Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming

Hiroshi Yamashita* and Hiroshi Yabe†

January 26, 2009

Abstract

In this paper, we consider a primal-dual interior point method for solving nonlinear semidefinite programming problems. We propose primal-dual interior point methods based on the unscaled and scaled Newton methods, which correspond to the AHO, HRVW/KSH/M and NT search directions in linear SDP problems. We analyze local behavior of our proposed methods and show their local and superlinear convergence properties.

Key words. nonlinear semidefinite programming, primal-dual interior point method, local and superlinear convergence

1 Introduction

We consider the following nonlinear semidefinite programming (SDP) problem:

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

where the functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $X : \mathbb{R}^n \to \mathbb{S}^p$ are sufficiently smooth, and $\mathbb{S}^p$ 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.

If all the functions f and g are linear and the matrix $X(x)$ is defined by

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

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

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

with given matrices $A_i \in \mathbb{S}^p, i = 1, \dots, n$, and $B \in \mathbb{S}^p$, then problem (1) reduces to the linear SDP problem. The linear SDP problems include linear programming problems, convex quadratic programming problems and second order cone programming problems, and they have many applications. As numerical methods for linear SDP problems, interior point methods have been studied extensively by many researchers, see for example [19, 22] and the references therein.

On the other hand, researches on theoretical properties and numerical methods for nonlinear SDP are much more recent. Nonlinear SDP problems also have been attracting a great deal of research attention, because such problems arise from several application fields, which include control theory, eigenvalue problems, finance and so forth. For this reason, it is desired to develop a numerical method for solving nonlinear SDP problems. Recently Yamashita, Yabe and Harada [23] proposed a primal-dual interior point method for solving problem (1) and proved its global convergence. Their computational experiments show that the proposed method performs well in practice.

In this paper, we analyze local behavior of primal-dual interior point methods based on the unscaled and scaled Newton methods, which correspond to the AHO direction [1], the HRVW/KSH/M direction [7, 10, 12] and the NT direction [13, 14] in the linear SDP problems. Researches on the rate of convergence of the primal-dual interior point methods for linear SDP problems can be found in [8, 9, 10, 11, 15]. However, in our knowledge, there are few similar researches for nonlinear SDP problems. Existing literatures include [5] and [6] both of which analyze SQP type method.

The present paper is organized as follows. In Section 2, the optimality conditions for problem (1) and some notations are described. In Section 3, we briefly review the primal-dual interior point method proposed by Yamashita et al. [23], and introduce the AHO, HRVW/KSH/M and NT directions. In Section 4, we present some definitions that are necessary for analysis in the subsequent sections. Sections 5 and 6 are devoted to showing local and superlinear convergence properties of our proposed methods. Specifically, in Section 5, we prove local and superlinear convergence of the primal-dual interior point method based on the unscaled Newton method, which corresponds to the AHO search direction. In Sections 6.1 and 6.2, we prove local and two-step superlinear convergence properties of the primal-dual interior point methods based on the scaled Newton methods, which correspond to the HRVW/KSH/M and the NT search directions, respectively.

2 Optimality conditions and notations

In this section, we define some notations used in this paper, and we give optimality conditions for problem (1).

We first define the inner product $\langle X, Z \rangle$ by $\langle X, Z \rangle = \mathrm{tr}(XZ)$ for any matrices X and Z in $\mathbb{S}^p$, where $\mathrm{tr}(M)$ denotes the trace of the matrix M. Let the Lagrangian function of problem (1) be defined by

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

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

We also define matrices

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

for $i = 1, \ldots, n$. Then Karush-Kuhn-Tucker (KKT) conditions for optimality of problem (1) are given by the following (see [4]):

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

and

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

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

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

and $\mathcal{A}^*(x)$ is an operator which yields

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

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

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

and

$$X(x) \succ 0, \qquad Z \succ 0. \tag{5}$$

To obtain a symmetrized form, we use the multiplication $X(x) \circ Z$ as follows

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

which will be used in the Newton method discussed later. It is known that $X(x) \circ Z = \mu I$ is equivalent to the relation $X(x)Z = ZX(x) = \mu I$ for any $\mu \geq 0$. By using this multiplication, we also define the notation $r_S(w)$ by

$$r_S(w, \mu) = \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ X(x) \circ Z - \mu I \end{pmatrix}, \tag{6}$$

and we denote $r_S(w,0)$ by $r_{0S}(w)$.

For $U \in \mathbb{S}^p$, nonsingular $P \in \mathbb{R}^{p\times p}$ and $Q \in \mathbb{R}^{p\times p}$, we define the operator

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

and the symmetrized Kronecker product

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

where the operator svec is defined by

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

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

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

and

$$\|U\|_F = \|\mathrm{svec}(U)\|_2$$

hold.

In the following, $(v)_i$ denotes the i-th element of the vector v. Let $\{a_k\}$ and $\{b_k\}$ be sequences of vectors or matrices. If there exists a positive constant ξ_0 such that $\|a_k\| \le \xi_0\|b_k\|$ for all k and for some vector norm or some matrix norm, then we write $a_k = \mathrm{O}(\|b_k\|)$. If there exist positive constants ξ_1 and ξ_2 such that $\xi_1\|b_k\| \le \|a_k\| \le \xi_2\|b_k\|$ for all k, then we write $a_k = \Theta(\|b_k\|)$. If $\|a_k\| \to 0$, $\|b_k\| \to 0$ and $\|a_k\|/\|b_k\| \to 0$, we write $a_k = \mathrm{o}(\|b_k\|)$. For vectors v, v_1, v_2 and matrices G, G_1, G_2, if $v = v_1 + v_2$ with $\|v_2\| = \mathrm{O}(h)$ or $G = G_1 + G_2$ with $\|G_2\| = \mathrm{O}(h)$, we write $v = v_1 + \mathrm{O}(h)$ or $G = G_1 + \mathrm{O}(h)$ respectively.

3 Algorithm for finding a KKT point

In this section, we briefly describe a procedure for finding a KKT point by using the BKKT conditions (4) and (5). We define the norms $\|r(w,\mu)\|$ and $\|r_S(w,\mu)\|$ by

$$\|r(w,\mu)\| = \sqrt{\left\| \begin{pmatrix} \nabla_x L(w) \\ g(x) \end{pmatrix} \right\|_2^2 + \|X(x)Z - \mu I\|_F^2}$$

and

$$\|r_S(w,\mu)\| = \sqrt{\left\| \begin{pmatrix} \nabla_x L(w) \\ g(x) \end{pmatrix} \right\|_2^2 + \|X(x)\circ Z - \mu I\|_F^2},$$

respectively, where $\|\cdot\|_2$ denotes the l_2 norm for vectors and $\|\cdot\|_F$ denotes the Frobenius norm for matrices. We note that $\|r_S(w,\mu)\| \le \|r(w,\mu)\|$ is satisfied because of $\|X(x)\circ Z - \mu I\|_F \le \|X(x)Z - \mu I\|_F$. In what follows, we denote $X(x)$ simply by X if it is not confusing.

In the paper [23], the authors used the following algorithm SDPIP as an outer iteration for solving the nonlinear SDP problem (1).

Algorithm SDPIP

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

Step 1. (Termination) If $\|r_0(w_k)\| \le \varepsilon$, then stop.

Step 2. (Approximate BKKT point) Find an interior point w_{k+1} that satisfies the approximate BKKT condition

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

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

In Step 2 of Algorithm SDPIP, an approximate BKKT point can be found by applying the Newton-like method. As in the case of linear SDP problems, we define a scaling matrix $T \in \mathbb{R}^{p\times p}$ and scale the primal-dual pair $(X(x), Z)$ by

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

respectively. Let the Newton directions for the primal and dual variables by $\Delta x \in \mathbb{R}^n$ and $\Delta Z \in \mathbb{S}^p$, respectively, at the point w. We define $\Delta X = \sum_{i=1}^n \Delta x_i A_i(x)$ and note that $\Delta X \in \mathbb{S}^p$. We also scale ΔX and ΔZ by

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

Following [23], we consider the following scaled Newton equations

$$\nabla_x^2 L(w)\Delta x - A_0(x)^T\Delta y - \mathcal{A}^*(x)\Delta Z = -\nabla_x L(x, y, Z) \tag{7}$$

$$A_0(x)\Delta x = -g(x) \tag{8}$$

$$\frac{1}{2}(\Delta\widetilde{X}\widetilde{Z} + \widetilde{Z}\Delta\widetilde{X} + \widetilde{X}\Delta\widetilde{Z} + \Delta\widetilde{Z}\widetilde{X}) = \mu I - \frac{1}{2}(\widetilde{X}\widetilde{Z} + \widetilde{Z}\widetilde{X}). \tag{9}$$

We denote the Newton equations above by

$$\widetilde{J}_S(w)\Delta w = -\tilde{r}_S(w, \mu), \tag{10}$$

where $\widetilde{J}_S(w)$ is a linear operator from $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$ to $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$ and $\tilde{r}_S(w, \mu)$ is obtained from (6) by replacing $X \circ Z$ by $\widetilde{X} \circ \widetilde{Z}$. If we choose $T = I$, we call the above equations the unscaled Newton equations and use $J_S(w)$ instead of $\widetilde{J}_S(w)$ in this case.

By using the operator $\odot$ defined in Section 2, the matrices $\widetilde{X}$, $\widetilde{Z}$, $\Delta\widetilde{X}$ and $\Delta\widetilde{Z}$ can be represented by

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

We note that equation (9) can also rewritten by the expression

$$(\widetilde{Z} \odot I)\Delta\widetilde{X} + (\widetilde{X} \odot I)\Delta\widetilde{Z} = \mu I - \widetilde{X} \circ \widetilde{Z}.$$

Thus, by using the operator svec and the symmetrized Kronecker product, the Newton equations (7) – (9) are represented by the form

$$
\begin{pmatrix} \nabla_x^2 L(w) & -A_0(x)^T & -A(x)^T \\ A_0(x) & 0 & 0 \\ (\widetilde{Z} \otimes_S I)(T \otimes_S T)A(x) & 0 & (\widetilde{X} \otimes_S I)(T^{-T} \otimes_S T^{-T}) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \mathrm{svec}(\Delta Z) \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x,y,Z) \\ -g(x) \\ \mathrm{svec}(\mu I - \widetilde{X} \circ \widetilde{Z}) \end{pmatrix}, \tag{11}
$$

where

$$A(x) = [\mathrm{svec}(A_1(x)), \dots, \mathrm{svec}(A_n(x))] \in \mathbb{R}^{p(p+1)/2 \times n}.$$

We use the same notation $\widetilde{J}_S(w)$ for the coefficient matrix in (11) for convenience. In particular, we denote $\widetilde{J}_S(w)$ by $J_S(w)$ in case of $T = I$.

In [23], it is shown that the direction $\Delta\widetilde{Z} \in \mathbb{S}^p$ is given by the form

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

or equivalently

$$\Delta Z = \mu X^{-1} - Z - (T^T \odot T^T)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)\Delta X, \tag{12}$$

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

$$
\begin{pmatrix} \nabla_x^2 L(w) + H & -A_0(x)^T \\ -A_0(x) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} \nabla f(x) - A_0(x)^T y - \mu \mathcal{A}^*(x) X^{-1} \\ -g(x) \end{pmatrix},
$$

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

$$H_{ij} = \left\langle \widetilde{A}_i(x), (\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)\widetilde{A}_j(x) \right\rangle \tag{13}$$

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

In [23], the authors also proposed the primal-dual merit function

$$F(x, Z) = F_{BP}(x) + \nu F_{PD}(x, Z) \tag{14}$$

with

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

where ν and ρ are positive parameters and $\|g(x)\|_1$ denotes the l_1-norm of $g(x)$, and they proved the global convergence property within the line search strategy under the assumption that the scaling matrix T was chosen so that $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$ was satisfied.

In this paper, we are interested in the local behavior of the above Newton method. For this purpose, we consider the three kinds of choices of the scaling matrix T, which are given as follows:

Choices of T

(i) We first consider the choice $T = I$, which corresponds to the AHO direction for linear SDP problems [1]. We will discuss its superlinear convergence property in Section 5.

(ii) If we set $T = X^{-1/2}$, then we have $\widetilde{X} = I$ and $\widetilde{Z} = X^{1/2}ZX^{1/2}$, which corresponds to HRVW/KSH/M direction for linear SDP problems [7, 10, 12]. We will discuss its two-step superlinear convergence property in Section 6.1.

(iii) If we set $T = W^{-1/2}$ with $W = X^{1/2}(X^{1/2}ZX^{1/2})^{-1/2}X^{1/2}$, then we have $\widetilde{X} = W^{-1/2}XW^{-1/2} = W^{1/2}ZW^{1/2} = \widetilde{Z}$, which corresponds to the NT direction for linear SDP problems [13, 14]. We will discuss its two-step superlinear convergence property in Section 6.2.

4 Preliminaries for analysis of local behavior

In this section, we briefly present some definitions that are necessary for analysis of local behavior of our proposed methods.

First we introduce the definitions of the stationary point, the Mangasarian-Fromovitz constraint qualification condition, the quadratic growth condition, the strict complementarity condition and the nondegeneracy condition, and then we give the second order necessary / sufficient conditions for optimality. More comprehensive description can be found in [2, 16, 17].

A point x^* is said to be a stationary point of problem (1) if there exist Lagrange multipliers (y, Z) such that (x^*, y, Z) satisfies the KKT conditions (2) and (3). Let $\Lambda(x^*)$ denote the set of Lagrange multipliers (y, Z) such that (x^*, y, Z) satisfies the KKT conditions. We say that the Mangasarian-Fromovitz constraint qualification (MFCQ) condition holds at a point x^* if the matrix $A_0(x^*)$ is of full rank and there exists a nonzero vector $v \in \mathbb{R}^n$ such that

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

The second order necessary condition for local optimality of x^* under the MFCQ condition is given by

$$\sup_{(y,Z)\in\Lambda(x^*)} h^T(\nabla_x^2 L(x^*, y, Z) + \hat{H}(x^*, Z))h \geq 0$$

for all $h \in C(x^*)$. Here $\hat{H}(x, Z)$ is a matrix whose (i, j)-th element is

$$(\hat{H}(x, Z))_{ij} = 2\mathrm{tr}(A_i(x)X(x)^{\dagger}A_j(x)Z) \qquad (15)$$

and $\dagger$ denotes the Moore-Penrose generalized inverse, and $C(x^*)$ denotes the critical cone of (1) at x^*, which is defined by

$$C(x^*) = \left\{ h \;\middle|\; A_0(x^*)h = 0,\ \sum_{i=1}^{n} h_i A_i(x^*) \in T_{\mathbb{S}^p_+}(X(x^*)),\ \nabla f(x^*)^T h = 0 \right\},$$

and $T_{\mathbb{S}^p_+}(X(x^*))$ denotes the tangent cone of $\mathbb{S}^p$ at $X(x^*)$, which is defined by

$$T_{\mathbb{S}^p}(X(x^*)) = \{D \mid \text{dist}(X(x^*) + tD, \mathbb{S}^p_+) = o(t),\ t \geq 0\},$$

where $\text{dist}(P, \mathbb{S}^p_+) = \inf\{\|P - Q\|_F, Q \in \mathbb{S}^p_+\}$, and $\mathbb{S}^p_+$ denotes the set of p-th order symmetric positive semidefinite matrices.

It is said that the quadratic growth condition holds at a feasible point x^* of problem (1) if there exists $c > 0$ such that the following inequality holds

$$f(x) \geq f(x^*) + c\|x - x^*\|_2^2$$

for any feasible point x in a neighborhood of x^*. The quadratic growth condition implies that x^* is a strict local optimal solution of problem (1). Suppose that the MFCQ condition holds. Then the quadratic growth condition holds if and only if the following second order sufficient conditions for optimality are satisfied

$$\sup_{(y,Z)\in\Lambda(x^*)} h^T(\nabla_x^2 L(x^*, y, Z) + \hat{H}(x^*, Z))h > 0 \tag{16}$$

for all $h \in C(x^*)\backslash\{0\}$.

We say that the strict complementarity condition holds at x^* if there exists $(y^*, Z^*) \in \Lambda(x^*)$ such that

$$\text{rank}(X(x^*)) + \text{rank}(Z^*) = p$$

is satisfied. Since the matrices $X(x^*)$ and Z^* commute, they can be simultaneously diagonalized. Thus if the strict complementarity condition holds at x^*, we can assume without loss of generality that the matrix $X(x^*)$ and Z^* are represented by

$$X(x^*) = \begin{pmatrix} X_B^* & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad Z^* = \begin{pmatrix} 0 & 0 \\ 0 & Z_N^* \end{pmatrix} \tag{17}$$

respectively, where X_B^* and Z_N^* are diagonal and positive definite matrices with $\text{rank}(X_B^*) + \text{rank}(Z_N^*) = p$. Corresponding to (17), we partition the matrices $X(x)$ and Z as

$$X(x) = \begin{pmatrix} X_B & X_U \\ X_U^T & X_N \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} Z_B & Z_U \\ Z_U^T & Z_N \end{pmatrix}$$

in the neighborhood of $w^* = (x^*, y^*, Z^*)$. Similarly, we partition the matrix $A_i(x)$ as

$$A_i(x) = \begin{pmatrix} A_{Bi}(x) & A_{Ui}(x) \\ A_{Ui}(x)^T & A_{Ni}(x) \end{pmatrix}$$

for $i = 1, \ldots, n$. Then the critical cone at x^* can be specifically represented by

$$C(x^*) = \left\{ h \mid A_0(x^*)h = 0,\ \sum_{i=1}^n h_i A_{Ni}(x^*) = 0 \right\}.$$

We say that the nondegeneracy condition holds at x^* if the n dimensional vectors

$$\nabla g_i(x^*), i = 1, \ldots, m \quad \text{and} \quad \begin{pmatrix} (A_{N1}(x^*))_{ij} \\ \vdots \\ (A_{Nn}(x^*))_{ij} \end{pmatrix},\ i, j = 1, \ldots, |N|$$

are linearly independent, where $|N|$ denotes the size of Z_N^*. If the strict complementarity condition holds at x^*, then $\Lambda(x^*)$ is a singleton if and only if the nondegeneracy condition is satisfied. It is known that the nondegeneracy condition is stronger than the MFCQ condition, i.e., if the nondegeneracy condition holds at x^*, then the MFCQ condition also holds at x^*.

Throughout this paper, we make the following assumptions.

Assumptions

(A1) The second derivatives of the functions f, $g_i, i = 1, ..., m$, and X are Lipschitz continuous at x^*.

(A2) The second order sufficient condition (16) for optimality of problem (1) holds at x^*.

(A3) The strict complementarity condition holds at x^*.

(A4) The nondegeneracy condition is satisfied at x_*.

□

We note that the set $\Lambda(x^*)$ becomes a singleton, i.e., $\Lambda(x^*) = \{(y^*, Z^*)\}$, under assumptions (A3) and (A4). In the following, we denote a KKT point (x^*, y^*, Z^*) by w^*.

Under assumptions (A1)-(A4), we can show the nonsingularity of the matrix $J_S(w)$ at w^* as follows.

Theorem 1 *Suppose that assumptions* (A1)-(A4) *hold. Then the matrix* $J_S(w^*)$ *is nonsingular.*

Proof. We prove this theorem by showing that $J_S(w^*)\Delta w = 0$ implies $\Delta w = 0$ for $\Delta w = (\Delta x, \Delta y, \Delta Z)^T \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$ instead of showing that

$$J_S(w^*) \begin{pmatrix} \Delta x \\ \Delta y \\ \mathrm{svec}(\Delta Z) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that $(\Delta x, \Delta y, \mathrm{svec}(\Delta Z))^T = (0, 0, 0)^T$, because they are equivalent. For this purpose, we consider the linear system of equations

$$\nabla_x^2 L(w^*)\Delta x - A_0(x^*)^T \Delta y - \mathcal{A}^*(x^*)\Delta Z = 0 \tag{18}$$

$$A_0(x^*)\Delta x = 0 \tag{19}$$

$$\Delta X Z^* + Z^* \Delta X + X^* \Delta Z + \Delta Z X^* = 0, \tag{20}$$

where $\Delta X = \sum_{i=1}^{n} (\Delta x)_i A_i(x^*)$. Following (17), we define diagonal and positive definite matrices X_B^* and Z_N^*, and we denote ΔX and ΔZ by

$$\Delta X = \begin{pmatrix} \Delta X_B & \Delta X_U \\ \Delta X_U^T & \Delta X_N \end{pmatrix} \quad \text{and} \quad \Delta Z = \begin{pmatrix} \Delta Z_B & \Delta Z_U \\ \Delta Z_U^T & \Delta Z_N \end{pmatrix}$$

Then equation (20) can be written by the form

$$\begin{pmatrix} X_B^*\Delta Z_B + \Delta Z_B X_B^* & \Delta X_U Z_N^* + X_B^* \Delta Z_U \\ Z_N^* \Delta X_U^T + \Delta Z_U^T X_B^* & \Delta X_N Z_N^* + Z_N^* \Delta X_N \end{pmatrix} = 0. \tag{21}$$

Since

$$(X_B^*)^{-1}\Delta Z_B X_B^* = -\Delta Z_B = -\Delta Z_B^T = X_B^* \Delta Z_B (X_B^*)^{-1},$$

we have

$$\Delta Z_B (X_B^*)^2 = (X_B^*)^2 \Delta Z_B,$$

which implies that $\Delta Z_B X_B^* = X_B^* \Delta Z_B$. Thus the (1,1) block of equation (21) yields $\Delta Z_B = 0$. Similarly we have $\Delta X_N = 0$ from the (2,2) block of (21), which implies that $\sum_{i=1}^{n} (\Delta x)_i A_{Ni}(x^*) = 0$. Since $A_0(x^*)\Delta x = 0$ is satisfied, we have $\Delta x \in C(x^*)$.

Furthermore by the (1,2) block of (21), we obtain

$$\Delta Z_U = -(X_B^*)^{-1} \Delta X_U Z_N^*. \tag{22}$$

By premultiplying (18) by Δx^T and using (19), we have

$$\Delta x^T \nabla_x^2 L(w^*) \Delta x - \Delta x^T \mathcal{A}^*(x^*) \Delta Z = 0 \tag{23}$$

Since the following relations hold

$$\begin{aligned} \Delta x^T \mathcal{A}^*(x^*) \Delta Z &= \mathrm{tr}(\Delta X \Delta Z) \\ &= \mathrm{tr} \begin{pmatrix} \Delta X_B & \Delta X_U \\ \Delta X_U^T & 0 \end{pmatrix} \begin{pmatrix} 0 & \Delta Z_U \\ \Delta Z_U^T & \Delta Z_N \end{pmatrix} \\ &= 2\mathrm{tr}(\Delta X_U \Delta Z_U^T), \end{aligned}$$

equation (22) implies

$$\Delta x^T \mathcal{A}^*(x^*) \Delta Z = -2\mathrm{tr}(\Delta X_U Z_N^* \Delta X_U^T (X_B^*)^{-1}).$$

On the other hand, the definition of $\hat{H}(x, Z)$ in (15) gives

$$\begin{aligned} \Delta x^T \hat{H}(x^*, Z^*) \Delta x &= 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{tr}(A_i(x^*) X(x^*)^\dagger A_j(x^*) Z^*)(\Delta x)_i (\Delta x)_j \\ &= 2\mathrm{tr}(\Delta X X(x^*)^\dagger \Delta X Z^*) \\ &= 2\mathrm{tr} \begin{pmatrix} 0 & \Delta X_B (X_B^*)^{-1} \Delta X_U Z_N^* \\ 0 & \Delta X_U^T (X_B^*)^{-1} \Delta X_U Z_N^* \end{pmatrix} \\ &= 2\mathrm{tr}(\Delta X_U Z_N^* \Delta X_U^T (X_B^*)^{-1}). \end{aligned}$$

Then equation (23) yields

$$\Delta x^T \left(\nabla_x^2 L(w^*) + \hat{H}(x^*, Z^*) \right) \Delta x = 0.$$

Since $\Delta x \in C(x^*)$, the second order sufficient condition (16) yields $\Delta x = 0$, which implies $\Delta Z_U = 0$. By (18), we have

$$A_0(x^*)^T \Delta y + \mathcal{A}^*(x^*) \begin{pmatrix} 0 & 0 \\ 0 & \Delta Z_N \end{pmatrix} = 0,$$

which implies that

$$\sum_{i=1}^{m} (\Delta y)_i \nabla g_i(x^*) + \sum_{i,j=1}^{|N|} (\Delta Z_N)_{ji} \begin{pmatrix} (A_{N1}(x^*))_{ij} \\ \vdots \\ (A_{Nn}(x^*))_{ij} \end{pmatrix} = 0,$$

because the l -th element of the vector $\mathcal{A}^*(x^*) \begin{pmatrix} 0 & 0 \\ 0 & \Delta Z_N \end{pmatrix}$ is given by $\mathrm{tr}(A_{Nl}(x^*)\Delta Z_N) = \sum_{i,j=1}^{|N|} (A_{Nl}(x^*))_{ij} (\Delta Z_N)_{ji}$. Thus the nondegeneracy condition yields $\Delta y = 0$ and $\Delta Z_N = 0$. Therefore we obtain $(\Delta x, \Delta y, \Delta Z) = (0, 0, 0)$, and then we prove the theorem. □

In the following, we will discuss local behavior of the unsymmetric residual $r_0(w)$ in (2) or $r(w, \mu)$ in (4). For this purpose, we define a linear operator $J : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{p\times p}$ at w by

$$J(w)\Delta w = \begin{pmatrix} \nabla_x^2 L(w)\Delta x - A_0(x)^T \Delta y - \mathcal{A}^*(x)\Delta Z \\ A_0(x)\Delta x \\ \Delta X Z + X \Delta Z \end{pmatrix}$$

for $\Delta w = (\Delta x, \Delta y, \Delta Z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$, which is an estimate of the first order change of $r_0(w + \Delta w)$ or $r(w + \Delta w, \mu)$. We note that $J(w)\Delta w$ can be represented by the matrix-vector form:

$$(24) \quad J(w)\Delta w = \begin{pmatrix} \nabla_x^2 L(w) & -A_0(x)^T & -A(x)^T \\ A_0(x) & 0 & 0 \\ (Z \otimes I)M^T A(x) & 0 & (I \otimes X)M^T \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \mathrm{svec}(\Delta Z) \end{pmatrix},$$

where $Z \otimes I \in \mathbb{R}^{p^2 \times p^2}$ and $I \otimes X \in \mathbb{R}^{p^2 \times p^2}$ denote the Kronecker products of Z and I, and I and X, respectively, and M is an $p(p+1) \times p^2$ matrix such that $M\mathrm{vec}(U) = \mathrm{svec}(U)$ and $M^T \mathrm{svec}(U) = \mathrm{vec}(U)$ hold for all $U \in \mathbb{S}^p$ (see Appendix of [20]). Here the operator vec is defined by

$$\mathrm{vec}(U) = (U_{11}, U_{21}, \ldots, U_{p1}, U_{12}, \ldots, U_{pp})^T \in \mathbb{R}^{p^2}.$$

We also use the same notation $J(w)$ for the rectangular coefficient matrix in (24) for convenience.

In the same way as the proof of the preceding theorem, we can show the nonsingularity of the linear operator $J(w)$ at w^*.

Corollary 1 *Suppose that assumptions* (A1)-(A4) *hold. Then the matrix* $J(w^*)$ *is left invertible.*

We note that the related analysis can be found in [3] and [18].

The following lemma will be a useful tool in the subsequent sections.

Lemma 1 *Suppose that assumptions* (A1)-(A4) *hold and that w is sufficiently close to w^*. Let μ be zero or a sufficiently small positive number. Then there exists a continuously differentiable function $\bar{w}(\mu) = (\bar{x}(\mu), \bar{y}(\mu), \bar{Z}(\mu))$ such that*

$$\bar{w}(0) = w^*, \quad r(\bar{w}(\mu), \mu) = r_S(\bar{w}(\mu), \mu) = 0 \quad \text{for} \quad \mu \geq 0, \tag{25}$$

and

$$\bar{X}(\mu) \succ 0 \quad \text{and} \quad \bar{Z}(\mu) \succ 0 \quad \text{for} \quad \mu > 0, \tag{26}$$

where $\bar{X}(\mu) = \sum_{i=1}^{n} (\bar{x}(\mu))_i A_i(\bar{x}(\mu))$.

Furthermore, if w is sufficiently close to $\bar{w}(\mu)$, then the following relation holds

$$r(w, \mu) = \Theta(\|w - \bar{w}(\mu)\|) \quad \text{and} \quad r_S(w, \mu) = \Theta(\|w - \bar{w}(\mu)\|) \quad \text{for} \quad \mu \geq 0. \tag{27}$$

Proof. Since $J_S(w^*)$ is nonsingular by Theorem 1, the implicit function theorem and assumption (A1) guarantee (25), and $J_S(\bar{w}(\mu))$ is nonsingular. Furthermore, the facts $\bar{X}(\mu)\bar{Z}(\mu) = \mu I$, $\bar{X}(0) = X(x^*)$ and $\bar{Z}(0) = Z^*$ guarantee (26), where $X(x^*)$ and Z^* are defined in (17).

It follows that

$$\begin{aligned} r_S(w, \mu) &= r_S(\bar{w}(\mu), \mu) + J_S(\bar{w}(\mu))(w - \bar{w}(\mu)) + \mathrm{O}(\|w - \bar{w}(\mu)\|^2) \\ &= J_S(\bar{w}(\mu))(w - \bar{w}(\mu)) + \mathrm{O}(\|w - \bar{w}(\mu)\|^2), \end{aligned}$$

and then the nonsingularity of $J_S(\bar{w}(\mu))$ guarantees $r_S(w, \mu) = \Theta(\|w - \bar{w}(\mu)\|)$. Similarly we obtain $r(w, \mu) = \Theta(\|w - \bar{w}(\mu)\|)$.

Therefore the proof is complete. □

We note that the preceding lemma also implies $r_0(w) = \Theta(\|r_{0S}(w)\|)$.

5 Superlinear convergence of unscaled Newton method

In this section, we consider the local behavior of the unscaled Newton method, which is the case $T_k = I$. Then the Newton equations (10) can be represented by

$$J_S(w)\Delta w = -r_S(w, \mu). \tag{28}$$

In the following, we present our algorithm and show its superlinear convergence property.

Algorithm unscaledSDPIP

Step 0. (Initialize) Set $\varepsilon > 0$ and $0 < \tau < 1$. Choose $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$ $(X(x_0) \succ 0, Z_0 \succ 0)$. Set $k = 0$.

Step 1. (Termination) If $\|r_0(w_k)\| \le \varepsilon$, then stop.

Step 2. (Newton step) Choose a barrier parameter μ_k such that

$$\mu_k = \xi_k \|r_0(w_k)\|^{1+\tau} \tag{29}$$

with $\xi_k = \Theta(1)$. Calculate the direction Δw_k by solving the Newton equations (28). Set $w_{k+1} = w_k + \Delta w_k$.

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

By Theorem 1, if the iterate w_k is sufficiently close to w^*, the Jacobian matrix $J_S(w_k)$ is nonsingular and its inverse is uniformly bounded. Thus the Newton equations have a unique solution and the following relations hold

$$\Delta w_k = \Theta(\|r_S(w_k, \mu_k)\|) = \mathrm{O}(\|r_{0S}(w_k)\|) + \mathrm{O}(\mu_k) = \mathrm{O}(\|r_0(w_k)\|), \tag{30}$$

where the last equality can be obtained by equation (29).

We give a lemma which plays an important role in showing superlinear convergence property of Algorithm unscaledSDPIP.

Lemma 2 *Suppose that assumptions* (A1)-(A4) *hold. Assume that w is an interior point which is sufficiently close to w^* and satisfies the approximate BKKT condition $\|r(w, \mu_-)\| \le M_c \mu_-$ for a given positive number μ_-, where M_c is a constant satisfying $0 < M_c < 1$. Let μ be a positive number defined by*

$$\mu = \xi \|r_0(w)\|^{1+\tau}$$

with $\xi = \Theta(1)$, where τ is a constant satisfying $0 < \tau < 1$. If Δw satisfies the Newton equations (28)*, then the new iterate $w + \Delta w$ satisfies*

$$\|r(w + \Delta w, \mu)\| \le M_c \mu, \quad X(x + \Delta x) \succ 0 \quad \text{and} \quad Z + \Delta Z \succ 0. \tag{31}$$

Proof. Let the eigenvalues of the matrix $X(x+\alpha\Delta x) \circ (Z + \alpha \Delta Z)$ be $\lambda_1(\alpha) \le \ldots \le \lambda_p(\alpha)$ for any $\alpha \in [0,1]$. Since $\Delta X = \mathrm{O}(\|r_0(w)\|)$ and $\Delta Z = \mathrm{O}(\|r_0(w)\|)$ hold by (30), we have

$$\begin{aligned}
X(x+\alpha\Delta x) \circ (Z + \alpha\Delta Z) &= (X(x) + \alpha \Delta X + \alpha^2 \mathrm{O}(\|r_0(w)\|^2)) \circ (Z + \alpha \Delta Z) \\
&= X(x) \circ Z + \alpha(\Delta X \circ Z + X(x) \circ \Delta Z) + \alpha^2 \mathrm{O}(\|r_0(w)\|^2) \\
&= X(x) \circ Z + \alpha(\mu I - X(x) \circ Z) + \alpha^2 \mathrm{O}(\|r_0(w)\|^2) \\
&= (1-\alpha) X(x) \circ Z + \alpha \mu I + \alpha^2 \mathrm{O}(\|r_0(w)\|^2).
\end{aligned}$$

Thus we have that

$$\begin{aligned}
&\|X(x+\alpha\Delta x) \circ (Z + \alpha\Delta Z) - ((1-\alpha)\mu_- + \alpha\mu) I\|_F \\
&\le (1-\alpha)\|X(x) \circ Z - \mu_- I\|_F + \alpha^2 \mathrm{O}(\|r_0(w)\|^2) \\
&\le (1-\alpha)\|X(x) Z - \mu_- I\|_F + \alpha^2 \mathrm{O}(\|r_0(w)\|^2) \\
&\le (1-\alpha) M_c \mu_- + \alpha^2 \mathrm{O}(\|r_0(w)\|^2) \\
&\le M_c((1-\alpha)\mu_- + \alpha\mu). \qquad (32)
\end{aligned}$$

The last inequality follows from the definition of μ. By combining (32) and the following relation

$$\|X(x+\alpha\Delta x)\circ(Z+\alpha\Delta Z)-((1-\alpha)\mu_-+\alpha\mu)I\|_F^2=\sum_{i=1}^{p}(\lambda_i(\alpha)-((1-\alpha)\mu_-+\alpha\mu))^2,$$

we have

$$(\lambda_i(\alpha)-((1-\alpha)\mu_-+\alpha\mu))^2\le M_c^2((1-\alpha)\mu_-+\alpha\mu)^2 \quad \text{for} \quad i=1,\ldots,p.$$

Then we obtain

$$0<(1-M_c)((1-\alpha)\mu_-+\alpha\mu)\le\lambda_i(\alpha) \quad \text{for} \quad i=1,\ldots,p.$$

Thus the matrix $X(x+\alpha\Delta x)\circ(Z+\alpha\Delta Z)$ is symmetric positive definite for all $\alpha\in[0,1]$. Since the matrices $X(x)$ and Z are symmetric positive definite, the above results imply that the matrices $X(x+\alpha\Delta x)$ and $Z+\alpha\Delta Z$ are also symmetric positive definite for all $\alpha\in[0,1]$. This guarantees that $w+\Delta w$ is an interior point.

It follows from the Newton equation and equation (30) that

$$\begin{aligned}\|r_S(w+\Delta w,\mu)\| &= \Theta(\|r_S(w,\mu)+J_S(w)\Delta w+\mathrm{O}(\|\Delta w\|^2)\|)\\ &= \mathrm{O}(\|\Delta w\|^2)\\ &= \mathrm{O}(\|r_0(w)\|^2).\end{aligned}$$

Thus Lemma 1 yields

$$\begin{aligned}\|r(w+\Delta w,\mu)\| &= \mathrm{O}(\|r_0(w)\|^2)\\ &= \mathrm{o}(\|r_0(w)\|^{1+\tau})\\ &= \mathrm{o}(\mu)\\ &\le M_c\mu,\end{aligned}$$

which proves (31).

Therefore the proof of this theorem is complete. □

We note that in the previous lemma, a positive number μ_- can be arbitrarily chosen.

Now we show the superlinear convergence of Algorithm unscaledSDPIP in the following theorem.

Theorem 2 *Suppose that assumptions* (A1)-(A4) *hold. Assume that an initial interior point w_0 is sufficiently close to w^* such that the approximate BKKT condition $\|r(w_0,\mu_{-1})\|\le M_c\mu_{-1}$ is satisfied for given $\mu_{-1}>0$ and $0<M_c<1$. Then the sequence $\{w_k\}$ generated by Algorithm unscaledSDPIP satisfies*

$$\|r(w_k,\mu_{k-1})\|\le M_c\mu_{k-1}, \quad X(x_k)\succ 0 \quad \text{and} \quad Z_k\succ 0 \tag{33}$$

for all $k\ge 0$ and converges locally and superlinearly to w^*.

Proof. To prove this theorem by the mathematical induction, we assume that (33) holds at w_k. Then it follows directly from Lemma 2 that the next point w_{k+1} also satisfies (33). Thus we have

$$\|r_0(w_{k+1})\| = \left\| r(w_{k+1}, \mu_k) + \begin{pmatrix} 0 \\ 0 \\ \mu_k I \end{pmatrix} \right\| \leq (M_c + \sqrt{n})\mu_k.$$

Similarly we have

$$\|r_0(w_{k+1})\| \geq \left\| \begin{pmatrix} 0 \\ 0 \\ \mu_k I \end{pmatrix} \right\| - \|r(w_{k+1}, \mu_k)\| \geq (\sqrt{n} - M_c)\mu_k.$$

The above two inequalities and (29) imply that

$$\|r_0(w_{k+1})\| = \Theta(\|r_0(w_k)\|^{1+\tau}).$$

It follows from (27) and (30) that if w_k is sufficiently close to w^*, then the following hold

$$\begin{aligned} \|w_{k+1} - w^*\| &\leq \|w_k - w^*\| + \|\Delta w_k\| \\ &= \|w_k - w^*\| + \mathrm{O}(\|r_0(w_k)\|) \\ &= \mathrm{O}(\|w_k - w^*\|). \end{aligned}$$

Thus w_{k+1} is also sufficiently close to w^*, and we obtain by (27)

$$\|w_{k+1} - w^*\| = \Theta(\|r_0(w_{k+1})\|) = \Theta(\|r_0(w_k)\|^{1+\tau}) = \Theta(\|w_k - w^*\|^{1+\tau}).$$

Therefore the local and superlinear convergence property is proved. □

6 Two-step superlinear convergence of scaled Newton method

In this section, we discuss local and superlinear convergence properties of interior point methods that use the scaled Newton equations. Specifically we show local and two-step superlinear convergence properties of two kinds of primal-dual interior point methods which use the HRVW/KSH/M and the NT directions.

We first prove the following lemma that estimates the inverse matrices of $X(x)$ and Z.

Lemma 3 *Suppose that assumptions* (A1) – (A4) *hold and that* w *is an interior point which is sufficiently close to* w^*. *Assume that* $\|r(w, \mu)\| = \mathrm{o}(\mu)$ *is satisfied for a positive number* μ. *Then the following relations hold*

$$X(x) = \begin{pmatrix} X_B & X_U \\ X_U^T & X_N \end{pmatrix} = \begin{pmatrix} \Theta(1) & \mathrm{O}(\mu) \\ \mathrm{O}(\mu) & \Theta(\mu) \end{pmatrix},$$

$$Z = \begin{pmatrix} Z_B & Z_U \\ Z_U^T & Z_N \end{pmatrix} = \begin{pmatrix} \Theta(\mu) & \mathrm{O}(\mu) \\ \mathrm{O}(\mu) & \Theta(1) \end{pmatrix},$$

$$X(x)^{-1} = \begin{pmatrix} \Theta(1) & \mathrm{O}(1) \\ \mathrm{O}(1) & \Theta(\mu^{-1}) \end{pmatrix} = \mathrm{O}(\mu^{-1}) \quad \text{and} \quad Z^{-1} = \begin{pmatrix} \Theta(\mu^{-1}) & \mathrm{O}(1) \\ \mathrm{O}(1) & \Theta(1) \end{pmatrix} = \mathrm{O}(\mu^{-1}).$$

Proof. Since $X(x)$ and Z are sufficiently close to $X(x^*) = \begin{pmatrix} X_B^* & 0 \\ 0 & 0 \end{pmatrix}$ and $Z^* = \begin{pmatrix} 0 & 0 \\ 0 & Z_N^* \end{pmatrix}$, respectively, it is clear that $X_B = \Theta(1)$ and $Z_N = \Theta(1)$. Since the following hold

$$\begin{aligned} w - w^* &= J(w^*)^{-1} r_0(w) + \mathrm{O}(\|w - w^*\|^2) \\ &= \mathrm{O}(\|r(w,\mu)\|) + \mathrm{O}(\mu) + \mathrm{O}(\|w - w^*\|^2) \\ &= \mathrm{O}(\mu) + \mathrm{O}(\|w - w^*\|^2), \end{aligned}$$

we have

$$w - w^* = \mathrm{O}(\mu),$$

and then we obtain

$$X(x) = \begin{pmatrix} \Theta(1) & \mathrm{O}(\mu) \\ \mathrm{O}(\mu) & \mathrm{O}(\mu) \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} \mathrm{O}(\mu) & \mathrm{O}(\mu) \\ \mathrm{O}(\mu) & \Theta(1) \end{pmatrix}.$$

It follows from the relation $r(w,\mu) = \mathrm{o}(\mu)$ that

$$X_B Z_B + X_U Z_U^T - \mu I = \mathrm{o}(\mu),$$

which yields

$$X_B Z_B = \mu I + \mathrm{o}(\mu).$$

Thus we obtain

$$Z_B = \mu X_B^{-1} + \mathrm{o}(\mu) = \Theta(\mu).$$

Similarly we have

$$X_N = \Theta(\mu).$$

Therefore we obtain

$$X(x) = \begin{pmatrix} \Theta(1) & \mathrm{O}(\mu) \\ \mathrm{O}(\mu) & \Theta(\mu) \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} \Theta(\mu) & \mathrm{O}(\mu) \\ \mathrm{O}(\mu) & \Theta(1) \end{pmatrix}.$$

Next we estimate the inverse matrices $X(x)^{-1}$ and Z^{-1}. Setting

$$R = X_N - X_U^T X_B^{-1} X_U,$$

we have

$$X(x)^{-1} = \begin{pmatrix} X_B^{-1} + X_B^{-1} X_U R^{-1} X_U^T X_B^{-1} & -X_B^{-1} X_U R^{-1} \\ -R^{-1} X_U^T X_B^{-1} & R^{-1} \end{pmatrix}.$$

Noting that $R = \Theta(\mu) + \Theta(1)\mathrm{O}(\mu^2) = \Theta(\mu)$ and then $R^{-1} = \Theta(\mu^{-1})$, we obtain

$$X(x)^{-1} = \begin{pmatrix} \Theta(1) + \mathrm{O}(\mu^2)\Theta(\mu^{-1}) & \Theta(\mu^{-1})\mathrm{O}(\mu) \\ \Theta(\mu^{-1})\mathrm{O}(\mu) & \Theta(\mu^{-1}) \end{pmatrix} = \begin{pmatrix} \Theta(1) & \mathrm{O}(1) \\ \mathrm{O}(1) & \Theta(\mu^{-1}) \end{pmatrix} = \mathrm{O}(\mu^{-1}).$$

Similarly we have

$$Z^{-1} = \begin{pmatrix} \Theta(\mu^{-1}) & \mathrm{O}(1) \\ \mathrm{O}(1) & \Theta(1) \end{pmatrix} = \mathrm{O}(\mu^{-1}).$$

Therefore the proof is complete. □

In the following, we present the algorithm called scaledSDPIP which calculates a KKT point by using the scaled Newton method.

Algorithm scaledSDPIP

Step 0. (Initialize) Set $\varepsilon > 0$ and $0 < \tau < 1$. Choose $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$ $(X(x_0) \succ 0, Z_0 \succ 0)$. Set $k = 0$.

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

Step 2. (Scaled Newton steps)

Step 2.1 Choose $\mu_k = \xi_k \|r_0(w_k)\|^{1+\tau}$ with $\xi_k = \Theta(1)$.

Step 2.2 Calculate the direction Δw_k by solving the scaled Newton equations $\widetilde{J}_S(w_k)\Delta w_k = -\tilde{r}_S(w_k, \mu_k)$ at w_k. Set $w_{k+\frac{1}{2}} = w_k + \Delta w_k$.

Step 2.3 Calculate the direction $\Delta w_{k+\frac{1}{2}}$ by solving the scaled Newton equations $\widetilde{J}_S(w_{k+\frac{1}{2}})\Delta w_{k+\frac{1}{2}} = -\tilde{r}_S(w_{k+\frac{1}{2}}, \mu_k)$ at $w_{k+\frac{1}{2}}$. Set $w_{k+1} = w_{k+\frac{1}{2}} + \Delta w_{k+\frac{1}{2}}$.

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

Now we prove two-step superlinear convergence of Algorithm scaledSDPIP. In the following, we will consider two kinds of scaled Newton methods. In Section 6.1, we first deal with the scaled Newton method with $T_k = X_k^{-1/2}$ (HRVW/KSH/M direction), and then in Section 6.2, we deal with the scaled Newton method with $T_k = W_k^{-1/2}$ (NT direction).

6.1 Scaled Newton method with $T_k = X_k^{-1/2}$

For the choice of $T_k = X_k^{-1/2}$, we have

$$\widetilde{X}_k = I, \qquad \widetilde{Z}_k = X_k^{1/2} Z_k X_k^{1/2}$$

and (12) and (13) reduce to

$$\Delta Z_k = \mu_k X_k^{-1} - Z_k - \frac{1}{2}(X_k^{-1}\Delta X_k Z_k + Z_k \Delta X_k X_k^{-1}). \tag{34}$$

and

$$(H_k)_{ij} = \mathrm{tr}\left(A_i(x_k) X_k^{-1} A_j(x_k) Z_k\right).$$

The following lemma estimates the Newton step Δw_k near the solution w^*.

Lemma 4 *Suppose that assumptions* (A1)–(A4) *hold. Let τ' be a positive constant. Suppose that w is an interior point which is sufficiently close to w^* and that $r(w, \mu_-) = \mathrm{O}(\mu_-^{1+\tau'})$ is satisfied for a positive number μ_-. Let μ be a positive number. Then the Newton step from $\tilde{J}_S(w)\Delta w = -\tilde{r}_S(w, \mu)$ satisfies*

$$\Delta w = \mathrm{O}(\|r(w, \mu)\|).$$

Proof. By letting $E = XZ - \mu_- I$, we have

$$XZ\Delta X X^{-1} = \Delta X Z + E\Delta X X^{-1} - \Delta X X^{-1} E.$$

Thus equation (34) yields

$$\begin{aligned} X\Delta Z &= \mu I - XZ - \frac{1}{2}(\Delta X Z + XZ\Delta X X^{-1}) \\ &= \mu I - XZ - \Delta X Z - \frac{1}{2}(E\Delta X X^{-1} - \Delta X X^{-1} E), \end{aligned}$$

which implies that

$$X\Delta Z + \Delta X Z = \mu I - XZ - \frac{1}{2}(E\Delta X X^{-1} - \Delta X X^{-1} E). \tag{35}$$

By transposing the matrices in the both sides above, we have

$$Z\Delta X + \Delta Z X = \mu I - ZX - \frac{1}{2}(X^{-1}\Delta X E^T - (X^{-1}E)^T \Delta X).$$

By adding the above two equations, we obtain

$$\begin{aligned} &X\Delta Z + \Delta X Z + Z\Delta X + \Delta Z X \\ &\quad = 2\mu I - (XZ + ZX) - \frac{1}{2}(E\Delta X X^{-1} + X^{-1}\Delta X E^T) + \frac{1}{2}(\Delta X X^{-1} E + (X^{-1}E)^T \Delta X)), \end{aligned}$$

which implies that

$$\begin{aligned} &X\Delta Z + \Delta X Z + Z\Delta X + \Delta Z X + (E \odot X^{-1})\Delta X - (I \odot X^{-1}E)\Delta X \\ &\quad = 2\mu I - (XZ + ZX). \end{aligned} \tag{36}$$

We write equations (7), (8) and (36) by

$$J_S'(w)\Delta w = -r_S(w, \mu). \tag{37}$$

We note that any solution to the Newton equations $\tilde{J}_S(w)\Delta w = -\tilde{r}_S(w, \mu)$ satisfies the linear system of equations (37).

Now we prove the nonsingularity of $J_S'(w)$. Since equation (36) implies

$$J_S'(w) - J_S(w) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ (E \otimes_S X^{-1} - I \otimes_S X^{-1}E)A(x) & 0 & 0 \end{pmatrix},$$

we have

$$\|J'_S(w) - J_S(w)\|_F \leq \|(E \otimes_S X^{-1})A(x)\|_F + \|(I \otimes_S X^{-1}E)A(x)\|_F.$$

Since Lemma 3 and the definition of E imply $X^{-1} = \mathrm{O}(\mu_-^{-1})$ and $E = \mathrm{O}(\mu_-^{1+\tau'})$, and each $A_i(x)$ is bounded, we have

$$\begin{aligned}
\|(E \otimes_S X^{-1})A(x)\|_F &\leq \|E \otimes_S X^{-1}\|_F \|A(x)\|_F \\
&= \mathrm{O}(\|E \otimes X^{-1} + X^{-1} \otimes E\|_F) \\
&= \mathrm{O}(\|E\|_F)\|\mathrm{O}(\|X^{-1}\|_F) \\
&= \mathrm{O}(\mu_-^{1+\tau'})\mathrm{O}(\mu_-^{-1}) \\
&= \mathrm{O}(\mu_-^{\tau'}).
\end{aligned}$$

Similarly we have

$$\|(I \otimes_S X^{-1}E)A(x)\|_F = \mathrm{O}(\mu_-^{\tau'}).$$

Thus it follows from the inequalities above that

$$\|J'_S(w) - J_S(w)\|_F = \mathrm{O}(\mu_-^{\tau'}).$$

Since w is sufficiently close to w^*, the matrix $J_S(w)$ is nonsingular and its inverse matrix is uniformly bounded, so is the matrix $J'_S(w)$. Thus equation (37) guarantees that $\Delta w = \Theta(\|r_S(w,\mu)\|) = \mathrm{O}(\|r(w,\mu)\|)$ hold. Therefore the lemma is proved. □

We give the following theorem, which plays an important role in showing superlinear convergence property of Algorithm scaledSDPIP.

Lemma 5 *Suppose that assumptions* (A1)-(A4) *hold and that w is an interior point which is sufficiently close to w^*. Let M_c be a positive constant, and let τ and τ' be positive constants that satisfy*

$$1 > \tau' > \tau \quad \text{and} \quad \tau' > \frac{2\tau}{1-\tau}.$$

Let μ_- be a given number that satisfies

$$\left(\frac{1}{2M_c}\right)^{1/\tau'} \geq \mu_- > 0. \tag{38}$$

Assume that w satisfies the approximate BKKT condition

$$\|r(w,\mu_-)\| \leq M_c \mu_-^{1+\tau'}. \tag{39}$$

Let μ be a positive number defined by

$$\mu = \xi \|r_0(w)\|^{1+\tau} \tag{40}$$

with $\xi = \Theta(1)$. If Δw is obtained by solving the scaled Newton equations $\widetilde{J}_S(w)\Delta w = -\tilde{r}_S(w,\mu)$, then the iterate $w_{\frac{1}{2}} = w + \Delta w$ satisfies

$$r(w_{\frac{1}{2}},\mu) = \mathrm{O}(\mu^{1+\frac{\tau'-\tau}{1+\tau}}), \quad X(x_{\frac{1}{2}}) \succ 0 \quad \text{and} \quad Z_{\frac{1}{2}} \succ 0.$$

Furthermore, if $\Delta w_{\frac{1}{2}}$ *is obtained by solving the scaled Newton equations* $\widetilde{J}_S(w_{\frac{1}{2}})\Delta w_{\frac{1}{2}} = -\tilde{r}_S(w_{\frac{1}{2}}, \mu)$*, then the iterate* $w_+ = w_{\frac{1}{2}} + \Delta w_{\frac{1}{2}}$ *satisfies*

$$\|r(w_+, \mu)\| \le M_c \mu^{1+\tau'}, \quad X(w_+) \succ 0 \quad \text{and} \quad Z_+ \succ 0. \tag{41}$$

Proof. We first note that condition (39) yields $r_0(w) = \Theta(\mu_-)$. We let the eigenvalues of the matrix $X(x + \alpha\Delta x) \circ (Z + \alpha\Delta Z)$ be $\lambda_1(\alpha) \le \ldots \le \lambda_p(\alpha)$ for each $\alpha \in [0, 1]$. Since $\|X \circ Z - \mu_- I\|_F \le \|XZ - \mu_- I\|_F \le M_c \mu_-^{1+\tau'}$, we have

$$(\lambda_i(0) - \mu_-)^2 \le \sum_{j=1}^{p} (\lambda_j(0) - \mu_-)^2 \le (M_c \mu_-^{1+\tau'})^2,$$

which implies by (38)

$$\lambda_i(0) \ge \mu_- - M_c \mu_-^{1+\tau'} \ge \frac{1}{2}\mu_- > 0, \quad i = 1, \ldots, p. \tag{42}$$

Let $E = XZ - \mu_- I$. Then condition (39) and Lemma 3 guarantee

$$E = \mathrm{O}(\mu_-^{1+\tau'}), \quad X^{-1} = \mathrm{O}(\mu_-^{-1})$$

and Lemma 4 and (40) imply

$$\Delta X = \mathrm{O}(\|r(w, \mu)\|) = \mathrm{O}(\|r_0(w)\|) + \mathrm{O}(\mu) = \mathrm{O}(\|r_0(w)\|) = \mathrm{O}(\mu_-).$$

Similarly we have

$$\Delta Z = \mathrm{O}(\mu_-).$$

Since equation (35) yields

$$\begin{aligned} & X(x + \alpha\Delta x)(Z + \alpha\Delta Z) \\ & \quad = (X(x) + \alpha\Delta X + \alpha^2 \mathrm{O}(\mu_-^2))(Z + \alpha\Delta Z) \\ & \quad = X(x)Z + \alpha(\Delta X Z + X(x)\Delta Z) + \alpha^2 \mathrm{O}(\mu_-^2) \\ & \quad = X(x)Z + \alpha(\mu I - X(x)Z) + \alpha \mathrm{O}(\|E\|_F)\mathrm{O}(\|X^{-1}\|_F)\mathrm{O}(\|\Delta X\|_F) + \alpha^2 \mathrm{O}(\mu_-^2) \\ & \quad = (1 - \alpha)X(x)Z + \alpha\mu I + \alpha \mathrm{O}(\mu_-^{1+\tau'}) + \alpha^2 \mathrm{O}(\mu_-^2) \\ & \quad = (1 - \alpha)X(x)Z + \alpha\mu I + \alpha \mathrm{O}(\mu_-^{1+\tau'}), \end{aligned} \tag{43}$$

we have

$$\|X(x + \alpha\Delta x) \circ (Z + \alpha\Delta Z) - (1 - \alpha)X \circ Z - \alpha\mu I\|_F = \alpha \mathrm{O}(\mu_-^{1+\tau'}).$$

By considering the eigenvalues $\lambda_1(\alpha) \le \ldots \le \lambda_p(\alpha)$ of the matrix $X(x + \alpha\Delta x) \circ (Z + \alpha\Delta Z)$ and the eigenvalues $(1 - \alpha)\lambda_1(0) + \alpha\mu \le \ldots \le (1 - \alpha)\lambda_p(0) + \alpha\mu$ of the matrix $(1 - \alpha)X \circ Z + \alpha\mu I$, we obtain the following inequality

$$\begin{aligned} & \sum_{i=1}^{p} |\lambda_i(\alpha) - (1 - \alpha)\lambda_i(0) - \alpha\mu|^2 \\ & \quad \le \|X(x + \alpha\Delta x) \circ (Z + \alpha\Delta Z) - (1 - \alpha)X \circ Z - \alpha\mu I\|_F^2 \end{aligned}$$

by the Hoffman and Wielandt theorem (see p.104 of [21] for example). Thus the above relations yield

$$
\begin{aligned}
(44) \qquad \alpha \mathrm{O}(\mu_-^{1+\tau'}) &= \|X(x+\alpha\Delta x)\circ(Z+\alpha\Delta Z)-(1-\alpha)X\circ Z-\alpha\mu I\|_F \\
&\geq |\lambda_i(\alpha)-(1-\alpha)\lambda_i(0)-\alpha\mu| \\
&\geq (1-\alpha)\lambda_i(0)+\alpha\mu-|\lambda_i(\alpha)|
\end{aligned}
$$

for $i=1,\dots,p$. In order to prove $\lambda_1(\alpha)>0$ for all $\alpha\in(0,1]$, we suppose that there exists $\hat{\alpha}$ satisfying $\lambda_1(\hat{\alpha})=0$ and $\hat{\alpha}\in(0,1]$. Then by (42), we have

$$\frac{1}{2}(1-\hat{\alpha})\mu_- + \hat{\alpha}\mu \leq (1-\hat{\alpha})\lambda_1(0)+\hat{\alpha}\mu-|\lambda_1(\hat{\alpha})| \leq \hat{\alpha}\mathrm{O}(\mu_-^{1+\tau'}),$$

which yields a contradiction because of $\mu=\Theta(\|r_0(w)\|^{1+\tau})=\Theta(\mu_-^{1+\tau})$ and $1>\tau'>\tau$. Thus we obtain $X(x+\alpha\Delta x)\circ(Z+\alpha\Delta Z)\succ 0$, and then $X(x+\alpha\Delta x)\succ 0$ and $Z+\alpha\Delta Z\succ 0$ for all $\alpha\in[0,1]$. By setting $\alpha=1$ in (43), we have

$$(45) \qquad \|X_{\frac{1}{2}}Z_{\frac{1}{2}}-\mu I\|_F = \mathrm{O}(\mu_-^{1+\tau'}) = \mathrm{O}(\mu^{1+\frac{\tau'-\tau}{1+\tau}}),$$

where $X_{\frac{1}{2}}=X(x_{\frac{1}{2}})$. Furthermore, the Newton equations yield

$$(46) \qquad \nabla_x L(w+\Delta w)=\mathrm{O}(\|\Delta w\|^2) \quad \text{and} \quad g(w+\Delta w)=\mathrm{O}(\|\Delta w\|^2).$$

On the other hand, Lemma 4 and the definition of μ yield $\Delta w=\mathrm{O}(\|r(w,\mu)\|)=\mathrm{O}(\|r_0(w)\|)$. Thus equations (45) and (46) imply that the following relation holds

$$(47) \qquad r(w_{\frac{1}{2}},\mu)=\mathrm{O}(\mu^{1+\frac{\tau'-\tau}{1+\tau}}),$$

which proves the first part of this theorem.

Next we show that (41) is satisfied. In the same way as above, we can show the second part of this theorem. In fact, μ and $\frac{\tau'-\tau}{1+\tau}$ in (47) correspond to μ_- and τ' in (39), respectively. Let the eigenvalues of the matrix $X(x_{\frac{1}{2}}+\alpha\Delta x_{\frac{1}{2}})\circ(Z_{\frac{1}{2}}+\alpha\Delta Z_{\frac{1}{2}})$ be $\lambda'_1(\alpha)\leq\dots\leq\lambda'_p(\alpha)$ for each $\alpha\in[0,1]$. Since $\|X_{\frac{1}{2}}\circ Z_{\frac{1}{2}}-\mu I\|_F\leq\|X_{\frac{1}{2}}Z_{\frac{1}{2}}-\mu I\|_F\leq\eta\mu^{1+\frac{\tau'-\tau}{1+\tau}}$ for some positive number η, we have

$$(48) \qquad \lambda'_i(0)\geq\frac{1}{2}\mu>0, \quad i=1,\dots,p$$

as described in (42). Let $E_{\frac{1}{2}}=X_{\frac{1}{2}}Z_{\frac{1}{2}}-\mu I$. Equation (45) and Lemma 3 imply $E_{\frac{1}{2}}=\mathrm{O}(\mu^{1+\frac{\tau'-\tau}{1+\tau}})$ and $X_{\frac{1}{2}}^{-1}=\mathrm{O}(\mu^{-1})$, and Lemma 4 and equation (47) imply

$$\Delta w_{\frac{1}{2}}=\mathrm{O}(\|r(w_{\frac{1}{2}},\mu)\|)=\mathrm{O}(\mu^{1+\frac{\tau'-\tau}{1+\tau}}).$$

Thus equation (35) yields

$$
\begin{aligned}
&X(x_{\frac{1}{2}}+\alpha\Delta x_{\frac{1}{2}})(Z_{\frac{1}{2}}+\alpha\Delta Z_{\frac{1}{2}}) \\
&\quad = X(x_{\frac{1}{2}})Z_{\frac{1}{2}}+\alpha(\mu I-X(x_{\frac{1}{2}})Z_{\frac{1}{2}})+\alpha\mathrm{O}(\|E_{\frac{1}{2}}\|_F)\mathrm{O}(\|X_{\frac{1}{2}}^{-1}\|_F)\mathrm{O}(\|\Delta X_{\frac{1}{2}}\|_F) \\
&\qquad +\alpha^2\mathrm{O}(\mu^{2(1+\frac{\tau'-\tau}{1+\tau})}) \\
&\quad = (1-\alpha)X(x_{\frac{1}{2}})Z_{\frac{1}{2}}+\alpha\mu I+\alpha\mathrm{O}(\mu^{1+2\frac{\tau'-\tau}{1+\tau}})+\alpha^2\mathrm{O}(\mu^{2(1+\frac{\tau'-\tau}{1+\tau})}) \\
(49) \qquad &\quad = (1-\alpha)X(x_{\frac{1}{2}})Z_{\frac{1}{2}}+\alpha\mu I+\alpha\mathrm{O}(\mu^{1+2\frac{\tau'-\tau}{1+\tau}}),
\end{aligned}
$$

which corresponds to (43). Thus as in (44), we have

$$\alpha \mathrm{O}(\mu^{1+2\frac{\tau'-\tau}{1+\tau}}) \geq (1-\alpha)\lambda_i'(0) + \alpha\mu - |\lambda_i'(\alpha)|$$

for $i = 1, \ldots, p$. In order to prove $\lambda_1'(\alpha) > 0$ for all $\alpha \in (0,1]$, we suppose that there exists $\hat{\alpha}$ satisfying $\lambda_1'(\hat{\alpha}) = 0$ and $\hat{\alpha} \in (0,1]$. Then by (48), we have

$$\frac{1}{2}(1-\hat{\alpha})\mu + \hat{\alpha}\mu \leq \mathrm{O}(\mu^{1+2\frac{\tau'-\tau}{1+\tau}}),$$

which yields a contradiction. Thus the fact $X(x_{\frac{1}{2}} + \alpha\Delta x_{\frac{1}{2}}) \circ (Z_{\frac{1}{2}} + \alpha\Delta Z_{\frac{1}{2}}) \succ 0$ implies $X(x_{\frac{1}{2}} + \alpha\Delta x_{\frac{1}{2}}) \succ 0$ and $Z_{\frac{1}{2}} + \alpha\Delta Z_{\frac{1}{2}} \succ 0$ for all $\alpha \in [0,1]$, which means that w_+ is an interior point. Setting $\alpha = 1$ in (49) and using the condition $\tau' > 2\tau/(1-\tau)$ yield

$$\|X(x_{\frac{1}{2}} + \Delta x_{\frac{1}{2}})(Z_{\frac{1}{2}} + \Delta Z_{\frac{1}{2}}) - \mu I\|_F = \mathrm{O}(\mu^{1+2\frac{\tau'-\tau}{1+\tau}}) = \mathrm{o}(\mu^{1+\tau'}) \leq M_c\mu^{1+\tau'}. \tag{50}$$

Furthermore, the Newton equations yield

$$\nabla_x L(w_{\frac{1}{2}} + \Delta w_{\frac{1}{2}}) = \mathrm{O}(\|\Delta w_{\frac{1}{2}})\|^2) = \mathrm{O}(\mu^{2(1+\frac{\tau'-\tau}{1+\tau})}) \tag{51}$$

and

$$g(w_{\frac{1}{2}} + \Delta w_{\frac{1}{2}}) = \mathrm{O}(\|\Delta w_{\frac{1}{2}})\|^2) = \mathrm{O}(\mu^{2(1+\frac{\tau'-\tau}{1+\tau})}). \tag{52}$$

Thus equations (50) - (52) imply that the following relation holds

$$\|r(w_+, \mu)\| \leq M_c\mu^{1+\tau'},$$

which proves the second part of this theorem.

Therefore the proof of this theorem is complete. □

Now we show the two-step superlinear convergence of Algorithm scaledSDPIP in the following theorem.

Theorem 3 *Suppose that assumptions* (A1)-(A4) *hold. Let M_c be a positive constant, and let τ and τ' be positive constants that satisfy*

$$1 > \tau' > \tau \quad \text{and} \quad \tau' > \frac{2\tau}{1-\tau}.$$

Let μ_{-1} be a given number that satisfies

$$\left(\frac{1}{2M_c}\right)^{1/\tau'} \geq \mu_{-1} > 0.$$

Assume that an initial interior point w_0 is sufficiently close to w^* such that the approximate BKKT condition $\|r(w_0, \mu_{-1})\| \leq M_c\mu_{-1}^{1+\tau'}$ is satisfied. Then the sequence $\{w_k\}$ generated by Algorithm scaledSDPIP with $T_k = X_k^{-1/2}$ satisfies

$$\|r(w_k, \mu_{k-1})\| \leq M_c\mu_{k-1}^{1+\tau'}, \quad X(x_k) \succ 0 \quad \text{and} \quad Z_k \succ 0$$

for all $k \geq 0$ and converges two-step superlinearly to w^* in the sense that

$$\|w_k + \Delta w_k + \Delta w_{k+\frac{1}{2}} - w^*\| = \mathrm{O}(\|w_k - w^*\|^{1+\tau'}) \quad \text{for all } k.$$

We can prove this theorem in the same way as the proof of Theorem 2, so we omit it.

6.2 Scaled Newton method with $T_k = W_k^{-1/2}$

Next we consider the case $T_k = W_k^{-1/2}$ given in Section 3, where the matrix W_k is defined by

$$W_k = X_k^{1/2}(X_k^{1/2} Z_k X_k^{1/2})^{-1/2} X_k^{1/2}.$$

We will also show that the point $w_{k+1} = w_k + \Delta w_k + \Delta w_{k+\frac{1}{2}}$ satisfies $\|r(w_{k+1}, \mu_k)\| \le M_c \mu_k^{1+\tau'}$ if $\|r(w_k, \mu_{k-1})\| \le M_c \mu_{k-1}^{1+\tau'}$ holds. This implies the two-step superlinear convergence.

For the choice of T_k, we have

$$\widetilde{X}_k = \widetilde{Z}_k \quad (\text{i.e. } W_k^{-1} X_k W_k^{-1} = Z_k)$$

and (12) and (13) reduce to

$$\Delta Z_k = \mu_k X_k^{-1} - Z_k - W_k^{-1} \Delta X_k W_k^{-1} \tag{53}$$

and

$$(H_k)_{ij} = \operatorname{tr}\left\{ A_i(x_k) W_k^{-1} A_j(x_k) W_k^{-1} \right\}.$$

The following lemma estimates the Newton step Δw_k near the solution w^*.

Lemma 6 *Suppose that assumptions* (A1)–(A4) *hold. Let τ' be a positive constant. Suppose that w is an interior point which is sufficiently close to w^* and that $r(w, \mu_-) = \mathrm{O}(\mu_-^{1+\tau'})$ is satisfied for a positive number μ_-. Let μ be a positive number. Then the Newton step of $\tilde{J}_S(w)\Delta w = -\tilde{r}_S(w, \mu)$ satisfies the following relation*

$$\Delta w = \mathrm{O}(\|r(w, \mu)\|).$$

Proof. By letting $E = XZ - \mu_- I$, we have

$$X^{-1} = \mu_-^{-1}(Z - X^{-1}E). \tag{54}$$

It follows from the definition of W that

$$\begin{aligned} W^{-1} &= X^{-1/2}(X^{1/2} Z X^{1/2})^{1/2} X^{-1/2} \\ &= \mu_-^{1/2} X^{-1/2}(I + \mu_-^{-1} X^{-1/2} E X^{1/2})^{1/2} X^{-1/2} \\ &= \mu_-^{1/2} X^{-1/2} \left(I + \frac{1}{2}\mu_-^{-1} X^{-1/2} E X^{1/2} + M \right) X^{-1/2} \\ &= \mu_-^{1/2} X^{-1} + \frac{1}{2}\mu_-^{-1/2} X^{-1} E + \mu_-^{1/2} X^{-1/2} M X^{-1/2}, \end{aligned} \tag{55}$$

where

$$M = \mathrm{O}(\mu_-^{-2}\|X^{-1/2} E X^{1/2}\|_F^2) = \mathrm{O}(\mu_-^{-2}\|E\|_F^2).$$

The last equality can be obtained from the fact $\|X^{-1/2} E X^{1/2}\|_F = \|E\|_F$. Substituting (54) into (55) yields

$$W^{-1} = \mu_-^{-1/2} Z - \frac{1}{2}\mu_-^{-1/2} X^{-1} E + \mu_-^{1/2} X^{-1/2} M X^{-1/2}. \tag{56}$$

Since we have by (54) and (56)

$$
\begin{aligned}
&XW^{-1}\Delta XW^{-1}\\
&\quad = \left(\mu_-\Delta X + \frac{1}{2}E\Delta X + \mu_- X^{1/2}MX^{-1/2}\Delta X\right)\left(\mu_-^{-1}Z - \frac{1}{2}\mu_-^{-1}X^{-1}E + X^{-1/2}MX^{-1/2}\right),
\end{aligned}
$$

equation (53) yields

$$
\begin{aligned}
X\Delta Z &= \mu I - XZ - XW^{-1}\Delta XW^{-1}\\
&= \mu I - XZ - \{\Delta XZ - \frac{1}{2}\Delta XX^{-1}E + \mu_-\Delta X(X^{-1/2}MX^{1/2})X^{-1} + \frac{1}{2}\mu_-^{-1}E\Delta XZ\\
&\qquad -\frac{1}{4}\mu_-^{-1}E\Delta XX^{-1}E + \frac{1}{2}E\Delta X(X^{-1/2}MX^{1/2})X^{-1} + (X^{1/2}MX^{-1/2})\Delta XZ\\
&\qquad -\frac{1}{2}(X^{1/2}MX^{-1/2})\Delta XX^{-1}E + \mu_-(X^{1/2}MX^{-1/2})\Delta X(X^{-1/2}MX^{1/2})X^{-1}\}\\
&= \mu I - XZ - \Delta XZ + \mathrm{O}(\mu_-^{\tau'})\mathrm{O}(\|\Delta X\|_F),
\end{aligned}
$$

because Lemma 3 implies $X^{-1} = \mathrm{O}(\mu_-^{-1})$, and we have $E = \mathrm{O}(\mu_-^{1+\tau'})$ and $M = \mathrm{O}(\mu_-^{2\tau'})$. This implies that

$$X\Delta Z + \Delta XZ = \mu I - XZ + \mathrm{O}(\mu_-^{\tau'})\mathrm{O}(\|\Delta X\|_F).$$

Thus we obtain

$$X\Delta Z + \Delta XZ + Z\Delta X + \Delta ZX = 2\mu I - (XZ + ZX) + \mathrm{O}(\mu_-^{\tau'})\mathrm{O}(\|\Delta X\|_F). \tag{57}$$

We write equations (7), (8) and (57) by

$$J_S'(w)\Delta w = -r_S(w,\mu),$$

which corresponds to (37). Therefore the lemma can be proved in the same way as the proof of Lemma 4. □

Since we obtain the same lemma as Lemma 5, we can show the following theorem in the same way as Theorem 3.

Theorem 4 *Suppose that assumptions* (A1)-(A4) *hold. Let M_c be a positive constant, and let τ and τ' be positive constants that satisfy*

$$1 > \tau' > \tau \quad \text{and} \quad \tau' > \frac{2\tau}{1-\tau}.$$

Let μ_{-1} be a given number that satisfies

$$\left(\frac{1}{2M_c}\right)^{1/\tau'} \geq \mu_{-1} > 0.$$

Assume that an initial interior point w_0 is sufficiently close to w^* such that the approximate BKKT condition $\|r(w_0,\mu_{-1})\| \leq M_c\mu_{-1}^{1+\tau'}$ is satisfied. Then the sequence $\{w_k\}$ generated by Algorithm scaledSDPIP with $T_k = W_k^{-1/2}$ satisfies

$$\|r(w_k,\mu_{k-1})\| \leq M_c\mu_{k-1}^{1+\tau'}, \quad X(x_k) \succ 0 \quad \text{and} \quad Z_k \succ 0$$

for all $k \geq 0$ and converges two-step superlinearly to w^* in the sense that

$$\|w_k + \Delta w_k + \Delta w_{k+\frac{1}{2}} - w^*\| = \mathrm{O}(\|w_k - w^*\|^{1+\tau'}) \quad \text{for all } k.$$

7 Concluding Remarks

In this paper, we have analyzed local behavior of primal-dual interior point methods for solving nonlinear semidefinite programming problems. We have first proposed a primal-dual interior point method based on the unscaled Newton method, called Algorithm unscaledSDPIP, and have showed its local and superlinear convergence. Next we have proposed two kinds of primal-dual interior point methods based on the scaled Newton method, called Algorithm scaledSDPIP, and have proved their local and two-step superlinear convergence properties.

In order to obtain a globally and superlinearly convergent method, we can combine Algorithm SDPIP described in Section 3 and the proposed methods in Sections 5 and 6. Specifically we propose the following method for modifying Algorithm unscaledSDPIP.

Algorithm unscaledSDPIP(Global)

Step 0. (Initialize) Set $\varepsilon > 0$, $0 < M_c < 1$, $\mu_{-1} > 0$, $0 < \delta < 1$ and $0 < \tau < 1$. Choose $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$ $(X(x_0) \succ 0, Z_0 \succ 0)$. Set $k = 0$.

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

Step 2. (Trial Newton step) If $\|r_0(w_k)\|$ is sufficiently small (i.e. w_k is close to a KKT point), execute the following steps. Otherwise go to Step 3.

Step 2.1 Choose $\mu_k = \xi_k \|r_0(w_k)\|^{1+\tau}$ with $\xi_k = \Theta(1)$. Calculate the direction Δw_k by solving the Newton equations $J_S(w_k)\Delta w_k = -r_S(w_k, \mu_k)$.

Step 2.2 If $\|r(w_k + \Delta w_k, \mu_k)\| \leq M_c \mu_k$, $X(x_k + \Delta x_k) \succ 0$ and $Z_k + \Delta Z_k \succ 0$, then set $w_{k+1} = w_k + \Delta w_k$ and go to Step 4. Otherwise go to Step 3.

Step 3. (Approximate BKKT point) Choose $\mu_k \in (0, \delta\mu_{k-1})$. Find an interior point w_{k+1} that satisfies

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

Step 4. (Update) Set $k := k + 1$ and go to Step 1. □

Next we propose the following method for Algorithm scaledSDPIP.

Algorithm scaledSDPIP(Global)

Step 0. (Initialize) Set $\varepsilon > 0$, $M_c > 0$, $\mu_{-1} > 0$, $0 < \delta < 1$, τ and τ' that satisfy

$$1 > \tau' > \tau > 0 \quad \text{and} \quad \tau' > \frac{2\tau}{1-\tau}.$$

Choose $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S}^p$ $(X(x_0) \succ 0, Z_0 \succ 0)$. Set $k = 0$.

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

Step 2. (Scaled Newton step) If $\|r_0(w_k)\|$ is sufficiently small (i.e. w_k is close to a KKT point), execute the following steps. Otherwise go to Step 3.

Step 2.1 Choose $\mu_k = \xi_k \|r_0(w_k)\|^{1+\tau}$ with $\xi_k = \Theta(1)$.

Step 2.2 Calculate the direction Δw_k by solving the scaled Newton equations $\widetilde{J}_S(w_k)\Delta w_k = -\tilde{r}_S(w_k, \mu_k)$ at w_k. If $X(x_k + \Delta x_k) \succ 0$ and $Z_k + \Delta Z_k \succ 0$, then go to Step 2.3. Otherwise go to Step 3.

Step 2.3 Set $w_{k+\frac{1}{2}} = w_k + \Delta w_k$. Calculate the direction $\Delta w_{k+\frac{1}{2}}$ by solving the scaled Newton equations $\widetilde{J}_S(w_{k+\frac{1}{2}})\Delta w_{k+\frac{1}{2}} = -\tilde{r}_S(w_{k+\frac{1}{2}}, \mu_k)$ at $w_{k+\frac{1}{2}}$. If $\|r(w_{k+\frac{1}{2}} + \Delta w_{k+\frac{1}{2}}, \mu_k)\| \leq M_c \mu_k^{1+\tau'}$, $X(x_{k+\frac{1}{2}} + \Delta x_{k+\frac{1}{2}}) \succ 0$ and $Z_{k+\frac{1}{2}} + \Delta Z_{k+\frac{1}{2}} \succ 0$, then set $w_{k+1} = w_{k+\frac{1}{2}} + \Delta w_{k+\frac{1}{2}}$ and go to Step 4. Otherwise go to Step 3.

Step 3. (Approximate BKKT point) Choose $\mu_k \in (0, \delta\mu_{k-1})$. Find an interior point w_{k+1} that satisfies

$$\|r(w_{k+1}, \mu_k)\| \leq M_c \mu_k^{1+\tau'}.$$

Step 4. (Update) Set $k := k+1$ and go to Step 1. □

We can expect the global convergence property of both algorithms because of the existence of Step 3 as a safeguard. See [23]. At the same time, we also expect that Step 3 is skipped near a KKT point and the superlinear convergence property is obtained as discussed in this paper.

References

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

[2] Bonnans, J. F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Verlag, New York, 2000.

[3] Bonnans, J. F., Ramirez, C. H.: *Strong regularity of semidefinite programming problems.* Technical Report, CMM-DIM B-05/06-137, Department of Mathematical Engineering, Universida de Chile, 2005.

[4] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, 2004.

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

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

[8] Kojima, M., Shida, M. and Shindoh, S.: Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs. Mathematical Programming **80**, 129-160 (1998).

[9] Kojima, M., Shida, M. and Shindoh, S.: A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction. SIAM Journal on Optimization **9**, 444-465 (1999).

[10] 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)

[11] Luo, Z., Sturm, J. F., Zhang, S.: Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. SIAM Journal on Optimization **8**, 59-81 (1998)

[12] Monteiro, R.D.C.: Primal-dual path-following algorithms for semidefinite programming. SIAM Journal on Optimization **7**, 663-678 (1997)

[13] Nesterov, Y.E., Todd, M.J.: Self-scaled barriers and interior-point methods for convex programming. Mathematics of Operations Research **22**, 1-42 (1997)

[14] Nesterov, Y.E., Todd, M.J.: Primal-dual interior-point methods for self-scaled cones. SIAM Journal on Optimization **8**, 324-364 (1998)

[15] Potra, F. A., Sheng, R.: Superlinear convergence of interior-point algorithms for semidefinite programming. Journal of Optimization Theory and Applications **99**, 103-119 (1998)

[16] Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. Mathematical Programming **77**, 301-320 (1997).

[17] Shapiro, A.: Duality, optimality conditions, and perturbation analysis. in "Handbook of Semidefinite Programming, Theory, Algorithms and Applications", 68-92, 2000.

[18] Sun, D.: The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Mathematics of Operations Research **31**, 761-776 (2006).

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

[20] 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).

[21] Wilkinson, J. H.: The Algebraic Eigenvalue Problem, Oxford University Press, Oxford, 1965.

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

[23] Yamashita, H., Yabe, H., Harada, K.: *A primal-dual interior point method for nonlinear semidefinite programming*, Technical Report, Mathematical Systems, Inc., September 2006 (revised November 2008).