A primal-dual interior point method for nonlinear semidefinite programming *

Hiroshi Yamashita[†], Hiroshi Yabe[‡] and Kouhei Harada[§]

September 3, 2006 (revised May 23, 2007)

Abstract

In this paper, we consider a primal-dual interior point method for solving nonlinear semidefinite programming problems. By combining the primal barrier penalty function and the primal-dual barrier function, a new primal-dual merit function is proposed within the framework of the line search strategy. We show the global convergence property of our method.

Key words. nonlinear semidefinite programming, primal-dual interior point method, barrier penalty function, primal-dual merit function, global convergence

1 Introduction

This paper is concerned with the following nonlinear semidefinite programming (SDP) problem:

$$\text{(1)} \qquad \begin{array}{lll} \text{minimize} & f(x), & x \in \mathbb{R}^n, \\ \text{subject to} & g(x) = 0, & X \equiv \mathcal{A}x - B \succeq 0 \end{array}$$

where we assume that the functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are sufficiently smooth. Here $\mathcal{A}$ is a linear operator defined by $\mathcal{A}x = \sum_{i=1}^{n} x_i A_i$ for $x \in \mathbb{R}^n$, and $A_i \in \mathbb{S}^p, i = 1, \ldots, n$, and $B \in \mathbb{S}^p$ are given matrices, where $\mathbb{S}^p$ denotes the set of pth order real symmetric matrices. By $X \succeq 0$ and $X \succ 0$, we mean that the matrix X is positive semidefinite and positive definite, respectively.

---

*The second author was supported in part by the Grant-in-Aid for Scientific Research (C) 16510123 of Japan Society for the Promotion of Science.

†Mathematical Systems Inc., 2-4-3, Shinjuku, Shinjuku-ku, Tokyo 160-0022, Japan. hy@msi.co.jp

‡Department of Mathematical Information Science, Faculty of Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan. yabe@rs.kagu.tus.ac.jp

§Mathematical Systems Inc., 2-4-3, Shinjuku, Shinjuku-ku, Tokyo 160-0022, Japan. harada@msi.co.jp

The problem (1) is an extension of the linear SDP problem. The linear SDP problems include linear programming problems, convex quadratic programming problems and second order cone programming problems, and they have many applications. Interior point methods for the linear SDP problems have been studied extensively by many researchers, see for example [1, 10, 11, 12, 13] and the references therein.

On the other hand, researches on numerical methods for nonlinear SDP are much more recent, and a few researchers have been studying these methods. For example, Kocvara and Stingl [9] developed a computer code PENNON for solving nonlinear SDP, in which the augmented Lagrangian function method was used. Correa and Ramirez [3] proposed an algorithm which used the sequentially linear SDP method. Related researches include Jarre [6], in which examples of nonlinear SDP problems were introduced, and Freund and Jarre [5]. Fares, Noll and Apkarian [4] applied the sequential linear SDP method to robust control problems. Recently Kanzow, Nagel, Kato and Fukushima [7] presented a successive linearization method with a trust region-type globalization strategy. However, no interior point type method for general nonlinear SDP problems has been proposed yet to our knowledge.

In this paper, we propose a globally convergent primal-dual interior point method for solving nonlinear SDP problems. The method is based on a line search algorithm in the primal-dual space. The present paper is organized as follows. In Section 2, the optimality conditions for problem (1) are described. In Sections 3 and 4, our primal-dual interior point method is proposed. Specifically, Section 3 presents the algorithm called SDPIP which constitutes the basic frame of primal-dual interior point methods. Section 4 gives the algorithm called SDPLS based on the line search strategy, which is an inner iteration of algorithm SDPIP given in Section 3. In Section 4.1, we describe the Newton method for solving nonlinear equations that are obtained by modifying the optimality conditions given in Section 2. In Section 4.2, we propose a new primal-dual merit function that consists of the primal barrier penalty function and the primal-dual barrier function. Then Section 4.3 presents algorithm SDPLS, and Section 5 shows its global convergence property. Furthermore, some numerical experiments are presented in Section 6. Finally, we give some concluding remarks in Section 7.

Throughout this paper, we define the inner product $\langle X, Z \rangle$ by $\langle X, Z \rangle = \mathrm{tr}(XZ)$ for any matrices X and Z in $\mathbb{S}^p$, where $\mathrm{tr}(M)$ denotes the trace of the matrix M. We also define an adjoint operator $\mathcal{A}^*$ of $\mathcal{A}$ such that $\mathcal{A}^* Z$ is an n dimensional vector whose ith element is $\mathrm{tr}\{A_i Z\}$. Then we have

$$\langle \mathcal{A}x, Z \rangle = x^T(\mathcal{A}^* Z) = (\mathcal{A}^* Z)^T x,$$

where the superscript T denotes the transpose of a vector or a matrix. In this paper, $(v)_i$ denotes the ith element of the vector v if necessary.

2 Optimality conditions

Let the Lagrangian function of problem (1) be defined by

$$L(w) = f(x) - y^T g(x) - \langle X, Z \rangle,$$

where $w = (x, y, Z)$, and $y \in \mathbb{R}^m$ and $Z \in \mathbb{S}^p$ are the Lagrange multiplier vector and matrix which correspond to the equality and positive semidefiniteness constraints, respectively. Then Karush-Kuhn-Tucker (KKT) conditions for optimality of problem (1) are given by the following (see [2]):

$$r_0(w) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ X \circ Z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{2}$$

and

$$X \succeq 0, \qquad Z \succeq 0. \tag{3}$$

Here $\nabla_x L(w)$ is given by

$$\begin{aligned} \nabla_x L(w) &= \nabla f(x) - A_0(x)^T y - \mathcal{A}^* Z, \\ A_0(x) &= \begin{pmatrix} \nabla g_1(x)^T \\ \vdots \\ \nabla g_m(x)^T \end{pmatrix} \in \mathbb{R}^{m \times n}, \end{aligned}$$

and the multiplication $X \circ Z$ is defined by

$$X \circ Z = \frac{XZ + ZX}{2}.$$

It is known that $X \circ Z = 0$ is equivalent to the relation $XZ = ZX = 0$.

We call $w = (x, y, Z)$ satisfying $X \succ 0$ and $Z \succ 0$ the interior point. The algorithm of this paper will generate such interior points. To construct an interior point algorithm, we introduce a positive parameter μ, and we replace the complementarity condition $X \circ Z = 0$ by $X \circ Z = \mu I$, where I denotes the identity matrix. Then we try to find a point that satisfies the barrier KKT (BKKT) conditions:

$$r(w, \mu) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ X \circ Z - \mu I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{4}$$

and

$$X \succ 0, \qquad Z \succ 0.$$

3 Algorithm for finding a KKT point

We first describe a procedure for finding a KKT point using the BKKT conditions. In this section, the subscript k denotes an iteration count of the outer iterations. We define the notation $\|r_0(w)\|_*$ and $\|r(w, \mu)\|_*$ by

$$\|r_0(w)\|_* = \left\| \begin{pmatrix} \nabla_x L(w) \\ g(x) \end{pmatrix} \right\| + \|X \circ Z\|_F$$

and

$$\|r(w, \mu)\|_* = \left\| \begin{pmatrix} \nabla_x L(w) \\ g(x) \end{pmatrix} \right\| + \|X \circ Z - \mu I\|_F,$$

respectively, where $\|\cdot\|$ denotes the l_2 norm for vectors and $\|\cdot\|_F$ denotes the Frobenius norm for matrices.

Now we present the algorithm called SDPIP which calculates a KKT point.

Algorithm SDPIP

Step 0. (Initialize) Set $\varepsilon > 0$, $M_c > 0$ and $k = 0$. Let a positive sequence $\{\mu_k\}, \mu_k \downarrow 0$ be given.

Step 1. (Approximate BKKT point) Find an interior point w_{k+1} that satisfies

$$\|r(w_{k+1}, \mu_k)\|_* \le M_c \mu_k. \tag{5}$$

Step 2. (Termination) If $\|r_0(w_{k+1})\|_* \le \varepsilon$, then stop.

Step 3. (Update) Set $k := k + 1$ and go to Step 1. □

We note that the barrier parameter sequence $\{\mu_k\}$ in Algorithm SDPIP needs not be determined beforehand. The value of each μ_k may be set adaptively as the iteration proceeds. We call condition (5) the approximate BKKT condition, and call a point that satisfies this condition the approximate BKKT point. If the matrix $A_0(x_*)$ is of full rank and there exists a nonzero vector $v \in \mathbb{R}^n$ such that

$$A_0(x_*)v = 0 \quad \text{and} \quad \sum_{i=1}^{n}((x_*)_i + v_i)A_i - B \succ 0,$$

then we say that the Mangasarian-Fromovitz constraint qualification (MFCQ) condition is satisfied at a point x_* (see [3] for example).

The following theorem shows the convergence property of Algorithm SDPIP.

Theorem 1 *Assume that the functions f and g are continuously differentiable. Let $\{w_k\}$ be an infinite sequence generated by Algorithm SDPIP. Suppose that the sequence $\{x_k\}$ is bounded and that the MFCQ condition is satisfied at any accumulation point of the sequence $\{x_k\}$. Then the sequences $\{y_k\}$ and $\{Z_k\}$ are bounded, and any accumulation point of $\{w_k\}$ satisfies KKT conditions* (2) *and* (3)*.*

Proof. To prove this theorem by contradiction, we suppose that either $\{y_k\}$ or $\{Z_k\}$ is not bounded, i.e.

$$\gamma_k \equiv \max\{|(y_k)_1|, \ldots, |(y_k)_m|, \lambda_{\max}(Z_k)\} \to \infty, \tag{6}$$

where $\lambda_{\max}(Z_k)$ denotes the largest eigenvalue of the matrix Z_k. It follows from (5) that the boundedness of $\{x_k\}$ implies

$$\limsup_{k\to\infty} \left\| A_0(x_k)^T y_k + \mathcal{A}^* Z_k \right\| < \infty.$$

Then we have $\|A_0(x_k)^T y_k/\gamma_k + \mathcal{A}^* Z_k/\gamma_k\| \to 0$. Letting an arbitrary accumulation point of $\{x_k, y_k/\gamma_k, Z_k/\gamma_k\}$ be (x_*, y_*, Z_*), we have

$$A_0(x_*)^T y_* + \mathcal{A}^* Z_* = 0 \quad \text{and} \quad X_* Z_* = Z_* X_* = 0. \tag{7}$$

We will prove that $Z_* = 0$. For this purpose, we assume that $\lambda_{\max}(Z_*) > 0$ holds. Since the matrices X_* and Z_* commute, they share the same eigensystem. Thus the matrices X_* and Z_* can be transformed to the diagonal matrices by using the same orthogonal matrix P as follows:

$$\overline{X}_* \equiv P X_* P^T = \operatorname{diag}(\lambda_1, ..., \lambda_p) \quad \text{and} \quad \overline{Z}_* \equiv P Z_* P^T = \operatorname{diag}(\tau_1, ..., \tau_p),$$

where $\lambda_1 \le \lambda_2 \le ... \le \lambda_p$ and $\tau_1 \le \tau_2 \le ... \le \tau_p$ are the nonnegative eigenvalues of X_* and Z_*, respectively. It follows from the assumption that there exists an integer p' such that $1 \le p' < p$, $\lambda_{p'} = 0$ and $\lambda_{p'+1} > 0$ hold. Furthermore, the MFCQ condition implies that there exists a nonzero vector $v \in \mathbb{R}^n$ which satisfies

$$A_0(x_*) v = 0 \quad \text{and} \quad \sum_{i=1}^n ((x_*)_i + v_i) A_i - B \succ 0.$$

Therefore, we get

$$\sum_{i=1}^n ((x_*)_i + v_i)(\overline{A}_i)_{jj} - \overline{B}_{jj} > 0 \tag{8}$$

for $j = 1, \dots, p$, where $\overline{A}_i = P A_i P^T$ and $\overline{B} = P B P^T$. Since the following holds

$$0 = \lambda_j = (\overline{X}_*)_{jj} = \sum_{i=1}^n (x_*)_i (\overline{A}_i)_{jj} - \overline{B}_{jj} \quad \text{for} \quad j = 1, \dots, p',$$

equation (8) yields

$$\sum_{i=1}^n v_i (\overline{A}_i)_{jj} > 0 \quad \text{for} \quad j = 1, ..., p'. \tag{9}$$

By premultiplying (7) by v^T, we have

$$\begin{aligned} 0 &= v^T A_0(x_*)^T y_* + v^T \mathcal{A}^* Z_* = v^T \mathcal{A}^* Z_* = \sum_{i=1}^n v_i \operatorname{tr}\{A_i Z_*\} \\ &= \sum_{i=1}^n v_i \operatorname{tr}\{\overline{A}_i \overline{Z}_*\} = \sum_{j=1}^p \sum_{i=1}^n v_i (\overline{A}_i)_{jj} \tau_j \\ &= \sum_{j=1}^{p'} \sum_{i=1}^n v_i (\overline{A}_i)_{jj} \tau_j + \sum_{j=p'+1}^{p} \sum_{i=1}^n v_i (\overline{A}_i)_{jj} \tau_j. \end{aligned}$$

Since the complementarity condition $\overline{X}_* \overline{Z}_* = 0$ implies $\tau_j = 0$ for $j = p' + 1, \dots, p$, the equation above yields

$$\sum_{j=1}^{p'} \sum_{i=1}^n v_i (\overline{A}_i)_{jj} \tau_j = 0.$$

By (9), we have $\tau_j = 0$ for $j = 1, \ldots, p'$, which contradicts the assumption $\lambda_{\max}(Z_*) > 0$. Therefore we obtain $Z_* = 0$, which yields $A_0(x_*)^T y_* = 0$ from (7). Since the matrix $A_0(x_*)$ is of full rank, we have $y_* = 0$. This contradicts the fact that some element of y_* or Z_* is not zero by (6). Therefore, the sequences $\{y_k\}$ and $\{Z_k\}$ are bounded.

Let $\hat{w}$ be any accumulation point of $\{w_k\}$. Since the sequences $\{w_k\}$ and $\{\mu_k\}$ satisfy (5) for each k and μ_k approaches zero, $r_0(\hat{w}) = 0$ follows from the definition of $r(w, \mu)$.

Therefore the proof is complete. $\square$

4 Algorithm for finding a barrier KKT point

As same as the case of linear SDP problems, we consider a scaling of the primal-dual pair (X, Z) in applying the Newton method to the system of equations (4). We define a transformation $T \in \mathbb{R}^{p \times p}$, and we scale X and Z by

$$\widetilde{X} = TXT^T \quad \text{and} \quad \widetilde{Z} = T^{-T} Z T^{-1}$$

respectively. Using the transformation T, we replace the equation $X \circ Z = \mu I$ by a form $\widetilde{X} \circ \widetilde{Z} = \mu I$, and deal with the modified BKKT conditions

$$\text{(10)} \qquad \tilde{r}(w, \mu) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ \widetilde{X} \circ \widetilde{Z} - \mu I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

instead of (4) to form Newton directions as described below.

4.1 Newton method

In this section, we consider a method for solving the BKKT conditions approximately for a given $\mu > 0$, which corresponds to the inner iterations of Step 1 of Algorithm SDPIP. Throughout this section, the index k denotes the inner iteration count for a given $\mu > 0$. We note again that $X_k \succ 0$ and $Z_k \succ 0$ are maintained for all k in the following.

We apply a Newton-like method to the system of equations (10). Let the Newton directions for the primal and dual variables by $\Delta x \in \mathbb{R}^n$ and $\Delta Z \in \mathbb{S}^p$, respectively. We define $\Delta X = \sum_{i=1}^n \Delta x_i A_i$ and we note that $\Delta X \in \mathbb{S}^p$. We also scale ΔX and ΔZ by

$$\Delta\widetilde{X} = T \Delta X T^T \quad \text{and} \quad \Delta\widetilde{Z} = T^{-T} \Delta Z T^{-1}.$$

Since $(\widetilde{X} + \Delta\widetilde{X}) \circ (\widetilde{Z} + \Delta\widetilde{Z}) = \mu I$ can be written as

$$(\widetilde{X} + \Delta\widetilde{X})(\widetilde{Z} + \Delta\widetilde{Z}) + (\widetilde{Z} + \Delta\widetilde{Z})(\widetilde{X} + \Delta\widetilde{X}) = 2\mu I,$$

neglecting the nonlinear parts $\Delta\widetilde{X}\Delta\widetilde{Z}$ and $\Delta\widetilde{Z}\Delta\widetilde{X}$ implies the Newton equation for (10)

$$\begin{aligned} &\text{(11)} & G\Delta x - A_0(x)^T \Delta y - \mathcal{A}^* \Delta Z &= -\nabla_x L(x, y, Z) \\ &\text{(12)} & A_0(x)\Delta x &= -g(x) \\ &\text{(13)} & \Delta\widetilde{X}\widetilde{Z} + \widetilde{Z}\Delta\widetilde{X} + \widetilde{X}\Delta\widetilde{Z} + \Delta\widetilde{Z}\widetilde{X} &= 2\mu I - \widetilde{X}\widetilde{Z} - \widetilde{Z}\widetilde{X}, \end{aligned}$$

where G denotes the Hessian matrix of the Lagrangian function $L(w)$ or its approximation (see Remark 2 in Section 4.3).

Similarly to usual primal-dual interior point methods for linear SDP problems, we derive an explicit form of the direction $\Delta Z \in \mathbb{S}^p$ from equation (13) and substitute it into equation (11) in order to obtain the Newton direction $\Delta w = (\Delta x, \Delta y, \Delta Z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$. For this purpose, we make use of relations described in [1] and Appendix of [11]. For $U \in \mathbb{S}^p$, nonsingular $P \in \mathbb{R}^{p\times p}$ and $Q \in \mathbb{R}^{p\times p}$, we define the operator

$$(P \odot Q)U = \frac{1}{2}(PUQ^T + QUP^T)$$

and the symmetrized Kronecker product

$$(P \otimes_S Q)\mathrm{svec}(U) = \mathrm{svec}((P \odot Q)U),$$

where the operator svec is defined by

$$\mathrm{svec}(U) = (U_{11}, \sqrt{2}U_{21}, \ldots, \sqrt{2}U_{p1}, U_{22}, \sqrt{2}U_{32}, \ldots, \sqrt{2}U_{p2}, U_{33}, \ldots, U_{pp})^T \in \mathbb{R}^{p(p+1)/2}.$$

We note that, for any $U, V \in \mathbb{S}^p$,

$$\langle U, V \rangle = \mathrm{tr}(UV) = \mathrm{svec}(U)^T \mathrm{svec}(V) \tag{14}$$

holds. By using the operator, the matrices $\widetilde{X}$, $\widetilde{Z}$, $\Delta\widetilde{X}$ and $\Delta\widetilde{Z}$ can be represented by

$$\widetilde{X} = (T \odot T)X, \qquad \widetilde{Z} = (T^{-T} \odot T^{-T})Z, \tag{15}$$

$$\Delta\widetilde{X} = (T \odot T)\Delta X \quad \text{and} \quad \Delta\widetilde{Z} = (T^{-T} \odot T^{-T})\Delta Z. \tag{16}$$

Let $P' \in \mathbb{R}^{p\times p}$ and $Q' \in \mathbb{R}^{p\times p}$ be nonsingular, and $V \in \mathbb{S}^p$. By denoting the inverse operator of svec by smat, we have

$$(P \odot Q)U = \mathrm{smat}\left((P \otimes_S Q)\mathrm{svec}(U)\right). \tag{17}$$

We also define

$$(P \odot Q)^{-1}U = \mathrm{smat}\left((P \otimes_S Q)^{-1}\mathrm{svec}(U)\right). \tag{18}$$

The expressions above give

$$\begin{aligned}(P \odot Q)(P' \odot Q')U &= \mathrm{smat}\left((P \otimes_S Q)\mathrm{svec}((P' \odot Q')U)\right)\\ &= \mathrm{smat}\left((P \otimes_S Q)(P' \otimes_S Q')\mathrm{svec}(U)\right)\end{aligned}$$

and

$$\{(P \odot Q)(P' \odot Q')\}^{-1}U = (P' \odot Q')^{-1}(P \odot Q)^{-1}U.$$

Furthermore, we get

$$\begin{aligned}\langle U, (P \odot Q)V \rangle &= \mathrm{tr}\left\{U(P \odot Q)V\right\}\\ &= \frac{1}{2}\mathrm{tr}\{U(PVQ^T + QVP^T)\}\\ &= \frac{1}{2}\mathrm{tr}\{Q^TUPV + P^TUQV\}\\ &= \mathrm{tr}\left\{((P^T \odot Q^T)U)V\right\}\\ &= \left\langle (P^T \odot Q^T)U, V \right\rangle \qquad (19)\end{aligned}$$

and

$$\begin{aligned}\left\langle U,(P\odot Q)^{-1}V\right\rangle &= \operatorname{tr}\left\{U(P\odot Q)^{-1}V\right\}\\ &= \operatorname{tr}\left\{((P^T\odot Q^T)(P^T\odot Q^T)^{-1}U)(P\odot Q)^{-1}V\right\}\\ &= \operatorname{tr}\left\{((P^T\odot Q^T)^{-1}U)(P\odot Q)(P\odot Q)^{-1}V\right\}\\ &= \operatorname{tr}\left\{((P^T\odot Q^T)^{-1}U)V\right\}\\ &= \left\langle (P^T\odot Q^T)^{-1}U,V\right\rangle.\end{aligned}$$

Now we have the following theorem that gives the Newton directions.

Theorem 2 *Suppose that the operator $\widetilde{X}\odot I$ is invertible. Then the direction $\Delta\widetilde{Z}\in\mathbb{S}^p$ is given by the form*

$$\Delta\widetilde{Z}=\mu\widetilde{X}^{-1}-\widetilde{Z}-(\widetilde{X}\odot I)^{-1}(\widetilde{Z}\odot I)\Delta\widetilde{X}, \tag{20}$$

or equivalently

$$\Delta Z=\mu X^{-1}-Z-(T^T\odot T^T)(\widetilde{X}\odot I)^{-1}(\widetilde{Z}\odot I)(T\odot T)\Delta X, \tag{21}$$

and the directions $(\Delta x,\Delta y)\in\mathbb{R}^n\times\mathbb{R}^m$ satisfy

$$\begin{pmatrix} G+H & -A_0(x)^T\\ -A_0(x) & 0\end{pmatrix}\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}=-\begin{pmatrix}\nabla f(x)-A_0(x)^Ty-\mu\mathcal{A}^*X^{-1}\\ -g(x)\end{pmatrix}, \tag{22}$$

where the elements of the matrix H are represented by the form

$$H_{ij}=\left\langle \widetilde{A}_i,(\widetilde{X}\odot I)^{-1}(\widetilde{Z}\odot I)\widetilde{A}_j\right\rangle \tag{23}$$

with $\widetilde{A}_i=TA_iT^T$.

Furthermore, if the matrix $G+H$ is positive definite and the matrix $A_0(x)$ is of full rank, then the Newton equations (11) – (13) *give a unique search direction $\Delta w=(\Delta x,\Delta y,\Delta Z)\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{S}^p$.*

Proof. By equation (13), we have

$$2(\widetilde{Z}\odot I)\Delta\widetilde{X}+2(\widetilde{X}\odot I)\Delta\widetilde{Z}=2\mu(\widetilde{X}\odot I)\widetilde{X}^{-1}-2(\widetilde{X}\odot I)\widetilde{Z},$$

which implies that

$$(\widetilde{X}\odot I)\left(\widetilde{Z}+\Delta\widetilde{Z}-\mu\widetilde{X}^{-1}\right)=-(\widetilde{Z}\odot I)\Delta\widetilde{X}.$$

Thus we obtain equation (20). Since $(T^{-T}\otimes_S T^{-T})^{-1}=(T^{-T})^{-1}\otimes_S(T^{-T})^{-1}=T^T\otimes_S T^T$ holds (see Appendix of [11]), it follows from (18) and (17) that for any $U\in\mathbb{S}^p$,

$$\begin{aligned}(T^{-T}\odot T^{-T})^{-1}U &= \operatorname{smat}\left((T^{-T}\otimes_S T^{-T})^{-1}\operatorname{svec}(U)\right)\\ &= \operatorname{smat}\left((T^T\otimes_S T^T)\operatorname{svec}(U)\right)\\ &= (T^T\odot T^T)U.\end{aligned}$$

By (15) and (16), equation (20) implies that

$$\Delta Z = \mu X^{-1} - Z - (T^T \odot T^T)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)\Delta X,$$

which means equation (21). Then we have

$$\begin{aligned}
\mathcal{A}^*\Delta Z &= \mu\mathcal{A}^* X^{-1} - \mathcal{A}^* Z - \mathcal{A}^*(T^T \odot T^T)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)\Delta X \\
&= \mu\mathcal{A}^* X^{-1} - \mathcal{A}^* Z - \sum_j \mathcal{A}^*(T^T \odot T^T)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)A_j\Delta x_j \\
&= \mu\mathcal{A}^* X^{-1} - \mathcal{A}^* Z - H\Delta x, \qquad (24)
\end{aligned}$$

where the elements of the matrix H are defined by the form

$$\begin{aligned}
H_{ij} &= \mathrm{tr}\left\{A_i(T^T \odot T^T)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)A_j\right\} \\
&= \mathrm{tr}\left\{((T \odot T)A_i)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)A_j\right\} \\
&= \mathrm{tr}\left\{\widetilde{A}_i(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)\widetilde{A}_j\right\} \\
&= \left\langle \widetilde{A}_i, (\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)\widetilde{A}_j \right\rangle
\end{aligned}$$

with $\widetilde{A}_i = TA_iT^T$. This implies (23). By substituting (24) into (11), the Newton equations reduce to equation (22).

Furthermore, it is well known that the coefficient matrix of the linear system of equations (22) becomes nonsingular if the matrix $G + H$ is positive definite and the matrix $A_0(x)$ is of full rank.

Therefore the proof is complete. □

We note that if the matrix G is updated by a positive definite quasi-Newton formula (see Remark 2 in Section 4.3) and the matrix H is chosen as a positive definite matrix, then Theorem 2 guarantees that the Newton direction is uniquely determined.

The following theorem shows the positive definiteness of the matrix H. In what follows, we assume that the matrices $A_1, \ldots, A_n$ are linearly independent, which means that $\sum_{i=1}^n v_iA_i = 0$ implies $v_i = 0, i = 1, \ldots, n$.

Theorem 3 *Assume that the operator $\widetilde{X} \odot I$ is invertible. Suppose that $\widetilde{X}$ and $\widetilde{Z}$ are symmetric positive definite, and that $\widetilde{X}\widetilde{Z} + \widetilde{Z}\widetilde{X}$ is symmetric positive semidefinite. Suppose that the matrices $A_i, i = 1, \ldots, n$ are linearly independent. Then the matrix H is positive definite.*

Furthermore, if $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$ holds, then H becomes a symmetric matrix.

Proof. Let $\widetilde{U} = \sum_{i=1}^n u_i\widetilde{A}_i$ for any $u(\neq 0) \in \mathbb{R}^n$. Since the linear independence of the matrices A_i for $i = 1, \ldots, n$ is equivalent to the linear independence of the matrices $\widetilde{A}_i$

for $i=1,\dots,n$, $u\neq 0$ guarantees that $\widetilde{U}\neq 0$. By defining $V=(\widetilde{X}\odot I)^{-1}\widetilde{U}\neq 0$, the quadratic form of H is written as

$$\begin{aligned} u^T H u &= \sum_{i=1}^{n}\sum_{j=1}^{n} u_i \mathrm{tr}\left\{\widetilde{A}_i(\widetilde{X}\odot I)^{-1}(\widetilde{Z}\odot I)\widetilde{A}_j\right\} u_j \\ &= \mathrm{tr}\left\{\widetilde{U}(\widetilde{X}\odot I)^{-1}(\widetilde{Z}\odot I)\widetilde{U}\right\} \\ &= \mathrm{tr}\left\{((\widetilde{X}\odot I)^{-1}\widetilde{U})(\widetilde{Z}\odot I)(\widetilde{X}\odot I)(\widetilde{X}\odot I)^{-1}\widetilde{U}\right\} \\ &= \mathrm{tr}\left\{V(\widetilde{Z}\odot I)(\widetilde{X}\odot I)V\right\} \\ &= \frac{1}{2}\left\{\mathrm{tr}\left\{V(\widetilde{Z}\odot I)(\widetilde{X}\odot I)V\right\}+\mathrm{tr}\left\{V(\widetilde{X}\odot I)(\widetilde{Z}\odot I)V\right\}\right\}. \end{aligned}$$

It follows from property 6 of Symmetrized Kronecker product in Appendix of [11] and relation (14) that

(25)
$$\begin{aligned} u^T H u &= \frac{1}{4}\left\{\mathrm{tr}\left\{V((\widetilde{Z}\widetilde{X}\odot I)+(\widetilde{Z}\odot\widetilde{X}))V\right\}+\mathrm{tr}\left\{V((\widetilde{X}\widetilde{Z}\odot I)+(\widetilde{X}\odot\widetilde{Z}))V\right\}\right\} \\ &= \frac{1}{4}\mathrm{svec}(V)^T\left(((\widetilde{X}\widetilde{Z}+\widetilde{Z}\widetilde{X})\otimes_S I)+(\widetilde{X}\otimes_S\widetilde{Z})+(\widetilde{Z}\otimes_S\widetilde{X})\right)\mathrm{svec}(V). \end{aligned}$$

It follows from Property 11 of Symmetrized Kronecker product in Appendix of [11] that if $\widetilde{X}$ and $\widetilde{Z}$ are symmetric positive definite, then $\widetilde{X}\otimes_S\widetilde{Z}$ and $\widetilde{Z}\otimes_S\widetilde{X}$ are symmetric positive definite. It also follows from Property 9 that if $\widetilde{X}\widetilde{Z}+\widetilde{Z}\widetilde{X}$ is symmetric positive semidefinite, then $(\widetilde{X}\widetilde{Z}+\widetilde{Z}\widetilde{X})\otimes_S I$ is symmetric positive semidefinite. Thus the matrix H is positive definite.

Next, we assume that $\widetilde{X}\widetilde{Z}=\widetilde{Z}\widetilde{X}$ holds. Since the relation $(\widetilde{X}\odot I)(\widetilde{Z}\odot I)=(\widetilde{Z}\odot I)(\widetilde{X}\odot I)$ holds, we have

$$(\widetilde{X}\odot I)^{-1}(\widetilde{Z}\odot I)=(\widetilde{Z}\odot I)(\widetilde{X}\odot I)^{-1}. \tag{26}$$

For any vectors $u,\ v\in\mathbb{R}^n$, we define

$$\widetilde{U}\equiv\sum_{i=1}^{n}u_i\widetilde{A}_i,\quad \widetilde{V}\equiv\sum_{i=1}^{n}v_i\widetilde{A}_i,\quad \widetilde{U}'=(\widetilde{X}\odot I)^{-1}\widetilde{U}\quad\text{and}\quad \widetilde{V}'=(\widetilde{X}\odot I)^{-1}\widetilde{V}.$$

Then in a similar way to the above, we obtain

$$\begin{aligned} u^T H v &= \mathrm{tr}\left\{\widetilde{U}(\widetilde{X}\odot I)^{-1}(\widetilde{Z}\odot I)\widetilde{V}\right\} \\ &= \mathrm{tr}\left\{\widetilde{U}(\widetilde{Z}\odot I)(\widetilde{X}\odot I)^{-1}\widetilde{V}\right\} \qquad \text{(from (26))} \\ &= \mathrm{tr}\left\{(\widetilde{Z}\odot I)(\widetilde{X}\odot I)^{-1}\widetilde{V}\widetilde{U}\right\} \\ &= \mathrm{tr}\left\{\widetilde{V}(\widetilde{X}\odot I)^{-1}(\widetilde{Z}\odot I)\widetilde{U}\right\} \qquad \text{(from (19))} \\ &= v^T H u. \end{aligned}$$

Letting $u = e_i$ and $v = e_j$ yields $H_{ij} = H_{ji}$, which implies that the matrix H is symmetric.

Therefore the theorem is proved. □

We note that Theorems 2 and 3 correspond to Theorem 3.1 in [11].

The following theorem claims that a BKKT point is obtained if the Newton direction satisfies $\Delta x = 0$.

Theorem 4 *Assume that* Δw *solves* (11) - (13). *If* $\Delta x = 0$, *then* $(x, y + \Delta y, Z + \Delta Z)$ *is a BKKT point.*

Proof. It follows from the Newton equations that

$$\begin{aligned} \nabla f(x) - A_0(x)^T(y + \Delta y) - \mathcal{A}^*(Z + \Delta Z) &= 0, \\ g(x) &= 0. \end{aligned}$$

Since equation (21) implies

$$Z + \Delta Z = \mu X^{-1},$$

we have

$$X \circ (Z + \Delta Z) = \mu I \qquad \text{and} \qquad Z + \Delta Z \succ 0.$$

Therefore the point $(x, y + \Delta y, Z + \Delta Z)$ satisfies the BKKT conditions. □

In the subsequent discussions, we assume that the nonsingular matrix T is chosen so that $\widetilde{X}$ and $\widetilde{Z}$ commute, i.e., $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$. In this case, the matrices $\widetilde{X}$ and $\widetilde{Z}$ share the same eigensystem. To end this section, we give the two concrete choices of the transformation T that satisfy such a condition (see [11]).

(i) If we set $T = X^{-1/2}$, then we have $\widetilde{X} = I$ and $\widetilde{Z} = X^{1/2}ZX^{1/2}$. In this case, the matrices H and ΔZ can be represented by the form:

$$\begin{aligned} H_{ij} &= \operatorname{tr}\left(A_i X^{-1} A_j Z\right), \\ \Delta Z &= \mu X^{-1} - Z - \frac{1}{2}(X^{-1}\Delta X Z + Z \Delta X X^{-1}). \end{aligned}$$

(ii) If we set $T = W^{-1/2}$ with $W = X^{1/2}(X^{1/2}ZX^{1/2})^{-1/2}X^{1/2}$, then we have $\widetilde{X} = W^{-1/2}XW^{-1/2} = W^{1/2}ZW^{1/2} = \widetilde{Z}$. Note that this choice is proposed by Nesterov and Todd. In this case, the matrices H and ΔZ can be represented by the form:

$$\begin{aligned} H_{ij} &= \operatorname{tr}\left\{A_i W^{-1} A_j W^{-1}\right\}, \\ \Delta Z &= \mu X^{-1} - Z - W^{-1}\Delta X W^{-1}. \end{aligned}$$

4.2 Primal-dual merit function

In what follows, we assume that the matrix T is so chosen that $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$ is satisfied. To force the global convergence of the algorithm described in Section 4, we use a merit function in the primal-dual space. For this purpose, we propose the following merit function:

$$F(x, Z) = F_{BP}(x) + \nu F_{PD}(x, Z), \tag{27}$$

where $F_{BP}(x)$ and $F_{PD}(x, Z)$ are the primal barrier penalty function and the primal-dual barrier function, respectively, and they are given by

$$\begin{aligned} F_{BP}(x) &= f(x) - \mu \log(\det X) + \rho \|g(x)\|_1, & (28)\\ F_{PD}(x, Z) &= \langle X, Z \rangle - \mu \log(\det X \det Z), & (29) \end{aligned}$$

where ν and ρ are positive parameters. It follows from the fact $\widetilde{X}\widetilde{Z} = TXZT^{-1}$ that $\left\langle \widetilde{X}, \widetilde{Z} \right\rangle = \langle X, Z \rangle$ and $F_{PD}(\widetilde{x}, \widetilde{Z}) = F_{PD}(x, Z)$ hold.

The following lemma gives a lower bound on the value of the primal-dual barrier function (29) and the asymptotic behavior of the function.

Lemma 1 *The primal-dual barrier function satisfies*

$$F_{PD}(x, Z) \geq p\mu(1 - \log \mu) \tag{30}$$

for any $X \succ 0$ and $Z \succ 0$. The equality holds in (30) *if and only if $X \circ Z = \mu I$ is satisfied. Furthermore, the following hold*

$$\lim_{\langle X,Z \rangle \downarrow 0} F_{PD}(x, Z) = \infty \qquad \text{and} \qquad \lim_{\langle X,Z \rangle \uparrow \infty} F_{PD}(x, Z) = \infty. \tag{31}$$

Proof. Let λ_i and τ_i for $i = 1, \ldots, p$ denote the eigenvalues of the matrices $\widetilde{X}$ and $\widetilde{Z}$, respectively. We note that the matrices $\widetilde{X}$ and $\widetilde{Z}$ share the same eigensystem. Then the matrix $\widetilde{X}\widetilde{Z}$ has eigenvalues $\lambda_i\tau_i,\ i = 1, \ldots, p$, and we have

$$\begin{aligned} F_{PD}(x, Z) &= \left\langle \widetilde{X}, \widetilde{Z} \right\rangle - \mu \log(\det\widetilde{X}\det\widetilde{Z}) \\ &= \sum_{i=1}^{p} \lambda_i\tau_i - \mu \log\left(\prod_{i=1}^{p} \lambda_i\tau_i\right) \\ &= \sum_{i=1}^{p} \left(\lambda_i\tau_i - \mu \log \lambda_i\tau_i\right). \end{aligned} \tag{32}$$

It is easily shown that the function $\phi(\xi) = \xi - \mu \log \xi$ $(\xi > 0)$ is convex and achieves a minimum value at $\xi = \mu$. Thus we obtain

$$\begin{aligned} F_{PD}(x, Z) &\geq \sum_{i=1}^{p} (\mu - \mu \log \mu) \\ &= p\,(\mu - \mu \log \mu). \end{aligned} \tag{33}$$

It is clear that the equality holds in inequality (33) if and only if $\lambda_i\tau_i = \mu,\ i = 1, \ldots, p$ are satisfied. Since $\widetilde{X}$ and $\widetilde{Z}$ commute, they can be represented by the forms $\widetilde{X} = PD_XP^T$ and $\widetilde{Z} = PD_ZP^T$ for an orthogonal matrix P, where D_X and D_Z are diagonal matrices whose diagonal elements are λ_i and τ_i, $i = 1, \ldots, p$, respectively. Thus, by noting the relations $\widetilde{X} \circ \widetilde{Z} = \widetilde{X}\widetilde{Z} = PD_X D_Z P^T$, we can show that $\widetilde{X} \circ \widetilde{Z} = \mu I$ is equivalent to the equations $\lambda_i \tau_i = \mu$, $i = 1, \ldots, p$. Furthermore, $\widetilde{X} \circ \widetilde{Z} = \mu I$ is equivalent to $X \circ Z = \mu I$. Therefore, the first part of this lemma is proved.

It follows from the algebraic and geometric mean $\frac{1}{p}\sum_{i=1}^{p} \lambda_i \tau_i \geq \left(\prod_{i=1}^{p} \lambda_i \tau_i\right)^{1/p}$ that

$$\begin{aligned} -\log\left(\prod_{i=1}^{p} \lambda_i \tau_i\right) &\geq -p\log\left(\sum_{i=1}^{p} \lambda_i \tau_i\right) + p\log p \\ &= -p\log\langle X, Z\rangle + p\log p. \end{aligned}$$

We use the inequality above and equation (32) to obtain

$$F_{PD}(x, Z) \geq \langle X, Z\rangle - \mu p \log\langle X, Z\rangle + \mu p \log p.$$

Therefore, the expressions (31) hold. This completes the proof. □

Now we introduce the first order approximation F_l of the merit function by

$$F_l(x, Z; \Delta x, \Delta Z) = F(x, Z) + \Delta F_l(x, Z; \Delta x, \Delta Z),$$

which is used in the line search procedure. Here $\Delta F_l(x, Z; \Delta x, \Delta Z)$ corresponds to the directional derivative and it is defined by the form

$$\Delta F_l(x, Z; \Delta x, \Delta Z) = \Delta F_{BPl}(x; \Delta x) + \nu \Delta F_{PDl}(x, Z; \Delta x, \Delta Z),$$

where

$$\begin{aligned} \Delta F_{BPl}(x; \Delta x) &= \nabla f(x)^T \Delta x - \mu \mathrm{tr}(X^{-1}\Delta X) \\ &\quad + \rho\left(\|g(x) + A_0(x)\Delta x\|_1 - \|g(x)\|_1\right), \\ \Delta F_{PDl}(x, Z; \Delta x, \Delta Z) &= \mathrm{tr}(\Delta X Z + X\Delta Z - \mu X^{-1}\Delta X - \mu Z^{-1}\Delta Z). \end{aligned} \tag{34}$$

We show that the search direction is a descent direction for both the primal barrier penalty function (28) and the primal-dual barrier function (29). We first give an estimate of $\Delta F_{BPl}(x; \Delta x)$ for the primal barrier penalty function.

Lemma 2 *Assume that Δw solves* (11) – (13)*. Then the following holds*

$$\Delta F_{BPl}(x; \Delta x) \leq -\Delta x^T (G + H)\Delta x - (\rho - \|y + \Delta y\|_\infty)\|g(x)\|_1.$$

Proof. It is clear from (12) and (34) that

$$\Delta F_{BPl}(x; \Delta x) = \nabla f(x)^T \Delta x - \mu \mathrm{tr}(X^{-1}\Delta X) - \rho\|g(x)\|_1. \tag{35}$$

It follows from (11) that

$$\nabla f(x)^T \Delta x = -\Delta x^T G \Delta x + \Delta x^T A_0(x)^T (y + \Delta y) + \Delta x^T \mathcal{A}^*(Z + \Delta Z).$$

Since $\mathcal{A}^*(Z + \Delta Z) = \mu \mathcal{A}^* X^{-1} - H\Delta x$ holds by (24), the preceding expression implies that

$$\nabla f(x)^T \Delta x = -\Delta x^T \left(G + H\right) \Delta x - g(x)^T (y + \Delta y) + \mu \Delta x^T \mathcal{A}^* X^{-1}.$$

By using the relations

$$\Delta x^T \mathcal{A}^* X^{-1} = \sum_{i=1}^{n} \Delta x_i \mathrm{tr}(A_i X^{-1}) = \mathrm{tr}((\sum_{i=1}^{n} \Delta x_i A_i) X^{-1}) = \mathrm{tr}(X^{-1} \Delta X),$$

equation (35) yields

$$\begin{aligned} \Delta F_{BPl}(x; \Delta x) &= -\Delta x^T \left(G + H\right) \Delta x - g(x)^T (y + \Delta y) \\ &\qquad + \mu \mathrm{tr}(X^{-1} \Delta X) - \mu \mathrm{tr}(X^{-1} \Delta X) - \rho \|g(x)\|_1 \\ &\leq -\Delta x^T \left(G + H\right) \Delta x - (\rho - \|y + \Delta y\|_\infty) \|g(x)\|_1. \end{aligned}$$

The proof is complete. □

Next we estimate the difference $\Delta F_{PDl}(x, Z; \Delta x, \Delta Z)$ for the primal-dual barrier function (29).

Lemma 3 *Assume that* Δw *solves* (11) – (13)*. Then the following holds*

$$\Delta F_{PDl}(x, Z; \Delta x, \Delta Z) \leq 0. \tag{36}$$

The equality holds in (36) *if and only if the matrices* X *and* Z *satisfy the relation* $X \circ Z = \mu I$*.*

Proof. It follows from the Newton equation (13) that

$$\begin{aligned} \Delta F_{PDl}(x, Z; \Delta x, \Delta Z) &= \mathrm{tr}\left\{(I - \mu \widetilde{X}^{-1} \widetilde{Z}^{-1})(\widetilde{Z} \Delta \widetilde{X} + \widetilde{X} \Delta \widetilde{Z})\right\} \\ &= \frac{1}{2} \mathrm{tr}\left\{(I - \mu \widetilde{X}^{-1} \widetilde{Z}^{-1})(\widetilde{Z} \Delta \widetilde{X} + \widetilde{X} \Delta \widetilde{Z} + \Delta \widetilde{X} \widetilde{Z} + \Delta \widetilde{Z} \widetilde{X})\right\} \\ &= \mathrm{tr}\left\{(I - \mu \widetilde{X}^{-1} \widetilde{Z}^{-1})(\mu I - \widetilde{X} \widetilde{Z})\right\} \\ &= -\mathrm{tr}\left\{\widetilde{X}^{-1} \widetilde{Z}^{-1} (\mu I - \widetilde{X} \widetilde{Z})^2\right\} \\ &= -\mathrm{tr}\left\{(\widetilde{X} \widetilde{Z})^{-1/2} (\mu I - \widetilde{X} \widetilde{Z})^2 (\widetilde{X} \widetilde{Z})^{-1/2}\right\}. \end{aligned}$$

Since the matrix $(\widetilde{X} \widetilde{Z})^{-1/2} (\mu I - \widetilde{X} \widetilde{Z})^2 (\widetilde{X} \widetilde{Z})^{-1/2}$ is symmetric positive semidefinite, we have

$$\Delta F_{PDl}(x, Z; \Delta x, \Delta Z) \leq 0.$$

It is clear that the equality holds in the above if and only if the matrix $\mu I - \widetilde{X}\widetilde{Z}$ becomes the zero matrix. Therefore the proof is complete. □

Now we obtain the following theorem by using the two lemmas given above. This theorem shows that the Newton direction Δw becomes a descent search direction for the proposed primal-dual merit function (27).

Theorem 5 *Assume that Δw solves* (11) – (13) *and that the matrix $G + H$ is positive definite. Suppose that the penalty parameter ρ satisfies $\rho > \|y+\Delta y\|_\infty$. Then the following hold:*

(i) *The direction Δw becomes a descent search direction for the primal-dual merit function $F(x, Z)$, i.e. $\Delta F_l(x, Z; \Delta x, \Delta Z) \leq 0$.*

(ii) *If $\Delta x \neq 0$, then $\Delta F_l(x, Z; \Delta x, \Delta Z) < 0$.*

(iii) *$\Delta F_l(x, Z; \Delta x, \Delta Z) = 0$ holds if and only if $(x, y + \Delta y, Z)$ is a BKKT point.*

Proof. (i) and (ii) : It follows directly from Lemmas 2 and 3 that

$$\begin{aligned} \Delta F_l(x, Z; \Delta x, \Delta Z) &\leq -\Delta x^T(G + H)\Delta x \\ &\qquad -(\rho - \|y + \Delta y\|_\infty)\|g(x)\|_1 \\ &\leq 0. \end{aligned} \tag{37}$$

The last inequality becomes a strict inequality if $\Delta x \neq 0$. Therefore the results hold.
(iii) If $\Delta F_l(x, Z; \Delta x, \Delta Z) = 0$ holds, then $\Delta F_{BPl}(x; \Delta x) = 0$ and $\Delta F_{PDl}(x, Z; \Delta x, \Delta Z) = 0$ are satisfied, and equation (37) yields

$$\Delta x = 0 \quad \text{and} \quad g(x) = 0.$$

It follows from Lemma 3 that $\Delta F_{PDl}(x, Z; \Delta x, \Delta Z) = 0$ implies $X \circ Z = \mu I$, i.e. $XZ = \mu I$. Thus equation (21) yields $\Delta Z = 0$. Then equation (11) implies that $\nabla f(x) - A_0(x)^T(y + \Delta y) - \mathcal{A}^* Z = 0$. Hence $(x, y + \Delta y, Z)$ is a BKKT point.

Conversely, suppose that $(x, y + \Delta y, Z)$ is a BKKT point. Equations (11) and (24) imply that

$$G\Delta x - \mathcal{A}^*\Delta Z = 0 \quad \text{and} \quad \mathcal{A}^*\Delta Z = -H\Delta x.$$

It follows that $(G + H)\Delta x = 0$ holds, which yields $\Delta x = 0$. Using equation (35) and Lemma 3, we have

$$\Delta F_{BPl}(x; \Delta x) = 0 \quad \text{and} \quad \Delta F_{PDl}(x, Z; \Delta x, \Delta Z) = 0,$$

which implies $\Delta F_l(x, Z; \Delta x, \Delta Z) = 0$. Therefore, the theorem is proved. □

4.3 Algorithm SDPLS that uses the line search procedure

To obtain a globally convergent algorithm to a BKKT point for a fixed $\mu > 0$, we modify the basic Newton iteration. Our iterations take the form

$$x_{k+1} = x_k + \alpha_k \Delta x_k, \quad Z_{k+1} = Z_k + \alpha_k \Delta Z_k \quad \text{and} \quad y_{k+1} = y_k + \Delta y_k$$

where α_k is a step size determined by the line search procedure described below.

The main iteration is to decrease the value of the merit function (27) for fixed μ. Thus the step size is determined by the sufficient decrease rule of the merit function. Specifically, we adopt Armijo's rule. At the current point (x_k, Z_k), we calculate the maximum allowed step to the boundary of the feasible region by

$$\alpha_{xk\max} = -\frac{1}{\lambda_{\min}(X^{-1}\Delta X)} \tag{38}$$

and

$$\alpha_{zk\max} = -\frac{1}{\lambda_{\min}(Z^{-1}\Delta Z)}, \tag{39}$$

where $\lambda_{\min}(M)$ denotes the minimum eigenvalue of the matrix M. If the minimum eigenvalue in either expression is positive, we ignore the corresponding term.

A step to the next iterate is given by

$$\alpha_k = \bar{\alpha}_k \beta^{l_k}, \qquad \bar{\alpha}_k = \min\left\{\gamma\alpha_{xk\max}, \gamma\alpha_{zk\max},\ 1\right\},$$

where $\gamma \in (0,1)$ and $\beta \in (0,1)$ are fixed constants and l_k is the smallest nonnegative integer such that

$$F(x_k + \bar{\alpha}_k\beta^{l_k}\Delta x_k, Z_k + \bar{\alpha}_k\beta^{l_k}\Delta Z_k) \le F(x_k, Z_k) + \varepsilon_0\bar{\alpha}_k\beta^{l_k}\Delta F_l(x_k, Z_k; \Delta x_k, \Delta Z_k), \tag{40}$$

where $\varepsilon_0 \in (0,1)$. Lemma 4 (ii) given below guarantees that an integer l_k exists.

Now we give a line search algorithm called Algorithm SDPLS. This algorithm should be regarded as the inner iteration of Algorithm SDPIP (see Step 1 of Algorithm SDPIP). We also note that ε' given below corresponds to $M_c\mu$ in Algorithm SDPIP.

Algorithm SDPLS

Step 0. (Initialize) Let $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$ $(X_0 \succ 0, Z_0 \succ 0)$, and $\mu > 0$, $\rho > 0$, $\nu > 0$. Set $\varepsilon' > 0$, $\gamma \in (0,1)$, $\beta \in (0,1)$ and $\varepsilon_0 \in (0,1)$. Let $k = 0$.

Step 1. (Termination) If $\|r(w_k, \mu)\|_* \le \varepsilon'$, then stop.

Step 2. (Compute direction) Calculate the matrix G_k and the transformation T_k. Determine the direction Δw_k by solving (11) – (13).

Step 3. (Step size) Find the smallest nonnegative integer l_k that satisfies the criterion (40), and calculate

$$\alpha_k = \bar{\alpha}_k \beta^{l_k}.$$

Step 4. (Update variables) Set

$$x_{k+1} = x_k + \alpha_k \Delta x_k, \quad Z_{k+1} = Z_k + \alpha_k \Delta Z_k \quad \text{and} \quad y_{k+1} = y_k + \Delta y_k.$$

Step 5. Set $k := k + 1$ and go to Step 1. □

Remark 1. Theorem 2 can be used to calculate the direction Δw_k in Step 2. Specifically, we compute the directions $(\Delta x_k, \Delta y_k)$ by solving linear system of equations (22), and we obtain ΔZ_k from equation (21). It follows from Theorem 4 that if $\Delta x_k = 0$ is obtained, then we can get the BKKT point $(x_k, y_k + \Delta y_k, Z_k + \Delta Z_k)$ and stop the procedure of the algorithm.

Remark 2. When the matrix G_k approximates the Hessian matrix $\nabla_x^2 L(w_k)$ of the Lagrangian function by using the quasi-Newton updating formula in Step 2, we have the following secant condition

$$G_{k+1} s_k = q_k,$$

where $s_k = x_{k+1} - x_k$ and

$$\begin{aligned} q_k &= \nabla_x L(x_{k+1}, y_{k+1}, Z_{k+1}) - \nabla_x L(x_k, y_{k+1}, Z_{k+1}) \\ &= (\nabla f(x_{k+1}) - A_0(x_{k+1})^T y_{k+1} - \mathcal{A}^* Z_{k+1}) - (\nabla f(x_k) - A_0(x_k)^T y_{k+1} - \mathcal{A}^* Z_{k+1}) \\ &= \nabla f(x_{k+1}) - \nabla f(x_k) - (A_0(x_{k+1}) - A_0(x_k))^T y_{k+1}. \end{aligned}$$

We note that it is easy to calculate the vector q_k. In order to preserve the positive definiteness of the matrix G_k, we can use the modified BFGS update proposed by Powell, which is given by the form

$$G_{k+1} = G_k - \frac{G_k s_k s_k^T G_k}{s_k^T G_k s_k} + \frac{\hat{q}_k \hat{q}_k^T}{s_k^T \hat{q}_k},$$

where

$$\begin{aligned} \hat{q}_k &= \psi_k q_k + (1 - \psi_k) G_k s_k, \\ \psi_k &= \begin{cases} 1 \quad \text{if} \quad s_k^T q_k \geq 0.2 s_k^T G_k s_k \\ \dfrac{0.8 s_k^T G_k s_k}{s_k^T (G_k s_k - q_k)} \quad \text{otherwise.} \end{cases} \end{aligned}$$

Remark 3. From the viewpoint of the numerical accuracy and computational cost, we calculate the minimum eigenvalue of the symmetric matrix $L^{-1} \Delta X L^{-T}$ in computing the minimum eigenvalue in (38) based on the fact that the spectrum of the nonsymmetric matrix $X^{-1} \Delta X$ is same as that of the symmetric matrix $L^{-1} \Delta X L^{-T}$, where $X = LL^T$ is the Cholesky factorization of X. We can also calculate the minimum eigenvalue of the matrix $Z^{-1} \Delta Z$ in equation (39) in the same way.

5 Global convergence to a barrier KKT point

In this section, we prove global convergence of Algorithm SDPLS. For this purpose, we make the following assumptions.

Assumptions

(A1) The functions f and $g_i, i = 1, ..., m$, are twice continuously differentiable.

(A2) The sequence $\{x_k\}$ generated by Algorithm SDPLS remains in a compact set Ω of $\mathbb{R}^n$.

(A3) The matrix $A_0(x_k)$ is of full rank for all k on Ω. The matrices $A_1, \ldots, A_n$ are linearly independent.

(A4) The matrix G_k is uniformly bounded and positive semidefinite.

(A5) The transformation T_k is chosen such that $\widetilde{X}_k$ and $\widetilde{Z}_k$ commute, and both of the sequences $\{T_k\}$ and $\{T_k^{-1}\}$ are bounded.

(A6) The penalty parameter ρ is sufficiently large so that $\rho > \|y_k + \Delta y_k\|_\infty$ holds for all k.

□

Assumption (A2) assures the existence of an accumulation point of the generated sequence $\{x_k\}$. The boundedness of the generated sequence $\{x_k\}$ is derived if there exist upper and lower bounds on the variable x, which is a reasonable assumption in practice. We should note that if a quasi-Newton approximation is used for computing the matrix G_k, then we only need the continuity of the first order derivatives of functions in assumption (A1).

In order to show the global convergence property, we first present the following lemma that gives a base for Armijo's line search rule. The merit function is differentiable except for the part $\|g(x)\|_1$, so we can prove this lemma in the same way as Lemmas 2 and 3 in [14].

Lemma 4 *Let $d_x \in \mathbb{R}^n$ and $D_z \in \mathbb{R}^{p\times p}$ be given. Define $F'(x, Z; d_x, D_z)$ by*

$$F'(x, Z; d_x, D_z) = \lim_{t\downarrow 0} \frac{F(x + td_x, Z + tD_z) - F(x, Z)}{t}.$$

Then the following hold:
(i) *There exists a $\theta \in (0, 1)$ such that*

$$F(x + d_x, Z + D_z) \le F(x, Z) + F'(x + \theta d_x, Z + \theta D_z; d_x, D_z),$$

whenever $X + \mathcal{A}d_x \succ 0$ and $Z + D_z \succ 0$.
(ii) *Let $\varepsilon_0 \in (0, 1)$ be given. If $\Delta F_l(x, Z; d_x, D_z) < 0$, then*

$$F(x + \alpha d_x, Z + \alpha D_z) - F(x, Z) \le \varepsilon_0 \alpha \Delta F_l(x, Z; d_x, D_z),$$

for sufficiently small $\alpha > 0$. □

The following lemma shows the boundedness of the sequence $\{w_k\}$ and the uniformly positive definiteness of the matrix H_k.

Lemma 5 *Suppose that assumptions* (A1), (A2) *and* (A6) *are satisfied. Let the sequence* $\{w_k\}$ *be generated by Algorithm SDPLS. Then the following hold.*
(i) $\liminf_{k\to\infty} \det(X_k) > 0$ *and* $\liminf_{k\to\infty} \det(Z_k) > 0$.
(ii) *The sequence* $\{w_k\}$ *is bounded.*
In addition, if assumptions (A3), (A4) *and* (A5) *are satisfied, the following hold.*
(iii) *There exists a positive constant* M *such that*

$$\frac{1}{M}\|v\|^2 \le v^T(G_k + H_k)v \le M\|v\|^2 \quad \text{for any } v \in \mathbb{R}^n$$

for all $k \ge 0$.
(iv) *The sequence* $\{\Delta w_k\}$ *is bounded.*

Proof. (i) Since the sequence $\{F_{PD}(x_k, Z_k)\}$ is bounded below from Lemma 1, the sequence $\{F_{BP}(x_k)\}$ is bounded above, because the function value of $F(x_k, Z_k)$ decreases monotonically. Therefore it follows from the log barrier term in $F_{BP}(x)$ that $\det X_k$ is bounded away from zero, and we have $\liminf_{k\to\infty} \det X_k > 0$. This implies that $\liminf_{k\to\infty} \det Z_k > 0$ also holds, because $\{F_{PD}(x_k, Z_k)\}$ is bounded above and below and $\langle X_k, Z_k\rangle \ge 0$ is satisfied.

(ii) The boundedness of the sequences $\{Z_k\}$ and $\{y_k\}$ follows from assumptions (A2), (A6) and the monotone decreasing of $F(x_k, Z_k)$. Therefore the sequence $\{w_k\}$ is bounded.

(iii) It follows from Appendix 9 of [11] that the operator $\widetilde{X} \odot I$ is invertible. For the vector V defined in the proof of Theorem 3, svec(V) can be represented by the form

$$\begin{aligned} \mathrm{svec}(V) &= \mathrm{svec}\left(\mathrm{smat}((\widetilde{X} \otimes_S I)^{-1}\widetilde{U})\right) \\ &= (\widetilde{X} \otimes_S I)^{-1} \sum_{i=1}^{n} u_i \mathrm{svec}(\widetilde{A}_i), \end{aligned}$$

where $\widetilde{U} \equiv \sum_{i=1}^{n} u_i \widetilde{A}_i \ne 0$. Letting

$$\widetilde{A} = \left(\mathrm{svec}(\widetilde{A}_1), \ldots, \mathrm{svec}(\widetilde{A}_n)\right) \in \mathbb{R}^{p(p+1)/2 \times n}$$

and

$$u = (u_1, \ldots, u_n)^T,$$

we have

$$\mathrm{svec}(V) = (\widetilde{X} \otimes_S I)^{-1}\widetilde{A}u.$$

Therefore it follows from (25) that

$$u^T H_k u = u^T \widetilde{A}^T((\widetilde{X}_k \otimes_S I)^{-1})^T \hat{H}_k (\widetilde{X}_k \otimes_S I)^{-1}\widetilde{A}u,$$

where
$$\hat{H}_k = ((\widetilde{X}_k\widetilde{Z}_k + \widetilde{Z}_k\widetilde{X}_k) \otimes_S I) + (\widetilde{X}_k \otimes_S \widetilde{Z}_k) + (\widetilde{Z}_k \otimes_S \widetilde{X}_k).$$
The boundedness of the sequence $\{w_k\}$ guarantees the uniformly positive definiteness and boundedness of the matrix $((\widetilde{X}_k \otimes_S I)^{-1})^T \hat{H}_k(\widetilde{X}_k \otimes_S I)^{-1}$. Since the linear independence of the matrices A_i for $i = 1, \dots, n$ is equivalent to the linear independence of the vectors $\mathrm{svec}(\widetilde{A}_i)$ for $i = 1, \dots, n$, the matrix $\widetilde{A}$ is of column full rank. This implies that there exist positive constants λ_{min} and λ_{max}, which are independent of k, such that
$$\lambda_{min}\|u\|^2 \le u^T H_k u \le \lambda_{max}\|u\|^2$$
holds. Thus by assumption (A4), we obtain the result.

(iv) Since, by results (ii) and (iii) shown above, the sequence $\{w_k\}$ is bounded and $\{G_k + H_k\}$ is uniformly bounded and positive definite, Theorem 2 guarantees the desired result. □

By Theorem 4, $\Delta x_k = 0$ guarantees that $(x_k, y_k + \Delta y_k, Z_k + \Delta Z_k)$ is a BKKT point. Thus in what follows, we assume that $\Delta x_k \ne 0$ for any $k \ge 0$. The following theorem gives the global convergence of an infinite sequence generated by Algorithm SDPLS.

Theorem 6 *Suppose that assumptions* (A1) – (A6) *hold. Let an infinite sequence* $\{w_k\}$ *be generated by Algorithm SDPLS. Then there exists at least one accumulation point of* $\{w_k\}$*, and any accumulation point of the sequence* $\{w_k\}$ *is an BKKT point.*

Proof. In the proof, we define the following notations
$$u_k = \begin{pmatrix} x_k \\ Z_k \end{pmatrix} \quad \text{and} \quad \Delta u_k = \begin{pmatrix} \Delta x_k \\ \Delta Z_k \end{pmatrix}$$
for simplicity. By Lemma 5 (ii), the sequence $\{w_k\}$ has at least one accumulation point. The boundedness of the sequence $\{w_k\}$ implies that all eigenvalues of X_k and Z_k are bounded above. It follows from Lemma 5 (i) that each smallest eigenvalue of X_k and Z_k is bounded away from zero. Furthermore, by Lemma 5 (iv), $\|\Delta w_k\|_*$ is uniformly bounded above. Hence, we have $\liminf_{k\to\infty} \bar{\alpha}_k > 0$.

It follows from Lemma 5 (iii) that there exists a positive constant M such that
$$\frac{1}{M}\|v\|^2 \le v^T(G_k + H_k)v \le M\|v\|^2$$
for any $v \in \mathbb{R}^n$ and all $k \ge 0$. Thus by (37), we have
$$\Delta F_l(u_k; \Delta u_k) \le -\frac{\|\Delta x_k\|^2}{M} < 0,$$
and inequality (40) yields
$$\begin{aligned} F(u_{k+1}) - F(u_k) &\le \varepsilon_0\bar{\alpha}_k\beta^{l_k}\Delta F_l(u_k; \Delta u_k) \\ &\le -\varepsilon_0\bar{\alpha}_k\beta^{l_k}\frac{\|\Delta x_k\|^2}{M} \\ &< 0. \end{aligned} \tag{41}$$

Because the sequence $\{F(u_k)\}$ is monotonically decreasing and bounded below, the left-hand side of (41) converges to 0, which implies that

$$\lim_{k\to\infty} \beta^{l_k} \Delta F_l(u_k; \Delta u_k) = 0.$$

If there exists a finite number N such that $l_k < N$ for all k, then we have $\lim_{k\to\infty} \Delta F_l(u_k; \Delta u_k) = 0$. Now we suppose that there exists a subsequence $K \subset \{0, 1, \cdots\}$ such that $l_k \to \infty, k \in K$. Then we can assume $l_k > 0$ for sufficiently large $k \in K$ without loss of generality, which means that the point $u_k + \alpha_k \Delta u_k/\beta$ does not satisfy condition (40). Thus, we get

$$F(u_k + \alpha_k \Delta u_k/\beta) - F(u_k) > \varepsilon_0 \alpha_k \Delta F_l(u_k; \Delta u_k)/\beta. \tag{42}$$

By Lemma 4, there exists a $\theta_k \in (0, 1)$ such that for $k \in K$,

$$\begin{aligned} F(u_k + \alpha_k \Delta u_k/\beta) - F(u_k) &\leq \alpha_k F'(u_k + \theta_k \alpha_k \Delta u_k/\beta; \Delta u_k)/\beta \\ &\leq \alpha_k \Delta F_l(u_k + \theta_k \alpha_k \Delta u_k/\beta; \Delta u_k)/\beta. \end{aligned} \tag{43}$$

Then, from (42) and (43), we see that

$$\varepsilon_0 \Delta F_l(u_k; \Delta u_k) < \Delta F_l(u_k + \theta_k \alpha_k \Delta u_k/\beta; \Delta u_k).$$

This inequality yields

$$\Delta F_l(u_k + \theta_k \alpha_k \Delta u_k/\beta; \Delta u_k) - \Delta F_l(u_k; \Delta u_k) > (\varepsilon_0 - 1) \Delta F_l(u_k; \Delta u_k) > 0. \tag{44}$$

Thus by the fact $l_k \to \infty$, $k \in K$, we have $\alpha_k \to 0$ and thus $\|\theta_k \alpha_k \Delta u_k/\beta\|_* \to 0, k \in K$, because $\|\Delta u_k\|_*$ is uniformly bounded. Here $\|\Delta u_k\|_*$ is defined by $\|\Delta u_k\|_* = \|\Delta x_k\| + \|\Delta Z_k\|_F$. This implies that the left-hand side of (44) and therefore $\Delta F_l(u_k; \Delta u_k)$ converges to zero when $k \to \infty, k \in K$.

By the discussions above, we have proved that

$$\lim_{k\to\infty} \Delta F_l(u_k; \Delta u_k) = 0. \tag{45}$$

Since equation (45) implies that

$$\Delta F_{BPl}(x_k; \Delta x_k) \to 0 \quad \text{and} \quad \Delta F_{PDl}(x_k, z_k; \Delta x_k, \Delta z_k) \to 0,$$

it follows from equations (37), (12) and Lemma 3 that

$$\Delta x_k \to 0, \quad g(x_k) \to 0, \quad X_k \circ Z_k \to \mu I \quad (\tilde{X}_k \circ \tilde{Z}_k \to \mu I).$$

Therefore, equation (21) yields

$$\Delta Z_k \to 0.$$

By equation (11), we have

$$\nabla_x L(x_k, y_k + \Delta y_k, Z_k) \to 0,$$

which implies that

$$r(x_k, y_k + \Delta y_k, Z_k, \mu) \to 0.$$

Since $x_{k+1} = x_k + \alpha_k \Delta x_k, Z_{k+1} = Z_k + \alpha_k \Delta Z_k, \Delta x_k \to 0, \Delta Z_k \to 0$ and $y_{k+1} = y_k + \Delta y_k$, the result follows. Therefore, the theorem is proved. □

The preceding theorem guarantees that any accumulation point of the sequence $\{(x_k, y_k, Z_k)\}$ satisfies the BKKT conditions. If we adopt a common step size α_k as $w_{k+1} = w_k + \alpha_k \Delta w_k$ in Step 4 of Algorithm SDPLS, where α_k is determined in Step 3, then the result of the theorem is replaced by the statement that any accumulation point of the sequence $\{(x_k, y_k + \Delta y_k, Z_k)\}$ satisfies the BKKT conditions.

6 Numerical Experiments

The algorithm of this paper is implemented and brief numerical experiments are done in order to verify the theoretical results of the proposed algorithm. The program is written in C++, and is run on 3.2GHz Pentium IV PC with LINUX OS.

In the following experiments, initial values of various quantities are set as follows: $\mu_0 = 1.0, X_0 = I, Z_0 = I$. The barrier parameter is updated by the rule $\mu_{k+1} = \mu_k/10.0$ after an approximate barrier KKT point is obtained in Step 1 of Algorithm SDPIP where we set $M_c = 0.1$, and the transformation is set to be $T = X^{-1/2}$ at each iteration of Algorithm SDPLS.

The first problem is Rosen-Suzuki problem combined with the positive semidefinite constraint, which is defined as follows:

$$\begin{array}{ll} \text{minimize} & {x_1}^2 + {x_2}^2 + 2{x_3}^2 + {x_4}^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4, \\ \text{subejct to} & 8 - {x_1}^2 - {x_2}^2 - {x_3}^2 - {x_4}^2 - x_1 + x_2 - x_3 + x_4 \geq 0, \\ & 10 - {x_1}^2 - 2{x_2}^2 - {x_3}^2 - 2{x_4}^2 + x_1 + x_4 \geq 0, \\ & 5 - 2{x_1}^2 - {x_2}^2 - {x_3}^2 - 2x_1 + x_2 + x_4 \geq 0, \\ & \begin{pmatrix} x_2 + x_3 & 0 & 0 & 0 \\ 0 & 2x_4 & x_1 & 0 \\ 0 & x_1 & 2x_4 & 0 \\ 0 & 0 & 0 & x_2 + x_3 \end{pmatrix} \succeq 0. \end{array}$$

The proposed algorithm solved this problem with 13 total inner iteration counts and ended with the final KKT residual norm $\|r_0(w)\|_* = 5.4 \times 10^{-7}$.

The second problem is Hock and Shittkowski No. 71 combined with the positive semidefinite constraint, which is defined by

$$\begin{array}{ll} \text{minimize} & x_1 x_4 (x_1 + x_2 + x_3) + x_3, \\ \text{subject to} & x_1 x_2 x_3 x_4 - 25 \geq 0, \\ & {x_1}^2 + {x_2}^2 + {x_3}^2 + {x_4}^2 - 40 \geq 0, \\ & 1 \leq x_i \leq 5, \; i = 1, \cdots, 4, \\ & \begin{pmatrix} x_1 & x_2 & 0 & 0 \\ x_2 & x_4 & x_2 + x_3 & 0 \\ 0 & x_2 + x_3 & x_4 & x_3 \\ 0 & 0 & x_3 & x_1 \end{pmatrix} \succeq 0. \end{array}$$

Our method solved this problem with 26 total inner iteration counts and ended with the final KKT residual norm $\|r_0(w)\|_* = 5.2 \times 10^{-7}$.

The third problem is a real financial one and taken from [8]. The model is to discriminate failure and non-failure companies by a Logit model using a positive semidefinite quadratic discriminant function. The problem for learning is defined by

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{M}(y_i z(x_i) - \log(1+e^{z(x_i)})),\, a \in \mathbb{R},\, b \in \mathbb{R}^q,\, Q \in \mathbb{S}^q, \\ \text{subject to} & Q \succeq 0, \end{array}$$

where $z(x) = a + b^T x + \frac{1}{2} x^T Q x$, and $x_i = (x_1, ..., x_q)_i$ gives financial data of each company $i = 1, ..., M$. The value of y_i gives failure or non-failure as follows:

$$\begin{array}{rcl} y_i & = & 0 \Leftrightarrow (x_1, ..., x_q)_i \in M_0(\mathrm{non-failure}), \\ y_i & = & 1 \Leftrightarrow (x_1, ..., x_q)_i \in M_1(\mathrm{failure}). \end{array}$$

In [8], Konno et.al. proposed a method that used a cutting plane approximation of positive semidefinite condition and solved resulting linearly constrained problems using an interior point NLP algorithm in NUOPT. In the following tables, we list two examples. Because this problem is convex, it is possible to use exact Hessian for the matrix G. The following tables show the results with both BFGS update (bfgs) and exact Hessian (hesse). In each table, the algorithms used, the final objective function value, the minimum eigenvalue of the obtained matrix Q, the total inner iteration counts and the run time (sec) are given. The learning experiments were done by Japan Credit Rating Agency, Ltd. with their own financial data including the data provided by Tokyo Shoko Research, Ltd. These tables show that our methods solve the problems efficiently and that our method (hesse) performs better than our method (bfgs).

Example 1: number of variables = 28, $q = 6$, $M = 6084$, $M_0 = 6053$				
algorithm	final objective	final $\lambda_{\min}(Q)$	iteration	time (sec)
cutting plane	-153.0808	-9.59e-05	–	7.77
ours (bfgs)	-153.0828	1.76e-09	117	1.65
ours (hesse)	-153.0828	1.77e-09	27	0.80

Example 2: number of variables = 45, $q = 8$, $M = 6084$, $M_0 = 6053$				
algorithm	final objective	final $\lambda_{\min}(Q)$	iteration	time (sec)
cutting plane	-143.7445	-9.17e-05	–	30.3
ours (bfgs)	-143.7468	3.88e-09	233	4.2
ours (hesse)	-143.7468	4.01e-09	30	1.5

The following tables show the required iteration counts for each value of μ. It is clear that majority of iterations are required at the first few values of μ.

Example 1			Example 2		
μ	bfgs	hesse	μ	bfgs	hesse
1.0e0	75	17	1.0e0	150	19
1.0e-1	25	2	1.0e-1	35	3
1.0e-2	14	2	1.0e-2	23	2
1.0e-3	4	2	1.0e-3	9	1
1.0e-4	3	1	1.0e-4	11	2
1.0e-5	3	2	1.0e-5	3	2
1.0e-6	1	1	1.0e-6	2	1
1.0e-7	1	1	1.0e-7	1	1

From the above brief experiments, we think the proposed method works as described in this paper, and hope the method is similarly efficient as existing primal-dual interior point methods for ordinary nonlinear programming [14].

7 Concluding Remarks

In this paper, we have proposed a primal-dual interior point method for solving nonlinear semidefinite programming problems. Within the line search strategy, we have proposed the primal-dual merit function that consists of the primal barrier penalty function and the primal-dual barrier function, and we have proved the global convergence property of our method. Brief numerical experiments show the practical efficiency of our method.

Analysis of the rate of convergence and more extensive numerical experiments of our method are under further research. In addition, we plan to construct a method within the framework of the trust region globalization strategy.

Acknowledgement. The authors would like to thank Japan Credit Rating Agency, Ltd. for providing us the results of their numerical experiments.

References

[1] F. Alizadeh, J. A. Haeberly and M. L. Overton, Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results, *SIAM Journal on Optimization*, 8 (1998), pp.746-768.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[3] R. Correa and C. H. Ramirez, A global algorithm for nonlinear semidefinite programming, *SIAM Journal on Optimization*, 15 (2004), pp.303-318.

[4] B. Fares, D. Noll and P. Apkarian, Robust control via sequential semidefinite programming, *SIAM Journal on Control and Optimization*, 40 (2002), pp.1791-1820.

[5] R. W. Freund and F. Jarre, *A sensitivity analysis and a convergence result for a sequential semidefinite programming method*, Numerical Analysis Manuscript No. 03-4-08, Bell Laboratories, April 2003.

[6] F. Jarre, *A QQP-minimization method for semidefinite and smooth nonconvex programs*, Technical report, Institute für Angewandte Mathematik und Statistik, Universität Würzburg, August 1998.

[7] C. Kanzow, C. Nagel, H. Kato and M. Fukushima, Successive linearization methods for nonlinear semidefinite programs, *Computational Optimization and Applications*, 31 (2005), pp.251-273.

[8] H. Konno, N. Kawadai and D. Wu, Estimation of failure probability using semidefinite Logit model, *Computational Management Science*, 1 (2004), pp.59-73.

[9] M. Kocvara and M. Stingl, PENNON : A code for convex nonlinear and semidefinite programming, *Optimization Methods and Software*, 18 (2003), pp.317-333.

[10] M. J. Todd, Semidefinite optimization, *Acta Numerica*, 10 (2001), pp.515-560.

[11] M. J. Todd, K. C. Toh and R. H. Tütüncü, On the Nesterov-Todd direction in semidefinite programming, *SIAM Journal on Optimization*, 8 (1998), pp.769-796.

[12] L. Vandenberghe and S. Boyd, Semidefinite programming, *SIAM Review*, 38 (1996), pp.49-95.

[13] H. Wolkowitz, R. Saigal and L. Vandenberghe, eds., *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*, Kluwer International Series in Operations Research and Management Science, Kluwer, Boston, 2000.

[14] H. Yamashita, A globally convergent primal-dual interior point method for constrained optimization, *Optimization Methods and Software*, 10 (1998), pp.443-469.