

A TWO-STEP PRIMAL-DUAL INTERIOR POINT METHOD FOR NONLINEAR SEMIDEFINITE PROGRAMMING PROBLEMS AND ITS SUPERLINEAR CONVERGENCE

Yuya Yamakawa Nobuo Yamashita
Kyoto University

(Received March 27, 2014; Revised September 6, 2014)

Abstract In this paper, we propose a primal-dual interior point method for nonlinear semidefinite programming problems and show its superlinear convergence. This method is based on generalized shifted barrier Karush-Kuhn-Tucker (KKT) conditions, which include barrier KKT conditions and shifted barrier KKT conditions as a special case. This method solves two Newton equations in a single iteration to guarantee superlinear convergence. We replace the coefficient matrix of the second Newton equation with that of the first to reduce the computational time of the single iteration. We show that the superlinear convergence of the proposed method with the replacement under the usual assumptions.

Keywords: Optimization, barrier KKT conditions, shifted barrier KKT conditions, generalized shifted barrier KKT conditions, Newton method, nonlinear semidefinite programming problem, primal-dual interior point method, superlinear convergence

1. Introduction

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

$$\begin{aligned} & \underset{x \in \mathbf{R}^n}{\text{minimize}} && f(x), \\ & \text{subject to} && g(x) = 0, \quad X(x) \succeq 0, \end{aligned} \tag{1.1}$$

where $f : \mathbf{R}^n \rightarrow \mathbf{R}$, $g : \mathbf{R}^n \rightarrow \mathbf{R}^m$, $X : \mathbf{R}^n \rightarrow \mathbf{S}^p$ are twice continuously differentiable functions, and \mathbf{S}^p denotes a set of $p \times p$ real symmetric matrices. Moreover, $X(x) \succeq 0$ means that $X(x)$ is positive semidefinite. Nonlinear SDP includes a wide class of mathematical programming problems, such as linear programming, second-order cone programming, linear semidefinite programming and nonlinear programming.

Various methods have been proposed for nonlinear SDP. This paper focuses on the interior point method. Yamashita, Yabe and Harada [10] proposed a primal-dual interior point method based on the barrier Karush-Kuhn-Tucker (KKT) conditions and the L_1 penalty function. They showed its global convergence. Moreover, Yamashita and Yabe [9] also proposed a two-step primal-dual interior point method based on the barrier KKT conditions and showed its superlinear convergence. Note that "two-step" indicates that two Newton equations are solved in a single iteration. On the other hand, Yamakawa and Yamashita [8] proposed a primal-dual interior point method based on the shifted barrier KKT conditions. They guaranteed its global convergence using a differentiable merit function for the shifted barrier KKT conditions. However, they have not shown the superlinear convergence of their method.

In this paper, we propose a new two-step primal-dual interior point method and show its superlinear convergence. First, we define the generalized shifted barrier KKT conditions,

which include the barrier KKT conditions and the shifted barrier KKT conditions as a special case. The proposed method is based on the generalized shifted barrier KKT conditions. It solves two Newton equations derived from the generalized shifted barrier KKT conditions in each iteration. However, in order to reduce calculations, we replace the coefficient matrix in the second equation with that in the first. Thus, we can solve the second equation more rapidly using some computational results in solving the first equation. Despite this change, we show the superlinear convergence under the same assumptions of the two-step primal-dual interior point method [9].

The present paper is organized as follows. In Section 2, we introduce some operators and their properties. We also present some optimality conditions for (1.1) and a general framework of primal-dual interior point methods. In Section 3, we propose a two-step primal-dual interior point method that uses the same coefficient matrix. In Section 4, we prove the superlinear convergence of the proposed method. Finally, we provide some concluding remarks.

Next, we define some mathematical notations throughout this paper. For two matrices $A, B \in \mathbf{R}^{n_1 \times n_2}$, we define an inner product as $\langle A, B \rangle := \text{tr}(AB^\top)$, where $\text{tr}(M)$ represents the trace of a square matrix $M \in \mathbf{R}^{n_1 \times n_1}$, and the superscript \top denotes the transposition of a vector or a matrix. In addition, for a vector w , w_i denotes the i -th element of the vector w , and $\|w\| := \sqrt{\langle w, w \rangle}$. Similarly, for a matrix W , W_{ij} denotes the (i, j) -th element of the matrix W , and $\|W\|_F$ and $\|W\|_2$ denote the Frobenius norm and the operator norm, respectively. Note that $\|W\|_F := \sqrt{\langle W, W \rangle}$ and $\|W\|_2 := \sup_{x \neq 0} \frac{\|Wx\|}{\|x\|}$. Let \mathbf{S}_{++}^p (\mathbf{S}_+^p) denote a set of $p \times p$ real symmetric positive (semi)definite matrices. For a matrix $M \in \mathbf{S}^p$, $M \succeq 0$ and $M \succ 0$ imply that $M \in \mathbf{S}_+^p$ and $M \in \mathbf{S}_{++}^p$, respectively. Note that for two matrices $A, B \in \mathbf{S}^p$, $A \succeq B$ and $A \succ B$ indicate that $A - B \succeq 0$ and $A - B \succ 0$, respectively. For a matrix $U \in \mathbf{S}^p$, $\lambda_1(U), \dots, \lambda_p(U)$ denote the eigenvalues of U . In particular, $\lambda_{\min}(U)$ and $\lambda_{\max}(U)$ denote the minimum and maximum eigenvalues of U . For a matrix $V \in \mathbf{S}_+^p$, $V^{\frac{1}{2}}$ denotes the symmetric positive semidefinite matrix such that $V = V^{\frac{1}{2}}V^{\frac{1}{2}}$. We denote the $m \times m$ unit matrix as I_m , and the j -th column vector of I_m as e_j . Let $\Psi : P_1 \times P_2 \rightarrow \mathbf{R}$, where P_1 and P_2 are open sets. We denote a Fréchet derivative of Ψ as $\nabla \Psi$. We also denote a Fréchet derivative of Ψ with respect to a variable $Z \in P_1$ as $\nabla_Z \Psi$. For a given vector $x \in \mathbf{R}^n$, $y \in \mathbf{R}^m$, $z \in \mathbf{R}^{\frac{p(p+1)}{2}}$, we denote the vector $w = [x^\top y^\top z^\top]^\top$ as $w = [x, y, z]$ for simplicity.

2. Preliminaries

In this section, we first introduce some operators. We also present the KKT conditions for nonlinear SDP (1.1), and propose the generalized shifted barrier KKT conditions related to a primal-dual interior point method.

2.1. Some operators

In this subsection, let $P, Q \in \mathbf{R}^{p \times p}$. We define the following notations.

- (i) We define a partial derivative of $X : \mathbf{R}^n \rightarrow \mathbf{S}^p$ with respect to x_i as $A_i(x) := \frac{\partial}{\partial x_i} X(x)$.
- (ii) We define an operator $\mathcal{A}(x) : \mathbf{R}^n \rightarrow \mathbf{S}^p$ as $\mathcal{A}(x)w := w_1 A_1(x) + \dots + w_n A_n(x)$ for all $w \in \mathbf{R}^n$.
- (iii) We define the adjoint operator of $\mathcal{A}(x) : \mathbf{R}^n \rightarrow \mathbf{S}^p$ as $\mathcal{A}^*(x) : \mathbf{S}^p \rightarrow \mathbf{R}^n$, that is, $\mathcal{A}^*(x)U := [\langle A_1(x), U \rangle, \dots, \langle A_n(x), U \rangle]^\top$ for all $U \in \mathbf{S}^p$.

- (iv) We define an operator $P \odot Q : \mathbf{S}^p \rightarrow \mathbf{S}^p$ as $(P \odot Q)U := \frac{1}{2}(PUQ^\top + QUP^\top)$ for all $U \in \mathbf{S}^p$.
- (v) We define an operator $\text{svec} : \mathbf{S}^p \rightarrow \mathbf{R}^{\frac{p(p+1)}{2}}$ as
- $$\text{svec}(U) := [U_{11}, \sqrt{2}U_{21}, \dots, \sqrt{2}U_{p1}, U_{22}, \sqrt{2}U_{32}, \dots, \sqrt{2}U_{p2}, U_{33}, \dots, U_{pp}]^\top \text{ for all } U \in \mathbf{S}^p.$$
- (vi) We denote the symmetrized Kronecker product as $P \otimes_S Q : \mathbf{R}^{\frac{p(p+1)}{2}} \rightarrow \mathbf{R}^{\frac{p(p+1)}{2}}$ which satisfies that $(P \otimes_S Q)\text{svec}(U) := \text{svec}((P \odot Q)U)$ for all $U \in \mathbf{S}^p$.
- (vii) We define $A(x) \in \mathbf{R}^{\frac{p(p+1)}{2} \times n}$ as $A(x) := [\text{svec}(A_1(x)), \dots, \text{svec}(A_n(x))]$.
- (viii) We define $U \circ V := \frac{UV+VU}{2}$ for all $U, V \in \mathbf{S}^p$.

Note that $\langle U, V \rangle = \text{tr}(UV) = \text{svec}(U)^\top \text{svec}(V)$ and $\|U\|_F = \|\text{svec}(U)\|$. For further details, see [7, 10].

2.2. Optimality conditions for nonlinear SDP

We introduce the first-order optimality conditions (KKT conditions) for nonlinear SDP (1.1). We define the Lagrangian function $L : \mathbf{R}^l \rightarrow \mathbf{R}$ as

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

where $l := n + m + \frac{p(p+1)}{2}$ and $w := [x, y, \text{svec}(Z)] \in \mathbf{R}^l$. Note that $y \in \mathbf{R}^m$ and $Z \in \mathbf{S}^p$ are Lagrange multipliers for $g(x) = 0$ and $X(x) \succeq 0$, respectively. Then, the gradient of L with respect to x is written as

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

where $J_g(x)$ denotes the Jacobian matrix of g at x . The KKT conditions of (1.1) are given by

$$\nabla_x L(w) = 0, \quad g(x) = 0, \quad \text{svec}(X(x) \circ Z) = 0, \quad X(x) \succeq 0, \quad Z \succeq 0. \quad (2.1)$$

For further details, see [10].

In this paper, we consider the following generalized shifted barrier KKT conditions with parameters $\mu \geq 0$ and $\kappa \in [0, \infty)$:

$$r_\kappa(w, \mu) := \begin{bmatrix} \nabla_x L(w) \\ g(x) + \kappa \mu y \\ \text{svec}(X(x) \circ Z - \mu I) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad X(x) \succeq 0, \quad Z \succeq 0. \quad (2.2)$$

In what follows, we call (2.2) the generalized shifted barrier KKT conditions. If $\mu = 0$, the generalized shifted barrier KKT conditions (2.2) are reduced to the KKT conditions (2.1). Note that when $\mu > 0$, the conditions $X(x) \succeq 0$ and $Z \succeq 0$ in (2.2) are equivalent to $X(x) \succ 0$ and $Z \succ 0$. Moreover, if $\kappa = 0$ or $\kappa = 1$, then (2.2) are reduced to the barrier KKT conditions [9, 10] or the shifted barrier KKT conditions [2, 8], respectively. Until now, several primal-dual interior point methods based on the condition (2.2) have been proposed [2, 8–10].

We define a set \mathcal{W} as

$$\mathcal{W} := \{ w = [x, y, \text{svec}(Z)] \in \mathbf{R}^l \mid X(x) \succ 0, \quad Z \succ 0 \}.$$

We call a point $w \in \mathcal{W}$ an interior point. For a given $\xi \geq 0$, a point w such that $\|r_\kappa(w, 0)\| \leq \xi$, $X(x) \succeq 0$ and $Z \succeq 0$ is called an approximate KKT point. Similarly, if w satisfies that $\|r_\kappa(w, \mu)\| \leq \xi$ with $\mu > 0$, $X(x) \succ 0$ and $Z \succ 0$, we call w an approximate generalized shifted barrier KKT point.

2.3. Primal-dual interior point method

We present a primal-dual interior point method based on the generalized shifted barrier KKT conditions.

Algorithm 2.1.

Step 0. (Initialize) Choose parameters $\kappa \geq 0$ and $\epsilon \in (0, 1)$, and give a sequence $\{\mu_k\}$ such that $\lim_{k \rightarrow \infty} \mu_k = 0$ and $\mu_k > 0$. Set $k := 0$.

Step 1. (Termination) If $\|r_\kappa(w_k, 0)\| \leq \epsilon$, then stop.

Step 2. (Newton step) Find an approximate generalized shifted barrier KKT point w_{k+1} such that $\|r_\kappa(w_{k+1}, \mu_k)\| \leq \mu_k$, $X(x_{k+1}) \succ 0$ and $Z_{k+1} \succ 0$.

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

The global convergence in the case where $\kappa = 0$ or $\kappa = 1$ has already shown in [2, 8, 10]. Since the global convergence for any $\kappa \in [0, \infty)$ can be also shown similarly, we omit its proof.

In this paper, we investigate the rate of local convergence. In Section 3, we propose a two-step primal-dual interior point method that can find w_{k+1} in Step 2. We also show that the proposed method can find w_{k+1} in a single iteration if w_k is sufficiently close to the KKT point.

3. Two-Step Primal-Dual Interior Point Method

In this section, we propose a two-step primal-dual interior point method, which is a special case of Algorithm 2.1 when w_k is close to a solution. To this end, we first develop a Newton equation with a scaling in Subsection 3.1. We then provide an actual algorithm in Subsection 3.2.

3.1. Newton equation with a scaling

We adopt a Newton method to find an approximate generalized shifted barrier KKT point w_{k+1} in Step 2 of Algorithm 2.1. As seen in existing methods [9, 10], we exploit a nonsingular scaling matrix T such that $\tilde{X}(x)\tilde{Z} = \tilde{Z}\tilde{X}(x)$ where $\tilde{X}(x) := TX(x)T^\top = (T \odot T)X(x)$, $\tilde{Z} := T^{-\top}ZT^{-1} = (T^{-\top} \odot T^{-\top})Z$, respectively. In the following, $X(x)$ and $\tilde{X}(x)$ are denoted by X and \tilde{X} , respectively, for simplicity.

We consider the following generalized shifted barrier KKT conditions with the scaling, which are obtained by replacing X and Z with \tilde{X} and \tilde{Z} , respectively, in the generalized shifted barrier KKT conditions (2.2):

$$\tilde{r}_\kappa(w, \mu) := \begin{bmatrix} \nabla_x L(w) \\ g(x) + \kappa\mu y \\ \text{svec}(\tilde{X} \circ \tilde{Z} - \mu I) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \tilde{X} \succ 0, \quad \tilde{Z} \succ 0.$$

It is known that these conditions are equivalent to (2.2). For details, see [8–10]

Next, we apply the Newton method to the nonlinear equations $\tilde{r}_\kappa(w, \mu) = 0$. Then, the Newton equations are given by

$$\nabla_{xx}^2 L(w)\Delta x - J_g(x)^\top \Delta y - A(x)^\top \text{svec}(\Delta Z) = -\nabla_x L(w), \quad (3.1)$$

$$J_g(x)\Delta x + \kappa\mu\Delta y = -g(x) - \kappa\mu y, \quad (3.2)$$

$$(\tilde{Z} \otimes_S I)(T \otimes_S T)A(x)\Delta x + (\tilde{X} \otimes_S I)(T^{-\top} \otimes_S T^{-\top})\text{svec}(\Delta Z) = \text{svec}(\mu I - \tilde{X} \circ \tilde{Z}). \quad (3.3)$$

Yamashita and Yabe [9] proposed the following two-step primal-dual interior point method based on the Newton equations (3.1)–(3.3) in the case where $\kappa = 0$.

Algorithm 3.1. [9, scaled SDPIP]

Step 0. (Initialize) Choose parameters $\epsilon > 0$ and $\tau \in (0, 1)$, and give an initial interior point $w_0 = [x_0, y_0, \text{svec}(Z_0)] \in \mathcal{W}$. Set $k := 0$.

Step 1. (Termination) If $\|r_0(w_k, 0)\| \leq \epsilon$, then stop.

Step 2. (Newton steps)

Step 2.1 Set $\mu_k := \|r_0(w_k, 0)\|^{1+\tau}$.

Step 2.2 Calculate the Newton direction Δw_k by solving the Newton equations (3.1)–(3.3) at w_k , and set $\hat{w}_k := w_k + \Delta w_k$.

Step 2.3 Calculate the Newton direction $\Delta \hat{w}_k$ by solving the Newton equations (3.1)–(3.3) at \hat{w}_k , and set $w_{k+1} := \hat{w}_k + \Delta \hat{w}_k$.

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

Yamashita and Yabe [9] showed the superlinear convergence of Algorithm 3.1 under some appropriate assumptions (see Assumption 2 of Section 4). Note that Step 2 in this method has to solve two linear equations with different coefficient matrices.

Yamashita and Yabe [9] also showed that if T is a special matrix such as $T = X^{-\frac{1}{2}}$ and $T = W^{-\frac{1}{2}}$ ($W = X^{\frac{1}{2}}(X^{\frac{1}{2}}ZX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}}$), the Newton equation (3.3) is written as

$$((Z \otimes_S I)A(x) + P(w))\Delta x + (X \otimes_S I)\text{svec}(\Delta Z) = \text{svec}(\mu I - X \circ Z), \quad (3.4)$$

where the matrix $P(w) \in \mathbf{R}^{\frac{p(p+1)}{2} \times n}$ depends on T . For further details, see [9] or Appendix C. Note that for the general matrix T , there is no matrix $P(w)$ that satisfies (3.4). Thus, we make the following assumption on T in the rest of the paper.

Assumption 1. The scaling matrix T satisfies the following (S1):

(S1) There exists a matrix $P(w) \in \mathbf{R}^{\frac{p(p+1)}{2} \times n}$ such that the equation (3.3) is equivalent to the equation (3.4).

See Appendix C for the scaling matrix T that satisfies assumption (S1).

3.2. Two-step primal-dual interior point method with the same coefficient matrix

We now propose a new algorithm. The proposed algorithm has a similar procedure to Algorithm 3.1, i.e., there exist two Newton steps in a single iteration.

First, we calculate $\hat{w}_k := w_k + \Delta w_k$ by solving the Newton equations (3.1)–(3.3) at w_k as Step 2.2 of Algorithm 3.1. From Assumption 1, the Newton equations are written as

$$\begin{bmatrix} \nabla_{xx}^2 L(w_k) & -J_g(x_k)^\top & -A(x_k)^\top \\ J_g(x_k) & \kappa \mu_k I & 0 \\ (Z_k \otimes_S I)A(x_k) + P(w_k) & 0 & (X_k \otimes_S I) \end{bmatrix} \begin{bmatrix} \Delta x_k \\ \Delta y_k \\ \text{svec}(\Delta Z_k) \end{bmatrix} = \begin{bmatrix} -\nabla_x L(w_k) \\ -g(x_k) - \kappa \mu_k y_k \\ \text{svec}(\mu_k I - X_k \circ Z_k) \end{bmatrix}, \quad (3.5)$$

where we define $X_k := X(x_k)$ for simplicity.

Recall that the next step of Algorithm 3.1, i.e., Step 2.3, solves the Newton equations (3.1)–(3.3) at \hat{w}_k in order to find $\Delta\hat{w}_k$. The coefficient matrix of these equations differs from the coefficient matrix of (3.5). Thus, the computational costs for Step 2.3 are almost the same as those for Step 2.2.

To reduce the computational costs of the second step, we generate a direction $\Delta\hat{w}_k$ by solving the following equation, which has the same coefficient matrix as that in (3.5).

$$\begin{bmatrix} \nabla_{xx}^2 L(w_k) & -J_g(x_k)^\top & -A(x_k)^\top \\ J_g(x_k) & \kappa\mu_k I & 0 \\ (Z_k \otimes_S I)A(x_k) + P(w_k) & 0 & (X_k \otimes_S I) \end{bmatrix} \begin{bmatrix} \Delta\hat{x}_k \\ \Delta\hat{y}_k \\ \text{svec}(\Delta\hat{Z}_k) \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\hat{w}_k) \\ -g(\hat{x}_k) - \kappa\mu_k \hat{y}_k \\ \text{svec}(\mu_k I - \hat{X}_k \circ \hat{Z}_k) \end{bmatrix}. \quad (3.6)$$

Note that \hat{w}_k appears only in the right-hand side of (3.6). Summing up the above ideas, we give a new two-step primal-dual interior point method.

Algorithm 3.2.

Step 0. (Initialize) Choose parameters $\kappa \geq 0$, $\epsilon > 0$ and $\tau \in (0, \frac{1}{2})$, and give an initial interior point $w_0 = [x_0, y_0, \text{svec}(Z_0)] \in \mathcal{W}$. Set $k := 0$.

Step 1. (Termination) If $\|r_\kappa(w_k, 0)\| \leq \epsilon$, then stop.

Step 2. (Newton steps)

Step 2.1 Set $\mu_k := \|r_\kappa(w_k, 0)\|^{1+\tau}$.

Step 2.2 Calculate the Newton direction Δw_k by solving the Newton equation (3.5), and set $\hat{w}_k := w_k + \Delta w_k$.

Step 2.3 Calculate the Newton direction $\Delta\hat{w}_k$ by solving the Newton equation (3.6), and set $w_{k+1} := \hat{w}_k + \Delta\hat{w}_k$.

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

In the following, we discuss the computational costs of Step 2, i.e., the calculations of Δw_k and $\Delta\hat{w}_k$. First, note that the equation (3.5) can be reduced to

$$\begin{bmatrix} \nabla_{xx}^2 L(w_k) + H_k & -J_g(x_k)^\top \\ J_g(x_k) & \kappa\mu_k I \end{bmatrix} \begin{bmatrix} \Delta x_k \\ \Delta y_k \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) + J_g(x_k)^\top y_k + \mu_k \mathcal{A}^*(x_k) X_k^{-1} \\ -g(x_k) - \kappa\mu_k y_k \end{bmatrix}, \quad (3.7)$$

$$\Delta Z_k = \mu_k X_k^{-1} - Z_k - (T_k^\top \circ T_k^\top)(\tilde{X}_k \circ I)^{-1}(\tilde{Z}_k \circ I)(T_k \circ T_k)\mathcal{A}(x_k)\Delta x_k,$$

where the (i, j) -th element of $H_k \in \mathbf{R}^{n \times n}$ is given by

$$(H_k)_{ij} = \langle A_i(x_k), (T_k^\top \circ T_k^\top)(\tilde{X}_k \circ I)^{-1}(\tilde{Z}_k \circ I)(T_k \circ T_k)A_j(x_k) \rangle,$$

and T_k is the scaling matrix at the k -th iteration. Similarly, we can rewrite (3.6) as

$$\begin{bmatrix} \nabla_{xx}^2 L(w_k) + H_k & -J_g(x_k)^\top \\ J_g(x_k) & \kappa\mu_k I \end{bmatrix} \begin{bmatrix} \Delta\hat{x}_k \\ \Delta\hat{y}_k \end{bmatrix} = \begin{bmatrix} \mathcal{A}^*(x_k)(\mu_k X_k^{-1} - (X_k \circ I)^{-1}(\hat{X}_k \circ \hat{Z}_k)) - \nabla_x L(\hat{w}_k) \\ -g(\hat{x}_k) - \kappa\mu_k \hat{y}_k \end{bmatrix}, \quad (3.8)$$

$$\Delta\hat{Z}_k = \mu_k X_k^{-1} - (T_k^\top \circ T_k^\top)(\tilde{X}_k \circ I)^{-1}(\tilde{Z}_k \circ I)(T_k \circ T_k)\mathcal{A}(x_k)\Delta\hat{x}_k - (X_k \circ I)^{-1}(\hat{X}_k \circ \hat{Z}_k).$$

From these equations, we see that the main calculations of Δw_k and $\Delta \hat{w}_k$ are a construction of the matrix H in (3.7) and (3.8). In Algorithm 3.1, it is necessary to calculate the matrix H twice during Steps 2.2 and 2.3. By contrast, in Algorithm 3.2, we use the same matrix H in Steps 2.2 and 2.3. Thus, we can expect that Algorithm 3.2 can find the next point w_{k+1} faster than Algorithm 3.1.

4. Local and Superlinear Convergence of Algorithm 3.2

In this section, we show the local and superlinear convergence of Algorithm 3.2. First, we give some assumptions for the convergence and we define some neighborhoods of the generalized shifted barrier KKT point. Next, under these assumptions, we show that the sequence generated by Algorithm 3.2 is included in the neighborhoods of the generalized shifted barrier KKT point. Finally, we show the superlinear convergence of Algorithm 3.2.

4.1. Assumptions and some resulting properties

In this subsection, we first give assumptions required for the proof of the superlinear convergence. To this end, let $M(w, \mu)$ be the Jacobian matrix of the linear equations (3.1)–(3.3) with $T = I$, i.e.,

$$M(w, \mu) := M_0(w) + \kappa\mu M_I, \tag{4.1}$$

where

$$M_0(w) := \begin{bmatrix} \nabla_{xx}^2 L(w) & -J_g(x)^\top & -A(x)^\top \\ J_g(x) & 0 & 0 \\ (Z \otimes_S I)A(x) & 0 & (X \otimes_S I) \end{bmatrix}, \quad M_I := \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_m & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

We will show the superlinear convergence of Algorithm 3.2 under the following assumptions, which are the same as [9]. Details of the following assumptions are given in Appendix A.

Assumption 2. Let $w^* = [x^*, y^*, \text{svec}(Z^*)]$ be a KKT point of nonlinear SDP (1.1).

- (A1) There exists a positive constant ν_L such that M_0 is Lipschitz continuous on $\mathcal{V}_L := \{w \in \mathbf{R}^l \mid \|w - w^*\| \leq \nu_L\}$.
- (A2) The second-order sufficient condition holds at x^* .
- (A3) The strict complementarity condition holds at x^* .
- (A4) The nondegeneracy condition holds at x^* .

Note that if (A3) holds, then (A4) holds if and only if the Lagrange multipliers $y^* \in \mathbf{R}^m$ and $Z^* \in \mathbf{S}^p$ corresponding to $x^* \in \mathbf{R}^n$ are unique. For details, see Yamashita and Yabe [9].

Assumption 2 (A1) implies that there exists a positive constant L_M such that

$$\|M_0(w_1) - M_0(w_2)\|_F \leq L_M \|w_1 - w_2\| \quad \text{for all } w_1, w_2 \in \mathcal{V}_L. \tag{4.2}$$

It follows from the definition of M_0 that

$$\|(Z_1 \otimes_S I)A(x_1) - (Z_2 \otimes_S I)A(x_2)\|_F \leq \|M_0(w_1) - M_0(w_2)\|_F \quad \text{for all } w_1, w_2 \in \mathcal{V}_L, \tag{4.3}$$

$$\|X(x_1) \otimes_S I - X(x_2) \otimes_S I\|_F \leq \|M_0(w_1) - M_0(w_2)\|_F \quad \text{for all } w_1, w_2 \in \mathcal{V}_L. \tag{4.4}$$

Moreover, we have from (4.2) and [4, 3.2.12] that

$$\|r_\kappa(w_1, 0) - r_\kappa(w_2, 0) - M_0(w_2)(w_1 - w_2)\| \leq L_M \|w_1 - w_2\|^2 \quad \text{for all } w_1, w_2 \in \mathcal{V}_L.$$

Since $r_\kappa(w_1, 0) - r_\kappa(w_2, 0) - M_0(w_2)(w_1 - w_2) = r_\kappa(w_1, \mu) - r_\kappa(w_2, \mu) - M(w_2, \mu)(w_1 - w_2)$ for all $w_1, w_2 \in \mathcal{V}_L, \mu \geq 0$, it then follows that

$$\|r_\kappa(w_1, \mu) - r_\kappa(w_2, \mu) - M(w_2, \mu)(w_1 - w_2)\| \leq L_M \|w_1 - w_2\|^2 \quad (4.5)$$

for all $w_1, w_2 \in \mathcal{V}_L, \mu \geq 0$. Furthermore, by the definition of M , we obtain

$$\begin{aligned} \|\text{svec}(X(x_1) \circ Z_1 - X(x_2) \circ Z_2) - (Z_2 \otimes_S I)A(x_2)(x_1 - x_2) - (X(x_2) \otimes_S I)\text{svec}(Z_1 - Z_2)\| \\ \leq L_M \|w_1 - w_2\|^2 \end{aligned} \quad (4.6)$$

for all $w_1, w_2 \in \mathcal{V}_L$.

Yamashita and Yabe [9] showed that $M_0(w^*)$ is nonsingular under Assumption 2 (A2)–(A4).

Theorem 4.1. [9, Theorem 1] *Suppose that Assumption 2 (A2)–(A4) hold. Then, the matrix $M_0(w^*)$ is invertible.* \square

Then, this theorem and the implicit function theorem guarantee that there exist a positive constant ζ and a continuously differentiable function $\bar{w} : (-\zeta, \zeta) \rightarrow \mathbf{R}^l$ such that $r_\kappa(\bar{w}(\mu), \mu) = 0$. Furthermore, the following lemma holds.

Lemma 4.1. [9, Lemma 1] *Suppose that Assumption 2 (A1)–(A4) hold. Then, there exist a positive constant γ and a continuously differentiable function $\bar{w} : [0, \gamma] \rightarrow \mathbf{R}^l$ such that*

$$\bar{w}(0) = w^*, \quad \bar{w}(\mu) := [\bar{x}(\mu), \bar{y}(\mu), \text{svec}(\bar{Z}(\mu))], \quad r_\kappa(\bar{w}(\mu), \mu) = 0 \quad \text{for any } \mu \in [0, \gamma].$$

Furthermore, $X(\bar{x}(\mu)) \succ 0$ and $\bar{Z}(\mu) \succ 0$ for any $\mu \in (0, \gamma]$. \square

We call $\{\bar{w}(\mu) | \mu \in [0, \gamma]\}$ the central path of (1.1).

Since $M_0(w^*)$ is invertible, there exists $\varepsilon \in (0, 1)$ such that any matrix $G \in \mathbf{R}^{l \times l}$ that satisfies

$$\|G - M_0(w^*)\|_F < \varepsilon \quad (4.7)$$

is nonsingular. From the continuity of M_0 at w^* , there exists positive constant ν_M such that

$$\|M_0(w) - M_0(w^*)\|_F \leq \frac{1}{4}\varepsilon \quad \text{for any } w \text{ such that } \|w - w^*\| \leq \nu_M. \quad (4.8)$$

Thus, it follows from (4.7) that $M_0(w)$ is nonsingular if $\|w - w^*\| \leq \nu_M$.

Let $\nu := \min\{\nu_M, \nu_L\}$. Then, we define a subset of \mathcal{V}_L .

$$\mathcal{V} := \{ w \in \mathbf{R}^l \mid \|w - w^*\| \leq \nu \} \subset \mathcal{V}_L.$$

Note that M_0 is Lipschitz continuous on \mathcal{V} .

Next, we give a condition on μ under which $M(w, \mu)$ is invertible for any $w \in \mathcal{V}$. Now, let $w \in \mathcal{V}$. By the definition of M and the triangle inequality, we have $\|M(w, \mu) - M_0(w^*)\|_F = \|M_0(w) + \kappa\mu M_I - M_0(w^*)\|_F \leq \|M_0(w) - M_0(w^*)\|_F + \|\kappa\mu M_I\|_F$. It then follows from (4.8) and $\|M_I\|_F = \|I_m\|_F = \sqrt{m}$ that $\|M_0(w) - M_0(w^*)\|_F + \|\kappa\mu M_I\|_F \leq \frac{1}{4}\varepsilon + \kappa\mu\sqrt{m}$. If $\mu \leq s := \frac{\varepsilon}{4(\kappa+1)\sqrt{m}}$, then we have

$$\|M(w, \mu) - M_0(w^*)\|_F \leq \frac{1}{4}\varepsilon + \kappa\mu\sqrt{m} = \left(\frac{1}{4} + \frac{1}{4}\right)\varepsilon = \frac{1}{2}\varepsilon \quad \text{for all } w \in \mathcal{V}. \quad (4.9)$$

Thus, it follows from (4.7) that the matrix $M(w, \mu)$ is invertible for all $w \in \mathcal{V}$ and $\mu \in [0, s]$. Moreover, we may define

$$U_M := \sup_{w \in \mathcal{V}, \mu \in [0, s]} \|M(w, \mu)^{-1}\|_F, \quad U_y := \sup_{w \in \mathcal{V}} \|y\|^2.$$

Note that $U_M < \infty$ from (4.7) and (4.9). Note also that $U_y < \infty$ from the boundedness of \mathcal{V} . Since

$$\|r_\kappa(w, 0)\| = \left\| r_\kappa(w, \mu) - \mu \begin{bmatrix} 0 \\ \kappa y \\ \text{svec}(-I) \end{bmatrix} \right\|, \quad \|r_\kappa(w, \mu)\| = \left\| r_\kappa(w, 0) + \mu \begin{bmatrix} 0 \\ \kappa y \\ \text{svec}(-I) \end{bmatrix} \right\|,$$

we have

$$\mu\sqrt{p} - \|r_\kappa(w, \mu)\| \leq \|r_\kappa(w, 0)\| \leq \|r_\kappa(w, \mu)\| + \mu U_1 \quad \text{for all } w \in \mathcal{V}, \mu \in [0, s], \quad (4.10)$$

$$\mu\sqrt{p} - \|r_\kappa(w, 0)\| \leq \|r_\kappa(w, \mu)\| \leq \|r_\kappa(w, 0)\| + \mu U_1 \quad \text{for all } w \in \mathcal{V}, \mu \in [0, s], \quad (4.11)$$

where $U_1 := \sqrt{\kappa U_y + p}$.

The differentiability of r_κ and X , and the boundedness of \mathcal{V}_L and $[0, s]$ imply that there exist positive constants L_r and L_{XZ} such that

$$\|r_\kappa(w_1, \mu) - r_\kappa(w_2, \mu)\| \leq L_r \|w_1 - w_2\| \quad \text{for all } w_1, w_2 \in \mathcal{V}_L, \mu \in [0, s], \quad (4.12)$$

$$\|X(x_1)Z_1 - X(x_2)Z_2\|_F \leq L_{XZ} \|w_1 - w_2\| \quad \text{for all } w_1, w_2 \in \mathcal{V}_L, \mu \in [0, s]. \quad (4.13)$$

Next, we define a neighborhood of the central path. Let

$$\nu_N := \min \left\{ \nu, \frac{3}{8L_M U_M}, \left[\frac{1}{5L_r^{1+\tau} U_M (1 + U_1)} \right]^{\frac{1}{\tau}} \right\}. \quad (4.14)$$

Then, we define a subset of \mathcal{V} .

$$\mathcal{V}_N := \{ w \in \mathbf{R}^l \mid \|w - w^*\| \leq \nu_N \} \subset \mathcal{V}. \quad (4.15)$$

Note that $\tau \in (0, \frac{1}{2})$ is the constant given in Algorithm 3.2. Secondly, we define two subsets of \mathcal{V}_N .

$$\mathcal{N}_1(\mu) := \{ w \in \mathcal{V}_N \mid \|r_\kappa(w, \mu)\| \leq \mu^{1+\sigma}, X(x) \succeq 0, Z \succeq 0 \},$$

$$\mathcal{N}_2(\mu) := \{ w \in \mathcal{V}_N \mid \|r_\kappa(w, \mu)\| \leq \mu^{1+\rho}, X(x) \succeq 0, Z \succeq 0 \},$$

where σ and ρ are positive constants such that

$$\max \left\{ \frac{\tau}{1-\tau}, \frac{1}{2} \right\} < \rho < 1, \quad \sigma < \frac{\rho - \tau}{1 + \tau}. \quad (4.16)$$

Since $0 < \sigma < \frac{\rho - \tau}{1 + \tau} < \frac{\rho}{1 + \tau} < \rho$, we have $\mathcal{N}_2(\mu) \subset \mathcal{N}_1(\mu)$ for $\mu \in [0, 1]$.

Lemma 4.1 shows that the generalized shifted barrier KKT point $\bar{w}(\mu)$ is unique for $\mu \in [0, \gamma]$. Then, we may regard $\mathcal{N}_1(\mu)$ and $\mathcal{N}_2(\mu)$ as the neighborhoods of the generalized shifted barrier KKT point $\bar{w}(\mu)$. Thus, we define the following neighborhoods of the central path by using $\mathcal{N}_1(\mu)$ and $\mathcal{N}_2(\mu)$.

$$\Theta_1(\theta) := \cup_{\mu \in [0, \theta]} \mathcal{N}_1(\mu), \quad \Theta_2(\theta) := \cup_{\mu \in [0, \theta]} \mathcal{N}_2(\mu) \quad \text{for any } \theta \in [0, \min\{\gamma, s\}].$$

Note that since $0 < s < 1$, we have $0 \leq \theta < 1$. Then,

$$\Theta_2(\theta) \subset \Theta_1(\theta) \subset \mathcal{V}_N \quad \text{for all } \theta \in [0, \min\{\gamma, s\}]. \quad (4.17)$$

We can consider $\Theta_1(\theta)$ and $\Theta_2(\theta)$ as the neighborhoods of the central path. Moreover, we define

$$U_w(\theta) := \sup_{w \in \Theta_1(\theta), \mu \in [0, \theta]} \|w - \bar{w}(\mu)\| \quad \text{for any } \theta \in [0, \min\{\gamma, s\}],$$

which expresses the supremum of a distance between a point in $\Theta_1(\theta)$ and the central path. Now, we briefly show that there exists $\theta_1 > 0$ such that $1 - L_M U_M U_w(\theta) \geq \frac{1}{4}$ for all $\theta \in [0, \theta_1]$ and $\bar{w}(\mu) \in \mathcal{V}_N$ for all $\mu \in [0, \theta_1]$. Since \bar{w} is continuous on $[0, \gamma]$ by Lemma 4.1, there exists $\theta_0 > 0$ such that

$$\|\bar{w}(\mu) - \bar{w}(0)\| \leq \nu_N \quad \text{for all } \mu \in [0, \theta_0]. \quad (4.18)$$

Using $\bar{w}(0) = w^*$, (4.17) and (4.18),

$$\begin{aligned} U_w(\theta) &= \sup_{w \in \Theta_1(\theta), \mu \in [0, \theta]} \|w - w^* + w^* - \bar{w}(\mu)\| \\ &\leq \sup_{w \in \Theta_1(\theta)} \|w - w^*\| + \sup_{\mu \in [0, \theta]} \|\bar{w}(\mu) - \bar{w}(0)\| \\ &\leq 2\nu_N \end{aligned} \quad (4.19)$$

for all $\theta \in [0, \min\{\gamma, s, \theta_0\}]$. Then (4.14) and (4.19) imply that $L_M U_M U_w(\theta) \leq \frac{3}{4}$ for all $\theta \in [0, \min\{\gamma, s, \theta_0\}]$. Thus,

$$1 - L_M U_M U_w(\theta) \geq \frac{1}{4} \quad \text{for all } \theta \in [0, \min\{\gamma, s, \theta_0\}]. \quad (4.20)$$

Moreover, from (4.18) and $\bar{w}(0) = w^*$,

$$\bar{w}(\mu) \in \mathcal{V}_N \quad \text{for all } \mu \in [0, \min\{\gamma, s, \theta_0\}]. \quad (4.21)$$

Hence, letting $\theta_1 := \min\{\gamma, s, \theta_0\}$, we have the desired results. Then, we give a condition under which $r_\kappa(w, \mu)$ provides an error bound of the generalized shifted barrier KKT point.

Lemma 4.2. *Suppose that Assumption 2 holds, and that $\theta \in [0, \theta_1]$. Then,*

$$\|w - \bar{w}(\mu)\| \leq U_r \|r_\kappa(w, \mu)\|, \quad \|XZ - \mu I\|_F \leq U_R \|r_\kappa(w, \mu)\|$$

for all $w \in \Theta_1(\theta)$ and $\mu \in [0, \theta]$, where $U_r := 4U_M$ and $U_R := 4L_{XZ}U_M$.

Proof. Let $w \in \Theta_1(\theta)$ and $\mu \in [0, \theta]$. From (4.21), $\bar{w}(\mu) \in \mathcal{V}_N$ for all $\mu \in [0, \theta]$. Note that $\Theta_1(\theta) \subset \mathcal{V}_N \subset \mathcal{V} \subset \mathcal{V}_L$. Substituting $w_1 = w \in \mathcal{V}_L$ and $w_2 = \bar{w}(\mu) \in \mathcal{V}_L$ into (4.5),

$$L_M \|w - \bar{w}(\mu)\|^2 \geq \|M(\bar{w}(\mu), \mu)(w - \bar{w}(\mu)) - r_\kappa(w, \mu)\| \geq \|M(\bar{w}(\mu), \mu)(w - \bar{w}(\mu))\| - \|r_\kappa(w, \mu)\|$$

from $r_\kappa(\bar{w}(\mu), \mu) = 0$. Since $\|M(\bar{w}(\mu), \mu)(w - \bar{w}(\mu))\| \geq \frac{\|w - \bar{w}(\mu)\|}{\|M(\bar{w}(\mu), \mu)^{-1}\|_F}$, it then follows from $U_w(\theta) \geq \|w - \bar{w}(\mu)\|$ and $U_M \geq \|M(\bar{w}(\mu), \mu)^{-1}\|_F$ that

$$L_M U_w(\theta) \|w - \bar{w}(\mu)\| \geq \frac{\|w - \bar{w}(\mu)\|}{\|M(\bar{w}(\mu), \mu)^{-1}\|_F} - \|r_\kappa(w, \mu)\| \geq \frac{\|w - \bar{w}(\mu)\|}{U_M} - \|r_\kappa(w, \mu)\|.$$

As the result, we have

$$\frac{1 - L_M U_M U_w(\theta)}{U_M} \|w - \bar{w}(\mu)\| \leq \|r_\kappa(w, \mu)\|.$$

Then, since $1 - L_M U_M U_w(\theta) \geq \frac{1}{4}$ from $0 < \theta \leq \theta_1 = \min\{\gamma, s, \theta_0\}$ and (4.20), we obtain

$$\|w - \bar{w}(\mu)\| \leq 4U_M \|r_\kappa(w, \mu)\|. \quad (4.22)$$

By $U_r = 4U_M$, we have the first inequality.

Next, we show the second inequality. We have $X(\bar{x}(\mu)) \circ \bar{Z}(\mu) = \mu I$ by $r_\kappa(\bar{w}(\mu), \mu) = 0$. Since $X(\bar{x}(\mu)) \succeq 0$ and $\bar{Z}(\mu) \succeq 0$, it follows from [8] that $X(\bar{x}(\mu)) \circ \bar{Z}(\mu) = \mu I$ is equivalent to $X(\bar{x}(\mu))\bar{Z}(\mu) = \mu I$. Then, (4.13) yields that

$$L_{XZ} \|w - \bar{w}(\mu)\| \geq \|\text{svec}[XZ - X(\bar{x}(\mu))\bar{Z}(\mu)]\| = \|XZ - \mu I\|_F.$$

Combining this inequality and (4.22), we have $\|XZ - \mu I\|_F \leq 4L_{XZ}U_M \|r_\kappa(w, \mu)\|$. Since $U_R = 4L_{XZ}U_M$, we obtain the desired inequality. \square

From Lemma 4.2, we can show that w^* is an isolated KKT point.

Theorem 4.2. *Suppose that Assumption 2 holds. If $\tilde{w} \in \mathcal{N}_2(0)$, then $\tilde{w} = w^*$.*

Proof. Note that $\tilde{w} \in \mathcal{N}_2(0) = \mathcal{N}_1(0) = \Theta_1(0)$. It then follows from the definition of $\mathcal{N}_2(0)$ that $r_\kappa(\tilde{w}, 0) = 0$. Furthermore, we have from Lemma 4.2 that $\|\tilde{w} - w^*\| = \|\tilde{w} - \bar{w}(0)\| \leq U_r \|r_\kappa(\tilde{w}, 0)\| = 0$, that is, $\tilde{w} = w^*$. \square

4.2. Proof of superlinear convergence

We show the superlinear convergence of Algorithm 3.2 by using the properties given in Subsection 4.1.

First, we give an assumption related to the matrix $P(w)$, which is included in (3.5) and (3.6). To this end, we define $\theta_2 := \min\{\theta_1, (\frac{3}{4U_R})^{\frac{1}{\rho}}\}$ and

$$\Gamma(\theta) := \{ (w, \eta) \in \mathbf{R}^l \times \mathbf{R} \mid w \in \mathcal{N}_2(\eta) \subset \Theta_2(\theta), w \in \mathcal{W}, \eta \in (0, \theta] \} \quad \text{for } \theta \in (0, \theta_2].$$

Then, we make the following assumption on the matrix $P(w)$.

Assumption 3. *The scaling matrix T satisfies Assumption 1 (S1), that is, there exists $P(w)$ such that (3.3) is equivalent to (3.4). The matrix $P(w)$ satisfies the following (S2):*

(S2) *If $\theta \in (0, \theta_2]$, then there exists $U_P > 0$ such that $\|P(w)\|_F \leq U_P \eta^\rho$ for $(w, \eta) \in \Gamma(\theta)$.*

When $T_k = I$, Assumption 3 (S2) holds since $P(w_k) := 0$. Furthermore, when $T_k = X_k^{-\frac{1}{2}}$ or $T_k = W_k^{-\frac{1}{2}}$ ($W_k = X_k^{\frac{1}{2}}(X_k^{\frac{1}{2}}Z_kX_k^{\frac{1}{2}})^{-\frac{1}{2}}X_k^{\frac{1}{2}}$), which are well-known scaling matrices of linear SDP, there exists the matrix $P(w_k)$ such that Assumption 3 (S2) holds. These proofs are given in Appendix C.

Assumption 1 (S1) means that the Newton equations of Steps 2.2 and 2.3 in Algorithm 3.2 are reduced to

$$M_P(w_k, \mu_k) \Delta w_k = -r_\kappa(w_k, \mu_k), \quad M_P(w_k, \mu_k) \Delta \hat{w}_k = -r_\kappa(\hat{w}_k, \mu_k), \quad (4.23)$$

respectively, where

$$M_P(w_k, \mu_k) := M(w_k, \mu_k) + N(w_k), \quad N(w_k) := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ P(w_k) & 0 & 0 \end{bmatrix}. \quad (4.24)$$

It also follows that

$$\|M_P(w_k, \mu_k) - M(w_k, \mu_k)\|_F = \|N(w_k)\|_F = \|P(w_k)\|_F. \quad (4.25)$$

To establish the superlinear convergence of Algorithm 3.2, we first show that $M_P(w_k, \mu_k)$ is nonsingular if $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$ (Lemma 4.4). Then, we show that $\hat{w}_k \in \mathcal{N}_1(\theta)$ and $\hat{w}_k \in \mathcal{W}$ if $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$ (Lemmas 4.5 and 4.7). Finally, we prove that $w_{k+1} \in \mathcal{N}_2(\theta)$ and $w_{k+1} \in \mathcal{W}$ if $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$ (Lemmas 4.8 and 4.9). From these results, we can easily obtain the desired theorem (Theorem 4.3).

In the following two lemmas, we assume that

$$0 < \theta \leq \theta_3, \quad \theta_3 := \min \left\{ \theta_2, \left[\frac{3}{4(1+U_1)^{1+\tau}} \right]^{\frac{1}{\tau}}, \left[\frac{\varepsilon(\sqrt{p}-1)^\rho}{4U_P} \right]^{\frac{1+\tau}{\rho}} \right\}. \quad (4.26)$$

Lemma 4.3. *Suppose that Assumptions 1, 2 and 3 hold, and that θ satisfies (4.26). If $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$, then there exists $\eta_k \in (0, \theta]$ such that $(w_k, \eta_k) \in \Gamma(\theta)$ and $\eta_k \leq U_\eta \mu_k^{\frac{1}{1+\tau}}$, where $U_\eta := (\sqrt{p}-1)^{-1}$. Moreover, $0 < \mu_k < \theta$.*

Proof. Note that $w_k \neq w^*$ because $w_k \in \mathcal{W}$. Then, Theorem 4.2 implies that $w_k \notin \mathcal{N}_2(0)$. Since $w_k \in \Theta_2(\theta)$ and $w_k \notin \mathcal{N}_2(0)$, there exists $\eta_k \in (0, \theta]$ such that $w_k \in \mathcal{N}_2(\eta_k) \subset \Theta_2(\theta)$, i.e., $(w_k, \eta_k) \in \Gamma(\theta)$. It follows from $w_k \in \Theta_2(\theta) \subset \Theta_1(\theta) \subset \mathcal{V}_N \subset \mathcal{V}$, $0 < \eta_k \leq \theta \leq \theta_2 \leq \theta_1 \leq s < 1$ and (4.11) that $\eta_k \sqrt{p} - \|r_\kappa(w_k, 0)\| \leq \|r_\kappa(w_k, \eta_k)\| \leq \eta_k^{1+\rho} \leq \eta_k$. Thus, we have $\eta_k(\sqrt{p}-1) \leq \|r_\kappa(w_k, 0)\| = \mu_k^{\frac{1}{1+\tau}}$, and hence we obtain $\eta_k \leq U_\eta \mu_k^{\frac{1}{1+\tau}}$.

Next, we prove that $0 < \mu_k < \theta$. Since $\eta_k \in (0, \theta]$ and $\eta_k \leq U_\eta \mu_k^{\frac{1}{1+\tau}}$, we get $0 < (\frac{\eta_k}{U_\eta})^{1+\tau} \leq \mu_k$. Moreover, we have from the first part of this proof that $\|r_\kappa(w_k, \eta_k)\| \leq \eta_k^{1+\rho}$. It then follows from (4.10) that $\|r_\kappa(w_k, 0)\| \leq \|r_\kappa(w_k, \eta_k)\| + \eta_k U_1 \leq (\eta_k^\rho + U_1)\eta_k \leq (1+U_1)\theta$. Since $\theta \leq [\frac{3}{4(1+U_1)^{1+\tau}}]^{\frac{1}{\tau}}$ by (4.26), we obtain $\mu_k = \|r_\kappa(w_k, 0)\|^{1+\tau} \leq (1+U_1)^{1+\tau} \theta^\tau \leq \frac{3}{4} \theta < \theta$. Therefore, $0 < \mu_k < \theta$. \square

We have from Lemma 4.3 that if $\theta \in (0, \theta_3]$, $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$, then

$$\exists \eta_k \in (0, \theta] \quad \text{s.t.} \quad (w_k, \eta_k) \in \Gamma(\theta), \quad \eta_k \leq U_\eta \mu_k^{\frac{1}{1+\tau}}, \quad (4.27)$$

and

$$0 < \mu_k < \theta. \quad (4.28)$$

Then, Assumption 3 (S2) and (4.27) imply that

$$\|P(w_k)\|_F \leq U_P \eta_k^\rho \leq \mathcal{U}_P \mu_k^{\frac{\rho}{1+\tau}}, \quad (4.29)$$

where $\mathcal{U}_P := U_\eta^\rho U_P$.

By using (4.28) and (4.29), we prove that the Jacobian matrix $M_P(w_k, \mu_k)$ of (4.23) is nonsingular.

Lemma 4.4. *Suppose that Assumptions 1, 2 and 3 hold, and that θ satisfies (4.26). If $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$, then $M_P(w_k, \mu_k)$ is nonsingular.*

Proof. We have from (4.24) that $\|M_P(w_k, \mu_k) - M_0(w^*)\|_F \leq \|M(w_k, \mu_k) - M_0(w^*)\|_F + \|N(w_k)\|_F$. Then (4.25) yields that

$$\|M_P(w_k, \mu_k) - M_0(w^*)\|_F \leq \|M(w_k, \mu_k) - M_0(w^*)\|_F + \|P(w_k)\|_F. \quad (4.30)$$

We can easily see that $w_k \in \mathcal{V}$ and $\mu_k \in [0, s]$ from $w_k \in \Theta_2(\theta)$, (4.26) and (4.28). Thus, (4.9) yields that

$$\|M(w_k, \mu_k) - M_0(w^*)\|_F \leq \frac{1}{2}\varepsilon. \quad (4.31)$$

On the other hand, it follows from (4.26), (4.28), $U_\eta = (\sqrt{p} - 1)^{-1}$ and $\mathcal{U}_P = U_\eta^\rho U_P$ that $\mu_k \leq \theta \leq \theta_3 \leq \left(\frac{\varepsilon}{4M_P}\right)^{\frac{1+\tau}{\rho}}$, that is, $\mathcal{U}_P \mu_k^{\frac{\rho}{1+\tau}} \leq \frac{1}{4}\varepsilon$. Then, we have from (4.29) that

$$\|P(w_k)\|_F \leq \frac{1}{4}\varepsilon. \quad (4.32)$$

By (4.30), (4.31) and (4.32), $\|M_P(w_k, \mu_k) - M_0(w^*)\|_F \leq \frac{3}{4}\varepsilon$. Therefore, $M_P(w_k, \mu_k)$ is nonsingular from (4.7). \square

We define

$$U_{\mathcal{M}} := \sup \left\{ \|M_P(w, \mu)^{-1}\|_F \mid w \in \Theta_2(\theta_3), w \in \mathcal{W}, \mu := \|r_\kappa(w, 0)\|^{1+\tau} \right\}.$$

It then follows from Lemma 4.4 that if $\theta \in (0, \theta_3]$, $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$, then

$$\|M_P(w_k, \mu_k)^{-1}\|_F \leq U_{\mathcal{M}} < \infty. \quad (4.33)$$

Now, we show that if $w_0 \in \Theta_2(\theta)$ and $w_0 \in \mathcal{W}$ for sufficiently small $\theta > 0$, then $\{w_k\} \subset \Theta_2(\theta)$ and $\{w_k\} \subset \mathcal{W}$. To this end, we first show that \hat{w}_k generated by Step 2.2 of Algorithm 3.2 satisfies $\hat{w}_k \in \mathcal{N}_1(\mu_k) \subset \Theta_1(\theta)$ and $\hat{w}_k \in \mathcal{W}$ if $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$. In what follows, we assume that

$$0 < \theta \leq \theta_4, \quad \theta_4 := \min \left\{ \theta_3, \left[\frac{\nu_N}{U_r + U_2} \right]^{1+\tau}, \left(\frac{1}{U_3} \right)^{\frac{1}{h_1}} \right\}, \quad (4.34)$$

where

$$h_1 := \frac{\rho - \tau}{1 + \tau} - \sigma, \quad U_2 := U_{\mathcal{M}}(1 + U_1), \quad U_3 := U_2(L_M U_2 + \mathcal{U}_P).$$

Note that $h_1 > 0$ from (4.16).

In order to prove $\hat{w}_k \in \mathcal{N}_1(\mu_k)$ and $\hat{w}_k \in \mathcal{W}$, we have to show that $\hat{w}_k \in \mathcal{V}_N$, $\|r_\kappa(\hat{w}_k, \mu_k)\| \leq \mu_k^{1+\sigma}$, $X(\hat{x}_k) \succ 0$ and $\hat{Z}_k \succ 0$. Thus, we first show that $\hat{w}_k \in \mathcal{V}_N$ and $\|r_\kappa(\hat{w}_k, \mu_k)\| \leq \mu_k^{1+\sigma}$. Note that $\mu_k = \|r_\kappa(w_k, 0)\|^{1+\tau}$ and $\hat{w}_k = [\hat{x}_k, \hat{y}_k, \text{svec}(\hat{Z}_k)]$ are generated by Steps 2.1 and 2.2 of Algorithm 3.2, respectively.

Lemma 4.5. *Suppose that Assumptions 1, 2 and 3 hold, and that θ satisfies (4.34). If $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$, then*

$$\|\Delta w_k\| \leq U_2 \mu_k^{\frac{1}{1+\tau}}, \quad \|\hat{w}_k - w^*\| \leq \nu_N, \quad \|r_\kappa(\hat{w}_k, \mu_k)\| \leq U_3 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}, \quad \|r_\kappa(\hat{w}_k, \mu_k)\| \leq \mu_k^{1+\sigma}.$$

Proof. First, we show that $\|\Delta w_k\| \leq U_2 \mu_k^{\frac{1}{1+\tau}}$ and $\|\hat{w}_k - w^*\| \leq \nu_N$. It is clear that $w_k \in \mathcal{V}$ and $\mu_k \in [0, s]$ by $w_k \in \Theta_2(\theta)$, (4.28) and (4.34). Thus, it follows from (4.11) that

$$\|r_\kappa(w_k, \mu_k)\| \leq \|r_\kappa(w_k, 0)\| + \mu_k U_1 = \mu_k^{\frac{1}{1+\tau}} + \mu_k U_1 \leq (1 + U_1) \mu_k^{\frac{1}{1+\tau}}. \quad (4.35)$$

Meanwhile, we have from (4.23) and Lemma 4.4 that $\Delta w_k = -M_P(w_k, \mu_k)^{-1} r_\kappa(w_k, \mu_k)$. Furthermore, (4.33) and (4.35) yield that

$$\|\Delta w_k\| \leq \|M_P(w_k, \mu_k)^{-1}\|_F \|r_\kappa(w_k, \mu_k)\| \leq U_2 \mu_k^{\frac{1}{1+\tau}}. \quad (4.36)$$

By Lemma 4.2 and (4.36),

$$\|\hat{w}_k - w^*\| \leq \|w_k - w^*\| + \|\Delta w_k\| \leq U_r \|r_\kappa(w_k, 0)\| + U_2 \mu_k^{\frac{1}{1+\tau}} = (U_r + U_2) \mu_k^{\frac{1}{1+\tau}}.$$

Then (4.28) and (4.34) imply that $\|\hat{w}_k - w^*\| \leq (U_r + U_2) \mu_k^{\frac{1}{1+\tau}} \leq (U_r + U_2) \theta^{\frac{1}{1+\tau}} \leq \nu_N$.

Secondly, we show that $\|r_\kappa(\hat{w}_k, \mu_k)\| \leq U_3 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}$ and $\|r_\kappa(\hat{w}_k, \mu_k)\| \leq \mu_k^{1+\sigma}$. We easily see that $w_k, \hat{w}_k \in \mathcal{V}_L$. It then follows from (4.5) that

$$\begin{aligned} L_M \|\Delta w_k\|^2 &\geq \|r_\kappa(\hat{w}_k, \mu_k) - r_\kappa(w_k, \mu_k) - M(w_k, \mu_k) \Delta w_k\| \\ &\geq \|r_\kappa(\hat{w}_k, \mu_k)\| - \|r_\kappa(w_k, \mu_k) + M(w_k, \mu_k) \Delta w_k\|. \end{aligned}$$

Moreover, (4.23), (4.25) and (4.29) yield that

$$\begin{aligned} L_M \|\Delta w_k\|^2 &\geq \|r_\kappa(\hat{w}_k, \mu_k)\| - \|(M(w_k, \mu_k) - M_P(w_k, \mu_k)) \Delta w_k\| \\ &\geq \|r_\kappa(\hat{w}_k, \mu_k)\| - \|P(w_k)\|_F \|\Delta w_k\| \\ &\geq \|r_\kappa(\hat{w}_k, \mu_k)\| - \mathcal{U}_P \mu_k^{\frac{\rho}{1+\tau}} \|\Delta w_k\|. \end{aligned} \quad (4.37)$$

Thus, we get $\|r_\kappa(\hat{w}_k, \mu_k)\| \leq L_M U_2^2 \mu_k^{1+\frac{1-\tau}{1+\tau}} + \mathcal{U}_P U_2 \mu_k^{1+\frac{\rho-\tau}{1+\tau}} \leq U_3 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}$ by using (4.36), (4.37) and $\mu_k \in (0, 1)$. Using (4.28), we have $\|r_\kappa(\hat{w}_k, \mu_k)\| \leq U_3 \mu_k^{h_1} \mu_k^{1+\sigma} \leq U_3 \theta^{h_1} \mu_k^{1+\sigma}$. Note that $h_1 = \frac{\rho-\tau}{1+\tau} - \sigma > 0$ by (4.16). Since $U_3 \theta^{h_1} \leq 1$ from (4.34), we get $\|r_\kappa(\hat{w}_k, \mu_k)\| \leq \mu_k^{1+\sigma}$. \square

Next, we show that $\hat{w}_k \in \mathcal{W}$ if we choose θ such that

$$0 < \theta \leq \theta_5, \quad \theta_5 := \min \left\{ \theta_4, \left(\frac{3}{4} \right)^{\frac{1}{\rho}}, \left(\frac{3}{4U_3} \right)^{\frac{1+\tau}{\rho-\tau}} \right\}. \quad (4.38)$$

For this purpose, we present the following lemma. See Appendix B for its proof.

Lemma 4.6. *The following three properties hold.*

- (a) *Let μ , α and K_1 be positive numbers. Furthermore, let A be a matrix included in \mathbf{S}^p . If $\mu \in (0, (\frac{3}{4K_1})^{\frac{1}{\alpha}}]$ and $\|A - \mu I\|_F \leq K_1 \mu^{1+\alpha}$, then $A \succ 0$.*

- (b) Let μ , β and K_2 be positive numbers. Furthermore, let $\Phi : [0, 1] \rightarrow \mathbf{S}^p$ be a function. If $\mu \in (0, (\frac{3}{4K_2})^{\frac{1}{\beta}}]$, $\Phi(0) \succ 0$ and $\|t^{-1}[\Phi(t) - (1-t)\Phi(0)] - \mu I\|_F \leq K_2\mu^{1+\beta}$ for any $t \in (0, 1]$, then $\Phi(t) \succ 0$ for all $t \in (0, 1]$.
- (c) Let $w \in \mathcal{W}$, $d_x \in \mathbf{R}^n$ and $D_Z \in \mathbf{S}^p$. Furthermore, let $\Phi : [0, 1] \rightarrow \mathbf{S}^p$ be defined by $\Phi(t) := X(x + td_x) \circ (Z + tD_Z)$. If $\Phi(t) \succ 0$ for all $t \in (0, 1]$, then $X(x + d_x) \succ 0$ and $Z + D_Z \succ 0$. \square

Lemma 4.7. Suppose that Assumptions 1, 2 and 3 hold, and that θ satisfies (4.38). If $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$, then $\hat{w}_k \in \mathcal{W}$.

Proof. Let $\Phi : [0, 1] \rightarrow \mathbf{S}^p$ be defined by $\Phi(t) := X(x_k + t\Delta x_k) \circ (Z_k + t\Delta Z_k)$. From $w_k \in \mathcal{W}$ and Lemma 4.6 (c), it suffices to show that $\Phi(t) \succ 0$ for all $t \in (0, 1]$. Now, we see that $w_k \in \mathcal{V}_L$ by $w_k \in \Theta_2(\theta)$. Moreover, since Lemma 4.5 yield that $\hat{w}_k \in \mathcal{V}_L$, we have $w_k + t\Delta w_k \in \mathcal{V}_L$ for all $t \in (0, 1]$. Thus, substituting $w_1 = w_k + t\Delta w_k$, $w_2 = w_k$ into (4.6),

$$\begin{aligned} t^2 L_M \|\Delta w_k\|^2 &\geq \|\text{svec}[X(x_k + t\Delta x_k) \circ (Z_k + t\Delta Z_k) - X(x_k) \circ Z_k] \\ &\quad - t[(Z_k \otimes_S I)A(x_k)\Delta x_k + (X(x_k) \otimes_S I)\text{svec}(\Delta Z_k)]\| \\ &= \|\text{svec}[\Phi(t) - (1-t)\Phi(0) - t\mu_k I] + tP(w_k)\Delta x_k\| \\ &\geq \|\Phi(t) - (1-t)\Phi(0) - t\mu_k I\|_F - t\|P(w_k)\|_F \|\Delta x_k\|, \end{aligned} \quad (4.39)$$

where the equality follows from $(Z_k \otimes_S I)A(x_k)\Delta x_k + (X(x_k) \otimes_S I)\text{svec}(\Delta Z_k) = \text{svec}(\mu_k I - X(x_k) \circ Z_k) - P(w_k)\Delta x_k$ in the Newton equation (3.5). It follows from $\|\Delta x_k\| \leq \|\Delta w_k\|$, (4.29) and (4.39) that

$$t^2 L_M \|\Delta w_k\|^2 \geq \|\Phi(t) - (1-t)\Phi(0) - t\mu_k I\|_F - t\mathcal{U}_P \mu_k^{\frac{\rho}{1+\tau}} \|\Delta w_k\|.$$

Since $\|\Delta w_k\| \leq U_2 \mu_k^{\frac{1}{1+\tau}}$ by Lemma 4.5, we have from $t, \mu_k \in (0, 1]$ that

$$\|\Phi(t) - (1-t)\Phi(0) - t\mu_k I\|_F \leq tL_M U_2^2 \mu_k^{\frac{2}{1+\tau}} + t\mathcal{U}_P U_2 \mu_k^{\frac{1+\rho}{1+\tau}} \leq tU_3 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}.$$

Dividing both sides by $t \in (0, 1]$, we obtain

$$\left\| \frac{\Phi(t) - (1-t)\Phi(0)}{t} - \mu_k I \right\|_F \leq U_3 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}. \quad (4.40)$$

Meanwhile, we have from (4.27) and the definition of $\Gamma(\theta)$ that there exists $\eta_k \in (0, \theta]$ such that $w_k \in \mathcal{N}_2(\eta_k)$. In addition, $\eta_k \in (0, (\frac{3}{4})^{\frac{1}{\beta}}]$ by (4.38). Then, the definitions of $r_\kappa(w_k, \eta_k)$ and $\mathcal{N}_2(\eta_k)$ imply that $\|\Phi(0) - \eta_k I\|_F \leq \|r_\kappa(w_k, \eta_k)\| \leq \eta_k^{1+\rho}$. Thus, $\Phi(0) \succ 0$ by Lemma 4.6 (a). Moreover, (4.28) and (4.38) yield that $\mu_k \in (0, (\frac{3}{4U_3})^{\frac{1+\tau}{\rho-\tau}}]$. It then follows from (4.40) and Lemma 4.6 (b) that $\Phi(t) \succ 0$ for all $t \in (0, 1]$. \square

We summarize the results of Lemmas 4.5 and 4.7. Suppose that $\theta \in (0, \theta_5]$, $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$. Lemmas 4.5 and 4.7 imply that

$$\hat{w}_k \in \mathcal{N}_1(\mu_k), \hat{w}_k \in \mathcal{W}, \|\Delta w_k\| \leq U_2 \mu_k^{\frac{1}{1+\tau}}, \|r_\kappa(\hat{w}_k, \mu_k)\| \leq U_3 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}. \quad (4.41)$$

Note that $w_k, \hat{w}_k \in \mathcal{V}_L$. We have from (4.2) and (4.41) that

$$\|M_0(\hat{w}_k) - M_0(w_k)\|_F = \|M_0(w_k + \Delta w_k) - M_0(w_k)\|_F \leq L_M \|\Delta w_k\| \leq \mathcal{U}_M \mu_k^{\frac{1}{1+\tau}}, \quad (4.42)$$

where $\mathcal{U}_M := L_M U_2$.

Next, we show that the sequence $\{w_k\}$ generated by Algorithm 3.2 is included in $\Theta_2(\theta)$ and \mathcal{W} . In what follows, suppose that θ satisfies

$$0 < \theta \leq \theta_6, \quad \theta_6 := \min \left\{ \theta_5, \frac{\nu_N}{U_5}, \left(\frac{1}{U_6} \right)^{\frac{1}{h_2}} \right\}, \quad (4.43)$$

where

$$h_2 := \frac{2\rho - \tau}{1 + \tau} - \rho, \quad U_4 := U_{\mathcal{M}} U_3, \quad U_5 := U_r(U_1 + U_3) + U_4, \quad U_6 := U_4(L_M U_4 + \mathcal{U}_M + \mathcal{U}_P).$$

Note that $h_2 > 0$ from (4.16). First of all, we show $w_{k+1} \in \mathcal{V}_N$ and $\|r_\kappa(w_{k+1}, \mu_k)\| \leq \mu_k^{1+\rho}$.

Lemma 4.8. *Suppose that Assumptions 1, 2 and 3 hold, and that θ satisfies (4.43). If $w_k \in \Theta_2(\theta)$ and $w_k \in \mathcal{W}$, then*

$$\begin{aligned} \hat{w}_k &\in \mathcal{N}_1(\mu_k) \subset \Theta_1(\theta), \quad \mathcal{N}_2(\mu_k) \subset \Theta_2(\theta), \\ \|\Delta \hat{w}_k\| &\leq U_4 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}, \quad \|w_{k+1} - w^*\| \leq \nu_N, \quad \|r_\kappa(w_{k+1}, \mu_k)\| \leq \mu_k^{1+\rho}. \end{aligned}$$

Proof. Note that $w_{k+1} = \hat{w}_k + \Delta \hat{w}_k$. First, we show that $\hat{w}_k \in \mathcal{N}_1(\mu_k) \subset \Theta_1(\theta)$ and $\mathcal{N}_2(\mu_k) \subset \Theta_2(\theta)$. Since $0 < \mu_k < \theta$ by (4.28), the definitions of $\Theta_1(\theta)$ and $\Theta_2(\theta)$ imply that $\mathcal{N}_1(\mu_k) \subset \cup_{\mu \in [0, \theta]} \mathcal{N}_1(\mu) = \Theta_1(\theta)$ and $\mathcal{N}_2(\mu_k) \subset \cup_{\mu \in [0, \theta]} \mathcal{N}_2(\mu) = \Theta_2(\theta)$, respectively. Furthermore, using $\hat{w}_k \in \mathcal{N}_1(\mu_k)$ in (4.41), we have the desired result.

Next, we prove that $\|\Delta \hat{w}_k\| \leq U_4 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}$ and $\|w_{k+1} - w^*\| \leq \nu_N$. We have from (4.23) and Lemma 4.4 that $\Delta \hat{w}_k = -M_P(w_k, \mu_k)^{-1} r_\kappa(\hat{w}_k, \mu_k)$. Moreover, (4.33) and (4.41) yield that

$$\|\Delta \hat{w}_k\| \leq \|M_P(w_k, \mu_k)^{-1}\|_F \|r_\kappa(\hat{w}_k, \mu_k)\| \leq U_{\mathcal{M}} \|r_\kappa(\hat{w}_k, \mu_k)\| \leq U_4 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}. \quad (4.44)$$

On the other hand, it is clear that $\hat{w}_k \in \mathcal{V}$ and $\mu_k \in [0, s]$ from $w_k \in \Theta_2(\theta)$, (4.28) and (4.43). Thus, substituting $w = \hat{w}_k$ and $\mu = \mu_k$ into (4.10), and using (4.41) and $\mu_k \in (0, 1)$, we get

$$\|r_\kappa(\hat{w}_k, 0)\| \leq U_1 \mu_k + \|r_\kappa(\hat{w}_k, \mu_k)\| \leq U_1 \mu_k + U_3 \mu_k^{1+\frac{\rho-\tau}{1+\tau}} \leq (U_1 + U_3) \mu_k. \quad (4.45)$$

It follows from Lemma 4.2 that $\|w_{k+1} - w^*\| \leq \|\hat{w}_k - w^*\| + \|\Delta \hat{w}_k\| \leq U_r \|r_\kappa(\hat{w}_k, 0)\| + \|\Delta \hat{w}_k\|$. Using (4.44), (4.45) and $\mu_k \in (0, 1)$, we obtain $\|w_{k+1} - w^*\| \leq [U_r(U_1 + U_3) + U_4] \mu_k = U_5 \mu_k$. Moreover, since $U_5 \mu_k \leq U_5 \theta \leq \nu_N$ by (4.28) and (4.43), we have $\|w_{k+1} - w^*\| \leq \nu_N$.

Finally, we prove that $\|r_\kappa(w_{k+1}, \mu_k)\| \leq \mu_k^{1+\rho}$. It follows from (4.23) and (4.24) that $r_\kappa(\hat{w}_k, \mu_k) = -M_P(w_k, \mu_k) \Delta \hat{w}_k = -(M(w_k, \mu_k) + N(w_k)) \Delta \hat{w}_k$. Then, since $\hat{w}_k, w_{k+1} \in \mathcal{V}_L$ and $\mu_k \geq 0$, we substitute $w_1 = w_{k+1}$, $w_2 = \hat{w}_k$ and $\mu = \mu_k$ into (4.5), that is,

$$\begin{aligned} L_M \|\Delta \hat{w}_k\|^2 &\geq \|r_\kappa(w_{k+1}, \mu_k) - r_\kappa(\hat{w}_k, \mu_k) - M(\hat{w}_k, \mu_k) \Delta \hat{w}_k\| \\ &\geq \|r_\kappa(w_{k+1}, \mu_k)\| - \|M(\hat{w}_k, \mu_k) - M(w_k, \mu_k) - N(w_k)\|_F \|\Delta \hat{w}_k\| \\ &\geq \|r_\kappa(w_{k+1}, \mu_k)\| - \|M(\hat{w}_k, \mu_k) - M(w_k, \mu_k)\|_F \|\Delta \hat{w}_k\| - \|N(w_k)\|_F \|\Delta \hat{w}_k\| \\ &= \|r_\kappa(w_{k+1}, \mu_k)\| - \|M_0(\hat{w}_k) - M_0(w_k)\|_F \|\Delta \hat{w}_k\| - \|P(w_k)\|_F \|\Delta \hat{w}_k\|, \end{aligned}$$

where the last equality follows from (4.1) and (4.25). Using (4.29) and (4.42), we get $\|r_\kappa(w_{k+1}, \mu_k)\| \leq L_M \|\Delta \hat{w}_k\|^2 + \mathcal{U}_M \mu_k^{\frac{1}{1+\tau}} \|\Delta \hat{w}_k\| + \mathcal{U}_P \mu_k^{\frac{\rho}{1+\tau}} \|\Delta \hat{w}_k\|$, and hence $\|r_\kappa(w_{k+1}, \mu_k)\| \leq L_M U_4^2 \mu_k^{2+\frac{2(\rho-\tau)}{1+\tau}} + \mathcal{U}_M U_4 \mu_k^{1+\frac{1+\rho-\tau}{1+\tau}} + \mathcal{U}_P U_4 \mu_k^{1+\frac{2\rho-\tau}{1+\tau}} \leq U_6 \mu_k^{1+\frac{2\rho-\tau}{1+\tau}}$ from (4.44) and $\mu_k \in (0, 1)$. Since (4.43) implies $U_6 \theta^{h_2} \leq 1$, we have from (4.28) that $U_6 \mu_k^{1+\frac{2\rho-\tau}{1+\tau}} = U_6 \mu_k^{h_2} \mu_k^{1+\rho} \leq U_6 \theta^{h_2} \mu_k^{1+\rho} \leq \mu_k^{1+\rho}$. Note that $h_2 = \frac{2\rho-\tau}{1+\tau} - \rho > 0$ by (4.16). Therefore, we obtain $\|r_\kappa(w_{k+1}, \mu_k)\| \leq \mu_k^{1+\rho}$. \square

Finally, we prove that the sequence $\{w_k\}$ generated by Algorithm 3.2 is included in \mathcal{W} . Let $\tilde{\theta}$ be defined by

$$\tilde{\theta} := \min \left\{ \theta_6, \left(\frac{3}{4} \right)^{\frac{1}{\sigma}}, \left(\frac{3}{4U_7} \right)^{\frac{1+\tau}{2\rho-\tau}} \right\}, \quad (4.46)$$

where $U_7 := U_4(L_M U_4 + 2\mathcal{U}_M + \mathcal{U}_P)$.

Lemma 4.9. *Suppose that Assumptions 1, 2 and 3 hold. If $w_k \in \Theta_2(\tilde{\theta})$ and $w_k \in \mathcal{W}$, then $w_{k+1} \in \mathcal{W}$.*

Proof. Let $\Phi : [0, 1] \rightarrow \mathbf{S}^p$ be defined by $\Phi(t) := X(\hat{x}_k + t\Delta \hat{x}_k) \circ (\hat{Z}_k + t\Delta \hat{Z}_k)$. We see that $\hat{w}_k \in \mathcal{W}$ by (4.41). Then, from Lemma 4.6 (c), it suffices to prove that $\Phi(t) \succ 0$ for all $t \in (0, 1]$. We easily see that $\hat{w}_k, \hat{w}_k + \Delta \hat{w}_k \in \mathcal{V}_L$ by Lemmas 4.5 and 4.8. It then follows that $\hat{w}_k + t\Delta \hat{w}_k \in \mathcal{V}_L$ for all $t \in (0, 1]$. Thus, substituting $w_1 = \hat{w}_k + t\Delta \hat{w}_k$, $w_2 = \hat{w}_k$ into (4.6), we have

$$\begin{aligned} t^2 L_M \|\Delta \hat{w}_k\|^2 &\geq \|\text{svec}[X(\hat{x}_k + t\Delta \hat{x}_k) \circ (\hat{Z}_k + t\Delta \hat{Z}_k) - ((1-t) + t)X(\hat{x}_k) \circ \hat{Z}_k \\ &\quad - t[(\hat{Z}_k \otimes_S I)A(\hat{x}_k)\Delta \hat{x}_k + (X(\hat{x}_k) \otimes_S I)\text{svec}(\Delta \hat{Z}_k)]]\| \\ &= \|\text{svec}[\Phi(t) - (1-t)\Phi(0)] - t\text{svec}[X(\hat{x}_k) \circ \hat{Z}_k \\ &\quad - t[(\hat{Z}_k \otimes_S I)A(\hat{x}_k)\Delta \hat{x}_k + (X(\hat{x}_k) \otimes_S I)\text{svec}(\Delta \hat{Z}_k)]]\| \\ &\geq \|\Phi(t) - (1-t)\Phi(0) - t\mu_k I\|_F - t\|P(w_k)\|_F \|\Delta \hat{w}_k\| \\ &\quad - t\|(\hat{Z}_k \otimes_S I)A(\hat{x}_k) - (Z_k \otimes_S I)A(x_k)\|_F \|\Delta \hat{w}_k\| \\ &\quad - t\|X(\hat{x}_k) \otimes_S I - X(x_k) \otimes_S I\|_F \|\Delta \hat{w}_k\| \\ &\geq \|\Phi(t) - (1-t)\Phi(0) - t\mu_k I\|_F - t\|P(w_k)\|_F \|\Delta \hat{w}_k\| \\ &\quad - 2t\|M_0(\hat{w}_k) - M_0(w_k)\|_F \|\Delta \hat{w}_k\|, \end{aligned}$$

where the second inequality follows from $\text{svec}[X(\hat{x}_k) \circ \hat{Z}_k] = \text{svec}(\mu_k I) - (Z_k \otimes_S I)A(x_k)\Delta \hat{x}_k - (X(x_k) \otimes_S I)\text{svec}(\Delta \hat{Z}_k) - P(w_k)\Delta \hat{x}_k$ in the Newton equation (3.6), and the last inequality follows from (4.3) and (4.4). Then, we exploit (4.29), (4.42) and Lemma 4.8, i.e., $\|P(w_k)\|_F \leq \mathcal{U}_P \mu_k^{\frac{\rho}{1+\tau}}$, $\|M_0(\hat{w}_k) - M_0(w_k)\|_F \leq \mathcal{U}_M \mu_k^{\frac{1}{1+\tau}}$ and $\|\Delta \hat{w}_k\| \leq U_4 \mu_k^{1+\frac{\rho-\tau}{1+\tau}}$. As the result, we get

$$\begin{aligned} \|\Phi(t) - (1-t)\Phi(0) - t\mu_k I\|_F &\leq t^2 L_M \|\Delta \hat{w}_k\|^2 + t\|P(w_k)\|_F \|\Delta \hat{w}_k\| \\ &\quad + 2t\|M_0(\hat{w}_k) - M_0(w_k)\|_F \|\Delta \hat{w}_k\| \\ &\leq t^2 L_M U_4^2 \mu_k^{2+\frac{2(\rho-\tau)}{1+\tau}} + 2t\mathcal{U}_M U_4 \mu_k^{1+\frac{1+\rho-\tau}{1+\tau}} + t\mathcal{U}_P U_4 \mu_k^{1+\frac{2\rho-\tau}{1+\tau}} \\ &\leq tU_7 \mu_k^{1+\frac{2\rho-\tau}{1+\tau}}, \end{aligned}$$

where the last inequality follows from $t, \mu_k \in (0, 1]$. Dividing both sides by $t \in (0, 1]$, we obtain

$$\left\| \frac{\Phi(t) - (1-t)\Phi(0)}{t} - \mu_k I \right\|_F \leq U_7 \mu_k^{1 + \frac{2\rho - \tau}{1 + \tau}}. \quad (4.47)$$

On the other hand, we have from (4.28) and (4.46) that $0 < \mu_k < \min\{(\frac{3}{4})^{\frac{1}{\sigma}}, (\frac{3}{4U_7})^{\frac{1+\tau}{2\rho-\tau}}\}$. In addition, Lemma 4.8 yields that $\hat{w}_k \in \mathcal{N}_1(\mu_k)$. Then, the definitions of $r_\kappa(\hat{w}_k, \mu_k)$ and $\mathcal{N}_1(\mu_k)$ imply that $\|\Phi(0) - \mu_k I\|_F \leq \|r_\kappa(\hat{w}_k, \mu_k)\| \leq \mu_k^{1+\sigma}$. Thus, $\Phi(0) \succ 0$ by Lemma 4.6 (a). It then follows from (4.47) and Lemma 4.6 (b) that $\Phi(t) \succ 0$ for all $t \in (0, 1]$. \square

Using Lemmas 4.8 and 4.9, we prove that $\{w_k\}$ converges to w^* superlinearly.

Theorem 4.3. *Suppose that Assumptions 1, 2 and 3 hold. If $w_0 \in \Theta_2(\tilde{\theta})$ and $w_0 \in \mathcal{W}$, the sequence $\{w_k\}$ generated by Algorithm 3.2 converges to w^* superlinearly.*

Proof. Note that $\tilde{\theta} \leq \theta_i$ ($i = 1, \dots, 6$) from the definitions of $\theta_1, \dots, \theta_6$ and $\tilde{\theta}$. First, we show the following relations by the mathematical induction.

$$w_k \in \mathcal{N}_2(\mu_{k-1}) \subset \Theta_2(\tilde{\theta}), \quad w_k \in \mathcal{W} \quad \text{for all positive } k. \quad (4.48)$$

Since $w_0 \in \Theta_2(\tilde{\theta})$ and $w_0 \in \mathcal{W}$, we have from Lemmas 4.8 and 4.9 that $w_1 \in \mathcal{N}_2(\mu_0) \subset \Theta_2(\tilde{\theta})$ and $w_1 \in \mathcal{W}$. Next, let $k \geq 2$. Suppose that $w_k \in \mathcal{N}_2(\mu_{k-1}) \subset \Theta_2(\tilde{\theta})$ and $w_k \in \mathcal{W}$. Then, it follows from Lemmas 4.8 and 4.9 that $w_{k+1} \in \mathcal{N}_2(\mu_k) \subset \Theta_2(\tilde{\theta})$ and $w_{k+1} \in \mathcal{W}$. Therefore, the proof of (4.48) is complete.

Secondly, we prove that $\{w_k\}$ converges to w^* superlinearly. Let k be an arbitrary positive integer. Note that $0 < \mu_k < \tilde{\theta} \leq \theta_1 = \min\{\gamma, s, \theta_0\} < 1$ from (4.28). Then, note also that $w_{k+1} \in \mathcal{N}_2(\mu_k) \subset \Theta_2(\tilde{\theta}) \subset \Theta_1(\tilde{\theta}) \subset \mathcal{V}_N \subset \mathcal{V}$ by (4.15), (4.17) and (4.48). Lemma 4.2 and (4.10) yield that $\|w_{k+1} - w^*\| \leq U_r \|r_\kappa(w_{k+1}, 0)\| \leq U_r (\|r_\kappa(w_{k+1}, \mu_k)\| + U_1 \mu_k) \leq U_r (\mu_k^\rho + U_1) \mu_k \leq U_r (1 + U_1) \mu_k$. Thus, we obtain $\|w_{k+1} - w^*\| \leq U_r (1 + U_1) \|r_\kappa(w_k, 0) - r_\kappa(w^*, 0)\|^{1+\tau}$ by $\mu_k = \|r_\kappa(w_k, 0)\|^{1+\tau}$ and $r_\kappa(w^*, 0) = 0$. It then follows from $w_k, w^* \in \mathcal{V}_L$ and (4.12) that

$$\|w_{k+1} - w^*\| \leq U_r (1 + U_1) \|r_\kappa(w_k, 0) - r_\kappa(w^*, 0)\|^{1+\tau} \leq L_r^{1+\tau} U_r (1 + U_1) \|w_k - w^*\|^{1+\tau}. \quad (4.49)$$

Using (4.14), $w_k \in \mathcal{V}_N$ and $U_r = 4U_M$ that

$$L_r^{1+\tau} U_r (1 + U_1) \|w_k - w^*\|^\tau \leq L_r^{1+\tau} U_r (1 + U_1) \nu_N^\tau \leq \frac{U_r}{5U_M} = \frac{4}{5}. \quad (4.50)$$

It follows from (4.49) and (4.50) that $\|w_{k+1} - w^*\| \leq \frac{4}{5} \|w_k - w^*\|$, and hence $\{\|w_k - w^*\|\}$ converges to 0. Since $\lim_{k \rightarrow \infty} L_r^{1+\tau} U_r (1 + U_1) \|w_k - w^*\|^\tau = 0$, we have from (4.49) that

$$\lim_{k \rightarrow \infty} \frac{\|w_{k+1} - w^*\|}{\|w_k - w^*\|} = 0.$$

Therefore, $\{w_k\}$ converges to w^* superlinearly. \square

5. Concluding Remarks

In this paper, we proposed a new two-step primal-dual interior point method (Algorithm 3.2) based on the generalized shifted barrier KKT conditions (2.2) for the nonlinear SDP and proved the superlinear convergence of the proposed method. In particular, in order to reduce calculations, we replaced the coefficient matrix in the second equation with that in the first. Therefore, we can expect that the proposed method can find the next point faster than existing methods [3, 5, 9].

Acknowledgements

The authors would like to thank the anonymous referees for their helpful comments and suggestions.

References

- [1] G. Golub, and L.C. Van: *Matrix Computations* (The Johns Hopkins University Press, 1989).
- [2] A. Kato, H. Yabe, and H. Yamashita: An interior point method with a primal-dual quadratic barrier penalty function for nonlinear semidefinite programming. *Technical Report*, (Department of Mathematical Information Science, Tokyo University of Science, 2012).
- [3] Z.-Q. Luo, J.F. Sturm, and S. Zhang: Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. *SIAM Journal on Optimization*, **8** (1998), 59–81.
- [4] J. Ortega, and W. Rheinboldt: *Iterative solution of nonlinear equations in several variables* (Academic Press, 1970).
- [5] F.A. Potra, and R. Sheng: A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming. *SIAM Journal on Optimization*, **8** (1998), 1007–1028.
- [6] G. Stewart: *Matrix algorithms* (Society for Industrial and Applied Mathematics, 1998).
- [7] M.J. Todd, K.C. Toh, and R.H. Tütüncü: On the Nesterov-Todd direction in semidefinite programming. *SIAM Journal on Optimization*, **8** (1998), 769–796.
- [8] Y. Yamakawa, and N. Yamashita: A differentiable merit function for the shifted perturbed KKT conditions of the nonlinear semidefinite programming. *Technical Report*, (Department of Applied Mathematics and Physics, Kyoto University, 2013).
- [9] H. Yamashita, and H. Yabe: Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming*, **132** (2012), 1–30.
- [10] H. Yamashita, H. Yabe, and K. Harada: A primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming*, **135** (2012), 89–121.

Appendix A.

In Appendix A, we describe the second-order sufficient condition, the strict complementarity condition and the nondegeneracy condition.

Let x^* be a stationary point of nonlinear SDP (1.1), and let $\Lambda(x^*)$ be a set of corresponding Lagrange multipliers, that is,

$$\Lambda(x^*) := \{ (y, Z) \in \mathbf{R}^m \times \mathbf{S}^p \mid [x^*, y, \text{svec}(Z)] \text{ satisfies (2.1) } \}.$$

First, we describe the second-order sufficient condition for nonlinear SDP (1.1). Let $C(x^*)$ be the critical cone of (1.1) at x^* , that is,

$$C(x^*) := \left\{ h \in \mathbf{R}^n \mid J_g(x^*)h = 0, \sum_{i=1}^n h_i A_i(x^*) \in \mathcal{T}_{\mathbf{S}_+^p}(X(x^*)), \nabla f(x^*)^\top h = 0 \right\},$$

where $\mathcal{T}_{\mathbf{S}_+^p}(X(x^*)) := \{ D \in \mathbf{S}^p \mid \text{dist}(X(x^*) + tD, \mathbf{S}_+^p) = o(t), t \geq 0 \}$ and $\text{dist}(P, \mathbf{S}_+^p) := \inf \{ \|P - Q\|_F \mid Q \in \mathbf{S}_+^p \}$. Then, we say that the second-order sufficient condition holds

at x^* if

$$\sup_{(y,Z) \in \Lambda(x^*)} h^\top \left(\nabla_{xx}^2 L(x^*, y, \text{svec}(Z)) + \hat{H}(x^*, Z) \right) h > 0 \quad \text{for all } h \in C(x^*) \setminus \{0\},$$

where the (i, j) -th element of $\hat{H}(x^*, Z)$ is $(\hat{H}(x^*, Z))_{ij} := 2\text{tr}(A_i(x^*)X(x^*)^\dagger A_j(x^*)Z)$, and $X(x^*)^\dagger$ denotes the Moore-Penrose generalized inverse of $X(x^*)$.

Next, we describe the strict complementarity condition and the nondegeneracy condition. We say that the strict complementarity condition holds at x^* if there exists $(y^*, Z^*) \in \Lambda(x^*)$ such that $\text{rank}(X(x^*)) + \text{rank}(Z^*) = p$. Then, without loss of generality, we may assume that $X(x^*)$ and Z^* are written as

$$X(x^*) = \begin{bmatrix} \bar{X}^* & 0 \\ 0 & 0 \end{bmatrix}, \quad Z^* = \begin{bmatrix} 0 & 0 \\ 0 & \underline{Z}^* \end{bmatrix},$$

where $\bar{X}^* \in \mathbf{S}_{++}^q$ and $\underline{Z}^* \in \mathbf{S}_{++}^r$, and q and r are positive integers such that $q + r = p$. Then, for each $i \in \{1, \dots, n\}$, let $\underline{A}_i(x) \in \mathbf{S}^r$ be a submatrix of $A_i(x)$ such that

$$A_i(x) = \begin{bmatrix} \bar{A}_i(x) & \hat{A}_i(x) \\ \hat{A}_i(x)^\top & \underline{A}_i(x) \end{bmatrix},$$

where $\bar{A}_i(x)$ and $\hat{A}_i(x)$ are appropriate submatrices of $A_i(x)$. We further define

$$B(x) := [\text{svec}(\underline{A}_1(x)), \dots, \text{svec}(\underline{A}_n(x))] \in \mathbf{R}^{\frac{r(r+1)}{2} \times n}, \quad K(x) := \begin{bmatrix} J_g(x) \\ B(x) \end{bmatrix} \in \mathbf{R}^{(m + \frac{r(r+1)}{2}) \times n}.$$

We say that the nondegeneracy condition holds at x^* if $\text{rank}(K(x^*)) = m + \frac{r(r+1)}{2}$.

Appendix B.

In Appendix B, we give the proof of Lemma 4.6.

Proof of Lemma 4.6. We first prove (a). Since $\mu \in (0, (\frac{3}{4K_1})^{\frac{1}{\alpha}}]$ and $\|A - \mu I\|_F \leq K_1 \mu^{1+\alpha}$, we have $\|A - \mu I\|_2 \leq \|A - \mu I\|_F \leq K_1 \mu^{1+\alpha} = K_1 \mu^\alpha \mu \leq \frac{3}{4} \mu$, where the extreme left-hand side inequality follows from [1, 2.3.2]. Thus, we have $v^\top A v = v^\top (A - \mu I) v + \mu \|v\|^2 \geq (\mu - \|A - \mu I\|_2) \|v\|^2 \geq \frac{1}{4} \mu \|v\|^2 > 0$ for all $v (\neq 0) \in \mathbf{R}^p$, where the first inequality follows from the Cauchy-Schwarz inequality and the definition of $\|\cdot\|_2$. Therefore, this inequality implies that $A \succ 0$.

Secondly, we show (b). It follows from (a) that $t^{-1}[\Phi(t) - (1-t)\Phi(0)] \succ 0$ for all $t \in (0, 1]$. If $t = 1$, then $\Phi(1) \succ 0$. On the other hand, if $t \in (0, 1)$, then $\Phi(t) \succ (1-t)\Phi(0) \succ 0$ for all $t \in (0, 1)$. Therefore, $\Phi(t) \succ 0$ for all $t \in (0, 1]$.

Finally, we give the proof of (c), that is, we show that $X(x + td_x) \succ 0$ for any $t \in (0, 1]$. To this end, suppose the opposite, i.e., there exists $\bar{t} \in (0, 1]$ such that $X(x + \bar{t}d_x)$ is not positive definite. Note that $w \in \mathcal{W}$ implies that $X(x) \succ 0$. It follows from the continuity of eigenvalues of $X(x)$ that there exists $\tilde{t} \in (0, \bar{t}]$ such that $\lambda_{\min}(X(x + \tilde{t}d_x)) = 0$. Thus, $X(x + \tilde{t}d_x)$ is singular, that is, there exists $v_0 \neq 0$ such that $X(x + \tilde{t}d_x)v_0 = 0$. Then, we obtain $v_0^\top \Phi(\tilde{t})v_0 = \frac{1}{2} [v_0^\top X(x + \tilde{t}d_x)(Z + \tilde{t}D_Z)v_0 + v_0^\top (Z + \tilde{t}D_Z)X(x + \tilde{t}d_x)v_0] = 0$. However, this contradicts $\Phi(\tilde{t}) \succ 0$ for any $t \in (0, 1]$. Similarly, $Z + tD_Z \succ 0$ for any $t \in (0, 1]$. Therefore, $t = 1$ implies that $X(x + d_x) \succ 0$ and $Z + D_Z \succ 0$. \square

Appendix C.

In Appendix C, we show that there exists $P(w)$ such that Assumption 1 holds for $T = X^{-\frac{1}{2}}$ and $T = W^{-\frac{1}{2}}$, and Assumption 3 also holds.

In what follows, we define $E(\eta) := XZ - \eta I$ ($\eta \in \mathbf{R}$). First, we give two inequalities which evaluate $E(\eta)$ and X^{-1} over $\Theta_2(\theta)$. Secondly, we also give an inequality which evaluates $A \otimes_S B$ for any $A, B \in \mathbf{R}^{p \times p}$. These inequalities play important roles in evaluation of $P(w)$.

Lemma C.1. *Suppose that Assumption 2 holds, and that $\theta \in (0, \theta_2]$. Then, we obtain $\|E(\eta)\|_F \leq U_R \eta^{1+\rho}$ and $\|X^{-1}\|_F \leq U_X \eta^{-1}$ for any $(w, \eta) \in \Gamma(\theta)$, where $U_X := 4pU_Z$ and $U_Z := \sup\{\|Z\|_F | w \in \Theta_2(\theta_2), w \in \mathcal{W}\}$.*

Proof. For any $(w, \eta) \in \Gamma(\theta)$, we have from the definition of $\Gamma(\theta)$ that $w \in \mathcal{N}_2(\eta) \subset \Theta_2(\theta)$, $w \in \mathcal{W}$ and $\eta \in (0, \theta]$. Thus, we also have from the definition of $\mathcal{N}_2(\eta)$ that $\|r_\kappa(w, \eta)\| \leq \eta^{1+\rho}$. Moreover, $w \in \Theta_2(\theta) \subset \Theta_1(\theta)$, $\eta \in (0, \theta] \subset [0, \theta_1]$ and Lemma 4.2 yield that $\|E(\eta)\|_F \leq U_R \|r_\kappa(w, \eta)\| \leq U_R \eta^{1+\rho}$. It then follows from $\eta \leq \theta \leq \theta_2 \leq (\frac{3}{4U_R})^{\frac{1}{\rho}}$ that $\|I - \eta^{-1}XZ\|_F = \eta^{-1}\|E(\eta)\|_F \leq U_R \eta^\rho \leq \frac{3}{4}$. Thus, $\eta \frac{\|X^{-1}\|_F}{\|Z\|_F} \leq \eta \|Z^{-1}X^{-1}\|_F = \|(I - (I - \eta^{-1}XZ))^{-1}\|_F \leq \frac{p}{1 - \|I - \eta^{-1}XZ\|_F} \leq 4p$, where the second inequality follows from [6, Theorem 4.18]. Since $w \in \Theta_2(\theta) \subset \Theta_2(\theta_2)$ and $w \in \mathcal{W}$, we obtain $0 < \|Z\|_F \leq U_Z$. Hence, we get $\|X^{-1}\|_F \leq 4pU_Z \eta^{-1}$. Letting $U_X = 4pU_Z$, we obtain the desired inequality. \square

Lemma C.2. *Let $C_1 := \sqrt{\frac{p(p+1)}{2}}$. Then, $\|A \otimes_S B\|_F \leq C_1 \|A\|_F \|B\|_F$ for any $A, B \in \mathbf{R}^{p \times p}$.*

Proof. It follows from [1, 2.3.2] that $\|A \otimes_S B\|_F \leq C_1 \|A \otimes_S B\|_2$. Let $U \in \mathbf{S}^p$. The properties of the symmetrized Kronecker product and svec operator yield that $\|(A \otimes_S B) \text{svec}(U)\| = \|(A \odot B)U\|_F = \frac{1}{2} \|AUB^\top + BUA^\top\|_F \leq \|A\|_F \|B\|_F \|U\|_F$. Hence,

$$\|A \otimes_S B\|_2 = \sup_{\text{svec}(U) \neq 0} \frac{\|(A \otimes_S B) \text{svec}(U)\|}{\|\text{svec}(U)\|} \leq \sup_{U \neq 0} \frac{\|A\|_F \|B\|_F \|U\|_F}{\|U\|_F} = \|A\|_F \|B\|_F.$$

To sum up the above discussion, we obtain that $\|A \otimes_S B\|_F \leq C_1 \|A\|_F \|B\|_F$. \square

(i) HRVW/KSH/M ($T = X^{-\frac{1}{2}}$)

First, we discuss the case where $T = X^{-\frac{1}{2}}$. We define

$$F(w, \eta) := \frac{1}{2} [E(\eta) \otimes_S X^{-1} - I \otimes_S (X^{-1}E(\eta))] A(x).$$

It then follows from Yamashita and Yabe [9, Lemma 4] that $P(w) := F(w, \eta)$ satisfies Assumption 1 (S1). Note that we can choose η arbitrarily.

Next, we show that (S2) holds. Suppose that Assumption 2 holds, and that $\theta \in (0, \theta_2]$. For any $(w, \eta) \in \Gamma(\theta)$, it follows from Lemma C.2 that

$$\|F(w, \eta)\|_F \leq \frac{1}{2} U_A [\|E(\eta) \otimes_S X^{-1}\|_F + \|I \otimes_S (X^{-1}E(\eta))\|_F] \leq C_2 U_A \|E(\eta)\|_F \|X^{-1}\|_F, \quad (\text{C.1})$$

where $U_A := \sup\{\|A(x)\|_F | w \in \Theta_2(\theta_2), w \in \mathcal{W}\}$ and $C_2 := \frac{1+\sqrt{p}}{2} \sqrt{\frac{p(p+1)}{2}}$. We get $\|P(w)\|_F = \|F(w, \eta)\|_F \leq C_2 U_A U_R U_X \eta^\rho$ from Lemma C.1 and (C.1). Therefore, letting $U_P = C_2 U_A U_R U_X$, we see that Assumption 3 (S2) holds.

(ii) **NT** ($T = W^{-\frac{1}{2}}$)

In this part, we discuss the case where we choose $T = W^{-\frac{1}{2}}$, where $W = X^{\frac{1}{2}}(X^{\frac{1}{2}}ZX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}}$. We define

$$\begin{aligned} G(w, \eta) := & -\frac{1}{2}(I \otimes_S (E(\eta)^\top X^{-1}))A(x) + \eta(I \otimes_S (X^{-\frac{1}{2}}H(w, \eta)X^{-\frac{1}{2}}))A(x) + \frac{1}{2\eta}(E(\eta) \otimes_S Z)A(x) \\ & - \frac{1}{4\eta}(E(\eta) \otimes_S (E(\eta)^\top X^{-1}))A(x) + \frac{1}{2}(E(\eta) \otimes_S (X^{-\frac{1}{2}}H(w, \eta)X^{-\frac{1}{2}}))A(x) \\ & + ((X^{\frac{1}{2}}H(w, \eta)X^{-\frac{1}{2}}) \otimes_S Z)A(x) - \frac{1}{2}((X^{\frac{1}{2}}H(w, \eta)X^{-\frac{1}{2}}) \otimes_S (E(\eta)^\top X^{-1}))A(x) \\ & + \eta((X^{\frac{1}{2}}H(w, \eta)X^{-\frac{1}{2}}) \otimes_S (X^{-\frac{1}{2}}H(w, \eta)X^{-\frac{1}{2}}))A(x), \end{aligned}$$

where $H(w, \eta) := \left(I + \frac{1}{\eta}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}\right)^{\frac{1}{2}} - I - \frac{1}{2\eta}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}$. Then, letting $P(w) := G(w, \eta)$, we see that Assumption 1 (S1) holds.

Next, we prove that (S2) holds. Suppose that Assumption 2 holds, and that $\theta \in (0, \theta_2]$. For any $(w, \eta) \in \Gamma(\theta)$, it follows from Lemma C.2 that there exists $C_3 > 0$ such that

$$\begin{aligned} & \|G(w, \eta)\|_F \\ & \leq C_3 U_A (\|I\|_F \|E(\eta)\|_F \|X^{-1}\|_F + \eta \|I\|_F \|X^{-\frac{1}{2}}\|_F^2 \|H(w, \eta)\|_F + \eta^{-1} \|E(\eta)\|_F \|Z\|_F \\ & \quad + \eta^{-1} \|E(\eta)\|_F^2 \|X^{-1}\|_F + \|X^{-\frac{1}{2}}\|_F^2 \|H(w, \eta)\|_F \|E(\eta)\|_F + \|Z\|_F \|X^{\frac{1}{2}}\|_F \|H(w, \eta)\|_F \|X^{-\frac{1}{2}}\|_F \\ & \quad + \|X^{\frac{1}{2}}\|_F \|H(w, \eta)\|_F \|X^{-\frac{1}{2}}\|_F \|E(\eta)\|_F \|X^{-1}\|_F + \eta \|X^{-\frac{1}{2}}\|_F^3 \|H(w, \eta)\|_F^2 \|X^{\frac{1}{2}}\|_F). \end{aligned} \quad (\text{C.2})$$

In what follows, we evaluate $\|X^{-\frac{1}{2}}\|_F$ and $\|H(w, \eta)\|_F$. First, it follows from Lemma C.1 that

$$\|X^{-\frac{1}{2}}\|_F = \sqrt{\text{tr}(X^{-1})} = \sqrt{\langle X^{-1}, I \rangle} \leq \sqrt{\|I\|_F \|X^{-1}\|_F} \leq \sqrt{p^{\frac{1}{2}} U_X \eta^{-\frac{1}{2}}}. \quad (\text{C.3})$$

Next, we evaluate $\|H(w, \eta)\|_F$. For this purpose, we first evaluate $\eta^{-1}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}$ which constitutes $H(w, \eta)$. Since $E(\eta) = XZ - \eta I$ implies $X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}} = X^{\frac{1}{2}}ZX^{\frac{1}{2}} - \eta I \in \mathbf{S}^p$, we obtain $\|X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}\|_F = \sqrt{\text{tr}(X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}})} = \sqrt{\text{tr}(E(\eta)^2)} = \sqrt{\langle E(\eta)^\top, E(\eta) \rangle}$. Then, the Cauchy-Schwarz inequality yields that

$$\|\eta^{-1}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}\|_F \leq \eta^{-1} \sqrt{\|E(\eta)^\top\|_F \|E(\eta)\|_F} = \eta^{-1} \|E(\eta)\|_F. \quad (\text{C.4})$$

Since $\eta^{-1}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}$ is symmetric, all eigenvalues $\lambda_1, \dots, \lambda_p$ are real numbers. Moreover, there exists an orthogonal matrix V such that $\eta^{-1}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}} = VDV^\top$, where $D = \text{diag}[\lambda_1, \dots, \lambda_p]$. From Lemma C.1, $0 \leq \eta \leq \theta \leq \theta_2 \leq (\frac{3}{4U_R})^{\frac{1}{\rho}}$ and (C.4),

$$1 > \frac{3}{4} \geq U_R \eta^\rho \geq \eta^{-1} \|E(\eta)\|_F \geq \|\eta^{-1}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}\|_F = \sqrt{\sum_{i=1}^p \lambda_i^2} \geq |\lambda_i|,$$

for $i = 1, \dots, p$. As the result, $I + D$ is symmetric positive definite, i.e., the existence of

$(I + D)^{\frac{1}{2}}$ is guaranteed. Considering the diagonalization of $\eta^{-1}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}}$,

$$\begin{aligned}
\|H(w, \eta)\|_F &= \left\| (VV^\top + VDV^\top)^{\frac{1}{2}} - VV^\top - \frac{1}{2}VDV^\top \right\|_F \\
&= \left\| V(I + D)^{\frac{1}{2}}V^\top - VV^\top - \frac{1}{2}VDV^\top \right\|_F \\
&\leq \|V\|_F^2 \left\| (I + D)^{\frac{1}{2}} - I - \frac{1}{2}D \right\|_F \\
&= p \sqrt{\sum_{i=1}^p \left(\sqrt{1 + \lambda_i} - 1 - \frac{1}{2}\lambda_i \right)^2}. \tag{C.5}
\end{aligned}$$

Let $\varphi : (-1, 1) \rightarrow \mathbf{R}$ be defined by $\varphi(u) := \sqrt{1 + u}$. Since φ is a twice continuously differentiable function defined on the bounded convex set, it follows from [4, 3.3.6] that there exists a positive constant L_φ such that $L_\varphi u^2 \geq |\varphi(u) - \varphi(0) - \varphi'(0)u| = \left| \sqrt{1 + u} - 1 - \frac{1}{2}u \right|$. Then, Lemma C.1, (C.4) and (C.5) yield that

$$\begin{aligned}
\|H(w, \eta)\|_F &\leq pL_\varphi \sqrt{\sum_{i=1}^p \lambda_i^4} \\
&= pL_\varphi \left\| \left(\eta^{-1}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}} \right)^2 \right\|_F \\
&\leq pL_\varphi \left\| \eta^{-1}X^{-\frac{1}{2}}E(\eta)X^{\frac{1}{2}} \right\|_F^2 \\
&\leq pL_\varphi \|E(\eta)\|_F^2 \eta^{-2} \\
&\leq pL_\varphi U_R^2 \eta^{2\rho}.
\end{aligned}$$

This inequality, Lemma C.1, (C.2), (C.3), the boundedness of $\|Z\|_F$ and $\|X^{\frac{1}{2}}\|_F$ imply that there exists $C_4 > 0$ such that $\|G(w, \eta)\|_F \leq C_4(\eta^\rho + \eta^{2\rho} + \eta^{3\rho} + \eta^{2\rho - \frac{1}{2}} + \eta^{3\rho - \frac{1}{2}} + \eta^{4\rho - \frac{1}{2}}) \leq 6C_4\eta^\rho$, where the second inequality follows from $\rho \leq 2\rho - \frac{1}{2}$ in (4.16) and $0 < \eta \leq \theta_2 \leq \theta_1 \leq s < 1$. Letting $U_P = 6C_4$, Assumption 3 (S2) holds.

These results show that $T = X^{-\frac{1}{2}}$ and $T = W^{-\frac{1}{2}}$ satisfy Assumptions 1 and 3.

Yuya Yamakawa
Kyoto University
Yoshidahommachi, Sakyo-ku, Kyoto-shi, Ky-
oto 606-8501, Japan
E-mail : yamakawa@amp.i.kyoto-u.ac.jp