Hamburger Beiträge
zur Angewandten Mathematik

Semidefinite Programs: New Search Directions, Smoothing-Type Methods, and Numerical Results

Christian Kanzow and Christian Nagel

Reihe A
Preprint 163
June 2001

Hamburger Beiträge zur Angewandten Mathematik

Reihe A	Preprints
Reihe B	Berichte
Reihe C	Mathematische Modelle und Simulation
Reihe D	Elektrische Netzwerke und Bauelemente
Reihe E	Scientific Computing
Reihe F	Computational Fluid Dynamics and Data Analysis

SEMIDEFINITE PROGRAMS: NEW SEARCH DIRECTIONS, SMOOTHING-TYPE METHODS, AND NUMERICAL RESULTS

Christian Kanzow and Christian Nagel

University of Hamburg
Department of Mathematics
Center for Optimization and Approximation
Bundesstrasse 55
20146 Hamburg
Germany

e-mail: kanzow@math.uni-hamburg.de
nagel@math.uni-hamburg.de

June 5, 2001

Abstract. Motivated by some results for linear programs and complementarity problems, this paper gives some new characterizations of the central path conditions for semidefinite programs. Exploiting these characterizations, some smoothing-type methods for the solution of semidefinite programs are derived. The search directions generated by these methods are automatically symmetric, and the overall methods are shown to be globally and locally superlinearly convergent under suitable assumptions. Some numerical results are also included which indicate that the proposed methods are very promising and comparable to several interior-point methods. Moreover, the current method seems to be superior to the recently proposed smoothing method by Chen and Tseng [8].

Key Words. Semidefinite programs, smoothing-type methods, Newton's method, global convergence, superlinear convergence.

1 Introduction

In this paper we describe an algorithm for the solution of semidefinite programs (SDPs). Using some standard notation that will be defined formally at the end of this section, a semidefinite program is a constrained optimization problem that is typically given in primal form by

$$\min C \bullet X \quad \text{s.t.} \quad A_i \bullet X = b_i,\ i = 1, \dots m,\ X \succeq 0 \tag{1}$$

or in its dual form by

$$\max b^T \lambda \quad \text{s.t.} \quad \sum_{i=1}^{m} \lambda_i A_i + S = C,\ S \succeq 0; \tag{2}$$

here, the vector $b \in \mathbb{R}^m$ as well as the symmetric matrices $C \in \mathbb{R}^{n \times n}$ and $A_i \in \mathbb{R}^{n \times n}$ ($i = 1, \dots, m$) are the given data, whereas the symmetric matrix $X \in \mathbb{R}^{n \times n}$ denotes the variable for the primal semidefinite program (1) and the vector $\lambda \in \mathbb{R}^m$ together with the symmetric matrix $S \in \mathbb{R}^{n \times n}$ denote the variables of the dual semidefinite program (2).

It is easy to see that the (primal) semidefinite program is a convex minimization problem. Under a suitable constraint qualification, this semidefinite program is therefore equivalent to its optimality conditions. These optimality conditions can be written as follows:

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

Motivated by the groundbreaking work of Nesterov and Nemirovskii [22], several authors suggest to solve the optimality conditions (3) by (primal-dual) interior-point methods. These interior-point methods typically consider the following perturbation of the optimality conditions (3), usually called the central path conditions:

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

where τ denotes a positive parameter (note that we parameterize the central path conditions by τ^2 instead of τ). Typical interior-point methods now apply a Newton-type method to (a symmetrized version of) the equations within the central path conditions, dealing with the $X \succ 0$ and $S \succ 0$ constraints explicitly by a suitable line search. The interested reader is referred to the papers [15, 2, 25, 29], for example.

The method to be discussed here is also a Newton-type method. However, before applying Newton's method, we first reformulate the optimality conditions or the central path conditions as a nonlinear system of equations. This reformulated system does not contain any explicit inequality constraints like $X \succeq 0, S \succeq 0$ or $X \succ 0, S \succ 0$, and Newton's method applied to this system will automatically generate symmetric search directions without any further transformations (unlike interior-point methods).

We believe that our method is of particular interest for the solution of some difficult combinatorial optimization problems. In fact, semidefinite programs are known to provide very good lower bounds for some combinatorial problems. However, solving such a semidefinite relaxation by an interior-point method within a branch-and-bound strategy may not result in the most efficient way to solve the underlying combinatorial problem since the solution of one semidefinite relaxation may not be used as a starting point for a neighbouring problem due to the fact that interior-point methods require strictly feasible starting points. On the other hand, the method to be presented here does not have such a restriction regarding its starting point.

Our method may be viewed as a generalization of some smoothing-type methods for linear programs and complementarity problems to the framework of semidefinite programs. While such a generalization has already been suggested in a recent paper by Chen and Tseng [8], we stress that there are still a couple of differences between that paper and ours. For example, we present a new characterization of the central path conditions which may be viewed as the basis for our method. Furthermore, our method is based on an essentially smooth reformulation of the optimality conditions (3) themselves (and this is what we really want to solve), while Chen and Tseng [8] consider a reformulation of the central path conditions. This may also explain why our approach seems to give better numerical results than the one from [8].

The organization of this paper is as follows. Section 2 contains some new characterizations of the central path conditions (4). These characterizations are based on a certain function ϕ whose further properties are discussed in Section 3. Our algorithm is described in Section 4, and its global and local convergence properties are analyzed in Sections 5 and 6, respectively. We then present some very promising numerical results in Section 7 and close this manuscript with some final remarks in Section 8.

Throughout this paper, we use the following notation: For two matrices $A, B \in \mathbb{R}^{n\times n}$, we define the scalar product

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

where $\operatorname{tr}(C) := \sum_{i=1}^n c_{ii}$ denotes the trace of a matrix $C \in \mathbb{R}^{n\times n}$. (Warning: The related symbol $\circ$ is used for the composition of two mappings; it does not denote the Hadamard product of two matrices!) We denote by $\mathcal{S}^{n\times n}, \mathcal{S}^{n\times n}_+$, and $\mathcal{S}^{n\times n}_{++}$ the sets of symmetric, symmetric positive semidefinite, and symmetric positive definite matrices of dimension $n \times n$, respectively. We also write $A \succeq 0$ and $A \succ 0$ in order to indicate that A belongs to $\mathcal{S}^{n\times n}_+$ and $\mathcal{S}^{n\times n}_{++}$, respectively. Furthermore, $A \succeq B$ or $A \succ B$ means that $A - B \succeq 0$ or $A - B \succ 0$. If $A \succeq 0$, we denote by $A^{1/2}$ the unique positive semidefinite square root of A. In our analysis, we will use both the spectral norm $\|A\|_2$ and the Frobenius norm $\|A\|_F$ for a matrix $A \in \mathbb{R}^{n\times n}$. We endow the vector space $\mathbb{R}^{n\times n} \times \mathbb{R}^m \times \mathbb{R}^{n\times n}$ with the norm

$$|||(X, \lambda, S)||| := \sqrt{\|X\|_F^2 + \|\lambda\|_2^2 + \|S\|_F^2}.$$

We use the same symbol for the norm

$$|||(X, \lambda, S, \tau)||| := \sqrt{\|X\|_F^2 + \|\lambda\|_2^2 + \|S\|_F^2 + \tau^2}$$

in the vector space $\mathbb{R}^{n\times n} \times \mathbb{R}^m \times \mathbb{R}^{n\times n} \times \mathbb{R}$. Note that these two norms correspond to the standard Euclidian norm if all entries in the underlying vector spaces are viewed as long vectors.

2 Reformulations of the Central Path

The aim of this section is to give two new reformulations of the central path conditions (4) for semidefinite programs. These reformulations can be obtained by generalizing existing reformulations for linear programs and complementarity problems in a suitable way.

Before we deal with the central path conditions (4), however, we first consider the optimality conditions (3). In order to motivate our approach, let us define a mapping $\varphi : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ by

$$\varphi(a, b) := a + b - \sqrt{a^2 + b^2}.$$

This mapping was introduced by Fischer [13] and is usually called the Fischer-Burmeister function. It is well-known (and easy to verify) that it has the following property:

$$\varphi(a, b) = 0 \Longleftrightarrow a \geq 0, b \geq 0, ab = 0. \tag{5}$$

Now let us define a mapping $\phi : \mathcal{S}^{n\times n} \times \mathcal{S}^{n\times n} \to \mathcal{S}^{n\times n}$ by

$$\phi(X, S) := X + S - (X^2 + S^2)^{1/2} \tag{6}$$

which is an obvious extension of the definition of φ with the arguments begin symmetric matrices rather than two real numbers. It has been shown by Tseng [27, Lemma 6.1] that the mapping ϕ has a property similar to (5), namely

$$\phi(X, S) = 0 \Longleftrightarrow X \succeq 0, S \succeq 0, XS = 0. \tag{7}$$

In the following, we will include a proof for this equivalence. We stress that our proof is somewhat different from the one given by Tseng [27] and that a similar technique will later be used in order to prove our new characterizations of the central path conditions. To verify the equivalence (7), we will exploit the following simple result from Alizadeh [1, Lemma 2.9].

Lemma 2.1 *Let $X, S \in \mathcal{S}_+^{n\times n}$ be two symmetric positive semidefinite matrices. Then $XS = 0$ if and only if $X \bullet S = 0$.*

Lemma 2.1 allows us to state the following result.

Proposition 2.2 *Let ϕ be the Fischer-Burmeister function defined in (6). Then*

$$\phi(X, S) = 0 \iff X \succeq 0, S \succeq 0, XS = 0.$$

Proof. First assume that $X \succeq 0, S \succeq 0, XS = 0$ holds. This implies $XS + SX = 0$ and therefore

$$(X + S)^2 = X^2 + S^2.$$

Using the fact that X and S are symmetric positive semidefinite, it follows that

$$X + S = (X^2 + S^2)^{1/2},$$

since the square root of a symmetric and positive semidefinite matrix is uniquely defined within the space of symmetric and positive semidefinite matrices. Obviously, this implies $\phi(X, S) = 0$.

Conversely, assume that $\phi(X, S) = 0$ holds for two symmetric matrices $X, S \in \mathcal{S}^{n\times n}$. This means that $X + S = (X^2 + S^2)^{1/2}$. Squaring both sides of this equation gives

$$X^2 + S^2 = (X + S)^2 \quad \text{and} \quad X + S \in \mathcal{S}_+^{n\times n}.$$

This is equivalent to

$$XS + SX = 0 \quad \text{and} \quad X + S \in \mathcal{S}_+^{n\times n}. \tag{8}$$

Let $X = Q^T DQ$ with $Q \in \mathbb{R}^{n\times n}$ orthogonal and $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ be the spectral decomposition of the symmetric matrix X. Then (8) can be rewritten as

$$Q^T DQS + SQ^T DQ = 0 \quad \text{and} \quad Q^T DQ + S \in \mathcal{S}_+^{n\times n}.$$

If we premultiply this equation by Q and postmultiply it by Q^T, we obtain

$$DQSQ^T + QSQ^T D = 0 \quad \text{and} \quad D + QSQ^T \in \mathcal{S}_+^{n\times n}.$$

Using the abbreviation $A := QSQ^T$, we get

$$DA + AD = 0 \quad \text{and} \quad D + A \in \mathcal{S}_+^{n\times n}. \tag{9}$$

Componentwise, this can be rewritten as

$$(\lambda_i + \lambda_j)a_{ij} = 0 \quad \text{and} \quad D + A \in \mathcal{S}_+^{n\times n} \tag{10}$$

for all $i, j = 1, \dots, n$. In particular, we obtain for $i = j$

$$2\lambda_i a_{ii} = 0 \quad \text{and} \quad \lambda_i + a_{ii} \geq 0$$

for all $i = 1, \dots, n$. Obviously, this implies $\lambda_i \geq 0$ for all $i = 1, \dots, n$, which in turn means that X is positive semidefinite.

Using a symmetric argument (based on a spectral decomposition of S), we see that S is also positive semidefinite.

In order to verify that $XS = 0$, we observe that (8) implies

$$X \bullet S = \mathrm{tr}[XS] = \frac{1}{2}\,\mathrm{tr}[XS + SX] = 0.$$

In view of Lemma 2.1, we therefore have $XS = 0$. $\square$

We now want to modify the definition of ϕ so that it can be used in order to characterize the central path conditions (4). To this end, let $\tau \geq 0$ be any nonnegative number which will be viewed as a parameter within this section. Then define $\varphi_\tau : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

$$\varphi_\tau(a,b) := a + b - \sqrt{a^2 + b^2 + 2\tau^2}.$$

This is the so-called smoothed Fischer-Burmeister function since it is obviously continuously differentiable for every $\tau > 0$ and since it coincides with the Fischer-Burmeister function φ for $\tau = 0$. The mapping φ_τ was introduced in [19] and has the following interesting property:

$$\varphi_\tau(a,b) = 0 \iff a \geq 0, b \geq 0, ab = \tau^2.$$

This simple observation was made in [19], and it shows that several smoothing-type methods for linear programs and related problems are closely related to interior-point methods.

We now generalize the smoothed Fischer-Burmeister function φ_τ in an obvious way: Define $\phi_\tau : \mathcal{S}^{n\times n} \times \mathcal{S}^{n\times n} \to \mathcal{S}^{n\times n}$ by

$$\phi_\tau(X,S) := X + S - (X^2 + S^2 + 2\tau^2 I)^{1/2}. \tag{11}$$

Then we can state the following result.

Proposition 2.3 *Let $\tau > 0$ be any positive number and let ϕ be defined by* (11). *Then*

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

Proof. First assume that $X \succ 0, S \succ 0, XS = \tau^2 I$ holds. This implies $XS + SX = 2\tau^2 I$ and therefore

$$(X+S)^2 = X^2 + S^2 + 2\tau^2 I.$$

Using the fact that X and S are symmetric positive definite, it follows that

$$X + S = (X^2 + S^2 + 2\tau^2 I)^{1/2}.$$

This, in turn, implies $\phi_\tau(X,S) = 0$.

Conversely, let $\phi_\tau(X,S) = 0$ for two symmetric matrices $X, S \in \mathcal{S}^{n\times n}$. This means that $X + S = (X^2 + S^2 + 2\tau^2 I)^{1/2}$. Squaring both sides of this equation gives

$$X^2 + S^2 + 2\tau^2 I = (X+S)^2 \quad \text{and} \quad X + S \in \mathcal{S}^{n\times n}_{++}.$$

This is equivalent to

$$XS + SX = 2\tau^2 I \quad \text{and} \quad X + S \in \mathcal{S}^{n\times n}_{++}. \tag{12}$$

Let $X = Q^T D Q$ with $Q \in \mathbb{R}^{n\times n}$ orthogonal and $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ be the spectral decomposition of the symmetric matrix X. Following the proof of Proposition 2.2 and using the abbreviation $A := QSQ^T$, we see that (12) can be rewritten as

$$DA + AD = 2\tau^2 I \quad \text{and} \quad D + A \in \mathcal{S}^{n\times n}_{++}. \tag{13}$$

Componentwise, this becomes

$$(\lambda_i + \lambda_j)a_{ij} = 2\tau^2 \delta_{ij} \quad \text{and} \quad D + A \in \mathcal{S}^{n\times n}_{++} \tag{14}$$

for all $i, j = 1, \ldots, n$, where δ_{ij} is the standard Kronecker symbol, i.e.,

$$\delta_{ij} := \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

In particular, we obtain for $i = j$

$$2\lambda_i a_{ii} = 2\tau^2 \quad \text{and} \quad \lambda_i + a_{ii} > 0$$

for all $i = 1, \ldots, n$. Obviously, this implies $\lambda_i > 0$ for all $i = 1, \ldots, n$. Hence the symmetric matrix X is positive definite.

In a similar way (using a spectral decomposition of S), we can show that S is also positive definite.

In order to verify that $XS = \tau^2 I$, we observe that (14) implies $a_{ij} = 0$ for all $i \neq j$ since $\lambda_i + \lambda_j > 0$ according to our previous argument. Hence A is a diagonal matrix. In particular, we therefore have $DA = AD$. Consequently, we obtain from (13) that $DA = \tau^2 I$. Premultiplying this equation by Q^T and postmultiplying it by Q gives $XS = Q^T DQS = Q^T DAQ = \tau^2 I$. □

We next want to introduce a second function with similar properties as the (smoothed) Fischer-Burmeister function. To this end, let

$$\varphi(a, b) := 2 \min\{a, b\}$$

for $a, b \in \mathbb{R}$. For obvious reasons, this mapping is called the minimum function. It is easy to see that it satisfies the equivalence

$$\varphi(a, b) = 0 \iff a \geq 0, b \geq 0, ab = 0.$$

In order to extend its definition to the class of symmetric matrices, it is helpful to reformulate the minimum function in the following way:

$$\varphi(a, b) = 2 \min\{a, b\} = a + b - |a - b| = a + b - \sqrt{(a - b)^2}.$$

Motivated by the expression on the right-hand side, we now define the function $\phi : \mathcal{S}^{n\times n} \times \mathcal{S}^{n\times n} \to \mathcal{S}^{n\times n}$ by

$$\phi(X, S) := X + S - ((X - S)^2)^{1/2}. \tag{15}$$

It turns out that this function shares the property (7) with the Fischer-Burmeister function from (6). This observation is similar to the one made by Tseng [27, Lemma 2.1] and can alternatively be verified by following the proof of Proposition 2.2. We skip the details here and just state the result.

Proposition 2.4 *Let ϕ be the minimum function defined in* (15). *Then*

$$\phi(X, S) = 0 \iff X \succeq 0, S \succeq 0, XS = 0.$$

We now want to modify the definition of the minimum function in such a way that we get a characterization of the central path conditions (4). To this end, we first recall that there is a suitable modification of the minimum function for scalar variables, namely

$$\varphi_\tau(a, b) := a + b - \sqrt{(a - b)^2 + 4\tau^2},$$

where τ denotes a nonnegative number. This smoothed minimum function is usually called the Chen-Harker-Kanzow-Smale smoothing function in the literature [6, 19, 23], and it was noted in [19] that it has the following property for each $\tau > 0$:

$$\varphi_\tau(a, b) = 0 \iff a > 0, b > 0, ab = \tau^2.$$

This observation motivates to define a mapping $\phi_\tau : \mathcal{S}^{n\times n} \times \mathcal{S}^{n\times n} \to \mathcal{S}^{n\times n}$ by

$$\phi_\tau(X, S) := X + S - ((X - S)^2 + 4\tau^2 I)^{1/2}. \tag{16}$$

It turns out that this function has the desired property.

Proposition 2.5 *Let $\tau > 0$ be any positive number and let ϕ be defined by* (16). *Then*

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

Proof. First assume that $X \succ 0, S \succ 0, XS = \tau^2 I$ holds. This implies $XS + SX = 2\tau^2 I$ and therefore

$$(X + S)^2 = (X - S)^2 + 4\tau^2 I.$$

Consequently, we have

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

i.e., $\phi_\tau(X, S) = 0$.

Conversely, if $\phi_\tau(X, S) = 0$ for two matrices $X, S \in \mathcal{S}^{n\times n}$, we get $X + S = ((X - S)^2 + 4\tau^2 I)^{1/2}$ and therefore

$$(X - S)^2 + 4\tau^2 I = (X + S)^2 \quad \text{and} \quad X + S \in \mathcal{S}^{n\times n}_{++}.$$

This is equivalent to

$$XS + SX = 2\tau^2 I \quad \text{and} \quad X + S \in \mathcal{S}^{n\times n}_{++}.$$

Hence we can follow the argument from the proof of Proposition 2.3 in order to show that $X \succ 0, S \succ 0$, and $XS = \tau^2 I$ holds. $\square$

Let ϕ_τ denote either the smoothed Fischer-Burmeister function from (11) or the smoothed minimum function from (16). Then define a mapping $\Phi_\tau : \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n} \to \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n}$ by

$$\Phi_\tau(X, \lambda, S) := \begin{pmatrix} \sum_{i=1}^m \lambda_i A_i + S - C \\ A_i \bullet X - b_i \ \ (i = 1, \dots, m) \\ \phi_\tau(X, S) \end{pmatrix}. \tag{17}$$

Then Propositions 2.3 and 2.5 immediately give the following new characterization of the central path conditions (4) for semidefinite programs.

Theorem 2.6 *Let Φ_τ be defined by (17) with ϕ given by (11) or (16), and let $\tau > 0$. Then the following statements are equivalent:*

(a) (X, λ, S) satisfies the central path conditions (4).

(b) (X, λ, S) is a solution of the nonlinear system of equations $\Phi_\tau(X, \lambda, S) = 0$.

3 Properties of ϕ

In this section we will state some properties of the functions ϕ_τ introduced in the previous section. In particular, we will show that these functions are differentiable (in the sense of Fréchet).

However, in contrast to our previous section, we will view the nonnegative number τ as an independent variable from now on. In order to make this clear in our notation, we set $\phi(X, S, \tau) := \phi_\tau(X, S)$, i.e., we now write

$$\phi(X, S, \tau) := X + S - (X^2 + S^2 + 2\tau^2 I)^{1/2} \tag{18}$$

for the smoothed Fischer-Burmeister function from (11), and

$$\phi(X, S, \tau) := X + S - ((X - S)^2 + 4\tau^2 I)^{1/2} \tag{19}$$

for the smoothed minimum function from (16). Taking τ as a variable rather than a parameter is motivated by some computational reasons and will be explained in some more detail in our next section when we present our smoothing-type method for the solution of the optimality conditions (3).

We begin our analysis of the functions ϕ with the following result whose proof can be found in [8, Lemma 1].

Lemma 3.1 *Let ϕ denote one of the functions defined in (18) or (19). Then, for any $X, S \in \mathcal{S}^{n\times n}$ and any $\tau > \nu > 0$, we have*

$$\begin{aligned} \kappa\,(\tau - \nu) I &\succeq \phi(X, S, \nu) - \phi(X, S, \tau) \succ 0, \\ \kappa\,\tau I &\succeq \phi(X, S, 0) - \phi(X, S, \tau) \succ 0, \end{aligned}$$

where κ denotes a positive constant independent of X, S, τ, and ν.

We stress that the constant κ from Lemma 3.1 is actually known: $\kappa = 2$ for the smoothed minimum function, and $\kappa = \sqrt{2}$ for the smoothed Fischer-Burmeister function.

Corollary 3.2 *Let ϕ be given by* (18) *or* (19)*, and let κ be the constant from Lemma 3.1. Then the following statements hold:*

(a) The inequality

$$\|\phi(X,S,\nu) - \phi(X,S,\tau)\|_F \le \kappa\sqrt{n}(\tau - \nu)$$

holds for all $X, S \in \mathcal{S}^{n\times n}$ and all $\tau > \nu > 0$.

(b) The inequality

$$\|\phi(X,S,0) - \phi(X,S,\tau)\|_F \le \kappa\sqrt{n}\,\tau$$

holds for all $X, S \in \mathcal{S}^{n\times n}$ and all $\tau > 0$.

Proof. Let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of the symmetric matrix $\phi(X,S,\nu) - \phi(X,S,\tau)$. By Lemma 3.1 we have $\kappa(\tau - \nu) \ge \lambda_i > 0$. We therefore get

$$\|\phi(X,S,\nu) - \phi(X,S,\tau)\|_F = \sqrt{\lambda_1^2 + \ldots + \lambda_n^2} \le \kappa\sqrt{n}(\tau - \nu).$$

This proves part (a). The second statement can be derived from Lemma 3.1 in essentially the same way as part (a). $\square$

We next want to show that the two functions ϕ from (18) and (19) are continuously differentiable in their arguments X, S, and τ, at least under suitable assumptions. This result was essentially given by Chen and Tseng [8, Lemma 2] (who, however, view τ as a parameter) and can alternatively be derived from the recent paper [24] by Sun and Sun.

Here we give a somewhat different proof for the differentiability of the functions ϕ. The reason is that, at least in our opinion, the proof given in, e.g., [8] is not very constructive in the sense that it is not clear how to obtain the somewhat complicated formulas for the derivatives of the functions ϕ. We hope that the reader will find our approach more constructive. It is based on the following lemma from [16, Section 7.2].

Lemma 3.3 *Let $A \in \mathcal{S}^{n\times n}_{++}$, $B \in \mathcal{S}^{n\times n}_{+}$ be two given matrices. Then*

$$\left\|A^{1/2} - B^{1/2}\right\|_2 \le \left\|A^{-1/2}\right\|_2 \cdot \|A - B\|_2 .$$

We are now in the position to derive a formula for the derivatives of the mappings ϕ. To be specific, assume that ϕ denotes the smoothed Fischer-Burmeister function from (18). We have to show that

$$\|\phi(X+U, S+V, \tau+\mu) - \phi(X,S,\tau) - \nabla\phi(X,S,\tau)(U,V,\mu)\|_2 = o(\|\!|(U,V,\mu)|\!\|)$$

holds for all $(U, V, \mu) \in \mathcal{S}^{n\times n} \times \mathcal{S}^{n\times n} \times \mathbb{R}$ tending to $(0,0,0)$, where $\nabla\phi(X,S,\tau)$ denotes a suitable linear operator standing for the derivative of ϕ at the point (X,S,τ). To this end, we decompose the mapping ϕ into

$$\phi(X,S,\tau) = \phi_1(X,S,\tau) - \phi_2(X,S,\tau)$$

with

$$\begin{aligned} \phi_1(X,S,\tau) &:= X+S, \\ \phi_2(X,S,\tau) &:= (X^2+S^2+2\tau^2 I)^{1/2}. \end{aligned}$$

Then it is easy to see that ϕ_1 is differentiable with

$$\nabla\phi_1(X,S,\tau)(U,V,\mu) = U+V.$$

The situation for ϕ_2 is more complicated. Let us define

$$E := \left(X^2+S^2+2\tau^2 I\right)^{1/2}$$

and assume that E is positive definite. Let

$$L_E[X] := EX + XE \tag{20}$$

denote the corresponding Lyapunov operator. Then the positive definiteness of E guarantees that the Lyapunov equation

$$L_E[X] = H$$

has a unique solution within the set of symmetric matrices for every $H \in \mathcal{S}^{n\times n}$, cf. [17, Theorem 2.2.3]. Hence we can define the inverse L_E^{-1} of L_E, i.e., $L_E^{-1}[H]$ denotes the unique element X satisfying $EX + XE = H$. Let us further define the matrix

$$D := \left((X+U)^2 + (S+V)^2 + 2(\tau+\mu)^2 I\right)^{1/2}.$$

An easy calculation shows that $D^2 - E^2 = L_E[D-E] + (D-E)^2$. Applying L_E^{-1} to this equation and rearranging terms yields

$$\begin{aligned} E - D &= L_E^{-1}[(D-E)^2 - (D^2-E^2)] \\ &= L_E^{-1}[(E-D)^2 - (XU+UX+SV+VS+4\tau\mu I + U^2+V^2+2\mu^2 I)]. \end{aligned}$$

Using the linearity of L_E^{-1} then gives

$$\begin{aligned} &\phi_2(X+U,S+V,\tau+\mu) - \phi_2(X,S,\tau) - \nabla\phi_2(X,S,\tau)(U,V,\mu) \\ =\; &-\nabla\phi_2(X,S,\tau)(U,V,\mu) - (E-D) \\ =\; &-\nabla\phi_2(X,S,\tau)(U,V,\mu) + L_E^{-1}[XU+UX+SV+VS+4\tau\mu I] \\ &\quad + L_E^{-1}[U^2+V^2+2\mu^2 I] - L_E^{-1}[(E-D)^2]. \end{aligned} \tag{21}$$

Obviously, we have

$$\|L_E^{-1}[U^2+V^2+2\mu^2 I]\|_F = O(\|\!|(U,V,\mu)|\!\|^2).$$

In view of Lemma 3.3, we also have

$$\begin{aligned}
\left\|(E-D)^2\right\|_F &\le \|E-D\|_F^2 \\
&\le \gamma_1 \left\|E^2-D^2\right\|_F^2 \\
&= \gamma_1 \left\|XU+UX+SV+VS+4\tau\mu I+U^2+V^2+2\mu^2 I\right\|_F^2 \\
&= O(\|\!|(U,V,\mu)|\!\|^2)
\end{aligned}$$

for some constant $\gamma_1 > 0$ independent of U, V, and μ. This implies

$$\|L_E^{-1}[(E-D)^2]\|_F = O(\|\!|(U,V,\mu)|\!\|^2).$$

Therefore, setting

$$\nabla\phi_2(X,S,\tau)(U,V,\mu) := L_E^{-1}[XU+UX+SV+VS+4\tau\mu I],$$

it follows immediately from (21) that ϕ_2 is differentiable at (X,S,τ). This, in turn, implies that ϕ itself is differentiable at this point. This proves the main part of the first statement in the following result.

Theorem 3.4 *Let $X, S \in \mathcal{S}^{n\times n}$ be two given matrices and $\tau \in \mathbb{R}_+$.*

(a) If ϕ is given by (18) and $X^2+S^2+2\tau^2 I \succ 0$, then ϕ is continuously differentiable in (X,S,τ) with

$$\nabla\phi(X,S,\tau)(U,V,\mu) = U+V-L_E^{-1}\left[XU+UX+SV+VS+4\tau\mu I\right], \tag{22}$$

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

(b) If ϕ is given by (19) and $(X-S)^2+4\tau^2 I \succ 0$, then ϕ is continuously differentiable in (X,S,τ) with

$$\nabla\phi(X,S,\tau)(U,V,\mu) = U+V-L_E^{-1}\left[(X-S)(U-V)+(U-V)(X-S)+8\tau\mu I\right], \tag{23}$$

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

Proof. (a) The differentiability of the smoothed Fischer-Burmeister function follows from our preceding discussion.

Since $E = (X^2+S^2+2\tau^2 I)^{1/2} \succ 0$ is continuous in (X,S,τ) by Lemma 3.3, it is readily seen that $\nabla\phi(X,S,\tau)$ is continuous in (X,S,τ), see also [8]. Hence ϕ is continuously differentiable in (X,S,τ).

(b) Let $D := ((X - S + U - V)^2 + 4(\tau + \mu)^2 I)^{1/2}$. Using $D^2 - E^2 = L_E[D - E] + (D - E)^2$ and applying L_E^{-1} to this equation yields

$$E - D = L_E^{-1}[(E - D)^2 - ((X - S)(U - V) + (U - V)(X - S) + (U - V)^2 + 8\tau\mu I + 4\mu^2 I)].$$

Since L_E^{-1} is linear, this implies

$$\begin{aligned} &\phi(X + U, S + V, \tau + \mu) - \phi(X, S, \tau) - \nabla\phi(X, S, \tau)(U, V, \mu) \\ &= U + V - D + E - \nabla\phi(X, S, \tau)(U, V, \mu) \\ &= E - D + L_E^{-1}[(X - S)(U - V) + (U - V)(X - S) + 8\tau\mu I] \\ &= L_E^{-1}[(E - D)^2] - L_E^{-1}[(U - V)^2 + 4\mu^2 I]. \end{aligned}$$

In view of Lemma 3.3 we have

$$\begin{aligned} \left\|(E - D)^2\right\|_F &\leq \|E - D\|_F^2 \\ &\leq \gamma_2 \left\|E^2 - D^2\right\|_F^2 \\ &= \gamma_2 \left\|(X - S)(U - V) + (U - V)(X - S) + (U - V)^2 + 8\tau\mu I + 4\mu^2 I\right\|_F^2 \\ &= O(\|\!|(U, V, \mu)|\!\|^2) \end{aligned}$$

for some constant $\gamma_2 > 0$ independent of U, V, and τ. This implies

$$\begin{aligned} &\|\phi(X + U, S + V, \tau + \mu) - \phi(X, S, \tau) - \nabla\phi(X, S, \tau)(U, V, \mu)\|_F \\ &\leq \gamma_3 \left\|(E - D)^2\right\|_F + \gamma_3 \left\|(U - V)^2 + 4\mu^2 I\right\|_F \\ &= O(\|\!|(U, V, \mu)|\!\|^2) \end{aligned}$$

for some constant $\gamma_3 > 0$ independent of U, V, and τ. This shows that ϕ is differentiable with $\nabla\phi$ given by (23).

Since $E = ((X - S)^2 + 4\tau^2 I)^{1/2} \succ 0$ is continuous in (X, S, τ), it is readily seen that $\nabla\phi(X, S, \tau)$ is continuous in (X, S, τ), cf. [8]. □

Note that Theorem 3.4 implies that, if $\tau > 0$, then both functions ϕ are continuously differentiable everywhere.

We close this section by noting that both functions ϕ were shown to satisfy the relation

$$\|\phi(X + U, S + V, \tau + \mu) - \phi(X, S, \tau) - \nabla\phi(X, S, \tau)(U, V, \mu)\|_2 = O(\|\!|(U, V, \mu)|\!\|^2), \tag{24}$$

while it would have been enough to show that

$$\|\phi(X + U, S + V, \tau + \mu) - \phi(X, S, \tau) - \nabla\phi(X, S, \tau)(U, V, \mu)\|_2 = o(\|\!|(U, V, \mu)|\!\|)$$

holds in order to see that the functions ϕ are differentiable at the point (X, S, τ). The reader might therefore ask whether the stronger relation (24) already implies that ϕ is continuously differentiable. However, this is not true as indicated by the following counterexample: Let $f(x) := x^2 \sin(1/x)$ for $x \neq 0$ and $f(0) := 0$. Then it is easy to see that f satisfies

$$|f(x) - f(0) - f'(0)x| = O(|x|^2),$$

with $f'(0) := 0$, i.e., (24) holds. However, f is only differentiable in the origin, but not continuously differentiable.

4 Description of Algorithm

We now want to exploit our previous results in order to obtain a suitable algorithm for the solution of the optimality conditions (3) and, therefore, for the solution of the underlying primal and dual semidefinite programs. The most obvious way would be to utilize the mapping

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

with ϕ being the Fischer-Burmeister function (6) or the minimum function (15), since then Propositions 2.2 and 2.4 immediately imply that

$$(X^*, \lambda^*, S^*) \text{ solves (3)} \iff (X^*, \lambda^*, S^*) \text{ solves } \Phi(X, \lambda, S) = 0.$$

However, solving the nonlinear system of equations $\Phi(X, \lambda, S) = 0$ is a nontrivial task because ϕ and, therefore, Φ is nonsmooth in general. Hence we do not follow this idea here although some recent theoretical results [24, 7, 14] indicate that such an approach might be possible.

The next idea is to replace the nondifferentiable mapping Φ by the smooth function

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

where ϕ_τ denotes either the smoothed Fischer-Burmeister function from (11) or the smoothed minimum function from (16). This (specialized to the framework of semidefinite programs) is precisely the approach followed by Chen and Tseng [8] although they have not observed the equivalence between the nonlinear system of equations $\Phi_\tau(X, \lambda, S) = 0$ on the one hand and the central path conditions (4) on the other hand, cf. Theorem 2.6.

In this paper, however, we follow an idea by Jiang [18] (in the context of nonlinear complementarity problems) and view τ as an independent variable. To this end, we define the mapping $\Theta : \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n} \times \mathbb{R} \to \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n} \times \mathbb{R}$ by

$$\Theta(X, \lambda, S, \tau) := \begin{pmatrix} \sum_{i=1}^m \lambda_i A_i + S - C \\ A_i \bullet X - b_i \ \ (i = 1, \ldots, m) \\ \phi(X, S, \tau) \\ \tau \end{pmatrix}, \tag{25}$$

where ϕ denotes one of the functions given by (18) or (19). Apart from the fact that τ is an independent variable rather than a parameter, the function Θ differs from the function Φ_τ also because we have added one more line so that

$$\Theta(X, \lambda, S, \tau) = 0 \tag{26}$$

becomes a square system of equations. This additional line immediately implies $\tau = 0$ so that the system (26) is equivalent to the optimality conditions (3) themselves (and not to the central path conditions (4)). This might be an advantage compared with the reformulation $\Phi_\tau(X,\lambda,S)=0$ since we really want to solve the optimality conditions (3) and not the central path conditions (4). Furthermore, it follows from Theorem 3.4 that Θ is a continuously differentiable function at any point (X,λ,S,τ) with $\tau>0$, and the positivity of τ will automatically be guaranteed by our method. This is an advantage compared with the nonsmooth reformulation $\Phi(X,\lambda,S)=0$. Moreover, according to our numerical experience done with some related methods for the solution of linear programs (cf. [9, 10, 11]), the reformulation (26) has the best numerical behaviour. It also has some better theoretical properties in the context of linear complementarity problems, see Burke and Xu [5, 4], although it is currently not clear whether this can be extended to semidefinite programs.

The main idea of our algorithm is to solve the system of equations (26) by Newton's method. Global convergence of this method is achieved by following a suitable neighbourhood of the central path. The neighbourhood used here is given by

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

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 as well as the notation used in the subsequent analysis, 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 are now in the position to give a formal statement of our smoothing-type method for the solution of semidefinite programs.

Algorithm 4.1

(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*

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

Choose $\tau_0>0$, $\beta>0$ *with* $\|\phi(X^0,S^0,\tau_0)\|_F\le\beta\tau_0$ *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*

$$\nabla\Theta(W^k,\tau_k)\begin{pmatrix}\Delta W\\ \Delta\tau\end{pmatrix}=-\Theta(W^k,\tau_k).\tag{27}$$

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*

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

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

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

(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*

$$\nabla\Theta(\hat{W}^k, \hat{\tau}_k)\begin{pmatrix} \Delta\hat{W} \\ \Delta\hat{\tau} \end{pmatrix} = -\Theta(\hat{W}^k, \hat{\tau}_k) + \begin{pmatrix} 0 \\ (1-\hat{\sigma})\hat{\tau}_k \end{pmatrix}. \tag{28}$$

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

$$\left\|\phi(\hat{X}^k+\hat{\eta}_k\Delta\hat{X}^k, \hat{S}^k+\hat{\eta}_k\Delta\hat{S}^k, \hat{\tau}_k+\hat{\eta}_k\Delta\hat{\tau}_k)\right\|_F \leq (1-\hat{\sigma}\hat{\eta}_k)\beta\hat{\tau}_k. \tag{29}$$

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).

It can easily be seen that all iterates (X^k, λ^k, S^k) and $(\hat{X}^k, \hat{\lambda}^k, \hat{S}^k)$ generated by Algorithm 4.1 are feasible for the optimality conditions (3) in the sense that

$$\sum_{i=1}^m \lambda_i^k A_i + S^k = C, \qquad A_i \bullet X^k = b_i \quad (i=1,\dots,m) \tag{30}$$

and

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

hold for all $k \in \mathbb{N}$. Moreover, we will see below that all matrices X^k, S^k and $\hat{X}^k, \hat{S}^k$ are symmetric, cf. Section 7. In contrast to interior-point methods, however, these matrices are not necessarily positive definite or positive semidefinite.

The termination criterion used in (S.1) is justified by Propositions 2.2 and 2.4: Together with our previous note on the feasibility of the iterates, these results imply that

$$\left\|\phi(X^k+\Delta X^k, S^k+\Delta S^k, 0)\right\|_F = 0 \iff W^k+\Delta W^k \text{ is a solution of (3).}$$

For our theoretical analysis of Algorithm 4.1, we will always assume that this criterion never holds so that Algorithm 4.1 generates an infinite sequence. Furthermore, the updating rule for τ_{k+1} in (S.2) is equivalent to the more standard formula

$$\tau_{k+1} = \hat{\tau}_k + \hat{\eta}_k \Delta\hat{\tau}_k;$$

this observation follows immediately from the last row of the linear system (28) in the corrector step which gives $\Delta\hat{\tau}_k = -\hat{\sigma}\hat{\tau}_k$.

Finally, we stress that we have to solve a linear system of equations in both the predictor and 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, an easy inspection of our subsequent analysis shows that all convergence results remain true for the following modification of Algorithm 4.1: 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 linear system of equations in the predictor step, or it has to solve two systems, but then these two systems have the same coefficient matrix. This modification has been implemented in order to obtain the numerical results in Section 7.

We now start to analyse the properties of Algorithm 4.1 more formally. The aim of the remaining part of this section will be to show that Algorithm 4.1 is well-defined. To this end, we first want to show that the linear systems (27) and (28) have a unique solution. In order to verify this statement, we need some further properties of the Lyapunov operator from (20). These properties are therefore summarized in our next result.

Lemma 4.2 *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 defined by (20), with L_A^{-1}, L_B^{-1} denoting their inverses. Then the following statements hold:*

(a) L_A and L_B are self-adjoint.

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

(c) $L_A \circ L_B$ and $L_B \circ L_A$ are strongly monotone.

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

Proof. (a) We only have to verify that L_A is self-adjoint. This follows directly from the fact that

$$\begin{aligned} L_A[X] \bullet Y &= \operatorname{tr}(L_A[X]Y) \\ &= \operatorname{tr}\big((AX + XA)Y\big) \\ &= \operatorname{tr}(AXY) + \operatorname{tr}(XAY) \\ &= \operatorname{tr}(XYA) + \operatorname{tr}(XAY) \\ &= \operatorname{tr}\big(X(AY + YA)\big) \\ &= \operatorname{tr}(XL_A[Y]) \\ &= X \bullet L_A[Y] \end{aligned}$$

for all $X, Y \in \mathcal{S}^{n\times n}$.

(b) We show that L_A^{-1} is self-adjoint. Noting that L_A^{-1} is the inverse of L_A and exploiting part (a), we obtain

$$\begin{aligned} L_A^{-1}[X] \bullet Y &= L_A^{-1}[X] \bullet L_A\big[L_A^{-1}[Y]\big] \\ &= L_A\big[L_A^{-1}[X]\big] \bullet L_A^{-1}[Y] \\ &= X \bullet L_A^{-1}[Y] \end{aligned}$$

for all $X, Y \in \mathcal{S}^{n\times n}$.

(c) Using the first statement, we obtain

$$\begin{aligned} \big(L_A \circ L_B[X]\big) \bullet X &= L_B[X] \bullet L_A[X] \\ &= \operatorname{tr}(L_B[X]L_A[X]) \\ &= \operatorname{tr}\big((BX+XB)(AX+XA)\big) \\ &= \operatorname{tr}\big(BXAX + XBAX + BXXA + XBXA\big) \\ &= \operatorname{tr}\big(2BXAX + X^2(BA+AB)\big) \\ &= 2\operatorname{tr}(BXAX) + \operatorname{tr}\big(X^2(BA+AB)\big) \\ &= 2\left\|B^{1/2}XA^{1/2}\right\|_F^2 + \operatorname{tr}\big(X(BA+AB)X\big) \end{aligned} \tag{31}$$

for all $X \in \mathcal{S}^{n\times n}$. Since $BA\,(AB)$ is similar to $B^{1/2}AB^{1/2}$ $(A^{1/2}BA^{1/2})$ and A, B are symmetric positive definite, it follows that BA and AB have real and positive eigenvalues. Hence the symmetric matrix $BA + AB$ is positive definite. Consequently, $X(BA + AB)X$ is positive semi-definite so that

$$\operatorname{tr}\big(X(BA+AB)X\big) \geq 0 \tag{32}$$

for all $X \in \mathcal{S}^{n\times n}$. Furthermore, since the mapping $X \mapsto \left\|B^{1/2}XA^{1/2}\right\|_F$ defines a norm and all norms are equivalent in finite dimensional spaces, there exists a constant $\mu > 0$ such that

$$\left\|B^{1/2}XA^{1/2}\right\|_F \geq \mu \left\|X\right\|_F \tag{33}$$

for all $X \in \mathcal{S}^{n\times n}$. Putting together the inequalities (31)–(33), we obtain

$$\big(L_A \circ L_B[X]\big) \bullet X \geq 2\left\|B^{1/2}XA^{1/2}\right\|_F^2 \geq 2\mu^2 \left\|X\right\|_F^2,$$

i.e., $L_A \circ L_B$ is strongly monotone on $\mathcal{S}^{n\times n}$. In order to see that $L_B \circ L_A$ is also strongly monotone, we just have to change the roles of A and B.

(d) Since L_A is self-adjoint by part (a), we obtain for every $X \in \mathcal{S}^{n\times n}$ (by setting $Y := L_A^{-1}[X]$)

$$\begin{aligned} \big(L_A^{-1} \circ L_B[X]\big) \bullet X &= \big(L_A^{-1} \circ L_B \circ L_A[Y]\big) \bullet L_A[Y] \\ &= \big(L_B \circ L_A[Y]\big) \bullet Y. \end{aligned}$$

However, $L_B \circ L_A$ is strongly monotone by part (c). Hence the fourth statement is a direct consequence of (c). $\square$

In order to see that the linear systems (27) and (28) have a unique solution, we will show that the linear mapping $\nabla\Theta(X, \lambda, S, \tau)$ is invertible. To this end, we state the following standard assumption.

Assumption 4.3 *The matrices A_i $(i = 1, \dots, m)$ are linearly independent, i.e.,*

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

Exploiting Lemma 4.2 and Assumption 4.3, we are now able to show that $\nabla\Theta(X, \lambda, S, \tau)$ is a bijection, i.e., it is both one-to-one and onto. Note that this implies that the predictor direction $(\Delta X^k, \Delta\lambda^k, \Delta S^k, \Delta\tau_k)$ and the corrector direction $(\Delta\hat{X}^k, \Delta\hat{\lambda}^k, \Delta\hat{S}^k, \Delta\hat{\tau}_k)$ are well-defined.

Proposition 4.4 *Suppose that Assumption 4.3 holds. Then the linear mapping $\nabla\Theta(X, \lambda, S, \tau)$ with ϕ given by* (18) *or* (19) *is bijective for all $(X, \lambda, S, \tau) \in \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n} \times \mathbb{R}_{++}$.*

Proof. We only consider the case where the function ϕ is given by (18). The proof for the smoothed minimum function is similar and therefore omitted here.

Let $(X, \lambda, S, \tau) \in \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n} \times \mathbb{R}_{++}$ be fixed. Since $\nabla\Theta(X, \lambda, S, \tau)$ is a linear mapping from the finite-dimensional vector space $\mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n} \times \mathbb{R}$ into itself, we only have to verify that this mapping is one-to-one. To this end, it is sufficient to show that the system

$$\nabla\Theta(X, \lambda, S, \tau)(\Delta X, \Delta\lambda, \Delta S, \Delta\tau) = (0, 0, 0, 0)$$

or, equivalently, the system

$$\begin{aligned} \sum_{i=1}^{m} \Delta\lambda_i A_i + \Delta S &= 0, && (34) \\ A_i \bullet \Delta X &= 0 \quad (i = 1, \dots, m), && (35) \\ \nabla\phi(X, S, \tau)(\Delta X, \Delta S, \Delta\tau) &= 0, && (36) \\ \Delta\tau &= 0 && (37) \end{aligned}$$

has $(\Delta X, \Delta\lambda, \Delta S, \Delta\tau) = (0, 0, 0, 0)$ as its only solution. From (37) we immediately obtain $\Delta\tau = 0$. Setting $E := (X^2 + S^2 + 2\tau^2 I)^{1/2}$, we therefore get from (36) and Theorem 3.4 that

$$\Delta X + \Delta S - L_E^{-1}\left[X\Delta X + \Delta X\, X + S\Delta S + \Delta S\, S\right] = 0.$$

Applying L_E to both sides of the equation and rearranging terms yields

$$L_{E-X}[\Delta X] + L_{E-S}[\Delta S] = 0.$$

Since $E - S \succ 0$ (see [27, Lemma 6.1(c)] for a formal proof), the inverse L_{E-S}^{-1} exists, and we get

$$L_{E-S}^{-1} \circ L_{E-X}[\Delta X] + \Delta S = 0. \tag{38}$$

Using (34) and (35) and taking the scalar product with ΔX yields

$$0 = L_{E-S}^{-1} \circ L_{E-X}[\Delta X] \bullet \Delta X - \sum_{i=1}^{m} \Delta\lambda_i \underbrace{A_i \bullet \Delta X}_{=0} = L_{E-S}^{-1} \circ L_{E-X}[\Delta X] \bullet \Delta X. \tag{39}$$

Using the fact that $E - X \succ 0$ and $E - S \succ 0$, it follows from Lemma 4.2 (d) that the operator $L_{E-S}^{-1} \circ L_{E-X}$ is strongly monotone. Therefore, (39) immediately gives $\Delta X = 0$. This implies $\Delta S = 0$ by (38). The assumed linear independence of the matrices A_i and (34) show that $\Delta\lambda = 0$, and this completes the proof. □

Based on the previous results, we are now in the position to show that Algorithm 4.1 is well-defined under Assumption 4.3.

Theorem 4.5 *Algorithm 4.1 is well-defined under Assumption 4.3. Furthermore, the iterates $W^k = (X^k, \lambda^k, S^k)$ and τ_k and $\hat{W}^k = (\hat{X}^k, \hat{\lambda}^k, \hat{S}^k)$ and $\hat{\tau}_k$ belong to the neighbourhood $\mathcal{N}(\beta)$.*

Proof. In view of Proposition 4.4, it remains to show that the two backtracking strategies in (S.1) and (S.2) of Algorithm 4.1 are well-defined.

To this end, we first consider (S.1). Since we assume that Algorithm 4.1 generates an infinite sequence, the termination criterion in (S.1) is not satisfied for any $k \in \mathbb{N}$. Hence we have $\left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, 0)\right\|_F > 0$. Since the mapping $\tau \mapsto \|\phi(X, S, \cdot)\|_F$ is continuous, this implies that the backtracking strategy in (S.1) terminates in a finite number of inner loops. Thus, $\hat{W}^k$ and $\hat{\tau}_k$ are well defined and satisfy the neighbourhood condition $(\hat{W}^k, \hat{\tau}_k) \in \mathcal{N}(\beta)$.

We next consider the steplength procedure from (S.2). Let us define the mapping $\psi(X, S, \tau) := \|\phi(X, S, \tau)\|_F$. Then standard rules from differential calculus imply

$$\begin{aligned} \psi'(\hat{X}^k, \hat{S}^k, \hat{\tau}_k)(\Delta\hat{X}^k, \Delta\hat{S}^k, \Delta\hat{\tau}_k) &= \frac{\phi(\hat{X}^k, \hat{S}^k, \hat{\tau}_k) \bullet \nabla\phi(\hat{X}^k, \hat{S}^k, \hat{\tau}_k)(\Delta\hat{X}^k, \Delta\hat{S}^k, \Delta\hat{\tau}_k)}{\left\|\phi(\hat{X}^k, \hat{S}^k, \hat{\tau}_k)\right\|_F} \\ &\overset{(28)}{=} -\left\|\phi(\hat{X}^k, \hat{S}^k, \hat{\tau}_k)\right\|_F. \end{aligned} \tag{40}$$

Now suppose that the computation of the steplength $\hat{\eta}_k$ does not terminate in a finite number of inner loops in (S.2). Then we have

$$\left\|\phi(\hat{X}^k + \alpha_2^t \Delta\hat{X}^k, \hat{S}^k + \alpha_2^t \Delta\hat{S}^k, \hat{\tau}_k + \alpha_2^t \Delta\hat{\tau}_k)\right\|_F > (1 - \hat{\sigma}\alpha_2^t)\beta\hat{\tau}_k$$

for all $t \in \mathbb{N}$. Since $\beta\hat{\tau}_k \geq \left\|\phi(\hat{X}^k, \hat{S}^k, \hat{\tau}_k)\right\|_F$, this implies

$$\left\|\phi(\hat{X}^k + \alpha_2^t \Delta\hat{X}^k, \hat{S}^k + \alpha_2^t \Delta\hat{S}^k, \hat{\tau}_k + \alpha_2^t \Delta\hat{\tau}_k)\right\|_F > (1 - \hat{\sigma}\alpha_2^t) \left\|\phi(\hat{X}^k, \hat{S}^k, \hat{\tau}_k)\right\|_F$$

or, equivalently,

$$\begin{aligned} & \frac{\left\|\phi(\hat{X}^k+\alpha_2^t\Delta\hat{X}^k,\hat{S}^k+\alpha_2^t\Delta\hat{S}^k,\hat{\tau}_k+\alpha_2^t\Delta\hat{\tau}_k)\right\|_F-\left\|\phi(\hat{X}^k,\hat{S}^k,\hat{\tau}_k)\right\|_F}{\alpha_2^t} \\ > \quad & -\hat{\sigma}\left\|\phi(\hat{X}^k,\hat{S}^k,\hat{\tau}_k)\right\|_F \end{aligned}$$

for all $t\in\mathbb{N}$. Taking the limit $t\to\infty$ and using (40) yields

$$-\left\|\phi(\hat{X}^k,\hat{S}^k,\hat{\tau}_k)\right\|_F=\psi'(\hat{X}^k,\hat{S}^k,\hat{\tau}_k)(\Delta\hat{X}^k,\Delta\hat{S}^k,\Delta\hat{\tau}_k)\geq-\hat{\sigma}\left\|\phi(\hat{X}^k,\hat{S}^k,\hat{\tau}_k)\right\|_F.$$

Since $\hat{\sigma}\in(0,1)$, this implies $\phi(\hat{X}^k,\hat{S}^k,\hat{\tau}_k)=0$. Hence we have

$$\|\phi(\hat{X}^k+\alpha_2^t\Delta\hat{X}^k,\hat{S}^k+\alpha_2^t\Delta\hat{S}^k,\hat{\tau}_k+\alpha_2^t\Delta\hat{\tau}_k)\|_F\to 0\quad\text{for } t\to\infty,$$

whereas

$$(1-\hat{\sigma}\alpha_2^t)\beta\hat{\tau}_k\to\beta\hat{\tau}_k\quad\text{for } t\to\infty,$$

contradicting the assumption that the computation of $\hat{\eta}_k$ does not terminate in a finite number of inner loops. Hence the line search in (S.2) is also well-defined.

The statement regarding the neighbourhood $\mathcal{N}(\beta)$ follows immediately from the updating rules in Algorithm 4.1. □

5 Global Convergence

Having seen that our Algorithm 4.1 is well-defined, we now want to show that it is also globally convergent. To this end, we assume throughout this section that Algorithm 4.1 generates an infinite sequence. Under this blanket assumption, we will prove that every accumulation point of a sequence $\{W^k\}=\{(X^k,\lambda^k,S^k)\}$ generated by Algorithm 4.1 is a solution of the optimality conditions (3). To verify this result, we need the following proposition.

Proposition 5.1 *If the sequence $\{W^k\}=\{(X^k,\lambda^k,S^k)\}$ generated by Algorithm 4.1 has an accumulation point, then the sequence $\{\tau_k\}$ converges to zero.*

Proof. Since the sequence $\{\tau_k\}$ is monotonically decreasing and bounded from below by zero, it converges to a nonnegative number τ_*. If $\tau_*=0$, we are done.

So assume that $\tau_*>0$. Then the updating rules in (S.1) of Algorithm 4.1 immediately give

$$\hat{W}^k=W^k,\hat{\tau}_k=\tau_k,\ \text{and } \eta_k=1 \tag{41}$$

for all $k \in \mathbb{N}$ sufficiently large. Subsequencing if necessary, we assume without loss of generality that (41) holds for all $k \in \mathbb{N}$. Then we obtain from the updating rules in (S.2) that

$$\tau_k = \tau_0 \prod_{j=0}^{k-1} \big(1 - \hat{\sigma}\hat{\eta}_j\big).$$

Since $\tau_k \to \tau_* > 0$ by assumption, it therefore follows that $\lim_{k\to\infty} \hat{\eta}_k = 0$. Hence the stepsize $\hat{\rho}_k := \hat{\eta}_k/\alpha_2$ does not satisfy the line search criterion (29) for all $k \in \mathbb{N}$ sufficiently large. We therefore have

$$\left\| \phi(\hat{X}^k + \hat{\rho}_k \Delta\hat{X}^k, \hat{S}^k + \hat{\rho}_k \Delta\hat{S}^k, \hat{\tau}_k + \hat{\rho}_k \Delta\hat{\tau}_k) \right\|_F > (1 - \hat{\sigma}\hat{\rho}_k)\beta\hat{\tau}_k \tag{42}$$

for all these $k \in \mathbb{N}$.

Now let $W^* = (X^*, \lambda^*, S^*)$ be an accumulation point of the sequence $\{W^k\}$, and let $\{W^k\}_K$ be a subsequence converging to W^*. Since $\tau_* > 0$, it follows from (41) and Proposition 4.4 that the corresponding subsequence $\{(\Delta\hat{W}^k, \Delta\hat{\tau}_k)\}_K$ converges to some $(\Delta\hat{W}^*, \Delta\hat{\tau}_*) = (\Delta\hat{X}^*, \Delta\hat{\lambda}^*, \Delta\hat{S}^*, \Delta\hat{\tau}_*)$, where $(\Delta\hat{W}^*, \Delta\hat{\tau}_*)$ is a solution of the linear system

$$\nabla\Theta(W^*, \tau_*) \begin{pmatrix} \Delta\hat{W} \\ \Delta\hat{\tau} \end{pmatrix} = -\Theta(W^*, \tau_*) + \begin{pmatrix} 0 \\ (1 - \hat{\sigma})\tau_* \end{pmatrix}, \tag{43}$$

cf. (28). In particular, the sequence $\{(\Delta\hat{W}^k, \Delta\hat{\tau}_k)\}_K$ is bounded. Using $\{\hat{\rho}_k\}_K \to 0$ and taking the limit $k \to \infty$ on the subset K, we then obtain from (41), (42) and the continuity of the function $\phi(\cdot,\cdot,\cdot)$ that

$$\|\phi(X^*, S^*, \tau_*)\|_F \geq \beta\tau_*. \tag{44}$$

On the other hand, we obtain from (41), (42), the definition of the neighbourhood $\mathcal{N}(\beta)$ and Theorem 4.5

$$\begin{aligned} \left\| \phi(\hat{X}^k + \hat{\rho}_k \Delta\hat{X}^k, \hat{S}^k + \hat{\rho}_k \Delta\hat{S}^k, \hat{\tau}_k + \hat{\rho}_k \Delta\hat{\tau}_k) \right\|_F &> (1 - \hat{\sigma}\hat{\rho}_k)\beta\hat{\tau}_k \\ &= (1 - \hat{\sigma}\hat{\rho}_k)\beta\tau_k \\ &\geq (1 - \hat{\sigma}\hat{\rho}_k) \left\| \phi(X^k, S^k, \tau_k) \right\|_F \end{aligned}$$

for all $k \in \mathbb{N}$ sufficiently large. In view of (41), this implies

$$\frac{\left\| \phi(X^k + \hat{\rho}_k \Delta\hat{X}^k, S^k + \hat{\rho}_k \Delta\hat{S}^k, \tau_k + \hat{\rho}_k \Delta\hat{\tau}_k) \right\|_F - \left\| \phi(X^k, S^k, \tau_k) \right\|_F}{\hat{\rho}_k} > -\hat{\sigma} \left\| \phi(X^k, S^k, \tau_k) \right\|_F.$$

Letting $\psi(X, S, \tau) := \|\phi(X, S, \tau)\|_F$ and using (40), we therefore obtain for $k \to \infty$ on the subset K that

$$-\|\phi(X^*, S^*, \tau_*)\|_F \geq -\hat{\sigma} \|\phi(X^*, S^*, \tau_*)\|_F,$$

since ψ is continuously differentiable at (X^*, S^*, τ_*) (recall that $\tau_* > 0$). Since $\hat{\sigma} \in (0,1)$, this implies $\|\phi(X^*, S^*, \tau_*)\|_F = 0$, a contradiction to (44). $\square$

As a simple consequence of Proposition 5.1, we now obtain the following global convergence result for Algorithm 4.1. Note that this global convergence result depends on the corrector step only, while the precise choice of the predictor step in Algorithm 4.1 has no influence on this result.

Theorem 5.2 *Every accumulation point of a sequence $\{W^k\} = \{(X^k, \lambda^k, S^k)\}$ generated by Algorithm 4.1 is a solution of the optimality conditions* (3).

Proof. Let $W^* = (X^*, \lambda^*, S^*)$ be an accumulation point of a sequence $\{W^k\} = \{(X^k, \lambda^k, S^k)\}$ generated by Algorithm 4.1, and let $\{W^k\}_K$ be a subsequence converging to W^*. In view of Proposition 5.1, we have $\lim_{k\to\infty} \tau_k = 0$. Since all iterates belong to the neighbourhood $\mathcal{N}(\beta)$ by Theorem 4.5, we therefore obtain

$$\|\phi(X^*, S^*, 0)\|_F = \lim_{k\in K} \left\|\phi(X^k, S^k, \tau_k)\right\|_F \le \lim_{k\in K} \beta\tau_k = 0.$$

In view of (30) and Propositions 2.2 and 2.4, this implies that $W^* = (X^*, \lambda^*, S^*)$ is a solution of the optimality conditions (3). $\square$

6 Local Superlinear Convergence

This section investigates the local properties of Algorithm 4.1. Our aim is to show that the sequence $\{\tau_k\}$ converges superlinearly to zero. Since this result depends on certain properties of the predictor step in Algorithm 4.1, we first state the following assumption.

Assumption 6.1 *The sequence $\{\tau_k\}$ generated by Algorithm 4.1 converges to zero, and we have*

$$\left|\left|\left| \begin{pmatrix} \Delta W^k \\ \Delta\tau_k \end{pmatrix} \right|\right|\right| = O(\tau_k), \tag{45}$$

where $(\Delta W^k, \Delta\tau_k)$ denotes the search direction computed in (27).

In order to justify Assumption 6.1, we first note that Proposition 5.1 provides a sufficient condition for the sequence $\{\tau_k\}$ to converge to zero. To understand the second condition, assume that the sequence of inverse operators $\nabla\Theta(W^k, \tau_k)^{-1}$ remains bounded for $k \to \infty$. Then we obtain from the linear system (27) that (45) holds provided that the right-hand side in (27) is of the order $O(\tau_k)$. This, however, is rather obvious since the feasibility of the iterates (cf. (30)) together with the fact that all iterates belong to the neighbourhood $\mathcal{N}(\beta)$ (cf. Theorem 4.5) show that

$$|||\Theta(W^k, \tau_k)||| = \sqrt{\|\phi(X^k, S^k, \tau_k)\|_F^2 + \tau_k^2} \le \left\|\phi(X^k, S^k, \tau_k)\right\|_F + \tau_k \le \beta\tau_k + \tau_k = O(\tau_k).$$

In addition, such a relation also holds if we replace the right-hand side in (27) by $-\Theta(W^k, 0)$ since then Corollary 3.2 and Theorem 4.5 imply

$$\begin{aligned} \|\Theta(W^k,0)\| &= \|\phi(X^k,S^k,0)\|_F \\ &\le \|\phi(X^k,S^k,\tau_k) - \phi(X^k,S^k,0)\|_F + \|\phi(X^k,S^k,\tau_k)\|_F \\ &\le \kappa\sqrt{n}\tau_k + \beta\tau_k \\ &= O(\tau_k), \end{aligned}$$

where $\kappa > 0$ denotes the constant from Lemma 3.1. In particular, all global and local convergence properties of Algorithm 4.1 remain true if we use this modification of the right-hand side in (27).

In order to state a sufficient condition for Assumption 6.1 to be satisfied, we introduce the following assumption.

Assumption 6.2 *Let (X^*, λ^*, S^*) be a solution of the optimality conditions (3) such that*

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

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

$$\sum_{i=1}^m \Delta\lambda_i A_i + \Delta S = 0 \qquad \textit{and} \qquad A_i \bullet \Delta X = 0 \ \ (i = 1, \ldots, m),$$

the following implication holds:

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

Assumption 6.2 (a) is rather standard, and Assumption 6.2 (b) was introduced by Kojima et al. [20]. As noted in [20], Haeberly showed that this assumption is equivalent to the primal and dual nondegeneracy condition considered by Alizadeh et al. [2].

The next result implies that Assumption 6.1 holds under Assumptions 4.3 and 6.2 provided that the iterates (X^k, λ^k, S^k) generated by Algorithm 4.1 converge to a solution (X^*, λ^*, S^*) satisfying these two conditions. The convergence of the iterates to this single point is not at all restrictive since it is known that the two Assumptions 4.3 and 6.2 together imply that (X^*, λ^*, S^*) is the unique solution of the optimality conditions (3).

Theorem 6.3 *Suppose that Assumptions 4.3 and 6.2 hold at a solution (X^*, λ^*, S^*) of (3). Then the linear mapping $\nabla\Theta(X^*, \lambda^*, S^*, 0)$ is bijective.*

Proof. We only consider the case where ϕ is defined via the smoothed Fischer-Burmeister function from (18). The proof for the smoothed minimum function from (19) is similar.

Let us define $E := \big((X^*)^2 + (S^*)^2\big)^{1/2}$. In view of the assumed strict complementarity, it is easy to see that E is a positive definite matrix. Hence Theorem 3.4 implies that Θ

is continuously differentiable at $(X^*, \lambda^*, S^*, 0)$. In order to see that $\nabla\Theta(X^*, \lambda^*, S^*, 0)$ is bijective, we only have to verify that it is one-to-one. To this end, we consider the equation

$$\nabla\Theta(X^*, \lambda^*, S^*, 0) \begin{pmatrix} \Delta X \\ \Delta\lambda \\ \Delta S \\ \Delta\tau \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and show that $(\Delta X, \Delta\lambda, \Delta S, \Delta\tau) = (0, 0, 0, 0)$ is its only solution. The last row immediately gives

$$\Delta\tau = 0. \tag{46}$$

Taking this into account and using Theorem 3.4, the first three block rows can be rewritten as follows:

$$\begin{aligned} \sum_{i=1}^{m} \Delta\lambda_i A_i + \Delta S &= 0, & (47)\\ A_i \bullet \Delta X &= 0 \quad (i = 1, \ldots, m), & (48)\\ \Delta X + \Delta S - L_E^{-1}[X^*\Delta X + \Delta X X^* + S^*\Delta S + \Delta S S^*] &= 0. & (49) \end{aligned}$$

Equation (49) implies

$$L_{E-X^*}[\Delta X] + L_{E-S^*}[\Delta S] = 0, \tag{50}$$

cf. the proof of Proposition 4.4. Now, using the fact that (X^*, λ^*, S^*) is a strictly complementary solution of (3) so that, in particular, we have $X^*S^* = 0$, i.e., X^* and S^* commute, it follows that these two matrices can be diagonalized simultaneously by an orthogonal transformation. This means that we can find a single orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ and diagonal matrices $D_X \in \mathbb{R}^{n\times n}$ and $D_S \in \mathbb{R}^{n\times n}$ such that $X^* = Q^T D_X Q$ and $S^* = Q^T D_S Q$. Taking this into account, an easy calculation shows that

$$E - X^* = S^* \quad \text{and} \quad E - S^* = X^*.$$

Hence (50) can be rewritten as

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

Using (47), (48) and Assumption 6.2, we therefore obtain from [21, Lemma 6.2] that $(\Delta X, \Delta S) = (0, 0)$. Since the matrices A_i are linearly independent by Assumption 4.3, it follows from (47) that $\Delta\lambda = 0$. In view of (46), this completes the proof. □

We stress that Theorem 6.3 provides only a sufficient condition for Assumption 6.1 to be satisfied. Since the assumptions used in Theorem 6.3 do imply that the solution set of the optimality conditions (3) is just a singleton, Theorem 6.3 is somewhat restrictive. However, some recent results obtained for linear programs and complementarity problems indicate that Assumption 6.1 may also hold under weaker conditions which do not necessarily imply the unique solvability of (3), cf. Tseng [28] and [10].

We now start to analyze the local behaviour of Algorithm 4.1, and begin with the following technical result.

Lemma 6.4 *Suppose Assumption 6.1 holds. Then we have*

$$\left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \tau_k + \Delta\tau_k)\right\|_F = o(\tau_k).$$

Proof. Since

$$\nabla\phi(X^k, S^k, \tau_k)\begin{pmatrix}\Delta X^k\\ \Delta S^k\\ \Delta\tau_k\end{pmatrix} = -\phi(X^k, S^k, \tau_k)$$

by (27), we obtain from the integral mean value theorem that

$$\begin{aligned}
&\left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \tau_k + \Delta\tau_k)\right\|_F \\
= &\left\|\int_0^1 \nabla\phi(X^k + \eta\Delta X^k, S^k + \eta\Delta S^k, \tau_k + \eta\Delta\tau_k)\begin{pmatrix}\Delta X^k\\ \Delta S^k\\ \Delta\tau_k\end{pmatrix} \mathrm{d}\eta + \phi(X^k, S^k, \tau_k)\right\|_F \\
= &\left\|\int_0^1 \nabla\phi(X^k + \eta\Delta X^k, S^k + \eta\Delta S^k, \tau_k + \eta\Delta\tau_k)\begin{pmatrix}\Delta X^k\\ \Delta S^k\\ \Delta\tau_k\end{pmatrix} \mathrm{d}\eta - \nabla\phi(X^k, S^k, \tau_k)\begin{pmatrix}\Delta X^k\\ \Delta S^k\\ \Delta\tau_k\end{pmatrix}\right\|_F \\
\leq &\int_0^1 \left\|\left[\nabla\phi(X^k + \eta\Delta X^k, S^k + \eta\Delta S^k, \tau_k + \eta\Delta\tau_k) - \nabla\phi(X^k, S^k, \tau_k)\right]\begin{pmatrix}\Delta X^k\\ \Delta S^k\\ \Delta\tau_k\end{pmatrix}\right\|_F \mathrm{d}\eta \\
= &\; o\left(\left\|\begin{pmatrix}\Delta X^k\\ \Delta S^k\\ \Delta\tau_k\end{pmatrix}\right\|\right),
\end{aligned}$$

where the last equality follows from the continuous differentiability of the mapping ϕ. Taking into account Assumption 6.1, we therefore get $\left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \tau_k + \Delta\tau_k)\right\|_F = o(\tau_k)$. $\square$

The main step in order to prove local superlinear convergence of the sequence $\{\tau_k\}$ is contained in the following result.

Lemma 6.5 *Suppose Assumption 6.1 holds, and let the constant β satisfy the inequality $\beta > \kappa\sqrt{n}$, where κ denotes the constant from Lemma 3.1. Then the sequence $\{\eta_k\}$ converges to zero.*

Proof. Let $\varepsilon > 0$ be arbitrarily given. Using the fact that $\Delta\tau_k = -\tau_k$ because of (27), we obtain from Lemma 6.4 that

$$\left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, 0)\right\| = \left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \tau_k + \Delta\tau_k)\right\|_F = o(\tau_k).$$

Hence there is an index $K_\varepsilon \in \mathbb{N}$ such that

$$\left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, 0)\right\|_F \leq \varepsilon\tau_k \qquad \forall k \geq K_\varepsilon.$$

Then we get for all $\eta > 0$ and all $k \geq K_\varepsilon$

$$\begin{aligned}
& \left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \eta\tau_k)\right\|_F \\
\leq\ & \left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, 0)\right\|_F + \\
& \left\|\phi(X^k + \Delta X^k, S^k + \Delta S^k, \eta\tau_k) - \phi(X^k + \Delta X^k, S^k + \Delta S^k, 0)\right\|_F \\
\leq\ & \varepsilon\tau_k + \kappa\sqrt{n}\eta\tau_k,
\end{aligned}$$

where the last inequality follows from Corollary 3.2. Since the inequality

$$\varepsilon\tau_k + \kappa\sqrt{n}\eta\tau_k \leq \beta\eta\tau_k$$

holds for all $\eta \geq \frac{\varepsilon}{\beta - \kappa\sqrt{n}}$, the definition of η_k shows that $\eta_k\alpha_1$ does not satisfy this inequality, i.e.,

$$\eta_k < \frac{\varepsilon}{(\beta - \kappa\sqrt{n})\alpha_1}.$$

Since $\beta - \kappa\sqrt{n} > 0$ by assumption and $\varepsilon > 0$ was chosen arbitrarily, this implies $\eta_k \to 0$. □

We are now in the position to state the main local convergence result for Algorithm 4.1.

Theorem 6.6 *Under Assumption 6.1 we have $\tau_{k+1} = o(\tau_k)$, i.e., the smoothing parameter converges locally superlinearly to zero.*

Proof. Using Lemma 6.5 and the definition of τ_{k+1} and $\hat{\tau}_k$ in Algorithm 4.1, we obtain

$$\tau_{k+1} \leq \hat{\tau}_k = \eta_k\tau_k = o(\tau_k),$$

i.e., $\tau_k \to 0$ superlinearly. □

We close this section by noting that Theorem 6.6 still holds if we would replace the right-hand side in (27) by $-\Theta(W^k, 0)$. This follows from the analysis carried out in [8] (so we skip the details here) and does not follow immediately from our previous discussion since both Lemma 6.4 and Lemma 6.5 depend on the fact that the right-hand side of (27) is given by $-\Theta(W^k, \tau_k)$.

7 Numerical Results

In order to test the numerical performance of Algorithm 4.1, we implemented the method in Matlab. To simplify the programming work, we borrow the data structure, problem input, and some linear algebra routines from the SDPT^3 (version 2.1) Matlab code by Toh, Todd, and Tütüncü [26].

In our Matlab implementation of Algorithm 4.1 we choose ϕ to be the smoothed minimum function from (19). (The results for the smoothed Fischer-Burmeister function seem to be similar.) Furthermore, we take $\alpha_1 = \alpha_2 = 0.5$. The centering parameter $\hat{\sigma}$ gets updated dynamically using a procedure suggested in [12] for the solution of linear programs.

In order to see how the Newton directions can be computed, let us first consider one iteration of the predictor step. Dropping the superscript k and using the abbreviation

$$R_d = C - \sum_{j=1}^m \lambda_j A_j - S,$$

the predictor step (27) (with the modification mentioned in Section 6 that the right-hand side $-\Theta(W,\tau)$ gets replaced by $-\Theta(W,0)$) becomes

$$\begin{aligned} \sum_{j=1}^m \Delta\lambda_j A_j + \Delta S &= R_d, && (51)\\ A_i \bullet \Delta X &= b_i - A_i \bullet X \quad (i=1,\dots,m), && (52)\\ \nabla\phi(X,S,\tau)(\Delta X,\Delta S,\Delta\tau) &= -\phi(X,S,0), && (53)\\ \Delta\tau &= 0\,. && (54)\end{aligned}$$

Writing $E := ((X-S)^2 + 4\tau^2 I)^{1/2}$ (cf. Theorem 3.4), applying the corresponding Lyapunov operator L_E on both sides of equation (53) and using (54), we obtain

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

or, equivalently,

$$\Delta X = -L^{-1}_{E-(X-S)}\left[L_{E+(X-S)}[\Delta S] + L_E[\phi(X,S,0)]\right]\,. \tag{55}$$

Substituting ΔS from (51) and rearranging terms yields

$$\begin{aligned}\Delta X \;=\; & \sum_{j=1}^m \Delta\lambda_j L^{-1}_{E-(X-S)}\left[L_{E+(X-S)}[A_j]\right]\\ & - L^{-1}_{E-(X-S)}\left[L_{E+(X-S)}[R_d] + L_E[\phi(X,S,0)]\right]\,.\end{aligned}$$

Taking inner products with A_i $(i=1,\dots,m)$ and using the fact that $L^{-1}_{E-(X-S)}$ is self-adjoint by Lemma 4.2 (b), we obtain from (52)

$$\begin{aligned}&\sum_{j=1}^m \Delta\lambda_j L_{E+(X-S)}[A_j] \bullet L^{-1}_{E-(X-S)}[A_i] = b_i - A_i \bullet X\\ &\quad + \big(L_{E+(X-S)}[R_d] + L_E[\phi(X,S,0)]\big) \bullet L^{-1}_{E-(X-S)}[A_i], \quad i = 1,\dots,m\,.\end{aligned} \tag{56}$$

This is a linear equation in the variables $\Delta\lambda \in \mathbb{R}^m$. After solving this system, we immediately get ΔS from (51). Note that ΔS is obviously symmetric since R_d and all A_i are symmetric. In view of (55), ΔX can then be obtained as a solution of a Lyapunov equation with a symmetric right-hand side and is therefore also symmetric, cf. [17, Theorem 2.2.3]. The solution of this Lyapunov equation may be computed by using a spectral decomposition of $(X-S)^2+4\tau^2 I$, see [17, p. 100].

The computation of the search direction in the corrector step (28) is similar to the one of the predictor step. The main difference is that we compute the vector $\Delta\hat{\lambda}$ by solving the linear system

$$\begin{aligned}\sum_{j=1}^{m} \Delta\hat{\lambda}_j L_{\hat{E}+(\hat{X}-\hat{S})}[A_j] \bullet L^{-1}_{\hat{E}-(\hat{X}-\hat{S})}[A_i] &= b_i - A_i \bullet \hat{X} \\ + \Big(L_{\hat{E}+(\hat{X}-\hat{S})}[\hat{R}_d] + L_{\hat{E}}[\phi(\hat{X},\hat{S},\hat{\tau})] - 8\sigma\hat{\tau}^2 I\Big) &\bullet L^{-1}_{\hat{E}-(\hat{X}-\hat{S})}[A_i], \quad i=1,\dots,m,\end{aligned} \tag{57}$$

rather than (56), where, of course, we have used the notation

$$\begin{aligned}(\hat{X},\hat{\lambda},\hat{S}) &:= (\hat{X}^k,\hat{\lambda}^k,\hat{S}^k),\\ \hat{R}_d &:= C - \sum_{j=1}^{m}\hat{\lambda}_j A_j - \hat{S},\\ \hat{E} &:= ((\hat{X}-\hat{S})^2+4\hat{\tau}^2 I)^{1/2}.\end{aligned}$$

Note, however, that the corrector step is not carried out when the predictor step was accepted with $\eta_k < 1$. Hence, either the algorithm uses only a predictor step in one iteration, or the two matrices in (56) and (57) coincide.

In order to describe the way we compute our starting point (X^0,λ^0,S^0), let us call a triple (X,λ,S) *feasible* for the optimality conditions (3) if it satisfies the linear equations $\sum_{i=1}^m \lambda_i A_i + S = C$ (this will be called *dual feasibility*) and $A_i \bullet X = b_i$ $(i=1,\dots,m)$ (this will be called *primal feasibility*). Note that we do not require $X \succeq 0$ or $S \succeq 0$ for such a feasible triple. Of course, our starting point (X^0,λ^0,S^0) should be feasible in this sense.

To this end, we define a symmetric matrix $\mathcal{A}\in\mathbb{R}^{m\times m}$ by

$$\mathcal{A}_{ij} = A_i \bullet A_j\,, \qquad i,j=1,\dots,m,$$

and solve the linear system $\mathcal{A}y=b$ to obtain $y^0\in\mathbb{R}^m$. Then we define

$$X^0 = \sum_{i=1}^{m} y_i^0 A_i$$

and compute λ^0 as a solution of the system $\mathcal{A}\lambda = (A_1\bullet C,\dots,A_m\bullet C)^T$. Finally, setting

$$S^0 = C - \sum_{i=1}^{m}\lambda_i^0 A_i\,,$$

we obtain a starting point (X^0,λ^0,S^0) that is obviously feasible. Note, however, that both X^0 and S^0 may have negative eigenvalues.

Having computed this starting point, the remaining parameters of Algorithm 4.1 are initialized by

$$\tau_0 = \left\|\phi(X^0,S^0,0)\right\|/5 \ \text{ and } \beta = \max\left\{2.1\cdot\sqrt{n}, 1.5\cdot\|\phi(X,S,\tau_0)\|/\tau_0\right\}.$$

We terminate the iteration if $\tau_k/n < 10^{-6}$ (recall that we parameterize the central path conditions by τ^2) and if the feasibility measure

$$\max\left\{\frac{\left\|\left[b_i - A_i \bullet X^k\right]_{i=1}^m\right\|_2}{\max\{1, \|b\|_2\}}, \frac{\left\|C - S^k - \sum_{i=1}^m \lambda_i^k A_i\right\|_F}{\max\{1, \|C\|_2\}}\right\}$$

is smaller than 10^{-10}. The reason for dividing τ_k by n is based on the fact that $\|\phi(X^k, S^k, 0)\|_F = O(\tau_k)$. Since we want to have $\|\phi(X^k, S^k, 0)\|_F$ small, it seems reasonable to terminate if τ_k gets small. However, getting $\|\phi(X^k, S^k, 0)\|_F$ small becomes increasingly more difficult the larger the dimension of the matrices X^k and S^k are since we take the Frobenius norm. In order to make our termination criterion more or less independent of the dimension of X^k and S^k, we therefore decided to use the above-mentioned stopping rule.

Note that, theoretically, this feasibility measure is always zero for our method. Numerically, however, the situation is different. While the dual feasibility does not really cause any troubles (mainly because S^k gets defined in such a way that the dual feasibility is zero), we sometimes observed difficulties with respect to the primal feasibility. In order to decrease the primal infeasibility, we therefore exploit a projection technique also used in SDPT^3: After computing a Newton direction $(\Delta X, \Delta\lambda, \Delta S, \Delta\tau)$, we check whether the inequality $\|[A_i \bullet (X + \Delta X)]_{i=1}^m - b\| > \|[A_i \bullet X]_{i=1}^m - b\|$ holds. If this inequality is satisfied, we replace ΔX by its orthogonal projection onto the nullspace $\{U \in \mathcal{S}^{n\times n} \mid A_i \bullet U = 0,\ i = 1, \ldots, m\}$. As a consequence of this procedure, the feasibility stays close to the machine precision for all test problems.

In the SDPT^3 code, there are eight test problems. The results for different sizes are shown in Tables 1 – 4. To compare the results with those from interior point methods, the number of iterations for the infeasible path following algorithm from the SDPT^3 package are also printed; more precisely, we present the results for the three most popular interior point methods, namely those based on the AHO-, HKM-, and NT-directions, see, e.g., [26] for some further details.

In Table 1 we tabulate the single run iteration counts (niter). Because it is not garantueed by Algorithm 4.1 that X and S are positive semi-definite, we also report the minimal eigenvalues of X and S at the final iterate. Table 1 indicates that Algorithm 4.1 has better iteration counts than all interior-point methods except for the Lovasz problem. However, we should note that one iteration of Algorithm 4.1 is (usually) more expensive than one iteration of an interior-point method due to the fact that we have to calculate a matrix square root.

In Tables 2 – 4 we report the average iteration counts for the first ten instances of each problem using different problem dimensions. (Note that all test problems depend on some random numbers, so we decided to give the average results over ten runs for each problem.) In general, it seems that Algorithm 4.1 needs less many iterations than all interior-point methods for smaller problems, whereas the number of iterations is comparable to interior-point methods for larger problems. Of course, the precise behaviour also depends on the particular test problem, and it might be an interesting future research topic to investigate the reasons why our smoothing-type method behaves particularly well on some problems while it has more difficulties for some other problems. In any case, we stress that the results we obtain for Algorithm 4.1 seem to be considerably better than those reported for a related

			AHO	HKM	NT	Alg. 4.1		
Problem	**n**	**m**	**niter**	**niter**	**niter**	**niter**	$\lambda_{\min}(\mathbf{X})$	$\lambda_{\min}(\mathbf{S})$
random	10	10	9	16	15	7	-1.91e-08	-2.77e-09
Norm min	20	6	9	10	10	8	-1.97e-09	-6.49e-09
Cheby	20	11	7	10	10	6	-2.58e-13	-3.17e-08
Maxcut	10	10	7	8	8	6	-8.04e-10	-9.35e-10
ETP	20	10	11	14	12	6	-5.86e-07	-4.26e-07
Lovasz	10	20	9	9	9	11	-9.42e-09	-1.12e-06
LogCheby	60	6	10	11	11	11	-9.06e-08	-2.68e-10
ChebyC	40	11	7	8	9	5	-4.74e-12	3.03e-10

Table 1: Number of iterations for small SDP (single run).

			AHO	HKM	NT	Alg. 4.1
Problem	**n**	**m**	**niter**	**niter**	**niter**	**niter**
random	10	10	8.2	13.5	12.6	6.5
Norm min	20	6	8.0	9.3	10.1	6.7
Cheby	20	11	7.9	9.8	9.9	5.9
Maxcut	10	10	7.4	8.2	8.4	5.5
ETP	20	10	11.8	14.7	12.2	10.8
Lovasz	10	≈ 25	7.6	8.7	8.7	9.1
LogCheby	60	6	10.4	10.9	11.1	10.8
ChebyC	40	11	7.6	8.5	9.0	5.2

Table 2: Average number of iterations for small SDP.

method by Chen and Tseng [8] (it should be noted, however, that the termination criteria are different and not directly comparable).

Finally, Table 5 gives some results for Algorithm 4.1 being applied to some test problems from the SDPLIB, cf. Borchers [3]. In this table, we present for each test problem the number of iterations, the final value of the smoothing parameter τ, the relative duality gap as well as the feasibility measure at the final iterate. Note that the duality gap is negative for many test problems because the matrices generated by our method are not necessarily positive semi-definite. (This, in fact, was the reason why we had to take a different termination criterion than interior-point methods.)

8 Final Remarks

We have presented two new characterizations of the central path conditions for semidefinite programs. These characterizations were used in order to derive a smoothing-type method for the solution of semidefinite programs. The search directions generated by these methods are automatically symmetric, and the method was shown to be globally and locally superlinearly convergent under suitable assumptions. The numerical results are very promising and it is certainly worth to do some more work in order to improve these methods. For example, it

			AHO	HKM	NT	Alg. 4.1
Problem	**n**	**m**	**niter**	**niter**	**niter**	**niter**
random	20	20	10.2	14.4	13.1	8.9
Norm min	40	11	8.5	10.1	10.7	7.7
Cheby	40	21	7.7	9.7	10.0	6.1
Maxcut	21	21	8.2	9.6	9.6	6.3
ETP	40	20	12.6	16.7	13.3	14.0
Lovasz	21	≈ 105	9.7	10.1	10.4	12.7
LogCheby	120	11	12.3	13.2	13.1	13.5
ChebyC	80	21	8.3	9.2	9.4	6.1

Table 3: Average number of iterations for medium sized SDP.

			AHO	HKM	NT	Alg. 4.1
Problem	**n**	**m**	**niter**	**niter**	**niter**	**niter**
random	50	50	10.4	15.6	13.7	10.9
Norm min	100	26	9.4	10.7	11.2	8.8
Cheby	100	27	9.3	10.4	11.4	7.1
Maxcut	50	50	9.0	10.0	10.5	6.7
ETP	100	50	13.7	18.4	15.1	19.1
Lovasz	30	≈ 220	10.3	10.6	10.7	15.3
LogCheby	300	51	13.6	14.0	13.7	13.6
ChebyC	200	41	9.0	9.8	10.0	6.8

Table 4: Average number of iterations for large SDP.

Problem	n	m	niter	τ	rel. gap.	inf. measure
arch0	335	174	44	5.9e-05	-1.037695e-05	3.499271e-13
arch2	335	174	43	9.4e-05	5.195117e-05	6.384054e-13
arch4	335	174	47	1.3e-04	7.104246e-05	4.250081e-13
arch8	335	174	78	1.1e-04	-3.542500e-06	9.052898e-13
gpp100	100	101	18	9.9e-05	-3.042351e-06	2.123789e-15
gpp124-1	124	125	19	1.0e-04	-1.912270e-05	1.195466e-14
gpp124-2	124	125	19	7.0e-05	-2.615799e-06	2.176315e-15
gpp124-3	124	125	16	1.1e-04	-2.025964e-06	1.887616e-15
gpp124-4	124	125	20	6.5e-05	-4.630696e-06	1.331054e-14
gpp250-1	250	250	19	2.1e-04	-1.294246e-05	2.476407e-14
gpp250-2	250	250	17	1.9e-04	-5.718274e-06	2.130110e-14
gpp250-3	250	250	16	1.7e-04	-3.424270e-06	1.258957e-14
gpp250-4	250	250	17	2.2e-04	-2.061362e-06	3.875924e-14
mcp100	100	100	10	1.8e-06	-2.068888e-09	6.683366e-16
mcp124-1	124	124	15	2.6e-05	-5.421892e-09	5.389812e-16
mcp124-2	124	124	10	8.6e-05	-2.899952e-07	7.948236e-16
mcp124-3	124	124	9	2.9e-05	-8.401942e-08	6.089117e-16
mcp124-4	124	124	9	5.8e-07	-1.672955e-09	7.768182e-16
mcp250-1	250	250	14	5.7e-05	-9.849429e-08	9.333300e-16
mcp250-2	250	250	11	1.1e-04	-4.597849e-07	1.003759e-15
mcp250-3	250	250	11	8.7e-05	-1.589781e-07	1.081043e-15
mcp250-4	250	250	11	4.9e-05	-1.219282e-07	1.007461e-15
mcp500-1	500	500	26	3.7e-04	-5.734057e-07	1.087701e-15
mcp500-2	500	500	14	1.4e-04	-3.508263e-07	1.429353e-15
mcp500-3	500	500	11	3.7e-04	-1.295995e-06	1.526598e-15
mcp500-4	500	500	10	5.6e-05	-7.792705e-08	1.574678e-15
theta1	50	104	13	3.9e-05	-1.307451e-07	6.261965e-17
theta2	100	498	15	1.7e-05	-1.766186e-07	1.049632e-14
theta3	150	1106	15	7.8e-05	-1.009075e-06	1.998401e-15
theta4	200	1949	15	9.5e-06	-1.006602e-07	3.996803e-15
truss1	13	6	8	3.3e-09	-3.003989e-09	3.621438e-15
truss2	133	58	13	1.3e-05	-7.869353e-06	2.209316e-14
truss3	31	27	14	5.0e-06	-5.614971e-10	2.660288e-15
truss4	19	12	7	1.3e-05	-3.397473e-05	1.324462e-15
truss5	331	208	16	2.1e-04	-6.840872e-07	1.803378e-14
truss6	451	172	21	2.5e-04	-2.430952e-04	4.601362e-13
truss7	301	86	25	5.9e-05	-2.316906e-07	3.795188e-13
truss8	628	496	20	1.7e-04	-4.805725e-06	3.038575e-14

Table 5: Selected problems from SDPLIB

is interesting to investigate the question how the matrix square roots can be computed in a more efficient way.

References

[1] F. Alizadeh: *Interior point methods in semidefinite programming with applications to combinatorial optimization.* SIAM Journal on Optimization 5, 1995, pp. 13–51.

[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] B. Borchers: *SDPLIB 1.2, A library of semidefinite programming test problems.* Optimization Methods and Software 11, 1999, pp. 597–611.

[4] J.V. Burke and S. Xu: *A non-interior predictor-corrector path-following method for LCP.* In M. Fukushima and L. Qi (eds.): *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods.* Kluwer Academic Publishers, Dordrecht, Netherlands, 1999, pp. 45–63.

[5] J.V. Burke and S. Xu: *A non-interior predictor-corrector path following algorithm for the monotone linear complementarity problem.* Mathematical Programming 87, 2000, pp. 113–130.

[6] B. Chen and P.T. Harker: *A non-interior-point continuation method for linear complementarity problems.* SIAM Journal on Matrix Analysis and Applications 14, 1993, pp. 1168–1190.

[7] X. Chen, H. Qi and P. Tseng: *Analysis of nonsmooth symmetric-matrix functions with applications to semidefinite complementarity problems.* Technical Report, Department of Mathematics, University of Washington, Seattle, WA, October 2000.

[8] 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.

[9] S. Engelke and C. Kanzow: *On the solution of linear programs by Jacobian smoothing methods.* Annals of Operations Research, to appear.

[10] S. Engelke and C. Kanzow: *Predictor-corrector smoothing methods for the solution of linear programs.* Preprint 153, Department of Mathematics, University of Hamburg, Hamburg, March 2000.

[11] S. Engelke and C. Kanzow: *Improved smoothing-type methods for the solution of linear programs.* Numerische Mathematik, to appear.

[12] S. Engelke and C. Kanzow: *Predictor-corrector smoothing methods for linear programs with a more flexible update of the smoothing parameter.* Preprint 162, Department of Mathematics, University of Hamburg, Hamburg, February 2001.

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

[14] M. Fukushima, Z.-Q. Luo and P. Tseng: *Smoothing functions for second-order-cone complementarity problems.* Technical Report, Department of Mathematics, University of Washington, Seattle, WA, November 2000.

[15] 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.

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

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

[18] H. Jiang: *Smoothed Fischer-Burmeister equation methods for the complementarity problem.* Technical Report, Department of Mathematics, University of Melbourne, Melbourne, Australia, June 1997.

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

[20] 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.

[21] 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.

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

[23] S. Smale: *Algorithms for solving equations.* In: *Proceedings of the International Congress of Mathematicians.* AMS, Providence, 1987, pp. 172–195.

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

[25] 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.

[26] K.C. Toh, M.J. Todd and R.H. Tütüncü: *SDPT3 — a Matlab software package for semidefinite programming, version 2.1.* Optimization Methods and Software 11, 1999, pp. 545–581.

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

[28] P. Tseng: *Error bounds and superlinear convergence analysis of some Newton-type methods in optimization.* In: G. Di Pillo and F. Giannessi (eds.): *Nonlinear Optimization and Related Topics.* Kluwer Academic Publishers, to appear.

[29] 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.