SOME PRACTICAL ASPECTS OF A NEWTON-TYPE METHOD FOR SEMIDEFINITE PROGRAMS

To Jochem Zowe on the occasion of his 60th birthday.

Christian Kanzow and Christian Nagel

Preprint 243 November 2001

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

Abstract. We discuss some practical issues of a Newton-type method for the solution of semidefinite programs. This Newton-type method is obtained by reformulating the semidefinite program (or its optimality conditions) as a nonlinear system of equations $\Theta(W) = 0$, to which a modification of Newton's method can be applied. Two reformulations of this kind can be obtained via the smoothed minimum function and via the smoothed Fischer-Burmeister function. When applying Newton's method to $\Theta(W) = 0$, we have to solve at each iteration a linearized system $\nabla\Theta(W)\Delta W = -\Theta(W)$, and the solution of this linearized system is the main computational burden for the Newton-type method. In this manuscript, we therefore take a closer look at this linearized system and use different approaches in order to show how it can be decomposed in an appropriate way. It turns out that, in the end, one only has to solve a linear system of equations of relatively small size. This linear system, however, looks rather strange in the beginning, but we show that the corresponding coefficient matrix is symmetric and positive definite when the smoothed minimum function is used in the reformulation. Surprisingly, it turns out that the corresponding matrix is only positive definite (not necessarily symmetric) when the smoothed Fischer-Burmeister function is used. This difference implies that the underlying Newton-type method is significantly cheaper to carry out for the smoothed minimum function, not only because the linear system can be solved more efficiently, but also since because the (nontrivial) calculation of the coefficient matrix is much more expensive for the smoothed Fischer-Burmeister function due to the lack of symmetry.

Key Words. Semidefinite programs, Newton's method, smoothing methods, symmetric matrices, positive definite matrices.

1 Introduction

During the last few years, semidefinite programs have received much attention due to their wide applicability to combinatorial optimization, control theory and many engineering problems. The interested reader is referred to the recent book [2] by Ben-Tal and Nemirovski which describes many of these applications. Formally, a semidefinite program (in its primal form) is a minimization problem of the form

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

where $C, A_1, \dots, A_m \in \mathcal{S}^{n\times n}$ 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.

Corresponding to the primal problem (1), there is also the dual semidefinite program

$$\max_{\lambda \in \mathbb{R}^m, S \in \mathcal{S}^{n\times n}} b^T\lambda \quad \text{s.t.} \quad \sum_{i=1}^m \lambda_i A_i + S = C, \; S \succeq 0. \tag{2}$$

Similar to linear programs, there is a duality theory relating the primal and dual semidefinite programs. In particular, it follows from this duality theory that, under mild assumptions, both the primal problem (1) and the dual problem (2) have a solution if and only if the *optimality conditions*

$$\begin{array}{rcl} \sum_{i=1}^m \lambda_i A_i + S & = & C, \\ A_i \bullet X & = & b_i \quad \forall i = 1, \dots, m, \\ X \succeq 0, \; S \succeq 0, \; XS & = & 0 \end{array} \tag{3}$$

have a solution.

Therefore, many interior-point methods try to solve a semidefinite program by solving these optimality conditions, see, e.g., [1, 5, 10, 12, 14]. Usually this is done by following the *central path*

$$\tau \mapsto \big(X_\tau, \lambda_\tau, S_\tau\big)$$

for $\tau \to 0$, where $\big(X_\tau, \lambda_\tau, S_\tau\big)$ denotes a (typically unique) solution of the *central path conditions*

$$\begin{array}{rcl} \sum_{i=1}^m \lambda_i A_i + S & = & C, \\ A_i \bullet X & = & b_i \quad \forall i = 1, \dots, m, \\ X \succ 0, \; S \succ 0, \; XS & = & \tau^2 I, \end{array} \tag{4}$$

Note that the central path conditions are obtained by introducing a slight perturbation of the complementarity condition $XS = 0$ within (3) via a parameter $\tau > 0$.

Another class of methods, which is related to interior-point methods because it may also be viewed as a path-following method for the central path, is the class of *smoothing methods*. These smoothing methods have very recently been extended from linear programs and complementarity problems to semidefinite programs, see, e.g., [13, 3, 11, 9]. Basically, the idea of these smoothing methods is to reformulate either the optimality conditions (3) or the central path conditions (4) as a nonlinear system of equations $\Theta(W) = 0$, to which Newton's method or a suitable modification of it gets applied.

In Section 2, we review one of these methods which has been investigated and tested numerically by the authors in [9]. While [9] describes two methods based on two different reformulations, one based on the smoothed minimum function and one based on the smoothed Fischer-Burmeister function, we first consider the smoothed minimum function in Sections 2, 3, and 4. The extensions for the smoothed Fischer-Burmeister function will be discussed in Section 5.

When applying Newton's method to the reformulated system $\Theta(W) = 0$, we have to solve at each iteration a linearized system of the form $\nabla\Theta(W)\Delta W = -\Theta(W)$. The solution of this system is the main computational effort for Newton's method. While it has been shown in [9] that this system is always solvable, the structure of this system has not been investigated further in [9].

The aim of this paper is therefore to take a closer look at this structure. To this end, we use two different approaches, the first approach is described in Section 3 and based on the Lyapunov operator, the second one is presented in Section 4 and uses an ordinary matrix-vector formulation of the linearized system $\nabla\Theta(W)\Delta W = -\Theta(W)$ (note that W and ΔW include matrices rather than vectors). It turns out that this linearized system can be decomposed in such a way that we have to solve only one linear system of equations of dimension m, and the coefficient matrix of this system, which looks rather strange in the beginning, has very nice properties. In fact, we will show that it is symmetric and positive definite.

However, when we try to translate the investigations from Sections 3 and 4, which were carried out for the smoothed minimum function, to the smoothed Fischer-Burmeister function in Section 5, it turns out that the analysis does not go through. In fact, we obtain the somewhat surprising result that the corresponding matrix is still positive definite, but no longer symmetric (in general). As far as we are aware of, this is the first time that there is a significant theoretical difference between the smoothed minimum function and the smoothed Fischer-Burmeister function.

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

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

where $\mathrm{tr}(C) := \sum_{i=1}^n c_{ii}$ 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 $\mathbb{R}^{n\times n}$. (Warning: The related symbol $\circ$ is used for the composition of two mappings; it does not denote the Hadamard product of two matrices!) 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 positive semidefinite square root of A.

Finally, if $E \succ 0$ is a given symmetric positive definite matrix, the corresponding *Lyapunov operator* L_E is defined by

$$L_E[X] := EX + XE \quad (X \in \mathcal{S}^{n\times n}).$$

Then it is well-known (see [7]) 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 Newton-type Method

In this section, we review the Newton-type method from [9] (which, in turn, is similar to a previous method considered by Chen and Tseng [3]).

In order to derive this Newton-type method, let us introduce the function

$$\phi(X,S) := X + S - \big((X-S)^2\big)^{1/2} \quad (X,S \in \mathcal{S}^{n\times n}). \tag{5}$$

The mapping ϕ is usually called the *minimum function* since, for $n = 1$, it reduces to

$$\phi(x,s) = x + s - \sqrt{(x-s)^2} = x + s - |x-s| = 2\min\{x,s\}.$$

It was noted in [13, 9] that the minimum function has the following property:

$$\phi(X,S) = 0 \Longleftrightarrow X \succeq 0, S \succeq 0, XS = 0. \tag{6}$$

Hence ϕ may be used in an obvious way in order to reformulate the optimality conditions (3) as a nonlinear system of equations.

However, the mapping ϕ is nonsmooth, making a reformulation of the optimality conditions (3) less favourable from a numerical point of view since we cannot apply Newton's method directly. To overcome this problem, we use the following approximation of ϕ:

$$\phi(X,S,\tau) := X + S - \sqrt{(X-S)^2 + 4\tau^2 I}, \tag{7}$$

where, for the moment, $\tau > 0$ denotes a fixed parameter. This *smoothed minimum function* has the following properties.

Proposition 2.1 *Let ϕ be defined by* (7) *with $\tau > 0$ fixed. Then the following statements hold for any two matrices $X, S \in \mathcal{S}^{n\times n}$:*

(a) ϕ satisfies the equivalence

$$\phi(X,S,\tau) = 0 \Longleftrightarrow X \succ 0, S \succ 0, XS = \tau^2 I.$$

(b) ϕ is continuously differentiable (in the sense of Fréchet) with

$$\nabla\phi(X,S,\tau)(U,V,\mu) = U + V - L_E^{-1}\big[(X-S)(U-V) + (U-V)(X-S) + 8\tau\mu I\big],$$

where $E := \big((X-S)^2 + 4\tau^2 I\big)^{1/2}$.

The proof of part (a) was given in [9], whereas part (b) may be found in [3, 9]. Part (a) may be used in order to reformulate the central path conditions (4), whereas part (b) guarantees that ϕ from (7) is a smooth approximation of the minimum function from (5).

From now on, we will view the parameter τ in the definition of the smoothed minimum function (7) as an independent variable. The algorithm to be reviewed in this section is a Newton-type method applied to the nonlinear system of equations

$$\Theta(X, \lambda, S, \tau) = 0, \tag{8}$$

where

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

is defined by

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

Then it is easy to see from (6) that (X, λ, S) is a solution of the optimality conditions (3) if and only if (X, λ, S, τ) satisfies the nonlinear system of equations (8). (Note that the definition of Θ immediately gives $\tau = 0$.)

The Newton-type method for the system (8) is a predictor-corrector method. Global convergence of this method is achieved by following a suitable neighbourhood of the central path. The neighbourhood used here is given by

$$\mathcal{N}(\beta) = \left\{ (X, \lambda, S, \tau) \,\middle|\, A_i \bullet X = b_i \ \forall i = 1, \dots, m, \sum_{i=1}^m \lambda_i A_i + S = C, \|\phi(X, S, \tau)\|_F \leq \beta\tau \right\},$$

where β denotes a positive number. Local fast convergence will be guaranteed by using a suitable predictor step. In order to simplify the formulation of our algorithm, let us introduce the abbreviations

$$W := (X, \lambda, S) \qquad \text{and} \qquad W^k := (X^k, \lambda^k, S^k),$$

where k denotes the iteration index. We now give a detailed statement of the Newton-type method from [9]. For some further explanations, the interested reader is referred to that reference.

Algorithm 2.2

(S.0) (Initializiation)

Choose $W^0 = (X^0, \lambda^0, S^0) \in \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n}$ *with*

$$\sum_{i=1}^m \lambda_i^0 A_i + S^0 = C \quad \text{and} \quad A_i \bullet X^0 = b_i \quad (i = 1, \dots, m).$$

Choose $\tau_0 > 0$, $\beta > 0$ *with* $\|\phi(X^0, S^0, \tau_0)\|_F \leq \beta\tau_0$ *and set* $k := 0$. *Choose* $\hat{\sigma}, \alpha_1, \alpha_2 \in (0, 1)$.

(S.1) (Predictor step)

Let $(\Delta W^k, \Delta\tau_k) = (\Delta X^k, \Delta\lambda^k, \Delta S^k, \Delta\tau_k) \in \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n} \times \mathbb{R}$ *be a solution of the system*

$$\nabla\Theta(W^k, \tau_k) \begin{pmatrix} \Delta W \\ \Delta\tau \end{pmatrix} = -\Theta(W^k, \tau_k). \tag{9}$$

If $\left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, 0)\right\|_F = 0$*: STOP.*
Otherwise, if $\left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \tau_k)\right\|_F > \beta\tau_k$*, then let*

$$\hat{W}^k := W^k, \quad \hat{\tau}_k := \tau_k \quad \textit{and} \quad \eta_k := 1,$$

else let $\eta_k = \alpha_1^s$*, where* s *is the natural number with*

$$\begin{aligned} \left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \alpha_1^r\tau_k)\right\|_F &\leq \beta\tau_k\alpha_1^r, \quad r = 0, 1, 2, \dots, s, \\ \left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \alpha_1^{s+1}\tau_k)\right\|_F &> \beta\tau_k\alpha_1^{s+1}, \end{aligned}$$

and set

$$\hat{\tau}_k := \eta_k\tau_k \quad \textit{and} \quad \hat{W}^k := \begin{cases} W^k, & \textit{if } s = 0, \\ W^k + \Delta W^k, & \textit{otherwise.} \end{cases}$$

(S.2) (Corrector step)
Let $(\Delta\hat{W}^k, \Delta\hat{\tau}_k) = (\Delta\hat{X}^k, \Delta\hat{\lambda}^k, \Delta\hat{S}^k, \Delta\hat{\tau}_k)$ *be a solution of*

$$\nabla\Theta(\hat{W}^k, \hat{\tau}_k)\begin{pmatrix} \Delta\hat{W} \\ \Delta\hat{\tau} \end{pmatrix} = -\Theta(\hat{W}^k, \hat{\tau}_k) + \begin{pmatrix} 0 \\ (1-\hat{\sigma})\hat{\tau}_k \end{pmatrix}. \tag{10}$$

Let $\hat{\eta}_k$ *be the maximum of the numbers* $1, \alpha_2, \alpha_2^2, \dots$ *with*

$$\left\|\phi(\hat{X}^k + \hat{\eta}_k\Delta\hat{X}^k, \hat{S}^k + \hat{\eta}_k\Delta\hat{S}^k, \hat{\tau}_k + \hat{\eta}_k\Delta\hat{\tau}_k)\right\|_F \leq (1 - \hat{\sigma}\hat{\eta}_k)\beta\hat{\tau}_k. \tag{11}$$

Set

$$W^{k+1} := \hat{W}^k + \hat{\eta}_k\Delta\hat{W}^k, \; \tau_{k+1} := (1 - \hat{\sigma}\hat{\eta}_k)\hat{\tau}_k, \; k \leftarrow k+1,$$

and go to (S.1).

Note that Algorithm 2.2 assumes that we solve two Newton systems at each iteration, one in the predictor step and one in the corrector step, with possibly different matrices $\nabla\Theta(W, \tau)$, and this is more costly than what is usually done by interior-point methods. However, all convergence results remain true for the following modification of Algorithm 2.2: If the predictor step has been accepted with $\eta_k < 1$, then skip the corrector step, i.e., set $W^{k+1} := W^k + \Delta W^k, \tau_{k+1} := \eta_k\tau_k, k \leftarrow k+1$, and return to (S.1). This modified algorithm either has to solve only one Newton system (in the predictor step), or it has to solve two systems, but then these two systems involve the same Jacobian of Θ.

Apart from this modification, it is also possible to replace the Newton system (9) in the predictor step (S.1) by

$$\nabla\Theta(W^k, \tau_k)\begin{pmatrix} \Delta W \\ \Delta\tau \end{pmatrix} = -\Theta(W^k, 0), \tag{12}$$

the difference to (9) being that τ_k got replaced by 0 on the right-hand side.

We next want to summarize the properties of Algorithm 2.2. To this end, we formulate a set of assumptions.

Assumption 2.3 *Let* (X^*, λ^*, S^*) *be a solution of the optimality conditions* (3)*.*

(a) *(Linear independence)*
The matrices A_i $(i = 1, \dots, m)$ *are linearly independent, i.e.,*

$$\sum_{i=1}^{m} \alpha_i A_i = 0 \quad \wedge \quad \alpha_i \in \mathbb{R} \quad \Longrightarrow \quad \alpha_i = 0 \ \forall i = 1, \dots, m.$$

(b) *(Strict complementarity)*
$X^* + S^* \succ 0$;

(c) *(Nondegeneracy)*
For any $(\Delta X, \Delta\lambda, \Delta S)$ *satisfying*

$$\sum_{i=1}^{m} \Delta\lambda_i A_i + \Delta S = 0 \quad \text{and} \quad A_i \bullet \Delta X = 0 \ (i = 1, \dots, m),$$

the following implication holds:

$$X^* \Delta S + \Delta X S^* = 0 \Longrightarrow (\Delta X, \Delta S) = (0, 0).$$

We next summarize the properties shown in [9] for the Newton-type method from Algorithm 2.2.

Theorem 2.4 *Let Assumption 2.3 (a) be satisfied. Then the following statements hold:*

(a) *Algorithm 2.2 is well-defined, in particular, the Newton systems* (9) *(or, alternatively,* (12)*) and* (10) *have a unique solution.*

(b) *All iterates* (W^k, τ_k) *and* $(\hat{W}^k, \hat{\tau}_k)$ *generated by Algorithm 2.2 belong to the neighbourhood* $\mathcal{N}(\beta)$.

(c) *Algorithm 2.2 is globally convergent in the sense that every accumulation point of the sequence* $\{W^k\}$ *is a solution of the optimality conditions* (3).

(d) *If, in addition, Assumptions 2.3 (b) and (c) hold, then Algorithm 2.2 converges locally superlinearly in the sense that* $\tau_{k+1} = o(\tau_k)$.

Note that Theorem 2.4, in particular, guarantees that the Newton system (9) in the predictor step (or, alternatively, the variant from (12)) and the Newton system (10) from the corrector step have a unique solution. However, knowing that there is such a solution and computing this solution are two different things. On the other hand, the computation of these solutions is of significant practical interest, especially because it is the main computational effort we have to carry out when applying Algorithm 2.2 to a semidefinite program.

In the following two sections, we therefore provide some strategies for the solution of the Newton systems (9) (or (12)) and (10). We also prove some interesting properties of these Newton systems.

3 Lyapunov-Formulation of Newton Systems

In this section, we show that the Newton systems (9), (12), and (10) may be solved by using the inverse Lyapunov function. To this end, we consider the Newton system (12) and provide all the details for this system. The arguments for (9) and (10) are similar and therefore not considered here.

Let us introduce the residuals

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

Then the Newton system (12) becomes

$$\sum_{j=1}^{m} \Delta\lambda_j A_j + \Delta S = R_C, \tag{13}$$

$$A_i \bullet \Delta X = r_{b,i} \qquad (i = 1, \ldots, m), \tag{14}$$

$$\nabla\phi(X, S, \tau)(\Delta X, \Delta S, \Delta\tau) = -\phi(X, S, 0), \tag{15}$$

$$\Delta\tau = 0, \tag{16}$$

where we dropped the iteration index k in order to simplify the notation. Using Proposition 2.1 (b) and taking into account (16), it follows from (15) that

$$\Delta X + \Delta S - L_E^{-1}\big[(X - S)(\Delta X - \Delta S) + (\Delta X - \Delta S)(X - S)\big] = -\phi(X, S, 0), \tag{17}$$

where, of course,

$$E := \big((X - S)^2 + 4\tau^2 I\big)^{1/2}$$

is the symmetric positive definite matrix from Proposition 2.1. Applying the Lyapunov operator L_E on both sides of (17) yields

$$L_E[\Delta X] + L_E[\Delta S] - (X - S)(\Delta X - \Delta S) - (\Delta X - \Delta S)(X - S) = -L_E\big[\phi(X, S, 0)\big].$$

Rearranging terms gives

$$L_{E-(X-S)}[\Delta X] + L_{E+(X-S)}[\Delta S] = -L_E\big[\phi(X, S, 0)\big].$$

Using the notation

$$A_E := E - (X - S) \quad \text{and} \quad B_E := E + (X - S), \tag{18}$$

this equation may be rewritten as

$$L_{A_E}[\Delta X] + L_{B_E}[\Delta S] = -L_E\big[\phi(X, S, 0)\big]. \tag{19}$$

Noting that A_E from (18) is symmetric positive definite (the same holds for B_E), we obtain

$$\Delta X = -L_{A_E}^{-1}\Big[L_{B_E}[\Delta S] + L_E\big[\phi(X,S,0)\big]\Big]. \tag{20}$$

Substituting ΔS from (13) and rearranging terms yields

$$\Delta X = \sum_{j=1}^{m} \Delta\lambda_j L_{A_E}^{-1}\big[L_{B_E}[A_j]\big] - L_{A_E}^{-1}\Big[L_{B_E}[R_C] + L_E\big[\phi(X,S,0)\big]\Big].$$

Taking inner products with A_i $(i = 1, \dots, m)$ and using (14), we obtain

$$\begin{aligned} \textstyle\sum_{j=1}^{m} \Delta\lambda_j L_{A_E}^{-1}\big[L_{B_E}[A_j]\big] \bullet A_i &= A_i \bullet \Delta X + L_{A_E}^{-1}\Big[L_{B_E}[R_C] + L_E\big[\phi(X,S,0)\big]\Big] \bullet A_i \\ &= r_{b,i} + L_{A_E}^{-1}\Big[L_{B_E}[R_C] + L_E\big[\phi(X,S,0)\big]\Big] \bullet A_i \end{aligned} \tag{21}$$

for $i = 1, \dots, m$. This expression may be reformulated further by applying the following result which states some properties of the Lyapunov operator.

Lemma 3.1 *Let $A, B \in \mathcal{S}_{++}^{n\times n}$ be two symmetric positive definite matrices and L_A, L_B be the corresponding Lyapunov operators, with L_A^{-1}, L_B^{-1} denoting their inverses. Then the following statements hold:*

(a) L_A, L_B, L_A^{-1}, and L_B^{-1} are self-adjoint.

(b) $L_A^{-1} \circ L_B$ and $L_B^{-1} \circ L_A$ are strongly monotone.

(c) If A and B commute, then $L_A \circ L_B = L_B \circ L_A$.

Proof. Statements (a) and (b) have been shown in, e.g., [9]. Hence we consider part (c) only. Let A and B two commuting matrices, i.e., $AB = BA$. Then we obtain

$$\begin{aligned} \big(L_A \circ L_B\big)[X] &= L_A\big[L_B[X]\big] \\ &= L_A\big[BX + XB\big] \\ &= A\big(BX + XB\big) + \big(BX + XB\big)A \\ &= ABX + AXB + BXA + XBA \\ &= BAX + BXA + AXB + XAB \\ &= B\big(AX + XA\big) + \big(AX + XA\big)B \\ &= \big(L_B \circ L_A\big)[X] \end{aligned}$$

for any matrix $X \in \mathcal{S}^{n\times n}$. Consequently, we have $L_A \circ L_B = L_B \circ L_A$. (Note that statement (c) is independent of the positive definiteness of A and B.) □

Applying Lemma 3.1 with $A := A_E$ and $B := B_E$ and recalling that these two matrices from (18) are symmetric positive definite, we see that (21) is equivalent to

$$\sum_{j=1}^{m} \Delta\lambda_j L_{B_E}[A_j] \bullet L_{A_E}^{-1}[A_i] = r_{b,i} + \Big(L_{B_E}[R_C] + L_E\big[\phi(X,S,0)\big]\Big) \bullet L_{A_E}^{-1}[A_i], \quad i = 1, \dots, m\,. \tag{22}$$

This is a linear equation in the variables $\Delta\lambda \in \mathbb{R}^m$ with coefficient matrix $M \in \mathbb{R}^{m\times m}$ defined elementwise by

$$m_{ij} := L_{B_E}[A_j] \bullet L_{A_E}^{-1}[A_i] \quad (i, j = 1, \ldots, m). \tag{23}$$

After solving this system, we immediately get ΔS from (13). Note that ΔS is obviously symmetric since R_C and all A_i are symmetric. In view of (20), ΔX can then be obtained as a solution of a Lyapunov equation with a symmetric right-hand side and is therefore also symmetric, cf. [7, Theorem 2.2.3]. The solution of this Lyapunov equation may be computed by using a spectral decomposition, see [7, p. 100].

We next take a closer look at the matrix $M = (m_{ij})$ defined by (23). Although, in the beginning, this matrix looks rather ugly, we will show that the matrix M has very nice properties. In fact, it is the aim of this section to show that M is symmetric and positive definite. The symmetry plays a crucial role because it implies that we can save a significant amount of work in calculating the matrix elements m_{ij}. Knowing that M is (symmetric and) positive definite, on the other hand, allows us to apply a Cholesky factorization (or a conjugate gradient method) in order to solve the linear system (22) for $\Delta\lambda$.

Surprisingly, it seems easier to prove the positive definiteness of M than its symmetry. We therefore begin with the following result.

Theorem 3.2 *Suppose Assumption 2.3 (a) holds. Then the matrix $M \in \mathbb{R}^{m\times m}$ with entries m_{ij} given by (23) is positive definite.*

Proof. Let $d \in \mathbb{R}^m$ be an arbitrary vector. Then

$$\begin{aligned}
d^T M d &= \sum_{i,j=1}^{m} d_i d_j m_{ij} \\
&= \sum_{i,j=1}^{m} d_i d_j \big(L_{B_E}[A_j] \bullet L_{A_E}^{-1}[A_i]\big) \\
&= \sum_{i,j=1}^{m} L_{B_E}[d_j A_j] \bullet L_{A_E}^{-1}[d_i A_i]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i,j=1}^{m} \big(L_{A_E}^{-1} \circ L_{B_E}\big)[d_j A_j] \bullet [d_i A_i] \\
&= \sum_{j=1}^{m} \big(L_{A_E}^{-1} \circ L_{B_E}\big)[d_j A_j] \bullet \Big(\sum_{i=1}^{m} d_i A_i\Big) \\
&= \big(L_{A_E}^{-1} \circ L_{B_E}\big)\Big[\sum_{j=1}^{m} d_j A_j\Big] \bullet \Big[\sum_{i=1}^{m} d_i A_i\Big] \\
&\geq 0
\end{aligned}$$

due to Lemma 3.1 (a), (b). Moreover, Lemma 3.1 (b) implies that equality can hold only for $\sum_{i=1}^{m} d_i A_i = 0$, and this immediately implies $d_i = 0$ for all $i = 1, \ldots, m$ in view of the assumed linear independence of the matrices $A_1, \dots, A_m$. $\square$

We now want to show that the matrix M is also symmetric. To this end, we recall the following standard result, see, e.g., [6].

Proposition 3.3 *Let $A, B \in \mathcal{S}^{n\times n}$ be two symmetric matrices. Then the following two statements are equivalent:*

(a) A and B commute.

(b) A and B have a simultaneous spectral decomposition, i.e., there is an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ and two diagonal matrices $D_A, D_B \in \mathbb{R}^{n\times n}$ such that $A = Q^T D_A Q$ and $B = Q^T D_B Q$.

In order to apply Lemma 3.1 (c), we next show that the two matrices A_E and B_E from (18) commute.

Lemma 3.4 *The two matrices $A_E = E - (X - S)$ and $B_E = E + (X - S)$ from (18) commute, where $E := \big((X - S)^2 + 4\tau^2 I\big)^{1/2}$.*

Proof. We have to show that $A_E B_E = B_E A_E$. In order to verify this equality, we invoke Proposition 3.3 and show that the two matrices A_E and B_E have a simultaneous spectral decomposition. To this end, let

$$X - S = Q^T D Q$$

be a spectral decomposition of the symmetric matrix $X - S$. Then

$$\big(X - S\big)^2 = Q^T D^2 Q$$

is a spectral decomposition of $\big(X - S\big)^2$, and we therefore obtain

$$\big(X - S\big)^2 + 4\tau^2 I = Q^T \big(D^2 + 4\tau^2 I\big) Q$$

and

$$E = \Big(\big(X - S\big)^2 + 4\tau^2 I\Big)^{1/2} = Q^T \big(D^2 + 4\tau^2 I\big)^{1/2} Q.$$

Consequently, we get

$$A_E = E - \big(X - S\big) = Q^T \Big(\big(D^2 + 4\tau^2 I\big)^{1/2} - D\Big) Q$$

and

$$B_E = E + \big(X - S\big) = Q^T \Big(\big(D^2 + 4\tau^2 I\big)^{1/2} + D\Big) Q,$$

showing that A_E and B_E have a simultaneous spectral decomposition. $\square$

We are now in the position to prove the symmetry of the matrix M.

Theorem 3.5 *Let $M = (m_{ij})$ be the matrix with entries m_{ij} from (23). Then M is symmetric.*

Proof. According to Lemma 3.4, the two symmetric positive definite matrices $A_E = E-(X-S)$ and $B_E = E+(X-S)$ commute. Hence Lemma 3.1 (c) implies that $L_{A_E} \circ L_{B_E} = L_{B_E} \circ L_{A_E}$ or, equivalently,

$$L_{A_E}^{-1} \circ L_{B_E} = L_{B_E} \circ L_{A_E}^{-1}. \tag{24}$$

Since (by Lemma 3.1 (a)) the symmetry of M may be rewritten as

$$\begin{aligned} m_{ij} = m_{ji} \quad &\Longleftrightarrow \quad L_{B_E}[A_j] \bullet L_{A_E}^{-1}[A_i] = L_{B_E}[A_i] \bullet L_{A_E}^{-1}[A_j] \\ &\Longleftrightarrow \quad \big(L_{A_E}^{-1} \circ L_{B_E}\big)[A_j] \bullet A_i = A_i \bullet \big(L_{B_E} \circ L_{A_E}^{-1}\big)[A_j] \\ &\Longleftrightarrow \quad \big(L_{A_E}^{-1} \circ L_{B_E}\big)[A_j] \bullet A_i = \big(L_{B_E} \circ L_{A_E}^{-1}\big)[A_j] \bullet A_i, \end{aligned}$$

the symmetry of M therefore follows from (24). $\square$

4 Matrix-Vector-Formulation of Newton Systems

In this section, we derive a formulation of the Newton systems (9), (12), and (10) as an ordinary linear system, i.e., we reformulate these Newton systems as matrix vector-products. To this end, 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

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

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

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

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.,

$$A \bullet B = \operatorname{svec}(A)^T \operatorname{svec}(B) \quad \forall A, B \in \mathcal{S}^{n\times n}. \tag{25}$$

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

$$\big(G \otimes_s K\big)\mathrm{svec}(H) := \frac{1}{2}\mathrm{svec}\big(KHG^T + GHK^T\big) \quad (S \in \mathcal{S}^{n\times n}). \tag{26}$$

Alternatively, the symmetric Kronecker product can also be defined by using the Kronecker product directly, see [12] for details.

Some properties of the symmetric Kronecker product are summarized in the following result. The proofs of these properties are elementary and may be found, among other things, in the appendix of the paper [12] by Todd, Toh, and Tütüncü.

Lemma 4.1 *The symmetric Kronecker product $\otimes_s$ defined by* (26) *has the following properties:*

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

(b) $G \otimes_s I$ *is symmmetric if and only if* G *is.*

(c) $\big(G \otimes_s K\big)\big(H \otimes_s L\big) = \frac{1}{2}\big(GH \otimes_s KL + GL \otimes_s KH\big)$.

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

We now consider the Newton system (12), i.e., we consider the system (13)–(16) from the previous section. The arguments for the Newton systems (9) and (10) are similar and therefore not presented here.

The first two equations (13) and (14) may be reformulated in matrix-vector notation in exactly the same way as described in [12], resulting in the two equations

$$\mathcal{A}^T\Delta\lambda + \mathrm{svec}(\Delta S) = \mathrm{svec}(R_C) \tag{27}$$

and

$$\mathcal{A}\,\mathrm{svec}(\Delta X) = r_b, \tag{28}$$

respectively, where

$$\mathcal{A} := \big(\mathrm{svec}(A_1), \dots, \mathrm{svec}(A_m)\big)^T \in \mathbb{R}^{m\times\frac{n(n+1)}{2}}. \tag{29}$$

Hence it remains to consider the third block (15). Following the previous section, we may reformulate this equation as

$$A_E\Delta X + \Delta X A_E + B_E\Delta S + \Delta S B_E = -L_E\big[\phi(X,S,0)\big],$$

cf. (19), where, of course, A_E and B_E denote the two symmetric and positive definite matrices from (18). Applying $\frac{1}{2}$svec on both sides then gives

$$\frac{1}{2}\mathrm{svec}\big(A_E\Delta X + \Delta X A_E\big) + \frac{1}{2}\mathrm{svec}\big(B_E\Delta S + \Delta S B_E\big) = -\frac{1}{2}\mathrm{svec}\big(L_E\big[\phi(X,S,0)\big]\big).$$

Using the definition (26) of svec, we have

$$\begin{aligned}\frac{1}{2}\text{svec}\big(A_E\Delta X + \Delta X A_E\big) &= \big(I\otimes_s A_E\big)\text{svec}(\Delta X),\\ \frac{1}{2}\text{svec}\big(B_E\Delta S + \Delta S B_E\big) &= \big(I\otimes_s B_E\big)\text{svec}(\Delta S).\end{aligned}$$

Setting

$$\mathcal{E} := I\otimes_s A_E \quad \text{and} \quad \mathcal{F} := I\otimes_s B_E, \tag{30}$$

we therefore get

$$\mathcal{E}\text{svec}(\Delta X) + \mathcal{F}\text{svec}(\Delta S) = -\frac{1}{2}\text{svec}\big(L_E\big[\phi(X,S,0)\big]\big). \tag{31}$$

Summarizing our discussion, we obtain the following result as a consequence of (27), (28), and (31).

Theorem 4.2 *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* (12) *if and only if the vector* $\big(svec(\Delta X), \Delta\lambda, svec(\Delta S)\big)$ *satisfies the linear system of equations*

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \mathcal{E} & 0 & \mathcal{F}\end{pmatrix}\begin{pmatrix} svec(\Delta X) \\ \Delta\lambda \\ svec(\Delta S)\end{pmatrix} = \begin{pmatrix} svec(R_C) \\ r_b \\ -\frac{1}{2}svec\big(L_E\big[\phi(X,S,0)\big]\big)\end{pmatrix}. \tag{32}$$

Note that the linear system (32) looks very similar to the one obtained for interior-point methods by Todd, Toh, and Tütüncü [12], however, the reader should be careful because the matrices $\mathcal{E}$ and $\mathcal{F}$ have a different meaning here.

Our next aim is to show that the linear system (32) has a unique solution, and how this solution can be computed. To this end, we need the following result.

Lemma 4.3 *The matrices* $\mathcal{E}$ *and* $\mathcal{F}$ *from* (30) *are symmetric positive definite. Furthermore, the matrix* $\mathcal{E}^{-1}\mathcal{F}$ *is also symmetric positive definite.*

Proof. We first recall that the two matrices A_E and B_E from (18) are symmetric positive definite. Hence it follows from the definitions of $\mathcal{E}$ and $\mathcal{F}$ in (30) together with Lemma 4.1 (a), (b), and (d) that $\mathcal{E}$ and $\mathcal{F}$ are also symmetric and positive definite.

Moreover, we know from Lemma 3.4 that the matrices A_E and B_E commute. Using Lemma 4.1 (c) and (a), this implies

$$\begin{aligned}\mathcal{E}\mathcal{F} &= \big(I\otimes_s A_E\big)\big(I\otimes_s B_E\big)\\ &= \frac{1}{2}\big(I\otimes_s A_E B_E + B_E\otimes_s A_E\big)\\ &= \frac{1}{2}\big(I\otimes_s B_E A_E + A_E\otimes_s B_E\big)\\ &= \big(I\otimes_s B_E\big)\big(I\otimes_s A_E\big)\\ &= \mathcal{F}\mathcal{E},\end{aligned}$$

so that $\mathcal{E}$ and $\mathcal{F}$ also commute. This, in turn, implies

$$\mathcal{F}\mathcal{E}^{-1} = \mathcal{E}^{-1}\mathcal{F}.$$

But then

$$\big(\mathcal{E}^{-1}\mathcal{F}\big)^T = \mathcal{F}^T\big(\mathcal{E}^{-1}\big)^T = \mathcal{F}\mathcal{E}^{-1} = \mathcal{E}^{-1}\mathcal{F},$$

i.e., the matrix $\mathcal{E}^{-1}\mathcal{F}$ is symmetric. Moreover, $\mathcal{E}^{-1}\mathcal{F}$ is positive definite since it is similar to the symmetric and positive definite matrix $\mathcal{F}^{1/2}\mathcal{E}^{-1}\mathcal{F}^{1/2}$, where $\mathcal{F}^{1/2}$ denotes the symmetric positive definite square root of $\mathcal{F}$. □

Using Lemma 4.3, we now obtain from Todd, Toh, and Tütüncü [12, Theorem 3.1] that the coefficient matrix of the linear system (32) is nonsingular.

Theorem 4.4 *Suppose Assumption 2.3 (a) holds. Then the matrix from* (32) *is nonsingular.*

We next show how our previous results may be used in order to solve the linear system (32). To this end, let us write down this 3×3 block system as

$$\mathcal{A}^T\Delta\lambda + \operatorname{svec}(\Delta S) = \operatorname{svec}(R_C), \tag{33}$$

$$\mathcal{A}\operatorname{svec}(\Delta X) = r_b, \tag{34}$$

$$\mathcal{E}\operatorname{svec}(\Delta X) + \mathcal{F}\operatorname{svec}(\Delta S) = -\frac{1}{2}\operatorname{svec}\big(L_E\big[\phi(X,S,0)\big]\big). \tag{35}$$

Since $\mathcal{E}$ is nonsingular by Lemma 4.3, we obtain from (35)

$$\operatorname{svec}(\Delta X) = -\mathcal{E}^{-1}\Big(\mathcal{F}\operatorname{svec}(\Delta S) + \frac{1}{2}\operatorname{svec}\big(L_E\big[\phi(X,S,0)\big]\big)\Big). \tag{36}$$

Substituting $\operatorname{svec}(\Delta S)$ from (33) gives

$$\operatorname{svec}(\Delta X) = -\mathcal{E}^{-1}\Big(\mathcal{F}\operatorname{svec}(R_C) - \mathcal{F}\mathcal{A}^T\Delta\lambda + \frac{1}{2}\operatorname{svec}\big(L_E\big[\phi(X,S,0)\big]\big)\Big).$$

Left-multiplication with $\mathcal{A}$ and using (34) yields

$$\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^T\Delta\lambda = r_b + \mathcal{A}\mathcal{E}^{-1}\Big(\mathcal{F}\operatorname{svec}(R_C) + \frac{1}{2}\operatorname{svec}\big(L_E\big[\phi(X,S,0)\big]\big)\Big). \tag{37}$$

The procedure for solving the linear system (32) is therefore as follows: First compute $\Delta\lambda$ from (37). The matrix of this linear system is symmetric and positive definite because of Lemma 4.3 and because $\mathcal{A}$ has full rank (under Assumption 2.3 (a)). Then we may obtain $\operatorname{svec}(\Delta X)$ and $\operatorname{svec}(\Delta S)$ from (36) and (33), respectively.

We next investigate the relation between the matrix

$$A := \mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^T \in \mathbb{R}^{m\times m} \tag{38}$$

occuring in (37) and the corresponding matrix M from Section 3 defined elementwise in (23). In fact, we claim that these two matrices are identical. To this end, let us calculate the entry a_{ij} of A. Using the definition of $\mathcal{A}$, we obtain

$$a_{ij} = \operatorname{svec}(A_i)^T\mathcal{E}^{-1}\mathcal{F}\operatorname{svec}(A_j).$$

Now it follows from (26) that

$$\mathcal{F}\text{svec}(A_j) = \big(I \otimes_s B_E\big)\text{svec}(A_j) = \frac{1}{2}\text{svec}\big(B_E A_j + A_j B_E\big) = \frac{1}{2}\text{svec}\big(L_{B_E}[A_j]\big).$$

Hence we obtain

$$a_{ij} = \frac{1}{2}\text{svec}(A_i)^T\big(I \otimes_s A_E\big)^{-1}\text{svec}(L_{B_E}[A_j]). \tag{39}$$

On the other hand, we have

$$A_E L_{A_E}^{-1}[A_i] + L_{A_E}^{-1}[A_i]A_E = A_i$$

in view of the definition of the inverse Lyapunov operator. Consequently, (26) yields

$$\begin{aligned} \frac{1}{2}\text{svec}(A_i) &= \frac{1}{2}\text{svec}\big(A_E L_{A_E}^{-1}[A_i] + L_{A_E}^{-1}[A_i]A_E\big) \\ &= \big(I \otimes_s A_E\big)\text{svec}\big(L_{A_E}^{-1}[A_i]\big). \end{aligned}$$

Hence we have

$$\frac{1}{2}\big(I \otimes_s A_E\big)^{-1}\text{svec}(A_i) = \text{svec}\big(L_{A_E}^{-1}[A_i]\big).$$

Substituting this into (39) and using (25) gives

$$\begin{aligned} a_{ij} &= \text{svec}\big(L_{A_E}^{-1}[A_i]\big)^T\text{svec}\big(L_{B_E}[A_j]\big) \\ &= \text{svec}\big(L_{B_E}[A_j]\big)^T\text{svec}\big(L_{A_E}^{-1}[A_i]\big) \\ &= L_{B_E}[A_j] \bullet L_{A_E}^{-1}[A_i]. \end{aligned}$$

However, the last expression is identical to m_{ij} from (23). Summarizing our previous discussion, we therefore get the following result.

Theorem 4.5 *The two matrices $M = (m_{ij})$ from* (23) *and $A = (a_{ij})$ from* (38) *are identical.*

5 Results for the Fischer-Burmeister Function

Algorithm 2.2 may also be applied to another operator Θ if we replace the smoothed minimum function from (7) by the mapping

$$\phi(X, S, \tau) := X + S - \big(X^2 + S^2 + 2\tau^2 I\big)^{1/2}. \tag{40}$$

For $n = 1$, this is the so-called *smoothed Fischer-Burmeister function*, see [4, 8]. It was noted in [9] that Proposition 2.1 (a) also holds for the smoothed Fischer-Burmeister function. Moreover, the smoothed Fischer-Burmeister function is also continuously differentiable with

$$\nabla\phi(X, S, \tau)(U, V, \mu) = U + V - L_E^{-1}\left[XU + UX + SV + VS + 4\tau\mu I\right], \tag{41}$$

where

$$E := \big(X^2 + S^2 + 2\tau^2 I\big)^{1/2}, \tag{42}$$

cf. [3, 9]. Now the question is which results still hold if we replace the smoothed minimum function everywhere by the smoothed Fischer-Burmeister function. Usually, these two functions are viewed as being equal in the sense that they have the same theoretical and similar numerical properties. Indeed, it was noted in [9] that the main convergence result from Theorem 2.4 also holds for the smoothed Fischer-Burmeister function.

Surprisingly, however, it turns out that not all of our new results from Sections 3 and 4 hold for the smoothed Fischer-Burmeister function. To this end, let us first consider the approach from Section 3. Using the expression (41), we may rewrite the block equation from (15) as

$$\Delta X + \Delta S - L_E^{-1}\big[X\Delta X + \Delta X X + S\Delta S + \Delta S S\big] = -\phi(X, S, 0),$$

where E denotes the matrix from (42). After some algebraic manipulations, similar to those in Section 3, we obtain

$$L_{A_E}[\Delta X] + L_{B_E}[\Delta S] = -L_E\big[\phi(X, S, 0)\big],$$

where

$$A_E := E - X \quad \text{and} \quad B_E := E - S. \tag{43}$$

Using these definitions of A_E and B_E, we see that we can compute $\Delta\lambda \in \mathbb{R}^m$ by solving the linear system (22) with matrix $M = (m_{ij}) \in \mathbb{R}^{m\times m}$ and elements m_{ij} defined by (23). Since A_E and B_E from (43) are obviously positive definite, it follows from Theorem 3.2 that the matrix M is positive definite also for the smoothed Fischer-Burmeister function.

However, the matrices A_E and B_E from (43) do not commute in general. This means that the approach used in Section 3 in order to show the symmetry of the matrix M is no longer applicable for the Fischer-Burmeister function. In fact, it turns out that the matrix M is not symmetric in general. To this end, consider the following counterexample.

Example 5.1 Choose $n = 3$, $m = 2$, $\tau = 1$, ϕ the smoothed Fischer-Burmeister function and

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Matrices A_i of this type occur in the constraints of the MAXCUT problem. We now consider the iterates

$$X = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad S = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

An easy computation shows that

$$E = (X^2 + S^2 + 2\tau^2 I)^{1/2} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

We therefore get

$$A_E = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B_E = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 2 \end{pmatrix}.$$

This yields

$$L_{B_E}[A_1] = \begin{pmatrix} 4 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad L_{B_E}[A_2] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Moreover, we have

$$L_{A_E}^{-1}[A_1] = \begin{pmatrix} \frac{7}{24} & \frac{1}{12} & 0 \\ \frac{1}{12} & \frac{1}{24} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad L_{A_E}^{-1}[A_2] = \begin{pmatrix} \frac{1}{24} & \frac{1}{12} & 0 \\ \frac{1}{12} & \frac{7}{24} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

because

$$\begin{aligned} & L_{A_E}^{-1}[A_1] \cdot A_E + A_E \cdot L_{A_E}^{-1}[A_1] \\ = & \begin{pmatrix} \frac{7}{24} & \frac{1}{12} & 0 \\ \frac{1}{12} & \frac{1}{24} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{7}{24} & \frac{1}{12} & 0 \\ \frac{1}{12} & \frac{1}{24} & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ = & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = A_1 \end{aligned}$$

and, similarly, for $L_{A_E}^{-1}[A_2]$. Then it is easy to verify that

$$M = \begin{pmatrix} \frac{7}{6} & \frac{1}{12} \\ \frac{1}{6} & \frac{7}{12} \end{pmatrix},$$

which is a non-symmetric matrix.

Now let us look at the results in Section 4, again with ϕ being the smoothed Fischer-Burmeister function from (40) and with A_E, B_E being the matrices from (43). Then we may follow all arguments from Section 4 up to Lemma 4.3. The proof of Lemm 4.3 is again based on the fact that A_E and B_E commute. Since this is no longer true for the Fischer-Burmeister function, it follows that the matrix $\mathcal{E}^{-1}\mathcal{F}$ is still positive definite, but not necessarily symmetric. Hence the corresponding linear system (32) still has a solution, and the matrix $\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^T$ from (37) is still positive definite, but Theorem 4.5 together with Example 5.1 shows that this matrix is no longer symmetric in general.

Consequently, the amount of work for computing either the matrix M from Section 3 or the matrix A from Section 4 is significantly cheaper when using the smoothed minimum function. Moreover, for the smoothed Fischer-Burmeister function it is no longer possible to apply a Cholesky factorization or a conjugate gradient method for the solution of these linear systems, i.e., the solution procedure itself is also more costly when using the smoothed Fischer-Burmeister function.

6 Final Remarks

In this manuscript we have shown that certain linearized systems occuring in the context of Newton-type methods for the solution of semidefinite programs can be decomposed in such a way that, in the end, only one linear system of equations of dimension m has to be solved. The corresponding matrix, although strange looking in the beginning, turned out to be symmetric positive definite when the smoothed minimum function is used in the Newton-type method, whereas it is only positive definite (not necessarily symmetric) when the smoothed Fischer-Burmeister function is taken.

These results suggest to use an inexact Newton method rather than Newton's method itself. However, this idea raises a couple of other interesting questions, namely how to compute the solution of Lyapunov equations by iterative solvers efficiently, and how to compute matrix square roots by iterative solvers. When answering these questions, one should take into account that, at least locally, our matrices do not differ much from one outer iteration to the next.

References

[1] 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.

[2] A. Ben-Tal and A. Nemirovski: *Lectures on Modern Convex Optimization.* MPS-SIAM Series on Optimization, SIAM, Philadelphia, PA, 2001.

[3] X. Chen and P. Tseng: *Non-interior continuation methods for solving semidefinite complementarity problems.* Technical Report, Department of Mathematics, University of Washington, Seattle, WA, 1999.

[4] A. Fischer: *A special Newton-type optimization method.* Optimization 24, 1992, pp. 269–284.

[5] 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.

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

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

[8] C. Kanzow: *Some noninterior continuation methods for linear complementarity problems.* SIAM Journal on Matrix Analysis and Applications 17, 1996, pp. 851–868.

[9] C. Kanzow and C. Nagel: *Semidefinite programs: New search directions, smoothing-type methods, and numerical results.* Preprint 163, Department of Mathematics, University of Hamburg, Hamburg, Germany, June 2001.

[10] Y. Nesterov and A. Nemirovskii: *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, Philadelphia, PA, 1994.

[11] D. Sun and J. Sun: *Semismooth matrix valued functions.* Technical Report, School of Mathematics, University of New South Wales, Sydney, Australia, November 1999.

[12] 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.

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

[14] 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.