Math. Program., Ser. A (2012) 135:89–121
DOI 10.1007/s10107-011-0449-z

FULL LENGTH PAPER

A primal–dual interior point method for nonlinear semidefinite programming

Hiroshi Yamashita · Hiroshi Yabe · Kouhei Harada

Received: 13 July 2007 / Accepted: 18 February 2011 / Published online: 29 March 2011

Abstract This paper is concerned with a primal–dual interior point method for solving nonlinear semidefinite programming problems. The method consists of the outer iteration (SDPIP) that finds a KKT point and the inner iteration (SDPLS) that calculates an approximate barrier KKT point. Algorithm SDPLS uses a commutative class of Newton-like directions for the generation of line search directions. By combining the primal barrier penalty function and the primal–dual barrier function, a new primal–dual merit function is proposed. We prove the global convergence property of our method. Finally some numerical experiments are given.

Keywords Nonlinear semidefinite programming · Primal–dual interior point method · Barrier penalty function · Primal–dual merit function · Global convergence

Mathematics Subject Classification (2000) 90C22 · 90C26 · 90C51

1 Introduction

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

H. Yamashita · K. Harada
Mathematical Systems Inc., 2-4-3, Shinjuku, Shinjuku-ku, Tokyo 160-0022, Japan
e-mail: hy@msi.co.jp

K. Harada
e-mail: harada@msi.co.jp

H. Yabe (✉)
Department of Mathematical Information Science, Faculty of Science,
Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan
e-mail: yabe@rs.kagu.tus.ac.jp

$$\begin{aligned} &\text{minimize} \;\; f(x), \qquad x \in \mathbf{R}^n, \\ &\text{subject to} \; g(x) = 0, \;\; X(x) \succeq 0 \end{aligned} \tag{1}$$

where the functions $f : \mathbf{R}^n \to \mathbf{R}$, $g : \mathbf{R}^n \to \mathbf{R}^m$ and $X : \mathbf{R}^n \to \mathbf{S}^p$ are sufficiently smooth, and $\mathbf{S}^p$ denotes the set of p-th order real symmetric matrices. By $X(x) \succeq 0$ and $X(x) \succ 0$, we mean that the matrix $X(x)$ is positive semidefinite and positive definite, respectively.

The problem (1) is an extension of the linear SDP problem. For the case of the linear SDP problems, all the functions f and g are linear and the matrix $X(x)$ is defined by

$$X(x) = \sum_{i=1}^{n} x_i A_i - B$$

with given matrices $A_i \in \mathbf{S}^p, i = 1, \ldots, n$, and $B \in \mathbf{S}^p$. 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 [20,23,25] and the references therein.

On the other hand, researches on theoretical properties and numerical methods for nonlinear SDP are much more recent. Nonlinear SDP problems have been attracting a great deal of research attention, because such problems arise from several application fields, which include control theory, eigenvalue problems, finance and so forth (see [5–7,12,18,24] for example). For this reason, it is desired to develop numerical methods for solving nonlinear SDP problems. A few researchers have been studying these methods. For example, Kocvara and Stingl [13] developed a computer code PENNON for solving nonlinear SDP, in which the augmented Lagrangian function method was used (see also Stingle [19]). Fares, Apkarian and Noll [5] applied an augmented Lagrangian method to a class of LMI-constraint problem. Correa and Ramirez [4] proposed an algorithm which used the sequential linear SDP method. Fares, Noll and Apkarian [6] applied the sequential linear SDP method to robust control problems. Freund, Jarre and Vogelbusch [7] also studied a sequential SDP method. Kanzow, Nagel, Kato and Fukushima [10] presented a successive linearization method with a trust region-type globalization strategy. These methods generalize the SLP and SQP methods for nonlinear programming to solve nonlinear SDP problems. Furthermore, Tseng [22] briefly stated an application of a primal interior point method for nonlinear programming to nonlinear SDP problems. Jarre [9] applied a primal predictor corrector interior method to nonconvex SDP problems. However, no primal–dual interior point type method for general nonlinear SDP problems has been proposed yet to our knowledge. We note that a preliminary technical report of our algorithm appears in [27].

In this paper, we propose a globally convergent primal–dual interior point method for solving problem (1). This method consists of the outer iteration (SDPIP) that finds a KKT point and the inner iteration (SDPLS) that calculates an approximate barrier KKT point. The present paper is organized as follows. In Sect. 2, the optimality conditions for problem (1) are described. In Sects. 3 and 4, our primal–dual interior point method is proposed. Specifically, Sect. 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 Sect. 3. In Sect. 4.1, we describe the Newton method for solving nonlinear equations that are obtained by modifying the optimality conditions given in Sect. 2. The method uses a commutative class of Newton-like directions. In Sect. 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 Sect. 4.3 presents algorithm SDPLS, and Sect. 5 shows its global convergence property within the framework of the line search strategy. Furthermore, some numerical experiments are presented in Sect. 6. Finally, we give some concluding remarks in Sect. 7.

2 Optimality conditions

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 $\mathbf{S}^p$, where $\mathrm{tr}(M)$ denotes the trace of the matrix M. The superscript T denotes the transpose of a vector or a matrix, and $(v)_i$ denotes the i-th element of the vector v if necessary.

This section introduces optimality conditions for problem (1) and a modification of KKT conditions that will be used in an interior point method proposed in the following sections. We first define the Lagrangian function of problem (1) by

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

where $w = (x, y, Z)$, and $y \in \mathbf{R}^m$ and $Z \in \mathbf{S}^p$ are the Lagrange multiplier vector and matrix which correspond to the equality and positive semidefiniteness constraints, respectively. We also define matrices

$$A_i(x) = \frac{\partial X}{\partial x_i}$$

for $i = 1, \ldots, n$. Then the Karush-Kuhn-Tucker (KKT) conditions for optimality of problem (1) are given by the following (see [3]):

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

and

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

Here $\nabla_x L(w)$ is a gradient vector of the Lagrangian function and is given by

$$\nabla_x L(w) = \nabla f(x) - A_0(x)^T y - \mathcal{A}^*(x)Z,$$

where $A_0(x)$ is a Jacobian matrix of $g(x)$:

$$A_0(x) = \begin{pmatrix} \nabla g_1(x)^T \\ \vdots \\ \nabla g_m(x)^T \end{pmatrix} \in \mathbf{R}^{m \times n},$$

and $\mathcal{A}^*(x)$ is an operator such that for Z,

$$\mathcal{A}^*(x)Z = \begin{pmatrix} \langle A_1(x), Z \rangle \\ \vdots \\ \langle A_n(x), Z \rangle \end{pmatrix}.$$

In the following, we will occasionally deal with the multiplication $X(x) \circ Z$ which is defined by

$$X(x) \circ Z = \frac{X(x)Z + ZX(x)}{2}$$

instead of $X(x)Z$. It is known that $X(x) \circ Z = 0$ is equivalent to the relation $X(x)Z = ZX(x) = 0$ for symmetric positive semidefinite matrices $X(x)$ and Z.

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

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

and

$$X(x) \succ 0, \qquad Z \succ 0.$$

3 Algorithm for finding a KKT point

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

$$\|r(w, \mu)\| = \sqrt{\left\| \begin{pmatrix} \nabla_x L(w) \\ g(x) \end{pmatrix} \right\|^2 + \|X(x)Z - \mu I\|_F^2},$$

where $\|\cdot\|$ on the righthand side denotes the l_2 norm for vectors and $\|\cdot\|_F$ denotes the Frobenius norm for matrices. We also define $\|r_0(w)\|$ by $\|r_0(w)\| = \|r(w, 0)\|$.

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 call condition (5) the approximate BKKT condition, and call a point that satisfies this condition the approximate BKKT point. 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.

Remark 1 The procedure in Step 1 of Algorithm SDPIP will be given as Algorithm SDPLS in Sect. 4.3. Thus Algorithm SDPIP is an outer iteration, while Algorithm SDPLS is its inner iteration in which the previous approximate BKKT point w_k can be used as an initial point to find a new approximate BKKT point w_{k+1}.

If the matrix $A_0(x_*)$ is of full rank and there exists a nonzero vector $v \in \mathbf{R}^n$ such that

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

then we say that the Mangasarian-Fromovitz constraint qualification (MFCQ) condition is satisfied at a point x_* (see [4] for example). The following theorem shows the convergence property of Algorithm SDPIP under the MFCQ condition.

Theorem 1 *Assume that the functions f, g and X 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

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

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

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

where $X_* = X(x_*)$. 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:

$$\overline{X}_* \equiv PX_*P^T = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) \quad \text{and} \quad \overline{Z}_* \equiv PZ_*P^T = \mathrm{diag}(\tau_1, \ldots, \tau_p),$$

where $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$ and $\tau_1 \le \tau_2 \le \cdots \le \tau_p$ 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 $v \in \mathbf{R}^n$ which satisfies

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

Therefore, we have

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

for $j = 1, \ldots, p$, where $\overline{A}_i(x_*) = PA_i(x_*)P^T$. Since the following holds

$$0 = \lambda_j = (\overline{X}_*)_{jj} \quad \text{for} \quad j = 1, \ldots, p',$$

Eq. (8) yields

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

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

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

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

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

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. □

4 Algorithm for finding a barrier KKT point

In this section, we propose an algorithm that approximately finds a BKKT point for a given fixed barrier parameter $\mu > 0$. The algorithm given below is to be used as an inner iteration of Algorithm SDPIP.

As in the case of linear SDP problems, we consider a scaling of the primal–dual pair $(X(x), Z)$ in applying the Newton method to the system of equations (4). In what follows, we denote $X(x)$ simply by X if it is not confusing. Throughout this section, we assume that $X \succ 0$ and $Z \succ 0$ hold. We introduce a nonsingular matrix $T \in \mathbf{R}^{p \times p}$ and scale X and Z by

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

respectively. Using the scaling matrix T, we replace the equation $XZ = \mu I$ by a form $\widetilde{X} \circ \widetilde{Z} = \mu I$, and deal with the scaled symmetrized residual:

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

instead of (4) to form Newton directions as described in Sect. 4.1. We will propose a new merit function in Sect. 4.2 and will summarize our algorithm in Sect. 4.3.

4.1 Newton method

This subsection describes a Newton-like method to the system of equations (10). Let the Newton directions for the primal and dual variables be $\Delta x \in \mathbf{R}^n$ and $\Delta Z \in \mathbf{S}^p$, respectively. We define $\Delta X = \sum_{i=1}^{n} \Delta x_i A_i(x)$, and note that $\Delta X \in \mathbf{S}^p$. 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 following Newton equations for (10):

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

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

Similarly to usual primal–dual interior point methods for linear SDP problems, we derive an explicit form of the direction $\Delta Z \in \mathbf{S}^p$ from Eq. (13) and substitute it into Eq. (11) to obtain the Newton direction $\Delta w = (\Delta x, \Delta y, \Delta Z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$. For this purpose, we make use of various useful relations described in [1] and Appendix of [21]. For $U \in \mathbf{S}^p$, nonsingular $P \in \mathbf{R}^{p\times p}$ and $Q \in \mathbf{R}^{p\times p}$, we define the operator

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

and the symmetrized Kronecker product

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

where the operator svec is defined by

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

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

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

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

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

We also define

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

The expressions above give

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

and

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

Furthermore, we have

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

and

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

Now we have the following theorem that gives the desired form of Newton directions.

Theorem 2 *Suppose that the operator* $\widetilde{X} \odot I$ *is invertible. Then the direction* $\Delta\widetilde{Z} \in \mathbf{S}^p$ *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 $(\Delta x, \Delta y) \in \mathbf{R}^n \times \mathbf{R}^m$ *satisfy*

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

where the elements of the matrix $H \in \mathbf{R}^{n\times n}$ *are represented by the form*

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

with $\widetilde{A}_i(x) = T A_i(x) T^T$.

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* $\Delta w = (\Delta x, \Delta y, \Delta Z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$.

Proof By Eq. (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 Eq. (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 [21]), it follows from (18) and (17) that for any $U \in \mathbf{S}^p$,

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

By (15) and (16), Eq. (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 Eq. (21). Then we have

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

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

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

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

Furthermore, it is well known that the coefficient matrix of Eq. (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 Sect. 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 $A_1(x), \ldots, A_n(x)$ are linearly independent, which means that $\sum_{i=1}^{n} v_i A_i(x) = 0$ implies $v_i = 0, i = 1, \ldots, n$.

Theorem 3 *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 $A_i(x), i = 1, \ldots, n$ 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 If $\widetilde{X}$ is symmetric positive definite, then the operator $\widetilde{X}\odot I$ is invertible (see Appendix 9 of [21]). Let $\widetilde{U} = \sum_{i=1}^{n} u_i\widetilde{A}_i(x)$ for any $u(\neq 0) \in \mathbf{R}^n$. Since the linear independence of the matrices $A_i(x)$ for $i = 1, \ldots, n$ is equivalent to the linear independence of the matrices $\widetilde{A}_i(x)$ for $i = 1, \ldots, n$, $u \neq 0$ 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

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

From Property 6 of symmetrized Kronecker product in Appendix of [21] and relation (14), we have

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

It follows from Property 11 of symmetrized Kronecker product in Appendix of [21] that if $\widetilde{X}$ and $\widetilde{Z}$ are symmetric positive definite, then so are $\widetilde{X} \otimes_S \widetilde{Z}$ and $\widetilde{Z} \otimes_S \widetilde{X}$. It also follows from Property 9 that if $\widetilde{X}\widetilde{Z} + \widetilde{Z}\widetilde{X}$ is symmetric positive semidefinite, then so is $(\widetilde{X}\widetilde{Z} + \widetilde{Z}\widetilde{X}) \otimes_S I$. 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 $u,\ v \in \mathbf{R}^n$, we define

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

Then in a similar way to the above, we obtain

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

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 [21].

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

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

Since Eq. (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 scaling matrix T that satisfy such a condition.

(i) HRVW/KSH/M direction
If we set $T = X^{-1/2}$, then we have $\widetilde{X} = I$ and $\widetilde{Z} = X^{1/2}ZX^{1/2}$, which corresponds to the HRVW/KSH/M direction for linear SDP problems [8,11,15]. In this case, the matrices H and ΔZ can be represented by the form:

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

(ii) NT direction
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 $\widetilde{X} = W^{-1/2}XW^{-1/2} = W^{1/2}ZW^{1/2} = \widetilde{Z}$, which corresponds to the NT direction for linear SDP problems [16,17]. In this case, the matrices H and ΔZ can be represented by the form:

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

4.2 Primal–dual merit function

In what follows, we assume that the scaling matrix T is so chosen that $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$ is satisfied. To obtain the global convergence property of the algorithm described in Sect. 4.1, we use the line search strategy and propose the following merit function in the primal–dual space:

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

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

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

where ν and ρ are positive parameters. Though the functions $F_{BP}(x)$ and $F_{PD}(x, Z)$ depend on the parameters ν, ρ and μ, we use the notation $F(x, Z)$ for simplicity. It follows from the fact $\widetilde{X}\widetilde{Z} = TXZT^{-1}$ that $\langle \widetilde{X}, \widetilde{Z} \rangle = \langle X, Z \rangle$ 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 $XZ = \mu I$ is satisfied. Furthermore, the following hold

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

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

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

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 $\widetilde{X} = P D_X P^T$ and $\widetilde{Z} = P D_Z P^T$ for an orthogonal matrix P, where D_X and D_Z are diagonal matrices whose diagonal elements are λ_i and $\tau_i,\ i = 1, \ldots, p$, respectively. Thus, by noting the relation $\widetilde{X}\widetilde{Z} = P D_X D_Z P^T$, we can show that $\widetilde{X}\widetilde{Z} = \mu I$ is equivalent to the equations $\lambda_i \tau_i = \mu,\ i = 1, \ldots, p$. Furthermore, $\widetilde{X}\widetilde{Z} = \mu I$ is equivalent to $XZ = \mu I$. 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 Eq. (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. □

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

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

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 estimate an upper bound 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

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

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

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

By using the relations

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

Eq. (35) yields

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

Therefore the lemma is proved. □

Next we show that $\Delta F_{PDl}(x, Z;\Delta x, \Delta Z)$ is nonpositive 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* $XZ = \mu I$.

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

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

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. □

By using the preceding two lemmas, we obtain the following theorem, which 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. $\Delta F_l(x, Z; \Delta x, \Delta Z) \le 0$.*
(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

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

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 Eq. (37) yields

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

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

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

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

Since this means $(G + H)\Delta x = 0$, we have $\Delta x = 0$. Using Eq. (35) and Lemma 3 yields

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

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

In order to construct a globally convergent algorithm to a BKKT point for a fixed $\mu > 0$, we should 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. Throughout this section, the index k denotes the inner iteration count for a given $\mu > 0$. We note that $X_k \succ 0$ and $Z_k \succ 0$ are maintained for all k in the following. We also denote $X(x_k)$ by X_k for simplicity.

Since the main iteration is to decrease the value of the merit function (27), 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 initial step size by

$$\bar{\alpha}_{xk} = \begin{cases} -\frac{\gamma}{\lambda_{\min}(X_k^{-1}\Delta X_k)} & \text{if } X(x) \text{ is linear} \\ 1 & \text{otherwise} \end{cases} \tag{38}$$

and

$$\bar{\alpha}_{zk} = -\frac{\gamma}{\lambda_{\min}(Z_k^{-1}\Delta Z_k)}, \tag{39}$$

where $\lambda_{\min}(M)$ denotes the minimum eigenvalue of the matrix M, and $\gamma \in (0, 1)$ is a constant. If the minimum eigenvalue in either expression (38) or (39) is positive, we ignore the corresponding term. A step to the next iterate is given by

$$\alpha_k = \bar{\alpha}_k \beta^{l_k}, \quad \bar{\alpha}_k = \min\{\bar{\alpha}_{xk}, \bar{\alpha}_{zk}, 1\},$$

where $\beta \in (0, 1)$ is a constant, and l_k is the smallest nonnegative integer such that the sufficient decrease condition

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

and the positive definiteness condition

$$X(x_k + \bar{\alpha}_k \beta^{l_k} \Delta x_k) \succ 0 \tag{41}$$

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

Now we give a line search algorithm called Algorithm SDPLS. Since this algorithm should be regarded as the inner iteration of Algorithm SDPIP (see Step 1 of Algorithm SDPIP), ε' given below corresponds to $M_c\mu$ and an initial point can be set to the approximate BKKT point obtained at the previous outer iteration in Algorithm SDPIP.

Algorithm SDPLS

Step 0. (Initialize) Let $w_0 \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ ($X_0 \succ 0$, $Z_0 \succ 0$), $\mu > 0$, $\rho > 0$ and $\nu > 0$ be given. Set $\varepsilon' > 0$, $\gamma \in (0, 1)$, $\beta \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$. Let $k = 0$.
Step 1. (Termination) If $\|r(w_k, \mu)\| \leq \varepsilon'$, then stop.
Step 2. (Compute direction) Calculate the matrix G_k and the scaling matrix 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 criteria (40) and (41), 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 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

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

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

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

Remark 3 If we want to use the Hessian matrix $\nabla_x^2 L(w_k)$ for the matrix G_k, we adopt the Levenberg-Marquardt type modification of $\nabla_x^2 L(w_k)$ to obtain a positive semidefinite G_k for global convergence property shown in the next section. Namely, we compute a parameter $\sigma \geq 0$ which gives a positive semidefinite matrix $\nabla_x^2 L(w_k) + \sigma I$. The procedure used in the numerical experiments in Sect. 6 is as follows (see also [19]):

Step 0. Calculate the Cholesky decomposition of $\nabla_x^2 L(w_k)$. If it is successful, set $\sigma = 0$, and stop. If not, set $\sigma = 1.0$, and go to Step 1.
Step 1. Calculate the Cholesky decomposition of $\nabla_x^2 L(w_k) + \sigma I$. If it is successful, go to Step 2. Otherwise go to Step 3.
Step 2. Repeat $\sigma := \sigma/2$ until the Cholesky decomposition fails. Set $\sigma := 2\sigma$, and stop.
Step 3. Repeat $\sigma := 2\sigma$ until the Cholesky decomposition succeeds. Stop. □

This method is used to solve large scale nonconvex problems in our experiments.

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 $f, g_i, i = 1, \ldots, m$, and X are twice continuously differentiable.
(A2) The sequence $\{x_k\}$ generated by Algorithm SDPLS remains in a compact set Ω of $\mathbf{R}^n$.
(A3) For all x_k in Ω, the matrix $A_0(x_k)$ is of full rank and the matrices $A_1(x_k), \ldots, A_n(x_k)$ are linearly independent.
(A4) The matrix G_k is uniformly bounded and positive semidefinite.
(A5) The scaling matrix 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. □

Remark 4 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). 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. Though (A6) requires that the penalty parameter ρ be sufficiently large, the value of ρ is increased if necessary in practical computation (see Sect. 5.1.7 of [29] for the detailed procedure).

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 [26].

Lemma 4 *Let $d_x \in \mathbf{R}^n$ and $D_z \in \mathbf{R}^{p\times p}$ 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 $X(x + d_x) \succ 0$ 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*

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

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 $\langle X_k, Z_k\rangle \geq 0$ 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) From Appendix 9 of [21], the operator $\widetilde{X} \odot I$ is invertible. For the vector V defined in the proof of Theorem 3, $\mathrm{svec}(V)$ can be represented by the form

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

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

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

and

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

we have

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

Therefore it follows from (25) that

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

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 $\{w_k\}$ and the uniformly positive definiteness of $\{X_k\}$ and $\{Z_k\}$ guarantee 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 $A_i(x_k)$ for $i = 1, \ldots, n$ is equivalent to the linear independence of the vectors $\mathrm{svec}(\widetilde{A}_i(x_k))$ for $i = 1, \ldots, n$, the matrix $\widetilde{A}(x_k)$ 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 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 $\Delta x_k \neq 0$ 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 a 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. By Lemma 5(iv), $\|\Delta w_k\|$ is uniformly bounded above. Hence, we have $\liminf_{k\to\infty} \bar{\alpha}_k > 0$. Furthermore, the sequence $\{l_k\}$ that satisfies $X(x_k + \bar{\alpha}_k \beta^{l_k} \Delta x_k) \succ 0$ is uniformly bounded above.

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 $v \in \mathbf{R}^n$ 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

$$\begin{aligned} F(u_{k+1}) - F(u_k) &\le \varepsilon_0 \bar{\alpha}_k \beta^{l_k} \Delta F_l(u_k; \Delta u_k) \\ &\le -\varepsilon_0 \bar{\alpha}_k \beta^{l_k} \frac{\|\Delta x_k\|^2}{M} \\ &< 0. \end{aligned} \tag{42}$$

Because the sequence $\{F(u_k)\}$ is monotonically decreasing and bounded below, the left-hand side of (42) converges to 0, which implies that

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

We consider the following two cases:

(i) If there exists a finite number N such that $l_k < N$ for all k, then we have $\lim_{k\to\infty} \Delta F_l(u_k; \Delta u_k) = 0$ clearly.

(ii) Next we consider the case where 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 $u_k + \theta'_k \alpha_k \Delta u_k/\beta$ does not satisfy condition (40) for some $\theta'_k \in (0, 1)$. Thus, we get

$$F(u_k + \theta'_k \alpha_k \Delta u_k/\beta) - F(u_k) > \varepsilon_0 \theta'_k \alpha_k \Delta F_l(u_k; \Delta u_k)/\beta. \tag{43}$$

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

$$\begin{aligned} F(u_k + \theta'_k \alpha_k \Delta u_k/\beta) - F(u_k) &\le \theta'_k \alpha_k F'(u_k + \theta_k \theta'_k \alpha_k \Delta u_k/\beta; \Delta u_k)/\beta \\ &\le \theta'_k \alpha_k \Delta F_l(u_k + \theta_k \theta'_k \alpha_k \Delta u_k/\beta; \Delta u_k)/\beta. \end{aligned} \tag{44}$$

Then, from (43) and (44), we see that

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

This inequality yields

$$\begin{aligned} &\Delta F_l(u_k + \theta_k \theta'_k \alpha_k \Delta u_k/\beta; \Delta u_k) - \Delta F_l(u_k; \Delta u_k) \\ &\quad > (\varepsilon_0 - 1)\Delta F_l(u_k; \Delta u_k) > 0. \end{aligned} \tag{45}$$

Thus by the fact $l_k \to \infty$, $k \in K$, we have $\alpha_k \to 0$ and then $\|\theta_k \theta'_k \alpha_k \Delta u_k/\beta\| \to 0, k \in K$, because $\|\Delta u_k\|$ is uniformly bounded. Here $\|\Delta u_k\|$ is defined by

$$\|\Delta u_k\| = \sqrt{\|\Delta x_k\|^2 + \|\Delta Z_k\|_F^2}.$$

This implies that the left-hand side of (45) 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

$$\lim_{k\to\infty} \Delta F_l(u_k; \Delta u_k) = 0. \tag{46}$$

Since Eq. (46) 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 Eqs. (37), (12) and Lemma 3 that

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

Therefore, Eq. (21) yields

$$\Delta Z_k \to 0.$$

By Eq. (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 desired 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 proposed algorithm of this paper is implemented and some numerical experiments are done in order to verify the theoretical results of the algorithm. The program is written in C++, and is run on 3.2 GHz 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 (outer iteration) where we set $M_c = 0.1$ and $\gamma = 0.9$, and the scaling matrix is set to be $T = X^{-1/2}$ at each iteration of Algorithm SDPLS (inner iteration). We solved various test problems (Problems (P1)–(P6)) in the following. In Problems (P5) and (P6), we used the Levenberg-Marquardt type algorithm given in Remark 3 in Sect. 4.3.

(P1) The first problem is Gaussian channel capacity problem which is described in [24]:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}(\log\det(X + R) - \log\det R), \\ &\text{subject to} \quad \frac{1}{n}\text{tr}(X) \le P, \ \ X \succeq 0, \end{aligned}$$

where noise covariance $R \in \mathbf{S}^n$ is known and given, and input covariance $X \in \mathbf{S}^n$ is the variable to be determined. The parameter $P \in \mathbf{R}$ gives a limit on the average total power in the input. If all channels are independent, i.e., all covariances are diagonal, and the noise covariance depends on X as $R_{ii} = r_i + a_i X_{ii},\ a_i > 0$ (case of near-end cross-talk), the above problem can be written as

$$\begin{aligned}\text{minimize} \quad & \frac{1}{2}\sum_{i=1}^{n} \log\left(1 + \frac{X_{ii}}{r_i + a_i X_{ii}}\right),\\ \text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n} X_{ii} \le P,\ X_{ii} \ge 0.\end{aligned}$$

This problem can be transformed to the following SDP:

$$\begin{aligned}\text{minimize} \quad & \frac{1}{2}\sum_{i=1}^{n} \log(1 + t_i),\\ \text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n} X_{ii} \le P,\ X_{ii} \ge 0,\ t_i \ge 0,\\ & \begin{pmatrix} 1 - a_i t_i & \sqrt{r_i} \\ \sqrt{r_i} & a_i X_{ii} + r_i \end{pmatrix} \succeq 0,\ i = 1, \ldots, n.\end{aligned}$$

In our experiment, r_i and a_i are set to uniform random numbers between 0 and 1. P is set to 1. We solved problems with $n = 10, 20, \ldots, 10240$ using the exact Hessian of the Lagrangian as the matrix G. The numerical results are shown in Table 1 in which the total inner iteration counts and the run time (sec) are given.

(P2) The second problem is minimization of the minimal eigenvalue problem defined as:

Table 1 Gaussian channel capacity problem

n	Iteration	CPU (s)
10	28	0.03
20	26	0.17
40	31	0.11
80	39	0.32
160	48	1.07
320	52	3.8
640	40	10.2
1,280	44	41.3
2,560	38	137
5,120	43	607
10,240	45	2,559

$$\begin{array}{ll}\text{minimize} & \lambda_{\min}(M(q)),\\ \text{subject to} & q \in Q,\end{array}$$

where $q \in \mathbf{R}^n$, $Q \subset \mathbf{R}^n$, and $M \in \mathbf{S}^p$ is a function of q. We formulate this problem as follows:

$$\begin{array}{ll}\text{minimize} & \mathrm{tr}(\Pi M(q)),\\ \text{subject to} & \mathrm{tr}(\Pi) = 1,\\ & \Pi \succeq 0,\\ & q \in Q,\end{array}$$

where $\Pi \in \mathbf{S}^p$ is an additional matrix variable. In our experiment, we set $q = (x, y)^T$, and $M = xyA + xB + yC$ with given $A, B, C \in \mathbf{S}^p$. The elements of matrices A, B and C are set from uniform random numbers in $[-5, 5]$. The constraint region Q for the variable q is set to $[-1, 1] \times [-1, 1]$. We solved problems with the sizes of M, Π, A, B, C equal to 10, 20, 40, 80 respectively, with the BFGS quasi-Newton update for the matrix G. The numerical results are shown in Table 2 in which the total inner iteration counts and the run time (sec) are given.

(P3) The third problem is a real financial one and taken from [12]. 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

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

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

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

Table 2 Minimization of the minimal eigenvalue problem

p	Iteration	CPU (s)
10	30	0.12
20	32	0.88
40	69	46.9
80	56	1, 176

Table 3 Logit model/Example 1: number of variables = 28, $q = 6$, $M = 6{,}084$, $M_0 = 6{,}053$

Algorithm	Final objective	Final $\lambda_{\min}(Q)$	Iteration	Time (s)
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

Table 4 Logit model/Example 2: number of variables = 45, $q = 8$, $M = 6{,}084$, $M_0 = 6{,}053$

Algorithm	Final objective	Final $\lambda_{\min}(Q)$	Iteration	Time (s)
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

In [12], 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 Tables 3 and 4, we list two examples. These tables show the results with both the BFGS update (bfgs) and the exact Hessian (hesse) for the matrix G. 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). Tables 5 and 6 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.

(P4) The fourth problem in our experiment is from the nearest correlation matrix problem:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|X - A\|_F\,, \\ \text{subject to} \quad & X \succeq \epsilon I, \\ & X_{ii} \;=\; 1,\ i = 1, \ldots, n, \end{aligned}$$

where $A \in \mathbf{S}^n$ is given, and we want to obtain $X \in \mathbf{S}^n$ which is nearest to A and satisfies the given constraints. In the above problem, eigenvalues of X should not be less than $\epsilon > 0$, and the diagonals of X is equal to 1. There exist special purpose algorithms for solving this type of problem (e.g., [18]). In our experiments, we add additional constraints which gives an upper bound on the condition number of the matrix X:

Table 5 Logit model: iteration counts for each μ in Example 1

μ	bfgs	hesse
1.0e0	75	17
1.0e−1	25	2
1.0e−2	14	2
1.0e−3	4	2
1.0e−4	3	1
1.0e−5	3	2
1.0e−6	1	1
1.0e−7	1	1

Table 6 Logit model: iteration counts for each μ in Example 2

1.0e0	150	19
1.0e−1	35	3
1.0e−2	23	2
1.0e−3	9	1
1.0e−4	11	2
1.0e−5	3	2
1.0e−6	2	1
1.0e−7	1	1

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|X - A\|_F\,, \\ \text{subject to} \quad & zI \preceq X \preceq yI, \\ & y \le \kappa z, \;\; z \ge \epsilon \\ & X \succeq \epsilon I, \\ & X_{ii} = 1, \quad i = 1, \ldots, n, \end{aligned}$$

where y and z denote the maximal and minimal eigenvalue of X respectively, and the upper bound of their ratio (condition number) $\kappa > 0$ is given. Elements of the matrix A are uniform random numbers in $[-1, 1]$ with $A_{ii} = 1, i = 1, \ldots, n$. We set $\epsilon = 10^{-3}$, $\kappa = 10.0$. Results of various values of n are given in Table 7, where the exact Hessian is used for the matrix G.

(P5) The fifth problem area is the so called static output feedback (SOF) problems from $COMPl_eib$ library [14]. The following is the SOF-$\mathcal{H}_2$ type problem:

Table 7 Nearest correlation matrix problem

n	Iteration	CPU (s)
10	22	0.05
20	19	0.80
40	18	24.88
80	19	594.08

$$\begin{aligned}&\text{minimize } \operatorname{tr}(X),\\ &\text{subject to } Q \succeq 0,\\ &\qquad A(F)Q + QA(F)^T + B_1B_1^T \preceq 0,\\ &\qquad \begin{pmatrix} X & C(F)Q \\ QC(F)^T & Q \end{pmatrix} \succeq 0,\end{aligned}$$

where $X \in \mathbf{S}^{n_z \times n_z}$, $F \in \mathbf{R}^{n_u \times n_y}$ and $Q \in \mathbf{S}^{n_x \times n_x}$ are variable matrices to be determined. The matrices $A \in \mathbf{R}^{n_x \times n_x}$, $B \in \mathbf{R}^{n_x \times n_u}$, $B_1 \in \mathbf{R}^{n_x \times n_w}$, $C \in \mathbf{R}^{n_y \times n_x}$, $C_1 \in \mathbf{R}^{n_z \times n_x}$, $D_{11} \in \mathbf{R}^{n_z \times n_w}$, $D_{12} \in \mathbf{R}^{n_z \times n_u}$ and $D_{21} \in \mathbf{R}^{n_y \times n_w}$ are given constant matrices, and form the matrices $A(F)$, $B(F)$, $C(F)$, $D(F)$ which appear in the problem definition as follows:

$$\begin{aligned} A(F) &= A + BFC,\\ B(F) &= B_1 + BFD_{21},\\ C(F) &= C_1 + D_{12}FC,\\ D(F) &= D_{11} + D_{12}FD_{21}.\end{aligned}$$

The initial interior points are not known for this type of problem, and it turns out that it is not easy to find them. So we try various starting points, and solve the problems for which we can find initial interior points. We list the results for these problems in Table 8. Iterations are stopped when the norm of KKT conditions is less than 10^{-6}. In [19], numerical results for these problems performed by PENBMI, a specialized BMI-version of PENNON is reported. We list CPU data of PENBMI multiplied by a factor 2.5/3.2 which is a ratio of CPU speeds used in two experiments. We note that the various conditions of these experiments are not equal, so the PENBMI's CPU data is listed to crudely observe how our algorithm performs compared with PENBMI. The CPU data with * means that the norm tolerance is set to 10^{-5}.

We next describe the results for SOF-$\mathcal{H}_\infty$ problem which is defined by the following:

$$\begin{aligned}&\text{minimize } \gamma,\\ &\text{subject to } Q \succeq 0,\\ &\qquad \gamma \geq 0,\\ &\qquad \begin{pmatrix} A(F)^TQ + QA(F) & QB(F) & C(F)^T \\ B(F)^TQ & -\gamma I & D(F)^T \\ C(F) & D(F) & -\gamma I \end{pmatrix} \succeq 0,\end{aligned}$$

where $Q \in \mathbf{S}^{n_x \times n_x}$ and $F \in \mathbf{R}^{n_u \times n_y}$ are variable matrices to be determined. As in the SOF-$\mathcal{H}_2$ type problems, we report the results for problems with feasible initial point obtained in Table 9.

(P6) The last set of problems is obtained from SDPLIB to check our algorithms for large scale problems. SDPLIB is a library for linear SDP problems (see [2]). We add the quadratic term $\frac{1}{2}x^TQx$ to the original linear objective function c^Tx to

Table 8 SOF-$\mathcal{H}_2$ problem

Problem	n	n_x	n_y	n_u	n_w	n_z	Iteration	CPU (s)	CPU (PENBMI)
AC1	27	5	3	3	3	2	38	0.11	0.62
AC2	39	5	3	3	3	5	138	0.64	1.25
AC3	38	5	4	2	5	5	41	0.19	0.56
AC6	64	7	4	2	7	7	68	0.69	2.53
AC17	22	4	2	1	4	4	117	0.26	0.27
HE1	15	4	1	2	2	2	174	0.31	0.17
HE2	24	4	2	2	4	4	33	0.09	0.59
HE3	115	8	6	4	1	10	269	7.94	1.53
REA1	26	4	3	2	4	4	76	0.21	0.74
DIS1	88	8	4	4	1	8	47	0.93	5.04
DIS2	16	3	2	2	3	3	43	0.08	0.18
DIS3	58	6	4	4	6	6	252	2.33	1.93
DIS4	66	6	6	4	6	6	30	0.38	2.91
AGS	160	12	2	2	12	12	43	2.28	130
BDT1	96	11	3	3	1	6	46	1.07∗	2.78
MFP	26	4	2	3	4	4	112	0.33	0.46
EB1	59	10	1	1	2	2	55	0.68	16.2
EB2	59	10	1	1	2	2	50	0.61	21.0
PSM	49	7	3	2	2	5	46	0.29	2.01
NN2	7	2	1	1	2	2	27	0.03	0.22
NN4	26	4	3	2	4	4	32	0.09	0.30
NN8	16	3	2	2	3	3	63	0.12	0.27
NN11	157	16	5	3	3	3	188	12.19∗	223
NN15	20	3	2	2	1	4	64	0.13	0.27
NN16	62	8	4	4	8	4	124	1.51	36.4

Table 9 SOF-$\mathcal{H}_\infty$ problems

Problem	n	n_x	n_y	n_u	n_w	n_z	Iteration	CPU (s)	CPU (PENBMI)
AC4	13	4	2	1	2	2	188	0.34	0.64
HE2	15	4	2	2	4	4	64	0.15	0.13
DIS2	11	3	2	2	3	3	156	0.24	8.00
AGS	83	12	2	2	12	12	116	6.84	3.27
MFP	17	4	2	3	4	4	102	0.27	0.42
EB1	57	10	1	1	2	2	277	4.63	1.43
EB2	57	10	1	1	2	2	74	1.21	1.79
PSM	35	7	3	2	2	5	78	0.39	0.58
NN2	5	2	1	1	2	2	27	0.03	0.06

Table 10 SDPLIB with nonlinear objective

Problem	n	p	Iteration	CPU (s)
arch8	174	335	51	14.10
control7	136	45	33	95.38
maxG11	800	800	27	252.44
mcp500-1	500	500	39	84.17
qap10	1,021	101	35	65.85
ss30	132	426	47	44.71
theta6	4,375	300	68	3,695.86
truss8	496	628	31	14.89

form nonlinear objective function $\frac{1}{2}x^T Qx + c^T x$ where the matrix Q is sparse and symmetric positive definite. The values of the diagonal elements of Q are set to 1, and those of the off-diagonal elements are uniform random numbers from [0, 1], and if generated random number is greater than 0.03, the value is set to 0. Therefore the density of nonzero elements of the matrix Q is approximately 3%. In Table 10, p denotes the size of the matrix that is constrained to be positive semidefinite.

From the above experiments for Problems (P1)–(P6), 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 [26].

7 Concluding remarks

In this paper, we have proposed a primal–dual interior point method for solving nonlinear semidefinite programming problems. Our method uses a commutative class of Newton-like directions. 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. Our numerical experiments show the practical efficiency of our method.

Analysis of the rate of convergence are studied by Yamashita and Yabe [28]. They showed the superlinear convergence of the primal–dual interior point method based on the unscaled Newton method, which corresponded to the case $T_k = I$, and the two-step superlinear convergence of the primal–dual interior point methods based on the scaled Newton methods, which corresponded to the cases (i) and (ii) discussed at the end of Sect. 4.1.

Acknowledgments The authors would like to thank Japan Credit Rating Agency, Ltd for providing us the results of their numerical experiments. The second author was supported in part by the Grant-in-Aid for Scientific Research (C) 21510164 of Japan Society for the Promotion of Science. The authors also would like to thank two anonymous referees for their valuable comments.

References

1. Alizadeh, F., Haeberly, J.A., Overton, M.L.: Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. SIAM J. Optim. **8**, 746–768 (1998)
2. Borchers, B.: SDPLIB 1.2, a library of semidefinite programming test problems. Optim. Methods Softw. **11** & **12**, 683–690 (1999)
3. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
4. Correa, R., Ramirez, C.H.: A global algorithm for nonlinear semidefinite programming. SIAM J. Optim. **15**, 303–318 (2004)
5. Fares, B., Apkarian, P., Noll, D.: An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. Int. J. Control **74**, 348–360 (2001)
6. Fares, B., Noll, D., Apkarian, P.: Robust control via sequential semidefinite programming. SIAM J. Control Optim. **40**, 1791–1820 (2002)
7. Freund, R.W., Jarre, F., Vogelbusch, C.H.: Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. Math. Program. **109**, 581–611 (2007)
8. Helmberg, C., Rendl, F., Vanderbei, R.J., Wolkowicz, H.: An interior-point method for semidefinite programming. SIAM J. Optim. **6**, 342–361 (1996)
9. Jarre, F.: An interior method for nonconvex semidefinite programs. Optim. Eng. **1**, 347–372 (2000)
10. Kanzow, C., Nagel, C., Kato, H., Fukushima, M.: Successive linearization methods for nonlinear semidefinite programs. Comput. Optim. Appl. **31**, 251–273 (2005)
11. Kojima, M., Shindoh, S., Hara, S.: Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. SIAM J. Optim. **7**, 86–125 (1997)
12. Konno, H., Kawadai, N., Wu, D.: Estimation of failure probability using semi-definite logit model. Comput. Manage. Sci. **1**, 59–73 (2003)
13. Kocvara, M., Stingl, M.: PENNON: a code for convex nonlinear and semidefinite programming. Optim. Methods Softw. **18**, 317–333 (2003)
14. Leibfritz, F., Lipinski, W.: Description of the benchmark examples in $COMPl_eib$ 1.0. Technical Report, Department of Mathematics, University of Trier, D-54286, Trier, Germany (2003)
15. Monteiro, R.D.C.: Primal-dual path-following algorithms for semidefinite programming. SIAM J. Optim. **7**, 663–678 (1997)
16. Nesterov, Y.E., Todd, M.J.: Self-scaled barriers and interior-point methods for convex programming. Math. Oper. Res. **22**, 1–42 (1997)
17. Nesterov, Y.E., Todd, M.J.: Primal-dual interior-point methods for self-scaled cones. SIAM J. Optim. **8**, 324–364 (1998)
18. Qi, H., Sun, D.: A quadratically convergent Newton method for computing the nearest correlation matrix. SIAM J. Matrix Anal. Appl. **28**, 360–385 (2006)
19. Stingl, M.: On the solution of nonlinear semidefinite programs by augmented Lagrangian methods, Doctoral Thesis. University of Erlangen (2005)
20. Todd, M.J.: Semidefinite optimization. Acta Numer. **10**, 515–560 (2001)
21. Todd, M.J., Toh, K.C., Tütüncü, R.H.: On the Nesterov-Todd direction in semidefinite programming. SIAM J. Optim. **8**, 769–796 (1998)
22. Tseng, P.: Convergent infeasible interior-point trust-region methods for constrained minimization. SIAM J. Optim. **13**, 432–469 (2002)
23. Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Rev. **38**, 49–95 (1996)
24. Vandenberghe, L., Boyd, S., Wu, S.-P.: Determinant maximization with linear matrix inequality constraints. SIAM J. Matrix Anal. Appl. **19**, 499–533 (1998)
25. Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.): Handbook of Semidefinite Programming: Theory, Algorithms and Applications, Kluwer International Series in Operations Research and Management Science. Kluwer, Boston (2000)
26. Yamashita, H.: A globally convergent primal–dual interior point method for constrained optimization. Optim. Methods Softw. **10**, 443–469 (1998)
27. Yamashita, H., Yabe, H.: A primal-dual interior point method for nonlinear semidefinite programming. Inst. Stat. Math. Cooper. Res. Report **203**, 296–317 (2007)
28. Yamashita, H., Yabe, H.: Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming. Math. Program. (2011). doi:10.1007s10107-010-354-x
29. Yamashita, H., Yabe, H., Tanabe, T.: A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization. Math. Program. **102**, 111–151 (2005)