

一类线性约束凸规划问题的一个原始-对偶内点算法

张 艺

(宁波大学 理学院, 浙江 宁波 315211)

摘要: 对一类具有线性约束的凸规划问题给出了一个原始-对偶内点算法, 该算法可在任一原始-对偶可行内点启动, 并且全局收敛. 当初始点靠近中心路径时, 便成为中心路径跟踪算法. 数值算例表明该算法是有效的.

关键词: 凸规则; 内点算法; 原始-对偶; 路径跟踪

中图分类号: O221.2

文献标志码: A

文章编号: 1001-5132 (2013) 02-0103-05

自从1984年求解线性规划的Karmarkar算法^[1]发表以来, 由于其理论上的多项式收敛性及实际计算的有效性, 人们对线性规划的内点算法作了大量研究, 取得了不少成果^[2], 其中的许多算法已推广应用到一些非线性规划问题^[4-8]. 笔者基于中心路径跟踪算法的思想, 对目标函数满足尺度Lipschitz条件的线性约束凸规划问题提出了一个原始-对偶内点算法, 并证明了算法的收敛性. 算法的主要特点是放宽了中心路径跟踪算法对初始点必须在中心路径附近的要求, 使算法可在任一原始-对偶可行内点启动, 如果初始点靠近中心路径, 则本算法便是中心路径跟踪算法.

1 问题描述

在叙述问题之前, 先给出一些记号, 本文记 $\|\cdot\|$, $\|\cdot\|_1$ 分别为向量(或矩阵)的2-范数和1-范数, $R_+^n = \{x | x > 0, x \in R^n\}$, $e = \{1, 1, \dots, 1\}^T \in R^n$, 当小写字母表示向量时, 对应的大写字母如没有给出明确的含义, 则表示以向量元素为对角线元素的对角矩阵, 如 $x = (x_1, x_2, \dots, x_n)^T \in R^n$, 则 $X = \text{diag}(x_1, x_2, \dots, x_n)$, 又如 E 是 n 阶单位矩阵.

考虑如下的线性约束凸规划问题:

$$\begin{aligned} & \min f(x), \\ \text{(P)} \quad & \text{s.t. } Ax = b, \\ & x \geq 0. \end{aligned}$$

其中 $A \in R^{m \times n}$, $b \in R^m$, $x \in R^n$, $f(x): R^n \rightarrow R$ 满足下列条件:

(i) $f(x)$ 二次连续可微, $\nabla^2 f(x)$ 半正定.

(ii) $f(x)$ 满足尺度 Lipschitz 条件, 即存在常数 $M > 0$, $0 < \beta < 1$, 使得:

$$\forall x \in R_+^n, \Delta x \in R^n, \text{ 当 } \|X^{-1}\Delta x\| \leq \beta \text{ 时, 有:}$$

$$\|X(\nabla f(x + \Delta x) - \nabla f(x) - \nabla^2 f(x)\Delta x)\| \leq$$

$$M(\Delta x)^T \nabla^2 f(x) \Delta x, \quad (1)$$

其中: $\nabla f(x)$, $\nabla^2 f(x)$ 分别表示 $f(x)$ 的梯度向量和 Hesse 矩阵.

满足上述条件的函数包含了一次函数, 二次函数及其他一些函数, 如熵函数 $f(x) = \sum_{i=1}^n x_i \ln x_i$ 及其正线性组合.

2 算法及其收敛性分析

问题(P)的 Lagrange 对偶问题:

$$\begin{aligned} & \max d(x, y) = b^T y + f(x) - (\nabla f(x))^T x, \\ \text{(D)} \quad & \text{s.t. } A^T y + s - \nabla f(x) = 0, \\ & s \geq 0. \end{aligned}$$

其中: $y \in R^m$, $s \in R^n$.

如果 $x, (x, y, s)$ 为问题(P)和(D)的可行解, 则对偶间隙为:

$$\begin{aligned} f(x) - d(x, y) &= f(x) - [b^T y + f(x) - (\nabla f(x))^T x] \\ &= (\nabla f(x))^T x - b^T y, \end{aligned}$$

记原始-对偶可行内点集为:

$$W = \{(x, y, s) | Ax = b, A^T y + s - \nabla f(x) = 0, x > 0, s > 0\}.$$

笔者作如下假定:

假设 1

- (1) A 为行满秩矩阵, 即 $R(A) = m$;
- (2) 原始-对偶内点集 W 非空有界.

由对偶理论知, $x, (x, y, s)$ 为问题(P)和问题(D)的解的充分必要条件是:

$$\begin{aligned} Ax &= b, \\ A^T y + s - \nabla f(x) &= 0, \\ Xs &= 0, \\ x, s &\geq 0. \end{aligned} \quad (2)$$

设 $u \in R_+^n$, 当 $\|u\|$ 充分小时, 方程组(PD(u)):

$$\begin{aligned} Ax &= b, \\ A^T y + s - \nabla f(x) &= 0, \\ s - X^{-1}u &= 0, \\ x, s &> 0, \end{aligned} \quad (3)$$

是方程组(2)的一个近似. 当 $\|u\| \rightarrow 0$ 时, 方程组(3)的解 $x(u)$ 和 $(x(u), y(u), s(u))$ 将分别收敛于(P)和(D)的最优解.

对当前迭代点 $w = (x, y, s) \in W$, 记

$$F(w) = \begin{pmatrix} Ax - b \\ A^T y + s - \nabla f(x) \\ s - X^{-1}u \end{pmatrix},$$

根据 Newton 方法, Newton 方向 $\Delta w = (\Delta x, \Delta y, \Delta s)$ 满足: $F'(w)\Delta w = -F(w)$, 即

$$A\Delta x = 0, \quad (4a)$$

$$-\nabla^2 f(x)\Delta x + A^T \Delta y + \Delta s = 0, \quad (4b)$$

$$X^{-2}U\Delta x + \Delta s = X^{-1}u - s, \quad (4c)$$

记 $P = \nabla^2 f(x) + X^{-2}U$, 通过计算得到:

$$\begin{aligned} \Delta y &= (AP^{-1}A^T)^{-1}AP^{-1}(s - X^{-1}u), \\ \Delta x &= P^{-1}A^T \Delta y - P^{-1}(s - X^{-1}u), \end{aligned} \quad (5)$$

$$\Delta s = X^{-1}u - X^{-2}U\Delta x - s.$$

我们取下一迭代点为 $w^+ = (x^+, y^+, s^+)$, 这里

$$\begin{aligned} x^+ &= x + \Delta x, \\ y^+ &= y + \Delta y, \\ s^+ &= s + \Delta s + s', \end{aligned} \quad (6)$$

其中: $s' = \nabla f(x^+) - \nabla f(x) - \nabla^2 f(x)\Delta x$. 可验证, 此时 $w^+ = (x^+, y^+, s^+)$ 满足(3)式中的前两式, 即

$$Ax^+ = b, A^T y^+ + s^+ - \nabla f(x^+) = 0.$$

为行文方便, 对 $u \in R_+^n, w = (x, y, s) \in W$, 再引入下面一些记号, 记

$$\theta = \min_{1 \leq i \leq n} \{u_i\}, \quad r = u - Xs,$$

$$\rho(x, s, u) = \|u - Xs\| / \theta = \|r\| / \theta, \quad d = X^{-1}\Delta x.$$

引理 1 设

$$A = \text{diag}(a_1, a_2, \dots, a_n) \in R^{n \times n},$$

$$a_i > 0 \quad (i = 1, 2, \dots, n), \quad B \in R^{n \times n},$$

为对称半正定矩阵, $C = A + B$, 则

$$\|C^{-1}v\| \leq \|v\| / \eta, \quad \forall v \in R^n,$$

其中: $\eta = \min_{1 \leq i \leq n} \{a_i\}$.

证明 由于

$$\begin{aligned} \|v\|^2 / \eta &\geq v^T A^{-1}v = (C^{-1}v)^T (CA^{-1}C)(C^{-1}v) = \\ &= (C^{-1}v)^T (A + 2B + BA^{-1}B)(C^{-1}v) \geq \\ &= (C^{-1}v)^T A(C^{-1}v) \geq \eta \|C^{-1}v\|^2, \end{aligned}$$

因此, 引理得证.

引理 2 设 $u \in R_+^n, w = (x, y, s) \in W$, Δx 由(5)式确定, 则

$$\|d\| = \|X^{-1}\Delta x\| \leq \rho(x, s, u). \quad (7)$$

证明 由(4a)和(4b)式得:

$$\begin{aligned} (\Delta x)^T \Delta s &= (\Delta x)^T \nabla^2 f(x) \Delta x = \\ &= (X^{-1}\Delta x)^T (X \nabla^2 f(x) X) (X^{-1}\Delta x) = \\ &= d^T (X \nabla^2 f(x) X) d. \end{aligned} \quad (8)$$

又由(4c)式知:

$$\begin{aligned} d^T r &= d^T (X^{-1}U\Delta x + X\Delta s) = \\ &= d^T X^{-1}U\Delta x + d^T X\Delta s = \\ &= d^T (U + X \nabla^2 f(x) X) d, \end{aligned} \quad (9)$$

记 $G = X \nabla^2 f(x) X + U$, 则 G 为正定矩阵, 由(9)式得:

$$d^T r = d^T G d = \|G^{1/2}d\|^2. \quad (10)$$

另一方面,

$$\begin{aligned} d^T r &= (G^{1/2}d)^T (G^{-1/2}r) \leq \|G^{1/2}d\| \cdot \|G^{-1/2}r\| \leq \\ &= \frac{1}{2}(\|G^{1/2}d\|^2 + \|G^{-1/2}r\|^2), \end{aligned} \quad (11)$$

由(10)式和(11)式得:

$$\begin{aligned} d^T G d &= \|G^{1/2}d\|^2 \leq \|G^{-1/2}r\|^2 = r^T G^{-1}r \leq \\ &= \|r\| \cdot \|G^{-1}r\|. \end{aligned}$$

由引理 1 得:

$$\theta \|d\|^2 \leq d^T G d \leq \|r\| \cdot \|G^{-1}r\| \leq \frac{1}{\theta} \|r\|^2, \quad (12)$$

从而有 $\|d\| \leq 1/\theta \|r\| = \rho(x, s, u)$.

引理证毕.

定理 1 设

$$u \in R_+^n, w = (x, y, s) \in W,$$

$$0 < \lambda \leq \min \left\{ \frac{1}{1+2M}, \beta \right\}, \rho(x, s, u) \leq \lambda,$$

则

$$w^+ = (x^+, y^+, s^+) \in W,$$

且 $\rho(x^+, s^+, u) \leq (1+2M)\lambda^2$.

证明 由(4c)式知:

$$\begin{aligned} u - X^+ s^+ &= u - (X + \Delta X)(s + \Delta s + s') = \\ &= u - (E + D)X(X^{-1}u - UX^{-2}\Delta x + s') = \\ &= u - (E + D)[(E - D)u + Xs'] = \\ &= D^2u - (E + D)Xs'. \end{aligned} \quad (13)$$

由引理 2 和定理条件知:

$$\|d\| = \|X^{-1}\Delta x\| \leq \rho(x, s, u) \leq \lambda \leq \beta < 1.$$

因此由(1)、(12)和(13)式可得:

$$\begin{aligned} \rho(x^+, s^+, u) &= \frac{\|D^2u - (E + D)Xs'\|}{\theta} \leq \\ &\leq \frac{1}{\theta} [\|D^2u\| + (\|E\| + \|D\|)\|Xs'\|] \leq \\ &\leq \frac{1}{\theta} [d^T U d + (1 + \|d\|)\|X(\nabla f(x + \Delta x) - \\ &\quad \nabla f(x) - \nabla^2 f(x)\Delta x)\|] \leq \\ &\leq \frac{1}{\theta} [d^T G d + (1 + \lambda)M\Delta x^T \nabla^2 f(x)\Delta x] = \\ &= \frac{1}{\theta} [d^T G d + (1 + \lambda)M d^T (X \nabla^2 f(x) X) d] \leq \\ &\leq \frac{1}{\theta} [1 + (1 + \lambda)M] d^T G d \leq \\ &\leq (1 + 2M)\rho^2(x, s, u) \leq (1 + 2M)\lambda^2. \end{aligned}$$

下证 $(x^+, y^+, z^+) \in W$.

由 $\|d\| \leq \rho(x, s, u) < 1$, 知 $x^+ = x + \Delta x = X(e + d) > 0$, 再由 θ 的定义及 $\rho(x^+, s^+, u) \leq (1+2M)\lambda^2 \leq \lambda < 1$, 即知 $s^+ > 0$, 因此, $w^+ \in W$.

引理证毕.

定理 2 在定理 1 的条件下, 取 α 满足:

$$\frac{\theta(1+2M)\lambda^2 + \|u\|}{\theta\lambda + \|u\|} \leq \alpha < 1,$$

则 $\rho(x^+, z^+, \alpha u) \leq \lambda < 1$.

证明 由定理 1 及 α 的选取知:

$$\rho(x^+, s^+, \alpha u) = \frac{\|X^+ s^+ - \alpha u\|}{\alpha\theta} \leq$$

$$\begin{aligned} &\frac{\|X^+ s^+ - u\|}{\alpha\theta} + \frac{(1-\alpha)\|u\|}{\alpha\theta} \leq \\ &\frac{(1+2M)\lambda^2}{\alpha} + \frac{(1-\alpha)\|u\|}{\alpha\theta} \leq \lambda < 1. \end{aligned}$$

定理证毕.

算法 1

Step 0 选取初始点

$$w^{(0)} = (x^{(0)}, y^{(0)}, s^{(0)}) \in W, u^{(0)} \in R_+^n,$$

$$0 < \lambda \leq \min \left\{ \frac{1}{1+2M}, \beta \right\},$$

并使 $\rho(x^{(0)}, s^{(0)}, u^{(0)}) \leq \lambda$, 允许误差:

$$\varepsilon > 0, \theta := \min_{1 \leq i \leq n} \{u_i^{(0)}\},$$

$$\alpha := \frac{\theta(1+2M)\lambda^2 + \|u^{(0)}\|}{\theta\lambda + \|u^{(0)}\|}, k := 0.$$

step 1 按(6)式计算新迭代点

$$w^{(k+1)} = (x^{(k+1)}, y^{(k+1)}, s^{(k+1)}).$$

step 2 如果 $(x^{(k+1)})^T s^{(k+1)} \leq \varepsilon$, 则终止计算, 输出问题(P)的近似解 $x^{(k+1)}$; 否则令 $u^{(k+1)} := \alpha u^{(k)}$, $k := k + 1$, 转 step 1.

定理 3 记 $\delta = \|u^{(0)}\|_1$, 则算法 1 的迭代次数不超过

$$\left\lceil \frac{\theta\lambda + \|u^{(0)}\|}{\theta\lambda[1 - (1+2M)\lambda]} \ln(2\delta/\varepsilon) \right\rceil,$$

其中: $\lceil \gamma \rceil$ 表示 γ 的最小整数上界.

证明 对任意的迭代指标 k , 由定理 2 知 $\rho(x^{(k)}, s^{(k)}, u^{(k)}) < 1$, $u^{(k)} = \alpha^k u^{(0)}$, 从而有 $X^{(k)} s^{(k)} < 2\alpha^k u^{(0)}$, $(x^{(k)})^T s^{(k)} < 2\alpha^k \delta$, 因此, 当 $k \geq \ln(\varepsilon/(2\delta))/\ln \alpha$ 时, 便可使算法终止.

由 $\ln(1-\gamma) \leq -\gamma, \forall \gamma \in (0, 1)$ 知:

$$\begin{aligned} \ln \alpha &= \ln \left[1 - \frac{\theta\lambda - (1+2M)\theta\lambda^2}{\theta\lambda + \|u^{(0)}\|} \right] \leq \\ &\leq -\frac{\theta\lambda[1 - (1+2M)\lambda]}{\theta\lambda + \|u^{(0)}\|}, \end{aligned}$$

所以

$$\begin{aligned} \ln(\varepsilon/(2\delta))/\ln \alpha &\leq \\ &\leq \frac{\lambda\theta + \|u^{(0)}\|}{\theta\lambda[1 - (1+2M)\lambda]} \ln(2\delta/\varepsilon). \end{aligned}$$

因此算法 1 的迭代次数不超过

$$\left\lceil \frac{\lambda\theta + \|u^{(0)}\|}{\theta\lambda[1 - (1+2M)\lambda]} \ln(2\delta/\varepsilon) \right\rceil.$$

定理证毕.

注 1 算法 1 实际上可在任一初始点 $w^{(0)} = (x^{(0)}, y^{(0)}, s^{(0)}) \in W$ 启动, 因为只要取

$$u^{(0)} = X^{(0)}s^{(0)}, \text{ 就有:}$$

$$\rho(x^{(0)}, s^{(0)}, u^{(0)}) = 0.$$

注 2 如果初始点 $w^{(0)}$ 靠近中心路径, 即 $u^{(0)} = \mu e$ 时, 算法 1 便成为中心路径跟踪算法^[2-3], 此时, 如果 $\mu = O(n)$, 则迭代界为 $O(\sqrt{n} \ln(n/\varepsilon))$.

由注 1 知道, 只要取

$$u^{(0)} = X^{(0)}s^{(0)}, \quad 0 < \lambda \leq \min\left\{\frac{1}{1+2M}, \beta\right\},$$

就可保证算法的收敛性, 但从收敛性的证明可看到, 如果 α 接近 1, 则可能造成收敛速度很慢. 因此, 需要对算法 1 作适当的修改: 当 α 接近 1 时, 对每一次迭代, 选取适当小的 α 进行试探, 如果下一迭代点不在 W 内, 则逐步扩大 α 直到下一迭点在 W 内为止. 具体算法如下:

算法 2

step 0 选取初始点

$$w^{(0)} = (x^{(0)}, y^{(0)}, s^{(0)}) \in W, u^{(0)} = X^{(0)}s^{(0)},$$

$$0 < \lambda \leq \min\left\{\frac{1}{1+2M}, \beta\right\},$$

允许误差

$$\varepsilon > 0, \theta := \min_{1 \leq i \leq n} \{u_i^{(0)}\},$$

$$\bar{\alpha} := \frac{\theta(1+2M)\lambda^2 + \|u^{(0)}\|}{\theta\lambda + \|u^{(0)}\|}, \quad k := 0, j := 2.$$

step 1 若 $\bar{\alpha} > 0.5$, 则 $\alpha := 0.5$, 否则 $\alpha := \bar{\alpha}$.

step 2 按(6)式计算新迭代点

$$w^{(k+1)} = (x^{(k+1)}, y^{(k+1)}, s^{(k+1)}).$$

step 3 如果 $w^{(k+1)} \in W$, 则转 step 4; 否则转 step 5.

step 4 如果 $(x^{(k+1)})^T s^{(k+1)} \leq \varepsilon$, 则终止计算, 输出问题(P)的近似解 $x^{(k+1)}$; 否则令 $u^{(k+1)} := \alpha u^{(k)}$, $k := k+1$, $j := 2$ 转 step 2.

step 5 $\eta := 1 - 1/2^j$, $u^{(k)} := \eta u^{(k-1)}$, $j := j+1$ 转 step 2.

3 数值算例

采用文献[8]中的算例, 即求解下列线性约束的凸二次规划问题:

$$\min f(x) = \frac{1}{2}x^T Qx + c^T x,$$

$$\text{s.t. } Ax = b, x \geq 0.$$

其中:

$$Q = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \end{pmatrix},$$

$$c = \begin{pmatrix} -8 \\ -8 \\ -8 \\ -8 \\ -8 \\ -8 \\ -8 \\ -8 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix},$$

$$A = \begin{pmatrix} 3 & 3 & 7 & 2 & 5 & 7 & 1 & 2 \\ 3 & 5 & 1 & 1 & 3 & 6 & 1 & 3 \\ 4 & 3 & 6 & 5 & 3 & 1 & 2 & 2 \\ 6 & 2 & 5 & 8 & 2 & 2 & 7 & 4 \\ 5 & 2 & 4 & 3 & 8 & 2 & 4 & 6 \\ 5 & 2 & 2 & 9 & 5 & 6 & 2 & 3 \end{pmatrix}, \quad b = \begin{pmatrix} 30 \\ 23 \\ 26 \\ 36 \\ 34 \\ 34 \end{pmatrix}.$$

用算法 2 在 Matlab 6.5 中进行了数值实验, 取 $\beta = 0.5$, $M = 1$, $\lambda = 0.25$, $\varepsilon = 10^{-6}$, 初始点与文献[8]相同, $x^{(0)} = s^{(0)} = e = (1, 1, 1, 1, 1, 1, 1, 1)^T$, $y^{(0)} = (0, 0, 0, 0, 0)^T$, 经 24 次迭代得近似最优解:

$$\bar{x} = (1.158358, 0.941372, 0.998811, 0.955113,$$

$$0.961895, 1.007498, 0.943895, 0.996524)^T,$$

最优目标函数数值为 -28.018267, 计算结果与调用 Matlab 6.5 中的函数 quadprog 所得结果一致.

值得指出的是, 文献[8]中 $b = (30, 27, 26, 36, 38, 34)^T$, 选取的 $x^{(0)} = (1, 1, 1, 1, 1, 1, 1, 1)^T$ 不满足 $Ax^{(0)} = b$, 经过 109 次迭代得到

$$x^* = (1.2334, 0.7485, 1.0546, 0.9608,$$

$$1.0151, 0.9954, 0.9416, 0.8829)^T,$$

也不满足 $Ax^* = b$ (或 $Ax^* \approx b$), 经验算

$$Ax^* = (30.002, 22.0661, 26.0004, \\ 36.0006, 33.9402, 34.0002)^T.$$

这说明文献[8]中的计算是错误的, 为保证 $Ax^{(0)} = b$, 笔者对 b 作了修改.

参考文献:

- [1] Karmarkar N. A new polynomial algorithm for linear program[J]. Combinatoric, 1984, 4:373-395.
- [2] 方述诚, 普森普拉 S. 线性优化及扩展理论与算法[M]. 北京: 科学出版社, 1994.
- [3] Monteiro R D C, Adler I. Interior path following primal-dual algorithms[J]. Mathemaical Programming, 1989, 44:27-41.
- [4] Monteiro R D C, Adler I. An extension of karmarkar type algorithm to a class of convex separable programming problems with global linear rate of convergence[J]. Mathematics of Operations Research, 1990, 15:408-422.
- [5] Jansen B, Roos C, Terlaky T, et al. Polynomiality of primal-dual affine scaling algorithm for nonlinear complementarity problems[J]. Mathematical Programming, 1997, 78:315-345.
- [6] Yu Q, Huang C C, Jiang Y. A polynomial predietor-corrector interior-point algorithm for convex quadratic programming[J]. Acta Mathematica Scientia, 2006, 26B(2):263-270.
- [7] Renegar J. A mathematical view of interior-poit method in convex optimization[M]. Philadelphia PA: Mathematical Programming Society, 2001.
- [8] 金正静, 白延琴, 韩伯顺. 求解凸二次规划问题的一种加权路径跟踪算法[J]. 运筹学学报, 2010, 14(1):55-65.

A Primal-dual Interior Point Algorithm for a Class of Convex Programming Problems with Linear Constraints

ZHANG Yi

(Faculty of Science, Ningbo University, Ningbo 315211, China)

Abstract: In this paper, a primal-dual interior point algorithm is presented for a class of convex programming problems with linear constrains. It can be implemented at any primal-dual interior feasible point and reaches the global convergence. If the initial point is close to the central path, it becomes a central path-following algorithm. The results of numerical tests show the effectiveness of the algorithm.

Key words: convex programming; interior point algorithm; primal-dual; path-following

(责任编辑 史小丽)