# QUADRATIC CONVERGENCE OF A NONSMOOTH NEWTON-TYPE METHOD FOR SEMIDEFINITE PROGRAMS WITHOUT STRICT COMPLEMENTARITY[1]

Christian Kanzow and Christian Nagel

Preprint 250 July 2003

University of Würzburg
Institute of Applied Mathematics and Statistics
Am Hubland
97074 Würzburg
Germany

e-mail: kanzow@mathematik.uni-wuerzburg.de
nagel@mathematik.uni-wuerzburg.de

July 10, 2003

[1]This research was partially supported by the DFG (Deutsche Forschungsgemeinschaft).

**Abstract:** We consider a Newton-type method for the solution of semidefinite programs. This Newton-type method is based on a semismooth reformulation of the semidefinite program as a nonsmooth system of equations. We establish local quadratic convergence of this method under a linear independence assumption and a slightly modified nondegeneracy condition. In contrast to previous investigations, however, the strict complementarity condition is not needed in our analysis.

**Keywords:** Semidefinite programs, Newton's method, quadratic convergence, nondegeneracy, strict complementarity.

# 1 Introduction

A semidefinite program is a minimization problem of the form

$$\min_{X \in \mathcal{S}^{n\times n}} C \bullet X \quad \text{s.t.} \quad A_i \bullet X = b_i \;\; (i = 1, \ldots, m), \;\; X \succeq 0, \tag{1}$$

where $C, A_1, \ldots, A_m \in \mathcal{S}^{n\times n}$ are symmetric matrices and $b \in \mathbb{R}^m$ is a given vector (the notation used here is standard in the semidefinite literature and will be defined at the end of this section). Hence a semidefinite program is a minimization problem with its variables being symmetric matrices rather than ordinary vectors.

Using some results from the corresponding duality theory, it is not difficult to see that, under mild assumptions, the semidefinite program (1) has a solution if and only if the following *optimality conditions*

$$\begin{array}{rcl} \sum_{i=1}^m \lambda_i A_i + S & = & C, \\ A_i \bullet X & = & b_i \quad \forall i = 1, \ldots, m, \\ X \succeq 0,\, S \succeq 0,\, XS & = & 0 \end{array} \tag{2}$$

are solvable. Many of the standard and efficient interior-point methods for the solution of semidefinite programs are based on this reformulation, see, e.g., [6, 2, 22, 20, 11, 14]. However, to the best of our knowledge, none of these interior-point methods is known to be locally superlinearly or quadratically convergent if the solution of (2) does not satisfy the strict complementarity condition.

Another approach is based on a reformulation of the optimality conditions (2) as a nonsmooth system of equations. This idea leads to a couple of semismooth and smoothing methods, see [3, 17, 9, 18, 10]. In order to prove local fast convergence for these methods, however, strict complementarity is also needed among some further assumptions. One exception is a result in [18], where local quadratic convergence of a nonsmooth Newton method is established for semidefinite complementarity problems without assuming strict complementarity, but using a positive definiteness assumption which is never satisfied when this result gets specialized to semidefinite programs. In the revision of the paper [18] (see [19]), the authors use an approach similar to ours and specialize their results to a Newton-type method applied to semidefinite programs, but then they have to assume strict complementarity in order to prove fast local convergence of their method.

The aim of this paper is now to have a closer look at the local convergence behaviour of a nonsmooth Newton-type method for the solution of the optimality conditions (2). It turns out that we can prove local quadratic convergence of this method under a linear independence condition and a certain nondegeneracy condition which is slightly different from a standard nondegeneracy condition used within the local analysis of some other methods for solving semidefinite programs. However, in contrast to these other methods, we do not need the strict complementarity condition.

The organization of the paper is as follows: In Section 2 we give a formal statement of our nonsmooth Newton method and present some background material. This Newton-type method has to solve at each iteration a linearized system. We present a reformulation of this system in the more standard matrix-vector form in Section 3. Based on this reformulation, we present our local convergence analysis in Section 4. Some numerical results illustrating the quadratic convergence behaviour are given in Section 5. We then close with some final remarks in Section 6.

A few words regarding the notation: For two matrices $A, B \in \mathbb{R}^{n\times n}$, we set

$$A \bullet B := \operatorname{tr}(AB^T),$$

where $\operatorname{tr}(C) := \sum_{i=1}^n c_{ii}$ denotes the trace of a matrix $C \in \mathbb{R}^{n\times n}$. It is easy to see that $\bullet$ defines a scalar product on the set of matrices $\mathbb{R}^{n\times n}$. We further write $\mathcal{S}^{n\times n}$ for the set of symmetric matrices in $\mathbb{R}^{n\times n}$, while $A \succeq 0$ and $A \succ 0$ indicate that $A$ is a symmetric positive semidefinite and symmetric positive definite matrix, respectively. If $A \succeq 0$, we denote by $A^{1/2}$ the unique symmetric positive semidefinite square root of $A$. Moreover, for any matrix $A \in \mathcal{S}^{n\times n}$ we set $|A| = (A^2)^{1/2}$. Finally, if $E \succ 0$ is a given symmetric positive definite matrix, the corresponding *Lyapunov operator* $L_E$ is defined by $L_E[X] := EX + XE\,(X \in \mathcal{S}^{n\times n})$. Then it is well-known (see [8]) that the resulting *Lyapunov equation* $L_E[X] = H$ has a unique solution for each symmetric $H \in \mathcal{S}^{n\times n}$, and we denote this solution by $L_E^{-1}[H]$.

# 2 Nonsmooth Newton Method

In order to formulate the optimality conditions (2) as a system of equations, let us introduce the function

$$\phi(X,S) := X + S - |X - S| \quad (X, S \in \mathcal{S}^{n\times n}). \tag{3}$$

The mapping $\phi$ is usually called the *minimum function.* Note that it is not differentiable everywhere. Nevertheless, it has a number of interesting properties which we collect in the following result.

**Proposition 2.1** *Let $\phi$ be defined by* (3)*. Then the following statements hold for any two matrices $X, S \in \mathcal{S}^{n\times n}$:*

*(a) $\phi$ satisfies the equivalence*

$$\phi(X,S) = 0 \Longleftrightarrow X \succeq 0, S \succeq 0, XS = 0. \tag{4}$$

*(b) If $E := |X - S|$ is nonsingular, then $\phi$ is continuously differentiable (in the sense of Fréchet) with*

$$\nabla\phi(X,S)(U,V) = U + V - L_E^{-1}\big[(X-S)(U-V) + (U-V)(X-S)\big].$$

*(c) The matrix $E = |X - S| = \big((X-S)^2\big)^{1/2}$ is nonsingular (or, equivalently, positive definite) if and only if the mapping $(A,B) \mapsto |A - B|$ is continuously differentiable at $(A,B) = (X,S)$.*

For a proof of part (a) we refer to [21, 9], part (b) may be found in [3, 9], and part (c) follows from [17].

Part (a) may be used in order to reformulate the optimality conditions (2) as the nonsmooth system of equations

$$\Theta(X, \lambda, S) = 0, \tag{5}$$

where

$$\Theta : \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n} \to \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n}$$

is defined by

$$\Theta(X, \lambda, S) := \begin{pmatrix} \sum_{i=1}^m \lambda_i A_i + S - C \\ A_i \bullet X - b_i \;\; (i = 1, \dots, m) \\ \phi(X, S) \end{pmatrix}.$$

A nonsmooth Newton method applied to the system (5) is an iterative method of the form

$$W^{k+1} := W^k + \Delta W^k \quad \forall k = 0, 1, 2, \dots$$

where we used the abbreviation

$$W^k := \big(X^k, \lambda^k, S^k\big),$$

and where $\Delta W^k$ is a solution of the linearized equation $H^k \Delta W^k = -\Theta(W^k)$ with $H^k$ being an element of the B-subdifferential of $\Theta$ in $W^k$, i.e., $H^k \in \partial_B \Theta(W^k)$, where

$$\partial_B \Theta(W) := \big\{ H \,\big|\, \exists \{W^k\} \subset D_\Theta : W^k \to W, \nabla\Theta(W^k) \to H \big\}$$

and where $D_\Theta$ denotes the set of points $W$ at which $\Theta$ is differentiable, cf. [15]. Note that this set is always nonempty because $\Theta$ is locally Lipschitz. Moreover, we have $\partial_B \Theta(W^k) = \{\nabla\Theta(W^k)\}$ whenever $\Theta$ is continuously differentiable at $W^k$.

We formally state our method in the following algorithm.

**Algorithm 2.2**

(S.0) Choose $W^0 := (X^0, \lambda^0, S^0) \in \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n}$, $\varepsilon \geq 0$, and set $k := 0$.

(S.1) If $\|\Theta(W^k)\| \leq \varepsilon$: STOP.

(S.2) Choose $H^k \in \partial_B \Theta(W^k)$ and find a solution $\Delta W^k = \big(\Delta X^k, \Delta\lambda^k, \Delta S^k\big)$ of the linearized system $H^k \Delta W = -\Theta(W^k)$.

(S.3) Set $W^{k+1} := W^k + \Delta W^k$, $k \leftarrow k + 1$, and go to (S.1).

We note that the solution $\Delta W^k = \big(\Delta X^k, \Delta\lambda^k, \Delta S^k\big)$ of the linearized equation in (S.2) automatically produces symmetric matrices $\Delta X^k$ and $\Delta S^k$, see [3, 9]. – The main local convergence result is as follows.

**Theorem 2.3** *Let $W^* := (X^*, \lambda^*, S^*)$ be a solution of the optimality conditions* (2) *and assume that all elements $H^* \in \partial_B \Theta(W^*)$ are invertible. Then Algorithm 2.2 is locally quadratically convergent.*

**Proof.** According to [17], the minimum function $\phi$ is strongly semismooth (see [16, 15] for the definition and some properties of strongly semismooth mappings). Hence $\Theta$ is strongly semismooth. The local quadratic convergence result therefore follows from a general theorem in [15], see also the recent books [4, 5] for more details on this subject. $\square$

In view of Theorem 2.3, the aim of this paper is to find suitable conditions under which the elements of the B-subdifferential $\partial_B \Theta(W^*)$ are invertible. This will be done in Section 4 after some preliminary discussions in the following section.

# 3 Matrix-Vector-Formulation of Newton System

Throughout this section, we assume that the mapping $\Theta$ is continuously differentiable at a current point $(X, \lambda, S) \in \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n}$. According to Proposition 2.1 (b), (c), this assumption is satisfied if and only if the matrix $X - S$ is nonsingular. Later, in Section 4, we will drop this assumption.

Since $\Theta$ is continuously differentiable at $(X, \lambda, S)$, Newton's method applied to $\Theta(X, \lambda, S) = 0$ has to solve the linearized equation

$$\nabla \Theta(X, \lambda, S)(\Delta X, \Delta\lambda, \Delta S) = -\Theta(X, \lambda, S) \tag{6}$$

at the current point. For our analysis in Section 4, it will be useful to write this linearized system in the usual matrix-vector format. We therefore present such a formulation in this section.

Since $\Theta$ is continuously differentiable at $(X, \lambda, S)$, the minimum function $\phi$ is continuously differentiable at $(X, S)$. In view of Proposition 2.1 (c), this means that the matrix

$$E := |X - S| = \big((X - S)^2\big)^{1/2} \tag{7}$$

is positive definite. Consequently, the corresponding Lyapunov operator $L_E$ is invertible. Defining the residuals

$$\begin{aligned} R_C &:= C - \sum_{j=1}^m \lambda_j A_j - S, \\ r_{b,i} &:= b_i - A_i \bullet X \qquad (i = 1, \ldots, m), \\ r_b &:= \big(r_{b,1}, \ldots, r_{b,m}\big)^T, \end{aligned}$$

the Newton system (6) can be rewritten as

$$\sum_{j=1}^m \Delta\lambda_j A_j + \Delta S = R_C, \tag{8}$$

$$A_i \bullet \Delta X = r_{b,i} \qquad (i = 1, \ldots, m), \tag{9}$$

$$\nabla \phi(X, S)(\Delta X, \Delta S) = -\phi(X, S). \tag{10}$$

In order to reformulate this system in the usual matrix-vector format, we need to transfrom matrices into vectors. For a general (not necessarily symmetric) matrix $A \in \mathbb{R}^{n\times n}$, this can be done by using the mapping vec : $\mathbb{R}^{n\times n} \to \mathbb{R}^{n^2}$ defined by

$$\operatorname{vec}(A) := \big(a_{11}, a_{21}, \dots, a_{n1}, a_{12}, a_{22}, \dots, a_{n2}, \dots, a_{nn}\big)^T \in \mathbb{R}^{n^2},$$

i.e., vec stacks the columns of $A$ into a vector of length $n^2$. For a symmetric matrix, we are not interested in all entries of $A$. It suffices to consider the lower triangular part of $A$, and the corresponding transformation can be done using the mapping svec : $\mathcal{S}^{n\times n} \to \mathbb{R}^{\frac{n(n+1)}{2}}$ defined by

$$\operatorname{svec}(A) := \big(a_{11}, \sqrt{2}a_{21}, \dots, \sqrt{2}a_{n1}, a_{22}, \sqrt{2}a_{32}, \dots, \sqrt{2}a_{n2}, \dots, a_{nn}\big)^T \in \mathbb{R}^{\frac{n(n+1)}{2}}.$$

The reason for the $\sqrt{2}$ factor in front of all nondiagonal elements is due to the fact that this is consistent with the inner product, i.e.,

$$A \bullet B = \operatorname{svec}(A)^T \operatorname{svec}(B) \quad \forall A, B \in \mathcal{S}^{n\times n}. \tag{11}$$

Having introduced vec and svec, the next question is how an ordinary matrix product can be expressed in terms of vec and svec. To this end, let us define the *Kronecker product* of two (not necessarily symmetric) matrices $G, K \in \mathbb{R}^{n\times n}$ by

$$G \otimes K := \big[g_{ij}K\big] \in \mathbb{R}^{n^2\times n^2}.$$

Then it can easily be verified that

$$\big(G \otimes K\big) \operatorname{vec}(H) = \operatorname{vec}\big(KHG^T\big) \qquad (H \in \mathbb{R}^{n\times n}).$$

Similarly, we define the *symmetric Kronecker product* by

$$\big(G \otimes_s K\big) \operatorname{svec}(H) := \frac{1}{2} \operatorname{svec}\big(KHG^T + GHK^T\big) \quad (H \in \mathcal{S}^{n\times n}). \tag{12}$$

Some properties of the symmetric Kronecker product are summarized in the following result. The proofs of these properties are elementary. In fact, statements (a), (b) and (c) can be found in [2, 20], whereas part (d) is just a reformulation of statement (c).

**Lemma 3.1** *The symmetric Kronecker product $\otimes_s$ defined by* (12) *has the following properties:*

*(a)* $G \otimes_s K = K \otimes_s G$.

*(b) If $G$ and $K$ are symmetric positive definite, then so is $G \otimes_s K$.*

*(c) If $G, K$ are two commuting symmetric matrices with eigenvalues $\sigma_1, \dots, \sigma_n$ and $\mu_1, \dots, \mu_n$, respectively, and if $q_1, \dots, q_n$ denotes a common set of orthonormal eigenvectors, then the $n(n+1)/2$ eigenvalues of $G \otimes_s K$ are given by*

$$\frac{1}{2}\big(\sigma_i\mu_j + \mu_i\sigma_j\big) \quad (1 \le j \le i \le n)$$

*with the corresponding set of orthonormal eigenvectors* $v_{ij}$ $(1 \leq j \leq i \leq n)$ *defined by*

$$v_{ij} := \begin{cases} \operatorname{svec}\left(q_i q_i^T\right), & \text{if } i = j, \\ \frac{1}{\sqrt{2}} \operatorname{svec}\left(q_i q_j^T + q_j q_i^T\right), & \text{if } j < i. \end{cases}$$

(d) *If* $G, K$ *are two commuting symmetric matrices with simultaneous spectral decompositions* $G = QD_GQ^T$ *and* $K = QD_KQ^T$ *for some orthogonal matrix* $Q \in \mathbb{R}^{n\times n}$ *and diagonal matrices* $D_G, D_K \in \mathbb{R}^{n\times n}$*, then* $G \otimes_s K = (Q \otimes_s Q)(D_G \otimes_s D_K)(Q \otimes_s Q)^T$ *is a spectral decomposition of* $G \otimes_s K$ *(in particular,* $Q \otimes_s Q$ *is an orthogonal matrix).*

We now consider the Newton system (6), i.e., we consider the system (8)–(10). The first two equations (8) and (9) may be reformulated in matrix-vector notation in exactly the same way as described in [20], resulting in the two equations

$$\mathcal{A}^T \Delta\lambda + \operatorname{svec}(\Delta S) = \operatorname{svec}(R_C) \tag{13}$$

and

$$\mathcal{A} \operatorname{svec}(\Delta X) = r_b, \tag{14}$$

respectively, where

$$\mathcal{A} := \big(\operatorname{svec}(A_1), \ldots, \operatorname{svec}(A_m)\big)^T \in \mathbb{R}^{m\times \frac{n(n+1)}{2}}. \tag{15}$$

Hence it remains to consider the third block (10). Using Proposition 2.1 (b) and the definition of $E$ from (7), equation (10) can be rewritten as

$$\Delta X + \Delta S - L_E^{-1}\big[(X - S)(\Delta X - \Delta S) + (\Delta X - \Delta S)(X - S)\big] = -\phi(X, S), \tag{16}$$

Applying the Lyapunov operator $L_E$ to both sides of (16) yields

$$L_E[\Delta X] + L_E[\Delta S] - (X - S)(\Delta X - \Delta S) - (\Delta X - \Delta S)(X - S) = -L_E\big[\phi(X, S)\big].$$

Rearranging terms gives

$$L_{E-(X-S)}[\Delta X] + L_{E+(X-S)}[\Delta S] = -L_E\big[\phi(X, S)\big].$$

Using the notation

$$A_E := E - (X - S) \quad \text{and} \quad B_E := E + (X - S), \tag{17}$$

this equation may be rewritten as

$$L_{A_E}[\Delta X] + L_{B_E}[\Delta S] = -L_E\big[\phi(X, S)\big]. \tag{18}$$

Applying $\frac{1}{2}$svec to both sides then gives

$$\frac{1}{2} \operatorname{svec}\left(L_{A_E}[\Delta X]\right) + \frac{1}{2} \operatorname{svec}\left(L_{B_E}[\Delta S]\right) = -\frac{1}{2} \operatorname{svec}\left(L_E\big[\phi(X, S)\big]\right).$$

Using the definition (12) of svec, we have the identity

$$\frac{1}{2}\,\mathrm{svec}\,\big(L_A[H]\big) = \frac{1}{2}\,\mathrm{svec}\,\big(AHI + IHA\big) = \big(I \otimes_s A\big)\,\mathrm{svec}(H),$$

for all symmetric matrices $A, H \in \mathcal{S}^{n\times n}$. Setting

$$\mathcal{E} := I \otimes_s A_E \quad \text{and} \quad \mathcal{F} := I \otimes_s B_E, \tag{19}$$

we therefore get

$$\mathcal{E}\,\mathrm{svec}(\Delta X) + \mathcal{F}\,\mathrm{svec}(\Delta S) = -(I \otimes_s E)\,\mathrm{svec}\,\big(\phi(X,S)\big). \tag{20}$$

Since $E$ and $I$ are both positive definite, it follows from Lemma 3.1 (b) that the matrix $(I \otimes_s E)$ is also positive definite and therefore, in particular, nonsingular. Hence (20) can be rewritten as

$$(I \otimes_s E)^{-1}\mathcal{E}\,\mathrm{svec}(\Delta X) + (I \otimes_s E)^{-1}\mathcal{F}\,\mathrm{svec}(\Delta S) = -\,\mathrm{svec}\,\big(\phi(X,S)\big). \tag{21}$$

Summarizing our discussion, we obtain the following result as a consequence of (13), (14), and (21).

**Theorem 3.2** *Let $\big(X,\lambda,S\big)$ be given such that $X - S$ is nonsingular. Then the triple $\big(\Delta X, \Delta\lambda, \Delta S\big) \in \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n}$ satisfies the Newton system (6) if and only if the vector $\big(\mathrm{svec}(\Delta X), \Delta\lambda, \mathrm{svec}(\Delta S)\big)$ satisfies the linear system of equations*

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ (I \otimes_s E)^{-1}\mathcal{E} & 0 & (I \otimes_s E)^{-1}\mathcal{F} \end{pmatrix} \begin{pmatrix} \mathrm{svec}(\Delta X) \\ \Delta\lambda \\ \mathrm{svec}(\Delta S) \end{pmatrix} = \begin{pmatrix} \mathrm{svec}(R_C) \\ r_b \\ -\,\mathrm{svec}\,\big(\phi(X,S)\big) \end{pmatrix}. \tag{22}$$

Note that the linear system (22) looks very similar to the one obtained for interior-point methods in [20], however, the reader should be careful because the matrices $\mathcal{E}$ and $\mathcal{F}$ have a different meaning here.

# 4 Local Convergence Analysis

Now let $(X^*, \lambda^*, S^*)$ be a solution of the optimality conditions (2). In order to motivate our approach, we first assume that the following conditions hold; in particular, we assume that strict complementarity holds at this solution.

**Assumption 4.1** *Let $(X^*, \lambda^*, S^*)$ be a solution of the optimality conditions* (2)*.*

*(A.1) (Linear independence)*
*The matrices $A_1, \ldots, A_m$ are linearly independent.*

*(A.2) (Nondegeneracy)*
*The following implication holds for any triple* $(\Delta X, \Delta\lambda, \Delta S)$*:*

$$\left.\begin{array}{l} \sum_{i=1}^m \Delta\lambda_i A_i + \Delta S = 0, \\ A_i \bullet \Delta X = 0 \ \ (i = 1, \dots, m), \\ X^* \Delta S + \Delta X S^* = 0 \end{array}\right\} \Longrightarrow (\Delta X, \Delta S) = (0, 0).$$

*(A.3) (Strict complementarity)*
$X^* + S^* \succ 0.$

These conditions are quite standard in order to prove local fast convergence of several methods for the solution of semidefinite programs, see, for example, [2, 11, 3, 9]. We next give some vector formulations of these conditions.

**Lemma 4.2** *Let* $(X^*, \lambda^*, S^*)$ *be a solution of the optimality conditions* (2)*. Then the following statements hold:*

*(a) Assumption (A.1) is equivalent to the full rank of the matrix* $\mathcal{A}$ *from* (15)*.*

*(b) Assumption (A.2) is equivalent to the following implication:*

$$\left.\begin{array}{l} \mathcal{A}^T \Delta\lambda + \operatorname{svec}(\Delta S) = 0, \\ \mathcal{A} \operatorname{svec}(\Delta X) = 0, \\ \big(I \otimes_s X^*\big) \operatorname{svec}(\Delta S) + \big(I \otimes_s S^*\big) \operatorname{svec}(\Delta X) = 0 \end{array}\right\} \Longrightarrow \left\{ \begin{array}{l} \operatorname{svec}(\Delta X) = 0, \\ \operatorname{svec}(\Delta S) = 0. \end{array}\right.$$

*(c) Assumption (A.2) is equivalent to the following implication:*

$$\left.\begin{array}{l} \mathcal{A}^T \Delta\lambda + \operatorname{svec}(\Delta S) = 0, \\ \mathcal{A} \operatorname{svec}(\Delta X) = 0, \\ V^* \big(I \otimes_s D_{S^*}\big)(V^*)^T \operatorname{svec}(\Delta X) + \dots \\ \quad V^* \big(I \otimes_s D_{X^*}\big)(V^*)^T \operatorname{svec}(\Delta S) = 0 \end{array}\right\} \Longrightarrow \left\{ \begin{array}{l} \operatorname{svec}(\Delta X) = 0, \\ \operatorname{svec}(\Delta S) = 0. \end{array}\right.$$

*Here,* $X^* = Q D_{X^*} Q^T$ *and* $S^* = Q D_{S^*} Q^T$ *denotes the simultaneous spectral decomposition of the two commuting matrices* $X^*, S^*$*, and* $V^* := Q \otimes_s Q$*.*

**Proof.** (a) This follows directly from the definition of the matrix $\mathcal{A}$.

(b) By applying the svec operator, we obviously have

$$\sum_{i=1}^m \Delta\lambda_i A_i + \Delta S = 0 \Longleftrightarrow \mathcal{A}^T \Delta\lambda + \operatorname{svec}(\Delta S) = 0$$

and

$$A_i \bullet \Delta X = 0 \, (i = 1, \dots, m) \Longleftrightarrow \mathcal{A} \operatorname{svec}(\Delta X) = 0.$$

Hence it remains to show the equivalence

$$X^*\Delta S + \Delta X S^* = 0 \Longleftrightarrow \big(I \otimes_s X^*\big)\,\mathrm{svec}(\Delta S) + \big(I \otimes_s S^*\big)\,\mathrm{svec}(\Delta X) = 0. \qquad (23)$$

To this end, first assume that $X^*\Delta S + \Delta X S^* = 0$. By taking the transpose and recalling that $\Delta S, \Delta X$ are automatically symmetric, we obtain $\Delta S X^* + S^*\Delta X = 0$. Adding these two equations, applying $\frac{1}{2}$ svec to the resulting equation and using (12) gives the right-hand side formulation in (23). Conversely, assume that

$$\big(I \otimes_s X^*\big)\,\mathrm{svec}(\Delta S) + \big(I \otimes_s S^*\big)\,\mathrm{svec}(\Delta X) = 0$$

holds. In view of (12), this is equivalent to

$$X^*\Delta S + \Delta S X^* + S^*\Delta X + \Delta X S^* = 0.$$

Proceeding as in the proof of [12, Lemma 6.2] (in this reference, strict complementarity is assumed, but it is easy to see that this assumption can be avoided here), it follows that $X^*\Delta S + \Delta X S^* = 0$, so that (23) holds.

(c) This statement follows immediately from part (b) by applying Lemma 3.1 (d). □

Part (b) of Lemma 4.2 will be used in our subsequent nonsingularity result, whereas the reformulation given in statement (c) is presented here in order to have a better comparison between the nondegeneracy condition from (A.2) and the one to be introduced later.

We next show how our previous results may be used in order to prove the nonsingularity of the matrix from Theorem 3.2 at a strictly complementary solution.

**Theorem 4.3** *Let* $(X^*, \lambda^*, S^*)$ *be a solution of the optimality conditions* (2) *satisfying Assumptions (A.1)–(A.3). Furthermore define the matrices (cf.* (17), (19)*)*

$$\begin{aligned} E^* &:= |X^* - S^*|, \\ \mathcal{E}^* &:= I \otimes_s A_{E^*} \quad \textit{with} \quad A_{E^*} := E^* - (X^* - S^*), \\ \mathcal{F}^* &:= I \otimes_s B_{E^*} \quad \textit{with} \quad B_{E^*} := E^* + (X^* - S^*). \end{aligned}$$

*Then* $(I \otimes_s E^*)^{-1}$ *exists, and the matrix*

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ (I \otimes_s E^*)^{-1}\mathcal{E}^* & 0 & (I \otimes_s E^*)^{-1}\mathcal{F}^* \end{pmatrix}$$

*(cf.* (22)*) is nonsingular.*

**Proof.** Since (A.3) holds, the matrix $E^*$ is positive definite. Hence $I \otimes_s E^*$ is also positive definite by Lemma 3.1 (b). Therefore, the inverse $(I \otimes_s E^*)^{-1}$ exists, and we are done if we are able to show that the matrix

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \mathcal{E}^* & 0 & \mathcal{F}^* \end{pmatrix}$$

is nonsingular. To this end, let $\big(\mathrm{svec}(\Delta X), \Delta\lambda, \mathrm{svec}(\Delta S)\big)$ be a given triple with

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \mathcal{E}^* & 0 & \mathcal{F}^* \end{pmatrix} \begin{pmatrix} \mathrm{svec}(\Delta X) \\ \Delta\lambda \\ \mathrm{svec}(\Delta S) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Blockwise this becomes

$$\begin{aligned} \mathcal{A}^T\Delta\lambda + \mathrm{svec}(\Delta S) &= 0, \\ \mathcal{A}\,\mathrm{svec}(\Delta X) &= 0, \\ \mathcal{E}^*\,\mathrm{svec}(\Delta X) + \mathcal{F}^*\,\mathrm{svec}(\Delta S) &= 0. \end{aligned} \tag{24}$$

Since $X^*S^* = 0$, the two matrices $X^*, S^*$ commute. Hence there is an orthogonal matrix $Q^* \in \mathbb{R}^{n\times n}$ and two diagonal matrices $D_{X^*}, D_{S^*} \succeq 0$ such that

$$X^* = Q^* D_{X^*} (Q^*)^T \quad \text{and} \quad S^* = Q^* D_{S^*} (Q^*)^T, \tag{25}$$

cf. [7]. This implies

$$X^*S^* = 0 \Longleftrightarrow D_{X^*}D_{S^*} = 0.$$

We therefore get

$$\begin{aligned} E^* &= |X^* - S^*| \\ &= \big((X^* - S^*)^2\big)^{1/2} \\ &= Q\big((D_{X^*} - D_{S^*})^2\big)^{1/2} Q^T \\ &= Q\big(D_{X^*} + D_{S^*}\big) Q^T \\ &= X^* + S^*. \end{aligned}$$

This yields

$$A_{E^*} = E^* - (X^* - S^*) = 2S^* \quad \text{and} \quad B_{E^*} = E^* + (X^* - S^*) = 2X^*.$$

Consequently, we obtain

$$\mathcal{E}^* = I \otimes_s A_{E^*} = 2\big(I \otimes_s S^*\big) \quad \text{and} \quad \mathcal{F}^* = I \otimes_s B_{E^*} = 2\big(I \otimes_s X^*\big).$$

Hence the last line from (24) may be rewritten as

$$\big(I \otimes_s X^*\big)\,\mathrm{svec}(\Delta S) + \big(I \otimes_s S^*\big)\,\mathrm{svec}(\Delta X) = 0.$$

Together with the first two lines from (24), it now follows from Assumption (A.2) and Lemma 4.2 (b) that $\mathrm{svec}(\Delta X) = 0$ and $\mathrm{svec}(\Delta S) = 0$. This, in turn, implies $\mathcal{A}^T\Delta\lambda = 0$ by (24). Since the columns of $\mathcal{A}^T$ are linearly independent by Lemma 4.2 (a), we finally get $\Delta\lambda = 0$. □

The previous result may also be obtained in a different way from [3, 9], however, the formulation given here is more convenient for our subsequent generalization.

If (A.3) does not hold at $W^* = (X^*, \lambda^*, S^*)$, then the matrix $E^*$ is not positive definite any longer. Hence there is no reason for $I \otimes E^*$ to be nonsingular. Nevertheless, it is possible to generalize the previous discussion. To this end, we recall that any element $H^* \in \partial_B \Theta(W^*)$ may be obtained by taking a sequence $W^k := (X^k, \lambda^k, S^k)$ converging to $W^*$ such that $\Theta$ is differentiable at each $W^k$ and such that $H^* = \lim_{k\to\infty} \nabla\Theta(W^k)$. In view of Proposition 2.1 (b), (c), however, $\Theta$ is continuously differentiable at $W^k$ if and only if $X^k - S^k$ is nonsingular. But this is precisely the assumption we used in our analysis of Section 3. Hence, using the corresponding notation (cf. (7), (17), (19))

$$\begin{aligned} E^k &:= \left|X^k - S^k\right|, \\ \mathcal{E}^k &:= I \otimes_s A_{E^k} \quad \text{with} \quad A_{E^k} := E^k - (X^k - S^k), \\ \mathcal{F}^k &:= I \otimes_s B_{E^k} \quad \text{with} \quad B_{E^k} := E^k + (X^k - S^k) \end{aligned}$$

it follows from Theorem 3.2 that the usual matrix formulation of the Jacobian $\nabla\Theta(W^k)$ is given by

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ (I \otimes_s E^k)^{-1}\mathcal{E}^k & 0 & (I \otimes_s E^k)^{-1}\mathcal{F}^k \end{pmatrix}. \tag{26}$$

We have to find the possible limiting elements of a sequence of matrices of this form. We therefore have to take a closer look at the last block row, i.e., at the limiting behaviour of the two matrices

$$(I \otimes_s E^k)^{-1}\mathcal{E}^k = (I \otimes_s E^k)^{-1}(I \otimes_s A_{E^k})$$

and

$$(I \otimes_s E^k)^{-1}\mathcal{F}^k = (I \otimes_s E^k)^{-1}(I \otimes_s B_{E^k}).$$

To this end, let

$$X^k - S^k = Q^k \Pi^k (Q^k)^T \quad \text{with} \quad \Pi^k = \operatorname{diag}(\pi_1^k, \ldots, \pi_n^k)$$

be a spectral decomposition of the symmetric matrix $X^k - S^k$, and let

$$\left|\Pi^k\right| = \operatorname{diag}\left(\left|\pi_1^k\right|, \ldots, \left|\pi_n^k\right|\right).$$

Then it is not difficult to see that

$$\begin{aligned} E^k &= Q^k \left|\Pi^k\right| (Q^k)^T, \\ A_{E^k} &= Q^k \left(\left|\Pi^k\right| - \Pi^k\right)(Q^k)^T, \\ B_{E^k} &= Q^k \left(\left|\Pi^k\right| + \Pi^k\right)(Q^k)^T. \end{aligned}$$

Let $q_i^k$ be the $i$th column of $Q^k$ and define the orthogonal matrix $V^k$ by

$$V^k := \left(\ldots, v_{ij}^k, \ldots\right)_{1 \le j \le i \le n}$$

with the columns $v_{ij}^k \in \mathbb{R}^{n(n+1)/2}$ given by

$$v_{ij}^k := \begin{cases} \operatorname{svec}\big(q_i^k(q_i^k)^T\big), & \text{if } i = j, \\ \frac{1}{\sqrt{2}} \operatorname{svec}\big(q_i^k(q_j^k)^T + q_j^k(q_i^k)^T\big), & \text{if } j < i, \end{cases}$$

Then Lemma 3.1 (c) yields

$$\begin{aligned} I \otimes_s E^k &= V^k \operatorname{diag}\left(\dots, \frac{1}{2}\big(|\pi_i^k| + |\pi_j^k|\big), \dots\right)(V^k)^T, \\ I \otimes_s A_{E^k} &= V^k \operatorname{diag}\left(\dots, \frac{1}{2}\big(|\pi_i^k| + |\pi_j^k| - \pi_i^k - \pi_j^k\big), \dots\right)(V^k)^T, \\ I \otimes_s B_{E^k} &= V^k \operatorname{diag}\left(\dots, \frac{1}{2}\big(|\pi_i^k| + |\pi_j^k| + \pi_i^k + \pi_j^k\big), \dots\right)(V^k)^T. \end{aligned}$$

Note that all diagonal elements $\pi_i^k$ of $\Pi^k$ are nonzero since $E^k$ was assumed to be nonsingular, so that $|\pi_i^k| + |\pi_j^k| \neq 0$ for all $1 \le j \le i \le n$. We therefore obtain

$$\begin{aligned} (I \otimes_s E^k)^{-1}(I \otimes_s A_{E^k}) &= V^k \operatorname{diag}\left(\dots, \frac{|\pi_i^k| + |\pi_j^k| - \pi_i^k - \pi_j^k}{|\pi_i^k| + |\pi_j^k|}, \dots\right)(V^k)^T \\ &=: V^k \Sigma_-^k (V^k)^T, \\ (I \otimes_s E^k)^{-1}(I \otimes_s B_{E^k}) &= V^k \operatorname{diag}\left(\dots, \frac{|\pi_i^k| + |\pi_j^k| + \pi_i^k + \pi_j^k}{|\pi_i^k| + |\pi_j^k|}, \dots\right)(V^k)^T \\ &=: V^k \Sigma_+^k (V^k)^T. \end{aligned}$$

Using the above identities we can rewrite the coefficient matrix from (26) as

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ V^k \Sigma_-^k (V^k)^T & 0 & V^k \Sigma_+^k (V^k)^T \end{pmatrix}.$$

Subsequencing if neccessary, we may assume without loss of generality that the orthogonal (and therefore bounded) matrix sequence $\{Q^k\}$ (and therefore $\{\Pi^k\}$) converge with

$$Q^* := \lim_{k \to \infty} Q^k \quad \text{and} \quad \Pi^* := \lim_{k \to \infty} \Pi^k.$$

Clearly, $Q^*$ is again an orthogonal matrix such that the identity $X^* - S^* = Q^* \Pi^* (Q^*)^T$ holds, i.e., $Q^*, \Pi^*$ correspond to a spectral decomposition of the symmetric matrix $X^* - S^*$. We also note that the corresponding subsequence of $V^k$ converges to

$$V^* := \big(\dots, v_{ij}^*, \dots\big)_{1 \le j \le i \le n}, \tag{27}$$

where

$$v_{ij}^* := \begin{cases} \operatorname{svec}\big(q_i^*(q_i^*)^T\big), & \text{if } i = j, \\ \frac{1}{\sqrt{2}} \operatorname{svec}\big(q_i^*(q_j^*)^T + q_j^*(q_i^*)^T\big), & \text{if } j < i, \end{cases} \tag{28}$$

and $q_i^*$ denotes the $i$th column of $Q^*$, cf. Lemma 3.1 (c). Furthermore, it is not difficult to see that all diagonal elements of the matrices $\Sigma_-^k$ and $\Sigma_+^k$ belong to the interval $[0,2]$ and that $\Sigma_-^k + \Sigma_+^k = 2I$. In particular, we may also assume without loss of generality that

$$\Sigma_-^k \to \Sigma_-^* \quad \text{and} \quad \Sigma_+^k \to \Sigma_+^*$$

for some diagonal matrices $\Sigma_-^*, \Sigma_+^*$. Then we have

$$\begin{array}{rcl} \big(I \otimes_s E^k\big)^{-1}\big(I \otimes_s A_{E^k}\big) & \to & V^*\Sigma_-^*(V^*)^T, \\ \big(I \otimes_s E^k\big)^{-1}\big(I \otimes_s B_{E^k}\big) & \to & V^*\Sigma_+^*(V^*)^T. \end{array} \tag{29}$$

In order to see the precise structure of the diagonal matrices $\Sigma_-^*$ and $\Sigma_+^*$, let us write $\Pi^* = \operatorname{diag}\big(\pi_1^*, \dots, \pi_n^*\big)$ and define the index sets

$$\begin{array}{rcl} \alpha & := & \big\{i \,\big|\, \pi_i^* > 0\big\} = \big\{i \,\big|\, \lambda_i(X^*) > 0, \lambda_i(S^*) = 0\big\}, \\ \beta & := & \big\{i \,\big|\, \pi_i^* = 0\big\} = \big\{i \,\big|\, \lambda_i(X^*) = 0, \lambda_i(S^*) = 0\big\}, \\ \gamma & := & \big\{i \,\big|\, \pi_i^* < 0\big\} = \big\{i \,\big|\, \lambda_i(X^*) = 0, \lambda_i(S^*) > 0\big\}, \end{array} \tag{30}$$

where $\lambda_i(X^*)$ and $\lambda_i(S^*)$ denote the eigenvalues of $X^*$ and $S^*$, respectively. Then an easy calculation shows that

$$\frac{\big|\pi_i^k\big| + \big|\pi_j^k\big| - \pi_i^k - \pi_j^k}{\big|\pi_i^k\big| + \big|\pi_j^k\big|} \to \sigma_{ij}^- \quad \text{with} \quad \sigma_{ij}^- \begin{cases} > 0, & \text{if } i \in \gamma \text{ or } j \in \gamma, \\ \in [0,2], & \text{if } i \in \beta \text{ and } j \in \beta, \\ 0, & \text{otherwise} \end{cases}$$

and, similarly,

$$\frac{\big|\pi_i^k\big| + \big|\pi_j^k\big| + \pi_i^k + \pi_j^k}{\big|\pi_i^k\big| + \big|\pi_j^k\big|} \to \sigma_{ij}^+ \quad \text{with} \quad \sigma_{ij}^+ \begin{cases} > 0, & \text{if } i \in \alpha \text{ or } j \in \alpha, \\ \in [0,2], & \text{if } i \in \beta \text{ and } j \in \beta, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, we have

$$\begin{array}{l} \sigma_{ij}^- = 0 \text{ if } (i,j) \in (\alpha \times \alpha) \cup (\alpha \times \beta) \cup (\beta \times \alpha), \\ \sigma_{ij}^+ = 0 \text{ if } (i,j) \in (\beta \times \gamma) \cup (\gamma \times \beta) \cup (\gamma \times \gamma). \end{array}$$

Further note that the limiting elements are not uniquely defined for all index pairs $(i,j) \in \beta \times \beta$, but that at least one of the elements $\sigma_{ij}^-, \sigma_{ij}^+$ is always positive.

Summarizing this discussion, we get the following result.

**Theorem 4.4** *Let $(X^*, \lambda^*, S^*)$ be a solution of the optimality conditions* (2)*, let $X^* - S^* = Q^*\Pi^*(Q^*)^T$ be a spectral decomposition of $X^* - S^*$, and let the index sets $\alpha, \beta, \gamma$ be defined as in* (30)*. Then every element from the B-subdifferential $\partial_B\Theta(X^*, \lambda^*, S^*)$ can, in matrix-vector notation, be written as*

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ V^*\Sigma_-^*(V^*)^T & 0 & V^*\Sigma_+^*(V^*)^T \end{pmatrix}$$

*with* $V^*$ *being the matrix from* (27), (28) *and with diagonal matrices*

$$\begin{aligned}\Sigma_-^* &= diag\big(\dots,\sigma_{ij}^-,\dots\big)_{1\le j\le i\le n},\\ \Sigma_+^* &= diag\big(\dots,\sigma_{ij}^+,\dots\big)_{1\le j\le i\le n}\end{aligned}$$

*having the following properties:*

$$\begin{aligned}&\Sigma_-^*\succeq 0,\ \ \Sigma_+^*\succeq 0,\ \ \Sigma_-^*+\Sigma_+^*=2I,\\ &\sigma_{ij}^-=0 \text{ if } (i,j)\in(\alpha\times\alpha)\cup(\alpha\times\beta)\cup(\beta\times\alpha),\\ &\sigma_{ij}^+=0 \text{ if } (i,j)\in(\beta\times\gamma)\cup(\gamma\times\beta)\cup(\gamma\times\gamma).\end{aligned}$$

Motivated by these considerations, we next state a slightly modified nondegeneracy condition.

**Assumption 4.5** *Let* $(X^*,\lambda^*,S^*)$ *be a solution of the optimality conditions* (2), $X^*-S^*=Q^*\Pi^*(Q^*)^T$ *be a spectral decomposition of* $X^*-S^*$, $\alpha,\beta,\gamma$ *the index sets from* (30), *and define the corresponding matrix* $V^*$ *as in* (27), (28).

*(A.4) Given diagonal matrices*

$$\begin{aligned}\Sigma_- &= diag\big(\dots,\sigma_{ij}^-,\dots\big)_{1\le j\le i\le n},\\ \Sigma_+ &= diag\big(\dots,\sigma_{ij}^+,\dots\big)_{1\le j\le i\le n}\end{aligned}$$

*such that*

$$\begin{aligned}&\Sigma_-\succeq 0,\ \ \Sigma_+\succeq 0,\ \ \Sigma_-+\Sigma_+\succ 0,\\ &\sigma_{ij}^-=0 \text{ if } (i,j)\in(\alpha\times\alpha)\cup(\alpha\times\beta)\cup(\beta\times\alpha),\\ &\sigma_{ij}^+=0 \text{ if } (i,j)\in(\beta\times\gamma)\cup(\gamma\times\beta)\cup(\gamma\times\gamma),\end{aligned}$$

*the following implication holds for any triple* $\big(\mathrm{svec}(\Delta X),\Delta\lambda,\mathrm{svec}(\Delta S)\big)$*:*

$$\left.\begin{array}{l}\mathcal{A}^T\Delta\lambda+\mathrm{svec}(\Delta S)=0\\ \mathcal{A}\,\mathrm{svec}(\Delta X)=0,\\ V^*\Sigma_-(V^*)^T\,\mathrm{svec}(\Delta X)+V^*\Sigma_+(V^*)^T\,\mathrm{svec}(\Delta S)=0\end{array}\right\}\Longrightarrow\left\{\begin{array}{l}\mathrm{svec}(\Delta X)=0,\\ \mathrm{svec}(\Delta S)=0.\end{array}\right.$$

In order to see the difference between our Assumption (A.4) and the nondegeneracy condition from Assumption (A.2), we first note that the precise value of the diagonal matrices $\Sigma_-$ and $\Sigma_+$ from Assumption (A.4) are unimportant. The only important thing is whether a diagonal element is zero or positive. Taking this into account and using Lemma 4.2 (c), it follows that Assumption (A.2) corresponds to the case where $\Sigma_-=I\otimes_s D_{S^*}$ and $\Sigma_+=I\otimes_s D_{X^*}$, where we used the notation from Lemma 4.2. However, if the index set $\beta$ is nonempty, it is not difficult to see that the assumption $\Sigma_-+\Sigma_+\succ 0$ does not hold in this case for all diagonal elements such that $(i,j)\in\beta\times\beta$. This is the main difference between our nondegeneracy condition and the one from Assumption (A.2). This difference is also the main reason why we are able to prove local quadratic convergence without assuming strict complementarity. This is the main consequence of the following result.

**Theorem 4.6** *Let $W^* = (X^*, \lambda^*, S^*)$ be a solution of the optimality conditions* (2) *satisfying (A.1) and (A.4). Then all elements $H^* \in \partial_B \Theta(W^*)$ are invertible.*

**Proof.** In view of our construction, each element $H^* \in \partial_B \Theta(W^*)$ has, in the usual matrix representation, the form of a matrix as given in Theorem 4.4. Taking into account that the two diagonal matrices $\Sigma_-^*$ and $\Sigma_+^*$ from Theorem 4.4 satisfy the conditions of Assumption (A.4), we can now follow the technique of proof from Theorem 4.3 in order to see that this matrix is indeed nonsingular under (A.1) and (A.4). □

As a direct consequence of Theorems 2.3 and 4.6, we obtain the following result.

**Theorem 4.7** *Let $(X^*, \lambda^*, S^*)$ be a solution of the optimality conditions* (2) *satisfying (A.1) and (A.4). Then Algorithm 2.2 is locally quadratically convergent.*

The next result also follows from Theorem 4.6 together with a result from [15].

**Corollary 4.8** *Let $(X^*, \lambda^*, S^*)$ be a solution of the optimality conditions* (2) *satisfying (A.1) and (A.4). Then $(X^*, \lambda^*, S^*)$ is the unique solution of* (2).

# 5 Illustrative Examples

In this section, we illustrate our theory developed in the previous section by using two examples. The first example is taken from [1] and contains a semidefinite program with a unique solution not satisfying strict complementarity, but where our nondegeneracy condition from Assumption (A.4) holds.

**Example 5.1** Let $n = 3$, $m = 3$ with $b = (1, 0, 0)^T$ and

$$C = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then Assumption (A.1) is obviously satisfied. Moreover, the corresponding semidefinite program has the unique solution

$$X^* = \operatorname{diag}(1, 0, 0), \quad \lambda^* = (0, 0, 0)^T, \quad S^* = \operatorname{diag}(0, 0, 1)$$

which does not satisfy the strict complementarity condition from (A.3). Moreover, it is not difficult to see that Assumption (A.2) is also violated. We now want to verify, however, that Assumption (A.4) holds. To this end, let

$$X^* - S^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} = Q^* \Pi^* (Q^*)^T \quad \text{with} \quad Q^* := I \in \mathbb{R}^{3 \times 3}$$

a spectral decomposition of $X^* - S^*$. Then

$$V^* = \big(v_{11}^*, v_{21}^*, v_{31}^*, v_{22}^*, v_{32}^*, v_{33}^*\big) = I \in \mathbb{R}^{6\times 6}.$$

Let $\Sigma_-, \Sigma_+$ be two diagonal matrices as described in Assumption (A.4). Noting that the index sets from (30) are given by $\alpha = \{1\}, \beta = \{2\}, \gamma = \{3\}$, it follows that

$$\begin{aligned} \Sigma_- &= \operatorname{diag}\big(0, 0, \sigma_{31}^-, \sigma_{22}^-, \sigma_{32}^-, \sigma_{33}^-\big), \\ \Sigma_+ &= \operatorname{diag}\big(\sigma_{11}^+, \sigma_{21}^+, \sigma_{31}^+, \sigma_{22}^+, 0, 0\big) \end{aligned} \tag{31}$$

for certain numbers satisfying

$$\begin{array}{l} \sigma_{31}^-, \sigma_{32}^-, \sigma_{33}^- > 0, \\ \sigma_{11}^+, \sigma_{21}^+, \sigma_{31}^+ > 0, \\ \sigma_{22}^-, \sigma_{22}^+ \geq 0, \sigma_{22}^- + \sigma_{22}^+ > 0. \end{array} \tag{32}$$

Now let $\Delta X, \Delta S \in \mathcal{S}^{n\times n}$ and $\Delta\lambda \in \mathbb{R}^m$ satisfy the equations

$$\begin{gathered} \mathcal{A}^T \Delta\lambda + \operatorname{svec}(\Delta S) = 0, \quad \mathcal{A}\operatorname{svec}(\Delta X) = 0, \\ V^*\Sigma_-(V^*)^T \operatorname{svec}(\Delta X) + V^*\Sigma_+(V^*)^T \operatorname{svec}(\Delta S) = 0, \end{gathered} \tag{33}$$

where

$$\mathcal{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{2} & 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 & 0 & 1 \end{pmatrix},$$

cf. (15). The system of equations (33) is equivalent to

$$\begin{aligned} \Delta\lambda_1 + \Delta S_{11} &= 0, & \Delta\lambda_3 + \Delta S_{21} &= 0, \\ \Delta\lambda_2 + \Delta S_{31} &= 0, & \Delta\lambda_2 + \Delta S_{22} &= 0, \\ \Delta S_{32} &= 0, & \Delta\lambda_3 + \Delta S_{33} &= 0, \\ \Delta X_{11} &= 0, & 2\Delta X_{31} + \Delta X_{22} &= 0, \\ 2\Delta X_{21} + \Delta X_{33} &= 0, & \sigma_{11}^+\Delta S_{11} &= 0, \\ \sigma_{21}^+\Delta S_{21} &= 0, & \sigma_{31}^-\Delta X_{31} + \sigma_{31}^+\Delta S_{31} &= 0, \\ \sigma_{22}^-\Delta X_{22} + \sigma_{22}^+\Delta S_{22} &= 0, & \sigma_{32}^-\Delta X_{32} &= 0, \\ \sigma_{33}^-\Delta X_{33} &= 0. \end{aligned}$$

Depending on whether $\sigma_{22}^- > 0$ or $\sigma_{22}^+ > 0$, it is now an easy exercise to verify that these conditions imply $\Delta X = 0$ and $\Delta S = 0$, so that Assumptions (A.4) holds.

In order to see that we really get fast local convergence, we applied the Newton-type method from [9] (which, basically, is a globalized version of Algorithm 2.2) to Example 5.1. The corresponding numerical results are given in Table 1. The columns of that table contain the absolute value of the relative gap between primal and dual objective functions, the norm of $\Theta$ at the current iterate $W^k = (X^k, \lambda^k, S^k)$ and the distance of $W^k$ to the

| $k$ | relative gap | $\|\Theta(W^k)\|$ | $\|W^k - W^*\|$ |
|---|---|---|---|
| 0 | 0.000000e+00 | 8.833707e-01 | 9.106836e-01 |
| 1 | 5.377397e-02 | 1.779061e-01 | 1.965437e-01 |
| 2 | 3.438035e-05 | 5.372091e-03 | 5.556943e-03 |
| 3 | 3.270881e-08 | 7.046664e-05 | 1.814441e-04 |
| 4 | 2.105543e-09 | 4.864869e-06 | 1.260400e-05 |
| 5 | 1.043839e-11 | 2.860393e-07 | 7.432387e-07 |

Table 1: Numerical results for Example 5.1

solution $W^* = (X^*, \lambda^*, S^*)$. (Note that the relative gap is zero at the starting point, but not optimal since feasibility is not satisfied.)

The second example is taken from [11]. Also this example has a unique solution not satisfying strict complementarity. Here, however, our Assumption (A.4) does not hold.

**Example 5.2** Let $n = 3$, $m = 2$ with $b = (-1, 0)^T$ and

$$C = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 0.5 \\ 0 & -0.5 & 0 \\ 0.5 & 0 & -1 \end{pmatrix}.$$

Assumption (A.1) obviously holds for this example. Moreover, the corresponding semidefinite program has the unique solution

$$X^* = \operatorname{diag}(1, 0, 0), \quad \lambda^* = (0, 0)^T, \quad S^* = \operatorname{diag}(0, 0, 1),$$

so that strict complementarity is violated. Similar to the previous example, we get $V^* = I, \alpha = \{1\}, \beta = \{2\}, \gamma = \{3\}$. Hence, if $\Sigma_-, \Sigma_+$ denote two matrices as described in Assumption (A.4), they satisfy (31), (32). Now let $(\Delta X, \Delta\lambda, \Delta S)$ any triple such that (33) holds with

$$\mathcal{A} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -0.5 & 0 & -1 \end{pmatrix}.$$

Componentwise, this may be rewritten as

$$\begin{aligned} -\Delta\lambda_1 + \Delta S_{11} &= 0, & \Delta S_{21} &= 0, \\ \frac{1}{\sqrt{2}}\Delta\lambda_2 + \sqrt{2}\Delta S_{31} &= 0, & -\frac{1}{2}\Delta\lambda_2 + \Delta S_{22} &= 0, \\ \Delta S_{32} &= 0, & -\Delta\lambda_2 + \Delta S_{33} &= 0, \\ -\Delta X_{11} &= 0, & \Delta X_{31} - \frac{1}{2}\Delta X_{22} - \Delta X_{33} &= 0, \\ \sigma_{11}^+ \Delta S_{11} &= 0, & \sigma_{21}^+ \Delta S_{21} &= 0, \\ \sigma_{31}^- \Delta X_{31} + \sigma_{31}^+ \Delta S_{31} &= 0, & \sigma_{22}^- \Delta X_{22} + \sigma_{22}^+ \Delta S_{22} &= 0, \\ \sigma_{32}^- \Delta X_{32} &= 0, & \sigma_{33}^- \Delta X_{33} &= 0 \end{aligned}$$

with certain elements $\sigma_{ij}^-, \sigma_{ij}^+$ satisfying (32). This is a homogenous linear system with 14 equations in 14 unknowns. Since the two equations $\Delta S_{21} = 0$ and $\sigma_{21}^+ \Delta S_{21} = 0$ are linearly dependent, this system has a nontrivial solution. Hence Assumption (A.4) does not hold in this case.

Despite the fact that Example 5.2 does not satisfy Assumption (A.4), it turns out that the Newton-type method from [9] applied to this example is still locally fast convergent. This is illustrated in Table 2. Hence, although Assumption (A.4) is a sufficient condition for local quadratic convergence, this indicates that it might not be a necessary condition.

| $k$ | relative gap | $\|\Theta(W^k)\|$ | $\|W^k - W^*\|$ |
|---|---|---|---|
| 0 | 0.000000e+00 | 8.002975e-01 | 1.327358e+00 |
| 1 | 1.370620e-01 | 3.130563e-01 | 6.265862e-01 |
| 2 | 1.772473e-03 | 7.147265e-03 | 2.687633e-02 |
| 3 | 2.393343e-05 | 4.128274e-04 | 1.526783e-03 |
| 4 | 7.823405e-08 | 3.397737e-05 | 1.207830e-04 |
| 5 | 5.080087e-10 | 2.105440e-06 | 7.471116e-06 |

Table 2: Numerical results for Example 5.2

Note that the previous two examples use $n = 3$ in order to get a semidefinite program with a unique solution not satisfying strict complementarity. The question is whether it is possible to illustrate our theory using a smaller dimensional example with $n = 2$. According to our following result, this is not possible, at least not under the Slater constraint qualification which states that there is a triple $(\hat{X}, \hat{\lambda}, \hat{S})$ such that the conditions

$$\sum_{i=1}^{m} \hat{\lambda}_i A_i + \hat{S} = C, \quad A_i \bullet \hat{X} = b_i \, (i = 1, \dots, m), \quad \hat{X} \succ 0, \quad \hat{S} \succ 0.$$

are satisfied.

**Lemma 5.3** *If $n = 2$ and the Slater constraint qualification holds for (2), then the optimality conditions (2) always have a strictly complementary solution.*

**Proof.** First note that (2) always has a solution under the Slater condition. Therefore we can take an arbitrary solution $(X^*, \lambda^*, S^*)$. Assume this solution does not satisfy the strict complementarity condition. Then it is easy to see that $X^* = 0$ or $S^* = 0$. Now let $(\hat{X}, \hat{\lambda}, \hat{S})$ be a triple satisfying the Slater constraint qualification. If $X^* = 0$, it then follows that $(X^*, \hat{\lambda}, \hat{S})$ is a strictly complementary solution of (2). On the other hand, if $S^* = 0$, it follows that $(\hat{X}, \lambda^*, S^*)$ is a solution of (2) satisfying strict complementarity. Hence, in either case, we can find a strictly complementarity solution. □

## 6 Final Remarks

We have shown that strict complementarity is not needed for a Newton-type method to be locally quadratically convergent when applied to a suitable reformulation of semidefinite programs. In order to obtain such a result, however, we had to introduce a modified nondegeneracy condition. This nondegeneracy condition was pointed out to be different from the one stated in Assumption (A.2), which was taken from Kojima, Shida and Shindoh [11]. On the other hand, Alizadeh, Haeberly and Overton [2] used another nondegeneracy condition which is known to be equivalent to the one from [11] if strict complementarity holds but is different without strict complementarity. Unfortunately, we do not know in how far the nondegeneracy condition from [2] is related to our Assumption (A.4).

## References


[1] F. Alizadeh, J.-P. Haeberly and M.L. Overton: *Complementarity and nondegeneracy in semidefinite programming.* Mathematical Programming 77, 1997, pp. 111–128.

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

[3] X. Chen and P. Tseng: *Non-interior continuation methods for solving semidefinite complementarity problems.* Mathematical Programming 95, 2003, pp. 431–474.

[4] F. Facchinei and J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems I.* Springer-Verlag, New York, NY, 2003.

[5] F. Facchinei and J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems II.* Springer-Verlag, New York, NY, 2003.

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

[7] R.A. Horn and C.R. Johnson: *Matrix Analysis.* Cambridge University Press, 1985.

[8] R.A. Horn and C.R. Johnson: *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[9] C. Kanzow and C. Nagel: *Semidefinite programs: New search directions, smoothing-type methods, and numerical results.* SIAM Journal on Optimization 13, 2003, pp. 1–23.

[10] C. Kanzow and C. Nagel: *Some practical aspects of a Newton-type method for semidefinite programs.* Preprint 243, Institute of Applied Mathematics and Statistics, University of Würzburg, Germany, November 2001 (revised December 2002).

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

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

[13] R.D.C. Monteiro: *Polynomial convergence of primal-dual algorithms for semidefinite programming based on the Monteiro and Zhang family of directions.* SIAM Journal on Optimization 8, 1998, pp. 797–812.

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

[15] L. Qi: *Convergence analysis of some algorithms for solving nonsmooth equations.* Mathematics of Operations Research 18, 1993, pp. 227–244.

[16] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–367.

[17] D. Sun and J. Sun: *Semismooth matrix valued functions.* Mathematics of Operations Research 27, 2002, pp. 150–169.

[18] D. Sun, J. Sun and L. Qi: *From strong semismoothness of the squared smoothing matrix function to semidefinite complementarity problems.* Applied Mathematics Report AMR 00/20, School of Mathematics, University of New South Wales, Sydney, Australia, October 2000.

[19] J. Sun, D. Sun and L. Qi: *Quadratic convergence of a squared smoothing Newton method for nonsmooth matrix equations and its applications in semidefinite optimization problems.* Preprint, Department of Decision Sciences, National University of Singapore, Singapore, April 2002 (revision of [18]).

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

[21] P. Tseng: *Merit functions for semi-definite complementarity problems.* Mathematical Programming 83, 1998, pp. 159–185.

[22] Y. Zhang: *On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming.* SIAM Journal on Optimization 8, 1998, pp. 365–386.