AMO-Advanced Modeling and Optimization

Volume 9, Number 1, 2007

Projection onto the solution set of semidefinite programs

Anhua Lin[*][†]

Email: alin@mtsu.edu

Department of Mathematical Sciences

Middle Tennessee State University

Murfreesboro, TN 37132, US

Abstract

Just as linear programs, a semidefinite program may have many solutions. In [Y.-B. Zhao and D. Li *SIAM J. Optim.* 12(4):893-912, 2002], a path-following method was proposed to project the origin onto the optimal solution set of a linear program, i.e., to find the least-2-norm solution of the linear program. In this paper we generalize this method to project any vector onto the optimal solution set of any semidefinite program.

Keywords: semidefinite program, projection, path-following algorithm

1 Introduction

Let S^n denote the space of symmetric real n by n matrices. Let S^n_+ be the cone of positive semidefinite symmetric matrices. Related to S^n_+ we define the partial ordering $\succeq$ via

$$A \succeq B \Leftrightarrow B \preceq A \Leftrightarrow A - B \in S^n_+, \forall A, B \in S^n_+.$$

[*]Partially supported by Middle Tennessee State University Summer and Academic Year Research Grant, 2006-2007

[†]Final version: April 24, 2007

We denote $A \succ 0$ or $0 \prec A$ if $A \in S^n_{++} \subset S^n$, the set of symmetric positive definite matrices. For any matrix/vector space, we use $\langle \cdot, \cdot \rangle$ to denote the usual inner-product, i.e., $\langle u, v \rangle = tr(u^T v)$. This inner-product induces the Euclidean-2-norm as $\|u\| = \sqrt{\langle u, u \rangle}$ on the corresponding space.

In this paper, we consider semidefinite program (SDP) in its standard primal-dual formulation:

$$
\begin{array}{lll}
\text{Primal Problem:} & & \\
\quad \text{Minimize} & \langle C, X \rangle & \\
\quad \text{subject to} & \mathcal{A}(X) := \begin{bmatrix} \langle A^1, X \rangle \\ \cdot \\ \cdot \\ \cdot \\ \langle A^i, X \rangle \\ \cdot \\ \cdot \\ \cdot \\ \langle A^m, X \rangle \end{bmatrix} = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_i \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix} = b, & \\
 & X \in S^n_+, & \\
\text{Dual Problem:} & & \\
\quad \text{Maximize} & \langle b, y \rangle & \\
\quad \text{subject to} & y \in R^m, \mathcal{A}^*(y) := \sum_{i=1}^m y_i A^i \preceq C, &
\end{array}
\qquad (1)
$$

where $C \in S^n$, $b \in R^m$, $\mathcal{A}(\cdot)$ is a linear operator from S^n to R^m defined by a linearly independent set $\{A^1, A^2, \ldots, A^m\} \subset S^n$, and $\mathcal{A}^*(\cdot)$ is the adjoint operator of $\mathcal{A}(\cdot)$.

It is well known that SDP is one of the most important generalization of linear programs (LP). A linear program may have many solutions. Much research has been done to find the least-2-norm solution of an LP, i.e., to project the origin onto the optimal solution set of the LP, see [Kanzow, 2003; Lucidi, 1987; Mangasarian, 1983; Mangasarian, 2004; Zhao and Li, 2002] for some algorithms for this regard. The one that is most relevant to this paper is a nice path-following method developed in [Zhao and Li, 2002].

Let S_p and S_d denote the primal solution set and dual solution set of the SDP (1), respectively.

If $S_p \neq \emptyset$ and $S_d \neq \emptyset$, then we say that SDP (1) is solvable. We only deal with solvalbe SDP in this paper. Just as LP, a solvable SDP may also have many solutions. For any $(Q, q) \in S^n \times R^m$, we let (X_Q, y_q) denote the projection of (Q, q) onto $S_p \times S_d$, i.e., $X_Q = \text{argmin}_{X \in S_p} \|X - Q\|$ and $y_q = \text{argmin}_{y \in S_d} \|y - q\|$. Notice that since S_p and S_q are both convex sets, X_Q and y_q are uniquely defined. In particular, (X_0, y_0) is just the least-2-norm primal-dual optimal solution pair. There is currently little existent research on finding the least-2-norm solution of an SDP. The purpose of this paper is fill this gap by generalizing the method of [Zhao and Li, 2002] to SDP. We will show that our algorithm can actually find (X_Q, y_q) for any $(Q, q) \in S^n \times R^m$.

Now we review some crucial concepts and facts about SDP.

The duality gap gap_{pd} is defined to be $\langle C, X\rangle - \langle b, y\rangle = \langle X, C - \mathcal{A}^*(y)\rangle$ where $X \in S_p$ and $y \in S_d$. Clearly $gap_{pd} \geq 0$ which is the weak-duality. When $gap_{pd} = 0$, then we say that the strong duality holds. If there are $X \in S^n_{++}$ and $y \in R^m$ such that $\mathcal{A}(X) = b$ and $S = C - \mathcal{A}^*(y) \succ 0$, then we say the SDP (1) is strictly feasible. It is well known that strict feasibility is a sufficient but not necessary condition for strong duality to hold.

If strong duality holds, then for any $(X, y) \in S_p \times S_d$ and $S = C - \mathcal{A}^*(y)$, we have

$$\begin{cases} XS = 0, \\ \mathcal{A}(X) = b, \\ \mathcal{A}^*(y) + S = C, \\ X, S \in S^n_+, y \in R^m. \end{cases} \tag{2}$$

On the other hand, if (X, y, S) satisfies (2), then strong duality holds and $(X, y) \in S_p \times S_d$. Most, if not all, algorithms for SDP require the strong duality assumption and try to find (X, y, S) approximately satisfying (2). Many of them actually need the strict feasibility assumption, especially those based on the central path.

The concept of central path lies at the heart of the study of semidefinite programs. This path is defined as the solution to the parametric nonlinear system:

$$\begin{cases} XS = \mu I, \\ \mathcal{A}(X) = b, \\ \mathcal{A}^*(y) + S = C, \\ X, S \in S^n_{++}, y \in R^m, \end{cases} \tag{3}$$

where $\mu > 0$ is the path parameter and I is the identity matrix. It is well known that strict feasibility holds if and only if for all $\mu > 0$, system (3) has a unique solution $(X(\mu), y(\mu), S(\mu))$ [Wolkowicz, Saigal and Vandenberghe, 2000]. As μ varies in R_{++}, the solutions form the central path. The most important property of the central path is that as $\mu \longrightarrow 0+$, $(X(\mu), y(\mu))$ converges to a primal-dual solution pair. Based on this property, a lot of so-called path-following algorithms [Wolkowicz, Saigal and Vandenberghe, 2000; Monteiro, 1997; Sturm and Zhang, 1998] have been developed to follow the central path to a solution of the SDP (1).

We are going to follow a different path. Let $(Q, q) \in S^n \times R^m$ be given. For any $\mu \geq 0$, $X, S \in S^n_+$, and $y \in R^m$, we define

$$F_\mu(X, y, S) := \begin{pmatrix} S^{\frac{1}{2}} X S^{\frac{1}{2}} - \mu I \\ -\mu^p (X - Q) + \mathcal{A}^*(y) + S - C \\ \mathcal{A}(X) + \mu^p (y - q) - b \end{pmatrix}.$$

Under the assumption of strong duality, it can be shown that for any $\mu > 0$, the system $F_\mu(X, y, S) = 0$ has a unique solution $(X(\mu), y(\mu), S(\mu))$, and as $\mu \longrightarrow 0+$, $(X(\mu), y(\mu)) \longrightarrow (X_Q, y_q)$. Although this property of the path motivates this research, its proof is not needed for our algorithm. We refer interested readers to [Lin, 2006]. However, we do need the assumption of strong duality for our analysis.

The paper is organized as follows. In Section 2 we describe the algorithm. In Section 3 we study the feasibility of the algorithm. Then in Section 4 we prove the convergence of the algorithm.

2 Algorithm

For any nonsingular matrix $T \in S^n$, we define $H_T(\cdot) : R^{n\times n} \to S^n$ as

$$H_T(M) := \frac{1}{2}\left(TMT^{-1} + T^{-1}M^tT\right),$$

where M^t denotes the transpose of M.

To measure the distance to the path, we define the "norm"of $F_\mu(X, y, S)$ as

$$\begin{aligned} & \|F_\mu(X,y,S)\| \\ := \quad & \max\left\{\left\|S^{\frac{1}{2}}XS^{\frac{1}{2}} - \mu I\right\|, \|-\mu^p(X-Q) + \mathcal{A}^*(y) + S - C\|, \|\mathcal{A}(X) + \mu^p(y-q) - b\|\right\}. \end{aligned}$$

All the iterates of the algorithm will be confined in a neighborhood of the path:

$$\mathcal{N}_\beta(\mu) := \{(X,y,S)|X \in S^n_+, y \in R^m, S \in S^n_+, \|F_\mu(X,y,S)\| \le \beta\mu\|\},$$

where $\beta \in (0,1)$ is a pre-picked constant.

Now we present the path-following method.

Algorithm 1 *1. Pick four numbers p, β, δ, and θ from $(0,1)$, set $k = 0$.*

Find $\mu_0 > 0$, $X^0 \succ 0$, $y^0 \in R^m$, $S^0 \succ 0$ such that $(X^0, y^0, S^0) \in \mathcal{N}_\beta(\mu_0)$.

2. At the k-th iteration, we have $(X^k, y^k, S^k) \in \mathcal{N}_\beta(\mu_k)$.

For simplicity, let $(X, y, S, \mu) = (X^k, y^k, S^k, \mu_k)$.

- *If $F_\mu(X, y, S) = 0$, then set $\alpha_k = 0$;*
- *otherwise, solve the Newton system of $F_\mu(X, y, S) = 0$ at (X, y, S) for $(\Delta X, \Delta y, \Delta S)$:*

$$\begin{cases} S^{\frac{1}{2}}(\Delta X)S^{\frac{1}{2}} + \frac{1}{2}S^{\frac{1}{2}}X(\Delta S)S^{-\frac{1}{2}} + \frac{1}{2}S^{-\frac{1}{2}}(\Delta S)XS^{\frac{1}{2}} &= \mu I - S^{\frac{1}{2}}XS^{\frac{1}{2}}, \\ -\mu^p(\Delta X) + \mathcal{A}^*(\Delta y) + \Delta S &= -(-\mu^p(X-Q) + \mathcal{A}^*(y) + S - C), \\ \mathcal{A}(\Delta X) + \mu^p(\Delta y) &= -(\mathcal{A}(X) + \mu^p(y-q) - b), \end{cases}$$

which is equivalent to (by multiplying $\sqrt{2}S^{\frac{1}{2}}$ to the first equation from both sides)

$$\left\{\begin{array}{rcl} 2S(\Delta X)S + SX(\Delta S) + (\Delta S)XS &=& r_c, \\ -\mu^p(\Delta X) + \mathcal{A}^*(\Delta y) + \Delta S &=& r_d, \\ \mathcal{A}(\Delta X) + \mu^p(\Delta y) &=& r_p, \end{array}\right. \tag{4}$$

where $r_c = 2\mu S - 2SXS$, $r_d = -(-\mu^p(X-Q) + \mathcal{A}^*(y) + S - C)$, and

$r_p = -(\mathcal{A}(X) + \mu^p(y-q) - b)$.

Let $\alpha_k = \min\left\{\frac{(1-\delta)\|F_\mu(X,y,S)\|}{\left\|H_{S^{\frac{1}{2}}}(\Delta X \Delta S)\right\|}, 1\right\} > 0$, *and* $(\Delta X^k, \Delta y^k, \Delta S^k) = (\Delta X, \Delta y, \Delta S)$.

Set $(X^{k+1}, y^{k+1}, S^{k+1}) = (X^k, y^k, S^k) + \alpha_k(\Delta X^k, \Delta y^k, \Delta S^k)$.

3. *Let γ^k be the first one among θ, θ^2, θ^3,..., satisfying*

$$(X^{k+1}, y^{k+1}, S^{k+1}) \in \mathcal{N}_\beta((1-\gamma^k)\mu_k),$$

i.e., $\left\|F_{(1-\gamma^k)\mu_k}(X^{k+1}, y^{k+1}, S^{k+1})\right\| \le \beta(1-\gamma^k)\mu_k$.

Set $\mu_{k+1} = (1-\gamma^k)\mu_k$.

4. *Set $k = k+1$, and go back to step 2.*

3 Feasibility of the algorithm

In this section we discuss how to perform every step of the algorithm. First we give a simple technical lemma which will be used several times later.

Lemma 3.1 *If $M \in S^n$, $\beta \in (0,1)$, $\mu > 0$, and $\|M - \mu I\| \le \beta\mu$, then $M \succ 0$.*

Proof. Let $\{\lambda_i | i = 1, \ldots, n\}$ be the n real eigenvalues of M. We have

$$\begin{array}{rcl} \beta\mu &\ge& \|M - \mu I\| \\ &=& \sqrt{\sum_{i=1}^n (\lambda_i - \mu)^2} \\ &\ge& |\lambda_i - \mu_i| \quad \text{for each } i, \\ &\ge& \mu - \lambda_i \quad \text{for each } i. \end{array}$$

Therefore $\lambda_i \geq (1-\beta)\mu > 0$ for each i. Hence $M \succ 0$. ■

3.1 Step 1

There are many ways to find $\mu_0 > 0$, $X^0 \succ 0$, $y^0 \in R^m$, and $S^0 \succ 0$ such that $(X^0, y^0, S^0) \in \mathcal{N}_\beta(\mu_0)$.

For example, let $\mu_0 = \max\left\{1, \left(\frac{\|\mathcal{A}(I)\|+\|b\|}{\beta}\right)^{\frac{2}{1+p}}, \left(\frac{\|Q\|+\|\mathcal{A}^*(q)-C\|}{\beta}\right)^{\frac{1}{1-p}}\right\}$, $X^0 = \mu_0^{\frac{1-p}{2}} I$, $y^0 = q$, and $S^0 = \mu_0^{\frac{1+p}{2}} I$. Clearly we have $\mu_0 > 0$, $X^0 \succ 0$, $y^0 \in R^m$, $S^0 \succ 0$, and

$$F_{\mu_0}(X^0, y^0, S^0) = \begin{pmatrix} (S^0)^{\frac{1}{2}} X^0 (S^0)^{\frac{1}{2}} - \mu_0 I \\ -\mu_0^p (X^0 - Q) + \mathcal{A}^*(y^0) + S^0 - C \\ \mathcal{A}(X^0) + \mu_0^p (y^0 - q) - b \end{pmatrix} = \begin{pmatrix} 0 \\ \mu_0^p Q + \mathcal{A}^*(q) - C \\ \mu_0^{\frac{1-p}{2}} \mathcal{A}(I) - b \end{pmatrix}.$$

Since $\mu_0 \geq 1$, then $\left\|\mu_0^{\frac{1-p}{2}} \mathcal{A}(I) - b\right\| \leq \mu_p^{\frac{1-p}{2}} \|\mathcal{A}(I)\| + \|b\| \leq \mu_0^{\frac{1-p}{2}} (\|\mathcal{A}(I)\| + \|b\|) \leq \beta \mu_0^{\frac{1+p}{2}} \mu_0^{\frac{1-p}{2}} = \beta\mu_0$. Similarly, $\|\mu_0^p Q + \mathcal{A}^*(q) - C\| \leq \mu_0^p(\|Q\| + \mu_0^{-p}\|\mathcal{A}^*(q) - C\|) \leq \mu_0^p(\|Q\| + \|\mathcal{A}^*(q) - C\|) \leq \beta\mu_0^p \mu_0^{1-p} = \beta\mu_0$. Hence $\|F_{\mu_0}(X^0, y^0, S^0)\| \leq \beta\mu_0$, so $(X^0, y^0, S^0) \in \mathcal{N}_\beta(\mu_0)$.

3.2 Step 2

Let $(X, y, S, \mu) = \left(X^k, y^k, S^k, \mu_k\right)$, $(X^+, y^+, S^+, \mu_+) = \left(X^{k+1}, y^{k+1}, S^{k+1}, \mu_{k+1}\right)$, and $H(\cdot) = H_{S^{\frac{1}{2}}}(\cdot)$.

We will use two matrix operators extensively in this section.

- $vec(\cdot)$: for any matrix M, $vec(M)$ denotes the vector obtained from stacking the columns of M one by one from the first to the last. Clearly, if the dimension of M is given, then it is very easy to get M from $vec(M)$, and vice versa.
- The Kronecker product $\otimes$: given $A, B \in R^{n\times n}$, $A \otimes B = [a_{ij}B] \in R^{n^2\times n^2}$.

These two operators have many useful properties. We will need the following (see chapter 4 of [Horn and Johnson, 1994], or the appendix of [Zhang,1998]):

1. $vec(AXB) = (B^t \otimes A) vec(X)$,
2. $(A \otimes B)^t = A^t \otimes B^t$,
3. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$,

4. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.

5. If λ_i's and ξ_j's are the eigenvalues of A and B, respectively, for $i, j = 1, 2, \ldots, n$, then the eigenvalues of $A \otimes B$ are all $\lambda_i \xi_j$'s.

Let $A = [vec(A_1), vec(A_2), \ldots, vec(A_m)]^t$, then the Newton system (4) is equivalent to

$$\begin{cases} 2vec(S(\Delta X)S) + vec(SX(\Delta S)) + vec((\Delta S)XS) & = vec(r_c), \\ -\mu^p vec(\Delta X) + A^t(\Delta y) + vec(\Delta S) & = vec(r_d), \\ A(vec(\Delta X)) + \mu^p(\Delta y) & = r_p. \end{cases} \tag{5}$$

Using $\otimes$, we can further rewrite (5) as

$$\begin{cases} (2S \otimes S)vec(\Delta X) & & +(I \otimes (SX) + (SX) \otimes I)vec(\Delta S) & = vec(r_c), \\ -\mu^p vec(\Delta X) & +A^t \Delta y & +vec(\Delta S) & = vec(r_d), \\ Avec(\Delta X) & +\mu^p \Delta y & & = r_p. \end{cases} \tag{6}$$

Let $E = 2S \otimes S$ and $F = I \otimes (SX) + (SX) \otimes I$, then we have $E^{-1} = \frac{1}{2} S^{-1} \otimes S^{-1}$ and $E^{-1}F = \frac{1}{2}(S^{-1} \otimes X + X \otimes S^{-1})$. It is easy to check that both E and $E^{-1}F$ are symmetric, and all their eigenvalues are positive, in other words, $E^{-1} \succ 0$ and $E^{-1}F \succ 0$.

Let $G^k = \begin{bmatrix} E & 0 & F \\ -\mu^p I & A^t & I \\ A & \mu^p I & 0 \end{bmatrix}$, then finally system (6) is equivalent to

$$G^k \begin{pmatrix} vec(\Delta X) \\ \Delta y \\ vec(\Delta S) \end{pmatrix} = \begin{pmatrix} vec(r_c) \\ vec(r_d) \\ r_p \end{pmatrix}. \tag{7}$$

The next lemma ensures that the Newton system is solvable.

Lemma 3.2 *G^k is nonsingular.*

Proof.

Suppose $G^k \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} Eu + Fw \\ -\mu^p u + A^t v + w \\ Au + \mu^p v \end{pmatrix} = 0$, then $Eu + Fw = 0 \Longrightarrow u = -E^{-1}Fw$, and $Au + \mu^p v = 0 \Longrightarrow u^t A^t = -\mu^p v^t$.

Therefore from $-\mu^p u + A^t v + w = 0$ we have

$$u^t(-\mu^p u + A^t v + w) = -\mu^p \|u\|^2 - \mu^p \|v\|^2 - w^t(E^{-1}F)w = 0,$$

hence $u = 0$, $v = 0$, and $w = 0$. So G^k is nonsingular. ■

Now we know that the solution $(\Delta X, \Delta y, \Delta S)$ to the system (6) (and (5)) exists and is unique. On the other hand, it is straightward to check that $((\Delta X)^t, \Delta y, (\Delta S)^t)$ is also a solution to system (5), so we must have $\Delta X = (\Delta X)^t$ and $\Delta S = (\Delta S)^t$. Therefore both ΔX and ΔS are symmetric. So $X(\alpha) := X + \alpha\Delta X \in S^n$ and $S(\alpha) := S + \alpha\Delta S \in S^n$.

The following lemma is fundamental for our analysis.

Lemma 3.3 *For all* $\alpha \in [0, \alpha_k]$, $\|H(X(\alpha)S(\alpha)) - \mu I\| \leq (1 - \delta\alpha)\|F_\mu(X, y, S)\|$.

Proof. We have

$$\begin{aligned}
& H(X(\alpha)S(\alpha)) - \mu I \\
= \ & H((X + \alpha\Delta X)(S + \alpha\Delta S)) - \mu I \\
= \ & H(XS + \alpha X\Delta S + \alpha\Delta XS + \alpha^2\Delta X\Delta S) - \mu I \\
= \ & H((1-\alpha)XS + \alpha(XS + X\Delta S + \Delta XS) + \alpha^2\Delta X\Delta S) - \mu I \\
= \ & (1-\alpha)H(XS) + \alpha H(XS + X\Delta S + \Delta XS) + \alpha^2 H(\Delta X\Delta S) - \mu I \\
= \ & (1-\alpha)S^{\frac{1}{2}}XS^{\frac{1}{2}} + \alpha\left(S^{\frac{1}{2}}XS^{\frac{1}{2}} + \frac{1}{2}S^{\frac{1}{2}}X(\Delta S)S^{-\frac{1}{2}} + \frac{1}{2}S^{-\frac{1}{2}}(\Delta S)XS^{\frac{1}{2}} + S^{\frac{1}{2}}(\Delta X)S^{\frac{1}{2}}\right) \\
& +\alpha^2 H(\Delta X\Delta S) - \mu I \\
= \ & (1-\alpha)S^{\frac{1}{2}}XS^{\frac{1}{2}} + \alpha\mu I - \mu I + \alpha^2 H(\Delta X\Delta S) \\
= \ & (1-\alpha)S^{\frac{1}{2}}XS^{\frac{1}{2}} - \mu(1-\alpha)I + \alpha^2 H(\Delta X\Delta S) \\
= \ & (1-\alpha)\left(S^{\frac{1}{2}}XS^{\frac{1}{2}} - \mu I\right) + \alpha^2 H(\Delta X\Delta S).
\end{aligned}$$

Therefore

$$\begin{aligned}
& \|H(X(\alpha)S(\alpha)) - \mu I\| \\
\leq\ & (1-\alpha)\left\|S^{\frac{1}{2}}XS^{\frac{1}{2}} - \mu I\right\| + \alpha^2\|H(\Delta X \Delta S)\| \\
\leq\ & (1-\alpha)\|F_\mu(X,y,S)\| + \alpha^2\|H(\Delta X \Delta S)\| \\
=\ & (1-\delta\alpha)\|F_\mu(X,y,S)\| + (\delta\alpha - \alpha)\|F_\mu(X,y,S)\| + \alpha^2\|H(\Delta X \Delta S)\| \\
=\ & (1-\delta\alpha)\|F_\mu(X,y,S)\| - \alpha((1-\delta)\|F_\mu(X,y,S)\| - \alpha\|H(\Delta X \Delta S)\|) \\
=\ & (1-\delta\alpha)\|F_\mu(X,y,S)\| - \alpha\|H(\Delta X \Delta S)\|\left(\frac{(1-\delta)\|F_\mu(X,y,S)\|}{\|H(\Delta X \Delta S)\|} - \alpha\right) \\
\leq\ & (1-\delta\alpha)\|F_\mu(X,y,S)\| - \alpha(\alpha_k - \alpha)\|H(\Delta X \Delta S)\| \\
\leq\ & (1-\delta\alpha)\|F_\mu(X,y,S)\|.
\end{aligned}$$

■

A simple but important consequence of Lemma 3.3 is the positive definiteness of $X(\alpha)$ and $S(\alpha)$.

Theorem 3.4 *For all $\alpha \in [0, \alpha_k]$, we have $X(\alpha) \succ 0$, and $S(\alpha) \succ 0$. In particular, $X^+ \succ 0$ and $S^+ \succ 0$.*

Proof. Using Lemma 3.3 and the fact that $(X, y, S) \in \mathcal{N}_\beta(\mu)$ we have

$$\begin{aligned}
\|H(X(\alpha)S(\alpha)) - \mu I\| &\leq (1-\delta\alpha)\|F_\mu(X,y,S)\| \\
&\leq (1-\delta\alpha)\beta\mu \\
&\leq \beta\mu.
\end{aligned}$$

Since $H(X(\alpha)S(\alpha)) \in S^n$, then Lemma 3.1 gives $H(X(\alpha)S(\alpha)) \succ 0$.

Now using (15) from [Monteiro, 1997] we know that $X(\alpha)S(\alpha)$ is nonsingular for all $\alpha \in [0, \alpha_k]$. Hence $X(\alpha)$ and $S(\alpha)$ are nonsingular symmetric matrices for all $\alpha \in [0, \alpha_k]$.

Because $X(0) \succ 0$, $S(0) \succ 0$, and the eigenvalues for $X(\alpha)$ $S(\alpha)$ are continuous functions of α, we must have $X(\alpha) \succ 0$ and $S(\alpha) \succ 0$ for all $\alpha \in [0, \alpha_k]$.

■

Now we show that (X^+, y^+, S^+) is closer to $(X(\mu), y(\mu), S(\mu))$ than (X, y, S) is in the sense of having a smaller $\|F_\mu\|$.

Theorem 3.5 $\|F_\mu(X^+, y^+, S^+)\| \le (1 - \delta\alpha_k)\|F_\mu(X, y, S)\|$.

Proof. By setting $\alpha = \alpha_k$ in Lemma 3.3 we get

$$\|H(X^+S^+) - \mu I\| \le (1 - \delta\alpha_k)\| F_\mu(X, y, S)\|.$$

Letting $D = (S^+)^{\frac{1}{2}} X^+ (S^+)^{\frac{1}{2}}$ and $W = S^{\frac{1}{2}}(S^+)^{-\frac{1}{2}}$, we have

$$\begin{aligned} H\left(X^+S^+\right) &= \frac{1}{2}\left(S^{\frac{1}{2}} X^+ S^+ S^{-\frac{1}{2}} + S^{-\frac{1}{2}} S^+ X^+ S^{\frac{1}{2}}\right) \\ &= \frac{1}{2}\left(S^{\frac{1}{2}}(S^+)^{-\frac{1}{2}}(S^+)^{\frac{1}{2}} X^+ (S^+)^{\frac{1}{2}} (S^+)^{\frac{1}{2}} S^{-\frac{1}{2}} + S^{-\frac{1}{2}}(S^+)^{\frac{1}{2}}(S^+)^{\frac{1}{2}} X^+ (S^+)^{\frac{1}{2}}(S^+)^{-\frac{1}{2}} S^{\frac{1}{2}}\right) \\ &= \frac{1}{2}\left(WDW^{-1} + (WDW^{-1})^t\right). \end{aligned}$$

So $H(X^+S^+) - \mu I = \frac{1}{2}\left(W(D - \mu I)W^{-1} + (W(D - \mu I)W^{-1})^t\right)$.

Using (21) from [Monteiro, 1997] we have

$$\begin{aligned} \left\|(S^+)^{\frac{1}{2}} X^+ (S^+)^{\frac{1}{2}} - \mu I\right\| &= \|D - \mu I\| \\ &\le \frac{1}{2}\|W(D - \mu I)W^{-1} + (W(D - \mu I)W^{-1})^t\| \\ &= \|H(X^+S^+) - \mu I\| \\ &\le (1 - \delta\alpha_k)\|F_\mu(X, y, S)\|. \end{aligned}$$

On the other hand, since

$$\begin{aligned} & -\mu^p(X^+ - Q) + \mathcal{A}^*(y^+) + S^+ - C \\ = \; & -\mu^p(X - Q + \alpha_k\Delta X) + \mathcal{A}^*(y + \alpha_k\Delta y) + (S + \alpha_k\Delta S) - C \\ = \; & (-\mu^p(X - Q) + \mathcal{A}^*(y) + S - C) + \alpha_k(-\mu^p\Delta X + \mathcal{A}^*(\Delta y) + \Delta S) \\ = \; & (1 - \alpha_k)(-\mu^p(X - Q) + \mathcal{A}^*(y) + S - C), \end{aligned}$$

and

$$
\begin{aligned}
\mathcal{A}(X^+) + \mu^p(y^+ - q) - b &= \mathcal{A}(X + \alpha_k \Delta X) + \mu^p(y + \alpha_k \Delta y - q) - b \\
&= (\mathcal{A}(X) + \mu^p(y - q) - b) + \alpha_k(\mathcal{A}(\Delta X) + \mu^p \Delta y) \\
&= (1 - \alpha_k)(\mathcal{A}(X) + \mu^p(y - q) - b),
\end{aligned}
$$

we have

$$
\begin{aligned}
\| - \mu^p(X^+ - Q) + \mathcal{A}^*(y^+) + S^+ - C\| &= (1 - \alpha_k)\| - \mu^p(X - Q) + \mathcal{A}^*(y) + S - C\| \\
&\leq (1 - \delta\alpha_k)\| - \mu^p(X - Q) + \mathcal{A}^*(y) + S - C\| \\
&\leq (1 - \delta\alpha_k)\|F_\mu(X, y, S)\|,
\end{aligned}
$$

and

$$
\begin{aligned}
\|\mathcal{A}(X^+) + \mu^p(y^+ - q) - b\| &= (1 - \alpha_k)\|\mathcal{A}(X) + \mu^p(y - q) - b\| \\
&\leq (1 - \delta\alpha_k)\|\mathcal{A}(X) + \mu^p(y - q) - b\| \\
&\leq (1 - \delta\alpha_k)\|F_\mu(X, y, S)\|.
\end{aligned}
$$

Therefore $\|F_\mu(X^+, y^+, S^+)\| \leq (1 - \delta\alpha_k)\|F_\mu(X, y, S)\|$.

■

3.3 Step 3

In step 3, we try to reduce μ. Again, we let $(X, y, S, \mu) = (X^k, y^k, S^k, \mu_k)$ and $(X^+, y^+, S^+, \mu^+) = (X^{k+1}, y^{k+1}, S^{k+1}, \mu_{k+1})$. We also define $d(\gamma) := \|F_{(1-\gamma)\mu}(X^+, y^+, S^+)\| - (1 - \gamma)\beta\mu$. Clearly $d(\gamma)$ is a continuous function. If $\alpha_k = 0$, then $\|F_\mu(X, y, S)\| = 0$ and

$$
\begin{aligned}
d(0) &= \|F_\mu(X^+, y^+, S^+)\| - \beta\mu \\
&= \|F_\mu(X, y, S)\| - \beta\mu \\
&= -\beta\mu
\end{aligned}
$$

$$< \quad 0;$$

if $\alpha_k > 0$, then

$$\begin{aligned} d(0) &= \|F_\mu(X^+, y^+, S^+)\| - \beta\mu \\ &\leq (1 - \delta\alpha_k)\|F_\mu(X, y, S)\| - \beta\mu \\ &\leq (1 - \delta\alpha_k)\beta\mu - \beta\mu \\ &= -\delta\alpha_k\beta\mu \\ &< 0. \end{aligned}$$

So when γ is sufficiently close to 0 but stays positive, we must have $d(\gamma) < 0$, i.e., $(X^+, y^+, S^+) \in \mathcal{N}_\beta((1-\gamma)\mu)$. Hence γ_k is well defined, and we have $0 < \gamma_k < 1$, $0 < \mu^+ = (1 - \gamma_k)\mu < \mu$.

4 Convergence

As mentioned in Section 1, we use (X_Q, y_q) to denote the projection of (Q, q) onto the optimal solution set of the SDP (1). Because it is a feasible solution pair, so $X_Q \succeq 0$, $\mathcal{A}^*(X_Q) = b$, and $S_q := C - \mathcal{A}^*(y_q) \succeq 0$. Since we also assume strong duality, then $\langle S_q, X_Q \rangle = 0$. Our goal in this section is to show that $(X^k, y^k, S^k) \longrightarrow (X_Q, y_q, S_q)$.

Let

$$\begin{aligned} U^k &= \frac{1}{\mu_k}\left((S^k)^{\frac{1}{2}} X^k (S^k)^{\frac{1}{2}} - \mu_k I\right), \\ V^k &= \frac{1}{\mu_k}\left(-\mu_k^p(X^k - Q) + \mathcal{A}^*(y^k) + S^k - C\right), \\ w^k &= \frac{1}{\mu_k}\left(\mathcal{A}(X^k) + \mu_k^p(y^k - q) - b\right). \end{aligned}$$

For simplicity, we will surpress the index k when there is no confusion, in other words, we let $(X, y, S, \mu, U, V, w) = (X^k, y^k, S^k, \mu_k, U^k, V^k, w^k)$.

Since $(X, y, S) \in N_\beta(\mu)$, then $\max\{\|U\|, \|V\|, \|w\|\} \leq \beta$ and $\langle S, X\rangle = tr(S^{\frac{1}{2}} X S^{\frac{1}{2}}) \leq 2n\mu$. From the definitions of V and w we also have $S = \mu^p(X - Q) - \mathcal{A}^*(y) + C + \mu V$, and $\mathcal{A}(X) = -\mu^p(y - q) + \mu w + b$.

First we show the boundedness of the iterates.

Lemma 4.1 (X, y, S) *is bounded.*

Proof. We have

$$
\begin{aligned}
& \langle S - S_q, X - X_Q \rangle \\
= \; & \langle \mu^p (X - Q) - \mathcal{A}^*(y) + C + \mu V - C + \mathcal{A}^*(y_q), X - X_Q \rangle \\
= \; & \langle -\mathcal{A}^*(y - y_q), X - X_Q \rangle + \mu^p \langle X - Q, X - X_Q \rangle + \mu \langle V, X - X_Q \rangle \\
= \; & \langle y - y_q, \mathcal{A}(X_Q) - \mathcal{A}(X) \rangle + \mu^p \langle X - Q, X - X_Q \rangle + \mu \langle V, X - X_Q \rangle \\
= \; & \langle y - y_q, b + \mu^p (y - q) - \mu w - b \rangle + \mu^p \langle X - Q, X - X_Q \rangle + \mu \langle V, X - X_Q \rangle \\
= \; & \mu^p \langle y - y_q, y - q \rangle - \mu \langle y - y_q, w \rangle + \mu^p \langle X - Q, X - X_Q \rangle + \mu \langle V, X - X_Q \rangle \\
= \; & \mu^p \langle (y - q) - (y_q - q), y - q \rangle - \mu \langle (y - q) - (y_q - q), w \rangle \\
& + \mu^p \langle X - Q, (X - Q) - (X_Q - Q) \rangle + \mu \langle V, (X - Q) - (X_Q - Q) \rangle \\
= \; & (\mu^p \|y - q\|^2 - \langle y - q, \mu^p (y_q - q) + \mu w \rangle + \mu \langle y_q - q, w \rangle \\
& + (\mu^p \|X - Q\|^2 - \langle X - Q, \mu^p (X_Q - Q) - \mu V \rangle - \mu \langle X_Q - Q, V \rangle \\
\geq \; & \mu^p \|y - q\|^2 - \mu^p \|y - q\| (\|y_q - q\| + \mu^{1-p} \|w\|) - \mu \|y_q - q\| \|w\| \\
& + \mu^p \|X - Q\|^2 - \mu^p \|X - Q\| (\|X_Q - Q\| + \mu^{1-p} \|V\|) - \mu \|X_Q - Q\| \|V\| \\
\geq \; & \mu^p \|y - q\|^2 - \mu^p \|y - q\| (\|y_q - q\| + \beta \mu^{1-p}) - \beta \mu \|y_q - q\| \\
& + \mu^p \|X - Q\|^2 - \mu^p \|X - Q\| (\|X_Q - Q\| + \beta \mu^{1-p}) - \beta \mu \|X_Q - Q\| \\
= \; & \mu^p \left(\|y - q\|^2 - \|y - q\| (\|y_q - q\| + \beta \mu^{1-p}) - \beta \mu^{1-p} \|y_q - q\| \right) \\
& + \mu^p \left(\|X - Q\|^2 - \|X - Q\| (\|X_Q - Q\| + \beta \mu^{1-p}) - \beta \mu^{1-p} \|X_Q - Q\| \right).
\end{aligned}
$$

On the other hand, we also have

$$
\begin{aligned}
\langle S - S_q, X - X_Q \rangle \; & = \; \langle S, X \rangle - \langle S, X_Q \rangle - \langle S_q, X \rangle + \langle S_q, X_Q \rangle \\
& \leq \; \langle S, X \rangle \leq 2n\mu.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& \mu^p \left(\|y-q\|^2 - \|y-q\|(\|y_q - q\| + \beta\mu^{1-p}) - \beta\mu^{1-p}\|y_q - q\| \right) \\
+ \ & \mu^p \left(\|X-Q\|^2 - \|X-Q\|(\|X_Q - Q\| + \beta\mu^{1-p}) - \beta\mu^{1-p}\|X_Q - Q\| \right) \\
\leq \ & 2n\mu,
\end{aligned}
$$

so

$$
\begin{aligned}
& \|y-q\|^2 - \|y-q\|(\|y_q - q\| + \beta\mu^{1-p}) - \beta\mu^{1-p}\|y_q - q\| \\
+ \ & \|X-Q\|^2 - \|X-Q\|(\|X_Q - Q\| + \beta\mu^{1-p}) - \beta\mu^{1-p}\|X_Q - Q\| \\
\leq \ & 2n\mu^{1-p}.
\end{aligned} \tag{8}
$$

Since $\mu_0 \geq \mu$, we have

$$
\begin{aligned}
& \|y-q\|^2 - \|y-q\|(\|y_q - q\| + \beta\mu_0^{1-p}) - \beta\mu_0^{1-p}\|y_q - q\| \\
+ \ & \|X-Q\|^2 - \|X-Q\|(\|X_Q - Q\| + \beta\mu_0^{1-p}) - \beta\mu_0^{1-p}\|X_Q - Q\| \\
\leq \ & 2n\mu_0^{1-p}.
\end{aligned}
$$

Notice that $\|y-q\|^2 - \|y-q\|(\|y_q - q\| + \beta\mu_0^{1-p}) - \beta\mu_0^{1-p}\|y_q - q\|$ and $\|X-Q\|^2 - \|X-Q\|(\|X_Q - Q\| + \beta\mu_0^{1-p}) - \beta\mu_0^{1-p}\|X_Q - Q\|$ are convex quadratic functions of $\|y-q\|$ and $\|X-Q\|$ respectively, they are both bounded from below, and will go to ∞ when $\|y-q\|$ and $\|X-Q\|$ go to ∞. Since $2n\mu_0^{1-p}$ is a constant, then (X, y) must be bounded, so is $S = \mu^p(X-Q) - \mathcal{A}^*(y) + C + \mu V$.

■

Now we show that μ_k decreases to 0.

Lemma 4.2

$$
\lim_{k \longrightarrow \infty} \mu_k = 0.
$$

Proof. Since $\mu_0 > 0$, $\mu_{k+1} = (1-\gamma_k)\mu_k$ and $\gamma_k \in (0,1)$, then μ_k is strictly decreasing, so $\mu_k \longrightarrow \hat{\mu} \geq 0$. Now we prove $\hat{\mu} = 0$ by contradiction.

Assume $\hat{\mu} > 0$, then we must have $\gamma_k \longrightarrow 0$.

From Lemma 4.1 we know that $\{(X^k, y^k, S^k)\}$ is bounded, so there exists a convergent subsequence $\left(X^{k_n}, y^{k_n}, S^{k_n}\right) \longrightarrow \left(\hat{X}, \hat{y}, \hat{S}\right)$, where $\hat{X} \succeq 0$, $\hat{y} \in R^m$, and $\hat{S} \succeq 0$. Moreover, since $\left(X^{k_n}, y^{k_n}, S^{k_n}\right) \in \mathcal{N}_\beta(\mu_{k_n})$, then $\left(\hat{X}, \hat{y}, \hat{S}\right) \in \mathcal{N}_\beta(\hat{\mu})$. So $\hat{S}^{\frac{1}{2}}\hat{X}\hat{S}^{\frac{1}{2}} \succ 0$ by Lemma 3.1. Hence $\hat{X} \succ 0$ and $\hat{S} \succ 0$.

Let $\hat{E} = 2\hat{S} \otimes \hat{S}$, $\hat{F} = I \otimes \left(\hat{S}\hat{X}\right) + \left(\hat{S}\hat{X}\right) \otimes I$, and $\hat{G} = \begin{bmatrix} \hat{E} & 0 & \hat{F} \\ -\hat{\mu}^p I & A^t & I \\ A & \hat{\mu}^p I & 0 \end{bmatrix}$. Then similar to the analysis of Lemma 3.2, we have $\hat{E} \succ 0$, $\hat{E}^{-1}\hat{F} \succ 0$, and $\hat{G}$ is nonsingular. Let $\hat{r}_c = 2\hat{\mu}\hat{S} - 2\hat{S}\hat{X}\hat{S}$, $\hat{r}_d = -\left(-\hat{\mu}^p\left(\hat{X} - Q\right) + \mathcal{A}^*(\hat{y}) + \hat{S} - C\right)$, and $\hat{r}_p = -\left(\mathcal{A}(\hat{X}) + \hat{\mu}^p(\hat{y} - q) - b\right)$. We have

$$\left(E^{k_n}, F^{k_n}, G^{k_n}, {G^{k_n}}^{-1}, r_c^{k_n}, r_d^{k_n}, r_p^{k_n}\right) \longrightarrow \left(\hat{E}, \hat{F}, \hat{G}, \hat{G}^{-1}, \hat{r_c}, \hat{r_d}, \hat{r_p}\right).$$

Let $\left(\Delta\hat{X}, \Delta\hat{y}, \Delta\hat{S}\right)$ be the (unique) solution to

$$\begin{cases} 2\hat{S}(\Delta\hat{X})\hat{S} + \hat{S}\hat{X}(\Delta\hat{S}) + (\Delta\hat{S})\hat{X}\hat{S} = \hat{r}_c, \\ -\hat{\mu}^p(\Delta\hat{X}) + \mathcal{A}^*(\Delta\hat{y}) + \Delta\hat{S} = \hat{r}_d, \\ \mathcal{A}(\Delta\hat{X}) + \hat{\mu}^p(\Delta\hat{y}) = \hat{r}_p. \end{cases}$$

We then have

$$\begin{pmatrix} vec(\Delta X^{k_n}) \\ \Delta y^{k_n} \\ vec(\Delta S^{k_n}) \end{pmatrix} = {G^{k_n}}^{-1} \begin{pmatrix} vec(r_c^{k_n}) \\ vec(r_d^{k_n}) \\ r_p^{k_n} \end{pmatrix} \longrightarrow \hat{G}^{-1} \begin{pmatrix} vec(\hat{r}_c) \\ vec(\hat{r}_d) \\ \hat{r}_p \end{pmatrix} = \begin{pmatrix} vec(\Delta\hat{X}) \\ \Delta\hat{y} \\ vec(\Delta\hat{S}) \end{pmatrix}.$$

Hence $\left(\Delta X^{k_n}, \Delta y^{k_n}, \Delta S^{k_n}\right) \longrightarrow \left(\Delta\hat{X}, \Delta\hat{y}, \Delta\hat{S}\right)$.

Set

$$\hat{\alpha} = \begin{cases} 0, & \text{if } \left\|F_\mu(\hat{X}, \hat{y}, \hat{S})\right\| = 0, \\ \min\left\{\frac{(1-\delta)\|F_{\hat{\mu}}(\hat{X}, \hat{y}, \hat{S})\|}{\|H_{\hat{S}^{\frac{1}{2}}}(\Delta\hat{X}\Delta\hat{S})\|}, 1\right\}, & \text{if } \left\|F_\mu(\hat{X}, \hat{y}, \hat{S})\right\| > 0, \end{cases}$$

and $(X^*, y^*, S^*) = (\hat{X}, \hat{y}, \hat{S}) + \hat{\alpha}(\Delta\hat{X}, \Delta\hat{y}, \Delta\hat{S})$.

Now we show that $\left(X^{k_n+1}, y^{k_n+1}, S^{k_n+1}\right) \longrightarrow (X^*, y^*, S^*)$ by considering two cases.

If $\left\|F_{\hat{\mu}}(\hat{X}, \hat{y}, \hat{S})\right\| = 0$, then $(\hat{r}_c, \hat{r}_d, \hat{r}_p) = 0$ and $\left(\Delta\hat{X}, \Delta\hat{y}, \Delta\hat{S}\right) = 0$. Thus $(\Delta X^{k_n}, \Delta y^{k_n}, \Delta S^{k_n}) \longrightarrow$

0. Using the fact that the α_k's are bounded, we have

$$\left(X^{k_n+1}, y^{k_n+1}, S^{k_n+1}\right) = \left(X^{k_n}, y^{k_n}, S^{k_n}\right) + \alpha_{k_n}\left(\Delta X^{k_n}, \Delta y^{k_n}, \Delta S^{k_n}\right) \longrightarrow \left(\hat{X}, \hat{y}, \hat{S}\right) = \left(X^*, y^*, S^*\right).$$

If $\left\|F_{\hat{\mu}}(\hat{X}, \hat{y}, \hat{S})\right\| > 0$, then $\left\|F_{\mu_{k_n}}(X^{k_n}, y^{k_n}, S^{k_n})\right\| > 0$ when k_n is sufficiently large, hence

$$\alpha_{k_n} = \min\left\{\frac{(1-\delta)\left\|F_{\mu_{k_n}}\left(X^{k_n}, y^{k_n}, S^{k_n}\right)\right\|}{\left\|H_{(S^{k_n})^{\frac{1}{2}}}(\Delta X^{k_n} \Delta S^{k_n})\right\|}, 1\right\} \longrightarrow \hat{\alpha},$$

by continuity, then

$$\begin{aligned}
\left(X^{k_n+1}, y^{k_n+1}, S^{k_n+1}\right) &= \left(X^{k_n}, y^{k_n}, S^{k_n}\right) + \alpha_{k_n}\left(\Delta X^{k_n}, \Delta y^{k_n}, \Delta S^{k_n}\right) \\
&\longrightarrow \left(\hat{X}, \hat{y}, \hat{S}\right) + \hat{\alpha}\left(\Delta\hat{X}, \Delta\hat{y}, \Delta\hat{S}\right) \\
&= (X^*, y^*, S^*).
\end{aligned}$$

From Section 3.3, we know that there exists a positive integer l such that

$$\left\|F_{(1-\theta^l)\hat{\mu}}(X^*, y^*, S^*)\right\| - \beta(1-\theta^l)\hat{\mu} < 0.$$

Since

$$\left\|F_{(1-\theta^l)\mu^{k_n}}(X^{k_n+1}, y^{k_n+1}, S^{k_n+1})\right\| - \beta(1-\theta^l)\mu^{k_n} \longrightarrow \left\|F_{(1-\theta^l)\hat{\mu}}(X^*, y^*, S^*)\right\| - \beta(1-\theta^l)\hat{\mu}$$

as k_n approaches ∞, then we must have

$$\left\|F_{(1-\theta^l)\mu^{k_n}}(X^{k_n+1}, y^{k_n+1}, S^{k_n+1})\right\| - \beta(1-\theta^l)\mu^{k_n} < 0.$$

when k_n is sufficiently large. Hence according to the definition of γ_k, we have $\gamma_{k_n} \geq \theta^l > 0$ which is a contradiction to $r_k \longrightarrow 0$.

So $\mu_k \longrightarrow 0$ as k approaches ∞.

■

Finally we can prove the main convergence theorem.

Theorem 4.3

$$\lim_{k\longrightarrow\infty}\left(X^k, y^k, S^k\right) = (X_Q, y_q, S_q).$$

Proof. Assume $\left(\hat{X}, \hat{y}, \hat{S}\right)$ is a limiting point (on subsequence $\{k_n\}$) of $\left(X^k, y^k, S^k\right)$.

Because $\mu_k \longrightarrow 0$ and $(X^k, y^k, S^k) \in \mathcal{N}_\beta(\mu_k)$, so $\left(\hat{X}, \hat{y}, \hat{S}\right) \in \mathcal{N}_\beta(0)$. Since $\hat{S}^{\frac{1}{2}}\hat{X}\hat{S}^{\frac{1}{2}} = 0 \Longleftrightarrow \hat{X}\hat{S} = 0$, then $\left(\hat{X}, \hat{y}, \hat{S}\right)$ satisfies the optimality condition (2). Hence $\left(\hat{X}, \hat{y}\right)$ is a primal-dual optimal solution pair. Since (X_Q, y_q) is the projection of (Q, q) onto the optimal solution set, we have $\|X_Q - Q\| \le \left\|\hat{X} - Q\right\|$ and $\|y_q - q\| \le \|\hat{y} - q\|$.

Using (8) we get

$$\begin{aligned} & \left\|y^{k_n} - q\right\|^2 - \left\|y^{k_n} - q\right\|\left(\|y_q - q\| + \beta\mu_{k_n}^{1-p}\right) - \beta\mu_{k_n}^{1-p}\|y_q - q\| \\ + & \left\|X^{k_n} - Q\right\|^2 - \left\|X^{k_n} - Q\right\|\left(\|X_Q - Q\| + \beta\mu_{k_n}^{1-p}\right) - \beta\mu_{k_n}^{1-p}\|X_Q - Q\| \\ \le & \; 2n\mu_{k_n}^{1-p}. \end{aligned}$$

Letting k_n go to ∞ in the previous inequality, we have

$$\begin{aligned} & \|\hat{y} - q\|^2 - \|\hat{y} - q\|\|y_q - q\| + \|\hat{X} - Q\|^2 - \|\hat{X} - Q\|\|X_Q - Q\| \\ = & \; \|\hat{y} - q\|(\|\hat{y} - q\| - \|y_q - q\|) + \|\hat{X} - Q\|(\|\hat{X} - Q\| - \|X_Q - Q\|) \\ \le & \; 0. \end{aligned}$$

Therefore $\|\hat{y} - q\| = \|y_q - q\|$ and $\|\hat{X} - Q\| = \|X_Q - Q\|$. So $(\hat{X}, \hat{y})$ is also the projection of (Q, q) onto $S_p \times S_d$. But such a pair is unique, then $(\hat{X}, \hat{y}, \hat{S}) = (X_Q, y_q, S_q)$.

Since $\left\{(X^k, y^k, S^k)\right\}$ is a bounded sequence with (X_Q, y_q, S_q) as the only limiting point, then we must have $(X^k, y^k, S^k) \longrightarrow (X_Q, y_q, S_q)$ as k approaches ∞.

■

References

[1] R. A. Horn and C. R. Johnson, (1994) *Topics in matrix analysis*, Cambridge University Press, Cambridge. Corrected reprint of the 1991 original.

[2] C. KANZOW, H. QI, AND L. QI, (2003) *On the minimum norm solution of linear programs*, J. Optim. Theory Appl., 116, no. 2, pp. 333-345.

[3] A. Lin, (2006) *On a special class of regularized central path for semidefinite programs*, manuscript submitted to Math. Prog. Ser. A.

[4] S. LUCIDI, (1987) *A new result in the theory and computation of the least-norm solution of a linear program*, J. Optim. Theory Appl., 55, pp. 103-117.

[5] S. LUCIDI, (1987) *A finite algorithm for the least two-norm solution of a linear program*, Optimization, 18, pp. 809-823.

[6] O. L. MANGASARIAN, (1983) *Least-norm linear programming solution as an unconstrained minimization problem*, J. Math. Anal. Appl., 92, pp. 240-251.

[7] O. L. MANGASARIAN, (2004) *A Newton method for linear programming*, J. Optim. Theory Appl., 121, no. 1, pp.1-18.

[8] R. D. C. MONTEIRO, (1997) *Primal-dual path-following algorithms for semidefinite programming*, SIAM J. Optim., 7 (3), pp. 663-678.

[9] J. F. STURM AND S. ZHANG, (1998) *On the Long Step Path-Following Method for Semidefinite Programming*, Operations Research Letters 22, pp. 145-150.

[10] H. WOLKOWICZ, R. SAIGAL, and L. VANDENBERGHE, (2000) editors, *Handbook of semidefinite programming*, volume 27 of *International Series in Operations Research & Management Science*. Kluwer Academic Publishers, Boston, MA.

[11] Y. ZHANG, (1998) *On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming*, SIAM J. Optim., 8(2), pp. 365-386.

[12] Y. B. ZHAO AND D. LI, (2002) *Locating the least 2-norm solution of linear programs via a path-following method*, SIAM J. Optim., 12, pp. 893-912.