

非线性等式和线性不等式约束优化的 信赖域方法*

陈中文 韩继业

(中国科学院应用数学研究所, 北京 100080)

摘要 对非线性等式和线性不等式约束的优化问题提出一个新的信赖域算法, 在通常假设条件下, 证明了算法的全局收敛性. 此外, 由于通过引进松弛变量, 可把非线性不等式约束转化为一个方程的形式, 因此, 该算法可用于求解一般非线性规划问题.

关键词 非线性等式约束 线性不等式约束 信赖域方法 全局收敛性

多年来, 许多无约束和等式约束优化问题的信赖域方法已表明十分有效^[1,2], 简单界或凸约束优化问题的信赖域方法也有一些讨论^[3,4], 但对一般约束优化问题的信赖域方法, 讨论不多^{[5]1)}.

本文考虑下述问题:

$$\begin{aligned} \min & f(x), \\ \text{s.t. } & c(x)=0, \quad x \in X, \end{aligned} \quad (1)$$

其中 $X = \{x \in \mathbb{R}^n : A^T x \geq b\}$, $A = (a_1, a_2, \dots, a_p) \in \mathbb{R}^{n \times p}$, $b \in \mathbb{R}^p$, $c(x) = (h_1(x), h_2(x), \dots, h_m(x))^T$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 以及 $c: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是 $X' \supseteq X$ 上的连续可微函数.

在问题(1)的特殊结构中, 所有非线性约束都包含在系统 $c(x)=0$ 中. 由于任何一个非线性不等式约束都可通过引入松弛变量而化为一个方程, 因此, 问题(1)具有一般性. 本文提出一个新的信赖域方法求解(1)式, 并在通常假设条件下, 证明了算法的全局收敛性.

1 算法的描述

首先引入下述记号: $f_k = f(x_k)$, $g_k = g(x_k) = \nabla f(x_k)$ 等. $\|\cdot\|$ 表示 $\mathbb{R}^n$ 中 Euclid 范数.

$$\begin{aligned} I^0(x) &= \{i | a_i^T x - b_i = 0, 1 \leq i \leq p\}, \\ I^\varepsilon(x) &= \{i | 0 \leq a_i^T x - b_i \leq \varepsilon, \varepsilon > 0, 1 \leq i \leq p\}, \end{aligned}$$

1995-09-28 收稿, 1995-12-29 收修改稿

* 国家自然科学基金资助项目

1) Omojokun E O. Trust region algorithms for optimization with nonlinear equality and inequality constraints. Ph D Thesis. University of Colorado at Boulder, 1989

$$\begin{aligned} A^0(x) &= (a_i | i \in I^0(x)), \quad A^e(x) = (a_i | i \in I^e(x)), \\ b^e(x) &= (b_i - a_i^T x | i \in I^e(x)), \\ A^c(x) &= (\nabla h_1(x), \nabla h_2(x), \dots, \nabla h_m(x)), \\ G(x) &= (A^c(x), A^e(x)). \end{aligned}$$

假设已有 $x_k \in X$, 考虑

$$\begin{aligned} \min \quad & g_k^T d + \frac{1}{2} d^T B_k d, \\ \text{s.t.} \quad & c_k + A_k^e d = 0, \\ & A^T(x_k + d) \geq b, \quad \|d\| \leq \Delta_k, \end{aligned} \quad (2)$$

其中 $B_k \in \mathbb{R}^{n \times n}$ 为对称阵, $\Delta_k > 0$ 为信赖域半径. 一般情况下, (2) 式未必有解. 因此, 在 G_k 满秩的条件下, 计算

$$s_k = -G_k(G_k^T G_k)^{-1} \begin{pmatrix} c_k \\ 0 \end{pmatrix}, \quad \alpha_k = \min \left\{ 1, \frac{0.4\Delta_k}{\|s_k\|} \right\} \quad (3)$$

以及 Lagrange 乘子 $\lambda_k^T = g_k^T G_k(G_k^T G_k)^{-1} = ((\lambda_k^e)^T, (\lambda_k^c)^T)$. 再求解

$$\begin{aligned} \min \quad & \lambda_k^{cT} t + \frac{1}{2} t^T M_k t, \\ \text{s.t.} \quad & t \geq b_k^e, \quad \|t\|_\infty \leq 1, \end{aligned} \quad (4)$$

其中 M_k 为半正定矩阵. 设 t_k 为 (4) 式的解, 计算

$$t'_k = G_k(G_k^T G_k)^{-1} \begin{pmatrix} 0 \\ t_k \end{pmatrix}, \quad \beta_k = \min \left\{ 1, \frac{0.4\Delta_k}{\|t'_k\|} \right\}. \quad (5)$$

令 $v_k = \alpha_k s_k + \beta_k t'_k$, Z_k 为 G_k^T 的零空间的正交基. 再令 $d = v_k + Z_k u$, 并考虑

$$\begin{aligned} \min \quad & \bar{g}_k^T Z_k u + \frac{1}{2} u^T Z_k^T B_k Z_k u, \\ \text{s.t.} \quad & \|u\| \leq \bar{\Delta}_k, \end{aligned} \quad (6)$$

其中 $\bar{g}_k = g_k + \alpha_k B_k s_k$, $\bar{\Delta}_k = \sqrt{\Delta_k^2 - \|v_k\|^2}$. 设 u_k 为 (6) 式的解, 则由此得到尝试步 d_k . 考虑效益函数:

$$P(x, \sigma) = f(x) + \sigma \left(\sum_{i=1}^m h_i^2(x) + \sum_{i=1}^p \min^2(a_i^T x - b_i, 0) \right)^{\frac{1}{2}},$$

其中 $\sigma > 0$ 为罚参数. 如果 $x_k + d_k \in X$, 则计算 $P(x, \sigma)$ 的实际下降量 $\text{are } d_k = P(x_k + d_k, \sigma_k) - P(x_k, \sigma_k)$ 及预测下降量

$$\begin{aligned} \text{pred } d_k &= \bar{g}_k^T Z_k u_k + \frac{1}{2} u_k^T Z_k^T B_k Z_k u_k + \alpha_k g_k^T s_k + \frac{1}{2} \alpha_k^2 s_k^T B_k s_k + \\ &\quad \beta_k \lambda_k^{cT} t_k + \sigma_k (\|c_k + A_k^e d_k\| - \|c_k\|). \end{aligned}$$

由此计算它们的比值

$$r_k = \frac{\text{ared}_k}{\text{pred}_k},$$

根据 r_k 的大小确定尝试步 d_k 是否可接受.

算法 A

步 0 $\forall x_0 \in X, B_0 \in \mathbb{R}^{n \times n}$ 对称, $\sigma_0 > 0, \Delta_0 > 0, \varepsilon_0 > 0, \rho > 0, 0 < \eta_1 < \eta_2 < 1, k := 0$.

步 1 确定 I_k^e , 若 $\text{rank}(A_k^e) = |I_k^e|$, 转步 2; 否则, $\varepsilon_k := \frac{1}{2} \varepsilon_0$, 转步 1.

步 2 计算 λ_k, t_k , 若 $\|c_k\| + |\lambda_k^T t_k| + \|Z_k^T g_k\| = 0$, 则终止.

步 3 计算 d_k , 若 $x_k + d_k \notin X$, 则 $x_{k+1} = x_k, \Delta_{k+1} = \frac{1}{4} \Delta_k, \sigma_{k+1} = \sigma_k, B_{k+1} = B_k, k := k+1$, 转步 1.

步 4 计算 pred_k , 若

$$\text{pred}_k \leq -\frac{1}{2} \sigma_k (\|c_k\| - \|c_k + A_k^{eT} d_k\|) \quad (7)$$

不成立, 则令

$$\sigma_k := \max \left\{ \sigma_k, 2 \frac{\alpha_k g_k^T s_k + \frac{1}{2} \alpha_k^2 s_k^T B_k s_k}{\|c_k\| - \|c_k + A_k^{eT} d_k\|} + \rho \right\}.$$

步 5 计算 r_k , 若 $r_k \geq \eta_1$, 则 $x_{k+1} = x_k + d_k$, 且

$$\Delta_{k+1} = \begin{cases} \max\{\Delta_k, 4\|d_k\|\}, & \text{若 } r_k > \eta_2, \\ \Delta_k, & \text{若 } \eta_1 \leq r_k \leq \eta_2, \end{cases}$$

$\sigma_{k+1} = \sigma_k$, 校正 $B_{k+1}, k := k+1$, 转步 1. 若 $r_k < \eta_1$, 则 $x_{k+1} = x_k, \Delta_{k+1} = \frac{1}{4} \Delta_k, \sigma_{k+1} = \sigma_k, B_{k+1} = B_k$,

$k := k+1$, 转步 1.

注 算法 A 中符号 $|I_k^e|$ 表示集合 I_k^e 中指标个数, 此外, 问题 (4) 中矩阵 M_k 理论上可为任何半正定矩阵, 因此, 一般取 M_k 为单位阵进行计算.

2 收敛性分析

本文都在下述假设下进行讨论.

A1. (a) 存在有界闭凸集 $\Omega \subseteq X$ 使得 $x_k, x_k + d_k \in \Omega, \forall k$.

(b) $A^e(x)$ 对 $\forall x \in \Omega$ 都满秩.

(c) $\{B_k\}$ 一致有界.

A2. $\text{rank}(A^0(x)) = |I^0(x)|, \forall x \in X$ 且 $\text{rank}(G(x)) = m + |I^e(x)|$.

引理 1 设 γ_1 满足 $\|G_k(G_k^T G_k)^{-1}\| \leq \gamma_1$, 则

$$\|c_k\| - \|c_k + A_k^{eT} d_k\| \geq \min \left\{ \|c_k\|, \frac{0.4}{\gamma_1} \Delta_k \right\}.$$

引理 2 设 t_k 为 (4) 式的解, 则 $\lambda_k^T t_k \leq 0$.

定理 1 若 $\|c_k\| + |\lambda_k^T t_k| + \|Z_k^T g_k\| = 0$, 则 x_k 为 (1) 式的 K - T 点.

引理 3 罚参数 $\{\sigma_k\}$ 有界.

证 根据算法, 仅当 $c_k \neq 0$ 且 $\alpha_k g_k^T s_k + \frac{1}{2} \alpha_k^2 s_k^T B_k s_k > 0$ 时, 才有 $\sigma_{k+1} > \sigma_k$, 而当

$$\sigma_k \geq 2 \frac{\alpha_k g_k^T s_k + \frac{1}{2} \alpha_k^2 s_k^T B_k s_k}{\|c_k\| - \|c_k + A_k^{cT} d_k\|} + \rho$$

时, (7) 式必成立, 从而 σ_k 不必增大. 因此, 只需证明当 $c_k \neq 0$ 且 $\alpha_k g_k^T s_k + \frac{1}{2} \alpha_k^2 s_k^T B_k s_k > 0$ 时,

$$\bar{\sigma}_k = \frac{\alpha_k g_k^T s_k + \frac{1}{2} \alpha_k^2 s_k^T B_k s_k}{\|c_k\| - \|c_k + A_k^{cT} d_k\|}$$

是有界的. 由假设 A1, 即知 $\bar{\sigma}_k$ 是有界的.

在下面的讨论中, 总假设 σ_k 固定不变, 其次, 称算法 A 中步 1 缩小 ε_k 的迭代为内迭代, 则下述引理可由文献 [6] 中引理 1 得到.

引理 4 算法 A 中步 1 的每次内迭代必有限终止, 且存在 $\varepsilon_0 > 0$ 使得 $I_k^e \supseteq I_k^{\varepsilon_0}$, $\forall x \in X$, 都有 $\text{rank}(A^{\varepsilon_0}(x)) = |I^{\varepsilon_0}(x)|$.

引理 5 存在 $\bar{\Delta} > 0$ 使得 $A^T(x_k + d_k) \geq b$, $\forall \Delta_k \leq \bar{\Delta}$.

引理 6 存在 $\gamma_2 > 0$, $\gamma_3 > 0$ 使得

$$|\text{are } d_k - \text{pre } d_k| \leq \gamma_2 \|d_k\|^2 + \gamma_3 \Delta_k^2.$$

引理 7 存在 $\gamma_4 > 0$, 使得

$$\bar{g}_k^T Z_k u_k + \frac{1}{2} u_k^T Z_k^T B_k Z_k u_k \leq -\gamma_4 \|Z_k^T \bar{g}_k\| \min \left\{ \bar{\Delta}_k, \frac{\|Z_k^T \bar{g}_k\|}{\|Z_k^T B_k Z_k\|} \right\}.$$

下述引理表明: 若迭代点 x_k 不满足终止条件, 则至多经有限次内迭代和有限次缩小 Δ_k 后必能得一可接受的尝试步 d_k , 即有 $x_k + d_k \in X$ 且 $r_k \geq \eta_1$.

引理 8 算法 A 有定义.

引理 9 若存在 $\delta > 0$, 使得 $\|c_k\| + |\lambda_k^T t_k| + \|Z_k^T g_k\| \geq \delta$, 则存在 $\gamma_5, \gamma_6 > 0$, 使得对于满足 $\|c_k\| \leq \gamma_5 \Delta_k$ 的迭代有不等式 $\text{pred}_k \leq -\gamma_6 \Delta_k$ 成立.

证 由引理 7 及 pred_k 的定义得

$$\text{pred}_k \leq -\gamma_4 \|Z_k^T \bar{g}_k\| \min \left\{ 0.6 \Delta_k, \frac{\|Z_k^T \bar{g}_k\|}{\|Z_k^T B_k Z_k\|} \right\} + \left(\|g_k\| + \frac{1}{2} \|B_k\| \|s_k\| \right) \gamma_1 \|c_k\| + \beta_k \lambda_k^T t_k.$$

设 $\Delta_k \leq \Delta^*$, 则当 $\gamma_5 \leq \delta (3\Delta^*)^{-1}$ 时, $\|c_k\| \leq \frac{1}{3} \delta$, 因此,

$$|\lambda_k^T t_k| + \|Z_k^T g_k\| \geq \frac{2}{3} \delta,$$

即 $|\lambda_k^{IT} t_k| \geq \frac{1}{3} \delta$ 或 $\|Z_k^T g_k\| \geq \frac{1}{3} \delta$. 若 $|\lambda_k^{IT} t_k| \geq \frac{1}{3} \delta$, 则有 $\text{pred}_k \leq \gamma'_1 \gamma_5 \Delta_k - \frac{1}{3} \delta \beta_k$, 其中

$$\gamma'_1 = \gamma_1 \left(\sup \|g_k\| + \frac{\gamma_1}{2} \sup \|B_k\| \sup \|c_k\| \right).$$

再由 β_k 的定义, 当 $0 < \gamma_5 \leq \frac{\delta}{6\gamma'_1} \min \left\{ \frac{1}{\Delta^*}, \frac{0.4}{\sqrt{n} \gamma_1} \right\}$ 时,

$$\text{pred}_k \leq -\frac{\delta}{6} \min \left\{ \frac{1}{\Delta^*}, \frac{0.4}{\sqrt{n} \gamma_1} \right\} \Delta_k.$$

类似地, 当 $\|Z_k^T g_k\| \geq \frac{1}{3} \delta$ 时, 有

$$\text{pred}_k \leq -\frac{\delta \gamma_4}{12} \min \left\{ 0.6, \frac{\delta}{6\gamma'_2} \right\} \Delta_k,$$

其中 $\Delta_k \|Z_k^T B_k Z_k\| \leq \gamma'_2$, 且 γ_5 满足

$$0 < \gamma_5 \leq \min \left\{ \frac{\delta \gamma_4}{12\gamma'_1} \min \left\{ 0.6, \frac{\delta}{6\gamma'_2} \right\}, \frac{\delta}{6\gamma_1} (\Delta^* \sup \|Z_k\|^2 \sup \|B_k\|)^{-1} \right\}.$$

因此, 引理成立.

下述定理表明, 在假设 A1 和 A2 下, 或者算法 A 有限终止, 或者至少有一个聚点是问题 (1) 的 K-T 点.

定理 2 设 $\{x_k\}$ 是算法 A 产生的无限序列, 则

$$\liminf (|c_k| + |\lambda_k^{IT} t_k| + \|Z_k^T g_k\|) = 0.$$

证 假设存在 $\delta > 0$, 使得

$$\|c_k\| + |\lambda_k^{IT} t_k| + \|Z_k^T g_k\| \geq \delta, \forall k.$$

可以证明: $\lim_{k \rightarrow \infty} \Delta_k = 0$. 另一方面, 如果 $\|c_k\| \leq \gamma_5 \Delta_k$, 则由引理 9, $\text{pred}_k \leq -\gamma_6 \Delta_k$. 而当 $\|c_k\| > \gamma_5 \Delta_k$ 时, 由引理 1 得

$$\text{pred}_k \leq -\frac{\sigma}{2} \min \left\{ \gamma_5, \frac{0.4}{\gamma_1} \right\} \Delta_k.$$

令

$$\gamma_7 = \min \left\{ \gamma_6, \frac{\sigma}{2} \min \left\{ \gamma_5, \frac{0.4}{\gamma_1} \right\} \right\},$$

则总有 $\text{pred}_k \leq -\gamma_7 \Delta_k, \forall k$. 由引理 6 知

$$|r_k - 1| \leq \frac{(\gamma_2 + \gamma_3) \Delta_k^2}{\gamma_7 \Delta_k} \rightarrow 0 \quad (k \rightarrow +\infty),$$

因此, 存在 $\bar{\Delta}' > 0$, 使得 $r_k \geq \eta_1$, 对所有 $\Delta_k \leq \bar{\Delta}'$ 成立. 令 $\Delta_{\min} = \min \{\bar{\Delta}, \bar{\Delta}'\}$, 则当 $\Delta_k \leq \Delta_{\min}$ 时, $r_k \geq \eta_1$, 且 $x_k + d_k \in \Omega$. 根据算法, Δ_k 单调增加, 此与 $\lim \Delta_k = 0$ 矛盾. 因此, 定理得证.

3 注记

● 在本文中, 我们对非线性等式约束和线性不等式约束优化问题提出了一个新的信赖域方

法. 由于任何一个非线性不等式约束都可通过引入松弛变量而转化为一个方程. 因此, 本文所提的信赖域方法可用于求解一般非线性规划问题.

我们在上节证明了算法 A 的全局收敛性, 但由于使用不可微精确罚函数为效益函数, 因此, 就象直线搜索法一样, 也会产生 Maratos 效应. 避免这种现象发生的一种方法是增加二阶校正, 例如, 可采用 Fletcher^[7] 所建议的方法. 在本文所提出的算法中, 当 $r_k < \eta_1$, $x_k + d_k \in X$ 且 $\|s_k\| \leq \theta \Delta_k$, $\theta \in (0, 1)$ 时, 计算

$$w_k = G_k (G_k^T G_k)^{-1} \begin{pmatrix} c(x_k + d_k) \\ 0 \end{pmatrix}, \quad (8)$$

$$r'_k = \frac{\text{ared}_k(d_k + w_k)}{\text{pred}_k}.$$

如果有 $x_k + d_k + w_k \in X$, $r'_k \geq \eta_1$, 则 $x_{k+1} = x_k + d_k + w_k$, 且

$$\Delta_{k+1} = \begin{cases} \max\{\Delta_k, 4 \min\{\|d_k\|, \|d_k + w_k\|\}\}, & \text{若 } r'_k > \eta_2, \\ \frac{1}{4} \Delta_k, & \text{若 } \eta_1 \leq r'_k \leq \eta_2. \end{cases}$$

再转下一步, 否则, 直接缩小 Δ_k 并转下一步.

当算法 A 中增加二阶校正后, 上节中的全局收敛性结论仍成立. 进一步, 可以在适当条件下, 分析算法 A 的收敛速度.

参 考 文 献

- 1 Lalee M, Nocedal J, Plantenga T. On the implementation of an algorithm for large scale equality constrained optimization. Technical Report, Evanston: Northwestern University, 1994
- 2 Powell M J D. On the global convergence of trust region algorithms for unconstrained optimization. Math Prog, 1984, 29: 297~303
- 3 Friedlander A, Martinez J M, Santos S A. A new trust region algorithm for bound constrained minimization. Appl Math Optim, 1994, 30: 235~266
- 4 Burke J V, More J J, Toraldo G. Convergence properties of trust region method for linear and convex constraints. Math Prog, 1990, 47: 305~336
- 5 Yuan Y. On the convergence of a new trust region algorithm. Numerische Mathematik, 1995, 70(4): 515~539
- 6 Du D, Sun J. A new gradient projection method. Mathematica Numerica Sinica, 1983, 5: 378~386
- 7 Fletcher R. Second order corrections for nondifferentiable optimization. In: Watson G A. ed. Numerical Analysis Proceedings, Dundee 1982. Lecture Notes in Mathematics 912. Berlin-New York: Springer-Verlag, 1982