中国科学 A 辑: 数学 2009 年 第 39 卷 第 8 期: 977～995
www.scichina.com math.scichina.com
《中国科学》杂志社 SCIENCE IN CHINA PRESS

解非线性半定规划的过滤集-逐次线性化方法

李成进 ①②, 孙文瑜 ①*

① 南京师范大学数学与计算机科学学院, 南京 210097
② 福建师范大学数学与计算机科学学院, 福州 350108
E-mail: chengjin98298@163.com, wysun@njnu.edu.cn

收稿日期: 2007-12-12; 接受日期: 2009-03-17; * 通信作者
国家自然科学基金 (批准号: 10871098) 和江苏省自然科学基金 (批准号: BK2006214) 资助项目

摘要 本文提出了解非线性半定规划的信赖域型过滤集-逐次线性化方法, 该方法基于 Fletcher 和 Leyffer 2002 年提出的解非线性规划的过滤集的概念. 本文给出了新的算法, 并在较弱的条件下证明了算法的总体收敛性. 最后, 我们报告了新方法的数值结果, 表明新方法是有效的.

关键词 半定规划 非线性规划 逐次线性化方法 过滤集方法 总体收敛性

MSC(2000) 主题分类 **65K05, 90C30**

1 引言

本文考虑解非线性半定规划 (或非线性 SDP) 问题:

$$\begin{aligned} \min_{X\in S^n}\ & f(X) \\ \text{s.t.}\ & g(X)\leqslant 0, \\ & X\succeq 0, \end{aligned} \tag{1.1}$$

其中, $f: S^n\to\mathbb{R}$, $g: S^n\to\mathbb{R}^m$ 是二次连续可微函数, S^n 表示 $\mathbb{R}^{n\times n}$ 中所有对称矩阵的子集, $X\succ 0$ ($\succeq 0$) 表示对称正定 (半正定) 矩阵.

非线性半定规划问题 (1.1) 是标准线性半定规划问题的推广, 而线性半定规划在过去十年中得到了广泛的研究和很大的发展, 例如见文献 [1–4]. 但是, 非线性半定规划的研究刚刚起步, 仍然处于一个初期阶段. 最近, 若干论文 (如文献 [5–13]) 从理论上和数值上研究了解非线性半定规划的不同的算法, 例如, 逐次线性化方法 [7], 序列半定规划方法 [5, 12], 势下降算法 [13], 光滑牛顿法 [9], 内点法 [6], 原始-对偶内点法 [14, 15], 非光滑牛顿型方法 [11], 增广 Lagrange 方法 [8], 等等. 文献 [10, 16–20] 对非线性半定规划的最优性条件和灵敏度分析进行了研究. 上述方法的一个共同特点是利用罚参数 (或评价函数) 作为接受准则和得到算法的总体收敛性.

引用格式: 李成进, 孙文瑜. 解非线性半定规划的过滤集-逐次线性化方法. 中国科学 A, 2009, 39(8): 977–995
Li C J, Sun W Y. On filter-successive linearization methods for nonlinear semidefinite programming. Sci China Ser A, 2009, 52, DOI: 10.1007/s11425-009-0168-6

最近, Fletcher 和 Leyffer 在他们的开创性工作 [21] 中提出了解非线性规划的一个新的总体化技术 — 过滤集技术. 考虑到在罚函数方法中罚参数的选择常常导致问题的数值困难和坏条件, 这个新技术的动机首先是避免采用罚函数和选择罚参数; 其次, 与通常方法相比, 过滤集型方法允许一定程度的非单调性, 放松了接受新迭代点的要求. 这个性质在算法实现中被认为是很有益的. 另外, 过滤集方法根据多目标优化的想法, 在接受准则中利用了目标函数和约束违反度函数, 这样, 算法在迭代过程中直接反映了原问题的信息, 因而过滤集-逐次线性化方法优于原来的逐次线性化方法. 现在, 过滤集型方法已经广泛而有效地用于约束优化和无约束优化中 (见文献 [22–28]), 其理论和数值结果非常令人鼓舞. 孙文瑜在文献 [29] 中给出了过滤集型最优化方法的详细评述.

受解非线性规划的过滤集方法的启发, 本文的目的是考虑将过滤集方法推广到非线性半定规划上. 我们打算结合逐次线性化方法和过滤集方法的思想和优点, 提出解非线性半定规划的新的方法, 这无论对非线性半定规划还是对过滤集型方法来说, 都是一个很有意义的尝试.

在解非线性半定规划的若干方法中, Kanzow 等在文献 [7] 中提出的逐次线性化方法 (SL 方法, Successive Linearization (SL) method) 是一个重要方法. 这个方法在每一次迭代中解一个二次半定规划子问题, 而这个二次半定规划子问题可以转化成一个含有二阶锥约束的线性半定规划子问题. 这样, 逐次线性化方法 (SL 方法) 本质上就成为一个计算上相对简单的一阶方法, 而现成的有效的软件 SDPT3 (见文献 [30–32]) 可以用来求解这个子问题, 这给我们提供了方便和优势. 但是, 在 SL 方法的接受准则中, 罚信息被采用了, 并且接受准则 $\xi^{(k)}=0$ 似乎太严厉, 这不利于算法的数值实现. 如果采用过滤集技术作为接受准则, 这个缺点可以被克服. 另外, SL 方法虽然简单, 但没有充分利用原函数提供的信息. 基于以上分析, 本文在逐次线性化的框架上引进过滤集技术. 为了便于利用现成的有效软件 SDPT3 (见文献 [30–32]), 本文仍然采用文献 [7] 中提出的线性化子问题.

本文基于过滤集技术和 SL 方法, 给出了解非线性半定规划问题的新方法. 在理论上, 我们提出了算法, 给出了收敛性证明, 这个证明的思路不同于文献 [7] 中 SL 方法的证明, 而是根据 Fletcher 等在文献 [23] 中关于解非线性规划的过滤集-逐次线性规划方法的框架进行推广得到的 (证明中的细节是不同的). 我们在较弱的假设条件下得到了过滤集-逐次线性化方法的总体收敛性, 即, 代替文献 [7] 中四个假设条件 (A.1)–(A.4), 我们仅需要利用三个假设条件 (A.1) (A.2) (A.4) 得到收敛性结果. 数值上, 在大部分试验问题上, 我们的方法比原来的 SL 方法更有竞争力.

下面, 我们引进本文所采用的一些符号和定义. 设函数 $g_i, i=1,\ldots,m$, 表示函数 g 的分量, $Df(X), D^2f(X)$ 和 $Dg_i(X)$, $D^2g_i(X)$ 分别表示 f 和 g_i 的一阶和二阶 Fréchet 导数, 其中 $D^2f(X)(\Delta X,\Delta X)$ 表示 $\langle D^2f(X)(\Delta X),\Delta X\rangle$, $\forall X$, $\Delta X\in S^n$. 矩阵 $X^*\in S^n$ 称为问题 (1.1) 的驻点, 如果存在 Lagrange 乘子 $(\lambda^*,U^*)\in\mathbb{R}^m\times S^n$ 满足下面的 Karush-Kuhn-Tucker (KKT) 条件:

$$\begin{aligned}
&Df(X^*)+\sum_{i=1}^m\lambda_i^*Dg_i(X^*)-U^*=0,\\
&g_i(X^*)\leqslant 0,\quad i=1,\ldots,m,\quad X^*\succeq 0,\\
&\lambda_i^*\geqslant 0,\quad U^*\succeq 0,
\end{aligned}\tag{1.2}$$

$$\lambda_i^* g_i(X^*) = 0, \quad i = 1, \ldots, m, \quad \langle X^*, U^* \rangle = 0,$$

其中 $\langle \cdot, \cdot \rangle$ 表示 $\mathbb{R}^{n\times n}$ 中的内积 (i.e., $\langle A, B\rangle = \mathrm{tr}(A^\top B)$), $\|X\|_F = \langle X, X\rangle^{\frac{1}{2}}$, $\forall\ X \in S^n$, $\|g\|_\infty = \max_{i=1,\ldots,m} |g_i|$, $\forall\ g \in \mathbb{R}^m$ (见文献 [33–35]). 另外, 我们定义三个数 a, t, b $(a \leqslant b)$ 的中间值为

$$\mathrm{mid}(a, t, b) := \begin{cases} a, & \text{若 } t < a, \\ t, & \text{若 } t \in [a, b], \\ b, & \text{若 } t > b. \end{cases}$$

除了 KKT 条件外, 本文还需要 Fritz-John 必要条件 (见文献 [37]). X^* 是问题 (1.1) 的解的必要条件是 X^* 是可行的, 且方向集

$$\{\Delta X \mid \langle \Delta X, Df(X^*)\rangle < 0,\ \langle \Delta X, Dg_i(X^*)\rangle < 0,\ \forall i \in \mathcal{A}^*,\ E^{*\top} \Delta X E^* \succ 0\} \tag{1.3}$$

是空集, 其中 $\mathcal{A}^*$ 表示 X^* 处的有效约束集, E^* 的列表示对应于 X^* 的零特征值的标准直交特征向量. 当上述两个条件满足时, 我们称 X^* 是一个 FJ 点.

我们在定理证明中还需要: 如果聚点 X^∞ 是有 s $(\leqslant n)$ 个零特征值的可行点, 但 X^∞ 不是 FJ 点, 那么, 存在 $\varepsilon > 0$ (我们可以假定 $\varepsilon < 1$) 和对称矩阵 ΔX, 满足 $\|\Delta X\|_F = 1$ 和对所有 $X \in \mathcal{N}^\infty$,

$$\langle \Delta X, Df(X)\rangle \leqslant -\varepsilon, \quad \langle \Delta X, Dg_i(X)\rangle \leqslant -\varepsilon, \quad \forall i \in \mathcal{A}^\infty, \quad E_X^\top \Delta X E_X \succeq \varepsilon I_s, \tag{1.4}$$

其中, $\mathcal{N}^\infty$ 是 X^∞ 的某个邻域, E_X 的列是 X 的对应于 s 个最小绝对值的特征值的标准直交特征向量. 为简单起见, 我们假定矩阵 X 有谱分解, 其特征值按非增顺序排列. 这样, 从 (1.3), $Df(X)$ 和 $Dg_i(X)$ 的连续性, 对应于 X 的特征值函数以及集合 $\mathcal{E}_X := \{E_X \mid \mathrm{X} \in \mathcal{N}^\infty\}$ 的有界性可以推出公式 (1.4).

问题 (1.1) 的罚问题是:

$$\begin{aligned} &\min_{X \in S^n} p_\alpha(X) \\ &\text{s.t.}\ \ X \succeq 0, \end{aligned} \tag{1.5}$$

其中 $p_\alpha : S^n \to \mathbb{R}$ 为精确 ℓ_1-罚函数

$$p_\alpha(X) := f(X) + \alpha \sum_{i=1}^m \max\{0, g_i(X)\}, \tag{1.6}$$

$\alpha > 0$ 是罚参数.

一个有用的结论是

引理 1.1[7] 当 $\sum_{i=1}^m \max\{0, g_i(X)\} = 0$ 时, (1.5) 的驻点是原问题 (1.1) 的驻点.

本文组织如下: 在第 2 节我们提出过滤集-逐次线性化算法, 并通过与文献 [7] 中的逐次线性化算法的比较来进行分析. 第 3 节研究我们算法的总体收敛性, 在较弱的条件下证明了算法所产生的序列的某个聚点是原问题 (1.1) 的驻点. 第 4 节报告了初步的有竞争力的数值结果. 在第 5 节我们给出了一个简单的结束语.

2 过滤集-逐次线性化算法

在这一节, 通过对照文献 [7] 中的算法 4.1, 我们提出了解非线性半定规划问题 (1.1) 的过滤集-逐次线性化算法. 首先, 我们定义文献 [7] 中给出的函数 $\Phi_\alpha : S^n \times S^n \to \mathbb{R}$:

$$\Phi_\alpha(X, \Delta X) := f(X) + \langle Df(X), \Delta X\rangle + \alpha \sum_{i=1}^{m} \max\{0, g_i(X) + \langle Dg_i(X), \Delta X\rangle\}, \tag{2.1}$$

它是 $p_\alpha(X + \Delta X)$ 的一阶近似. 设 $\xi^{(k)}$ 表示 m 维向量, 其分量为

$$\max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X^{(k)}\rangle\}, \quad \forall\, i = 1, \ldots, m,$$

其中 $X^{(k)}$ 是第 k 次迭代点, $\Delta X^{(k)}$ 是第 k 次的主迭代步.

设 $\Delta X_j^{(k)}$ 表示第 k 次主迭代中的第 j 个内循环迭代步, 设 $h(X) := \max\{0, \max_{i=1,\ldots,m}\{g_i(X)\}\}$, $h^{(k)} := h(X^{(k)})$, $h_j^{(k)} := h(X^{(k)} + \Delta X_j^{(k)})$, $f^{(k)} := f(X^{(k)})$, $f_j^{(k)} := f(X^{(k)} + \Delta X_j^{(k)})$, $\Delta f_j^{(k)} := f^{(k)} - f_j^{(k)}$, $\Delta l_j^{(k)} := -\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle$, $\xi_j^{(k)}$ 表示 m 维向量, 其分量为 $\max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\}$, $\forall\, i = 1, \ldots, m$.

现在, 根据多目标优化中 "占优" 的概念, 我们给出过滤集的几个基本概念.

定义 2.1 点对 $(f^{(k)}, h^{(k)})$ 被称为占优另一个点对 $(f^{(l)}, h^{(l)})$, 当且仅当 $f^{(k)} \leqslant f^{(l)}$ 和 $h^{(k)} \leqslant h^{(l)}$.

在非线性半定规划中, 这意味着 $X^{(k)}$ 关于 $\min f(X)$ 和 $\min h(X)$ 两个目标至少与 $X^{(l)}$ 一样好. 利用这个概念, Fletcher 和 Leyffer[21] 将过滤集用于信赖域型优化算法作为接受或拒绝试验步的准则 (见文献 [21]).

定义 2.2 过滤集是点对 $(f^{(l)}, h^{(l)})$ 的表, 在这个表中, 没有一个点对占优任何其他点对. 一个点对 $(f^{(k)}, h^{(k)})$ 称为可被接受包含在过滤集中, 如果它不被过滤集中任何点对占优.

应用上述定义, 一个点 X 称为可被接受包含到过滤集中, 如果它相应的点对 (f, h) 不被过滤集中任何点对占优. 这个条件可以写成:

$$h < h_t \text{ 或 } f < f_t, \quad \forall\, t \in \mathcal{F}, \tag{2.2}$$

其中 $\mathcal{F}$ 表示过滤集中点对 (或元素) 的当前指标集. 通常, 把点 X 包含到过滤集中是指把它相应的点对 (f, h) 加到过滤集的点对表中去. 过滤集中的任何点对如果被新的点对占优, 就将其去掉. 代替 (2.2), 一个更有效的接受准则为: 对所有 $t \in \mathcal{F}$,

$$h \leqslant \beta h_t \text{ 或 } f \leqslant f_t - \gamma h_t, \tag{2.3}$$

其中 β 和 γ 是预置参数, 满足 $1 > \beta > \gamma > 0$, β 靠近 1 而 γ 靠近 0 (见文献 [24]).

最后, 我们引进 f-型迭代和 h-型迭代的含义. 对某个迭代, 其迭代点被当前过滤集接受. 如果下面两个条件满足,

$$\Delta f_j^{(k)} \geqslant \sigma \Delta l_j^{(k)} \quad \text{和} \quad \Delta l_j^{(k)} \geqslant \delta (h^{(k)})^2, \tag{2.4}$$

其中 σ 和 δ 是正参数, 则称这个迭代为 f-型迭代. 如果上式右边的不等式不满足, 则这个迭代称为 h-型迭代. 当 h-型迭代存在时, 将当前的迭代点包含到过滤集中. 关于这一点的详细讨论见文献 [23].

现在, 我们叙述新算法, 它利用过滤集技术代替文献 [7] 中 SL 方法的算法 4.1 的 (S.2), (S.3), (S.4) 步.

算法 2.3 (过滤集-逐次线性化方法)

(S.0) 选择初始数据: $\alpha_0^{(0)} > 0$, $0 < \gamma < \beta < 1$, $\varepsilon^{(0)} > 0$, $\theta_1 > 1$, $\theta_2 > 0$, $0 < \theta_3 < \varepsilon^{(0)}$, $2\gamma < \sigma_1 < 1$, $\sigma_4, \sigma_5, \sigma_6 > 1$, $1 - \frac{\theta_3}{\varepsilon^{(0)}} < \sigma_7 < 1$, $0 < \sigma_2, \sigma_3, \sigma_8, \sigma_9, \sigma_{10} < 1$, $c_{\max} > c_{\min} > 0$, $c_0^{(0)} \in [c_{\min}, c_{\max}]$, $u > 0$, $X^{(0)} \succeq 0$, $X_+ := X^{(0)}$, $\mathcal{F}^{(0)} := \{0\}$, $(h_0, f_0) := (u, -\infty)$, $f^{(0)} := f(X^{(0)})$, $\overline{h} = h^{(0)} = h(X^0)$,

$$\delta := \begin{cases} 1, & \text{如果} \quad h^{(0)} > 0, \\ 0, & \text{如果} \quad h^{(0)} = 0, \end{cases}$$

并令 $k := 0$, $j := 0$.

(S.1) 求子问题的解: 求子问题

$$\begin{aligned} \min_{\Delta X \in S^n} \ & \frac{1}{2} c_j^{(k)} \langle \Delta X, \Delta X \rangle + \Phi_{\alpha_j^{(k)}}(X^{(k)}, \Delta X) \\ \text{s.t.} \quad & X^{(k)} + \Delta X \succeq 0 \end{aligned} \tag{2.5}$$

的解 $\Delta X_j^{(k)} \in S^n$.

(S.2) 判断: 如果 $\overline{h} \geqslant h(X^{(k)} + \Delta X_j^{(k)})$, 则 $\overline{h} := h(X^{(k)} + \Delta X_j^{(k)})$, $X_+ := X^{(k)} + \delta \Delta X_j^{(k)}$. 如果 $\|\mathrm{svec}(\Delta X_j^{(k)})\|_\infty \leqslant \varepsilon^{(k)}$, 转 (S.5); 否则, 如果 $\|\xi_j^{(k)}\|_\infty \geqslant \sigma_4 \|\Delta X_j^{(k)}\|_F^2$, 令 $c_{j+1}^{(k)} := \theta_1 c_j^{(k)}$, $\alpha_{j+1}^{(k)} := \alpha_j^{(k)} + \theta_2$, $j := j+1$, 转 (S.1); 否则转 (S.3).

(S.3) 过滤集策略: 如果 $h_j^{(k)} \leqslant \beta h_t$ 或 $f_j^{(k)} \leqslant f_t - \gamma h_t$ $(t \in \mathcal{F}^{(k)})$, 那么转 (S.4); 否则, 令 $c_{j+1}^{(k)} := \theta_1 c_j^{(k)}, \alpha_{j+1}^{(k)} := \alpha_j^{(k)} + \theta_2, j := j+1$, 转 (S.1).

(S.4) f-或 h-迭代: 如果 $\{\Delta f_j^{(k)} < \sigma_1 \Delta l_j^{(k)}$ 且 $\Delta l_j^{(k)} \geqslant \sigma_3 (h^{(k)})^2\}$ 或者 $\{h_j^{(k)} \geqslant h^{(k)}$ 且 $f_j^{(k)} \geqslant f^{(k)}\}$, 令 $c_{j+1}^{(k)} := \theta_1 c_j^{(k)}$, $\alpha_{j+1}^{(k)} := \alpha_j^{(k)} + \theta_2$, $j := j+1$, 转 (S.1); 否则转 (S.6).

(S.5) KKT 点: 如果 $\Delta X_j^{(k)} = 0$ 且 $h^{(k)}$(i.e., $\xi_j^{(k)}) = 0$, 停止; 否则 $X^{(k+1)} := X_+$,

$$c^{(k+1)} := \begin{cases} |\,(c_{\min}, c_j^{(k)}, \sigma_{10} c_{\max}), & \text{如果 } \delta = 1, \\ |\,(c_{\min}, \sigma_2 c_j^{(k)}, \sigma_{10} c_{\max}), & \text{如果 } \delta = 0, \end{cases}$$

$$\alpha^{(k+1)} := \begin{cases} \alpha_j^{(k)} + \sigma_5 \theta_2, & \text{如果 } \delta = 1, \\ \sigma_8 \alpha_j^{(k)}, & \text{如果 } \delta = 0, \end{cases}$$

$$\varepsilon^{(k+1)} = \max\{0, \varepsilon^{(k)} - \theta_3\},$$

$$\theta_3 := \begin{cases} \sigma_7 \theta_3, & \text{如果 } \varepsilon^{(k+1)} \neq 0, \\ \theta_3, & \text{否则}, \end{cases}$$

$k := k+1$, 转 (S.7).

(S.6) 主迭代: 令 $\Delta X^{(k)} := \Delta X_j^{(k)}, c^{(k)} := c_j^{(k)}, \alpha^{(k)} := \alpha_j^{(k)}, \Delta f^{(k)} := \Delta f_j^{(k)}$, $\Delta l^{(k)} := \Delta l_j^{(k)}$, $X^{(k+1)} := X^{(k)} + \Delta X^{(k)}$. 如果 $\Delta l^{(k)} < \sigma_3 (h^{(k)})^2$, 则校正当前过滤集, 并令 $(h_{k+1}, f_{k+1}) :=$

$(h^{(k+1)}, f^{(k+1)})$. 同时, 去掉 $\mathcal{F}^{(k+1)}$ 中被新点占优的点, 校正参数:

$$c^{(k+1)} := \begin{cases} |\,(c_{\min}, \sigma_2 c^{(k)}, c_{\max}), & \text{如果 } \Delta f^{(k)} \geqslant \sigma_9 \Delta l^{(k)}, \\ |\,(c_{\min}, \theta_1 c^{(k)}, c_{\max}), & \text{如果 } \Delta f^{(k)} < \sigma_1 \Delta l^{(k)}, \\ |\,(c_{\min}, c^{(k)}, c_{\max}), & \text{否则}, \end{cases}$$

$\alpha^{(k+1)} := \min\{\alpha^{(k)}, \sigma_6\}$, $\varepsilon^{(k+1)} := \theta_3 \varepsilon^{(k)}$, $k := k+1$.

(S.7) 建立新的内循环: 令 $j := 0, c_j^{(k)} := c^{(k)}, \alpha_j^{(k)} := \alpha^{(k)}, X_+ := X^{(k)}, h^{(k)} = h(X^{(k)})$, $f^{(k)} = f(X^{(k)})$, $\overline{h} = h(X^{(k)})$,

$$\delta := \begin{cases} 1, & \text{如果 } h^{(k)} > 0, \\ 0, & \text{如果 } h^{(k)} = 0, \end{cases}$$

转 (S.1).

注 1 子问题 (2.5) 是在闭凸集上极小化严格凸函数的极小化问题, 这样, 该子问题存在唯一解.

注 2 正像在第 1 节中所述, 子问题 (2.5) 可以转换成一个线性半定规划问题, 它可以利用软件 SDPT3 有效地求解 (见文献 [30–32]).

注 3 不同于逐次线性化方法中的 (S.2) 步, 在算法 2.3 中, (S.2) 和 (S.3) 步将试验当

$$\|\mathrm{svec}(\Delta X_j^{(k)})\|_\infty > \varepsilon^{(k)} \quad \text{和} \quad \|\xi_j^{(k)}\|_\infty \leqslant \sigma_4 \|\Delta X_j^{(k)}\|_F^2$$

时试探点能否被过滤集接受, 其中 svec 定义为: 对任何对称矩阵 $X \in S^n$,

$$\mathrm{svec}(X) = (X(1,1), \sqrt{2}X(1,2), X(2,2), \sqrt{2}X(1,3), \sqrt{2}X(2,3), X(3,3), \ldots)^\top \in \mathbb{R}^{\frac{n(n+1)}{2}},$$

其中 $X(i,j)$ 是 X 的 (i,j)-元素, 参数 σ_4 选为一个很大的实数 (例如, $1.0\mathrm{e}+8$), 使得条件 $\|\xi_j^{(k)}\|_\infty \leqslant \sigma_4 \|\Delta X_j^{(k)}\|_F^2$ 满足. 明显地, 过滤集中的点对是原始问题 (1.1) 在当前迭代点的目标函数和约束违反度函数. 因此, 它直接提供了原始问题在当前点的信息.

注 4 在算法 2.3 中, 步 (S.3) 用于试验新的试探点可否被当前过滤集接受, 步 (S.4) 中第二个准则用来保证试探点不差于前面的点, 而步 (S.4) 中第一个准则保证在新点处原问题的目标函数值有足够的下降. 这些连同步 (S.6) 中的准则一起区分 f-型迭代和 h-型迭代 (详见文献 [21, 23, 24]).

注 5 步 (S.6) 产生新的迭代点, 并将 h-型迭代的迭代点加入到当前过滤集中. 我们把新迭代点加入到过滤集的同时, 把过滤集中被新点占优的旧点去掉, 这与文献 [23] 的处理不同. 另外, 我们关于 $c^{(k)}$ 和 $\alpha^{(k)}$ 的校正准则也不同于 SL 方法中的校正准则.

注 6 步 (S.5) 用来判断驻点. 在实际执行中, 停机准则采用

$$\|\mathrm{svec}(\Delta X_j^{(k)})\|_\infty < 1.0\mathrm{e}-4 \quad \text{和} \quad h^{(k)} < 1.0\mathrm{e}-4.$$

注 7 当 $\delta = 1$ 时, 步 (S.5) 防止当前主迭代点 X_+ 的约束违反度函数值过大. 但当 $\delta = 0$ 时, 我们通过减少 $c^{(k)}$ 和 $\alpha^{(k)}$ 来减少原问题的目标函数值. 同时, 参数 $c_j^{(k)}$ 在 $[c_{\min},\ \sigma_{10} c_{\max}]$ 中重置.

注 8 在算法中, 内循环 $((S.1) \leftrightarrow (S.2) \leftrightarrow (S.3) \leftrightarrow (S.4))$ 用 j 表示, 主迭代用 k 表示. 下一节, 在适当的条件下, 我们将证明内循环中仅存在有限多个迭代. 参数 δ 用来判定在某个主迭代是否必要改善原问题的可行性.

3 收敛性分析

这一节讨论算法 2.3 的总体收敛性. 为此, 我们首先引进一些预备性的引理, 然后在引理 3.5–3.8 中证明一些必要的结果. 最后, 我们建立本文主要定理 3.9.

首先作如下假设:

(A.1) 由算法 2.3 产生的序列 $\{X^{(k)}\}$ 是有界的.

(A.2) 由算法 2.3 产生的序列 $\{c_j^{(k)}\Delta X_j^{(k)}\}$ 是有界的.

(A.3) 对任何给出的 $X^* \succeq 0$, 如果 $(\lambda^*, U^*) \in \mathbb{R}^m \times S^n$ 满足

$$\sum_{i:g_i(X^*)\geqslant 0} \lambda_i^* Dg_i(X^{(*)}) - U^* = 0, \quad \lambda^* \geqslant 0, \quad U^* \succeq 0, \quad \langle X^*, U^*\rangle = 0,$$

那么, $\lambda^* = 0$, $U^* = 0$.

上面的假设分别是文献 [7] 中的假设 (1), (2) 和 (4). 假设 (A.3) 对应于非线性规划推广的 Mangasarian-Fromovitz 约束规范条件.

引理 3.1 $\tau^{(k)} = \min_{t\in\mathcal{F}^{(k)}}\{h_t\} > 0$.

证明 我们需要证明 $h_t > 0, \forall\, t \in \mathcal{F}^{(k)}$. 由反证法, 如果存在某个 $t^* \in \mathcal{F}^{(k)}$ 使得在第 k^* 次迭代被接受的 h_{t^*} 为零, 那么我们只要证明 $\Delta l^{(k^*)} \geqslant 0$ 成立, 这与 h-迭代的定义矛盾.

由于 0 总是子问题的可行点, 则下面的不等式成立:

$$\begin{aligned}&\frac{1}{2}c^{(k^*)}\langle \Delta X^{(k^*)}, \Delta X^{(k^*)}\rangle + f(X^{(k^*)}) + \langle Df(X^{(k^*)}), \Delta X^{(k^*)}\rangle \\ &\quad + \alpha^{(k^*)}\sum_{i=1}^m \max\{0, g_i(X^{(k^*)}) + \langle Dg_i(X^{(k^*)}), \Delta X^{(k^*)}\rangle\} \leqslant f(X^{(k^*)}),\end{aligned} \tag{3.1}$$

其中, $\Delta X^{(k^*)}$ 是子问题 (2.5) 的最优解.

从 (3.1) 式左边第一项和第四项的非负性, 我们有

$$\Delta l^{(k^*)} = -\langle Df(X^{(k^*)}), \Delta X^{(k^*)}\rangle \geqslant 0.$$

于是, 证明完成.

引理 3.2[7] 设 $\alpha_j^{(k)} > 0$, 如果 $\Delta X_j^{(k)} = 0$ 是子问题 (2.5) 的 (唯一) 解, 那么 $X^{(k)}$ 是罚问题 (1.5) 的驻点, $\alpha_j^{(k)}$ 是罚参数. 反之, 如果 $X^{(k)}$ 是罚问题 (1.5) 的驻点, $\alpha_j^{(k)}$ 是罚参数, 那么, 对每一个 $c_j^{(k)} > 0$, $\Delta X_j^{(k)} = 0$ 是子问题 (2.5) 的 (唯一) 解.

引理 3.3[7] 假定假设 (A.1) 和 (A.2) 成立, 假设 (A.3) 在算法 2.3 产生的序列 $\{X^{(k)}\}$ 的任何聚点处成立, 如果由解 (2.5) 产生的序列 $\{\Delta X_j^{(k)}\}$ 收敛到 0, 那么在有限多次迭代以后 $\{\xi_j^{(k)}\}$ 将为 0.

引理 3.4[23] 如果假设 (A.1) 成立, 步 (S.5) 在有限次迭代后不再迭代, 步 (S.6) 无限次迭代, 那么, 序列 $\{X^{(k)}\}$ 的任何聚点 $(\forall k \in \mathcal{Q})$ 对于原问题 (1.1) 是可行的.

由序列 $\{X^{(k)}\}$ 的有界性, 序列 $\{X^{(k)}\}$ 的闭凸包 $\overline{\mathrm{co}}(\{X^{(k)}\})$ 是有界的. 于是连同 f 和 g_i $(i = 1, \ldots, m)$ 的二次连续可微的条件得到: 对所有 $X \in \overline{\mathrm{co}}(\{X^{(k)}\})$ 和所有满足 $\|S\|_F = 1$ 的矩阵 S, $\frac{1}{2}D^2 f(X)(S, S)$ 和 $\frac{1}{2}D^2 g_i(X)(S, S)\,(i = 1, \ldots, m)$ 是有界的. 再设 M 是所有这些项的上界.

明显地, 内循环 (S.1) $\leftrightarrow$ (S.2) $\leftrightarrow$ (S.3) $\leftrightarrow$ (S.4) 可以产生某个循环, 它仅与下标 j 有关. 算法 2.3 中另一个循环是 (S.1) $\rightarrow$ (S.2) $\rightarrow$ (S.5) $\rightarrow$ (S.7) $\rightarrow$ (S.1), 利用下面的定理, 在某些条件下, 它仅仅迭代有限多次.

引理 3.5 假定假设 (A.1) 和 (A.2) 成立, 假定 (A.3) 对算法 2.3 产生的序列 $\{X^{(k)}\}$ 的任何聚点成立. 那么, 在算法 2.3 中仅存在有限多个 $\Delta X_j^{(k)} = 0$.

证明 如果算法 2.3 在有限多次迭代后在步 (S.5) 终止, 那么引理成立. 从引理 3.2 和引理 1.1 可知, 满足终止准则的主迭代点是原问题的驻点 (KKT 点).

假定算法中存在无限多次迭代, 设 $Q := \{(k,j) \mid \Delta X_j^{(k)} = 0\}$. 我们用反证法证明这个引理. 假定在 Q 中有无限多个元素, 显然, 当 $k \to \infty$ 或 $j \to \infty$ 或 $k, j \to \infty$ 时, 序列 $\{\Delta X_j^{(k)}\}$ 收敛到 0. 这样, 由假设和引理 3.3 得到: 存在 $(k^*, j^*) \in Q$ 使得 $\xi_{j^*}^{(k^*)} = 0$. 于是, 算法在某次迭代在 (S.5) 步必定停止, 这与假设矛盾. 证毕.

利用上述定理和算法 2.3 中的 σ_7, 可以证明: 在有限多次迭代以后, 循环 (S.1) → (S.2) → (S.5)→ (S.7) → (S.1) 终止. 为了证明在内循环中存在有限多次迭代, 我们建立下面的引理.

引理 3.6 如果对某个主迭代 k, 内循环迭代无限多次, 那么, 当 $j \to \infty$ 时, $\{\Delta X_j^{(k)}\}$ 收敛到 0.

证明 假设内循环 (S.1) ↔ (S.2) ↔ (S.3)↔ (S.4) 迭代无限多次. 注意到在 (S.2), (S.3) 和 (S.4) 中 $c_j^{(k)}$ 和 $\alpha_j^{(k)}$ 的校正公式分别是 $c_{j+1}^{(k)} := \theta_1 c_j^{(k)}$ 和 $\alpha_{j+1}^{(k)} := \alpha_j^{(k)} + \theta_2$, 于是得到

$$c_j^{(k)} = (\theta_1)^j c_0^{(k)}, \qquad \alpha_j^{(k)} = \alpha_0^{(k)} + j\theta_2, \quad \forall\, j. \tag{3.2}$$

用反证法. 假定存在正数 η 和 $\{\Delta X_j^{(k)}\}$ 的子集满足

$$\|\Delta X_j^{(k)}\|_F \geqslant \eta. \tag{3.3}$$

不失一般性, 假定这个子集就是 $\{\Delta X_j^{(k)}\}$ 本身, 即 (2.3) 对所有 j 成立.

显然, $\Delta X_j^{(k)} = 0$ 是子问题 (2.5) 的可行解, 在这一点目标函数值为

$$f(X^{(k)}) + \alpha_j^{(k)} \sum_{i=1}^m \max\{0, g_i(X^{(k)})\}, \quad \forall\, j. \tag{3.4}$$

另一方面, 对某个 j, 子问题 (2.5) 的最优值是

$$\begin{aligned}
&\frac{1}{2} c_j^{(k)} \langle \Delta X_j^{(k)}, \Delta X_j^{(k)} \rangle + f(X^{(k)}) + \langle Df(X^{(k)}), \Delta X_j^{(k)} \rangle \\
&\quad + \alpha_j^{(k)} \sum_{i=1}^m \max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X_j^{(k)} \rangle\} \\
&\quad \geqslant \frac{1}{2} c_j^{(k)} \|\Delta X_j^{(k)}\|_F^2 + f(X^{(k)}) + \langle Df(X^{(k)}), \Delta X_j^{(k)} \rangle \\
&\quad \geqslant \frac{1}{2} c_j^{(k)} \|\Delta X_j^{(k)}\|_F^2 + f(X^{(k)}) - \|Df(X^{(k)})\|_F \|\Delta X_j^{(k)}\|_F,
\end{aligned} \tag{3.5}$$

其中 $\Delta X_j^{(k)}$ 是子问题的最优解. 下面, 我们将对两种不同情形分别推出矛盾.

第一种情形: $\{\Delta X_j^{(k)}\}$ 是有界的. 在这种情形, 可以假定 $\|\Delta X_j^{(k)}\|_F \leqslant \overline{\eta}, \forall j$. 设 $a := \frac{1}{2} c_0^{(k)} \eta^2 > 0$, $b := \theta_2 \sum_{i=1}^m \max\{0, g_i(X^{(k)})\}$, $c := \alpha_0^{(k)} \sum_{i=1}^m \max\{0, g_i(X^{(k)})\} + \|Df(X^{(k)})\|_F \overline{\eta}$, 它们都是常数. 明显地, 存在正数 j^* 使得对所有 $j \geqslant j^*$, 不等式 $\theta_1 > 1$,

$$a(\theta_1)^j > bj + c,$$

其中 $\theta_1 > 1$. 于是, 我们有

$$\begin{aligned}
&\frac{1}{2} \eta^2 c_0^{(k)} (\theta_1)^j - \|Df(X^{(k)})\|_F \overline{\eta} + f(X^{(k)}) \\
&\quad \geqslant f(X^{(k)}) + (\alpha_0^{(k)} + j\theta_2) \sum_{i=1}^m \{0, g_i(X^{(k)})\}
\end{aligned} \tag{3.6}$$

对所有 $j \geqslant j^*$ 成立. 组合 (3.2), (3.4)–(3.6) 得到

$$\begin{aligned}&\frac{1}{2}c_j^{(k)}\langle \Delta X_j^{(k)}, \Delta X_j^{(k)}\rangle + f(X^{(k)}) + \langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle \\ &\quad +\alpha_j^{(k)}\sum_{i=1}^m \max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\} \\ &\quad \geqslant f(X^{(k)}) + \alpha_j^{(k)}\sum_{i=1}^m \max\{0, g_i(X^{(k)})\}\end{aligned}$$

对所有 $j \geqslant j^*$ 成立, 这与 $\Delta X_j^{(k)}$ 的最优性矛盾.

第二种情形: $\{\Delta X_j^{(k)}\}$ 是无界的. 不失一般性, 假定

$$\lim_j \|\Delta X_j^{(k)}\|_F = +\infty.$$

由假设可知, 存在正数 j' 使得

$$\begin{aligned}&\frac{1}{2}c_j^{(k)}\|\Delta X_j^{(k)}\|_F^2 + f(X^{(k)}) - \|Df(X^{(k)})\|_F\|\Delta X_j^{(k)}\|_F \\ &\quad \geqslant \left(\frac{1}{2}c_j^{(k)} - 1\right)\|\Delta X_j^{(k)}\|_F^2 + f(X^{(k)}) \\ &\quad \geqslant \left(\frac{1}{2}c_j^{(k)} - 1\right)\eta^2 + f(X^{(k)})\end{aligned} \tag{3.7}$$

对所有 $j \geqslant j'$ 成立. 类似于第一种情形可知, 当 j 充分大时, (3.7) 的最后一项大于 (3.4), 这形成矛盾. 证明完成.

现在我们讨论关于内迭代有限终止的引理.

引理 3.7 在引理 3.5 的假设下, 内迭代有限多次终止.

证明 用反证法. 假定存在某个主迭代 k 使得内循环无限多次迭代.

从算法 2.3 的校正法则可知, 对所有 j, $\Delta X_j^{(k)} \neq 0$. 但从引理 3.6, 当 $j \to \infty$ 时, $\Delta X_j^{(k)} \to 0$. 于是, 由引理 3.3 可知, 步 (S.1) 和 (S.2) 之间的迭代有限终止. 因此, 如果循环 (S.1) $\leftrightarrow$ (S.2) $\leftrightarrow$ (S.3) 是无限多次迭代, 则循环 (S.1) $\leftrightarrow$ (S.3) 必定是无限多次迭代. 从 g_i $(i = 1, \ldots, m)$ 的二次连续可微性可知, 对 $i = 1, \ldots, m$,

$$g_i(X^{(k)} + \Delta X_j^{(k)}) = g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle + \circ(\|\Delta X_j^{(k)}\|).$$

于是,

$$\max\{0, g_i(X^{(k)} + \Delta X_j^{(k)})\} - \max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\} = \circ(\|\Delta X_j^{(k)}\|). \tag{3.8}$$

从引理 3.3 和 3.6, 我们有对充分大的 j, $\xi_j^{(k)} = 0$, 这样 (3.8) 可以写成

$$\max\{0, g_i(X^{(k)} + \Delta X_j^{(k)})\} = \circ(\|\Delta X_j^{(k)}\|), \quad i = 1, \ldots, m,$$

利用引理 3.6, 得到当 $j \to \infty$ 时, $h_j^{(k)} \to 0$. 注意到引理 3.1, 对于某个 k, $\tau^{(k)}$ 是个常数. 这样, 不等式 $h_j^{(k)} \leqslant \beta h_t$ 对充分大的 j 必定满足.

根据上面的讨论, 在内循环存在无限多次迭代 (S.1) $\leftrightarrow$ (S.4).

由于子问题的目标函数是严格凸的, 可行域是闭凸集, 于是, 子问题 (2.5) 有唯一解. 对于 $X^{(k)} \succeq 0$ 和 $X^{(k)} + \Delta X_j^{(k)} \succeq 0$, 其中 $\Delta X_j^{(k)}$ 是子问题的最优解, 则

$$X^{(k)} + \lambda \Delta X_j^{(k)} \succeq 0$$

对所有 $\lambda \in [0,1]$ 成立. 这样, $\lambda \Delta X_j^{(k)}$ $(\lambda \in [0,1])$ 是子问题的可行点, 注意到 $\Delta X_j^{(k)} \neq 0$, 于是得到对所有 $\lambda \in [0,1)$,

$$\begin{aligned}
&\frac{1}{2}c_j^{(k)}\langle \Delta X_j^{(k)}, \Delta X_j^{(k)}\rangle + f(X^{(k)}) + \langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle \\
&\quad +\alpha_j^{(k)}\sum_{i=1}^m \max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\} \\
&\quad < \frac{1}{2}\lambda^2 c_j^{(k)}\langle \Delta X_j^{(k)}, \Delta X_j^{(k)}\rangle + f(X^{(k)}) \\
&\quad\quad +\lambda\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle + \alpha_j^{(k)}\sum_{i=1}^m \max\{0, g_i(X^{(k)}) + \lambda\langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\}.
\end{aligned}$$

上面的不等式可以写成

$$\begin{aligned}
-(1-\lambda)\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle > &\frac{1}{2}(1-\lambda^2)c_j^{(k)}\langle \Delta X_j^{(k)}, \Delta X_j^{(k)}\rangle \\
&+\alpha_j^{(k)}\Bigg[\sum_{i=1}^m \max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\} \\
&-\sum_{i=1}^m \max\{0, g_i(X^{(k)}) + \lambda\langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\}\Bigg]. \qquad (3.9)
\end{aligned}$$

下面, 我们根据 $h^{(k)}$ 的不同值分两种情况进行讨论, 并推出矛盾.

情形 (I): $h^{(k)} = 0$. 在这种情形, 我们需要考虑 (3.9) 右边的方括号. 为讨论方便起见, 令

$$A_i := g_i(X^{(k)}), \quad B_i := \langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle, \quad i = 1, \ldots, m.$$

这样, (3.9) 式右边的方括号可写成

$$\sum_{i=1}^m \max\{0, A_i + B_i\} - \sum_{i=1}^m \max\{0, A_i + \lambda B_i\}. \qquad (3.10)$$

当 $B_i = 0$ 时, $\max\{0, A_i + B_i\} - \max\{0, A_i + \lambda B_i\} = 0$. 因此, 不失一般性, 我们假定对 $i = 1, \ldots, m$, $B_i \neq 0$. 根据 A_i 和 B_i 的不同的值, (3.10) 的不同的值的情形如下:

(1) 若 $B_i > 0$, 则 $A_i + B_i > A_i + \lambda B_i, \forall \lambda \in [0,1)$, 于是,

$$\max\{0, A_i + B_i\} - \max\{0, A_i + \lambda B_i\} \geqslant 0, \quad \forall \lambda \in [0,1).$$

(2) 若 $B_i < 0$, 则从引理 3.3 和 3.6 可知, 对于充分大的 j, $A_i + B_i > 0$ 将不发生.

(3) 若 $B_i < 0$ 和 $A_i + B_i < 0$, 则当 λ 靠近 1 时, $A_i + \lambda B_i < 0$. 于是有

$$\max\{0, A_i + B_i\} - \max\{0, A_i + \lambda B_i\} = 0$$

对所有 $\lambda \in [-\frac{A_i}{B_i}, 1)$ 成立. 设 $\lambda^i := \max\{0, -\frac{A_i}{B_i}\}$, $\mathcal{L} := \{i \in \{1, 2, \ldots, m\} \mid B_i < 0,\ A_i + B_i < 0\}$, $\lambda_j := \max_{i \in \mathcal{L}} \lambda^i \in [0,1)$.

(4) 若 $B_i < 0$ 和 $A_i + B_i = 0$, 则 $A_i = -B_i > 0$ 且

$$\max\{0, A_i + B_i\} - \max\{0, A_i + \lambda B_i\} = -(1-\lambda)A_i.$$

于是, 因为 $g_i(X^{(k)}) = A_i > 0$ 与 $h^{(k)} = 0$ 矛盾, 从而 (4) 不会发生.

从上述讨论可知, 存在正数 j' 使得

$$\begin{aligned}
&\sum_{i=1}^m \max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\} - \sum_{i=1}^m \max\{0, g_i(X^{(k)}) \\
&\quad +\lambda_j\langle Dg_i(X^{(k)}), \Delta X_j^{(k)}\rangle\} \geqslant 0 \qquad (3.11)
\end{aligned}$$

对所有 $j \geqslant j'$ 成立, 再利用 (3.9) 和 $\lambda_j < 1$ 得到

$$-\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle > \frac{1}{2}(1+\lambda_j)c_j^{(k)}\|\Delta X_j^{(k)}\|_F^2 \tag{3.12}$$

对所有 $j \geqslant j'$ 成立.

另一方面, 存在正数 j'' 使得

$$\frac{1}{2}(1+\lambda_j)c_j^{(k)}\|\Delta X_j^{(k)}\|_F^2 \geqslant (1-\sigma_1)^{-1}M\|\Delta X_j^{(k)}\|_F^2 \tag{3.13}$$

对所有 $j \geqslant j''$ 成立. 组合 (3.12) 和 (3.13), 我们得到

$$-\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle > (1-\sigma_1)^{-1}M\|\Delta X_j^{(k)}\|_F^2 \tag{3.14}$$

对所有 $j \geqslant j^* = \max\{j', j''\}$ 成立. 从 f 的二阶连续性和 $D^2f(X)(\Delta X, \Delta X)$ 的上界, 有

$$f(X^{(k)}) - f(X^{(k)} + \Delta X_j^{(k)}) \geqslant -\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle - M\|\Delta X_j^{(k)}\|_F^2,$$

这和 (3.14) 一起给出

$$\Delta f_j^{(k)} = f(X^{(k)}) - f(X^{(k)} + \Delta X_j^{(k)}) > -\sigma_1\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle = \sigma_1\Delta l_j^{(k)}$$

对所有 $j \geqslant j^*$ 成立. 由于 $-\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle > 0$, 故 $f(X^{(k)}) > f(X^{(k)} + \Delta X_j^{(k)})$. 这样, 在这种情形内迭代有限终止. 这产生了矛盾.

情形 (II): $h^{(k)} > 0$. 在这种情形, 根据反证法假设, 算法 2.3 中的步 (S.4) 和 $h_j^{(k)} \to 0$ $(j \to \infty)$ 得到

$$\Delta f_j^{(k)} < \sigma_1\Delta l_j^{(k)} \tag{3.15}$$

对所有充分大的 j 成立. 但由 (3.15), 引理 3.6, 和

$$f(X^{(k)} + \Delta X_j^{(k)}) = f(X^{(k)}) + \langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle + \circ(\|\Delta X_j^{(k)}\|),$$

我们有

$$\begin{aligned}-\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle &= f(X^{(k)}) - f(X^{(k)} + \Delta X_j^{(k)}) + \circ(\|\Delta X_j^{(k)}\|)\\ &\leqslant -\sigma_1\langle Df(X^{(k)}), \Delta X_j^{(k)}\rangle + \circ(\|\Delta X_j^{(k)}\|)\end{aligned}$$

对所有充分大的 j 成立. 这样, 对所有充分大的 j, 有 $(1-\sigma_1)\Delta l_j^{(k)} \leqslant \circ(\|\Delta X_j^{(k)}\|)$. 利用 $\|\Delta X_j^{(k)}\| \to 0$ $(j \to \infty)$, 得到: 存在充分大的 j, 使得

$$\Delta l_j^{(k)} \leqslant \sigma_3(h^{(k)})^2 \quad 和 \quad h_j^{(k)} < h^{(k)}$$

成立, 从而内循环终止, 这产生了矛盾. 证明完成.

引理 3.8 在假设条件 (A.3) 下, FJ 点满足 KKT 条件.

证明 假定 X^* 是原问题 (1.1) 的 FJ 点, 从 FJ 点的定义和广义 Farkas 引理 (见文献 [10, 33]) 可知, 存在 $\lambda_0' \geqslant 0$, $\lambda_i' \geqslant 0$, $i \in \mathcal{A}^*$, $s \times s$ 对称矩阵 $U \succeq 0$ 满足 $(\lambda_0', \lambda_{\mathcal{A}^*}', U) \neq 0$, 其中 $n-s$ 是 X^* 的秩, $\lambda_{\mathcal{A}^*}'$ 表示 $\{\lambda_i', i \in \mathcal{A}^*\}$ 的集合, 使得对所有 $\Delta X \in S^n$,

$$\lambda_0'\langle Df(X^*), \Delta X\rangle + \sum_{i\in\mathcal{A}^*}\lambda_i'\langle Dg_i(X^*), \Delta X\rangle - \langle E^*UE^{*\top}, \Delta X\rangle \geqslant 0.$$

因而,

$$\lambda_0' Df(X^*) + \sum_{i\in\mathcal{A}^*}\lambda_i' Dg_i(X^*) - E^*UE^{*\top} = 0.$$

用 U' 表示 $n \times n$ 对称矩阵 $E^*UE^{*\top}$. 显然, $U' \succeq 0$ 和 $\langle U', X^*\rangle = 0$. 另一方面, 由 X^* 的可行性, 有 $X^* \succeq 0$ 和 $g_i(X^*) \leqslant 0\,(i = 1, \ldots, m)$ 成立.

此外, 我们还有 $\lambda_0' \neq 0$, 因为否则的话, 由假设 (A.3) 得到 $\lambda_i' = 0\,(i \in \mathcal{A}^*)$, $U' = 0$, 因而 $U = 0$, 这与 $(\lambda_0',\ \lambda_{\mathcal{A}^*}',\ U) \neq 0$ 矛盾. 这样, 令 $\lambda_i^* := \lambda_0^{-1}\lambda_i'$ $(i \in \mathcal{A}^*)$, $\lambda_i^* := 0$ $(i \notin \mathcal{A}^*)$, $U^* := \lambda_0^{-1}U'$, 我们便得到在 X^* 处的 KKT 条件成立. 证明完成.

现在, 我们利用上面的引理给出算法 2.3 的总体收敛性定理.

定理 3.9 在引理 3.5 的假设下, 算法 2.3 或者在有限多次迭代后终止于 KKT 点, 或者迭代序列产生一个满足 KKT 条件的聚点.

证明 从引理 3.5 可知, 在有限多次迭代后算法产生的主迭代点是原问题的 KKT 点. 从引理 3.7 和引理 3.4 可知, 迭代序列 $(k \in \mathcal{Q})$ 的每一个聚点是原问题的可行点,$\mathcal{Q}$ 的定义将在下面给出.

当算法 2.3 无限多次迭代时存在两种可能的情形:

情形 (I). 对充分大的 k, 所有迭代都是 f-型迭代, 这时我们考虑 f-型迭代的子序列.

情形 (II). 存在 h-型迭代的无穷子序列.

我们将识别无穷子序列 $(k \in \mathcal{Q})$ 的存在性, 其任何聚点是 KKT 点. 在情形 (I), $\mathcal{Q}$ 恰是主序列的尾部, 它们由最后一个 h-型迭代后的所有迭代组成; 在情形 (II), $\mathcal{Q}$ 由所有 h-型迭代组成.

下面, 在反证法假设的条件下, 我们推导参数 ρ 的区间, 在此区间上, f-型迭代存在且 $c^{(k)}$ 充分大.

用反证法. 假定迭代序列 $\{X^{(k)}\}_{k\in\mathcal{Q}}$ 的可行聚点 X^∞ 不是一个 FJ 点. 从 (1.4) 可知, 存在 ΔX $(\|\Delta X\|_F = 1)$, 使得

$$\begin{cases} \langle Df(X^{(k)}), \Delta X\rangle \leqslant -\varepsilon, \\ \langle Dg_i(X^{(k)}), \Delta X\rangle \leqslant -\varepsilon, \quad \forall\, i \in \mathcal{A}^\infty, \\ E_k^\top \Delta X E_k \succeq \varepsilon I_S \end{cases} \tag{3.16}$$

对充分大的 $k \in \mathcal{Q}$ 成立, 其中 s 是矩阵 X^∞ 的零特征值的个数, E_k 的每一列是对应于 $X^{(k)}$ 的 s 个绝对值最小的特征值的标准直交特征向量. 设 V_k 是 $n \times (n-s)$ 矩阵, 使得 $\big(V_k\ E_k\big)$ 是对应于 $X^{(k)}$ 的直交矩阵. 不失一般性, 假定序列 $\big\{\big(V_k\ E_k\big)\big\}_{k\in\mathcal{Q}}$ 有极限 $\big(V\ E\big)$, 否则, 我们可以考虑序列 $\big\{\big(V_k\ E_k\big)\big\}_{k\in\mathcal{Q}}$ 的子序列. 这样, 我们有

$$\begin{pmatrix} V^\top \\ E^\top \end{pmatrix} X^\infty \big(V\ E\big) = \begin{pmatrix} V^\top X^\infty V & V^\top X^\infty E \\ E^\top X^\infty V & E^\top X^\infty E \end{pmatrix} = \begin{pmatrix} \Lambda_T & 0 \\ 0 & 0_S \end{pmatrix},$$

其中, $(n-s)\times(n-s)$ 对角阵 $\Lambda_T \succ 0$ 的对角元是 X^∞ 的按非增顺序排列的正特征值, 0_S 是 $s \times s$ 零矩阵.

显然, $X^{(k)} + \rho\Delta X \succeq 0$ 等价于

$$\begin{pmatrix} V_k^\top \\ E_k^\top \end{pmatrix} (X^{(k)} + \rho\Delta X)\big(V_k\ E_k\big) = \begin{pmatrix} \Lambda_T^k + \rho V_k^\top \Delta X V_k & \rho V_k^\top \Delta X E_k \\ \rho E_k^\top \Delta X V_k & \Lambda_S^k + \rho E_k^\top \Delta X E_k \end{pmatrix} \succeq 0, \tag{3.17}$$

其中, 对角矩阵 Λ_T^k 和 Λ_S^k 的对角元是 $X^{(k)}$ 的特征值, 因此, 当 $k \to \infty$ 时,

$$\Lambda_T^k \to \Lambda_T \succ 0, \quad \Lambda_S^k \to 0_S, \quad E_k \to E, \quad V_k \to V.$$

因为对所有 k, $X^{(k)}$ 是子问题的可行点, 故有

$$\begin{pmatrix} \Lambda_T^k & 0 \\ 0 & \Lambda_S^k \end{pmatrix} \succeq 0, \quad \forall\, k \in \mathcal{Q}.$$

因此,

$$\Lambda_S^k + \rho E_k^\top \Delta X E_k \succeq \rho E_k^\top \Delta X E_k \succ \rho\varepsilon I_S, \tag{3.18}$$

这意味着对所有充分大的 $k \in \mathcal{Q}$ 和 $\rho > 0$, $\Lambda_S^k + \rho E_k^\top \Delta X E_k$ 是正定矩阵. 注意到对所有充分大的 $k \in \mathcal{Q}$ 和 $\rho > 0$, (3.17) 等价于

$$\Lambda_T^k + \rho V_k^\top \Delta X V_k - \rho^2 (V_k^\top \Delta X E_k)(\Lambda_S^k + \rho E_k^\top \Delta X E_k)^{-1}(E_k^\top \Delta X V_k) \succeq 0, \tag{3.19}$$

于是, 上面两式给出

$$\begin{aligned} &\rho^2 (V_k^\top \Delta X E_k)(\Lambda_S^k + \rho E_k^\top \Delta X E_k)^{-1}(E_k^\top \Delta X V_k) \\ &\quad \preceq \rho^2 (V_k^\top \Delta X E_k)(\rho\varepsilon)^{-1} I_S (E_k^\top \Delta X V_k) \\ &\quad = \rho\varepsilon^{-1}(V_k^\top \Delta X E_k E_k^\top \Delta X V_k), \end{aligned}$$

这样, (3.17) 的充分条件为对所有充分大的 $k \in \mathcal{Q}$ 和 $\rho > 0$,

$$\Lambda_T^k + \rho(V_k^\top \Delta X V_k - \varepsilon^{-1} V_k^\top \Delta X E_k E_k^\top \Delta X V_k) \succeq 0.$$

为简单起见, 令 $G^k := V_k^\top \Delta X V_k - \varepsilon^{-1} V_k^\top \Delta X E_k E_k^\top \Delta X V_k$,

$$G^k \to G^* := V^\top \Delta X V - \varepsilon^{-1} V^\top \Delta X E E^\top \Delta X V.$$

如果 $G^* \succ 0$, 则对所有 $\rho \in (0, +\infty)$ 和充分大的 $k \in \mathcal{Q}$, (3.17) 成立. 否则, $\xi_{\min}^* \leqslant 0$, 其中 $\xi_{\min}^*$ 是 G^* 的最小特征值. 设 $\xi_{\min}^k$ 表示 G^k 的最小特征值, 我们有对充分大的 $k \in \mathcal{Q}$, $\xi_{\min}^k \geqslant \xi_{\min}^* - \Upsilon$, 其中 Υ 表示一个控制 $|\xi_{\min}^*|$ 的正数. 注意到 $\Lambda_T \succ 0$, 并设 $\lambda_{\min}^*$ 是 Λ_T 的最小特征值, 则 $\lambda_{\min}^* > 0$ 和 $\lambda_{\min}^k > \frac{1}{2}\lambda_{\min}^*$ (对所有充分大的 $k \in \mathcal{Q}$), 其中 $\lambda_{\min}^k$ 表示 Λ_T^k 的最小特征值. 因而对所有 $\rho \in \big(0, \frac{\lambda_{\min}^*}{2(\Upsilon - \xi_{\min}^*)}\big]$ 和充分大的 $k \in \mathcal{Q}$, 不等式 (3.17) 成立.

显然, 当 $\rho = 0$ 时, 对所有 $k \in \mathcal{Q}$, (3.17) 成立, 从而对所有 $\rho \in \big[0, \frac{\lambda_{\min}^*}{2(\Upsilon - \xi_{\min}^*)}\big]$ 和充分大的 $k \in \mathcal{Q}$, (3.17) 成立. 进而这时 $X^{(k)} + \rho\Delta X \succeq 0$ 成立.

对 X^∞ 处的有效约束, 由 (3.16) 有

$$g_i(X^{(k)}) + \rho\langle Dg_i(X^{(k)}), \Delta X\rangle \leqslant h^{(k)} - \rho\varepsilon, \quad \forall\, i \in \mathcal{A}^\infty.$$

对无效约束 $i \notin \mathcal{A}^\infty$, 如果 $k \in \mathcal{Q}$ 充分大, 利用在 $\{X^{(k)}\}$ 上 g_i 的连续性和 Dg_i 的有界性, 则存在与 k 无关的正常数 $\overline{c}$ 和 $\overline{a}$, 使得 $g_i(X^{(k)}) \leqslant -\overline{c}$ 和 $\langle Dg_i(X^{(k)}), \Delta X\rangle \leqslant \overline{a}$. 因此,

$$g_i(X^{(k)}) + \rho\langle Dg_i(X^{(k)}), \Delta X\rangle \leqslant -\overline{c} + \rho\overline{a}, \quad i \notin \mathcal{A}^\infty.$$

设 $\varphi = \min\{\frac{\overline{c}}{\overline{a}}, \frac{\lambda_{\min}^*}{2(\Upsilon - \xi_{\min}^*)}\}$, 如果 $k \in \mathcal{Q}$ 充分大, $\rho \in [\frac{h^{(k)}}{\varepsilon}, \varphi]$, 我们有

$$g_i(X^{(k)}) + \rho\langle Dg_i(X^{(k)}), \Delta X\rangle \leqslant 0, \quad i = 1, \ldots, m, \quad X^{(k)} + \rho\Delta X \succeq 0. \tag{3.20}$$

这样,

$$\begin{aligned} &\frac{1}{2}c^{(k)}\langle \Delta X^{(k)}, \Delta X^{(k)}\rangle + \langle Df(X^{(k)}), \Delta X^{(k)}\rangle + f(X^{(k)}) \\ &\quad + \alpha^{(k)} \sum_{i=1}^m \max\{0, g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X^{(k)}\rangle\} \\ &\quad \leqslant \frac{1}{2}c^{(k)}\rho^2 + \rho\langle Df(X^{(k)}), \Delta X\rangle + f(X^{(k)}) \end{aligned}$$

对所有充分大的 $k \in \mathcal{Q}$ 和 $\rho \in [\frac{h^{(k)}}{\varepsilon}, \varphi]$ 成立. 从上述不等式和 (3.16), 有

$$-\langle Df(X^{(k)}), \Delta X^{(k)}\rangle \geqslant -\rho\langle Df(X^{(k)}), \Delta X\rangle - \frac{1}{2}c^{(k)}\rho^2 \geqslant \rho\varepsilon - \frac{1}{2}c^{(k)}\rho^2.$$

故对所有充分大的 $k \in \mathcal{Q}$ (例如, $k \geqslant K_1$) 和 $\rho \in [\frac{h^{(k)}}{\varepsilon}, \min\{\varphi, \frac{\varepsilon}{c^{(k)}}\}]$,

$$\Delta l^{(k)} = -\langle Df(X^{(k)}), \Delta X^{(k)}\rangle \geqslant \frac{1}{2}\rho\varepsilon \tag{3.21}$$

和 (3.20) 成立. 同时, 由引理 3.4, 我们可以假定对所有满足 $K_1 \leqslant k \in \mathcal{Q}$ 的 k, $h^{(k)} \leqslant \frac{1}{2}$.

由 f 的二次连续可微性和 $\{X^{(k)}\}$ 的有界性, 我们有

$$f(X^{(k)} + \Delta X^{(k)}) = f(X^{(k)}) + \langle Df(X^{(k)}), \Delta X^{(k)}\rangle + \frac{1}{2}D^2 f(Y)(\Delta X^{(k)}, \Delta X^{(k)}),$$

其中 Y 位于 $X^{(k)}$ 和 $X^{(k)} + \Delta X^{(k)}$ 的线段上. 这样, 上式和 (3.21) 及 (A.2) 一起给出

$$\frac{\Delta f^{(k)}}{\Delta l^{(k)}} \geqslant 1 - \frac{2MT^2}{\rho\varepsilon(c^{(k)})^2},$$

其中 T 是序列 $\{c^{(k)}\|\Delta X^{(k)}\|_F\}$ 的上界. 设

$$T' := \frac{2MT^2}{\varepsilon(1-\sigma_1)},$$

则对于满足 $K_1 \leqslant k \in \mathcal{Q}$ 的所有 k 和满足

$$\rho \in \left[\max\left\{\frac{h^{(k)}}{\varepsilon}, \frac{T'}{(c^{(k)})^2}\right\}, \min\left\{\varphi, \frac{\varepsilon}{c^{(k)}}\right\}\right]$$

的 ρ, 下面的不等式

$$\Delta f^{(k)} \geqslant \sigma_1 \Delta l^{(k)} \tag{3.22}$$

成立.

另一方面, 由 g_i $(i = 1, \ldots, m)$ 的二次连续可微性, (A.2) 和算法 2.3 的 (S.2) 步, 我们有

$$\begin{aligned} g_i(X^{(k)} + \Delta X^{(k)}) &= g_i(X^{(k)}) + \langle Dg_i(X^{(k)}), \Delta X^{(k)}\rangle + \frac{1}{2}D^2 g_i(Y_i)(\Delta X^{(k)}, \Delta X^{(k)}) \\ &\leqslant \sigma_4\|\Delta X^{(k)}\|_F^2 + M\|\Delta X^{(k)}\|_F^2 \\ &\leqslant (M+\sigma_4)\frac{T^2}{(c^{(k)})^2}, \quad i = 1, \ldots, m, \end{aligned}$$

其中 Y_i 的定义类似于 Y. 于是, 我们有

$$h(X^{(k)} + \Delta X^{(k)}) \leqslant (M+\sigma_4)\frac{T^2}{(c^{(k)})^2}. \tag{3.23}$$

如果 $\frac{1}{c^{(k)}} \leqslant \sqrt{\frac{\beta}{(M+\sigma_4)T^2}}\sqrt{\tau^{(k)}}$, 那么从 (3.23) 得到 $h(X^{(k)} + \Delta X^{(k)}) \leqslant \beta\tau^{(k)}$. 这样, 当 $K_1 \leqslant k \in \mathcal{Q}$ 和 $c^{(k)}$ 充分大时,

$$\frac{1}{c^{(k)}} \leqslant \sqrt{\frac{\beta}{(M+\sigma_4)T^2}}\sqrt{\tau^{(k)}} \tag{3.24}$$

和

$$\max\left\{\frac{h^{(k)}}{\varepsilon}, \frac{T'}{(c^{(k)})^2}\right\} \leqslant \rho \leqslant \min\left\{\varphi, \frac{\varepsilon}{c^{(k)}}\right\} \tag{3.25}$$

成立. 因此, $h(X^{(k)} + \Delta X^{(k)}) \leqslant \beta\tau^{(k)}$ 和 $\Delta f^{(k)} \geqslant \sigma_1 \Delta l^{(k)}$. 从 (3.21) 和 (3.25) 得

$$\Delta l^{(k)} \geqslant \frac{1}{2}\rho\varepsilon \geqslant \frac{1}{2}h^{(k)},$$

注意到 $h^{(k)} \leqslant \frac{1}{2}$, $\sigma_3 < 1$, 于是, $\Delta l^{(k)} \geqslant \sigma_3 (h^{(k)})^2$. 进一步, 从 $\Delta f^{(k)} \geqslant \sigma_1 \Delta l^{(k)}$, $\Delta l^{(k)} \geqslant \frac{1}{2} h^{(k)}$ 及 $\sigma_1 > 2\gamma$ 得到 $f(X^{(k)} + \Delta X^{(k)}) \leqslant f^{(k)} - \gamma h^{(k)}$. 因此, 对所有使得 (3.24) 和 (3.25) 成立的 k ($K_1 \leqslant k \in \mathcal{Q}$) 和 $c^{(k)}$, f-型迭代的条件满足.

根据上述讨论, 我们在两种不同情形可以推出矛盾.

在情形 (I), 由于 $\mathcal{Q}$ 恰是主序列的尾部, 它由最后一个 h-型迭代后面的所有迭代组成, 于是, $\tau^{(k)}$ 是固定的, 存在 $j' \in N$ 使得对所有 $j \geqslant j'$, 不等式

$$\frac{1}{c_j^{(k)}} \leqslant \frac{1}{c_{\min}(\theta_1)^j} \leqslant \min\left\{\sqrt{\frac{\beta}{(M+\sigma_4)T^2}}\sqrt{\tau^{(k)}},\ \frac{\varphi}{\varepsilon}\right\}$$

成立. 另一方面, 存在 $j'' \in N$ 使得 $\forall j \geqslant j''$, 有 $\frac{T'}{c_{\min}(\theta_1)^{j-1}} < \varepsilon$, 因而对所有 $c_0^{(k)} \in [c_{\min}, c_{\max}]$,

$$\frac{T'}{(c_0^{(k)})^2(\theta_1)^{2j}} < \frac{1}{\theta_1} \cdot \frac{\varepsilon}{(c_0^{(k)})(\theta_1)^j}$$

成立.

设 $j^* = \max\{j', j''\}$, 则存在 $K_2 \in \mathbb{N}$, 使得对所有满足 $K_2 \leqslant k \in \mathcal{Q}$ 的 k, 不等式

$$\frac{h^{(k)}}{\varepsilon} < \frac{\varepsilon}{c_{\min}(\theta_1)^{j^*+1}}$$

成立. 再设 $K^* = \max\{K_1, K_2\}$, 则对所有满足 $K^* \leqslant k \in \mathcal{Q}$ 的 k, 参数

$$\rho = \frac{\varepsilon}{c_{j^*+1}^{(k)}}$$

满足 (3.24) 和 (3.25). 因此, 对这样的每一个主迭代 k, 存在内循环指标 $j \in \{0, 1, 2, \ldots, j^*+1\}$ 使得 f-型迭代在第 j 次内迭代产生. 这样, 从 (3.21) 和 (3.22) 可推出

$$\Delta f^{(k)} \geqslant \frac{1}{2}\sigma_1 \min\left\{\varphi, \frac{\varepsilon}{c_{\max}(\theta_1)^{j^*+1}}\right\} \tag{3.26}$$

对所有满足 $K^* \leqslant k \in \mathcal{Q}$ 的 k 成立. 另外, 由于 $\{X^{(k)}\}$ 的有界性, 对这样的 k, $f^{(k)}$ 是单调下降的, 下有界的, 因此我们得到 $\sum_{k\in\mathcal{Q}} \Delta f^{(k)}$ 是收敛的, 这与 (3.26) 矛盾.

最后, 在情形 (II), 我们考虑所有 h-型迭代的子序列. 因为对于 $k \in \mathcal{Q}$, 有 $h^{(k)} \to 0$, 因此 $\tau^{(k)} \to 0$, 并且存在 $\mathcal{Q}$ 的 h-型无穷子序列 $\mathcal{Q}'$ 满足 $\tau^{(k+1)} = h^{(k)} < \tau^{(k)}$. 类似于情形 (I), 存在 $j^\star, j^{\star\star} \in N$ 使得

$$\frac{\varepsilon}{c_{\min}(\theta_1)^j} \leqslant \varphi, \quad \forall j \geqslant j^\star,$$

和

$$\frac{T'}{c_{\min}(\theta_1)^{j-1}} \leqslant \varepsilon, \quad \forall j \geqslant j^{\star\star}.$$

设 $j^\circ = \max\{j^\star, j^{\star\star}\}$. 由于当 $k \in \mathcal{Q}'$ 和 $k \to \infty$ 时, $h^{(k)} \to 0$ 和 $\tau^{(k)} \to 0$. 因此, 存在 $k^* \in \mathcal{Q}'$, $k^* \geqslant K_1$, 使得

$$\sqrt{\frac{\beta}{(M+\sigma_4)T^2}}\sqrt{\tau^{(k^*)}} < \frac{1}{c_{\max}(\theta_1)^{j^\circ}}$$

和

$$\frac{h^{(k^*)}}{\varepsilon} < \frac{\varepsilon}{\theta_1}\sqrt{\frac{\beta}{(M+\sigma_4)T^2}}\sqrt{\tau^{(k^*)}}.$$

因此, 从 $\varepsilon, \frac{1}{\theta_1} < 1$, 存在 $j^{**} > j^{\circ}$ 使得

$$\frac{h^{(k^*)}}{\varepsilon^2} < \frac{1}{c_0^{(k^*)}(\theta_1)^{j^{**}}} \leqslant \sqrt{\frac{\beta}{(M+\sigma_4)T^2}}\sqrt{\tau^{(k^*)}}$$

成立, 于是 f-型迭代是采用的. 进一步, 对于 $j \in \{0, 1, \ldots, j^{**}\}$, 不可能产生一个 h-型步, 因为当 j 减少时 $\Delta l^{(k^*)}$ 单调上升. 这意味着主迭代 k^* 是 f-型迭代, 这与 $k^* \in \mathcal{Q}'$ 矛盾.

这样, 在情形 (I) 和情形 (II) 都得到了矛盾, 从而 X^{∞} 是一个 FJ 点. 再利用引理 3.8, 结论得到. 证明全部完成.

4 数值结果

为了将我们的算法与文献 [7] 中算法 4.1 进行比较, 我们用 Matlab 6.5 对这两个算法进行了编程计算, 并都采用文献 [30–32] 中的软件 SDPT3-Solver 4.0 解相应的子问题.

由于对非线性半定规划尚无标准的试验问题库, 于是, 我们像文献 [7] 中一样, 按照下述方法随机产生一系列试验问题: 我们预先定义解 $X^{\star}$ 为 $X^{\star} = A^{\top}A$, $A \in \mathbb{R}^{r\times n}$ 为随机生成的矩阵, $r \in \{1, \ldots, n\}$. 因此, $X^{\star}$ 是正半定的, 至少有 $n-r$ 个零特征值. 我们然后考虑如下含非线性约束的二次半定规划问题:

$$\min_{X\in S^n} \|X - X^{\star}\|_F^2 \qquad \text{s.t.} \quad g(X) \leqslant 0,\ X \succeq 0,$$

表 1 数值结果

		SL			Fil-SL		
(m,n)	r	ite	val	time (s)	ite	val	time (s)
(12,10)	8	7	4.2650e − 7	25.657	7	4.0753e − 7	44.354
		failure			32	2.1248e − 6	173.92
		32	8.9033e − 9	82.8090	9	1.5684e − 8	34.020
		8	3.4231e − 10	28.2610	6	2.7180e − 10	23.8864
		failure			failure		
	10	24	2.0844e − 10	88.8370	9	9.7521e − 12	46.256
		17	1.9348e − 10	67.3470	8	7.5167e − 11	41.510
		11	1.0988e − 10	34.0630	11	2.1715e − 12	35.4910
		23	6.5830e − 11	61.8190	16	3.2638e − 11	51.8150
		23	1.0020e − 10	66.586	12	3.0352e − 12	45.5350
(40,25)	15	366	7.7732e − 6	1.7229e + 3	31	7.7734e − 6	852.6260
		failure			failure		
		failure			24	1.2274e − 8	599.1210
		82	4.2865e − 10	1.1621e + 3	29	3.2807e − 10	507.089
		50	2.2503e − 4	660.309	13	2.2552e − 4	306.731
	20	38	1.4641e − 4	282.637	8	1.4641e − 4	189.913
		31	2.6326e − 8	532.976	20	2.829e − 8	444.419
		11	2.4234e − 9	211.945	13	2.8193e − 9	307.482
		46	8.2205e − 6	782.555	13	2.9009e − 7	302.756
		failure			56	5.5492e − 5	1.5354e + 3

续表

		SL			Fil-SL		
(m,n)	r	ite	val	time (s)	ite	val	time (s)
(40,30)	25	failure			failure		
		5	2.9965e − 9	156.315	7	1.445e − 7	308.514
		failure			7	8.7612e − 6	331.697
		46	1.6376e − 7	1.3331e + 3	21	1.6357e − 7	765.971
		failure			24	4.8783e − 7	909.515
	30	9	4.0429e − 12	266.393	15	4.792e − 13	569.038
		failure			25	4.4752e − 13	685.415
		309	8.0433e − 4	1.996e + 3	11	1.7267e − 7	388.689
		failure			39	2.7518e − 12	1.1465e + 3
		failure			failure		
(50,45)	35	failure			21	3.9676e − 5	2.115e + 3
		failure			failure		
		failure			21	5.9081e − 9	2.3072e + 3
		43	6.1546e − 4	3.8838e + 3	14	7.3508e − 6	1.8737e + 3
		28	1.390e − 4	2.7435e + 3	15	1.360e − 4	1.8769e + 3
	45	41	3.9085e − 11	3.6006e + 3	17	8.6475e − 5	2.0633e + 3
		19	4.1085e − 9	1.782e + 3	13	7.6185e − 4	1.8991e + 3
		28	5.1868e − 11	3.2226e + 3	25	1.7590e − 11	2.6439e + 3
		7	5.6282e − 6	671.625	10	3.0875e − 12	1.2754e + 3
		failure			12	5.4473e − 10	2.5786e + 3

其中 $g_i(X)\,(i\in\{1,\dots,m\})$ 是非线性项 $\exp(\cdot),\log(\cdot),\cos(\cdot),\sin(\cdot),\langle X,X\rangle,(\cdot)^k,\ k\neq 1$, 和线性项 $\langle X',\cdot\rangle,\mathrm{tr}(.),a*(\cdot)$ 的组合. 例如,

$$
\begin{aligned}
&g_1(X)=a*\cos(\mathrm{tr}(X))*\exp(\varepsilon*\langle X',X\rangle)+b*\langle X,X\rangle+c;\\
&g_2(X)=a*\log((\mathrm{tr}(X)^2))*\langle X,X\rangle+b*\langle X',X\rangle^2+c\langle X'',X\rangle+d;\\
&g_3(X)=a*\sin(\log(\langle X,X\rangle^2))+b*\exp(-\mathrm{tr}(X))+c;\\
&g_4(X)=a*\langle X',X\rangle/\langle X,X\rangle+b*(\mathrm{tr}(X))^{-1}+c*\langle X'',X\rangle+d;\\
&\cdots,
\end{aligned}
$$

其中 ε 取为固定的小的实数 $(1.0\mathrm{e}-10)$ 或 $(1.0\mathrm{e}-8)$, 对称系数矩阵 $X',X'',\dots$ 随机生成. 随机实系数 a,b,c,d 的选取使得 $X^\star$ 是可行的, 因此, 这个二次半定规划问题的最优解恰恰是 $X^\star$, 最优值是 0. 算法 2.3 中的参数设置如下:

$\alpha_0^{(0)}=50,\gamma=\min\{1.0\mathrm{e}-6,1/2n\},\beta=0.95,\theta_1=4,\theta_2=20,\theta_3=0.045,\varepsilon^{(0)}=0.05,\sigma_1=0.01,\sigma_2=0.5,\sigma_3=0.1,\sigma_4=1.0\mathrm{e}+8,\sigma_5=5,\sigma_6=30,\sigma_7=0.101,\sigma_8=0.05,\sigma_9=0.75,\sigma_{10}=0.04,C_{\min}=0.001,C_{\max}=100,c_0^{(0)}=1,X^{(0)}=I,h_0=u=\max\{1000,5*h(X^{(0)})\},f_0=-1.0\mathrm{e}+10$. 其中, I 表示适当阶数的单位矩阵.

如果 $\|\mathrm{svec}(\Delta X_j^{(k)})\|_\infty < 1.0\mathrm{e}-4$ 且 $h^{(k)} < 1.0\mathrm{e}-4$ 满足, 或者如果 SDPT3 解子问题失败, 或者如果迭代次数 $k > 500$, 算法 2.3 终止. 在数值结果表中, failure 代表最优值太大 ($\geqslant 1.0\mathrm{e}-3$), 或者 SDPT3 解子问题失败. ite 和 val 分别表示迭代次数和最优值. 对于 (n, m, r) 的每一个选择, 两个算法都采用随机生成数据执行 5 次 (见表 1). 从表中可知, 我们的算法是有竞争力的.

5 结束语

在本文中, 我们提出了解非线性半定规划的过滤集-逐次线性化方法, 它把过滤集技术推广到求解非线性半定规划问题. 我们证明了在适当的条件下, 新方法是总体收敛的. 初步的数值试验结果显示出新方法是有潜力和竞争力的. 进一步的更详细的数值研究工作仍有待进行.

致谢 作者对《中国科学》的编委和三位审稿人表示诚挚的谢意, 感谢他们对本文初稿提出的许多有价值的建议和评述, 使得本文的工作得到大大改进.

参考文献

1 Todd M J. Semidefinite optimization. *Acta Numer*, **10**: 515–560 (2001)
2 Vandenberghe L, Boyd S. Semidefinite programming. *SIAM Review*, **38**: 49–95 (1996)
3 Wolkowicz H, Saigal R, Vandenberghe L, eds. Handbook of Semidefinite Programming. Boston-Dordrecht-London: Kluwer Academic Publishers, 2000
4 Helmberg C. Semidefinite programming for combinatorial optimization. Semidefinite programming home page: http://www.zib.de/helmberg/semidef.html (2000)
5 Correa R, Ramirez H. A global algorithm for nonlinear semidefinite programming. *SIAM J Optim*, **15**: 303–318 (2004)
6 Jerre F. An interior method for nonconvex semidefinite programs. *Optim Eng*, **1**: 347–372 (2000)
7 Kanzow C, Nagel C, Kato H, et al. Successive linearization methods for nonlinear semidefinite programs. *Comput Optim Appl*, **31**: 251–273 (2005)
8 Sun D, Sun J, Zhang L. The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Math Program*, **114**: 349–391 (2008)
9 Sun J, Sun D, Qi L. A squared smoothing Newton method for nonsmooth matrix equations and its applications in semidefinite optimization problems. *SIAM J Optim*, **14**: 783–806 (2004)
10 Li C, Sun W. Some properties for nonconvex semidefinite programming. *Numer Math J Chinese Univ*, **30**: 184–192 (2008)
11 Li C, Sun W. A class of nonsmooth Newton-type methods for nonlinear semidefinite programming. *J Nanjing Norm Univ Nat Sci Ed*, **31**: 1–7 (2008)
12 Li C, Sun W. A filter-sequential semidefinite programming method for nonlinear semidefinite programming. Technical Report Opt-07-03, School of Mathematics and Computer Science, Nanjing Normal University, 2007. 中国科技论文在线, No. A200802-255, http://www.paper.edu.cn/paper.php?serial_number=200802-167
13 Nie J, Yuan Y. A potential reduction algorithm for an extended SDP problem. *Sci China Ser A*, **43**: 35–46 (2000)
14 Yamashita H, Yabe H, Harada K. A primal-dual interior point method for nonlinear semidefinite programming. Technical report, Department of Mathematical Information Science, Faculty of Science, Tokyo University of Science, 2007

15 Liu Z, Sun W. A filter primal-dual interior point method for nonlinear semidefinite programming. Technique report, School of Mathematics and Computer Science, Nanjing Normal University, 2008

16 Shapiro A. First and second order analysis of nonlinear semidefinite programs. *Math Program*, **77**: 301–120 (1997)

17 Sun D. The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Math Oper Res*, **31**: 761–776 (2006)

18 Bonnans J F, Shapiro A. Perturbation Analysis of Optimization Problems. New York: Springer, 2000

19 Bonnans J F, Commintti R, Shapiro A. Sensitivity analysis of optimization problem under second order regularity constraints. *Math Oper Res*, **23**: 803–832 (1998)

20 Xu D C, Xu S Z. Approximation bounds for quadratic maximization and max-cut problems with semidefinite programming relaxation. *Sci China Ser A*, **50**: 1583–1596 (2007)

21 Fletcher R, Leyffer S. Nonlinear programming without a penalty function. *Math Program*, **91**: 239–269 (2002)

22 Fletcher R, Gould N I M, Leyffer S, et al. Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming. *SIAM J Optim*, **13**: 635–659 (2002)

23 Fletcher R, Leyffer S, Toint Ph L. On the global convergence of an SLP-filter algorithm. Numerical Analysis Report NA/183, Department of Mathematics, University of Dundee, Scotland, 1998

24 Fletcher R, Leyffer S, Toint Ph L. On the global convergence of a filter-SQP algorithm. *SIAM J Optim*, **13**: 44–59 (2002)

25 Gould N I M, Sainvitu C, Toint Ph L. A filter-trust-region method for unconstrained optimization. *SIAM J Optim*, **16**: 341–357 (2005)

26 Chen Y N, Sun W. A dwindling filter line-search method for unconstrained optimization. Technical Report Opt-06-11, School of Mathematics and Computer Science, Nanjing Normal University, 2006

27 Miao W. On Filter and Nonmonotone Techniques in Numerical Optimization. PhD dissertation. School of Mathematics and Computer Science, Nanjing Normal University, 2006

28 Miao W, Sun W. A filter-trust-region method for unconstrained optimization. *Numer Math J Chinese Univ*, **29**: 88–96 (2007)

29 Sun W. On filter-type methods for optimization: motivation and development. An invited talk, 4th Sino-Japanese Optimization Meeting (SJOM2008), August 26–31, 2008. Cheng Kung University, Taiwan.

30 Toh K C, Tutuncu R H, Todd M J. SDPT3 version 4.0 (beta) – a MATLAB software for semidefinite-quadratic-linear programming. http://www.math.nus.edu.sg/~mattohkc/sdpt3.html, updated in 17 July, 2006

31 Toh K C, Tutuncu R H, Todd M J. On the implementation and usage of SDPT3–a MATLAB software package for semidefinite-quadratic-linear programming version 4.0 (Draft). http://www.math.nus.edu.sg/~mattohkc/guide4-0-draft.pdf, 17 July, 2006

32 Tutuncu R H, Toh K C, Todd M J. Solving semidefinite quadratic-linear programs using SDPT3. *Math Program*, **95**: 189–217 (2003)

33 Sun W, Yuan Y. Optimization Theory and Methods: Nonlinear Programming. New York: Springer, 2006

34 Sun W, Zhou Q Y. An unconstrained optimization method using nonmonotone second order Goldstein's line search. *Sci China Ser A*, **50**: 1389–1400 (2007)

35 Ke X W, Han J Y. A nonmonotone trust region algorithm for equality constrained optimization. *Sci China Ser A*, **38**: 683–695 (1995)

36 Ben-Tal A, Nemirovski A. Lectures On Modern Convex Optimization – Analysis, Algorithms, and Engineering Applications. MPS-SIAM Series on Optimization. Philadelphia: SIAM, 2001

37 Bazaraa M S, Shetty C M. Nonlinear Programming: Theory and Algorithms. New York: John Wiley and Sons, 1979

38 Borchers B. SDPLIB 1.2, a library of semidefinite programming test problems. *Optim Methods Softw*, **11**: 597–611 (1999)