Title

Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling

Journal

Mathematical Programming, 109(2-3)

ISSN

0025-5610

Authors

Freund, Roland W
Jarre, F
Vogelbusch, C H

Publication Date

2007-03-01

Peer reviewed

Mathematical Programming manuscript No.
(will be inserted by the editor)

Roland W. Freund · **Florian Jarre** · **Christoph H. Vogelbusch**

Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling

Received: date / Accepted: date

Abstract We consider the solution of nonlinear programs with nonlinear semidefiniteness constraints. The need for an efficient exploitation of the cone of positive semidefinite matrices makes the solution of such nonlinear semidefinite programs more complicated than the solution of standard nonlinear programs. This paper studies a sequential semidefinite programming (SSP) method, which is a generalization of the well-known SQP method for standard nonlinear programs. We present a sensitivity result for nonlinear semidefinite programs, and then based on this result, we give a self-contained proof of local quadratic convergence of the SSP method. We also describe a class of nonlinear semidefinite programs that arise in passive reduced-order modeling, and we report results of some numerical experiments with the SSP method applied to problems in that class.

Keywords nonlinear semidefinite programming · sequential quadratic programming · sensitivity · reduced-order modeling · positive realness

Mathematics Subject Classification (2000) 90C22 · 90C30 · 65K05

R. W. Freund
Department of Mathematics
University of California at Davis
One Shields Avenue
Davis, California 95616, U.S.A.
E-mail: freund@math.ucdavis.edu

F. Jarre and C. H. Vogelbusch
Institut für Mathematik
Universität Düsseldorf
Universitätsstraße 1
D–40225 Düsseldorf, Germany
E-mail: {jarre,vogelbuc}@opt.uni-duesseldorf.de

1 Introduction

In recent years, interior-point methods for solving linear semidefinite programs (SDPs) have received a lot of attention, and as a result, these methods are now very well developed; see, e.g., [36,38], the papers in [41], and the references given there. At each iteration of an interior-point method, the complementarity condition is relaxed, symmetrized, and linearized. Various symmetrization operators are known. The choice of the symmetrization operator and of the relaxation parameter determine the step length at each iteration, and thus the efficiency of the overall method. The computation of the search direction is often carried out by forming a Schur complement matrix. For the HKM search direction [16,19,22] and the NT search direction [24,25], this matrix is positive definite, and this is the essential property that guarantees a descent property of the search step. On the other hand, for the AHO search direction [2], the Schur complement matrix is not symmetric.

In this paper, we are concerned with the solution of nonlinear semidefinite programs (NLSDPs). We consider an approach that separates the linearization and the symmetrization in a natural way, namely a generalization of the sequential quadratic programming (SQP) method for standard nonlinear programs. Such a generalization has already been mentioned by Robinson [31] within the more general framework of nonlinear programs over convex cones. This framework includes NLSDPs as a special case. While Robinson did not discuss implementational issues of such a generalized SQP approach, the recent progress in the solution of linear SDPs makes this approach especially interesting for the solution of NLSDPs.

We first present a derivation of a generalized SQP method, namely the sequential semidefinite programming (SSP) method, for solving NLSDPs. In order to analyze the convergence of the SSP method, we present a sensitivity result for certain local optimal solutions of general, possibly nonconvex, quadratic semidefinite programs. We then use this result to derive a self-contained proof of local quadratic convergence of the SSP method under the assumptions that the optimal solution is locally unique, strictly complementary, and satisfies a second-order sufficient condition. Our analysis avoids the consideration of the symmetrization operator, because we assume that the linear or quadratic SDP subproblems are solved to optimality by using any of the common symmetrization techniques.

One of the first numerical approaches for solving a class of NLSDPs was given in [27,28]. Other recent approaches for solving NLSDPs are the program package LOQO [39] based on a primal-dual method; see also [40]. Another promising approach for solving large-scale SDPs is the modified-barrier method proposed in [18]. This modified-barrier approach does not require the barrier parameter to converge to zero, and may thus overcome some of the problems related to ill-conditioning in traditional interior-point methods. Further approaches to solving NLSDPs have been presented in [11,12,37]. In [11], the augmented Lagrangian method is applied to NLSDPs, while the approach proposed in [12] is based on an SQP method generalized to NLSDPs. The paper [12] also contains a proof of local quadratic convergence. However, in contrast to this paper, the algorithm [12] is not derived from a comparison with interior-point algorithms, and the proof of convergence does not use any differentiability properties of the optimal solutions. In [10], Correa and Ramírez present a proof of global convergence of a modifi-

cation of the method proposed in [12]. The modification employs certain merit functions to control the step lengths of the SQP algorithm.

The remainder of this paper is organized as follows. In Section 2, we introduce some notation. In Section 3, we describe a class of nonlinear semidefinite programs that arise in passive reduced-order modeling. In Section 4, we recall known results for linear SDPs in a form that can be easily transferred to NLSDPs. In Section 5, we discuss primal-dual systems for NLSDPs, and in Section 6, the SSP method is introduced as a generalized SQP method. In Section 7, we present sensitivity results, first for a certain class of quadratic SDPs, and then for general NLSDPs. Based on these sensitivity results, in Section 8, we give a self-contained proof of local quadratic convergence of the SSP method. In Section 9, we present results of some numerical experiments. Finally, in Section 10, we make some concluding remarks.

2 Notation

Throughout this article, all vectors and matrices are assumed to have real entries. As usual, $Y^T = \begin{bmatrix} y_{ji} \end{bmatrix}$ denotes the transpose of the matrix $Y = \begin{bmatrix} y_{ij} \end{bmatrix}$. The vector norm $\|x\| := \sqrt{x^T x}$ is always the Euclidean norm and $\|Y\| := \max_{\|x\|=1} \|Yx\|$ is the corresponding matrix norm. For vectors $x \in \mathbb{R}^n$, $x \geq 0$ means that all entries of x are nonnegative, and $\mathrm{Diag}(x)$ denotes the $n \times n$ diagonal matrix the diagonal entries of which are the entries of x. The $n \times n$ identity matrix is denoted by I_n.

The trace inner product on the space of real $n \times m$ matrices is given by

$$\langle Z, Y \rangle := Z \bullet Y := \mathrm{trace}(Z^T Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij} y_{ij}$$

for any pair $Y = \begin{bmatrix} y_{ij} \end{bmatrix}$ and $Z = \begin{bmatrix} z_{ij} \end{bmatrix}$ of $n \times m$ matrices. The space of real symmetric $m \times m$ matrices is denoted by $\mathcal{S}^m$. The notation $Y \succeq 0$ ($Y \succ 0$) is used to indicate that $Y \in \mathcal{S}^m$ is symmetric positive semidefinite (positive definite).

Semidefiniteness constraints are expressed by means of matrix-valued functions from $\mathbb{R}^n$ to $\mathcal{S}^m$. We use the symbol $\mathcal{A} : \mathbb{R}^n \to \mathcal{S}^m$ if such a function is *linear*, and the symbol $\mathcal{B} : \mathbb{R}^n \to \mathcal{S}^m$ if such a function is *nonlinear*.

Note that any linear function $\mathcal{A} : \mathbb{R}^n \to \mathcal{S}^m$ can be expressed in the form

$$\mathcal{A}(x) = \sum_{i=1}^{n} x_i A^{(i)} \quad \text{for all} \quad x \in \mathbb{R}^n, \tag{1}$$

with symmetric matrices $A^{(i)} \in \mathcal{S}^m$, $i = 1, 2, \ldots, n$. Based on the representation (1) we introduce the norm

$$\|\mathcal{A}\| := \left(\sum_{i=1}^{n} \left\| A^{(i)} \right\|^2 \right)^{\frac{1}{2}} \tag{2}$$

of $\mathcal{A}$. The adjoint operator $\mathcal{A}^* : \mathcal{S}^m \to \mathbb{R}^n$ with respect to the trace inner product is defined by

$$\langle \mathcal{A}(x), Y \rangle = \langle x, \mathcal{A}^*(Y) \rangle = x^T \mathcal{A}^*(Y) \quad \text{for all} \quad x \in \mathbb{R}^n \quad \text{and} \quad Y \in \mathcal{S}^m.$$

It turns out that

$$\mathcal{A}^*(Y) = \begin{bmatrix} A^{(1)} \bullet Y \\ \vdots \\ A^{(n)} \bullet Y \end{bmatrix} \quad \text{for all} \quad Y \in \mathcal{S}^m. \tag{3}$$

We assume that nonlinear functions $\mathcal{B} : \mathbb{R}^n \to \mathcal{S}^m$ are at least $\mathcal{C}^2$-differentiable. We denote by

$$B^{(i)}(x) := \frac{\partial}{\partial x_i}\mathcal{B}(x) \quad \text{and} \quad B^{(i,j)}(x) := \frac{\partial^2}{\partial x_i \partial x_j}\mathcal{B}(x), \quad i,j = 1,2,\ldots,n,$$

the first and second partial derivatives of $\mathcal{B}$, respectively. For each $x \in \mathbb{R}^n$, the derivative $D_x\mathcal{B}$ at x induces a linear function $D_x\mathcal{B}(x) : \mathbb{R}^n \to \mathcal{S}^m$, which is given by

$$D_x\mathcal{B}(x)[\Delta x] := \sum_{i=1}^{n} (\Delta x)_i B^{(i)}(x) \in \mathcal{S}^m \quad \text{for all} \quad \Delta x \in \mathbb{R}^n.$$

In particular,

$$\mathcal{B}(x + \Delta x) \approx \mathcal{B}(x) + D_x\mathcal{B}(x)[\Delta x], \quad \Delta x \in \mathbb{R}^n,$$

is the linearization of $\mathcal{B}$ at the point x. For any linear function $\mathcal{A} : \mathbb{R}^n \to \mathcal{S}^m$, we have

$$D_x\mathcal{A}(x)[\Delta x] = \mathcal{A}(\Delta x) \quad \text{for all} \quad x, \Delta x \in \mathbb{R}^n. \tag{4}$$

We always use the expression on the right-hand side of (4) to describe derivatives of linear functions.

We remark that for any fixed matrix $Y \in \mathcal{S}^m$, the map $x \mapsto \mathcal{B}(x) \bullet Y$ is a scalar-valued function of $x \in \mathbb{R}^n$. Its gradient at x is given by

$$\nabla_x \left(\mathcal{B}(x) \bullet Y\right) = \left(D_x \left(\mathcal{B}(x) \bullet Y\right)\right)^T = \begin{bmatrix} B^{(1)}(x) \bullet Y \\ \vdots \\ B^{(n)}(x) \bullet Y \end{bmatrix} \in \mathbb{R}^n \tag{5}$$

and its Hessian by

$$\nabla_x^2 \left(\mathcal{B}(x) \bullet Y\right) = \begin{bmatrix} B^{(1,1)}(x) \bullet Y & \cdots & B^{(1,n)}(x) \bullet Y \\ \vdots & & \vdots \\ B^{(n,1)}(x) \bullet Y & \cdots & B^{(n,n)}(x) \bullet Y \end{bmatrix} \in \mathcal{S}^n.$$

In particular, for any linear function $\mathcal{A} : \mathbb{R}^n \to \mathcal{S}^m$, in view of (1), (3), and (5), we have

$$\nabla_x \left(\mathcal{A}(x) \bullet Y\right) = \mathcal{A}^*(Y). \tag{6}$$

3 An application in passive reduced-order modeling

We remark that applications of linear SDPs include relaxations of combinatorial optimization problems and problems related to Lyapunov functions or the positive real lemma in control theory; we refer the reader to [1,3,8,14,36,38] and the references given there. In this section, we describe an application in passive reduced-order modeling that leads to a class of NLSDPs.

Roughly speaking, a system is called passive if it does not generate energy. For the special case of time-invariant linear dynamical systems, passivity is equivalent to positive realness of the frequency-domain transfer function associated with the system. More precisely, consider *transfer functions* of the form

$$Z_n(s) = B_2^T \big(G + sC\big)^{-1} B_1, \quad s \in \mathbb{C}, \tag{7}$$

where $G, C \in \mathbb{R}^{n\times n}$ and $B_1,\ B_2 \in \mathbb{R}^{n\times m}$ are given data matrices. The integer n is the *state-space dimension* of the time-invariant linear dynamical system, and m is the number of inputs and outputs of the system. In (7), the matrix pencil $G+sC$ is assumed to be *regular*, i.e., the matrix $G+sC$ is singular for only finitely many values of $s \in \mathbb{C}$. Note that Z_n is an $m \times m$-matrix-valued rational function of the complex variable $s \in \mathbb{C}$.

In *reduced-order modeling*, one is given a large-scale time-invariant linear dynamical system of state-space dimension N, and the problem is to construct a 'good' approximation of that system of state-space dimension $n \ll N$; see, e.g., [13] and the references given there. If the large-scale system is passive, then for certain applications, it is crucial that the reduced-order model of state-space dimension n preserves the passivity of the original system. Unfortunately, some of the most efficient reduced-order modeling techniques do not preserve passivity. However, the reduced-order models are often 'almost' passive, and passivity of the models can be enforced by perturbing the data matrices of the models. Next, we describe how the problem of constructing such perturbations leads to a class of NLSDPs.

An $m \times m$-matrix-valued rational function Z is called *positive real* if the following three conditions are satisfied:

(i) Z is analytic in $\mathbb{C}_+ := \{\, s \in \mathbb{C} \mid \mathrm{Re}(s) > 0 \,\}$;
(ii) $Z(\overline{s}) = \overline{Z(s)}$ for all $s \in \mathbb{C}$;
(iii) $Z(s) + \big(Z(\overline{s})\big)^T \succeq 0$ for all $s \in \mathbb{C}_+$.

For functions Z_n of the form (7) positive realness (and thus passivity of the system associated with Z_n) can be characterized via linear SDPs; see, e.g., [3,14] and the references given there. More precisely, if the linear SDP

$$\begin{aligned} P^T G + G^T P &\succeq 0, \\ P^T C = C^T P &\succeq 0, \\ P^T B_1 &= B_2, \end{aligned} \tag{8}$$

has a solution $P \in \mathbb{R}^{n\times n}$, then the transfer function (7), Z_n, is positive real. Conversely, under certain additional assumptions (see [14]), positive realness of Z_n implies the solvability of the linear SDP (8).

Now assume that Z_n in (7) is the transfer function of a non-passive reduced-oder model of a passive large-scale system. Our goal is to perturb some of the data matrices in (7) so that the perturbed transfer function is positive real. For the special case $m = 1$, such an approach is discussed in [5]. In this case, there is a simple eigenvalue-based characterization [4] of positive realness. However, this characterization cannot be extended to the general case $m \geq 1$. Another special case, which leads to linear SDPs, is described in [9].

In the general case $m \geq 1$, we employ perturbations X_G and X_C of the matrices G and C in (7). The resulting perturbed transfer function is then of the form

$$\tilde{Z}_n(s) = B_2^T \big(G + X_G + s(C + X_C)\big)^{-1} B_1, \tag{9}$$

and the problem is to construct the perturbations X_G and X_C such that $\tilde{Z}_n$ is positive real. Applying the characterization (8) of positive realness to (9), we obtain the following nonconvex NLSDP:

$$\begin{aligned} P^T(G + X_G) + (G + X_G)^T P &\succeq 0, \\ P^T(C + X_C) = (C + X_C)^T P &\succeq 0, \\ P^T B_1 &= B_2. \end{aligned} \tag{10}$$

Here, the unknowns are the matrices $P, X_G, X_C \in \mathbb{R}^{n \times n}$. If (10) has a solution P, X_G, X_C, then choosing the matrices X_G and X_C as the perturbations in (9) guarantees passivity of the reduced-order model given by the transfer function $\tilde{Z}_n$.

4 Linear semidefinite programs

In this section, we briefly review the case of linear semidefinite programs.

Given a linear function $\mathcal{A} : \mathbb{R}^n \to \mathcal{S}^m$, a vector $b \in \mathbb{R}^n$, and a matrix $C \in \mathcal{S}^m$, a pair of primal and dual linear semidefinite programs is as follows:

$$\text{maximize} \quad C \bullet Y \quad \text{subject to} \quad \begin{aligned} &Y \in \mathcal{S}^m,\ Y \succeq 0, \\ &\mathcal{A}^*(Y) + b = 0, \end{aligned} \tag{11}$$

and

$$\text{minimize} \quad b^T x \quad \text{subject to} \quad \begin{aligned} &x \in \mathbb{R}^n, \\ &\mathcal{A}(x) + C \preceq 0. \end{aligned} \tag{12}$$

We remark that this formulation is a slight variation of the standard pair of primal-dual programs. We chose the above version in order to facilitate the generalization of problems of the form (12) to nonlinear semidefinite programs in standard form.

If there exists a matrix $Y \succ 0$ that is feasible for (11), then we call Y strictly feasible for (11) and say that (11) satisfies Slater's condition. Likewise, if there exists a vector x such that $\mathcal{A}(x) + C \prec 0$, then we call (12) strictly feasible and say that (12) satisfies Slater's condition.

The following optimality conditions for linear semidefinite programs are well known; see, e.g., [33]. If problem (11) or (12) satisfies Slater's condition, then

the optimal values of (11) and (12) coincide. Furthermore, if in addition both problems are feasible, then optimal solutions $Y^{\rm opt}$ and $x^{\rm opt}$ of both problems exist and $Y := Y^{\rm opt}$ and $x := x^{\rm opt}$ satisfy the complementarity condition

$$YS = 0, \quad \text{where} \quad S := -C - \mathcal{A}(x). \tag{13}$$

Conversely, if Y and x are feasible points for (11) and (12), respectively, and satisfy the complementarity condition (13), then $Y^{\rm opt} := Y$ is an optimal solution of (11) and $x^{\rm opt} := x$ is an optimal solution of (12).

These optimality conditions can be summarized as follows. If problem (12) satisfies Slater's condition, then for a point $x \in \mathbb{R}^n$ to be an optimal solution of (12) it is necessary and sufficient that there exist matrices $Y \succeq 0$ and $S \succeq 0$ such that

$$\begin{aligned} \mathcal{A}(x) + C + S &= 0, \\ \mathcal{A}^*(Y) + b &= 0, \\ YS &= 0. \end{aligned} \tag{14}$$

Note that, in view of (6), the second equation in (14) can also be written in the form

$$\nabla_x \big((\mathcal{A}(x) + C) \bullet Y \big) + b = \mathcal{A}^*(Y) + b = 0. \tag{15}$$

Furthermore, the last equation in (14) is equivalent to its symmetric form, $YS + SY = 0$; see, e.g., [2]. In the case of strict complementarity, the derivatives of $YS = 0$ and $YS + SY = 0$ are also equivalent. For later use, we state these facts in the following lemma.

Lemma 1 *Let $Y, S \in \mathcal{S}^m$.*

a) *If $Y \succeq 0$ or $S \succeq 0$, then*

$$YS = 0 \quad \Longleftrightarrow \quad YS + SY = 0. \tag{16}$$

b) *If Y and S are strictly complementary, i.e., $Y, S \succeq 0$, $YS = 0$, and $Y + S \succ 0$, then for any $\dot{Y}, \dot{S} \in \mathcal{S}^m$,*

$$Y\dot{S} + \dot{Y}S = 0 \quad \Longleftrightarrow \quad Y\dot{S} + \dot{Y}S + \dot{S}Y + S\dot{Y} = 0. \tag{17}$$

Moreover, Y, S have representations of the form

$$Y = U \begin{bmatrix} Y_1 & 0 \\ 0 & 0 \end{bmatrix} U^T, \quad S = U \begin{bmatrix} 0 & 0 \\ 0 & S_2 \end{bmatrix} U^T, \tag{18}$$

where U is an $m \times m$ orthogonal matrix, $Y_1 \succ 0$ is a $k \times k$ diagonal matrix, and $S_2 \succ 0$ is an $(m-k) \times (m-k)$ diagonal matrix, and any matrices $\dot{Y}, \dot{S} \in \mathcal{S}^m$ satisfying (17) *are of the form*

$$\dot{Y} = U \begin{bmatrix} \dot{Y}_1 & \dot{Y}_3 \\ \dot{Y}_3^T & 0 \end{bmatrix} U^T, \quad \dot{S} = U \begin{bmatrix} 0 & \dot{S}_3 \\ \dot{S}_3^T & \dot{S}_2 \end{bmatrix} U^T, \quad \text{where} \quad Y_1 \dot{S}_3 + \dot{Y}_3 S_2 = 0. \tag{19}$$

Proof The equivalence (16) is well known; see, e.g., [2, Page 749].

We now turn to the proof of part b). The strict complementarity of Y and S readily implies that Y and S have representations of the form (18); see, e.g., [20, Page 62]. Then, any matrices $\dot{Y}, \dot{S} \in \mathcal{S}^m$ can be written in the form

$$\dot{Y} = U \begin{bmatrix} \dot{Y}_1 & \dot{Y}_3 \\ \dot{Y}_3^T & \dot{Y}_2 \end{bmatrix} U^T, \quad \dot{S} = U \begin{bmatrix} \dot{S}_1 & \dot{S}_3 \\ \dot{S}_3^T & \dot{S}_2 \end{bmatrix} U^T, \tag{20}$$

where U is the matrix from (18) and the block sizes in (20) are the same as in (18).

Using (18), it follows that the matrices (20) satisfy the equation on the left-hand side of (17) if, and only if,

$$Y_1 \dot{S}_1 = 0, \quad \dot{Y}_2 S_2 = 0, \quad Y_1 \dot{S}_3 + \dot{Y}_3 S_2 = 0.$$

Since Y_1 and S_2 are in particular nonsingular, these relations are equivalent to

$$\dot{S}_1 = 0, \quad \dot{Y}_2 = 0, \quad Y_1 \dot{S}_3 + \dot{Y}_3 S_2 = 0. \tag{21}$$

Similarly, using (18), it follows that the matrices (20) satisfy the equation on the right-hand side of (17) if, and only if,

$$Y_1 \dot{S}_1 + \dot{S}_1 Y_1 = 0, \quad \dot{Y}_2 S_2 + S_2 \dot{Y}_2 = 0, \quad Y_1 \dot{S}_3 + \dot{Y}_3 S_2 = 0.$$

Since $Y_1 \succ 0$ and $S_2 \succ 0$, these relations are also equivalent to (21). Thus, we have established the equivalence of the two equations in (17). Furthermore, in view of (20) and (21), any solutions of (17) are indeed of the form (19), and the proof is complete. □

5 Nonlinear semidefinite programs

In this section, we consider nonlinear semidefinite programs, which are extensions of the dual linear semidefinite programs (12).

Given a vector $b \in \mathbb{R}^n$ and a matrix-valued function $\mathcal{B} : \mathbb{R}^n \to \mathcal{S}^m$, we consider problems of the following form:

$$\begin{aligned} \text{minimize} \quad & b^T x \quad \text{subject to} \quad x \in \mathbb{R}^n, \\ & \qquad\qquad\quad \mathcal{B}(x) \preceq 0. \end{aligned} \tag{22}$$

Here, the function $\mathcal{B}$ is nonlinear in general, and thus (22) represents a class of nonlinear semidefinite programs. We assume that the function $\mathcal{B}$ is at least $\mathcal{C}^2$-differentiable.

For simplicity of presentation, we have chosen a simple form of problem (22). We stress that problem (22) may also include additional nonlinear equality and inequality constraints. The corresponding modifications are detailed at the end of this paper. Furthermore, the choice of the linear objective function $b^T x$ in (22) was made only to simplify notation. A nonlinear objective function can always be transformed into a linear one by adding one artificial variable and one more constraint. In particular, all statements about (22) in this paper can be modified so that they apply to additional nonlinear equality and inequality constraints and to nonlinear objective functions.

Note that the class (22) reduces to linear semidefinite programs of the form (12) if $\mathcal{B}$ is an affine function.

The Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathcal{S}^m \to \mathbb{R}$ of (22) is defined as follows:

$$\mathcal{L}(x,Y) := b^T x + \mathcal{B}(x) \bullet Y. \tag{23}$$

Its gradient with respect to x is given by

$$g(x,Y) := \nabla_x \mathcal{L}(x,Y) = b + \nabla_x \left(\mathcal{B}(x) \bullet Y\right) \tag{24}$$

and its Hessian by

$$H(x,Y) := \nabla_x^2 \mathcal{L}(x,Y) = \nabla_x^2 \left(\mathcal{B}(x) \bullet Y\right). \tag{25}$$

If the problem (22) is convex and satisfies Slater's condition [21], then for each optimal solution x of (22) there exists an $m \times m$ matrix $Y \succeq 0$ such that the pair (x,Y) is a saddle point of the Lagrangian (23), $\mathcal{L}$.

More generally, for nonconvex problems (22), let $x \in \mathbb{R}^n$ be a feasible point of (22), and assume that the Robinson or Mangasarian-Fromovitz constraint qualification [21,30,31] is satisfied at x, i.e., there exists a vector $\Delta x \neq 0$ such that $\mathcal{B}(x) + D_x \mathcal{B}(x)[\Delta x] \prec 0$. Then, if x is a local minimizer of (22), the first-order optimality condition is satisfied, i.e., there exist matrices $Y, S \in \mathcal{S}^m$ such that

$$\begin{aligned} \mathcal{B}(x) + S &= 0, \\ g(x,Y) &= 0, \\ YS &= 0, \\ Y, S &\succeq 0. \end{aligned} \tag{26}$$

The system (26) is a straightforward generalization of the optimality conditions (14) and (15), with the affine function $\mathcal{A}(x) + C$ in (15) replaced by the nonlinear function $\mathcal{B}(x)$.

Primal-dual interior-point methods for solving (22) roughly proceed as follows. For some sequence of duality parameters $\mu_k > 0$, $\mu_k \to 0$, the solutions of the perturbed primal-dual system,

$$\begin{aligned} \mathcal{B}(x) + S &= 0, \\ g(x,Y) &= 0, \\ YS &= \mu_k I_m, \\ Y, S &\succ 0, \end{aligned} \tag{27}$$

are approximated by some variant of Newton's method. Since, in general, Newton's method does not preserve inequalities, the parameters $\mu_k > 0$ are used to maintain strict feasibility, i.e., $Y, S \succ 0$ for all iterates.

The solutions of (27) coincide with the solutions of the standard logarithmic-barrier problems for (22). Moreover, the logarithmic-barrier approach for solving (22) can be interpreted as a certain choice of the 'symmetrization operator' for the equation, $YS = \mu_k I_m$, in the third row of (27); see Section 6 below. With this choice, the barrier function yields a very natural criterion for the step-size control in trust-region algorithms. The authors have implemented various versions of predictor-corrector trust-region barrier methods for solving (22). For a number

of examples, the running times of the resulting algorithms were comparable to the behavior of interior-point methods for convex programs. However, the authors also encountered several instances in which the number of iterations for these methods was very high compared to the typical number of iterations needed for solving linear SDPs. For such negative examples it may be more efficient to solve a sequence of linear SDPs in order to obtain an approximate solution of (22). This observation motivated the SSP method described in the next section.

6 An SSP method for nonlinear semidefinite programs

In this section, we introduce the sequential semidefinite programming (SSP) method, which is a generalization of the SQP method for standard nonlinear programs to nonlinear semidefinite programs of the form (22). For an overview of SQP methods for standard nonlinear programs, we refer the reader to [6] and the references given there.

In analogy to the SQP method, at each iteration of the SSP method one solves a subproblem that is slightly more difficult than the linearization of (26) at the current iterate. More precisely, let (x^k, Y^k) denote the current point at the beginning of the k-th iteration. One then determines corrections $(\Delta x, \Delta Y)$ and a matrix S such that

$$\begin{aligned} \mathcal{B}(x^k) + D_x\mathcal{B}(x^k)[\Delta x] + S &= 0, \\ b + H^k\Delta x + \nabla_x\big(\mathcal{B}(x^k) \bullet (Y^k + \Delta Y)\big) &= 0, \\ (Y^k + \Delta Y)S &= 0, \\ Y^k + \Delta Y,\, S &\succeq 0. \end{aligned} \tag{28}$$

Here and in the sequel, we use the notation

$$H^k := H\big(x^k, Y^k\big). \tag{29}$$

Recall from (24) and (25) that $g(x,Y)$ and $H(x,Y)$ denote the gradient and Hessian with respect to x, respectively, of the Lagrangian (23), $\mathcal{L}(x,Y)$, of the nonlinear semidefinite program (22). Moreover, from (24) it follows that the linearization of $g(x,Y)$ at the point (x^k, Y^k) is given by

$$g\big(x^k + \Delta x, Y^k + \Delta Y\big) \approx b + H^k\Delta x + \nabla_x\big(\mathcal{B}(x^k) \bullet (Y^k + \Delta Y)\big).$$

Thus, the second equation in (28) is just the linearization of the second equation in (26). Furthermore, the first equation of (28) is a straightforward linearization of the first equation in (26). This linearization is used in the same way in primal-dual interior-point methods.

The last two rows in (28) and (26) are identical when Y in (26) is rewritten as $Y = Y^k + \Delta Y$. In analogy to SQP methods for standard nonlinear programs, the problem of how to guarantee the nonnegativity constraints, namely $\mathcal{B}(x) \preceq 0$, is thus shifted to the subproblem (28). If the iterates x^k generated from (28) converge, then their limit x automatically satisfies $\mathcal{B}(x) \preceq 0$.

In contrast, interior-point methods use perturbations, symmetrizations, and linearizations for the last two rows in (26), resulting in cheaper linear subproblems

that are typically less 'powerful' than the subproblems (28). When using (28), both the problem of choosing a suitable symmetrization scheme and the problem of how to guarantee the nonnegativity constraints are shifted to the subproblem (28).

Note that the conditions (28) are the optimality conditions for the problem

$$\begin{aligned} \text{minimize} \quad & b^T \Delta x + \frac{1}{2}(\Delta x)^T H^k \Delta x \quad \text{subject to} \quad \Delta x \in \mathbb{R}^n, \\ & \mathcal{B}(x^k) + D_x \mathcal{B}(x^k)[\Delta x] \preceq 0. \end{aligned} \tag{30}$$

The conditions (28) and (30) have been considered in [31, Equations (2.1) and (2.2)], with the remark that they have "been found to be an appropriate approximation of" (22) "for numerical purposes".

In order to be able to solve the subproblem (28) efficiently, in practice, one replaces the matrix H^k in (28), respectively (30), by a positive semidefinite approximation $\hat{H}^k$ of H^k. As in the case of standard SQP methods, a BFGS update for the Hessian of the Lagrangian (23), $\mathcal{L}$, can be used to approximate H^k by some positive semidefinite matrix $\hat{H}^k$. Given an estimate $\hat{H}^{k-1}$ of H^{k-1} for the previous, $(k-1)$-th, SSP iteration, the quasi-Newton condition to generate a BFGS update $\hat{H}^k$ approximating the matrix H^k for the current, k-th, SSP iteration can be derived as follows:

$$\begin{aligned} \hat{H}^k \Delta x &= \nabla_x \big(\mathcal{B}(x^k) \bullet Y^k\big) - \nabla_x \big(\mathcal{B}(x^{k-1}) \bullet Y^k\big) \\ &= \nabla_x \mathcal{L}\big(x^k, Y^k\big) - \nabla_x \mathcal{L}\big(x^{k-1}, Y^k\big) \\ &\approx \nabla_x^2 \mathcal{L}\big(x^k, Y^k\big)\big(x^k - x^{k-1}\big). \end{aligned} \tag{31}$$

If $\hat{H}^{k-1}$ is positive semidefinite, the BFGS update with the above condition can be suitably damped such that $\hat{H}^k$ is also positive semidefinite; see, e.g., [29]. At each iteration of the SSP method, problem (30) is solved with H^k replaced by the matrix $\hat{H}^k$ that is obtained by the BFGS update of $\hat{H}^{k-1}$ from the previous SSP iteration. If $\hat{H}^k$ is positive semidefinite, problem (30) essentially reduces to a linear SDP, since the convex quadratic term in the objective function can be written as a semidefiniteness constraint or a second-order-cone constraint. While the formulation as a second-order-cone constraint is more efficient, and for example, can be specified as input for the program package SeDuMi [34] in order to solve (30), it was pointed out by [26] that it may be most efficient to use a program that is designed for SDPs with linear constraints and a convex quadratic objective function.

It seems that many results for standard SQP methods carry over in a straightforward fashion to the SSP method. For example, the SSP method can be augmented by a penalty term in case that the subproblems (30) become infeasible. In this case, the right-hand sides, "0", of the first three rows in (28) are replaced by weaker, penalized right-hand sides. Moreover, the convergence analysis of the method proposed in [10, 12] yields results that are comparable to the ones for standard SQP methods.

The standard analysis of quadratic convergence of SQP methods for nonlinear programs that satisfy strict complementarity conditions proceeds by first showing that the active constraints will be identified correctly in the final stages of the

algorithm and then using the equivalence of the SQP iteration and the Newton iteration for the simplified KKT-system in which only the active constraints are used.

For nonlinear semidefinite programs the situation is slightly more complicated since it is difficult to identify active constraints. The paper [12] presents a proof that is based on a new approach by Bonnans et al. [7] and uses some general results due to Robinson [31]. This approach does not require strict complementarity, and it allows the use of approximate Hessian matrices in (31).

In the next two sections, we present a more elementary and self-contained approach to analyze convergence of the SSP method under a strict complementarity condition.

7 Sensitivity results

In this section, we establish sensitivity results, first for the special case of quadratic semidefinite programs and then for general nonlinear semidefinite programs of the form (22). More precisely, we show that strictly complementary solutions of such problems depend smoothly on the problem data.

We start with quadratic semidefinite programs of the form

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad \begin{array}{l} x \in \mathbb{R}^n, \\ \mathcal{A}(x) + C \preceq 0. \end{array} \tag{32}$$

Here, $\mathcal{A} : \mathbb{R}^n \to \mathcal{S}^m$ is a linear function, $C \in \mathcal{S}^m$, and $f : \mathbb{R}^n \to \mathbb{R}$ is a quadratic function defined by $f(x) = b^T x + \frac{1}{2} x^T H x$, where $b \in \mathbb{R}^n$ and $H \in \mathcal{S}^n$. Note that we make no further assumptions on the matrix H. Thus, problem (32) is a general, possibly nonconvex, quadratic semidefinite program.

The problem (32) is described by the data

$$\mathcal{D} := [\mathcal{A}, b, C, H]. \tag{33}$$

In Theorem 1 below, we present a sensitivity result for the solutions (32) when the data $\mathcal{D}$ is changed to $\mathcal{D} + \Delta\mathcal{D}$ where

$$\Delta\mathcal{D} := [\Delta\mathcal{A}, \Delta b, \Delta C, \Delta H] \tag{34}$$

is a sufficiently small perturbation. We use the norm

$$\|\mathcal{D}\| := \big(\|\mathcal{A}\|^2 + \|b\|^2 + \|C\|^2 + \|H\|^2\big)^{\frac{1}{2}}$$

for data (33) and perturbations (34). Recall that $\|\mathcal{A}\|$ is defined in (2).

We denote by

$$\mathcal{L}^{(\mathrm{q})}(x, Y) := f(x) + (\mathcal{A}(x) + C) \bullet Y$$

the Lagrangian of problem (32). Note that $\nabla_x f(x) = b + Hx$. Together with (6), it follows that

$$\nabla_x \mathcal{L}^{(\mathrm{q})}(x, Y) = b + Hx + \mathcal{A}^*(Y) \quad \text{and} \quad \nabla_x^2 \mathcal{L}^{(\mathrm{q})}(x, Y) = H.$$

Recall that problem (32) is said to satisfy Slater's condition if there exists a vector $x \in \mathbb{R}^n$ with $\mathcal{A}(x) + C \prec 0$. Moreover, the triple $(\bar{x}, \bar{Y}, \bar{S})$, where $\bar{x} \in \mathbb{R}^n$ and $\bar{Y}$, $\bar{S} \in \mathcal{S}^m$, is called a *stationary point* of (32) if

$$\begin{aligned} \mathcal{A}(\bar{x}) + C + \bar{S} &= 0, \\ b + H\bar{x} + \mathcal{A}^*(\bar{Y}) &= 0, \\ \bar{Y}\bar{S} + \bar{S}\bar{Y} &= 0, \\ \bar{Y}, \bar{S} &\succeq 0. \end{aligned} \tag{35}$$

Here, we have used equivalence (16) of Lemma 1 and replaced $\bar{Y}\bar{S} = 0$ by its symmetric version, which is stated as the third equation of (35). If in addition to (35), one has

$$\bar{Y} + \bar{S} \succ 0, \tag{36}$$

then $(\bar{x}, \bar{Y}, \bar{S})$ is said to be a *strictly complementary* stationary point of (32).

Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of (32). We say that $h \in \mathbb{R}^n$, $h \neq 0$, is a *feasible direction* at $\bar{x}$ if $x = \bar{x} + \epsilon h$ is a feasible point of (32) for all sufficiently small $\epsilon > 0$. Following [31, Definition 2.1], we say that the *second-order sufficient condition* holds at $\bar{x}$ with modulus $\mu > 0$ if for all feasible directions $h \in \mathbb{R}^n$ at $\bar{x}$ with $h^T(b + H\bar{x}) = h^T \nabla_x f(\bar{x}) = 0$ one has

$$h^T H h = h^T \big(\nabla_x^2 \mathcal{L}^{(\mathrm{q})}(\bar{x}, \bar{Y})\big) h \geq \mu \|h\|^2. \tag{37}$$

After these preliminaries, our main result of this section can be stated as follows.

Theorem 1 *Assume that problem* (32) *satisfies Slater's condition. Let the point* $(\bar{x}, \bar{Y}, \bar{S})$ *be a locally unique and strictly complementary stationary point of* (32) *with data* (33)*,* $\mathcal{D}$*, and assume that the second-order sufficient condition holds at* $\bar{x}$ *with modulus* $\mu > 0$*. Then, for all sufficiently small perturbations* (34)*,* $\Delta\mathcal{D}$*, there exists a locally unique stationary point* $\big(\bar{x}(\mathcal{D} + \Delta\mathcal{D}), \bar{Y}(\mathcal{D} + \Delta\mathcal{D}), \bar{S}(\mathcal{D} + \Delta\mathcal{D})\big)$ *of the perturbed program* (32) *with data* $\mathcal{D} + \Delta\mathcal{D}$*. Moreover, the point* $\big(\bar{x}(\mathcal{D} + \Delta\mathcal{D}), \bar{Y}(\mathcal{D} + \Delta\mathcal{D}), \bar{S}(\mathcal{D} + \Delta\mathcal{D})\big)$ *is a differentiable function of the perturbation* (34)*, and for* $\Delta\mathcal{D} = 0$*, we have* $\big(\bar{x}(\mathcal{D}), \bar{Y}(\mathcal{D}), \bar{S}(\mathcal{D})\big) = (\bar{x}, \bar{Y}, \bar{S})$*. The derivative* $D_{\mathcal{D}}\big(\bar{x}(\mathcal{D}), \bar{Y}(\mathcal{D}), \bar{S}(\mathcal{D})\big)$ *of* $\big(\bar{x}(\mathcal{D}), \bar{Y}(\mathcal{D}), \bar{S}(\mathcal{D})\big)$ *with respect to* $\mathcal{D}$ *evaluated at* $(\bar{x}, \bar{Y}, \bar{S})$ *is characterized by the directional derivatives*

$$(\dot{x}, \dot{Y}, \dot{S}) := D_{\mathcal{D}}\big(\bar{x}(\mathcal{D}), \bar{Y}(\mathcal{D}), \bar{S}(\mathcal{D})\big)[\Delta\mathcal{D}]$$

for any $\Delta\mathcal{D}$*. Here* $(\dot{x}, \dot{Y}, \dot{S})$ *is the unique solution of the system of linear equations,*

$$\begin{aligned} \mathcal{A}(\dot{x}) + \dot{S} &= -\Delta C - \Delta\mathcal{A}(\bar{x}), \\ H\dot{x} + \mathcal{A}^*(\dot{Y}) &= -\Delta b - \Delta H\bar{x} - \Delta\mathcal{A}^*(\bar{Y}), \\ \bar{Y}\dot{S} + \dot{Y}\bar{S} + \dot{S}\bar{Y} + \bar{S}\dot{Y} &= 0, \end{aligned} \tag{38}$$

for the unknowns $\dot{x} \in \mathbb{R}^n$*,* $\dot{Y}$*,* $\dot{S} \in \mathcal{S}^m$*. Finally, the second-order sufficient condition holds at* $\bar{x}(\Delta\mathcal{D})$ *whenever* $\Delta\mathcal{D}$ *is sufficiently small.*

Remark 1 Theorem 1 is an extension of the sensitivity result for linear semidefinite programs presented in [15]. A related sensitivity result for linear semidefinite programs for a more restricted class of perturbations, but also under weaker assumptions, is given in [35]. A local Lipschitz continuity property of unique and strictly complementary solutions of linear semidefinite programs is derived in [23]. A weaker form of second-order necessary and sufficient conditions is given in Theorem 9 in [32]. Furthermore, in Theorem 11 of [32], the directional derivative $\dot{x}$ is expressed as the solution of a certain quadratic program.

Remark 2 While we did not explicitly state a linear independence constraint qualification, commonly referred to as LICQ, it is implied by our condition of uniqueness of the stationary point; see, e.g., [15]. Moreover, our assumptions on the stationary point $(\bar{x}, \bar{Y}, \bar{S})$ imply that $\bar{x}$ is a strict local minimizer of (32).

Remark 3 The first and third equations in (38) are symmetric $m \times m$ matrix equations, and so only their upper triangular parts have to be considered. Thus, the total number of scalar equations in (38) is $m^2 + m + n$. On the other hand, there are $m^2 + m + n$ unknowns, namely the entries of $\dot{x} \in \mathrm{Re}^n$ and of the upper triangular parts of $\dot{Y}, \dot{S} \in \mathcal{S}^m$. Hence, (38) is a square system.

Remark 4 In view of part b) of Lemma 1, the last equation of (38) is equivalent to

$$\bar{Y}\ \dot{S} + \dot{Y}\ \bar{S} = 0. \tag{39}$$

Thus, Theorem 1 can be stated equivalently with (39) in (38). However, the resulting system of equations (38) would then be overdetermined.

Proof (Theorem 1) The proof is divided into four steps.
Step 1. In this step, we establish the following result. If the perturbed program has a local solution that is a differentiable function of the perturbation, then the derivative is indeed a solution of (38).

Slater's condition is invariant under small perturbations of the problem data. Hence, if there exists a local solution $\bar{x} + \Delta x$, $\bar{S} + \Delta S$ of the perturbed problem near $\bar{x}$, $\bar{S}$, then the necessary first-order conditions of the perturbed problem apply at $\bar{x} + \Delta x$, $\bar{S} + \Delta S$, and state that there exists a matrix ΔY such that $\bar{Y} + \Delta Y \succeq 0$, $\bar{S} + \Delta S \succeq 0$, and

$$\begin{aligned} (\mathcal{A} + \Delta\mathcal{A})(\bar{x} + \Delta x) + C + \Delta C + \bar{S} + \Delta S &= 0, \\ b + \Delta b + (H + \Delta H)(\bar{x} + \Delta x) + (\mathcal{A}^* + \Delta\mathcal{A}^*)(\bar{Y} + \Delta Y) &= 0, \\ (\bar{Y} + \Delta Y)(\bar{S} + \Delta S) + (\bar{S} + \Delta S)(\bar{Y} + \Delta Y) &= 0. \end{aligned} \tag{40}$$

Subtracting from these equations the first three equations of (35) yields

$$\begin{aligned} (\mathcal{A} + \Delta\mathcal{A})(\Delta x) + \Delta S &= -\Delta C - \Delta\mathcal{A}(\bar{x}), \\ (H + \Delta H)\Delta x + (\mathcal{A}^* + \Delta\mathcal{A}^*)(\Delta Y) &= -\Delta b - \Delta H\bar{x} - \Delta\mathcal{A}^*(\bar{Y}), \\ \bar{Y}\,\Delta S + \Delta Y\,\bar{S} + \Delta S\,\bar{Y} + \bar{S}\,\Delta Y &= -\Delta Y\,\Delta S - \Delta S\,\Delta Y. \end{aligned} \tag{41}$$

Neglecting the second-order terms in (41), and using (39), we obtain the result claimed in (38). It still remains to verify the existence and differentiability of Δx, ΔY, ΔS.

Step 2. In this step, we prove that the system of linear equations (38) has a unique solution. To this end, we show that the homogeneous version of (38), i.e., the system

$$\begin{aligned} \mathcal{A}(\dot{x}) + \dot{S} &= 0, \\ H\dot{x} + \mathcal{A}^*(\dot{Y}) &= 0, \\ \bar{Y}\dot{S} + \dot{Y}\bar{S} + \dot{S}\bar{Y} + \bar{S}\dot{Y} &= 0, \end{aligned} \tag{42}$$

only has the trivial solution $\dot{x} = 0$, $\dot{Y} = \dot{S} = 0$.

Let $\dot{x} \in \mathbb{R}^n$, $\dot{Y}$, $\dot{S} \in \mathcal{S}^m$ be any solution of (42). Recall that, in view of part b) by Lemma 1 we may assume that $\bar{Y}$ and $\bar{S}$ are given in diagonal form:

$$\bar{Y} = \begin{bmatrix} \bar{Y}_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \bar{S} = \begin{bmatrix} 0 & 0 \\ 0 & \bar{S}_2 \end{bmatrix}, \tag{43}$$

where $\bar{Y}_1, \bar{S}_2 \succ 0$ and $\bar{Y}_1, \bar{S}_2$ are diagonal.

Indeed, this can be done by replacing, in (42), $\bar{Y}$, $\bar{S}$, $\dot{Y}$, $\dot{S}$, $\mathcal{A}(x)$ by $U^T\bar{Y}U$, $U^T\bar{S}U$, $U^T\dot{Y}U$, $U^T\dot{S}U$, $U^T\mathcal{A}(x)U$, respectively, where U is the matrix in (18), and then multiplying the first and third rows from the left by U and from the right by U^T. Furthermore, in view of part b) by Lemma 1, any matrices $\dot{Y}\,\dot{S} \in \mathcal{S}^m$ satisfying the last equation of (42) are then of the form

$$\dot{Y} = \begin{bmatrix} \dot{Y}_1 & \dot{Y}_3 \\ \dot{Y}_3^T & 0 \end{bmatrix}, \quad \dot{S} = \begin{bmatrix} 0 & \dot{S}_3 \\ \dot{S}_3^T & \dot{S}_2 \end{bmatrix}, \quad \text{where} \quad \dot{Y}_3\bar{S}_2 + \bar{Y}_1\dot{S}_3 = 0. \tag{44}$$

Next, we establish the inequality

$$\dot{x}^T H \dot{x} \geq \mu \|\dot{x}\|^2, \tag{45}$$

where $\mu > 0$ is the modulus of the second-order sufficient condition (37). Assume that $\dot{x} \neq 0$. Let $\breve{x} \in \mathbb{R}^n$ be a Slater point for problem (32). This guarantees that

$$M = \begin{bmatrix} M_1 & M_3 \\ M_3^T & M_2 \end{bmatrix} := -\big(\mathcal{A}(\breve{x}) + C\big) \succ 0, \tag{46}$$

where the block partitioning M is conforming with (44). For $\eta > 0$, set

$$h_\eta^+ := \dot{x} + \eta(\breve{x} - \bar{x}) \quad \text{and} \quad h_\eta^- := -\dot{x} + \eta(\breve{x} - \bar{x}). \tag{47}$$

Since $\dot{x} \neq 0$, there exists an $\eta_0 > 0$ such that $h_\eta^\pm \neq 0$ for all $0 < \eta \leq \eta_0$. Next, we prove that for all such η, both vectors h_η^+ and h_η^- are feasible directions for (32) at $\bar{x}$. Let $0 < \eta \leq \eta_0$ be arbitrary, but fixed. We then need to verify that $\mathcal{A}(\bar{x} + \epsilon h_\eta^\pm) + C \preceq 0$ for all sufficiently small $\epsilon > 0$. Recall that $\mathcal{A}$ is a linear function. Using (47), (43), (44), (46), the first equation of (35), and the first equation of (42), one readily verifies that

$$\begin{aligned} \mathcal{A}(\bar{x} + \epsilon h_\eta^\pm) + C &= (1 - \epsilon\eta)\big(\mathcal{A}(\bar{x}) + C\big) + \epsilon\eta\big(\mathcal{A}(\breve{x}) + C\big) \pm \epsilon\mathcal{A}(\dot{x}) \\ &= -\left(\begin{bmatrix} 0 & 0 \\ 0 & \bar{S}_2 \end{bmatrix} + \epsilon \begin{bmatrix} \eta M_1 & \eta M_3 \pm \dot{S}_3 \\ (\eta M_3 \pm \dot{S}_3)^T & \eta(M_2 - \bar{S}_2) \pm \dot{S}_2 \end{bmatrix} \right). \end{aligned} \tag{48}$$

Recall that $\eta > 0$ is fixed. Since, by (43) and (46), $\bar{S}_2 \succ 0$ and $M_1 \succ 0$, a standard Schur-complement argument shows that the matrix on the right-hand side of (48) is negative definite for all sufficiently small $\epsilon > 0$. Thus, the vectors (47) are feasible directions for (32) at $\bar{x}$ for any $\eta > 0$. This in turn implies

$$\dot{x}^T (b + H\bar{x}) = \dot{x}^T \nabla_x f(\bar{x}) = 0. \tag{49}$$

Indeed, suppose that $\dot{x}^T \nabla_x f(\bar{x}) < 0$. Then, for sufficiently small $\eta > 0$, the feasible direction h_η^+ also satisfies $\left(h_\eta^+\right)^T \nabla_x f(\bar{x}) < 0$, and thus h_η^+ is a descent direction for the objective function f of (32) at the point $\bar{x}$. This contradicts the local optimality of $\bar{x}$. Likewise, if $\dot{x}^T \nabla_x f(\bar{x}) > 0$, then, for sufficiently small $\eta > 0$, h_η^- is a descent direction for the objective function f of (32) at the point $\bar{x}$, leading to the same contradiction. The second-order sufficient condition (37) also holds true on the closure of the feasible directions. Since $\dot{x}$ is the limit of the feasible directions (47) for $\eta \to 0$ and $\dot{x}$ satisfies (49), the inequality (45) follows from (37).

Next recall from (43) that $\bar{Y}_1$ and $\bar{S}_2$ are positive definite diagonal matrices. The last relation in (44) thus implies that corresponding entries of the matrices $\dot{Y}_3$ and $\dot{S}_3$ are either zero or of opposite sign. It follows that $\langle \dot{Y}_3, \dot{S}_3 \rangle \leq 0$, and equality holds if, and only if, $\dot{Y}_3 = \dot{S}_3 = 0$. Using this inequality, together with the first two relations in (44), the first two equations of (42), and (45), one readily verifies that

$$\begin{aligned} 0 \geq 2\,\langle \dot{Y}_3, \dot{S}_3 \rangle &= \langle \dot{Y}, \dot{S} \rangle = -\langle \dot{Y}, \mathcal{A}(\dot{x}) \rangle \\ &= -\langle \mathcal{A}^*(\dot{Y}), \dot{x} \rangle = \langle H\dot{x}, \dot{x} \rangle = \dot{x}^T H \dot{x} \geq \mu \|\dot{x}\|^2. \end{aligned}$$

Since $\mu > 0$, this implies

$$\dot{x} = 0 \quad \text{and} \quad \dot{Y}_3 = \dot{S}_3 = 0. \tag{50}$$

By the first row of (42), it further follows that

$$\dot{S} = -\mathcal{A}(\dot{x}) = -\mathcal{A}(0) = 0. \tag{51}$$

Thus, it only remains to show that $\dot{Y} = 0$. In view of (44) and (51), we have

$$\dot{Y} = \begin{bmatrix} \dot{Y}_1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{52}$$

Now suppose that $\dot{Y}_1 \neq 0$. Then, by (43) and (52), we have

$$\bar{Y}_\epsilon := \bar{Y} + \epsilon \dot{Y} \succeq 0 \quad \text{and} \quad \bar{Y}_\epsilon \neq \bar{Y}$$

for all sufficiently small $|\epsilon|$. Moreover, using (35), (42), and (51), one readily verifies that the point $(\bar{x}, \bar{Y}_\epsilon, \bar{S})$ also satisfies (35) for all sufficiently small $|\epsilon|$. This contradicts the assumption that $(\bar{x}, \bar{Y}, \bar{S})$ is a locally unique stationary point. Hence $\dot{Y}_1 = 0$ and, by (52), $\dot{Y} = 0$.

This concludes the proof that the square system (42) is nonsingular.

Step 3. In this step, we show that the nonlinear system (35) has a local solution that depends smoothly on the perturbation $\Delta\mathcal{D}$. To this end, we apply the implicit function theorem to the system

$$\mathcal{A}(x) + C + S = 0, \quad Hx + b + \mathcal{A}^*(Y) = 0, \quad YS + SY = 0. \tag{53}$$

As we have just seen, the linearization of (53) at the point $(\bar{x}, \bar{Y}, \bar{S})$ is nonsingular, and hence (53) has a differentiable and locally unique solution $\big(\bar{x}(\Delta\mathcal{D}), \bar{Y}(\Delta\mathcal{D}), \bar{S}(\Delta\mathcal{D})\big)$. Furthermore, we have $\bar{Y}(\Delta\mathcal{D}), \bar{S}(\Delta\mathcal{D}) \succeq 0$. This semidefiniteness follows with a standard continuity argument: The optimality conditions of the nonlinear SDP coincide with the optimality conditions of the linearized SDP. Under our assumptions, the latter one has a unique optimal solution that depends continuously on small perturbations of the data; see, e.g., [15]. Hence the linearized problem at the data point $\mathcal{D} + \Delta\mathcal{D}$ has an optimal solution $\big(\tilde{x}, \tilde{Y}, \tilde{S}\big)$ that satisfies the same optimality conditions as $\big(\bar{x}(\Delta\mathcal{D}), \bar{Y}(\Delta\mathcal{D}), \bar{S}(\Delta\mathcal{D})\big)$. The solution of the linearized problem also satisfies $\tilde{Y} \succeq 0$, $\tilde{S} \succeq 0$. Since $\big(\bar{x}(\Delta\mathcal{D}), \bar{Y}(\Delta\mathcal{D}), \bar{S}(\Delta\mathcal{D})\big)$ is locally unique, it must coincide with $\big(\tilde{x}, \tilde{Y}, \tilde{S}\big)$, i.e., $\bar{Y}(\Delta\mathcal{D}), \bar{S}(\Delta\mathcal{D})$ satisfy the semidefiniteness conditions.

Step 4. In this step, we prove that the second-order sufficient condition is satisfied at the perturbed solution. Since feasible directions h are defined only up to a positive scalar factor, without loss of generality, one may require that $\|h\| = 1$. For the unperturbed problem (32), the second-order sufficient condition at $\bar{x}$ then states that $h^T H h \geq \mu$ for all $h \in \mathbb{R}^n$ with $\mathcal{A}(x + \epsilon h) + C \preceq 0$, $\epsilon = \epsilon(h) > 0$, $\|h\| = 1$, and $h^T(\bar{b} + H\bar{x}) = 0$. To prove that the second-order sufficient condition is invariant under small perturbations $\Delta\mathcal{D}$ of the problem data $\mathcal{D}$, we thus need to show that for some fixed $\tilde{\mu} > 0$, we have

$$h^T(H + \Delta H)h \geq \tilde{\mu} \tag{54}$$

for all solutions $h \in \mathbb{R}^n$ of the inequalities

$$\begin{aligned}(\mathcal{A} + \Delta\mathcal{A})\big(\bar{x}(\Delta\mathcal{D}) + \epsilon h\big) + C + \Delta C \preceq 0, \quad \epsilon = \epsilon(h) > 0,\\ \|h\| = 1, \quad h^T\big(b + \Delta b + (H + \Delta H)\bar{x}(\Delta\mathcal{D})\big) = 0.\end{aligned}$$

In view of the first two relations in (40), the above is equivalent to

$$\begin{aligned}&\epsilon(\mathcal{A} + \Delta\mathcal{A})(h) \preceq \bar{S}(\Delta\mathcal{D}), \quad \epsilon = \epsilon(h) > 0,\\ &\|h\| = 1, \quad h^T(\mathcal{A}^* + \Delta\mathcal{A}^*)(\bar{Y}(\Delta\mathcal{D})) = 0.\end{aligned} \tag{55}$$

It remains to show that the set of solutions h of (55) varies continuously with ΔD. Indeed, for any fixed $\tilde{\mu}$ with $0 < \tilde{\mu} < \mu$, the second-order condition at $\bar{x}$ then readily implies that (54) is satisfied for all solutions h of (55), provided $\|\Delta\mathcal{D}\|$ is sufficiently small.

In Step 2, we have shown that both $\bar{S}(\Delta\mathcal{D})$ and $\bar{Y}(\Delta\mathcal{D})$ are continuous functions of $\Delta\mathcal{D}$ and that the dimension of the null space of $\bar{S}(\Delta\mathcal{D})$ is constant, namely equal to k, for all sufficiently small $\|\Delta D\|$. Moreover, the null space of $\bar{S}(\Delta\mathcal{D})$ varies continuously with ΔD.

Let $\Delta\mathcal{D}_k$ be a sequence of perturbations with $\Delta\mathcal{D}_k \to 0$. Let h_k be a sequence of associated solutions of (55). It suffices to show that any accumulation point $\bar{h}$ of the sequence h_k satisfies (55) for $\Delta\mathcal{D} = 0$ and the associated values $\Delta\mathcal{A} = 0$, $\bar{Y}(0) = \bar{Y}$, $\bar{S}(0) = \bar{S}$. Since $\bar{Y}(\Delta\mathcal{D})$ and $\Delta\mathcal{A}^*$ vary continuously with $\Delta\mathcal{D}$, it follows that $\bar{h}$ satisfies the last two relations of (55) for $\Delta\mathcal{D} = 0$.

We now assume by contradiction that $\epsilon\mathcal{A}(\bar{h}) \not\preceq \bar{S}$ for any $\epsilon > 0$. Since $\bar{S} \succeq 0$, this implies that there exists a vector $z \in \mathbb{R}^m$ with $\|z\| = 1$, $z^T\mathcal{A}(\bar{h})z = \tilde{\epsilon} > 0$, and $z^T\bar{S}z = 0$. It follows that

$$z^T(\mathcal{A} + \Delta\mathcal{A}_k)(h_k)z \geq \frac{\tilde{\epsilon}}{2}$$

if k is sufficiently large. Since the null space of $\bar{S}(\Delta\mathcal{D})$ varies continuously with ΔD, we have

$$(z + \Delta z_k)^T\bar{S}(\Delta\mathcal{D}_k)(z + \Delta z_k) = 0$$

for some small $\Delta z_k \in \mathbb{R}^m$ whenever $\|\Delta\mathcal{D}_k\|$ is sufficiently small. We now choose $\|\Delta\mathcal{D}_k\|$ so small, i.e., k so large, that

$$(z + \Delta z_k)^T(\mathcal{A} + \Delta\mathcal{A}_k)(h_k)(z + \Delta z_k) \geq \frac{\tilde{\epsilon}}{4}.$$

This implies that h_k does not satisfy (55), and thus yields the desired contradiction. Hence $\bar{h}$ satisfies (55) for $\Delta\mathcal{D} = 0$. □

Theorem 1 can be sharpened slightly.

Corollary 1 *Under the assumptions of Theorem 1 there exists a small neighborhood $\mathcal{N}$ of zero in the data space of* (32) *such that for all perturbations $\Delta\mathcal{D} \in \mathcal{N}$ of the problem data* (33)*, $\mathcal{D}$, of* (32)*, there exists a local solution $(x_\Delta, Y_\Delta, S_\Delta)$ of* (40) *near $(\bar{x}, \bar{Y}, \bar{S})$, at which the assumptions of Theorem 1 are also satisfied. Moreover, the second derivatives*

$$\nabla^2_{\mathcal{D}}\big(x_\Delta, Y_\Delta, S_\Delta\big)[\Delta\mathcal{D}]$$

of such local solutions $(x_\Delta, Y_\Delta, S_\Delta)$ are uniformly bounded for all $\Delta\mathcal{D} \in \mathcal{N}$.

Proof The first part of the corollary is an immediate consequence of Theorem 1. For the second part observe that the second derivative is obtained by differentiating the system (38). For sufficiently small perturbations $\Delta\mathcal{D}$, the singular values of this system are uniformly bounded away from zero, and hence the second derivatives are uniformly bounded. □

Theorem 1 can be generalized to the class of NLSDPs of the form (22). Recall that, by (23) and (25), the Lagrangian of (22) and its Hessian are given by

$$\mathcal{L}(x, Y) = b^T x + \mathcal{B}(x) \bullet Y \quad \text{and} \quad H(x, Y) = \nabla^2_x\left(\mathcal{B}(x) \bullet Y\right), \tag{56}$$

respectively. The generalization of Theorem 1 to problems (22) is then as follows.

Theorem 2 *Let x^* be a local solution of* (22)*, and let Y^* be an associated Lagrange multiplier. Assume that the Robinson constraint qualification is satisfied at x^* and that the point (x^*, Y^*) is strictly complementary and locally unique. Finally, assume that the second-order sufficient condition holds at x^* with modulus $\mu > 0$. Then,* (22) *has a locally unique solution for small perturbations of the data $(\mathcal{B}, b)$, and the solution depends smoothly on the perturbations.*

Proof First, we define the linear function $\mathcal{A} := D_x\mathcal{B}(x^*) : \mathbb{R}^n \to \mathcal{S}^m$, and the matrices $C := \mathcal{B}(x^*)$ and $H := H(x^*, Y^*)$. Then, the SSP approximation (30) of (22) at the point (x^*, Y^*) is just the quadratic semidefinite problem

$$\begin{array}{lll} \text{minimize} \quad b^T \Delta x + \dfrac{1}{2}(\Delta x)^T H \Delta x & \text{subject to} & \Delta x \in \mathbb{R}^n, \\ & & \mathcal{A}(\Delta x) + C \preceq 0. \end{array} \tag{57}$$

Note that (57) is a problem of the form (32) with data (33), $\mathcal{D}$. Moreover, $\Delta\bar{x} := 0$, $\bar{Y} := Y^*$, and $\bar{S} := -\mathcal{A}(0) - C$ satisfy the conditions (35). These conditions coincide with the first-order conditions of (22), and thus the point $(\Delta\bar{x}, \bar{Y}, \bar{S})$ is also the unique solution of (35). Furthermore, the second-order sufficient condition for (57) and (22) coincide. This condition guarantees that $\Delta\bar{x}$ is a locally unique solution of (57). Finally, the Robinson constraint qualification for problem (22) at x^* implies that problem (57) satisfies Slater's condition. In particular, all assumptions of Theorem 1 are satisfied. Small perturbations $\Delta\mathcal{D}$ of the data of (22) result in small changes of the corresponding SSP problem (57). Since Theorem 1 allows for arbitrary changes in all of the data of (57), the claims follow. □

8 Convergence of the SSP method

In this section, we prove that the plain SSP method with step size 1 is locally quadratically convergent.

For pairs (x, Y), where $x \in \mathbb{R}^n$, $Y \in \mathcal{S}^m$, we use the norm

$$\big\|(x, Y)\big\| := \big(\|x\|^2 + \|Y\|^2\big)^{\frac{1}{2}}.$$

The main result of this section can then be stated as follows.

Theorem 3 *Assume that the function $\mathcal{B}$ in* (22) *is $\mathcal{C}^3$-differentiable and that problem* (22) *has a locally unique and strictly complementary solution $(\bar{x}, \bar{Y})$ that satisfies the Robinson constraint qualification and the second-order sufficient condition with modulus $\mu > 0$, cf.* (37)*. Let some iterate (x^k, Y^k) be given and let the next iterate*

$$(x^{k+1}, Y^{k+1}) := (x^k, Y^k) + (\Delta x, \Delta Y)$$

be defined as the local solution of (28)*, or, equivalently,* (30)*, that is closest to (x^k, Y^k). Then there exist $\epsilon > 0$ and $\gamma < 1/\epsilon$ such that*

$$\big\|(x^{k+1}, Y^{k+1}) - (\bar{x}, \bar{Y})\big\| \le \gamma \big\|(x^k, Y^k) - (\bar{x}, \bar{Y})\big\|^2$$

whenever $\big\|(x^k, Y^k) - (\bar{x}, \bar{Y})\big\| < \epsilon$.

Proof The proof is divided into three steps. In the first step, we establish the exact relation of problems (30) and (32). In a second step, we consider a point x^k near $\bar{x}$. We show that x^k is the optimal solution of an SSP subproblem the data of which is at most $\mathcal{O}(\|x^k - \bar{x}\|)$ away from the data of the SSP subproblem at $(\bar{x}, \bar{Y})$. We remark that x^k is always the optimal solution of the $(k-1)$-th subproblem, but the data of this subproblem lies $\mathcal{O}(\|x^{k-1} - \bar{x}\|)$ away from the SSP subproblem at $(\bar{x}, \bar{Y})$. In a third step, we then show by a perturbation analysis that the correction

$\Delta x = x^{k+1} - x^k$ is of size $\mathcal{O}(\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)$ and that the residual for the SSP subproblem in the $(k+1)$-th step is of size $\mathcal{O}((\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)^2)$.

Step 1. We first show how the SSP subproblem (30) can be written in the form (32). To this end, we define the linear function $\mathcal{A} := D_x\mathcal{B}(x^k) : \mathbb{R}^n \to \mathcal{S}^m$, and the matrices $C := \mathcal{B}(x^k)$ and $H := H(x^k, Y^k)$. Note that the linear constraint $\mathcal{A}(\Delta x) + C \preceq 0$ is just the linearization of the nonlinear constraint $\mathcal{B}(x^k + \Delta x) \preceq 0$ about the point x^k. Finally, let b be as in (22). The SSP subproblem (30) then takes the simple form (57), and in particular, it conforms with the format (32) of Theorem 2.

Step 2. Let any point x^k close to $\bar{x}$ be given. We show that $\Delta x = 0$ is a local solution of a problem of the form (57), where the data is 'close' to the data of (30) at $(\bar{x}, \bar{Y})$. Let $\Delta C := \mathcal{B}(\bar{x}) - \mathcal{B}(x^k)$. By continuity of $\mathcal{B}$, $\|\Delta C\|$ is of order $\mathcal{O}(\|x^k - \bar{x}\|)$. Let

$$\Delta b := -\nabla_x \left(\mathcal{B}(x^k) \bullet \bar{Y}\right) - b = -\mathcal{A}^*(\bar{Y}) - b.$$

From (24) and the second row of (26), it follows that $\|\Delta b\| = \mathcal{O}(\|x^k - \bar{x}\|)$. Moreover, the point $(0, \bar{Y}, \bar{S})$ satisfies the first-order conditions,

$$\mathcal{A}(0) + C + \Delta C + \bar{S} = 0, \quad b + \Delta b + H \cdot 0 + \mathcal{A}^*(\bar{Y}) = 0, \qquad \bar{Y}\,\bar{S} = 0,$$

for the quadratic semidefinite program

$$\begin{aligned} \text{minimize } (b + \Delta b)^T \Delta x + \frac{1}{2}(\Delta x)^T H \Delta x \;\; \text{subject to } \Delta x \in \mathbb{R}^n, \\ \mathcal{A}(\Delta x) + C + \Delta C \preceq 0. \end{aligned} \tag{58}$$

Let

$$\bar{\mathcal{A}} := D_x\mathcal{B}(\bar{x}), \quad \bar{b} := b, \quad \bar{C} := \mathcal{B}(\bar{x}), \quad \bar{H} := \nabla_x^2 \mathcal{L}(\bar{x}, \bar{Y}), \tag{59}$$

be the data of the SSP subproblem (30) at the point $(\bar{x}, \bar{Y})$. Then, the data of (58) differs from the data (59) by a perturbation of norm $\mathcal{O}(\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)$. Here, the term $\|Y^k - \bar{Y}\|$ reflects the fact that H and $\bar{H}$ also differ by the choice of Y. Note that the point $(\Delta x, Y, S) = (0, \bar{Y}, \bar{S})$ satisfies the first-order optimality conditions,

$$\bar{\mathcal{A}}(\Delta x) + \bar{C} + S = 0, \quad b + \bar{H}\Delta x + \bar{\mathcal{A}}^*(Y) = 0, \quad YS = 0, \tag{60}$$

of the quadratic problem (57) with data (59). As shown in Theorem 2, the assumptions for the nonlinear SDP (22) at $(\bar{x}, \bar{Y})$ imply that problem (57) with data (59) satisfies all conditions of Theorem 2 at $(0, \bar{Y}, \bar{S})$.

Since the second-order sufficient condition depends continuously on the data of (32), it follows that for (58), the second-order condition at $(0, \bar{Y}, \bar{S})$ is satisfied, provided that $\|x^k - \bar{x}\|$ and $\|Y^k - \bar{Y}\|$ are sufficiently small. Thus, problem (58) fulfills all assumptions of Theorem 2.

Step 3. By definition, $(\Delta x, Y) = (0, \bar{Y})$ is the optimal solution (with associated multiplier) of (58). Let (x^k, Y^k) be close to $(\bar{x}, \bar{Y})$. The SSP subproblem replaces the data Δb and ΔC of (58) by 0 (of the respective dimension). Thus, the data of (58) is changed by a perturbation of order $\mathcal{O}(\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)$. We assume that this perturbation lies in the neighborhood $\mathcal{N}$ about zero as guaranteed by

Corollary 1. Denote the optimal solution of the SSP subproblem by $(\Delta x, \bar{Y} + \Delta\hat{Y})$.

The SSP subproblem is then used to define (x^{k+1}, Y^{k+1}). Let

$$\mathcal{A}^+ := D_x\mathcal{B}(x^{k+1}), \quad C^+ := \mathcal{B}(x^{k+1}), \quad H^+ := \nabla_x^2\mathcal{L}(x^{k+1}, Y^{k+1})$$

be the data of the SSP subproblem at the next, $(k+1)$-th, iteration.

Corollary 1 states that $(\Delta x, \Delta\hat{Y})$ are given by the tangent equations (38) plus some uniformly bounded second-order terms. Thus, $(\Delta x, \Delta\hat{Y})$ are of the order $\mathcal{O}(\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)$. Here, $\Delta\hat{Y}$ is a correction of the unknown point $\bar{Y}$, while the correction $\Delta Y = Y^{k+1} - Y^k$ produced by the SSP subproblem has the form $\Delta Y = \Delta\hat{Y} + \bar{Y} - Y^k$. Obviously, also the norm $\|\Delta Y\|$ of this correction is of the order $\mathcal{O}(\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)$.

Next, we compute an upper bound on the size of the residuals of the first and second equations in (60) at $(x^{k+1}, Y^{k+1}, S^{k+1})$. Note that the residual term of the third equation in (60) is zero. By definition of $(\Delta x, Y^{k+1}, S^{k+1})$, it follows that

$$\mathcal{A}(\Delta x) + C + S^{k+1} = 0, \quad b + H\Delta x + \mathcal{A}^*(Y^{k+1}) = 0, \quad Y^{k+1}\,S^{k+1} = 0. \quad (61)$$

If the data of (22) is C^3-smooth, this implies that

$$\begin{aligned}
&(\mathcal{A}^+)^*(Y^{k+1}) + b \\
&= \mathcal{A}^*(Y^{k+1}) + b + \Delta\mathcal{A}^*(Y^{k+1}) \\
&= -H\Delta x + \Delta\mathcal{A}^*(Y^k) + \Delta\mathcal{A}^*(\Delta Y) \\
&= -\nabla_x^2\big(\mathcal{B}(x^k) \bullet Y^k\big)\,\Delta x + \big(\nabla_x\mathcal{B}(x^k + \Delta x) - \nabla_x\mathcal{B}(x^k)\big) \bullet Y^k + \Delta\mathcal{A}^*(\Delta Y) \\
&= \mathcal{O}(\|\Delta x\|^2 + \|\Delta Y\|^2),
\end{aligned}$$

where $\Delta\mathcal{A} := \mathcal{A}^+ - \mathcal{A}$. Likewise, it follows from (61) that

$$\begin{aligned}
C^+ + S^{k+1} &= \Delta C + C + S^{k+1} = \Delta C - \mathcal{A}(\Delta x) \\
&= \mathcal{B}(x^{k+1}) - \mathcal{B}(x^k) - D_x\mathcal{B}(x^k)[\Delta x] = \mathcal{O}(\|\Delta x\|^2),
\end{aligned}$$

where $\Delta C := C^+ - C$. Hence, we can define perturbations $\hat{b}$ and $\hat{C}^+$ of b and C^+ of order

$$\|\hat{b} - b\| + \|\hat{C}^+ - C^+\| = \mathcal{O}((\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)^2)$$

such that $(\Delta x, Y, S) = (0, Y^{k+1}, S^{k+1})$ is an optimal solution of the problem (57) with data $\mathcal{A}^+$, $\hat{b}$, $\hat{C}^+$, H^+. By the same derivation as above, the next SSP step has length $\mathcal{O}((\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)^2)$, and thus it generates residuals of order $\mathcal{O}((\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)^4)$. Repeating this process, a standard argument shows that $\|x^{k+1} - \bar{x}\|$ and $\|Y^{k+1} - \bar{Y}\|$ are of order $\mathcal{O}((\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)^2)$ as well. □

Remark 5 As mentioned before, one will typically choose to solve SSP subproblems with a positive semidefinite approximation $\hat{H}$ to the Hessian of the Lagrangian. A proof of convergence for such modifications is the subject of current

research; see, e.g., [10]. Since all the data enters in a continuous fashion in the preceding analysis, it follows that the SSP method with step size one is still locally superlinearly convergent if the matrices H^k in (29) are replaced by approximations $\hat{H}^k$ with $\|H^k - \hat{H}^k\| \to 0$.

Remark 6 The assumption of $\mathcal{C}^3$-differentiability of the function $\mathcal{B}$ in Theorem 3 can be weakened to $\mathcal{C}^2$-differentiability and a Lipschitz condition for the Hessian at $\bar{x}$.

The result of Theorem 3 can be extended to the following slightly more general class of NLSDPs. Given a vector $b \in \mathbb{R}^n$, a matrix-valued function $\mathcal{B} : \mathbb{R}^n \to \mathcal{S}^m$, and two vector-valued functions $c : \mathbb{R}^n \to \mathbb{R}^p$ and $d : \mathbb{R}^n \to \mathbb{R}^q$, we consider problems of the following form:

$$\begin{array}{lll} \text{minimize} \quad b^T x \quad \text{subject to} & x \in \mathbb{R}^n, \\ & \mathcal{B}(x) \preceq 0, \\ & c(x) \leq 0, \\ & d(x) = 0. \end{array} \tag{62}$$

The Lagrangian of problem (22) takes the form $\mathcal{L} : \mathbb{R}^n \times \mathcal{S}^m \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$:

$$\mathcal{L}(x, Y, u, v) := b^T x + \mathcal{B}(x) \bullet Y + u^T c(x) + v^T d(x). \tag{63}$$

Its gradient with respect to x is given by

$$\begin{aligned} g(x, Y, u, v) &:= \nabla_x \mathcal{L}(x, Y, u, v) \\ &= b + \nabla_x \left(\mathcal{B}(x) \bullet Y\right) + \nabla_x c(x)\, u + \nabla_x d(x)\, v \end{aligned} \tag{64}$$

and its Hessian by

$$\begin{aligned} H(x, Y, u, v) &:= \nabla_x^2 \mathcal{L}(x, Y, u, v) \\ &= \nabla_x^2 \left(\mathcal{B}(x) \bullet Y\right) + \sum_{i=1}^{p} u_i \nabla_x^2 c_i(x) + \sum_{j=1}^{q} v_j \nabla_x^2 d_j(x). \end{aligned} \tag{65}$$

Note that in (64), the gradients of the vector-valued functions $c(x)$ and $d(x)$ are defined as $\nabla_x c(x) := (D_x c(x))^T$ and $\nabla_x d(x) := (D_x d(x))^T$.

For NLSDPs (62), the SSP subproblems are of the form

$$\begin{array}{lll} \text{minimize} \quad b^T \Delta x + \dfrac{1}{2} (\Delta x)^T H^k \Delta x \quad \text{subject to} & \Delta x \in \mathbb{R}^n, \\ & \mathcal{B}(x^k) + D_x \mathcal{B}(x^k)[\Delta x] \preceq 0, \\ & c(x^k) + D_x c(x^k)\, \Delta x \leq 0, \\ & d(x^k) + D_x d(x^k)\, \Delta x = 0. \end{array} \tag{66}$$

The extension of Theorem 3 to problems (62) is as follows.

Theorem 4 *Assume that the functions $\mathcal{B}$, c, and d in* (62) *are $\mathcal{C}^3$-differentiable, and that problem* (62) *has a locally unique and strictly complementary solution $(\bar{x}, \bar{Y}, \bar{u}, \bar{v})$ that satisfies the Robinson constraint qualification and the second-order sufficient condition with modulus $\mu > 0$, cf.* (37)*. Let (x^k, Y^k, u^k, v^k) be some given iterate, and let the next iterate*

$$(x^{k+1}, Y^{k+1}, u^{k+1}, v^{k+1}) := (x^k, Y^k, u^k, v^k) + (\Delta x, \Delta Y, \Delta u, \Delta v)$$

be defined as the local solution of (66) *that is closest to (x^k, Y^k, u^k, v^k). Then there exist $\epsilon > 0$ and $\gamma < 1/\epsilon$ such that*

$$\big\|(x^{k+1}, Y^{k+1}, u^{k+1}, v^{k+1}) - (\bar{x}, \bar{Y}, \bar{u}, \bar{v})\big\| \leq \gamma \big\|(x^k, Y^k, u^k, v^k) - (\bar{x}, \bar{Y}, \bar{u}, \bar{v})\big\|^2$$

whenever $\big\|(x^k, Y^k, u^k, v^k) - (\bar{x}, \bar{Y}, \bar{u}, \bar{v})\big\| < \epsilon$.

Proof By our assumption on strict complementarity, all entries of the vector $\bar{v}$ of the Lagrange multipliers associated with the equality constraints $d(x) = 0$ of (22) are different from zero. Without loss of generality, we assume that $\bar{v} > 0$. Indeed, for any entry $\bar{v}_j < 0$ we replace the corresponding constraint $d_j(x) = 0$ by the equivalent constraint $-d_j(x) = 0$. These sign changes do not change the iterates generated by (30). Moreover, for $\big(x^k, Y^k, u^k, v^k\big)$ sufficiently close to $\big(\bar{x}, \bar{Y}, \bar{u}, \bar{v}\big)$ it follows from $\bar{v} > 0$ that the iterates do not change when the constraints $d(x) = 0$ are replaced by $d(x) \leq 0$. We can thus assume that $q = 0$, i.e., there are no equality constraints in (62).

We further assume that, without loss of generality, the matrix $\mathcal{B}$ is augmented to a 2×2 block diagonal matrix, where the $(2,2)$-block is the diagonal matrix $\mathrm{Diag}(c(x))$. Thus, for the analysis of the SSP method we may assume that $p = q = 0$ in (22), i.e., we only need to consider problems of the form (22). □

9 Numerical results

In this section, we present results of some numerical experiments with a Matlab implementation of the SSP method. Actually, our Matlab program is for a slightly more general class of nonlinear programs with conic constraints (NLCPs). The numerical experiments with our program illustrate the theoretical results of the preceding sections. In particular, quadratic convergence is observed for problems where the Hessian H of the Lagrangian at the optimal solution is positive semidefinite. In cases where H is not positive semidefinite, our implementation uses perturbations of the nonconvex SSP subproblems in order to obtain convex conic subproblems. In these cases, typically, the rate of convergence of the algorithm based on such perturbed problems is only linear.

The Matlab program generates its search directions by solving conic quadratic subproblems using Version 1.05R5 of SeDuMi [34]. SeDuMi allows free and positive variables as well as Lorentz-cone ('ice-cream cone') constraints, rotated Lorentz-cone constraints, and semidefinite cone constraints. The NLCPs can also be formulated in terms of these cones. In order to simplify the use of SeDuMi for the SSP subproblems, the NLCPs are rewritten in the following standard format:

$$\text{minimize} \quad c^T x \quad \text{subject to} \quad \begin{array}{l} x \in K, \\ F(x) = 0. \end{array} \tag{67}$$

Here, K is a Cartesian product of free variables and several cones of the types allowed in SeDuMi.

We tested the following techniques for generating positive semidefinite approximations $\hat{H}$ of H: a BFGS approach, the Hessian of the augmented Lagrangian, and the orthogonal projection of H onto the cone of positive semidefinite matrices. Our experiences with these techniques are as follows.

I.1 The BFGS approach can result in considerably more SSP iterations compared to the projection of the Hessian of the Lagrangian. Moreover, the BFGS approach strongly depends on the initial matrix $\hat{H}^0$. A good choice is $\hat{H}^0 := V \max(D, \epsilon I) V^T$, where $H(x^0, Y^0) = V D V^T$ is the eigenvalue decomposition of $H(x^0, Y^0)$.

I.2 The use of the Hessian of the augmented Lagrangian can be a good choice for some problems, but for most of our test problems the penalty parameter had to be very large to obtain a semidefinite Hessian. This, in turn, significantly reduced the precision of our computations.

I.3 In spite of not being affinely invariant, the use of the projection of the Hessian of the Lagrangian resulted in the most efficient overall algorithm.

We also tested different step length strategies.

II.1 The following penalty line search with a quadratic correction gave good results for all test cases. The SSP subproblem provides a search direction Δx for problem (67). By solving a least-squares problem, a vector q is computed satisfying $D_x F(x) q = -F(x + \Delta x)$. For $\lambda \in [0, 1]$, a line search along the points $x(\lambda) := x + \lambda \Delta x + \lambda^2 q$ is performed based on the penalty function

$$M \|F(x(\lambda))\| + c^T x(\lambda),$$

where $M > 0$ is a penalty parameter.

II.2 For some examples, the choice of a filter approach was slightly better. In the filter approach used here, a Euclidean trust-region radius was always set to be 1.5 times larger than the previous step, and non-acceptable steps were not discarded, but instead an Armijo-type step-length reduction was used to generate an acceptable step. The motivation for this modified filter strategy lies in the fact that the computation of a solution of a subproblem is very expensive, and therefore discarding the solution of a subproblem is avoided. The above filter approach led to very fast convergence, especially for convex problems.

II.3 For the examples presented here, the trust-region approach was the best choice. The SSP subproblem was restricted by an additional Euclidean trust-region constraint. For problems of the form (68) below, it was sufficient to apply the trust-region constraint only to the variables X_G and X_C, while P and S remain free. For these examples, an additional corrector step significantly accelerated the convergence. For this corrector step, X_G and X_C are kept fixed, and P and S are updated by solving an additional linear SDP. At each iteration, the ratio between predicted and actual reduction was computed. Depending on that ratio, the step was accepted and the trust region was increased or decreased, or the step was rejected and the trust region was decreased.

For our numerical examples, we use nonlinear nonconvex SDPs of the form (10), which we rewrite in the form (68) below. Recall that in (10), $G, C \in \mathbb{R}^{n \times n}$

and $B_1,\ B_2 \in \mathbb{R}^{n\times m}$ are given data matrices. The nonconvex NLSDP used for the numerical examples is then as follows:

$$\begin{aligned} \text{minimize} \quad \|S\| \quad \text{subject to} \quad & P \in \mathbb{R}^{n\times n},\ S \in \mathbb{R}^{n\times m}, \\ & X_G \in \mathbb{R}^{n\times n},\ \|X_G\| \le r_G, \\ & X_C \in \mathbb{R}^{n\times n},\ \|X_C\| \le r_C, \\ & P^T B_1 + S = B_2, \\ & P^T(G+X_G) + (G+X_G)^T P \succeq \varepsilon_G I, \\ & P^T(C+X_C) + (C+X_C)^T P \succeq \varepsilon_C I, \\ & P^T(C+X_C) - (C+X_C)^T P = 0. \end{aligned} \tag{68}$$

Furthermore, in (68), in addition to the constraints on the norms of the perturbations X_G and X_C, we restrict X_G and X_C to have possible nonzero entries only in certain positions, which depend on the nonzero structure of the given matrices G and C, respectively. For our numerical tests, the data matrices G and C in (68) are generated as follows. First, two matrices C_{org} and G_{org} of norm one were constructed such that the associated transfer function is guaranteed to be positive real. Then, certain entries of G_{org} and C_{org} were changed by adding random perturbations of norm at most $\tilde{\varepsilon}_G = \tilde{\varepsilon}_C \le 10^{-2}$ respectively, to define the data matrices G and C. The values ε_G and ε_C were chosen in the interval $[10^{-15}, 10^{-7}]$. In all our examples, the transfer functions of the systems given by the resulting matrices C and G were not positive real. As initial points for problem (68), we used the zero matrices $X_G^0 = 0$ and $X_C^0 = 0$. For fixed $X_G = 0$ and $X_C = 0$, problem (68) reduces to a linear SDP, the solution of which was taken as initial values P^0 and S^0 for P and S.

All our computations were run on a Xeon with a clock rate of 2.8 GHz and 3 GB RAM. All solutions were computed to a precision of 12 decimal digits. In Table 1, we list the problem dimension n, the total number $M(n)$ of equality constraints, the total number $N(n)$ of scalar unknowns, the number of iterations, and the cpu time (in seconds) for the SSP algorithm using the convexification I.3 and the step length control II.3 above.

Table 1 shows that the number of iterations is nearly independent of the dimension n of the problem, while—as expected—the cpu time increases with n. The total number of constraints is approximately $M(n) \approx \frac{3}{2}n^2$, and the total number of scalar variables is approximately $N(n) \approx 3n^2$. The number of iterations to solve the linear semidefinite subproblems not only depends on the dimension, but also on other properties of the problem as, for example, a comparison of the problems of dimension 32 and 33 shows. In this case, the iteration counts differ only by one, yet the cpu time quadruples, since SeDuMi needs more iterations to solve the subproblems. Some of the linear semidefinite subproblems are nearly infeasible, a situation for which SeDuMi (and other solvers) needs a higher number of interior-point iterations. In practical applications, the values $\tilde{\varepsilon}_G$ and $\tilde{\varepsilon}_C$ are small, but the required accuracy of the optimal solution of (68) is high. In our examples, the optimal perturbations X_G and X_G generated by the algorithm had a norm of about 10^{-3}. Thus, the starting point appears to be very close to the optimal solution; nevertheless, a plain primal predictor-corrector implementation with adaptive

Table 1 Numerical results for nonconvex NLSDPs of the form (68).

n	$M(n)$	$N(n)$	iter	cpu time	n	$M(n)$	$N(n)$	iter	cpu time
8	118	285	5	3.71	22	783	1713	4	416.31
9	146	348	5	4.43	23	853	1862	5	151.33
10	177	417	7	8.06	24	926	2013	6	683.54
11	211	492	5	7.18	25	1002	2176	3	145.60
12	248	573	8	16.05	26	1081	2337	5	612.22
13	288	660	4	10.88	27	1163	2508	7	518.92
14	331	753	6	20.77	28	1248	2685	5	789.41
15	377	852	7	30.12	29	1336	2868	4	475.52
16	426	957	6	34.38	30	1427	3057	7	4213.50
17	478	1068	5	37.40	31	1521	3252	4	784.34
18	533	1185	10	91.17	32	1618	3455	6	4659.64
19	591	1308	4	47.61	33	1718	3660	5	1130.44
20	652	1437	5	83.66	34	1821	3877	2	630.53
21	716	1572	4	289.48	35	1927	4092	6	1799.36

step length control failed to solve this type of problem. The quadratic rate of convergence of the SSP method anticipated by the theoretical results could not be observed in the numerical experiments. One reason is that the solutions X_C, X_G, P, S of (68) were computed to high accuracy where rounding errors significantly perturbed the solutions of the subproblems. Moreover, due to the convexification used, quadratic convergence is no longer guaranteed by the theoretical results of this paper.

10 Concluding remarks

We have discussed the SSP method, which is a generalization of the SQP method for standard nonlinear programs to nonlinear semidefinite programming problems. Such a generalization has already been studied in [12]; the main differences between our approach and the one in [12] are the motivation and the analysis. For the derivation of the SSP method, we have chosen a motivation that contrasts the SSP method with primal-dual interior-point methods. In interior-point methods that are applied directly to nonlinear semidefinite programs, the symmetrization procedure, the linearization, and the convexification are applied in one step. In our proposed method, these operations are separated, and the choice of the symmetrization scheme is shifted to the subproblems. The convergence analysis thus avoids the consideration of the symmetrization operator. Centering steps are not needed in this algorithm. Our convergence analysis differs from the convergence analyses of standard SQP methods in that it is based on a sensitivity result for certain optimal solutions of quadratic semidefinite programs. The derivation of this sensitivity result is also of independent interest.

Acknowledgements We would like to thank two anonymous referees whose constructive comments helped us to improve the presentation of the paper.

References

1. Alizadeh, F.: Interior point methods in semidefinite programming with applications to combinatorial optimization. SIAM J. Optim. **5**, 13–51 (1995)
2. Alizadeh, F., Haeberly, J.-P.A., Overton, M.L.: Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results, SIAM J. Optim. **8**, 746–768 (1998)
3. Anderson, B.D.O., Vongpanitlerd, S.: Network Analysis and Synthesis. Prentice-Hall, Englewood Cliffs, New Jersey (1973)
4. Bai, Z., Freund, R.W.: Eigenvalue-based characterization and test for positive realness of scalar transfer functions. IEEE Trans. Automat. Control **45**, 2396–2402 (2000)
5. Bai, Z., Freund, R.W.: A partial Padé-via-Lanczos method for reduced-order modeling, Linear Algebra Appl. **332–334**, 139–164 (2001)
6. Boggs, P.T., Tolle, J.W.: Sequential quadratic programming. Acta Numer. **4**, 1–51 (1995)
7. Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C.: Optimisation Numérique. Springer-Verlag, Berlin (1997)
8. Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V.: Linear Matrix Inequalities in System and Control Theory. SIAM Publications, Philadelphia (1994)
9. Coelho, C.P., Phillips, J.R., Silveira, L.M.: A convex programming approach for generating guaranteed passive approximations to tabulated frequency-data. IEEE Trans. Computer-Aided Design **23**, 293–301 (2004)
10. Correa, R., Ramírez C., H.: A global algorithm for nonlinear semidefinite programming. SIAM J. Optim. **15**, 303–318 (2004)
11. Fares, B., Apkarian, P., Noll, D.: An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. Internat. J. Control **74**, 348–360 (2001)
12. Fares, B., Noll, D., Apkarian, P.: Robust control via sequential semidefinite programming. SIAM J. Control Optim. **40**, 1791–1820 (2002)
13. Freund, R.W.: Model reduction methods based on Krylov subspaces. Acta Numer. **12**, 267–319 (2003)
14. Freund, R.W., Jarre, F.: An extension of the positive real lemma to descriptor systems. Optim. Methods Softw. **19**, 69–87 (2004)
15. Freund, R.W., Jarre, F.: A sensitivity result for semidefinite programs. Oper. Res. Lett. **32**, 126–132 (2004)
16. Helmberg, C., Rendl, F., Vanderbei, R., Wolkowicz, H.: An interior-point method for semidefinite programming, SIAM J. Optim. **6**, 342–361 (1996)
17. Hörmander, L.: An Introduction to Complex Analysis in Several Variables, Second Edition. North-Holland Publishing Company, Amsterdam (1973)
18. Kocvara, M., Stingl, M.: PENNON — A code for convex nonlinear and semidefinite programming. Optim. Methods Softw. **18**, 317–333 (2003)
19. Kojima, M., Shindoh, S., Hara, S.: Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices, SIAM J. Optim. **7**, 86–125 (1997)
20. Luo, Z.-Q., Sturm, J.F., Zhang, S.: Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming, SIAM J. Optim. **8**, 59–81 (1998)
21. Mangasarian, O.L.: Nonlinear Programming. MvGraw-Hill, New York (1969)
22. Monteiro, R.: Primal-dual path-following algorithms for semidefinite programming, SIAM J. Optim. **7**, 663–678 (1997)
23. Nayakkankuppam, M.V., Overton, M.L.: Conditioning of semidefinite programs. Math. Program. **85**, 525–540 (1999)
24. Y. Nesterov, Y., Todd, M.J.: Self-scaled barriers and interior-point methods in convex programming, Math. Oper. Res. **22**, 1–42 (1997)
25. Y. Nesterov, Y., Todd, M.J.: Primal-dual interior-point methods for self-scaled cones, SIAM J. Optim. **8**, 324–364 (1998)
26. D. Noll: personal communication, September 2004
27. Overton, M.L.: On minimizing the maximum eigenvalue of a symmetric matrix. SIAM J. Matrix Anal. Appl. **9**, 256–268 (1988)
28. Overton, M.L.: Large-scale optimization of eigenvalues. SIAM J. Optim. **2**, 88–120 (1992)
29. Powell, M.J.D.: A fast algorithm for nonlinearly constrained optimization calculations. In: Lecture Notes in Mathematics Vol. 630. Springer-Verlag, Berlin, pp. 144–157 (1978)
30. Robinson, S.M.: First order conditions for general nonlinear optimization. SIAM J. Appl. Math. **30**, 597–607 (1976)

31. Robinson, S.M.: Generalized equations and their solutions, part II: applications to nonlinear programming. Math. Prog. Study **19**, 200–221 (1982)
32. Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. Math. Program. **77**, 301–320 (1997)
33. Shapiro, A., Scheinberg, K.: Duality and optimality conditions. In: Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.) Handbook of Semidefinite Programming: Theory, Algorithms and Applications. Kluwer Academic Publishers, Boston, pp. 67–110 (2000)
34. Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optim. Methods Softw. **11–12**, 625–653 (1999)
35. Sturm, J.F., Zhang, S.: On sensitivity of central solutions in semidefinite programming. Math. Program. **90**, 205–227 (2001)
36. Todd, M.J.: Semidefinite optimization. Acta Numer. **10**, 515–560 (2001)
37. Tuan, H.D., Apkarian, P., Nakashima, Y.: A new Lagrangian dual global optimization algorithm for solving bilinear matrix inequalities. Internat. J. Robust Nonlinear Control **10**, 561–578 (2000)
38. Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Rev. **38**, 49–95 (1996)
39. Vanderbei, R.J.: LOQO user's manual—version 3.10. Report SOR 97-08, Princeton University, Princeton, New Jersey (1997)
40. Vanderbei, R.J., Benson, H.Y., Shanno, D.F.: Interior-point methods for nonconvex nonlinear programming: filter methods and merit functions. Comput. Optim. Appl. **23**, 257–272 (2002)
41. Wolkowicz, H., Saigal, R., Vandenberghe, L., eds.: Handbook of Semidefinite Programming: Theory, Algorithms and Applications. Kluwer Academic Publishers, Boston (2000)