应用数学和力学编委会编
重 庆 出 版 社 出 版

文章编号: 1000_0887(2005) 11_1351_08

一类光滑凸规划的牛顿法

姚奕荣, 张连生, 韩伯顺

(上海大学 数学系, 上海 200444)

(戴世强推荐)

摘要: 给出了一个求解一类光滑凸规划的算法, 利用光滑精确乘子罚函数把一个光滑凸规划的极小化问题化为一个紧集上强凸函数的极小化问题, 然后在给定的紧集上用牛顿法对这个强凸函数进行极小化

关 键 词: 凸规划; 牛顿法; KKT 乘子
中图分类号: O221.2 文献标识码: A

1 凸规划和精确罚函数

考虑如下凸规划:

(P) $\min f(x)$

$\quad$ s.t. $x \quad S = \left\{ x \quad R^n : g_i(x) \quad 0,\ i = 1, \quad , m \right\}$,

假定 S 是紧集 于是存在一个大的有界箱子 X, 使得

$$S = \left\{ x \quad R^n : g_i(x) \quad 0,\ i = 1, \quad , m \right\} \quad \mathrm{int} X$$

设 $f(x)$, $g_i(x)$, $i = 1, \quad , m$ 是两次连续可微的凸函数, 且至少其中之一是强凸的

(P) 的光滑精确函数 $Q(x, \quad , p)$ 如下:

$$(Q_p) \quad \lim_{x \quad X} Q(x, \quad , p) = f(x) + \frac{1}{p} \sum_{i=1}^{m} \mathrm{e}^{\quad_i p g_i(x)},$$

其中 $p > 0$, $\quad_i \quad 0$, $i = 1, \quad , m$

我们有如下结论

定理 1.1 假定 $G(P) \quad \mathrm{int} X$, $G(P) + \quad B(\quad, 1) \quad X$ 对 $\quad > 0$, 若 $\quad_i > 0$, $i = 1, \quad , m$ 有限

$$\quad_i \quad \frac{\ln([M + f(x^*)]p + m)}{p g_{i_0}(x_0)}, \qquad i = 1, \quad , m;\ p \quad \frac{m}{\quad_1},$$

其中 $0 < \quad_1 \quad \min_{x \quad (S + \quad_1 B(\quad, 1)) \setminus (G(P) + \quad B(\quad, 1))} (f(x) - f(x^*))$, $x^* \quad G(P)$, $0 < \quad_1 < \quad$, $B(\quad, 1) = \left\{ x \quad R^n : \quad x \quad < 1 \right\}$, $0 < g_{i_0}(x_{i_0}) = \min_{x \quad X \setminus (S + \quad_1 B(\quad, 1))} \max_i (g_i(x))$, 且 $|f(x)| \quad M$, 对所有 $x \quad X$, 则 $Q(x^*, \quad, P) < Q(x, \quad, P)$, 对所有 $x \quad X \setminus (G(P) + \quad B(\quad, 1))$

收稿日期: 2004_02_01; 修订日期: 2005_07_12
基金项目: 国家自然科学基金资助项目(10271073)
作者简介: 姚奕荣(1959) , 男, 上海人, 副教授, 博士(联系人. Tel: + 86_21_66132412; E_mail: yryao@staff.shu.edu.cn)

证明 由于 $f(x) > f(x^*)$, 对所有 $x \quad S \setminus G(P), x^* \quad G(P)$, 则对任何 $\ > 0, f(x) > f(x^*)$, 对所有 $x \quad S \setminus (G(P) + \ B(\ ,1))$ 进一步, $S \setminus (G(P) + \ B(\ ,1))$ 是一个紧集, 且 $f(x)$ 是连续函数, 这样存在 $_1 > 0,\ _1 < \quad$ 和 $\ _1 > 0$, 使得对所有 $x \quad (S + \ _1 B(\ ,1)) \setminus (G(P) + \ B(\ ,1))$, 成立

$$f(x) \quad f(x^*) + \quad _1$$

这蕴含着

$$\min_{x \quad (S+ \ _1 B(\ ,1)) \setminus (G(P)+ \ B(\ ,1))} (f(x) - f(x^*)) \qquad _1 > 0,$$

此外, $\max_i (g_i(x))$ 是 x 的连续函数, 且 $X \setminus (S + \ _1 B(\ ,1))$ 是紧集, 这样

$$\min_{x \quad X \setminus (S+ \ _1 B(\ ,1))} \max_i (g_i(x)) \quad g_{i_0}(x_{i_0}) > 0,$$

现在, 分两种情况:

1) 对 $x \quad (S + \ _1 B(\ ,1)) \setminus (G(P) + \ B(\ ,1))$, 我们有

$$Q(x,\ ,p) = f(x) + \frac{1}{p} \ _{i=1}^{m} e^{\ _i p g_i(x)} > f(x) \quad f(x^*) + \quad _1 \quad f(x^*) + \frac{m}{p}$$

$$f(x^*) + \frac{1}{p} \ _{i=1}^{m} e^{\ _i p g_i(x^*)} = Q(x^*,\ ,p);$$

2) 对 $x \quad X \setminus (S + \ _1 B(\ ,1))$ 我们有 $\max_i (g_i(x)) = g_j(x) \quad g_{i_0}(x_{i_0}) > 0$, 且

$$Q(x,\ ,p) = f(x) + \frac{1}{p} \ _{i=1}^{m} e^{\ _i p g_i(x)} > - M + \frac{1}{p} e^{\ _j p g_j(x)}$$

$$- M + \frac{1}{p} e^{\ _j p g_{i_0}(x_{i_0})} \quad - M + \frac{1}{p}((M + f(x^*)p) + m) = f(x^*) + \frac{m}{p}$$

$$f(x^*) + \frac{1}{p} \ _{i=1}^{m} e^{\ _i p g_i(x^*)} = Q(x^*,\ ,p);$$

因此, 当 $p \quad m/\ \ _1 > 0,\ \ _i \quad (\ln((M + f(x^*)p) + m)/(p g_{i_0}(x_{i_0})) > 0, i = 1, \quad , m$ 时, 我们有 $Q(x^*,\ ,p) < Q(x,\ ,p)$, 对所有 $x \quad X \setminus (G(P) + \ B(\ ,1))$

注 根据定理 1.1, 对某个小的 $\ > 0$ 满足 $G(P) + \ B(\ ,1) \quad X$, 我们有 $G(P_p) \quad G(P) + \ B(\ ,1)$, 其中 $p > 0,\ \ _i > 0, i = 1, \quad , m$ 是由定理 1.1 确定, 这蕴含着, 对 $x^*_p \quad G(Q_p)$ 存在 $x^* \quad G(P)$ 使得 $\ x^*_p - x^*$

定理 1.2 若 $f(x), g_i(x), i = 1, \quad , m$ 是两次连续可微的凸函数, 且至少其中之一是强凸的, 则 $Q(x,\ ,p) = f(x) + \frac{1}{p} \ _{i=1}^{m} e^{\ _i p g_i(x)}$ 是强凸的

证明是显现然的

定理 1.3 若 $x^* \quad \mathrm{int}X$ 是一个(P) 的 KKT 点, 则 x^* 是 (Q_p) 的稳定点 其中 $\ _i = \ ^*_i,\ \ ^*_i \quad 0, i = 1, \quad , m$ 是在 x^* 处(P) 的 KKT 乘子

证明 从 x^* 是(P) 的 KKT 点, 我们有

$$f(x^*) + \ _{i \quad I(x^*)} \ ^*_i \quad g_i(x^*) = 0,$$

$$\ ^*_i \quad 0,\ i \quad I(x^*) = \left\{ i: g_i(x^*) = 0, i = 1, \quad , m \right\}; \qquad \ ^*_i = 0, i \ / \ I(x^*),$$

$$g_i(x^*) \quad 0, \quad \ ^*_i g_i(x^*) = 0, \qquad i = 1, \quad , m$$

进一步,

$$\nabla Q(x^*, \lambda, p) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* e^{\lambda_i^* p g_i(x^*)} \nabla g_i(x^*) =$$
$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) = 0, \tag{1}$$

于是 $x^* \in \mathrm{int}X$ 是 (Q_p) 的稳定点.

定理 1.4 若 $p > 0$ 充分大, $\lambda_i > 0$, $i = 1, \dots, m$ 有限, 满足定理 1.1 的条件, 且 $x_p^* \in \mathrm{int}X$ 是 (Q_p) 的稳定点, $x^* \in G(P)$, $\lambda_i^* \geq 0$, $i = 1, \dots, m$, 是 x^* 的 KKT 乘子, 则 x_p^* 是(P) 的近似 KKT 点, 且 $\lambda_{i_0}^* = \lambda_i e^{\lambda_i p g_i(x_p^*)} \approx \lambda_i^* > 0$, 对 $i \in I(x_p^*) = \left\{ i: g_i(x_p^*) \approx 0, i = 1, \dots, m \right\} = I(x^*)$; $\lambda_{i_0}^* = \lambda_i e^{\lambda_i p g_i(x_p^*)} \approx 0 = \lambda_i^*$, 对 $i \notin I(x_p^*) = I(x^*)$.

证明 根据定理 1.1, 成立 $f(x_p^*) \to f(x^*)$, $\nabla f(x_p^*) \to \nabla f(x^*)$, $g_i(x_p^*) \to g_i(x^*)$, $\nabla g_i(x_p^*) \to \nabla g_i(x^*)$. 因此, $I(x_p^*) = \left\{ i: g_i(x_p^*) \approx 0, i = 1, \dots, m \right\} = I(x^*)$. 从 $x_p^* \in \mathrm{int}X$ 是 (Q_p) 的稳定点, 于是我们有

$$0 = \nabla Q(x_p^*, \lambda, p) = \nabla \left(f(x_p^*) + \frac{1}{p} \sum_{i=1}^{m} e^{\lambda_i p g_i(x_p^*)} \right) =$$
$$\nabla f(x_p^*) + \sum_{i \in I(x^*)} \lambda_i e^{\lambda_i p g_i(x_p^*)} \nabla g_i(x_p^*) + \sum_{i \notin I(x^*)} \lambda_i e^{\lambda_i p g_i(x_p^*)} \nabla g_i(x_p^*),$$

对 $i \notin I(x_p^*) = I(x^*)$, $g(x_p^*) \to g(x^*) < 0$, 因此 $\lambda_{i_0}^* = \lambda_i e^{\lambda_i p g_i(x_p^*)} \to 0$ 且

$$\nabla f(x_p^*) + \sum_{i \in I(x^*)} \lambda_i e^{\lambda_i p g_i(x_p^*)} \nabla g_i(x_p^*) \approx 0, \tag{2}$$
$$\lambda_{i_0}^* g_i(x_p^*) \approx 0, \quad g_i(x_p^*) \to g_i(x^*), \quad i = 1, \dots, m,$$

进一步, $\nabla g_i(x_p^*) \to \nabla g(x^*)$, 这样

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i e^{\lambda_i p g_i(x_p^*)} \nabla g_i(x^*) \approx 0, \tag{3}$$

且

$$\lambda_{i_0}^* = \lambda_i e^{\lambda_i p g_i(x_p^*)} \approx \lambda_i^* > 0, \quad 对\ i \in I(x^*)$$

从(2), x_p^* 是(P) 的近似 KKT 点. 它蕴含着 x_p^* 是(P) 的近似全局极小点.

定理 1.5 若 $x^* \in \mathrm{int}X$ 是(P) 的严格互补 KKT 点, $\nabla g_i(x^*)$, $i \in I(x^*)$ 是线性独立, $|I(x^*)| = n$, $\lambda_i^* \geq 0$, $i = 1, \dots, m$ 是在 x^* 处(P) 的 KKT 乘子. $\lambda_i = \lambda_i^* + \mu_i$, $\lambda_i^* > 0$, $i \in I(x^*)$, $p > 0$; $\mu_i > 0$, $i \in I(x^*)$, 则适当选取 μ_i, $i \in I(x^*)$, x^* 是 (Q_p) 的全局极小点.

证明 从 x^* 是(P) 的 KKT 点, 成立

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) = 0, \tag{4}$$

进一步,

$$\nabla Q(x^*, \lambda, p) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i e^{\lambda_i p g_i(x^*)} \nabla g_i(x^*) =$$
$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) +$$
$$\sum_{i \in I(x^*)} \mu_i \nabla g_i(x^*) + \sum_{i \notin I(x^*)} \lambda_i e^{\lambda_i p g_i(x^*)} \nabla g_i(x^*), \tag{5}$$

若 $\nabla g_j(x^*) = 0$, 对所有 $j \notin I(x^*)$, 则选取 $\Delta\lambda_i = 0$, 对所有 $i \in I(x^*)$, 若 $\nabla g_j(x^*) \neq 0$, 对某些 $j \notin I(x^*)$, 则由 $\nabla g_i(x^*), i \in I(x^*)$, $|I(x^*)| + = n$, 是线性独立的, 于是存在 $\alpha_{ij} \in R^1, i \in I(x^*)$, 使得 $\nabla g_j(x^*) = \sum_{i \in I(x^*)} \alpha_{ij} \nabla g_i(x^*)$, 且

$$\nabla Q(x^*, \lambda, p) = \sum_{i \in I(x^*)} \Delta\lambda_i \nabla g_i(x^*) + \sum_{j \notin I(x^*)} \lambda_j e^{\lambda_j p g_j(x^*)} \nabla g_j(x^*) =$$

$$\sum_{i \in I(x^*)} \Delta\lambda_i \nabla g_i(x^*) + \sum_{i \in I(x^*)} \sum_{j \notin I(x^*)} \alpha_{ij} \lambda_j e^{\lambda_j p g_j(x^*)} \nabla g_i(x^*), \tag{6}$$

若令 $\Delta\lambda_i = -\sum_{j \notin I(x^*)} \alpha_{ij} \lambda_j e^{\lambda_j p g_j(x^*)}$, $i \in I(x^*)$, 则 $\nabla Q(x^*, \lambda, p) = 0$, 且 x^* 是 (Q_λ) 的全局极小点。

2 关于(P)的牛顿法

根据定理1.2, 问题(P)等价于问题 (Q_λ), 若 $\lambda_i \geq 0, i = 1, \ldots, m$ 被适当地选取, 于是我们能用牛顿法求解 (Q_λ) 来代替求解(P) [1,2]。

而且, $Q(x, \lambda, p)$ 是强凸函数, 存在 $l_H, L_H, 0 < l_H \leq L_H$ 使得

$$l_H \|d\|^2 \leq d^T \nabla^2 Q(x, \lambda, p) d \leq L_H \|d\|^2, \quad \text{对所有 } x \in X, d \neq 0。$$

牛顿法如下:

步骤1 给定初始点 $x_0 \in \operatorname{int} X$, $p > 0$ 充分大, $\lambda_i > 0, i = 1, \ldots, m$ 有限, 计算 $Q(x_0, \lambda, p)$, $\nabla Q(x_0, \lambda, p)$, $\nabla^2 Q(x_0, \lambda, p)$, $k := 0$, 若 $\|\nabla Q(x_k, \lambda, p)\| = 0$, 则停止, 否则, 转步骤2。

步骤2 确定牛顿方向 d_k 如下:

$$\nabla^2 Q(x_k, \lambda, p) d_k = -\nabla Q(x_k, \lambda, p), \tag{7}$$

且

$$Q(x_k + t_k d_k, \lambda, p) \leq Q(x_k, \lambda, p) - (l_H/2) t_k d_k^T d_k \leq Q(x_k, \lambda, p) - c_k t_k d_k^T d_k,$$
$$t_k \in (0, 2l_H/L_H], 0 < c_k = l_H - (L_H/2) t_k < l_H。 \tag{8}$$

步骤3 令 $x_{k+1} = x_k + t_k d_k$, $k := k + 1$, 转步骤2。

定理2.1 若 $f(x), g_i(x), i = 1, \ldots, m$ 是凸的, 且至少其中之一是强凸的, 则 $Q(x, \lambda, p)$ 是强凸的, $\{x_k\} \subset X$ 是由上述牛顿法产生, 令 $x_k \xrightarrow{k} x^* \in \operatorname{int} X$, x^* 是 X 上 $Q(x, \lambda, p)$ 的极小点。

证明 由定理1.2和定理1.1的注, $Q(x, \lambda, p)$ 是强凸的, $x^* \in \operatorname{int} X$。

进一步, 由泰勒展开, 存在一个 $\theta \in (0, 1)$ 使得

$$\begin{aligned}
Q(x_{k+1}, \lambda, p) &= Q(x_k + t_k d_k, \lambda, p) = \\
&Q(x_k, \lambda, p) + t_k d_k^T \nabla Q(x_k, \lambda, p) + (t_k^2/2) d_k^T \nabla^2 Q(x_k + \theta t_k d_k, \lambda, p) d_k = \\
&Q(x_k, \lambda, p) - t_k d_k^T \nabla^2 Q(x_k, \lambda, p) d_k + (t_k^2/2) d_k^T \nabla^2 Q(x_k + \theta t_k d_k, \lambda, p) d_k \leq \\
&Q(x_k, \lambda, p) - t_k l_H d_k^T d_k + (t_k^2/2) L_H d_k^T d_k = \\
&Q(x_k, \lambda, p) - t_k \left(l_H - (t_k/2) L_H \right) d_k^T d_k = \\
&Q(x_k, \lambda, p) - c_k t_k d_k^T d_k,
\end{aligned} \tag{9}$$

其中 $0 < t_k < 2l_H/L_H$, 则 $c_k = l_H - (L_H/2) t_k \in (0, l_H)$。

而且, $d_k = -(\nabla^2 Q(x_k, \lambda, p))^{-1} \nabla Q(x_k, \lambda, p)$, 于是(9)简化为

$$Q(x_{k+1}, \lambda, p) \leq$$

$$Q(x_k, K, p) - c_k t_k \nabla^{\mathrm T} Q(x_k, K, p)(\nabla^2 Q(x_k, K, p))^{-2} \nabla Q(x_k, K, p) \le$$
$$Q(x_k, K, p) - c_k t_k (1/L_{\mathrm H}^2) \nabla^{\mathrm T} Q(x_k, K, p) \nabla Q(x_k, K, p), \tag{10}$$

其中 $t_k \in (0, 2l_{\mathrm H}/L_{\mathrm H})$, $c_k = l_{\mathrm H} - (L_{\mathrm H} t_k/2) \in (0, l_{\mathrm H})$。因为

$$\sum_{k=1}^{\infty} (Q(x_k, K, p) - Q(x_{k+1}, K, p)) = Q(x_1, K, p) - Q(x^*, K, p) < +\infty,$$

从(10), 可得到

$$\frac{c_0 t_0}{L_{\mathrm H}^2} \sum_{k=1}^{\infty} \| \nabla Q(x_k, K, p) \|^2 \le \sum_{k=1}^{\infty} (Q(x_k, K, p) - Q(x_{k+1}, K, p)) < +\infty,$$

其中 $t_0 = t_k \in (0, 2l_{\mathrm H}/L_{\mathrm H})$, $c_0 = c_k = l_{\mathrm H} - (L_{\mathrm H} t_k/2) \in (0, l_{\mathrm H})$。这蕴含着 $\| \nabla Q(x_k, K, p) \| \xrightarrow{k} 0$。由 $x_k \xrightarrow{k} x^*$, 我们得到 $\nabla Q(x^*, K, p) = 0$, 且 x^* 是 X 上 $Q(x, K, p)$ 的极小点。

引理 2.1 (中值定理)[3]。

设 U 是 E 中开集且 $x \in U$, 令 $y \in E$, $f: U \to F$ 是 c^1 映射。假定线段 $x + yt$, $0 \le t \le 1$ 被包含在 U 内。则

$$f(x + y) - f(x) = \int_0^1 f'(x + ty) y \,\mathrm{d}t = \int_0^1 f'(x + ty) \,\mathrm{d}t y \le$$
$$| y | \sup_t | f'(x + ty) |,$$

其中 E, F 表示完备赋范向量空间, $f'(x)$ 表示 f 在 x 处的导数。

引理 2.2 若 A、B 是 $n \times n$ 矩阵, 则 AB 和 BA 的特征值是相同的[4]。

引理 2.3 若 A、B 是半正定矩阵, 则

$$K_i(A) K_n(B) \le K_i(AB) \le K_i(A) K_1(B),$$
$$K_n(A) K_i(B) \le K_i(AB) \le K_1(A) K_i(B), \qquad i = 1, \cdots, n,$$

其中 $K_n(\#) \le \cdots \le K_i(\#) \le \cdots \le K_1(\#)$, $K_i(\#)$ 表示矩阵 $(\#)$ 第 i 个特征值[5]。

定理 2.2 若 $\{x_k\} \subset X$ 是由牛顿法产生, 则成立

$$\frac{Q(x_{k+1}, K, p) - Q(x^*, K, p)}{Q(x_k, K, p) - Q(x^*, K, p)} \le 1 - \frac{l_{\mathrm H}^4}{L_{\mathrm H}^4}, \tag{11}$$

其中 x^* 是 $Q(x, K, p)$ 的一个极小点, $t_k = l_{\mathrm H}/L_{\mathrm H}$。

证明 由(7)、(9)和 $\nabla Q(x^*, K, p)$, 且根据引理 2.1 至引理 2.3, 我们有

$$\frac{Q(x_{k+1}, K, p) - Q(x^*, K, p)}{Q(x_k, K, p) - Q(x^*, K, p)} \le$$
$$1 - c_k t_k \frac{d_k^{\mathrm T} d_k}{Q(x_k, K, p) - Q(x^*, K, p)} \le$$
$$1 - c_k t_k \frac{\nabla^{\mathrm T} Q(x^*, K, p)(\nabla^2 Q(x_k, K, p))^{-2} \nabla Q(x_k, K, p)}{0.5(x_k - x^*)^{\mathrm T} \nabla^2 Q(x^* + R(x_k - x^*), K, p)(x_k - x^*)} \le$$
$$1 - 2 c_k t_k \frac{1}{L_{\mathrm H}^2} \frac{\nabla^{\mathrm T} Q(x_k, K, p) \nabla Q(x_k, K, p)}{L_{\mathrm H}(x_k - x^*)^{\mathrm T}(x_k - x^*)} =$$
$$1 - (2 c_k t_k / L_{\mathrm H}^3)[(\nabla^{\mathrm T} Q(x_k, K, p) - \nabla^{\mathrm T} Q(x^*, K, p))(\nabla Q(x_k, K, p) -$$
$$\nabla Q(x^*, K, p)] \big/ [(x_k - x^*)^{\mathrm T}(x_k - x^*)] =$$
$$1 - \Big[2 c_k t_k \int_0^1 (x_k - x^*)^{\mathrm T} \nabla^2 Q(x^* + t(x_k - x^*), K, p) \,\mathrm{d}t \int_0^1 \nabla^2 Q(x^* +$$

$$t(x_k - x^*), K, p)(x_k - x^*)\,\mathrm{d}t\Big] \Big/ [L_H^3 (x_k - x^*)^T (x_k - x^*)] =$$

$$1 - \frac{2c_k t_k}{L_H^3 (x_k - x^*)^T (x_k - x^*)} \int_0^1\int_0^1 (x_k - x^*)^T \nabla^2 Q(x^* + t(x_k - x^*), K, p) \cdot$$

$$\nabla^2 Q(x^* + tc(x_k - x^*), K, p)(x_k - x^*)\,\mathrm{d}t\,\mathrm{d}c \le$$

$$1 - \frac{2c_k t_k l_H^2}{L_H^3} \frac{(x_k - x^*)^T (x_k - x^*)}{(x_k - x^*)^T (x_k - x^*)} = 1 - \frac{l_H^4}{L_H^4},$$

其中 $R \in (0,1)$, $t_k = \frac{l_H}{L_H} \in \left(0, \frac{2l_H}{L_H}\right)$, $c_k = l_H - \frac{L_H t_k}{2} = \frac{l_H}{2} \in (0, l_H)$.

注 按照(11)，我们有

$$Q(x_{k+1}, K, p) - Q(x^*, K, p) \le (Q(x_0, K, p) - Q(x^*, K, p))\left(1 - (l_H^4/L_H^4)\right)^{k+1} \le \left(1 - (l_H^4/L_H^4)\right)^{L},$$

于是关于算法迭代数的界是 $O(L)$，L 表示问题的规模.

此外，求解系统(7)每次迭代的计算量是 $O(n^3)$. 这样算法的复杂性是 $O(n^3 L)$，其中 L 依赖于函数 $Q(x, K, p)$[6].

定理 2.3 若 $\{x_k\}$ 是由牛顿法产生，则成立

$$\|x_{k+1} - x^*\| \le \left(1 - (l_H^4/L_H^4)\right)^{1/2} \|x_k - x^*\|, \quad 对所有 k,$$

其中

$$x_{k+1} = x_k + t_k d_k,\ 0 < t_k = \frac{l_H^3}{L_H^3} < \frac{2l_H}{L_H},\ \nabla^2 Q(x_k, K, p) d_k = -\nabla Q(x_k, K, p). \quad (12)$$

证明 由(12)，我们有

$$x_{k+1} = x_k + t_k d_k, \quad \nabla^2 Q(x_k, K, p) d_k = -\nabla Q(x_k, K, p),$$

于是 $x_{k+1} - x^* = x_k - x^* + t_k d_k$，且

$$\|x_{k+1} - x^*\|^2 = \|x_k - x^* + t_k d_k\|^2 =$$

$$\|x_k - x^*\|^2 + 2t_k (x_k - x^*)^T d_k + t_k^2 \|d_k\|^2 =$$

$$\|x_k - x^*\|^2 - 2t_k (x_k - x^*)^T (\nabla^2 Q(x_k, K, p))^{-1} \nabla Q(x_k, K, p) +$$

$$t_k^2 \nabla^T Q(x_k, K, p) (\nabla^2 Q(x_k, K, p))^{-2} \nabla Q(x_k, K, p) =$$

$$\|x_k - x^*\|^2 - 2t_k (x_k - x^*)^T (\nabla^2 Q(x_k, K, p))^{-1} (\nabla Q(x_k, K, p) -$$

$$\nabla Q(x^*, K, p)) + t_k^2 (\nabla^T Q(x_k, K, p) -$$

$$\nabla^T Q(x^*, K, p))(\nabla^2 Q(x_k, K, p))^{-2} (\nabla Q(x_k, K, p) - \nabla Q(x^*, K, p)) \le$$

$$\|x_k - x^*\|^2 - 2t_k (x_k - x^*)^T (\nabla^2 Q(x_k, K, p))^{-1} \cdot$$

$$\int_0^1 \nabla^2 Q(x^* + t(x_k - x^*), K, p)(x_k - x^*)\,\mathrm{d}t +$$

$$\frac{t_k^2}{l_H^2} \int_0^1 (x_k - x^*)^T \nabla^2 Q(x^* + t(x_k - x^*), K, p)\,\mathrm{d}t \cdot$$

$$\int_0^1 \nabla^2 Q(x^* + t(x_k - x^*), K, p)(x_k - x^*)\,\mathrm{d}t =$$

$$\|x_k - x^*\|^2 - 2t_k \int_0^1 (x_k - x^*)^T (\nabla^2 Q(x_k, K, p))^{-1} \nabla^2 Q(x^* +$$

$$t(x_k - x^*), K, p)(x_k - x^*)\,\mathrm{d}t + \frac{t_k^2}{l_H^2} \int_0^1\int_0^1 (x_k - x^*)^T \nabla^2 Q(x^* +$$

$$t(x_k - x^*), \lambda, p)\nabla^2 Q(x^* + tc(x_k - x^*), \lambda, p)(x_k - x^*)\,\mathrm{d}t\,\mathrm{d}c \le$$

$$\|x_k - x^*\|^2 - 2t_k \frac{l_H}{L_H}\|x_k - x^*\|^2 + t_k^2 \frac{L_H^2}{l_H^2}\|x_k - x^*\|^2 =$$

$$\left(1 - 2t_k \frac{l_H}{L_H} + t_k^2 \frac{L_H^2}{l_H^2}\right)\|x_k - x^*\|^2 =$$

$$\left(1 - \frac{l_H^4}{L_H^4}\right)\|x_k - x^*\|^2,$$

其中, 令 $0 < t_k = l_H^3/L_H^3 < l_H/L_H$.

3 讨 论

若(P)是凸规划, 并且没有一个函数是强凸的, 则我们可增加一个大的球 $B_R = \{x \in R^n : \|x\|^2 \le R^2\}$, $R > 0$ 充分大, 使得 $S \subset \mathrm{int} B_R$. 这样(P)等价于(P)

(P) $\quad \min f(x)$

$\quad\quad$ s.t. $x \in S = \{x \in R^n : g_i(x) \le 0, i = 1, \dots, m, \|x\|^2 \le R^2\} = S;$

相应的光滑精确乘子罚函数 (Q_λ)

$$(Q_\lambda)\quad \min_{x \in X} Q(x, \lambda, p) = f(x) + \frac{1}{p}\left(\sum_{i=1}^{m} e^{\lambda_i p g_i(x)} + e^{\lambda_{m+1} p(\|x\|^2 - R^2)}\right),$$

且 $Q(x, \lambda, p)$ 是强凸的, 我们能利用上述牛顿法来求解.

这样, 对线性规划(LP), 和半正定二次规划(QP):

(LP) $\quad \min c^T x,$

$\quad\quad$ s.t. $x \in S = \{x \in R^n : a_i^T x - b_i \le 0, i = 1, \dots, m\},$

(QP) $\quad \min x^T Q x + a^T x,$

$\quad\quad$ s.t. $x \in S = \{x \in R^n : a_i^T x - b_i \le 0, i = 1, \dots, m\},$

(QQ_i) $\quad \min x^T Q x + a^T x,$

$\quad\quad$ s.t. $x \in S = \{x \in R^n : x^T Q_i x + a_i^T x + b_i \le 0, i = 1, \dots, m\},$

Q, Q_i 是半正定矩阵, 我们能利用上述方法来求解.

4 实 例

例

$$\min f(x) = x_1^2 + 2x_2^2 - 2x_1x_2 - 2x_1 - 6x_2,$$

$$\text{s.t. } x_1 + x_2 \le 2,$$

$$-x_1 + 2x_2 \le 2,$$

$$x_1 \ge 0,\ x_2 \ge 0,$$

$$g_1(x) = x_1 + x_2 - 2,$$

$$g_2(x) = -x_1 + 2x_2 - 2,$$

$$X = \{(x_1, x_2) \mid 0 \le x_i \le 2; i = 1, 2\}, \quad x \in X,$$

初始点: $x_0 = (0, 0.666\,66)$,

$$Q(x, \lambda, p) = f(x) + (1/p)(e^{\lambda_1 p g_1(x)} + e^{\lambda_2 p g_2(x)}) =$$

$$f(x) + (1/p)(e^{\lambda_1 p(x_1 + x_2 - 2)} + e^{\lambda_2 p(-x_1 + 2x_2 - 2)}).$$

表 1 概括 3 次迭代的计算

k	X_k	$Q(x_k, K, p)$	P	K_1	K_2
0	(0, 0.666 660)	4.714 055	20.000 000	4.000 000	4.000 000
1	(0.764 050, 1.176 028)	3.887 479	20.200 000	3.063709	3.500 684
2	(0.799 770, 1.199 841)	0.231 317	20.402 000	3.035 991	3.259 452
3	(0.799 210, 1.199 472)	0.006 740	)	)	)

k	d_k	t	$f(x)$	$Q(x, K, p)$
0	(50.000 000, 3.333 340)	0.152 810	− 3.111 089	− 3.111 089
1	(0.326 703, 0.217 808)	0.109 334	− 7.031 498	− 7.030 285
2	(− 0.000 536, − 0.000 353)	1.043 645	− 7.198 911	− 7.151 063
3	)	)	− 7.196 313	− 7.151 620

按照表 1, 在 3 次迭代末, 点 $x_3^* = (0.799 210, 1.199 472)$ 被达到了目标函数 $f(x) = -$ 71196 313 注意到最优点是(0.8, 1.2), 其目标函数值是 − 7.2[7]#

[参 考 文 献]

[1] Di Pillo G. Exact Penalty Methods[A]. In: Spedicate E Ed. Algorithm for Continuous Optimization: The State of the Art[C]. Boston: Kluwer Academic Publishers, 1994, 209) 253.

[2] Yu Nesterov, Nemirovsky A. Interior_Point Polynomial Methods in Convex Programming[M]. America: SIAM, Philadelphia, PA, 1994, 50) 150.

[3] Serge Lang. Real Analysis[M]. Reading, Mass: Addison_Wesley Publisher, 1983, 106.

[4] Wilkinson J H. The Algebraic Eigenvalue Problem[M]. Oxford: Clarendon Press, 1965.

[5] 王松桂, 贯忠贞. 矩阵论中不等式[M]. 合肥: 安徽教育出版社, 1994, 129.

[6] G ler O. Complexity of Smooth Convex Programming and its applications[A]. In: Complexity in Numerical Optimization[C]. by P M Pardalas Ed. World Scientifc Publishing Co, 1993, 180) 202.

[7] Bazaraa M S, Shetty C M. Nonlinear Programming_Theory and Algorithms[M]. New York: John Wiley & Sons, 1993, 504.

Newton Method for Solving a Class of Smooth Convex Programming

YAO Yi_rong, ZHANG Lian_sheng, HAN Bo_shun

(Department of Mathematics, Shanghai University, Shanghai 200444, P. R. China)

Abstract: An algorithm for solving a class of smooth convex programming is given. Using smooth exact multiplier penalty function, a smooth convex programming is minimized to a minimizing strongly convex function on the compact set was reduced. Then the strongly convex function with a Newtonon method on the given compact set was minimized.

Key words: convex programming; Newton method; KKT multiplier