONZAGA C C. An Interior Trust Region Method for Linearly Constrained Optimization[J]. COAL BULL, 1991:55-65.

[6] MCCORMICK G. The Projective SUMT Method for Convex Programming Problems[J]. Math Oper Res, 1989, 14:203-223.

[7] SORENSEN D C. Newton's Method with a Model Trust Region Modification[J]. SIAM J Numer Anal,1982,19: 224-248.

A Nonmonotone Trust Region Interior Algorithm for Linearly Constrained Optimization

QIAN Chun-qing

(Mathematics and Sciences College, Shanghai Normal University, Shanghai 200234, China)

Abstract: We present a trust region interioralgorithm in association with nonmonotone line search technique for linearly constrained optimization. We prove that, under suitable hypotheses, the proposed algorithm converges globally, and analyze the unit stepsize will be asymptotically accepted.

Key words: trust region; nonmonotone line search; interior point