A sensitivity analysis and a convergence result for a sequential semidefinite programming method*

Roland W. Freund
Bell Laboratories, Lucent Technologies
Room 2C–525
700 Mountain Avenue
Murray Hill, New Jersey 07974–0636, U.S.A.
e-mail: freund@research.bell-labs.com

Florian Jarre
Institut für Mathematik, Universität Düsseldorf
Universitätsstraße 1, D–40225 Düsseldorf, Germany
e-mail: jarre@opt.uni-duesseldorf.de

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. In particular, a suitable symmetrization procedure needs to be chosen. The choice of the symmetrization procedure can be shifted in a very natural way to certain linear semidefinite subproblems, and can thus be reduced to a well-studied problem. The resulting sequential semidefinite programming (SSP) method is a generalization of the well-known SQP method for standard nonlinear programs. We present a sensitivity result for nonlinear semidefinite programs, and then use this result to derive an elementary and self-contained proof of local quadratic convergence of the SSP method.

Key words. semidefinite programming, nonlinear programming, sequential quadratic programming, quadratic semidefinite program, sensitivity, convergence

1 Introduction

In this paper we consider the solution of nonlinear semidefinite programs (NLSDPs). In recent years, interior-point methods for solving linear semidefinite programs (SDPs) have received a lot of attention and as a result, are now very well developed; see, e.g. [23, 25], the papers in [28], 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.

*Numerical Analysis Manuscript No. 03–4–08, Bell Laboratories, Murray Hill, New Jersey, April 2003. Available on WWW at `http://cm.bell-labs.com/cs/doc/03`.

The generalization of interior-point methods for SDPs to NLSDPs, however, is difficult as the step length now also depends on the quality of the linearization of the nonlinear functions. Moreover, the choice of the symmetrization procedure is considerably more complicated than in the linear case since the system matrix is no longer positive semidefinite. To address this difficulty, in this paper 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 [19] in the more general framework of nonlinear programs over convex cones. This framework includes nonlinear semidefinite programs as a special case. While Robinson did not discuss implementational issues of such a generalized SQP approach, the recent progress in linear semidefinite programming 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 an elementary and self-contained proof of local quadratic convergence of the SSP method under the assumption that the optimal solution is locally unique, strictly complementary, and satisfies a second-order sufficient condition.

One of the first numerical approaches for solving a class of NLSDPs was given in [16, 17]. Other recent approaches for solving NLSDPs are the program package LOQO [26] based on a primal-dual approach; see also [27]. Another promising approach for solving large-scale semidefinite programs is the modified barrier approach proposed in [12]. 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 [6, 7, 24]. In [6], the augmented Lagrangian method is applied to NLSDPs, while the approach proposed [7] is based on an SQP method generalized to NLSDPs. The paper [7] also contains a proof of local quadratic convergence. However, in contrast to this paper, the algorithm [7] 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 [5], Correa and Ramírez present a proof of global convergence of a modification of the method proposed in [7]. 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 recall known results for linear semidefinite programs in a form that can be easily transferred to nonlinear semidefinite programs. In Section 4, we discuss primal-dual systems for nonlinear semidefinite programs, and in Section 5, the SSP method is introduced as a generalized SQP method. In Section 6, we present sensitivity results, first for a certain class of quadratic semidefinite programs and then for general nonlinear semidefinite programs. Based on these sensitivity results, in Section 7, we give an elementary and self-contained proof of local quadratic convergence of the SSP method. Finally, in Section 8, we make some concluding remarks.

2 Notation

Throughout this article, all vectors and matrices are assumed to have real entries. As usual, $Y^T = [y_{ji}]$ denotes the transpose of the matrix $Y = [y_{ij}]$. 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 $x \in I\!R^n$, $x \geq 0$ 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 $Y = [y_{ij}]$ and $Z = [z_{ij}]$ 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 $I\!R^n$ to $\mathcal{S}^m$. We use the symbol $\mathcal{A} : I\!R^n \to \mathcal{S}^m$ if such a function is *linear*, and the symbol $\mathcal{B} : I\!R^n \to \mathcal{S}^m$ if such a function is *nonlinear*.

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

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

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

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

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

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

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 always assume that nonlinear functions $\mathcal{B} : I\!R^n \to \mathcal{S}^m$ 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 $x \in I\!R^n$, the derivative $D_x\mathcal{B}$ at x induces a linear function $D_x\mathcal{B}(x) : I\!R^n \to \mathcal{S}^m$, which is given by

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

In particular,

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

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

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

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 $x \in I\!R^n$. Its gradient at x is given by

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

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 $\mathcal{A} : I\!R^n \to \mathcal{S}^m$, in view of (1)–(5), we have

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

3 Linear semidefinite programs

In this section, we briefly review the case of linear semidefinite programs. We remark that applications of linear semidefinite programs 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, 4, 8, 23, 25] and the references given there.

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

$$\begin{array}{lll} \text{maximize} \quad C \bullet Y \quad \text{subject to} & Y \in \mathcal{S}^m,\ Y \succeq 0, \\ & \mathcal{A}^*(Y) + b = 0, \end{array} \tag{7}$$

and

$$\begin{array}{lll} \text{minimize} \quad b^T x \quad \text{subject to} & x \in I\!R^n, \\ & \mathcal{A}(x) + C \preceq 0. \end{array} \tag{8}$$

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 (8) to nonlinear semidefinite programs in standard form.

If there exists a matrix $Y \succ 0$ that is feasible for (7), then we call Y strictly feasible for (7) and say that (7) satisfies Slater's condition. Likewise, if there exists a vector x such that $\mathcal{A}(x) + C \prec 0$, then we call (8) strictly feasible and say that (8) satisfies Slater's condition.

The following optimality conditions for linear semidefinite programs are well known; see, e.g., [20]. If problem (7) or (8) satisfies Slater's condition, then the optimal values of (7) and (8) 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

$$YS = 0, \quad \text{where} \quad S := -C - \mathcal{A}(x). \tag{9}$$

Conversely, if Y and x are feasible points for (7) and (8), respectively, and satisfy the complementarity condition (9), then $Y^{\rm opt} := Y$ is an optimal solution of (7) and $x^{\rm opt} := x$ is an optimal solution of (8).

These optimality conditions can be summarized as follows. If problem (8) satisfies Slater's condition, then for a point $x \in I\!R^n$ to be an optimal solution of (8) it is necessary and sufficient that there exist matrices $Y \succeq 0$ and $S \succeq 0$ such that

$$\begin{aligned} \mathcal{A}(x) + C + S &= 0, \\ \mathcal{A}^*(Y) + b &= 0, \\ YS &= 0. \end{aligned} \tag{10}$$

Note that, in view of (6), the second equation in (10) can also be written in the form

$$\nabla_x \big((\mathcal{A}(x) + C) \bullet Y \big) + b = \mathcal{A}^*(Y) + b = 0. \tag{11}$$

4 Nonlinear semidefinite programs

In this section, we consider nonlinear semidefinite programs, which are extensions of the dual linear semidefinite programs (8).

Given a vector $b \in I\!R^n$, a matrix-valued function $\mathcal{B} : I\!R^n \to \mathcal{S}^m$, and two vector-valued functions $c : I\!R^n \to I\!R^p$ and $d : I\!R^n \to I\!R^q$, we consider problems of the following form:

$$\begin{array}{lll} \text{minimize} \quad b^T x \quad \text{subject to} & x \in I\!R^n, \\ & \mathcal{B}(x) \preceq 0, \\ & c(x) \leq 0, \\ & d(x) = 0. \end{array} \tag{12}$$

Here, the functions $\mathcal{B}$, c, and d are nonlinear in general, and thus (12) represents a class of nonlinear semidefinite programs. We assume that the functions $\mathcal{B}$, c, and d are at least $\mathcal{C}^2$-differentiable. We stress that the choice of the linear objective function $b^T x$ in (12) 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 (12) in this paper can be modified so that they apply to nonlinear objective functions.

Note that the class (12) reduces to general nonlinear programs if $m = 0$, and to linear semidefinite programs of the form (8) if $\mathcal{B}$ is an affine function and $p = q = 0$.

The Lagrangian $\mathcal{L} : I\!R^n \times \mathcal{S}^m \times I\!R^p \times I\!R^q \to I\!R$ of (12) is defined as follows:

$$\mathcal{L}(x, Y, u, v) := b^T x + \mathcal{B}(x) \bullet Y + u^T c(x) + v^T d(x). \tag{13}$$

Its gradient with respect to x is given by

$$g(x,Y,u,v) := \nabla_x \mathcal{L}(x,Y,u,v) = b + \nabla_x \left(\mathcal{B}(x) \bullet Y\right) + \nabla_x c(x)\, u + \nabla_x d(x)\, v \tag{14}$$

and its Hessian by

$$H(x,Y,u,v) := \nabla_x^2 \mathcal{L}(x,Y,u,v) = \nabla_x^2 \left(\mathcal{B}(x) \bullet Y\right) + \sum_{i=1}^{p} u_i \nabla_x^2 c_i(x) + \sum_{j=1}^{q} v_j \nabla_x^2 d_j(x). \tag{15}$$

Note that in (14), 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 $\nabla_x d(x) := (D_x d(x))^T$.

If the problem (12) is convex and satisfies Slater's condition [13], then for each optimal solution x of (12) there exists an $m \times m$ matrix $Y \succeq 0$ and vectors $u \in I\!R^p$, $u \geq 0$, and $v \in I\!R^q$ such that the quadruple (x, Y, u, v) is a saddle point of the Lagrangian (13), $\mathcal{L}$.

More generally, for nonconvex problems (12), let $x \in I\!R^n$ be a feasible point of (12), and assume that the Robinson or Mangasarian-Fromovitz constraint qualification [13, 18, 19] is satisfied at x, i.e., the matrix $D_x d(x)$ has full rank and there exists a vector $\Delta x \neq 0$ such that $\mathcal{B}(x) + D_x \mathcal{B}(x)[\Delta x] \prec 0$, $c(x) + D_x c(x) \Delta x < 0$, and $D_x d(x) \Delta x = 0$. Then, if x is a local minimizer of (12), the first-order optimality condition is satisfied, i.e., there exist matrices Y, $S \in \mathcal{S}^m$ and vectors u, $s \in I\!R^p$, $v \in I\!R^q$ such that

$$\begin{aligned}
\mathcal{B}(x) + S &= 0, \\
c(x) + s &= 0, \\
d(x) &= 0, \\
g(x,Y,u,v) &= 0, \\
Y S &= 0, \\
\mathrm{Diag}(u)\,\mathrm{Diag}(s) &= 0, \\
Y,\ S &\succeq 0, \\
u,\ s &\geq 0.
\end{aligned} \tag{16}$$

We remark that in the case $m = 0$, the system (16) reduces to the standard first-order optimality conditions [13, 18] for smooth nonlinear programming. In the case $p = q = 0$, the system (16) is a straightforward generalization of the optimality conditions (10) and (11), with the affine function $\mathcal{A}(x) + C$ in (11) replaced by the nonlinear function $\mathcal{B}(x)$.

Primal-dual interior-point methods for solving (12) 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,

$$\begin{aligned}
\mathcal{B}(x) + S &= 0, \\
c(x) + s &= 0, \\
d(x) &= 0, \\
g(x,Y,u,v) &= 0, \\
Y S &= \mu_k I_m, \\
\mathrm{Diag}(u)\,\mathrm{Diag}(s) &= \mu_k I_p, \\
Y,\ S &\succ 0, \\
u,\ s &> 0,
\end{aligned} \tag{17}$$

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

The solutions of (17) coincide with the solutions of the standard logarithmic-barrier problems for (12). Moreover, the logarithmic-barrier approach for solving (12) can be interpreted as a certain choice of the 'symmetrization operator' for the equation, $YS = \mu_k I_m$, in the fifth row of (17); see Section 5 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 (12). 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 (12). This observation motivated the SSP method described in the next section.

5 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 (12). For an overview of SQP methods for standard nonlinear programs, we refer the reader to [2] 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 (16) at the current iterate. More precisely, let (x^k, Y^k, u^k, v^k) denote the current point at the beginning of the k-th iteration. One then determines corrections $(\Delta x, \Delta Y, \Delta u, \Delta v)$ and variables S and s such that

$$\begin{aligned}
\mathcal{B}(x^k) + D_x\mathcal{B}(x^k)[\Delta x] + S &= 0,\\
c(x^k) + D_x c(x^k)\,\Delta x + s &= 0,\\
d(x^k) + D_x d(x^k)\,\Delta x &= 0,\\
b + H^k\Delta x + \nabla_x\big(\mathcal{B}(x^k) \bullet (Y^k + \Delta Y)\big) + \nabla_x c(x^k)\,(u^k + \Delta u) + \nabla_x d(x^k)\,(v^k + \Delta v) &= 0,\\
(Y^k + \Delta Y^k)S &= 0,\\
\mathrm{Diag}(u^k + \Delta u)\,\mathrm{Diag}(s) &= 0,\\
Y^k + \Delta Y,\ S &\succeq 0,\\
u^k + \Delta u,\ s &\geq 0.
\end{aligned} \tag{18}$$

Here and in the sequel, we use the notation

$$H^k := H\big(x^k, Y^k, u^k, v^k\big). \tag{19}$$

Recall from (14) and (15) that $g(x, Y, u, v)$ and $H(x, Y, u, v)$ denote the gradient and Hessian with respect to x, respectively, of the Lagrangian (13), $\mathcal{L}(x, Y, u, v)$, of the nonlinear semidefinite program (12). Moreover, from (14) it follows that the linearization of $g(x, Y, u.v)$ at the point (x^k, Y^k, u^k, v^k) is given by

$$\begin{aligned}
g\big(x^k + \Delta x, Y^k + \Delta Y, u^k + \Delta u, v^k + \Delta v\big) \approx\ & b + H^k\Delta x + \nabla_x\big(\mathcal{B}(x^k) \bullet (Y^k + \Delta Y)\big)\\
& + \nabla_x c(x^k)\,(u^k + \Delta u) + \nabla_x d(x^k)\,(v^k + \Delta v).
\end{aligned}$$

Thus the fourth equation in (18) is just the linearization of the fourth equation in (16). Furthermore, the first three equations of (18) are straightforward linearizations of the first three equations (16). These linearizations are used in the same way in primal-dual interior methods.

The last four rows in (18) and (16) are identical when Y and u in (16) are rewritten as $Y = Y^k + \Delta Y$ and $u = u^k + \Delta u$, respectively. In analogy to SQP methods for standard nonlinear programs, the problem of how to guarantee the nonnegativity constraints, namely $\mathcal{B}(x) \preceq 0$ and $c(x) \leq 0$ in (12), is thus shifted to the subproblem (18). If the iterates x^k generated from (18) converge, then their limit x automatically satisfies $\mathcal{B}(x) \preceq 0$ and $c(x) \leq 0$.

In contrast, interior methods use perturbations, symmetrizations, and linearizations for the last four rows in (16), resulting in cheaper linear subproblems that are typically less 'powerful' than the subproblems (18). An important aspect for interior methods for linear SDPs is the choice of the symmetrization procedure for the bilinear equation $YS = \mu_k I_m$; see, e.g., [14]. For convex SDPs, theoretical convergence analyses are well developed and also supported by numerical evidence of rapid convergence. However, the generalization of these convergence results to nonlinear nonconvex subproblems is far from obvious. The proposed SSP method allows to apply the symmetrization to linear SDPs, thus reducing this aspect to a well-studied topic.

In summary, both the problem of choosing a suitable symmetrization scheme and the problem of how to guarantee the nonnegativity constraints are shifted to the subproblem (18).

Note that the conditions (18) are the optimality conditions for the problem

$$\begin{array}{lll} \text{minimize} \quad b^T \Delta x + \frac{1}{2}(\Delta x)^T H^k \Delta x \quad \text{subject to} & \Delta x \in I\!R^n, \\ & \mathcal{B}(x^k) + D_x \mathcal{B}(x^k)[\Delta x] \preceq 0, \\ & c(x^k) + D_x c(x^k)\, \Delta x \leq 0, \\ & d(x^k) + D_x d(x^k)\, \Delta x = 0. \end{array} \tag{20}$$

Recall from (19) that H^k is the Hessian of the Lagrangian (13), $\mathcal{L}$, at the k-th iterate. The conditions (18) and (20) have been considered in [19, Equations (2.1) and (2.2)], with the remark that they have "been found to be an appropriate approximation of" (12) "for numerical purposes".

In order to be able to solve the subproblem (18) efficiently, in practice, one replaces the matrix H^k in (18), respectively (20), 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 (13), $\mathcal{L}$, can be used to approximate H^k by some positive semidefinite matrix $\hat{H}^k$. Given an estimate $\hat{H}^k$ of H^k for the current, k-th, SSP iteration, the quasi-Newton condition to generate a BFGS update $\hat{H}^{k+1}$ approximating the matrix H^{k+1} for the next, $(k+1)$-st, SSP iteration can be derived as follows:

$$\begin{aligned} \hat{H}^{k+1} \Delta x &= \nabla_x^2 \mathcal{L}\big(x^{k+1}, Y^{k+1}, u^{k+1}, v^{k+1}\big)\big(x^{k+1} - x^k\big) \\ &\approx \nabla_x \mathcal{L}\big(x^{k+1}, Y^{k+1}, u^{k+1}, v^{k+1}\big) - \nabla_x \mathcal{L}\big(x^k, Y^{k+1}, u^{k+1}, v^{k+1}\big) \\ &= \nabla_x \big(\mathcal{B}(x^{k+1}) \bullet Y^{k+1}\big) - \nabla_x \big(\mathcal{B}(x^k) \bullet Y^{k+1}\big) \\ &\quad + \big(\nabla_x c(x^{k+1}) - \nabla_x c(x^k)\big) u^{k+1} + \big(\nabla_x d(x^{k+1}) - \nabla_x d(x^k)\big) v^{k+1}. \end{aligned} \tag{21}$$

If $\hat{H}^k$ is positive semidefinite, the BFGS update with the above condition can be suitably damped such that $\hat{H}^{k+1}$ is also positive semidefinite. At each iteration of the SSP method, problem (20) is solved with H^k replaced by the matrix $\hat{H}^{k+1}$ that is obtained by the BFGS

update of $\hat{H}^k$ from the previous SSP iteration. If $\hat{H}^{k+1}$ is positive semidefinite, problem (20) 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. We stress that the formulation as a second-order-cone constraint is much more efficient, and for example, can be specified as input for the program package SEDUMI [21] in order to solve (20).

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 (20) become infeasible. In this case, the right-hand sides, "0", of the first three rows in (18) are replaced by weaker, penalized right-hand sides. Moreover, the convergence analysis of the method proposed in [5, 7] 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 the usual formulation of the optimality conditions is an overdetermined system of nonlinear equations. The paper [7] presents a proof that is based on a new approach by Bonnans et al. [3] and uses some general results due to Robinson [19]. It does not require strict complementarity and allows for approximate Hessian matrices in (21).

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.

6 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. More precisely, we show that the strictly complementary solutions of such problems depend smoothly on the problem data.

First, we consider quadratic semidefinite programs of the form

$$\begin{array}{lll} \text{minimize} \quad f(x) \quad \text{subject to} & x \in I\!R^n, \\ & \mathcal{A}(x) + C \preceq 0. \end{array} \tag{22}$$

Here, $\mathcal{A} : I\!R^n \to \mathcal{S}^m$ is a linear function, $C \in \mathcal{S}^m$, and $f : I\!R^n \to I\!R$ is a quadratic function defined by $f(x) = b^T x + \frac{1}{2} x^T H x$, where $b \in I\!R^n$ and $H \in \mathcal{S}^n$. Note that we make no further assumptions on the matrix H. Thus, problem (22) is a general, possibly nonconvex, quadratic semidefinite program.

The problem (22) is described by the data

$$\mathcal{D} := [\mathcal{A}, b, C, H]. \tag{23}$$

In Theorem 1 below, we present a sensitivity result for the solutions (22) when the data $\mathcal{D}$ is changed to $\mathcal{D} + \Delta\mathcal{D}$ where

$$\Delta\mathcal{D} := [\Delta\mathcal{A}, \Delta b, \Delta C, \Delta H] \tag{24}$$

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 (23) and perturbations (24). 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 the quadratic semidefinite problem (22). 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 (22) is said to satisfy Slater's condition if there exists a vector $x \in I\!R^n$ with $\mathcal{A}(x) + C \prec 0$. Moreover, the triple $(\bar{x}, \bar{Y}, \bar{S})$, where $\bar{x} \in I\!R^n$ and $\bar{Y}$, $\bar{S} \in \mathcal{S}^m$, is called a *stationary point* of (22) if

$$\begin{aligned} \mathcal{A}(\bar{x}) + C + \bar{S} &= 0, \\ b + H\bar{x} + \mathcal{A}^*(\bar{Y}) &= 0, \\ \bar{Y}\,\bar{S} &= 0, \\ \bar{Y},\ \bar{S} &\succeq 0. \end{aligned} \tag{25}$$

If in addition to (25), one has

$$\bar{Y} + \bar{S} \succ 0, \tag{26}$$

then $(\bar{x}, \bar{Y}, \bar{S})$ is said to be a *strictly complementary* stationary point of (22).

Let $\bar{x} \in I\!R^n$ be a feasible point of (22). We say that $h \in I\!R^n$, $h \neq 0$, is a *feasible direction* at $\bar{x}$ if $x = \bar{x} + \epsilon h$ is a feasible point of (22) for all sufficiently small $\epsilon > 0$. Following [19, Definition 2.1], we say that the *second-order sufficient condition* holds at $\bar{x}$ with modulus $\mu > 0$ if for all feasible directions $h \in I\!R^n$ at $\bar{x}$ with $h^T(b + H\bar{x}) = h^T \nabla_x f(\bar{x}) = 0$ one has

$$h^T H h = h^T \big(\nabla_x^2 \mathcal{L}^{(\mathrm{q})}(\bar{x}, \bar{Y})\big) h \geq \mu \|h\|^2. \tag{27}$$

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

Theorem 1 *Assume that problem* (22) *satisfies Slater's condition. Let* $(\bar{x}, \bar{Y}, \bar{S})$ *be a locally unique and strictly complementary stationary point of* (22) *with data* (23), $\mathcal{D}$, *and assume that the second-order sufficient condition holds at* $\bar{x}$ *with modulus* $\mu > 0$. *Then, for all sufficiently small perturbations* (24), $\Delta\mathcal{D}$, *there exists a locally unique stationary point* $\big(\bar{x}(\Delta\mathcal{D}), \bar{Y}(\Delta\mathcal{D}), \bar{S}(\Delta\mathcal{D})\big)$ *of the perturbed program* (22) *with data* $\mathcal{D} + \Delta\mathcal{D}$. *Moreover, the point* $\big(\bar{x}(\Delta\mathcal{D}), \bar{Y}(\Delta\mathcal{D}), \bar{S}(\Delta\mathcal{D})\big)$ *is a differentiable function of the perturbation* (24), *and for* $\Delta\mathcal{D} = 0$, $\big(\bar{x}(0), \bar{Y}(0), \bar{S}(0)\big) = (\bar{x}, \bar{Y}, \bar{S})$. *The derivative* $D_{\mathcal{D}}\big(\bar{x}(0), \bar{Y}(0), \bar{S}(0)\big)$ *at* $(\bar{x}, \bar{Y}, \bar{S})$ *is characterized by the directional derivatives*

$$(\dot{x}, \dot{Y}, \dot{S}) := D_{\mathcal{D}}\big(\bar{x}(0), \bar{Y}(0), \bar{S}(0)\big)[\Delta\mathcal{D}]$$

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

$$\begin{aligned} \mathcal{A}(\dot{x}) + \dot{S} &= -\Delta C - \Delta\mathcal{A}(\bar{x}), \\ H\dot{x} + \mathcal{A}^*(\dot{Y}) &= -\Delta b - \Delta H \bar{x} - \Delta\mathcal{A}^*(\bar{Y}), \\ \dot{Y}\,\bar{S} + \bar{Y}\,\dot{S} &= 0, \end{aligned} \tag{28}$$

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 [9]. A related sensitivity result for linear semidefinite programs for a more restricted class of perturbations, but also under weaker assumptions, is given in [22]. A local Lipschitz continuity property of unique and strictly complementary solutions of linear semidefinite programs is derived in [15].

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., [9]. Moreover, our assumptions on the stationary point $(\bar{x}, \bar{Y}, \bar{S})$ imply that $\bar{x}$ is a strict local minimizer of (22).

Proof. The proof of Theorem 1 is divided into five steps. As a first 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 the solution of (28). In a second step, we prove that the overdetermined system (28) is equivalent to a square system (i.e., a system with the same number of variables and equations), and that this square system has a unique solution. In a third step, we conclude that a nonlinear square system derived from the overdetermined nonlinear system (25) has a local solution that depends smoothly on the perturbation $\Delta\mathcal{D}$. In a fourth step, we show that this solution also satisfies (25), and in a fifth step, we prove that the second-order sufficient condition is satisfied at the perturbed solution.

Step 1. 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

$$\begin{aligned} (\mathcal{A} + \Delta\mathcal{A})(\bar{x} + \Delta x) + C + \Delta C + \bar{S} + \Delta S &= 0, \\ b + \Delta b + (H + \Delta H)(\bar{x} + \Delta x) + (\mathcal{A}^* + \Delta\mathcal{A}^*)(\bar{Y} + \Delta Y) &= 0, \\ (\bar{Y} + \Delta Y)(\bar{S} + \Delta S) &= 0. \end{aligned} \tag{29}$$

Subtracting from these equations the first three equations of (25) yields

$$\begin{aligned} (\mathcal{A} + \Delta\mathcal{A})(\Delta x) + \Delta S &= -\Delta C - \Delta\mathcal{A}(\bar{x}), \\ (H + \Delta H)\Delta x + (\mathcal{A}^* + \Delta\mathcal{A}^*)(\Delta Y) &= -\Delta b - \Delta H\bar{x} - \Delta\mathcal{A}^*(\bar{Y}), \\ \Delta Y\,\bar{S} + \bar{Y}\Delta S &= -\Delta Y \Delta S. \end{aligned} \tag{30}$$

Neglecting the second-order terms in (30), we obtain the result claimed in (28). It still remains to verify the existence and differentiability of Δx, ΔY, ΔS.

Step 2. In this step, we prove that the system (28) has a unique solution. To this end, we show that the homogeneous version of (28), i.e., the system

$$\begin{aligned} \mathcal{A}(\dot{x}) + \dot{S} &= 0, \\ H\dot{x} + \mathcal{A}^*(\dot{Y}) &= 0, \\ \dot{Y}\,\bar{S} + \bar{Y}\,\dot{S} &= 0, \end{aligned} \tag{31}$$

for the unknowns $\dot{x} \in \mathbb{R}^n$, $\dot{Y},\ \dot{S} \in \mathcal{S}^m$, only has the solution $\dot{x} = 0$, $\dot{Y} = \dot{S} = 0$.

First, recall that, by (25) and (26), the matrices $\bar{Y}$ and $\bar{S}$ satisfy

$$\bar{Y}\,\bar{S} = 0 \prec \bar{Y} + \bar{S}, \quad \bar{Y},\ \bar{S} \succeq 0. \tag{32}$$

In particular, the real symmetric $m \times m$ matrices $\bar{Y}$ and $\bar{S}$ commute, and thus they share a common set of real orthonormal eigenvectors. Let U be a real orthogonal $m \times m$ matrix whose columns are such common eigenvectors. Then by construction, both matrices $U^T \bar{Y} U$ and $U^T \bar{S} U$ are diagonal, and the diagonal entries are the eigenvalues of $\bar{Y}$ and $\bar{S}$, respectively. Moreover, the properties (32) imply that there exists an integer $0 \le k \le m$ such that exactly k of the diagonal entries of $U^T \bar{Y} U$ are positive, exactly $m-k$ of the diagonal entries of $U^T \bar{S} U$ are positive, and the two associated subsets of indices are disjoint. Thus, by ordering the columns of U properly, we may assume that the leading k diagonal entries of $U^T \bar{Y} U$ are positive and the trailing $m-k$ diagonal entries of $U^T \bar{S} U$ are positive. We denote by $\bar{Y}_1 \succ 0$ the leading $k \times k$ diagonal matrix of $U^T \bar{Y} U$, and by $\bar{S}_2 \succ 0$ the trailing $(m-k) \times (m-k)$ diagonal matrix of $U^T \bar{S} U$. Without loss of generality, we now assume that the matrices $\bar{Y}$ and $\bar{S}$ in (31) are already in such diagonal form:

$$\bar{Y} = \begin{bmatrix} \bar{Y}_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \bar{S} = \begin{bmatrix} 0 & 0 \\ 0 & \bar{S}_2 \end{bmatrix}, \quad \text{where} \quad \bar{Y}_1,\ \bar{S}_2 \succ 0 \quad \text{and} \quad \bar{Y}_1,\ \bar{S}_2 \quad \text{are diagonal.} \tag{33}$$

Indeed, this can be done by replacing, in (31), $\bar{Y}$, $\bar{S}$, $\dot{Y}$, $\dot{S}$, $\mathcal{A}(x)$ 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, and then multiplying the first and third rows from the left by U and from the right by U^T.

Let

$$\dot{Y} = \begin{bmatrix} \dot{Y}_1 & \dot{Y}_3 \\ \dot{Y}_3^T & \dot{Y}_2 \end{bmatrix} \quad \text{and} \quad \dot{S} = \begin{bmatrix} \dot{S}_1 & \dot{S}_3 \\ \dot{S}_3^T & \dot{S}_2 \end{bmatrix} \tag{34}$$

be partitionings of the symmetric matrices $\dot{Y}$ and $\dot{S}$ with blocks of the same sizes as in the partitionings of $\bar{Y}$ and $\bar{S}$ in (33). Then, by (33) and (34), we have

$$\dot{Y}\,\bar{S} + \bar{Y}\,\dot{S} = \begin{bmatrix} \bar{Y}_1\,\dot{S}_1 & \dot{Y}_3\,\bar{S}_2 + \bar{Y}_1\,\dot{S}_3 \\ 0 & \dot{Y}_2\,\bar{S}_2 \end{bmatrix}. \tag{35}$$

Since the matrices $\dot{S}_1$, $\dot{Y}_2$ are symmetric and the matrices $\bar{Y}_1$, $\bar{S}_2$ are diagonal, the relation (35) implies that the solution set of the system (31) does not change when the third row in (31) is replaced by $\Pi_{\mathrm{up}}\big(\dot{Y}\,\bar{S} + \bar{Y}\,\dot{S}\big) = 0$. Here, $\Pi_{\mathrm{up}}(Z)$ denotes the upper triangular part of an $m \times m$ matrix Z. We have thus shown that the homogeneous version of the system (28) is equivalent to the system

$$\begin{aligned} \mathcal{A}(\dot{x}) + \dot{S} &= 0, \\ H\dot{x} + \mathcal{A}^*(\dot{Y}) &= 0, \\ \Pi_{\mathrm{up}}\big(\dot{Y}\,\bar{S} + \bar{Y}\,\dot{S}\big) &= 0. \end{aligned} \tag{36}$$

Note that (36) is a square system of $n + m(m+1)$ linear equations for $n + m(m+1)$ unknowns, namely the entries of the vector $\dot{x}$ and of the symmetric matrices $\dot{Y}$ and $\dot{S}$.

Let $\dot{x} \in I\!R^n$, $\dot{Y}$, $\dot{S} \in \mathcal{S}^m$ be any solution of (36), and assume that the matrices $\dot{Y}$ and $\dot{S}$ are partitioned as in (34). We need to show that $\dot{x} = 0$ and $\dot{Y} = \dot{S} = 0$. For later use, we note that, by (33), (35), and the third row of (36),

$$\dot{S}_1 = 0, \quad \dot{Y}_2 = 0, \quad \dot{Y}_3\bar{S}_2 + \bar{Y}_1\dot{S}_3 = 0. \tag{37}$$

Next, we establish the inequality

$$\dot{x}^T H \dot{x} \geq \mu \|\dot{x}\|^2, \tag{38}$$

where $\mu > 0$ is the modulus of the second-order sufficient condition (27).

Assume that $\dot{x} \neq 0$. Let $\breve{x} \in I\!R^n$ be a Slater point for problem (22). This guarantees that

$$M = \begin{bmatrix} M_1 & M_3 \\ M_3^T & M_2 \end{bmatrix} := -\big(\mathcal{A}(\breve{x}) + C\big) \succ 0, \tag{39}$$

where the block partitioning M is conforming with (34). For $\eta > 0$, set

$$h_\eta^+ := \dot{x} + \eta(\breve{x} - \bar{x}) \quad \text{and} \quad h_\eta^- := -\dot{x} + \eta(\breve{x} - \bar{x}). \tag{40}$$

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 (22) 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 (40), (33), (34), (39), the first row of (25), the first row of (36), and the first relation of (37), one readily verifies that

$$\begin{aligned} \mathcal{A}(\bar{x} + \epsilon h_\eta^\pm) + C &= (1 - \epsilon\eta)\big(\mathcal{A}(\bar{x}) + C\big) + \epsilon\eta\big(\mathcal{A}(\breve{x}) + C\big) \pm \epsilon\mathcal{A}(\dot{x}) \\ &= -\left(\begin{bmatrix} 0 & 0 \\ 0 & \bar{S}_2 \end{bmatrix} + \epsilon \begin{bmatrix} \eta M_1 & \eta M_3 \pm \dot{S}_3 \\ (\eta M_3 \pm \dot{S}_3)^T & \eta(M_2 - \bar{S}_2) \pm \dot{S}_2 \end{bmatrix} \right). \end{aligned} \tag{41}$$

Recall that $\eta > 0$ is fixed. Since, by (33) and (39), $\bar{S}_2 \succ 0$ and $M_1 \succ 0$, a standard Schur-complement argument shows that the matrix on the right-hand side of (41) is negative definite for all sufficiently small $\epsilon > 0$. Thus the vectors (40) are feasible directions for (22) at $\bar{x}$ for any $\eta > 0$.

This in turn implies

$$\dot{x}^T(b + H\bar{x}) = \dot{x}^T \nabla_x f(\bar{x}) = 0. \tag{42}$$

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 $\big(h_\eta^+\big)^T \nabla_x f(\bar{x}) < 0$, and thus h_η^+ is a descent direction for the objective function f of (22) 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 (22) at the point $\bar{x}$, leading to the same contradiction.

The second-order sufficient condition (27) also holds true on the closure of the feasible directions. Since $\dot{x}$ is the limit of the feasible directions (40) for $\eta \to 0$ and $\dot{x}$ satisfies (42), the inequality (38) follows from (27).

Next recall from (33) that $\bar{Y}_1$ and $\bar{S}_2$ are positive definite diagonal matrices. The third relation in (37) 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 (34), the first two relations in (37), the first two rows of (36), and (38), one readily verifies that

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

Since $\mu > 0$, this implies

$$\dot{x} = 0 \quad \text{and} \quad \dot{Y}_3 = \dot{S}_3 = 0. \tag{43}$$

By the first row of (36), it further follows that

$$\dot{S} = -\mathcal{A}(\dot{x}) = -\mathcal{A}(0) = 0. \tag{44}$$

Thus it only remains to show that $\dot{Y} = 0$. In view of (34), (37), and (44), we have

$$\dot{Y} = \begin{bmatrix} \dot{Y}_1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{45}$$

Now suppose that $\dot{Y}_1 \neq 0$. Then, by (33) and (45), 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 (25), (36), and (44), one readily verifies that the point $(\bar{x}, \bar{Y}_\epsilon, \bar{S})$ also satisfies (25) 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 (45), $\dot{Y} = 0$.

This concludes the proof that the square system (36) is nonsingular. From the equivalence of (28) and (36), it follows that the overdetermined system (28) always has a unique solution.

Step 3. The result shown in Step 2 above also implies that the system,

$$\begin{aligned} \mathcal{A}(\dot{x}) + \dot{S} &= -\Delta C - \Delta\mathcal{A}(\bar{x}), \\ H\dot{x} + \mathcal{A}^*(\dot{Y}) &= -\Delta b - \Delta H \bar{x} - \Delta\mathcal{A}^*(\bar{Y}), \\ \Pi_{\rm up}\big(\dot{Y}\,\bar{S} + \bar{Y}\,\dot{S}\big) &= 0, \end{aligned} \tag{46}$$

has a unique solution. Note that (46) is obtained by replacing the last row of the system (28) by the equation $\Pi_{\rm up}\big(\dot{Y}\,\bar{S} + \bar{Y}\,\dot{S}\big) = 0$. We can now apply the implicit function theorem to the system

$$\mathcal{A}(x) + C + S = 0, \quad Hx + b + \mathcal{A}^*(Y) = 0, \quad \Pi_{\rm up}\big(Y\,S\big) = 0. \tag{47}$$

As we have just seen, the linearization of (47) at the point $(\bar{x}, \bar{Y}, \bar{S})$ is nonsingular, and hence (47) has a differentiable and locally unique solution.

Step 4. Let $(\bar{x} + \Delta x, \bar{Y} + \Delta Y, \bar{S} + \Delta S)$ be such a solution of (47). Recall from (33) that, without loss of generality, the matrices $\bar{Y}$ and $\bar{S}$ were assumed to be diagonal. In particular, both matrices have the identity matrix I_m as their eigenvector matrix. Moreover, in view of (33), $\bar{Y}$ has k positive eigenvalues, namely the diagonal entries of $\bar{Y}_1$, and the eigenvalue 0 with multiplicity $m - k$. It follows that, for sufficiently small $\|\Delta Y\|$, the eigenvalues of the perturbed matrix $\bar{Y} + \Delta Y \in \mathcal{S}^m$ consist of two disjoint sets Λ_1 and Λ_2, where the eigenvalues in

Λ_1 are perturbations of the diagonal entries of $\bar{Y}_1$ and the eigenvalues in Λ_2 are perturbations of the eigenvalue 0. A standard result (see, e.g., [10, Theorem 8.1.10]) on the sensitivity to perturbations of invariant subspaces of symmetric matrices then guarantees the existence of a real orthogonal $m \times m$ matrix $\hat{U} = \hat{U}(\Delta Y)$ such that

$$\hat{U}^T(\bar{Y} + \Delta Y)\hat{U} = \begin{bmatrix} \bar{Y}_1 + \Delta\hat{Y}_1 & 0 \\ 0 & \Delta\hat{Y}_2 \end{bmatrix}, \quad \Delta\hat{Y}_1 \in \mathcal{S}^{k\times k}, \quad \Delta\hat{Y}_2 \in \mathcal{S}^{(m-k)\times(m-k)}, \tag{48}$$

and $\|\hat{U} - I_m\|$, $\|\Delta\hat{Y}_1\|$, and $\|\Delta\hat{Y}_2\|$ are all small if $\|\Delta Y\|$ is small. We stress that the matrices $\Delta\hat{Y}_1$ and $\Delta\hat{Y}_2$ are not necessarily diagonal. For later use, we set

$$\hat{U}^T(\bar{S} + \Delta S)\hat{U} =: \begin{bmatrix} \Delta\hat{S}_1 & \Delta\hat{S}_3 \\ (\Delta\hat{S}_3)^T & \bar{S}_2 + \Delta\hat{S}_2 \end{bmatrix}, \tag{49}$$

where the block partitioning is the same as in (48) and $\bar{S}_2 \succ 0$ is the diagonal matrix from (33). Note that $\|\Delta\hat{S}_1\|$, $\|\Delta\hat{S}_2\|$, and $\|\Delta\hat{S}_3\|$ are all small if $\|\Delta S\|$ is small. In particular, since $\|\hat{U} - I_m\|$ is small, the system (46) with the last row replaced by $\Pi_{\mathrm{up}}\big(\hat{U}^T(\dot{Y}\,\bar{S} + \bar{Y}\,\dot{S})\hat{U}\big) = 0$ is again nonsingular, and therefore the solution of (47) with the last relation $\Pi_{\mathrm{up}}\,(Y\,S) = 0$ replaced by

$$\Pi_{\mathrm{up}}\big(\hat{U}^T\,Y\,S\,\hat{U}\big) = 0 \tag{50}$$

is locally unique. The solution of the original system (47) and the solution obtained with $\hat{U}$ are analytic functions [11, Theorem 2.1.2], and all derivatives at $(\bar{x}, \bar{Y}, \bar{S})$ can be obtained by differentiating the system (46). Since the original system (46) and the one with $\hat{U}$ coincide with each other (both coincide with the complete system (28)), it follows that all derivatives coincide with each other. By the identity principle for analytic functions, both solutions coincide.

In particular, $Y = \bar{Y} + \Delta Y$ and $S = \bar{S} + \Delta S$ satisfy (50). Thus, inserting Y and S into (50) and using (48) and (49), we obtain

$$\Pi_{\mathrm{up}}\left(\begin{bmatrix} \bar{Y}_1 + \Delta\hat{Y}_1 & 0 \\ 0 & \Delta\hat{Y}_2 \end{bmatrix}\begin{bmatrix} \Delta\hat{S}_1 & \Delta\hat{S}_3 \\ (\Delta\hat{S}_3)^T & \bar{S}_2 + \Delta\hat{S}_2 \end{bmatrix}\right) = 0.$$

This equation is equivalent to

$$\Pi_{\mathrm{up}}\big((\bar{Y}_1 + \Delta\hat{Y}_1)\Delta\hat{S}_1\big) = 0, \quad (\bar{Y}_1 + \Delta\hat{Y}_1)\Delta\hat{S}_3 = 0, \quad \Pi_{\mathrm{up}}\big(\Delta\hat{Y}_2(\bar{S}_2 + \Delta\hat{S}_2)\big) = 0. \tag{51}$$

Since, by (33), the matrices $\bar{Y}_1$ and $\bar{S}_2$ are fixed positive definite diagonal matrices and all the other blocks in (51) are small if $\|\Delta Y\|$ and $\|\Delta S\|$ are small, the relations (51) imply that

$$\Delta\hat{S}_1 = 0, \quad \Delta\hat{S}_3 = 0, \quad \Delta\hat{Y}_2 = 0. \tag{52}$$

Using (48), (49), and (52), it follows that $Y = \bar{Y} + \Delta Y$ and $S = \bar{S} + \Delta S$ satisfy

$$Y\,S = \hat{U}\begin{bmatrix} \bar{Y}_1 + \Delta\hat{Y}_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & \bar{S}_2 + \Delta\hat{S}_2 \end{bmatrix}\hat{U}^T = 0, \quad , Y,\, S \succeq 0, \quad Y + S \succ 0,$$

provided $\|\Delta Y\|$ and $\|\Delta S\|$ are sufficiently small. However, this means that $\bar{Y} + \Delta Y$ and $\bar{S} + \Delta S$ are again strictly complementary if $\|\Delta Y\|$ and $\|\Delta S\|$ are sufficiently small. In particular,

$(\bar{x}+\Delta x, \bar{Y}+\Delta Y, \bar{S}+\Delta S)$ is not only a solution of (47) but also of the overdetermined nonlinear system (29). Hence, $\big(\bar{x}(\Delta\mathcal{D}), \bar{Y}(\Delta\mathcal{D}), \bar{S}(\Delta\mathcal{D})\big) := (\bar{x}+\Delta x, \bar{Y}+\Delta Y, \bar{S}+\Delta S)$ is a locally unique stationary point of the perturbed program (22) with data $\mathcal{D}+\Delta\mathcal{D}$.

Step 5. 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 (22), the second-order sufficient condition at $\bar{x}$ then states that $h^T H h \geq \mu$ for all $h \in I\!R^n$ 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

$$h^T(H+\Delta H)h \geq \tilde{\mu} \tag{53}$$

for all of solutions $h \in I\!R^n$ 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 (29), the above is equivalent to

$$\epsilon(\mathcal{A}+\Delta\mathcal{A})(h) \preceq \bar{S}(\Delta\mathcal{D}), \quad \epsilon = \epsilon(h) > 0, \quad \|h\| = 1, \quad h^T(\mathcal{A}^*+\Delta\mathcal{A}^*)(\bar{Y}(\Delta\mathcal{D})) = 0. \tag{54}$$

It remains to show that the set of solutions h of (54) 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 (53) is satisfied for all solutions h of (54), provided $\|\Delta\mathcal{D}\|$ is sufficiently small.

In Step 4, 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 (54). It suffices to show that any accumulation point $\bar{h}$ of the sequence h_k satisfies (54) 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 (54) 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 $z \in I\!R^m$ 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 $\Delta z_k \in I\!R^m$ 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 (54), and thus yields the desired contradiction. Hence $\bar{h}$ satisfies (54) 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* (22) *such that for all perturbations $\Delta\mathcal{D} \in \mathcal{N}$ of the problem data* (23)*, $\mathcal{D}$, of* (22)*, there exists a local solution $(x_\Delta, Y_\Delta, S_\Delta)$ of* (29) *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 (46). 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 following class of nonlinear semidefinite programs:

$$\begin{array}{lll} \text{minimize} \quad b^T x \quad \text{subject to} & x \in I\!R^n, \\ & \mathcal{B}(x) \preceq 0. \end{array} \tag{55}$$

Here, $\mathcal{B} : I\!R^n \to \mathcal{S}^m$ is assumed to be a $\mathcal{C}^2$-differentiable function, and $b \in I\!R^n$. Note the class of problems (55) is just the special case $p = q = 0$ of the more general class of problems (12). In particular, in view of (13) and (15), the Lagrangian of (55) 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 (55) is as follows.

Theorem 2 *Let x^* be a local solution of* (55)*, 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,* (55) *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 $\mathcal{A} := D_x\mathcal{B}(x^*) : I\!R^n \to \mathcal{S}^m$, and the matrices $C := \mathcal{B}(x^*)$ and $H := H(x^*, Y^*)$. Then, the SSP approximation (20) (with $p = q = 0$) of (55) at the point (x^*, Y^*) is just the quadratic semidefinite problem

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

Note that (57) is a problem of the form (22) with data (23), $\mathcal{D}$. Moreover, $\Delta\bar{x} := 0$, $\bar{Y} := Y^*$, and $\bar{S} := -\mathcal{A}(0) - C$ satisfy the conditions (25). These conditions coincide with the first-order conditions of (55), and thus the point $(\Delta\bar{x}, \bar{Y}, \bar{S})$ is also the unique solution of (25). Furthermore, the second-order sufficient condition for (57) and (55) coincide. This condition guarantees that $\Delta\bar{x}$ is a locally unique solution of (57). Finally, the Robinson constraint qualification for problem (55) 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 (55) 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. □

7 Convergence of the SSP method

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

For quadruples (x, Y, u, v), where $x \in I\!R^n$, $Y \in \mathcal{S}^m$, $u \in I\!R^p$, $v \in I\!R^q$, we use the norm

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

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

Theorem 3 *Assume that the functions $\mathcal{B}$, c, d in (12) are $\mathcal{C}^3$-differentiable and that problem (12) 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. (27). Let some iterate (x^k, Y^k, u^k, v^k) be given 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 given as the local solution of (18), or, equivalently, (20), 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. The proof is divided into four steps. In the first step, we reduce the general problem (12) to one of the form (55). In the second step, we establish the exact relation of problems (20) and (22). In a third step, we consider x^k near $\bar{x}$ and with a residual of size ϵ in the SSP subproblem. We show that the correction $\Delta x = x^{k+1} - x^k$ is of size ϵ as well and that the residual for the SSP subproblem in the $(k+1)$-st step is of size ϵ^2. We remark that x^k is always the optimal solution of the $(k-1)$-st subproblem, but the data of this subproblem lies $\|x^{k-1} - \bar{x}\|$ away from the SSP subproblem at $(\bar{x}, \bar{Y})$. Instead, we show that x^k is the optimal solution of a subproblem with data only $\|x^k - \bar{x}\|$ away from the SSP subproblem at $(\bar{x}, \bar{Y})$. In a fourth, step we then show by a perturbation analysis of the formally locally optimal point x^k, that the SSP step has length $\mathcal{O}(\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)$ and that the next iterate x^{k+1} lies much closer to $\bar{x}$.

Step 1. 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 (12) 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 (20). 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 (12).

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 remaining part of our analysis we may assume that $p = q = 0$ in (12), i.e., we only need to consider problems of the form (55). Recall that the Lagrangrian, $\mathcal{L}(x, Y)$, and its Hessian,

$H(x,Y)$, of (55) are given by (56). In particular, the Lagrangrian is now only a function of x and Y.

Step 2. We first show how the SSP subproblem (20) (with $p = q = 0$) can be written in the form (22). To this end, we define the linear function $\mathcal{A} := D_x\mathcal{B}(x^k) : I\!R^n \to \mathcal{S}^m$, 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 (12). The SSP subproblem (20) then takes the simple form (57), and in particular, it conforms with the format (22) of Theorem 1.

Step 3. 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 (20) at $\left(\bar{x}, \bar{Y}\right)$. The differentiability of $\mathcal{B}$ guarantees that for each eigenvalue $\bar{\lambda}_i$ of $\mathcal{B}(\bar{x})$, there is an eigenvalue λ_i of $\mathcal{B}(x^k)$ with $|\lambda_i - \bar{\lambda}_i| = \mathcal{O}(\|x^k - \bar{x}\|)$. Moreover, the eigenvectors of separated eigenvalues are also locally Lipschitz continuous; see, e.g., [10, Theorem 8.1.10]. Thus, there exists a perturbation $\Delta C \in \mathcal{S}^m$ with norm $\|\Delta C\|$ of order $\mathcal{O}(\|x^k - \bar{x}\|)$ such that $\mathcal{B}(x^k) + \Delta C \preceq 0$ and such that the null space of $\mathcal{B}(x^k) + \Delta C$ coincides with the null space of $\mathcal{B}(\bar{x})$.

Next, set $\bar{S}^k := -\mathcal{B}(x^k) - \Delta C$. By definition of ΔC, the null spaces of the matrices $\bar{S}^k$ and $\bar{S}$ coincide, and $\bar{S}^k$ and $\bar{Y}$ are strictly complementary. Let

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

From (14) and the fourth row of (16), we obtain the estimate $\|\Delta b\| = \mathcal{O}(\|x^k - \bar{x}\|$. Then, the point $(0, \bar{Y}, \bar{S}^k)$ satisfies the first-order conditions,

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

for the quadratic semidefinite program

$$\begin{array}{lll} \text{minimize} \quad (b + \Delta b)^T \Delta x + \frac{1}{2}(\Delta x)^T H \Delta x & \text{subject to} & \Delta x \in I\!R^n, \\ & & \mathcal{A}(\Delta x) + C + \Delta C \preceq 0. \end{array} \tag{58}$$

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 (20) 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,

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

of the quadratic problem (57) with data (59). As shown in Theorem 2, the assumptions for the nonlinear SDP (55) at $(\bar{x}, \bar{Y})$ imply that problem (57) with data (59) satisfies all conditions of Theorem 1 at $(0, \bar{Y}, \bar{S})$.

Since the second-order sufficient condition depends continuously on the data of (22), it follows that for (58), the second-order condition at $(0, \bar{Y}, \bar{S}^k)$ is satisfied, provided that $\|x^k - \bar{x}\|$ and $\|Y^k - \bar{Y}\|$ are sufficiently small. Thus problem (58) fulfills all assumptions of Theorem 1.

Step 4. 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.

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, $(k+1)$-st iteration.

Corollary 1 states that the corrections $(\Delta x, \Delta\hat{Y})$ of the optimal solution $(0, \bar{Y})$ of (58) generated by the SSP step are given by the tangent equations (28) plus some uniformly bounded second-order terms. Thus, the corrections $(\Delta x, \Delta\hat{Y})$ 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 ΔY produced by the SSP subproblem has the form $\Delta Y = \Delta\hat{Y} + Y^k - \bar{Y}$. 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 rhe residual term of the third equation in (60) is zero. By definition of $(\Delta x, Y^{k+1}, S^{k+1})$, it follows that

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

If the data of (12) is C^3-smooth, this implies that

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

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 generates residuals of order $\mathcal{O}((\|x^k - \bar{x}\| + \|Y^k - \bar{Y}\|)^4)$. Repeating this process, it then follows by a standard argument 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 3 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., [5]. 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 (19) are replaced by approximations $\hat{H}^k$ with $\|H^k - \hat{H}^k\| \to 0$.

Remark 4 The assumption of $\mathcal{C}^3$-differentiability of the functions $\mathcal{B}$, c, d in Theorem 3 can be weakened to $\mathcal{C}^2$-differentiability and a Lipschitz condition for the Hessians at $\bar{x}$.

Remark 5 Recall that in Step 1 of the proof of Theorem 3, we first eliminate the equality constraints. The standard convergence analysis for SQP methods proceeds precisely the opposite way, by assuming without loss of generality that there are no inequality constraints. The approach in this paper to eliminate the equality constraints is merely for convenience to simplify the notation.

8 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. For the derivation of this generalization, we have chosen a motivation that contrasts the SSP method with primal-dual interior methods. For interior methods that are applied directly to nonlinear semidefinite programs, the choice of the symmetrization procedure is considerably more complicated than in the linear case since the system matrix is no longer positive semidefinite. In the proposed method, the choice of the symmetrization scheme is shifted in a very natural way to the subproblems, and is thus reduced to a well-studied problem. 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. Numerical comparisons with trust-region barrier methods are the subject of future research.

References

[1] Alizadeh, F. (1995): Interior point methods in semidefinite programming with applications to combinatorial optimization. SIAM J. Optim. **5**, 13–51

[2] Boggs, P.T., Tolle, J.W. (1995): Sequential quadratic programming. Acta Numer. **4**, 1–51

[3] Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C. (1997): Optimisation Numérique. Springer-Verlag, Berlin

[4] Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V. (1994): Linear Matrix Inequalities in System and Control Theory. SIAM Publications, Philadelphia

[5] Correa, R., Ramírez C., H. (2002): A global algorithm for nonlinear semidefinite programming. Research Report 4672, INRIA, Rocquencourt, France

[6] Fares, B., Apkarian, P., Noll, D. (2001): An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. Internat. J. Control **74**, 348–360

[7] Fares, B., Noll, D., Apkarian, P. (2002): Robust control via sequential semidefinite programming. SIAM J. Control Optim. **40**, 1791–1820

[8] Freund, R.W., Jarre, F. (2003): An extension of the positive real lemma to descriptor systems. Optim. Methods Softw., to appear

[9] Freund, R.W., Jarre, F. (2003): A sensitivity result for semidefinite programs. Numerical Analysis Manuscript No. 03–4–07, Bell Laboratories, Murray Hill, New Jersey

[10] Golub, G.H., Van Loan, C.F. (1996): Matrix Computations, Third Edition. The Johns Hopkins University Press, Baltimore

[11] Hörmander, L. (1973): An Introduction to Complex Analysis in Several Variables, Second Edition. North-Holland Publishing Company, Amsterdam

[12] Kocvara, M., Stingl, M. (2001): Augmented Lagrangian method for semidefinite programming. Technical Report, Institut für Angewandte Mathematik, Universität Erlangen-Nürnberg

[13] Mangasarian, O.L. (1969): Nonlinear Programming. MvGraw-Hill, New York

[14] Monteiro, R.D.C., Zhang, Y. (1998): A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. Math. Program. **81**, 281–299

[15] Nayakkankuppam, M.V., Overton, M.L. (1999): Conditioning of semidefinite programs. Math. Program. **85**, 525–540

[16] Overton, M.L. (1988): On minimizing the maximum eigenvalue of a symmetric matrix. SIAM J. Matrix Anal. Appl. **9**, 256–268

[17] Overton, M.L. (1992): Large-scale optimization of eigenvalues. SIAM J. Optim. **2**, 88–120

[18] Robinson, S.M. (1976): First order conditions for general nonlinear optimization. SIAM J. Appl. Math. **30**, 597–607

[19] Robinson, S.M. (1982): Generalized equations and their solutions, part II: applications to nonlinear programming. Math. Prog. Study **19**, 200–221

[20] Shapiro, A., Scheinberg, K. (2000): 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

[21] Sturm, J.F. (1999): Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optim. Methods Softw. **11–12**, 625–653

[22] Sturm, J.F., Zhang, S. (2001): On sensitivity of central solutions in semidefinite programming. Math. Program. **90**, 205–227

[23] Todd, M.J. (2001): Semidefinite optimization. Acta Numer. **10**, 515–560

[24] Tuan, H.D., Apkarian, P., Nakashima, Y. (2000): A new Lagrangian dual global optimization algorithm for solving bilinear matrix inequalities. Internat. J. Robust Nonlinear Control **10**, 561–578

[25] Vandenberghe, L., Boyd, S. (1996): Semidefinite programming. SIAM Rev. **38**, 49–95

[26] Vanderbei, R.J. (1997): LOQO user's manual—version 3.10. Report SOR 97-08, Princeton University, Princeton, New Jersey

[27] Vanderbei, R.J., Benson, H., Shanno, D. (2000): Interior-point methods for nonconvex nonlinear programming: filter methods and merit functions. Report ORFE 00-06, Princeton University, Princeton, New Jersey

[28] Wolkowicz, H., Saigal, R., Vandenberghe, L., eds. (2000): Handbook of Semidefinite Programming: Theory, Algorithms and Applications. Kluwer Academic Publishers, Boston