An interior point method with a primal-dual quadratic barrier penalty function for nonlinear semidefinite programming

Atsushi Kato [*] and   Hiroshi Yabe [†] and   Hiroshi Yamashita [‡]

August 20, 2011 (Revised June 1, 2012)

Abstract

In this paper, we consider an interior point method for nonlinear semidefinite programming. Yamashita, Yabe and Harada presented a primal-dual interior point method in which a nondifferentiable merit function was used. By using shifted barrier KKT conditions, we propose a differentiable primal-dual merit function within the framework of the line search strategy, and prove the global convergence property of our method.

Keyword.   Nonlinear semidefinite programming, Primal-dual interior point method, Primal-dual quadratic barrier penalty function, Global convergence

1   Introduction

In this paper, we consider the following nonlinear semidefinite programming (SDP) problem:

$$\begin{cases} \min & f(x), & x \in \mathbf{R}^n, \\ \text{s.t.} & g(x) = 0, & X(x) \succeq 0, \end{cases} \qquad (1.1)$$

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.

Numerical methods which solve the nonlinear SDP have been developed and studied by several authors [3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 17, 21]. Drummond, Iusem and Svaiter [4] studied the central path of a nonlinear convex SDP and discussed the behavior between the logarithmic barrier function and objective function. Leibfritz and Mostafa

---

[*]Department of Mathematical Information Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan. e-mail: j1410702@ed.kagu.tus.ac.jp

[†]Department of Mathematical Information Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan. e-mail: yabe@rs.kagu.tus.ac.jp

[‡]Mathematical Systems Inc., 2-4-3, Shinjuku, Shinjuku-ku, Tokyo 160-0022, Japan e-mail: hy@msi.co.jp

[15] proposed an interior point trust region algorithm. Huang, Yang and Teo [9] transformed a nonlinear SDP into a mathematical program with a matrix equality constraint and proposed the sequential quadratic penalty method. Correa and Ramirez [3] presented the sequentially semidefinite programming local method which is based on the sequential quadratic programming method with a line search strategy. Kanzow, Nagel, Kato and Fukushima [12] proposed the trust region-type method. This method solves the subproblem which can be converted to a linear semidefinite program at each iteration. Yamashita, Yabe and Harada [21] proposed a primal-dual interior point method with a line search strategy. Huang, Teo and Yang [10] introduced an approximate augmented Lagrangian function for nonlinear SDP. Huang, Yang and Teo [11] considered a nonlinear SDP with a matrix equality constraint and proposed the penalty method. Gomez and Ramirez [7] proposed a filter method for nonlinear SDP.

Though Yamashita, Yabe and Harada [21] derived a primal-dual merit function, their function was not differentiable. Thus it is significant to consider a differentiable primal-dual merit function. In this paper, following the idea of Yamashita and Yabe [20] for nonlinear programming problems, we propose a primal-dual interior point method whose merit function is differentiable for nonlinear SDP problems. Our primal-dual interior point method can obtain the global convergence property under weaker assumptions than those in [21].

This paper is organized as follows. In Section 2, we introduce the optimality conditions for problem (1.1) and some notations, and briefly review the primal-dual interior point method of Yamashita et al. [21]. In Section 3, we describe how to get a search direction and a step size. We prove the global convergence of the proposed method in Section 4.

In what follows, $\| \bullet \|$, $\| \bullet \|_1$ and $\| \bullet \|_\infty$ denote the l_2, l_1 and l_∞ norms for vectors. $\| \bullet \|_F$ denotes the Frobenius norm for matrices. We define the following norm

$$\left\| \begin{pmatrix} s \\ t \\ U \end{pmatrix} \right\|_* = \sqrt{\left\| \begin{pmatrix} s \\ t \end{pmatrix} \right\|^2 + \|U\|_F^2}$$

for $(s, t, U) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$. The superscript T denotes the transpose of a vector or a matrix. I denotes the identity matrix. $\mathrm{tr}(M)$ denotes the trace of the matrix M. We define the inner product $\langle M_1, M_2 \rangle$ by $\langle M_1, M_2 \rangle = \mathrm{tr}(M_1 M_2^T)$ for any matrices M_1 and M_2 in $\mathbf{R}^{p \times p}$.

2 Optimality conditions and algorithm for finding a KKT point

In this section, we introduce the optimality conditions for problem (1.1) and Algorithm SDPIP that finds a KKT point.

At first, we confirm the optimality conditions for problem (1.1). Define the Lagrangian function of problem (1.1) as follows.

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

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. The

first-order necessary conditions for optimality of problem (1.1), which is called the Karush-Kuhn-Tucker (KKT) conditions, are given by (see [2]):

$$r_0(w) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ X(x)Z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.1}$$

and

$$X(x) \succeq 0, \qquad Z \succeq 0, \tag{2.2}$$

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

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

Here the matrix $\nabla g(x)$ is given by

$$\nabla g(x) = (\nabla g_1(x) \; \cdots \; \nabla g_m(x)) \in \mathbf{R}^{n \times m},$$

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

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

where matrices $A_i(x) \in \mathbf{S}^p$ are defined by

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

for $i = 1, \cdots, n$.

Given a positive barrier parameter μ, interior point methods usually replace the complementarity condition $X(x)Z = 0$ by $X(x)Z = \mu I$ and deal with the barrier KKT conditions

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

and

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

instead of conditions (2.1) and (2.2). These conditions are derived from the necessary condition for optimality of the following problem:

$$\min_{X(x) \succ 0} \; F_{BP1}(x, \mu) = f(x) + \rho \|g(x)\|_1 - \mu \log(\det X(x)), \quad x \in \mathbf{R}^n, \tag{2.3}$$

where ρ is a given positive penalty parameter (see [19]).

In connection with (1.1), we consider the following problem:

$$\min_{X(x) \succ 0} \; F_{BP2}(x, \mu) = f(x) + \frac{1}{2\mu} \|g(x)\|^2 - \mu \log(\det X(x)), \quad x \in \mathbf{R}^n. \tag{2.4}$$

The necessary conditions for the optimality of this problem are given by

$$\nabla F_{BP2}(x, \mu) = \nabla f(x) + \frac{1}{\mu} \nabla g(x) g(x) - \mu \mathcal{A}^*(x) X^{-1}(x) = 0, \tag{2.5}$$

and $X(x) \succ 0$. We define the variables y and Z by $y = -g(x)/\mu$ and $Z = \mu X(x)^{-1}$. Then the above conditions are written as

$$r(w, \mu) = \begin{pmatrix} \nabla_x L(w) \\ g(x) + \mu y \\ X(x)Z - \mu I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.6}$$

and

$$X(x) \succ 0, \qquad Z \succ 0. \tag{2.7}$$

We call these conditions the shifted barrier KKT (SBKKT) conditions, the point which satisfies SBKKT conditions the SBKKT point and the point which satisfies conditions (2.7) the interior point. It should be noted that we deal with x, y and Z as independent variables and that condition (2.6) reduces to condition (2.1) when $\mu = 0$, i.e., $r_0(w) = r(w, 0)$.

In order to obtain a KKT point, we propose the following algorithm, which uses the SBKKT conditions.

Algorithm SDPIP

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

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

$$\|r(w_{k+1}, \mu_k)\|_* \leq M_c \mu_k. \tag{2.8}$$

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

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

In contrast to the interior point method in [21], stopping criterion (2.8) contains $g(x) + \mu y$. We should note that the global convergence property of Algorithm SDPIP can be shown in the same way as Theorem 1 of [21].

3 How to obtain an approximate SBKKT point

Algorithm SDPIP given in the previous section needs to find an approximate SBKKT point at each iteration. In this section, we propose an algorithm to obtain such a point, which will be described as Algorithm SDPLS at the end of this section. Algorithm SDPLS uses the following iterative scheme:

$$w_{k+1} = w_k + \alpha_k \Delta w_k,$$

where the subscript k denotes an iteration count of Algorithm SDPLS, $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta Z_k) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ is the k-th search direction and $\alpha_k > 0$ is the k-th step size.

In what follows, we denote $X(x)$ simply by X if it is not confusing. Note that an initial interior point w_0 and a fixed barrier parameter $\mu > 0$ are taken over from Algorithm SDPIP.

This section is organized as follows. In Section 3.1, we introduce a Newton-like method to calculate a search direction Δw_k. In Section 3.2, we propose our merit function. In Section 3.3, we introduce how to obtain a suitable step size α_k which guarantees that the new point w_{k+1} is an interior point. In Section 3.4, we give Algorithm SDPLS.

3.1 How to obtain a Newton direction

In this subsection, we omit the subscript k and we assume that $X \succ 0$ and $Z \succ 0$. To obtain a search direction Δw, we apply a Newton-like method to the system of equations (2.6). However, it is difficult to express the Newton direction explicitly because $XZ = ZX$ generally dose not hold. To overcome this matter, we introduce a nonsingular matrix $T \in \mathbf{S}^p$ and scale X and Z by

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

respectively, so that $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$ holds. We replace the equation $XZ = \mu I$ by a form $\widetilde{X} \circ \widetilde{Z} = \mu I$, where the multiplication $\widetilde{X} \circ \widetilde{Z}$ is defined by

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

Instead of (2.6), we consider the following scaled symmetrized residual:

$$\begin{pmatrix} \nabla_x L(w) \\ g(x) + \mu y \\ \widetilde{X} \circ \widetilde{Z} - \mu I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.1}$$

Remark 3.1. *To achieve $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$, the scaling matrix T can be chosen as follows.*

(i) *If we set $T = X^{-1/2}$, then we have $\widetilde{X} = I$ and $\widetilde{Z} = X^{-1/2}ZX^{-1/2}$, which correspond to the HRVW/KSH/M direction for linear SDP problems [8, 14, 16].*

(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 $\widetilde{X} = W^{-1/2}XW^{1/2} = W^{1/2}ZW^{1/2} = \widetilde{Z}$, which correspond to the NT direction for linear SDP problems [18].*

Now we apply the Newton method to the nonlinear equations (3.1) and we can obtain the Newton step Δw by solving the following linear system of equations:

$$G\Delta x - \nabla g(x)\Delta y - \mathcal{A}^*(x)\Delta Z = -\nabla_x L(x, y, Z) \tag{3.2}$$
$$\nabla g(x)^T \Delta x + \mu \Delta y = -g(x) - \mu y \tag{3.3}$$
$$\Delta\widetilde{X}\widetilde{Z} + \widetilde{Z}\Delta\widetilde{X} + \widetilde{X}\Delta\widetilde{Z} + \Delta\widetilde{Z}\widetilde{X} = 2\mu I - \widetilde{X}\widetilde{Z} - \widetilde{Z}\widetilde{X}, \tag{3.4}$$

where G denotes the Hessian matrix of the Lagrangian function $\nabla_x^2 L(w)$ or its approximation, $\Delta X = \sum_{i=1}^n \Delta x_i A_i(x) \in \mathbf{S}^p$ and

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

Remark 3.2. *In practice, the matrix G approximates the Hessian matrix $\nabla_x^2 L(w)$ by the Levenberg-Marquardt type modification or the quasi-Newton updating formula (see Remarks 2 and 3 of* [21]).

The explicit form of Newton directions are given by (see [21])

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

and

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

where the elements of the matrix $H \in \mathbf{S}^p$ are represented by the form

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

with $\widetilde{A}_i(x) = TA_i(x)T^T$ and the operator $P \odot Q$ means that

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

for $U \in \mathbf{S}^p$, nonsingular $P \in \mathbf{R}^{p \times p}$ and $Q \in \mathbf{R}^{p \times p}$. Therefore, the equations (3.2) – (3.4) are rewritten by

$$\{G + H + \frac{1}{\mu}\nabla g(x)\nabla g(x)^T\}\Delta x = -\nabla F_{BP2}(x, \mu), \tag{3.5}$$

$$\Delta y = -\frac{1}{\mu}\{g(x) + \mu y + \nabla g(x)^T \Delta x\} \tag{3.6}$$

and

$$\Delta Z = \mu X^{-1} - Z - (T^T \odot T^T)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)\Delta X. \tag{3.7}$$

When the matrices $A_i(x)$ $(i = 1, \ldots, n)$ are linearly independent, the matrix H is positive definite (see [21]). Note that we can calculate easily a direction ΔZ and a matrix H from Remark 3.1.

The following lemma is obtained directly from the relationships (3.5) – (3.7).

Lemma 3.1. *If the matrix $G + H + \frac{1}{\mu}\nabla g(x)\nabla g(x)^T$ is nonsingular, then the Newton equations (3.2) – (3.4) give a unique search direction $\Delta w = (\Delta x, \Delta y, \Delta Z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$, which is given by (3.5) – (3.7).*

To obtain an unique direction, the interior point method in [21] needs the assumption that $\nabla g(x)$ is of full rank, while our interior point method do not need such an assumption.

3.2 Differentiable primal-dual merit function

Yamashita et al. [21] proposed a primal-dual interior point method which consists of the outer iteration that finds a KKT point and the inner iteration that calculates an approximate barrier KKT point. To globalize the algorithm, they introduced the primal-dual merit function, which is defined by

$$F_{\ell 1}(x, Z, \mu) = F_{BP1}(x, \mu) + \nu F_{PD1}(x, Z, \mu),$$

where ν is a positive parameter, the function $F_{BP1}(x, \mu)$ is defined by (2.3) and $F_{PD1}(x, Z, \mu)$ is defined by

$$F_{PD1}(x, Z, \mu) = \langle X, Z \rangle - \mu \log\{\det(XZ)\}.$$

However this merit function is not differentiable and a dual variable y is not contained. By taking account of these issues, we propose the following differentiable merit function in the whole primal-dual space:

$$F_{\ell 2}(w, \mu) = F_{BP2}(x, \mu) + \nu F_{PD2}(w, \mu), \tag{3.8}$$

where $F_{BP2}(x, \mu)$ is defined by (2.4) and $F_{PD2}(w, \mu)$ is the primal-dual barrier function which is given by

$$F_{PD2}(w, \mu) = \frac{1}{2}\|g(x) + \mu y\|^2 + \log \frac{\langle X, Z \rangle/p + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{\{\det(XZ)\}^{1/p}}. \tag{3.9}$$

Note that $\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2$ is rewritten by

$$
\begin{aligned}
\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 &= \operatorname{tr}\{(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I)(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I)^T\} \\
&= \operatorname{tr}\{(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I)(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I)\} \\
&= \operatorname{tr}(Z^{\frac{1}{2}} X Z X Z^{\frac{1}{2}}) - 2\mu \operatorname{tr}(Z^{\frac{1}{2}} X Z^{\frac{1}{2}}) + \mu^2 \operatorname{tr}(I) \\
&= \operatorname{tr}(XZXZ) - 2\mu \operatorname{tr}(XZ) + p\mu^2. \tag{3.10}
\end{aligned}
$$

Let λ_i and τ_i for $i = 1, \cdots, p$ denote the eigenvalues of the matrices $\widetilde{X}$ and $\widetilde{Z}$, respectively. Since $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$ holds, the matrices $\widetilde{X}$ and $\widetilde{Z}$ share the same eigensystem. Therefore, the matrix $\widetilde{X}\widetilde{Z}$ has eigenvalues $\lambda_i \tau_i$, $i = 1, \cdots, p$. i.e., the following holds

$$\langle X, Z \rangle = \sum_{i=1}^{p} \lambda_i \tau_i, \quad \det(XZ) = \prod_{i=1}^{p} \lambda_i \tau_i. \tag{3.11}$$

The function $F_{PD2}(w, \mu)$ has the following properties.

Lemma 3.2. *The following relationships hold:*

(i) $F_{PD2}(w, \mu) \geq 0$.

(ii) $F_{PD2}(w, \mu) = 0$ *if and only if* $g(x) + \mu y = 0$ *and* $XZ - \mu I = 0$.

Proof. To show this lemma, we separate $F_{PD2}(w, \mu)$ in (3.9) into the two functions $F_S(x, y, \mu)$ and $F_P(x, Z, \mu)$. They are given by

$$F_S(x, y, \mu) = \frac{1}{2}\|g(x) + \mu y\|^2$$

and

$$F_P(x, Z, \mu) = \log \frac{\langle X, Z \rangle/p + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{\{\det(XZ)\}^{1/p}}.$$

(i) The following obviously holds

$$F_S(x, y, \mu) \geq 0.$$

The equations (3.11) and arithmetic and geometric means yield

$$\frac{1}{p}\sum_{i=1}^{p} \lambda_i \tau_i + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 \geq \left(\prod_{i=1}^{p} \lambda_i \tau_i\right)^{\frac{1}{p}} = \{\det(XZ)\}^{\frac{1}{p}}.$$

Thus we have

$$F_P(x, Z, \mu) \geq \log 1 = 0.$$

(ii) Suppose that $g(x) + \mu y = 0$ and $XZ - \mu I = 0$. From $g(x) + \mu y = 0$, we obtain $F_S(x, y, \mu) = 0$. $XZ - \mu I = 0$ means $Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I = 0$. Since $\lambda_1 \tau_1 = \cdots = \lambda_p \tau_p = \mu$ holds, we get

$$\frac{1}{p}\sum_{i=1}^{p} \lambda_i \tau_i = \left(\prod_{i=1}^{p} \lambda_i \tau_i\right)^{1/p},$$

which implies $F_P(x, Z, \mu) = \log 1 = 0$. Therefore, $F_{PD2}(w, \mu) = 0$.

Conversely suppose that $F_{PD2}(w, \mu) = 0$. This means that

$$F_S(x, y, \mu) = \frac{1}{2}\|g(x) + \mu y\|^2 = 0, \quad F_P(x, Z, \mu) = \log 1 = 0.$$

The equality $g(x) + \mu y = 0$ clearly holds. It follows from (3.11) and arithmetic and geometric means that

$$\frac{1}{p}\sum_{i=1}^{p} \lambda_i \tau_i + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 = \left(\prod_{i=1}^{p} \lambda_i \tau_i\right)^{1/p} \leq \frac{1}{p}\sum_{i=1}^{p} \lambda_i \tau_i,$$

which implies

$$\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 \leq 0.$$

We obtain $XZ - \mu I = 0$. Therefore, the proof is complete. $\square$

The next lemma helps derive the derivative of $F_{\ell 2}(w, \mu)$.

Lemma 3.3. *The following holds for $i = 1, \cdots, n$:*

$$\frac{\partial}{\partial x_i}\text{tr}(XZXZ) = 2\langle A_i(x), ZXZ \rangle.$$

Proof. From the definition of the matrix $A_i(x) \in \mathbf{S}^p$, we obtain

$$
\begin{aligned}
\frac{\partial}{\partial x_i}\mathrm{tr}(XZXZ) &= \mathrm{tr}(\frac{\partial X}{\partial x_i}ZXZ) + \mathrm{tr}(XZ\frac{\partial X}{\partial x_i}Z) \\
&= \mathrm{tr}(A_i(x)ZXZ) + \mathrm{tr}(XZA_i(x)Z) \\
&= \mathrm{tr}(A_i(x)ZXZ) + \mathrm{tr}(A_i(x)ZXZ) \\
&= 2\mathrm{tr}(A_i(x)ZXZ) \\
&= 2\langle A_i(x), ZXZ \rangle.
\end{aligned}
$$

Therefore, the lemma is proven. $\square$

From Lemma 3.2, (3.8) and (3.10), we calculate the derivative of the merit function which is given by

$$
\nabla F_{\ell 2}(w, \mu) = \begin{pmatrix} \nabla F_{BP2}(x, \mu) + \nu \nabla_x F_{PD2}(w, \mu) \\ \nu \nabla_y F_{PD2}(w, \mu) \\ \nu \nabla_Z F_{PD2}(w, \mu) \end{pmatrix}, \tag{3.12}
$$

where

$$
\begin{aligned}
\nabla_x F_{PD2}(w, \mu) &= \frac{\frac{1}{p}\mathcal{A}^*(x)Z + 2\{\mathcal{A}^*(x)(ZXZ) - \mu\mathcal{A}^*(x)Z\}}{\langle X, Z \rangle/p + \|Z^{\frac{1}{2}}XZ^{\frac{1}{2}} - \mu I\|_F^2} \\
&\quad - \frac{1}{p}\mathcal{A}^*(x)X^{-1} + \nabla g(x)\{g(x) + \mu y\}, \tag{3.13} \\
\nabla_y F_{PD2}(w, \mu) &= \mu\{g(x) + \mu y\}, \tag{3.14} \\
\nabla_Z F_{PD2}(w, \mu) &= \frac{\frac{1}{p}X + 2(XZX - \mu X)}{\langle X, Z \rangle/p + \|Z^{\frac{1}{2}}XZ^{\frac{1}{2}} - \mu I\|_F^2} - \frac{1}{p}Z^{-1} \tag{3.15}
\end{aligned}
$$

and

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

Note that $\nabla F_{BP2}(x, \mu)$ is given in (2.5) and that $\nabla_x F_{PD2}(w, \mu)$ and $\nabla_y F_{PD2}(w, \mu)$ are derivatives of $F_{PD2}(w, \mu)$ with respect to vectors x and y, respectively. $\nabla_Z F_{PD2}(w, \mu)$ is a derivative of $F_{PD2}(w, \mu)$ with respect to a matrix Z (see [1]).

The following lemma shows that an SBKKT point is equivalent to a stationary point of the function $F_{\ell 2}(w, \mu)$.

Lemma 3.4. *The following statements are equivalent:*

(a) $r(w, \mu) = 0$.

(b) $\nabla F_{BP2}(x, \mu) = 0$, $g(x) + \mu y = 0$ *and* $XZ - \mu I = 0$.

(c) $\nabla F_{\ell 2}(w, \mu) = 0$.

Proof. The equivalence of (a) and (b) is obvious from (2.5) and (2.6).

Therefore, it suffices to show the equivalence of (b) and (c). Suppose that (b) is satisfied. From the equality $g(x) + \mu y = 0$, $\nabla_y F_{PD2}(w, \mu) = 0$ clearly holds. Since the expression $XZ - \mu I = 0$ implies that $X = \mu Z^{-1}$, $Z = \mu X^{-1}$ and $Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I = 0$, we have

$$\nabla_Z F_{PD2}(w, \mu) = \frac{\frac{1}{p}\mu Z^{-1} + 2(\mu X - \mu X)}{\mu} - \frac{1}{p} Z^{-1} = 0.$$

By (3.13), the expressions $g(x) + \mu y = 0$ and $XZ - \mu I = 0$ also yield

$$\nabla_x F_{PD2}(w, \mu) = \frac{\frac{1}{p}\mathcal{A}^*(x)(\mu X^{-1}) + 2\{\mu \mathcal{A}^*(x) Z - \mu \mathcal{A}^*(x) Z\}}{\mu} - \frac{1}{p}\mathcal{A}^*(x) X^{-1} = 0.$$

Therefore, (c) holds by (3.12).

Conversely assume that (c) is satisfied. By the assumption $\nabla_y F_{PD2}(w, \mu) = 0$, $g(x) + \mu y = 0$ is obvious. From $\nabla_Z F_{PD2}(w, \mu) = 0$, we get

$$\frac{\frac{1}{p}X + 2(XZX - \mu X)}{\langle X, Z \rangle / p + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2} - \frac{1}{p} Z^{-1} = 0.$$

Multiplying both sides of the above equation by X^{-1} from the right, we have

$$\frac{\frac{1}{p}I + 2(XZ - \mu I)}{\langle X, Z \rangle / p + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2} - \frac{1}{p} Z^{-1} X^{-1} = 0,$$

which implies

$$2(XZ - \mu I) = \frac{1}{p}\left\{ \frac{\langle X, Z \rangle}{p} + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 \right\} Z^{-1} X^{-1} - \frac{1}{p}I.$$

Multiplying both sides of the above equation by $Z^{\frac{1}{2}}$ from the left and $Z^{-\frac{1}{2}}$ from the right, we have

$$2(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I) = \frac{1}{p}\left\{ \frac{\langle X, Z \rangle}{p} + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 \right\} (Z^{\frac{1}{2}} X Z^{\frac{1}{2}})^{-1} - \frac{1}{p}I.$$

Multiplying both sides of the above equation by $(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I)^T$ from the right, we obtain

$$2(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I)(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I)^T$$
$$= \frac{1}{p}\left\{ \frac{\langle X, Z \rangle}{p} + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 \right\} \{I - \mu (Z^{\frac{1}{2}} X Z^{\frac{1}{2}})^{-1}\} - \frac{1}{p}(Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I).$$

Considering the trace of the above equation, we have

$$2\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2$$
$$= \frac{1}{p}\left\{ \frac{\langle X, Z \rangle}{p} + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 \right\} [p - \mu \mathrm{tr}\{(Z^{\frac{1}{2}} X Z^{\frac{1}{2}})^{-1}\}] - \frac{1}{p}\{\mathrm{tr}(Z^{\frac{1}{2}} X Z^{\frac{1}{2}}) - \mu n\}$$
$$= \left[1 - \frac{\mu}{p}\mathrm{tr}\{(Z^{\frac{1}{2}} X Z^{\frac{1}{2}})^{-1}\}\right] \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 + \mu\left\{1 - \frac{1}{p^2}\langle X, Z \rangle \langle X^{-1}, Z^{-1} \rangle\}\right\}.$$

Then it follows from arithmetic and geometric means that

$$\left[1 + \frac{\mu}{p}\mathrm{tr}\{(Z^{\frac{1}{2}}XZ^{\frac{1}{2}})^{-1}\}\right]\|Z^{\frac{1}{2}}XZ^{\frac{1}{2}} - \mu I\|_F^2 \;=\; \mu\left\{1 - \frac{\langle \widetilde{X}, \widetilde{Z}\rangle}{p}\frac{\langle \widetilde{X}^{-1}, \widetilde{Z}^{-1}\rangle}{p}\right\}$$

$$\leq \;\mu\left\{1 - \frac{(\prod_{i=1}^p \lambda_i\tau_i)^{1/p}}{(\prod_{i=1}^p \lambda_i\tau_i)^{1/p}}\right\}$$

$$= \; 0.$$

Because $\mathrm{tr}\{(Z^{\frac{1}{2}}XZ^{\frac{1}{2}})^{-1}\} > 0$ holds, we obtain $XZ - \mu I = 0$. From $g(x) + \mu y = 0$ and $XZ - \mu I = 0$, we get

$$\nabla_x F_{PD2}(w,\mu) = \frac{\frac{\mu}{p}\mathcal{A}^*(x)X^{-1} + 2\{\mu\mathcal{A}^*(x)Z - \mu\mathcal{A}^*(x)Z\}}{\mu} - \frac{1}{p}\mathcal{A}^*(x)X^{-1} = 0.$$

$\nabla F_{\ell 2}(w,\mu) = 0$ and $\nabla_x F_{PD2}(w,\mu) = 0$ imply $\nabla F_{BP2}(x,\mu) = 0$. Therefore, the proof is complete. □

Now we introduce the directional derivative $D(F_{\ell 2}(w,\mu); \Delta w)$ of the merit function $F_{\ell 2}(w,\mu)$ at a point w in the search direction Δw which is defined by

$$D(F_{\ell 2}(w,\mu); \Delta w) = \nabla F_{PB2}(w,\mu)^T\Delta x + \nu D(F_{PD2}(w,\mu); \Delta w), \tag{3.16}$$

where

$$D(F_{PD2}(w,\mu); \Delta w) \;=\; \nabla_x F_{PD2}(w,\mu)^T\Delta x + \nabla_y F_{PD2}(w,\mu)^T\Delta y$$
$$+\langle \nabla_Z F_{PD2}(w,\mu), \Delta Z\rangle. \tag{3.17}$$

The next two lemmas evaluate the directional derivatives $D(F_{PD2}(w,\mu); \Delta w)$ and $D(F_{\ell 2}(w,\mu); \Delta w)$ in the search direction Δw.

Lemma 3.5. *The Newton direction Δw satisfies*

$$D(F_{PD2}(w,\mu); \Delta w) \leq -\frac{\|Z^{\frac{1}{2}}XZ^{\frac{1}{2}} - \mu I\|_F^2}{\langle X, Z\rangle/p + \|Z^{\frac{1}{2}}XZ^{\frac{1}{2}} - \mu I\|_F^2} - \|g(x) + \mu y\|^2.$$

Proof. From (3.13) – (3.15) and the definition of ΔX, we have

$$\nabla_x F_{PD2}(w,\mu)^T\Delta x \;=\; \frac{\langle \Delta X, Z\rangle/p + 2\{\langle \Delta X, ZXZ\rangle - \mu\langle \Delta X, Z\rangle\}}{\langle X, Z\rangle/p + \|Z^{\frac{1}{2}}XZ^{\frac{1}{2}} - \mu I\|_F^2}$$
$$-\frac{1}{p}\langle \Delta X, X^{-1}\rangle + \{\nabla g(x)^T\Delta x\}^T\{g(x) + \mu y\}, \tag{3.18}$$

$$\nabla_y F_{PD2}(w,\mu)^T\Delta y \;=\; \mu\Delta y^T\{g(x) + \mu y\} \tag{3.19}$$

and

$$\langle \nabla_Z F_{PD2}(w,\mu), \Delta Z\rangle \;=\; \frac{\langle X, \Delta Z\rangle/p + 2\{\langle XZX, \Delta Z\rangle - \mu\langle X, \Delta Z\rangle\}}{\langle X, Z\rangle/p + \|Z^{\frac{1}{2}}XZ^{\frac{1}{2}} - \mu I\|_F^2}$$
$$-\frac{1}{p}\langle Z^{-1}, \Delta Z\rangle. \tag{3.20}$$

11

It follows from (3.4) that

$$
\begin{aligned}
\langle \Delta X, Z \rangle + \langle X, \Delta Z \rangle &= \operatorname{tr}(\Delta X Z) + \operatorname{tr}(X \Delta Z) \\
&= \operatorname{tr}(\Delta \widetilde{X} \widetilde{Z} + \widetilde{X} \Delta \widetilde{Z}) \\
&= \frac{1}{2} \operatorname{tr}(\Delta \widetilde{X} \widetilde{Z} + \widetilde{X} \Delta Z + \Delta \widetilde{Z} \widetilde{X} + \widetilde{Z} \Delta X) \\
&= \frac{1}{2} \operatorname{tr}(2\mu I - \widetilde{X} \widetilde{Z} - \widetilde{Z} \widetilde{X}) \\
&= \operatorname{tr}(\mu I - Z^{\frac{1}{2}} X Z^{\frac{1}{2}}) \tag{3.21}
\end{aligned}
$$

and

$$
\begin{aligned}
\langle \Delta X, ZXZ \rangle + \langle XZX, \Delta Z \rangle &= \operatorname{tr}(\Delta XZXZ) + \operatorname{tr}(XZX\Delta Z) \\
&= \operatorname{tr}\{(\widetilde{X}\widetilde{Z})(\Delta \widetilde{X}\widetilde{Z} + \widetilde{X}\Delta \widetilde{Z})\} \\
&= \frac{1}{2} \operatorname{tr}\{(\widetilde{X}\widetilde{Z})(\Delta \widetilde{X}\widetilde{Z} + \widetilde{X}\Delta \widetilde{Z} + \Delta \widetilde{Z}\widetilde{X} + \widetilde{Z}\Delta \widetilde{X})\} \\
&= \frac{1}{2} \operatorname{tr}\{(\widetilde{X}\widetilde{Z})(2\mu I - \widetilde{X}\widetilde{Z} - \widetilde{Z}\widetilde{X})\} \\
&= \operatorname{tr}\{(Z^{\frac{1}{2}} X Z^{\frac{1}{2}})(\mu I - Z^{\frac{1}{2}} X Z^{\frac{1}{2}})\}.
\end{aligned}
$$

Thus we obtain that

$$
\begin{aligned}
&\langle \Delta X, ZXZ \rangle + \langle XZX, \Delta Z \rangle - \mu\{\langle \Delta X, Z \rangle + \langle X, \Delta Z \rangle\} \\
&= \operatorname{tr}\{(Z^{\frac{1}{2}} X Z^{\frac{1}{2}})(\mu I - Z^{\frac{1}{2}} X Z^{\frac{1}{2}})\} - \operatorname{tr}\{\mu I(\mu I - Z^{\frac{1}{2}} X Z^{\frac{1}{2}})\} \\
&= -\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2. \tag{3.22}
\end{aligned}
$$

By (3.4), (3.11) and (3.21), we get

$$
\langle \Delta X, Z \rangle + \langle X, \Delta Z \rangle = \operatorname{tr}(\mu I - \widetilde{X}\widetilde{Z}) = p\mu - \sum_{i=1}^{p} \lambda_i \tau_i \tag{3.23}
$$

and

$$
\begin{aligned}
\langle \Delta X, X^{-1} \rangle + \langle Z^{-1}, \Delta Z \rangle &= \operatorname{tr}(\Delta X X^{-1}) + \operatorname{tr}(Z^{-1} \Delta Z) \\
&= \operatorname{tr}(X^{-1} \Delta X Z Z^{-1}) + \operatorname{tr}(\Delta Z Z^{-1} X^{-1} X) \\
&= \operatorname{tr}\{(\Delta \widetilde{X}\widetilde{Z} + \widetilde{X}\Delta \widetilde{Z})(\widetilde{X}\widetilde{Z})^{-1}\} \\
&= \frac{1}{2} \operatorname{tr}\{(2\mu I - \widetilde{X}\widetilde{Z} - \widetilde{Z}\widetilde{X})(\widetilde{X}\widetilde{Z})^{-1}\} \\
&= \operatorname{tr}\{(\mu I - \widetilde{X}\widetilde{Z})(\widetilde{X}\widetilde{Z})^{-1}\} \\
&= \operatorname{tr}\{\mu(\widetilde{X}\widetilde{Z})^{-1} - I\} \\
&= \mu \sum_{i=1}^{p} \frac{1}{\lambda_i \tau_i} - p. \tag{3.24}
\end{aligned}
$$

We have from (3.6) that

$$
\{\nabla g(x)^T \Delta x + \mu \Delta y\}^T \{g(x) + \mu y\} = -\|g(x) + \mu y\|^2. \tag{3.25}
$$

12

For simplicity, we define the function $h(x, Z, \mu)$ by

$$h(x, Z, \mu) = \frac{\langle X, Z \rangle}{p} + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 = \sum_{i=1}^{p} \frac{\lambda_i \tau_i}{p} + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2.$$

It follows from (3.17) – (3.20) and (3.22) – (3.25) that

$$
\begin{aligned}
& D(F_{PD2}(w, \mu); \Delta w) \\
= {} & \frac{1}{h(x, Z, \mu)} \left[\frac{1}{p} \{ \langle \Delta X, Z \rangle + \langle X, \Delta Z \rangle \} \right. \\
& \left. + 2\{ \langle \Delta X, ZXZ \rangle + \langle XZX, \Delta Z \rangle - \mu(\langle \Delta X, Z \rangle + \langle X, \Delta Z \rangle) \} \right] \\
& - \frac{1}{p} \{ \langle \Delta X, X^{-1} \rangle + \langle Z^{-1}, \Delta Z \rangle \} + \{ \nabla g(x)^T \Delta x + \mu \Delta y \}^T \{ g(x) + \mu y \} \\
= {} & \frac{1}{h(x, Z, \mu)} \left\{ \mu - \sum_{i=1}^{p} \frac{\lambda_i \tau_i}{p} - 2\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2 \right\} \\
& - \frac{\mu}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i \tau_i} + 1 - \|g(x) + \mu y\|^2 \\
= {} & \frac{\mu - \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{h(x, Z, \mu)} - \frac{\mu}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i \tau_i} - \|g(x) + \mu y\|^2 \\
\leq {} & \frac{p\mu}{\sum_{i=1}^{p} \lambda_i \tau_i} - \frac{\mu}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i \tau_i} - \frac{\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{h(x, Z, \mu)} - \|g(x) + \mu y\|^2.
\end{aligned}
$$

From arithmetic and geometric means, we obtain

$$
\begin{aligned}
& \frac{p\mu}{\sum_{i=1}^{p} \lambda_i \tau_i} - \frac{\mu}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i \tau_i} - \frac{\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{h(x, Z, \mu)} - \|g(x) + \mu y\|^2 \\
\leq {} & \frac{p\mu}{\sum_{i=1}^{p} \lambda_i \tau_i} - \frac{\mu}{(\prod_{i=1}^{p} \lambda_i \tau_i)^{1/p}} - \frac{\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{h(x, Z, \mu)} - \|g(x) + \mu y\|^2 \\
\leq {} & - \frac{\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{h(x, Z, \mu)} - \|g(x) + \mu y\|^2.
\end{aligned}
$$

Therefore, the proof is complete. $\qquad \square$

Lemma 3.6. *The Newton direction Δw satisfies*

$$
\begin{aligned}
D(F_{\ell 2}(w, \mu); \Delta w) \leq {} & -\Delta x^T \{ G + H + \frac{1}{\mu} \nabla g(x) \nabla g(x)^T \} \Delta x \\
& - \nu \frac{\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{\langle X, Z \rangle / p + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2} - \nu \|g(x) + \mu y\|^2.
\end{aligned}
$$

Proof. From Lemma 3.5, (3.5) and (3.16), we obtain

$$
\begin{aligned}
D(F_{\ell 2}(w, \mu); \Delta w) \leq {} & -\Delta x^T \{ G + H + \frac{1}{\mu} \nabla g(x) \nabla g(x)^T \} \Delta x \\
& - \nu \frac{\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{\langle X, Z \rangle / p + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2} - \nu \|g(x) + \mu y\|^2.
\end{aligned}
$$

13

Therefore, the proof is complete. □

The following lemma claims that an SBKKT point is obtained if $\Delta x = 0$, $g(x) + \mu y = 0$ and $XZ - \mu I = 0$ hold.

Lemma 3.7. *If the matrix $G + H + \frac{1}{\mu}\nabla g(x)\nabla g(x)^T$ is positive definite, then the following hold:*

(i) *The Newton direction Δw is a descent direction for the merit function $F_{\ell 2}(w, \mu)$.*

(ii) *If $D(F_{\ell 2}(w, \mu); \Delta w) = 0$ holds, then the point w is an SBKKT point.*

Proof. (i) It is obvious from Lemma 3.6.
(ii) By Lemma 3.5 and the assumption, we obtain that

$$
\begin{aligned}
0 \;\leq\; & -\Delta x^T \{ G + H + \frac{1}{\mu}\nabla g(x)^T \nabla g(x) \}\Delta x \\
& -\nu \frac{\| Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{\langle X, Z\rangle/p + \| Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2} - \nu\|g(x) + \mu y\|^2 \leq 0,
\end{aligned}
$$

which implies that $\Delta x = 0$, $g(x) + \mu y = 0$ and $XZ - \mu I = 0$ hold. $\Delta x = 0$ means $\nabla F_{BP2}(x, \mu) = 0$ from (3.5). Thus from Lemma 3.4, $r(w, \mu) = 0$ follows. □

3.3 Algorithm SDPLS

To globalize our interior point method and guarantee that the new point w_{k+1} is an interior point, we use the line search strategy to obtain a suitable step size α_k. At first, we introduce the algorithm which finds a step size α_k by using the backtracking at the k-th iteration.

Algorithm LS

Step 0. (Initialization) Choose the parameters $\nu > 0$, $\gamma \in (0, 1)$, $\beta \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$. Calculate an initial step size $\bar{\alpha}_k$ by

$$
\bar{\alpha}_k = \min\{\bar{\alpha}_{xk}, \bar{\alpha}_{zk}, \; 1\},
$$

where $\bar{\alpha}_{xk}$ and $\bar{\alpha}_{zk}$ are calculated by

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

and

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

Here, $\lambda_{\min}(M)$ means the minimum eigenvalue of the matrix M. Set $\ell_k = 0$.

Step 1. (Backtracking) If the integer ℓ_k satisfies the conditions

$$F_{\ell 2}(w_k + \bar{\alpha}_k \beta^{\ell k} \Delta w_k, \mu) \leq F_{\ell 2}(w_k, \mu) + \varepsilon_0 \bar{\alpha}_k \beta^{\ell k} D(F_{\ell 2}(w_k, \mu); \Delta w_k) \qquad (3.26)$$

and

$$X(x_k + \bar{\alpha}_k \beta^{\ell k} \Delta x_k) \succ 0, \qquad (3.27)$$

then we set $\alpha_k = \bar{\alpha}_k \beta^{\ell k}$ and stop.

Step 2. (Update) Set $\ell_k := \ell_k + 1$ and return to Step 1. $\qquad \square$

Since the functions f, g and X are sufficiently smooth, there exists a step size $\alpha_k \neq 0$ which satisfies (3.26) and (3.27) at each k. Thus, Algorithm LS is terminated in a finite number of iteration counts.

Now we propose Algorithm SDPLS that finds an approximate SBKKT point.

Algorithm SDPLS

Step 0. (Initialization) Give an initial interior point w_0, the fixed barrier parameter $\mu > 0$ and a parameter $M_c > 0$. Set $k = 0$.

Step 1. (Termination) If $\|r(w_k, \mu)\|_* \leq M_c \mu$, then stop.

Step 2. (Newton direction) Calculate the matrix G_k and the scaling matrix T_k. Determine a search direction Δw_k by (3.5) – (3.7).

Step 3. (Step size) Find a step size α_k by Algorithm LS.

Step 4. (Update) Set

$$w_{k+1} = w_k + \alpha_k \Delta w_k$$

and $k := k + 1$. Return to Step 1. $\qquad \square$

It is notable that we can use a common step size α_k for all variables (x, y, Z) in Step 4 differently from the interior point method in [21].

4 Global convergence to an SBKKT point

This section shows the global convergence property of Algorithm SDPLS.
Since $D(F_{\ell 2}(w_k, \mu); \Delta w_k) = 0$ means that w_k is an SBKKT point from Lemma 3.7 (ⅱ), we assume Algorithm SDPLS generates an infinite sequence $\{w_k\}$ i.e., we assume $D(F_{\ell 2}(w_k, \mu); \Delta w_k) \neq 0$ for any $k \geq 0$. To prove global convergence, we make the following assumptions.

Assumptions

(A1) The functions f, $g_i, i = 1, ..., m$, 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) The matrix $G_k + H_k + \frac{1}{\mu}\nabla g(x_k)\nabla g(x_k)^T$ is uniformly positive definite and the matrix $G_k + H_k$ is uniformly bounded.

(A4) 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.

Assumption (A3) means that there exists a positive constant ϑ such that

$$\vartheta\|v\|^2 \leq v^T\left\{G_k + H_k + \frac{1}{\mu}\nabla g(x_k)\nabla g(x_k)^T\right\}v$$

holds for all $k \geq 0$ and any $v \in \mathbf{R}^n$. The interior point method in [21] assumed that the matrix $G_k + H_k$ was uniformly positive definite and bounded and that for all x_k in Ω, the matrix $\nabla g(x_k)$ was of full rank and the matrices $A_1(x_k), \ldots, A_n(x_k)$ were linearly independent. However assumption (A3) is much weaker than that. Moreover, the interior point method in [21] supposed that the penalty parameter ρ was sufficiently large so that $\rho > \|y_k + \Delta y_k\|_\infty$ held for all k, while our method dose not need such an assumption.

The following lemma shows the boundedness of the sequence $\{\Delta w_k\}$.

Lemma 4.1. *Suppose that assumptions* (A1) *–* (A3) *hold. Let the sequence* $\{w_k\}$ *be generated by Algorithm SDPLS. Then the followings 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.*

Moreover, if assumption (A4) *is satisfied, the following holds.*

(iii) *The sequence* $\{\Delta w_k\}$ *is bounded.*

Proof. (i) The logarithmic function term in (3.8) guarantees that all eigenvalues $(\lambda_k)_i$ and $(\tau_k)_i$ $(i = 1, \cdots, p)$ are bounded away from zero at each k.

(ii) Lemma 3.7 (i) and the condition (3.26) imply that the sequence $\{F_{\ell 2}(w_k, \mu)\}$ is monotonically decreasing, i.e., $F_{\ell 2}(w_0, \mu) \geq F_{\ell 2}(w_k, \mu)$. Since assumptions (A1) and (A2) hold, there exists a constant c such that $F_{BP2}(x_k, \mu) \geq c$ for all $k \geq 0$. Then, we obtain

$$F_{\ell 2}(w_0, \mu) \geq F_{\ell 2}(w_k, \mu) \geq F_{BP2}(x_k, \mu) + \frac{\nu}{2}\|g(x_k) + \mu y_k\|^2 \geq F_{BP2}(x_k, \mu) \geq c,$$

which implies that the sequences $\{y_k\}$ and $\{F_{\ell 2}(w_k, \mu)\}$ are bounded.

Next we prove the boundedness of $\{Z_k\}$ by contradiction. We suppose that there exists the sequence $\{(\tau_j)_k\}$ of the maximum eigenvalue of Z_k such that

$$\lim_{k\to\infty}(\tau_j)_k = \infty.$$

It follows from (3.9), (3.11) and arithmetic and geometric means that

$$
\begin{aligned}
F_{PD2}(w_k, \mu) &\geq \log \frac{\langle X_k, Z_k \rangle / p + \| Z_k^{\frac{1}{2}} X_k Z_k^{\frac{1}{2}} - \mu I \|_F^2}{\{\det(X_k Z_k)\}^{1/p}} \\
&\geq \log \frac{\langle X_k, Z_k \rangle / p + \| Z_k^{\frac{1}{2}} X_k Z_k^{\frac{1}{2}} - \mu I \|_F^2}{\langle X_k, Z_k \rangle / p} \\
&= \log \frac{\displaystyle\sum_{i=1}^{p} (\lambda_i)_k^2 (\tau_i)_k^2 + \frac{1 - 2p\mu}{p} \sum_{i=1}^{p} (\lambda_i)_k (\tau_i)_k + p\mu^2}{\frac{1}{p} \displaystyle\sum_{i=1}^{p} (\lambda_i)_k (\tau_i)_k} \\
&= \log(\bar{F}_1(w_k, \mu) + \bar{F}_2(w_k, \mu)), \tag{4.1}
\end{aligned}
$$

where $\bar{F}_1(w_k, \mu)$ and $\bar{F}_2(w_k, \mu)$ are defined by

$$
\bar{F}_1(w_k, \mu) = \frac{(\lambda_j)_k^2 (\tau_j)_k^2 + \frac{1 - 2p\mu}{p} (\lambda_j)_k (\tau_j)_k}{\frac{1}{p}(\lambda_j)_k(\tau_j)_k + \frac{1}{p}\displaystyle\sum_{i \neq j} (\lambda_i)_k (\tau_i)_k}
$$

and

$$
\bar{F}_2(w_k, \mu) = \frac{\displaystyle\sum_{i \neq j} (\lambda_i)_k^2 (\tau_i)_k^2 + \frac{1 - 2p\mu}{p} \sum_{i \neq j} (\lambda_i)_k (\tau_i)_k + p\mu^2}{\frac{1}{p}(\lambda_j)_k(\tau_j)_k + \frac{1}{p}\displaystyle\sum_{i \neq j} (\lambda_i)_k (\tau_i)_k}.
$$

Dividing the functions $\bar{F}_1(w_k, \mu)$ and $\bar{F}_2(w_k, \mu)$ by $(\tau_j)_k$ and taking limit as $k \to \infty$, we have from the assumption (A2)

$$
\frac{(\lambda_j)_k^2 (\tau_j)_k + \frac{1 - 2p\mu}{p} (\lambda_j)_k}{\frac{1}{p}(\lambda_j)_k + \frac{1}{p}\displaystyle\sum_{i \neq j} \frac{(\lambda_i)_k (\tau_i)_k}{(\tau_j)_k}} \to \infty
$$

and

$$
\frac{\displaystyle\sum_{i \neq j} \frac{(\lambda_i)_k^2 (\tau_i)_k^2}{(\tau_j)_k} + \frac{1 - 2p\mu}{p} \sum_{i \neq j} \frac{(\lambda_i)_k (\tau_i)_k}{(\tau_j)_k} + \frac{p\mu^2}{(\tau_j)_k}}{\frac{1}{p}(\lambda_j)_k + \frac{1}{p}\displaystyle\sum_{i \neq j} \frac{(\lambda_i)_k (\tau_i)_k}{(\tau_j)_k}} \to +0,
$$

which imply by (4.1) that $F_{\ell 2}(w_k, \mu) \to \infty$. This contradicts the boundedness of $\{F_{\ell 2}(w_k, \mu)\}$. Therefore, $\{Z_k\}$ is bounded.

(iii) By assumptions (A2) and (A3), the matrix $G_k + H_k + \frac{1}{\mu}\nabla g(x_k)\nabla g(x_k)^T$ is uniformly bounded. Thus we conclude that Δw_k is uniformly bounded by the result (ii), (3.5), (3.6) and (3.7). $\qquad \square$

Now we prove the global convergence property of an infinite sequence $\{w_k\}$ generated by Algorithm SDPLS.

Theorem 4.1. *Suppose assumptions* (A1) – (A4) *hold. Let* $\{w_k\}$ *be the sequence 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 SBKKT point.*

Proof. By Lemma 4.1 (ⅱ), the sequence $\{w_k\}$ has at least one accumulation point. From Lemma 3.6 and assumption (A3), we have

$$D(F_{\ell 2}(w_k, \mu); \Delta w_k) \leq -\vartheta \|\Delta x_k\|^2 + \nu D(F_{PD2}(w_k, \mu); \Delta w_k) < 0, \tag{4.2}$$

and from Lemma 3.4 and (3.26),

$$
\begin{aligned}
F_{\ell 2}(w_{k+1}, \mu) - F_{\ell 2}(w_k, \mu) &\leq \varepsilon_0 \bar{\alpha}_k \beta^{\ell_k} D(F_{\ell 2}(w_k, \mu); \Delta w_k) \\
&\leq -\varepsilon_0 \bar{\alpha}_k \beta^{\ell_k} \{\vartheta \|\Delta x_k\|^2 - \nu D(F_{PD2}(w_k, \mu); \Delta w_k)\} \\
&\leq \varepsilon_0 \bar{\alpha}_k \beta^{\ell_k} \nu D(F_{PD2}(w_k, \mu); \Delta w_k) \\
&< 0.
\end{aligned} \tag{4.3}
$$

Since the sequence $\{F_{\ell 2}(w_k, \mu_k)\}$ is monotonically decreasing and bounded below from the proof of Lemma 4.1 (ⅱ), the left-hand side of (4.3) converges to 0. From Lemma 4.1 (ⅰ), we have $\liminf_{k\to\infty} \bar{\alpha}_k > 0$, which yields by (4.3)

$$\lim_{k\to\infty} \beta^{\ell_k} D(F_{\ell 2}(w_k, \mu); \Delta w_k) = 0.$$

We will show

$$\lim_{k\to\infty} D(F_{\ell 2}(w_k, \mu); \Delta w_k) = 0. \tag{4.4}$$

For this purpose, we suppose that there exists an infinite subsequence $K \subset \{0, 1, \cdots\}$ and a positive constant δ such that

$$|D(F_{\ell 2}(w_k, \mu); \Delta w_k)| \geq \delta > 0 \quad \forall k \in K, \tag{4.5}$$

which implies

$$\lim_{k\to\infty, \ k \in K} \beta^{\ell_k} = 0.$$

Thus we can assume that $\ell_k > 0$ for sufficiently large $k \in K$ without loss of generality. Since the point $w_k + \alpha_k \Delta w_k / \beta$ dose not satisfy condition (3.26), we have

$$F_{\ell 2}(w_k + \alpha_k \Delta w_k / \beta, \mu) - F_{\ell 2}(w_k, \mu) > \varepsilon_0 \alpha_k D(F_{\ell 2}(w_k, \mu); \Delta w_k) / \beta. \tag{4.6}$$

Thus by assumption (A1) and the mean value theorem, there exists a $\theta_k \in (0, 1)$ such that

$$
\begin{aligned}
F_{\ell 2}(w_k + \alpha_k \Delta w_k / \beta, \mu) - F_{\ell 2}(w_k, \mu) \\
= \alpha_k D(F_{\ell 2}(w_k + \theta_k \alpha_k \Delta w_k / \beta, \mu); \Delta w_k) / \beta.
\end{aligned} \tag{4.7}
$$

Combining (4.6) and (4.7), we have

$$D(F_{\ell 2}(w_k + \theta_k \alpha_k \Delta w_k / \beta, \mu); \Delta w_k) > \varepsilon_0 D(F_{\ell 2}(w_k, \mu); \Delta w_k).$$

18

This inequality yields

$$D(F_{\ell 2}(w_k + \theta_k \alpha_k \Delta w_k / \beta, \mu); \Delta w_k) - D(F_{\ell 2}(w_k, \mu); \Delta w_k)$$
$$> -(1 - \varepsilon_0) D(F_{\ell 2}(w_k, \mu); \Delta w_k) > 0. \qquad (4.8)$$

Thus by the property $\ell_k \to \infty$ ($k \in K$) and Lemma 4.1 (ⅲ), we have $\alpha_k \to 0$ and $\|\theta_k \alpha_k \Delta w_k / \beta\| \to 0$, $k \in K$. Thus the left-hand side of (4.8) and therefore $D(F_{\ell 2}(w_k, \mu); \Delta w_k)$ converges to 0 when $k \to \infty$, $k \in K$. This contradicts assumption (4.5). Therefore, we get (4.4).

It follows from Lemma 3.5 and (4.2) that

$$D(F_{\ell 2}(w_k, \mu); \Delta w_k)$$
$$\leq -\vartheta_1 \|\Delta x_k\|^2 - \nu \frac{\|Z_k^{\frac{1}{2}} X_k Z_k^{\frac{1}{2}} - \mu I\|_F^2}{\langle X_k, Z_k \rangle / p + \|Z_k^{\frac{1}{2}} X_k Z_k^{\frac{1}{2}} - \mu I\|_F^2} - \nu \|g(x_k) + \mu y_k\|^2 < 0.$$

Since the boundedness of the sequence $\{w_k\}$ guarantees that the denominator of the above equation dose not approach infinity, (4.4) implies that

$$\Delta x_k \to 0, \quad g(x_k) + \mu y_k \to 0, \quad X_k Z_k - \mu I \to 0,$$

which implies by (3.5), assumption (A3) and Lemma 3.4 that $\nabla F_{BP2}(x_k, \mu) \to 0$ and any accumulation point of the sequence $\{w_k\}$ is an SBKKT point. $\qquad \square$

5 Concluding remarks

In this paper, we have proposed a new merit function of a primal-dual interior point method for nonlinear SDP problems and have proved the global convergence of our method. Yamashita et al [21] considered a nondifferentiable merit function in (x, Z)-space, while we have dealt with a differentiable primal-dual merit function in the whole space by using the shifted barrier KKT conditions. The global convergence of the present method have been shown under the weaker assumptions than those in [21].

References

[1] Bernstein, D.S.: Matrix Mathematics: Theory, Facts, and Formulas (Second Edition). Princeton University Press, Princeton (2009)

[2] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)

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

[4] Drummond, L.M.G., Iusem, A.N., Svaiter, B.F.: On the central path for nonlinear semidefinite programming. RAIRO-Recherche Operationnelle-Operations Research. **34**, 331–345 (2000)

[5] Fares, B., Apkarian, P., Noll, D.: An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. International Journal of Control **74**, 348-360 (2001)

[6] Freund, R. W., Jarre, F. and Vogelbusch, C. H.: Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. Mathematical Programming **109**, 581-611 (2007)

[7] Gomez, W., Ramirez, H.: A filter algorithm for nonlinear semidefinite programming. Computational and Applied Mathematics. **29**, 297–328 (2010)

[8] Helmberg, C., Rendl, F., Vanderbei, R.J., Wolkowcz, H.,: An interior point method for semidefinite programming. SIAM Journal on Optimization. **6**, 342-361 (1996)

[9] Huang, X.X., Yang, X.Q., Teo, K.L.: A sequential quadratic penalty method for nonlinear semidefinite programming. Optimization. **52**, 715–738 (2003)

[10] Huang, X.X., Teo, K.L., Yang, X.Q.: Approximate augmented Lagrangian functions and nonlinear semidefinite programs. ACTA Mathematica Sinica-English Series. **22**, 1283–1296 (2006)

[11] Huang, X.X., Yang, X.Q., Teo, K.L.: Lower-order penalization approach to nonlinear semidefinite programming. Journal of Optimization Theory and Applications. **132**, 1–20 (2007)

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

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

[14] Kojima, M., Shindoh, S., Hara, S.: Interior point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. SIAM Journal on Optimization. **7**, 86–125 (1997)

[15] Leibfritz, F., Mostafa, E.M.E.: An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems. SIAM Journal on Optimization. **12**, 1048–1074 (2002)

[16] Monteiro, R.D.C.: Primal-dual path following algorithms for semidefinite programming. SIAM Journal on Optimization. **8**, 769–796 (1998)

[17] Stingl, M.: On the Solution of Nonlinear Semidefinite Programs by Augmented Lagrangian Methods, Doctoral Thesis, University of Erlangen, 2005.

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

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

[20] Yamashita, H., Yabe, H.: An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization. SIAM Journal on Optimization. **14**, 479–499 (2003)

[21] Yamashita, H., Yabe, H., Harada, K.: A primal-dual interior point method for nonlinear semidefinite programming. to appear in Mathematical Optimization.