Two theoretical results for sequential semidefinite programming

Rodrigo Garcés * and Walter Gomez Bofill *
Florian Jarre †

Feb. 1, 2008

Abstract. We examine the local convergence of a sequential semidefinite programming approach for solving nonlinear programs with nonlinear semidefiniteness constraints. Known convergence results are extended to slightly weaker second order sufficient conditions and the resulting subproblems are shown to have local convexity properties that imply a weak form of self-concordance of the barrier subproblems.

Key words. semidefinite programming, second order sufficient condition, sequential quadratic programming, quadratic semidefinite program, self-concordance, sensitivity, convergence.

1 Introduction

The papers [6, 8] present sequential semidefinite programming algorithms for solving nonlinear semidefinite programs. The semidefinite subproblems are solved by standard interior point packages such as SEDUMI, and numerical examples in [6, 8] show a good overall convergence. Both papers also present a proof of local quadratic convergence under the assumption of a certain second order sufficient condition for a local minimizer. Of course, this local analysis is of practical relevance only if the subproblems are efficiently solvable—for example by polynomial time interior point algorithms.

The main contributions of the present paper are as follows:

- As pointed out by Shapiro [18], the second order condition used in [6, 8] is unnecessarily strong. The paper [5] gives an example of a perfectly well-conditioned problem that does not satisfy this strong second order sufficient condition. In the present paper we change the analysis of [8] to a weaker second order condition. Unfortunately, the subproblems to be solved in the sequential semidefinite programming approach are nonconvex, and as pointed out in [5], local superlinear convergence is lost, when convexifying the subproblems.
- We therefore also include a local analysis of nonconvex semidefinite subproblems as arising in the sequential semidefinite programming algorithms providing further theoretical

*Department of Mathematical Engineering, Universidad de La Frontera, Av. Francisco Salazar 01145, Temuco, Chile, *e-mail*: rgarc008@pinhue.ufro.cl; wgomez@ufro.cl

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

justification for the implementations in [6, 8]. To our knowledge this is the first extension of the concept of self-concordance to a class of practically important and "genuinely" nonconvex problems.

In contrast to our analysis below, the analysis in [6] does not require strict complementarity. The analysis in this paper is based on a sensitivity result that exploits strict complementarity. We believe that the sensitivity result also holds in a weaker form – local Lipschitz continuity in place of differentiability – when the strict complementarity assumption does not hold. Such a weaker sensitivity result still suffices to establish local quadratic converge of the sequential semidefinite programming approach.

The numerical solution of nonlinear semidefinite programming problems has also been successfully approached by augmented Lagrangian methods, see [10, 11]. Fast local convergence results of such approaches under weak second order conditions have been provided e.g. in [14, 19].

2 Notation and Preliminaries

By $\mathbb{S}^m$ we denote the linear space of $m \times m$ real symmetric matrices. The space $\mathbb{R}^{m\times n}$ is equipped with the inner product $A \bullet B := \text{trace}(A^T B) = \sum_{i=1}^m \sum_{j=1}^n A_{ij} B_{ij}$. The corresponding so called Forbenius norm is then defined by $\|A\|_F = \sqrt{A \bullet A}$. The negative semidefinite order $\preceq$ for $A, B \in \mathbb{S}^m$ is defined in the standard form, that is, $A \preceq B$ iff $A - B$ is a negative semidefinite matrix. The order relations $\prec$, $\succeq$ and $\succ$ are defined similarly. By $\mathbb{S}^m_+$ we denote the set of positive semidefinite matrices.

The following simple Lemma is used in the sequel.

Lemma 1 (See [8]) *Let $Y, S \in \mathbb{S}^m$.*

(a) If $Y, S \succeq 0$ then

$$YS + SY = 0 \iff YS = 0. \tag{1}$$

(b) If $Y + S \succ 0$ and $YS + SY = 0$ then $Y, S \succeq 0$.

(c) If $Y + S \succ 0$ and $YS + SY = 0$ then for any $\dot{Y}, \dot{S} \in \mathbb{S}^m$,

$$Y\dot{S} + \dot{Y}S = 0 \iff Y\dot{S} + \dot{Y}S + \dot{S}Y + S\dot{Y} = 0. \tag{2}$$

Moreover, Y, S have representations of the form

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

where U is an $m \times m$ orthogonal matrix , $Y_1 \succ 0$ is a $(m-r) \times (m-r)$ diagonal matrix and $S_2 \succ 0$ is a $r \times r$ diagonal matrix, and any matrices $\dot{Y}, \dot{S} \in \mathbb{S}^m$ satisfying (2) are of the form

$$\dot{Y} = U \begin{bmatrix} \dot{Y}_1 & \dot{Y}_3 \\ \dot{Y}_3^T & 0 \end{bmatrix} U^T \quad , \quad \dot{S} = U \begin{bmatrix} 0 & \dot{S}_3 \\ \dot{S}_3^T & \dot{S}_2 \end{bmatrix} U^T,$$

where

$$Y_1 \dot{S}_3 + \dot{Y}_3 S_2 = 0. \tag{3}$$

Proof: For (a), (c) see [8].

(b) By contradiction we assume that λ is a negative eigenvalue of S and u a corresponding eigenvector. The equality $YS + SY = 0$ implies that

$$\begin{aligned} u^T Y(Su) + (Su)^T Yu &= 0, \\ u^T Y(\lambda u) + (\lambda u)^T Yu &= 0, \\ \lambda(u^T Yu + u^T Yu) &= 0, \\ \Rightarrow u^T Yu &= 0, \text{ since } \quad \lambda < 0. \end{aligned}$$

Now using the fact that $Y + S \succ 0$, we have

$$0 < u^T(Y + S)u = u^T Yu + u^T Su = \lambda u^T u = \lambda \parallel u \parallel^2 < 0$$

which is a contradiction. Hence, $S \succeq 0$. The same arguments give us $Y \succeq 0$. ■

Remark 1 *Due to (3) and the positive definiteness of the diagonal matrices* Y_1 *and* S_2*, it follows that* $(\dot{Y}_3)_{ij}(\dot{S}_3)_{ij} < 0$ *whenever* $(\dot{Y}_3)_{ij} \neq 0$*. Hence, if in addition to (3) also* $\langle \dot{Y}_3, \dot{S}_3 \rangle = 0$ *holds true, then* $\dot{Y}_3 = \dot{S}_3 = 0$*.*

In the sequel we refer to the set of symmetric and strict complementary matrices

$$\mathcal{C} = \{(Y, S) \in \mathbb{S}^m \times \mathbb{S}^m \ \mid YS + SY = 0, \ Y + S \succ 0\}. \tag{4}$$

As a consequence of Lemma 1 (b), the set $\mathcal{C}$ is (not connected) contained in $\mathbb{S}^m_+ \times \mathbb{S}^m_+$. Moreover, Lemma 1 (c) implies that the rank of the matrices Y and S is locally constant on $\mathcal{C}$.

3 Nonlinear semidefinite programs

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

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad b^T x \quad \text{subject to} \quad G(x) \preceq 0. \tag{5}$$

Here, the function G is at least $\mathcal{C}^3$-differentiable.

For simplicity of presentation, we have chosen a simple form of problem (5). All statements about (5) in this paper can be modified so that they apply to additional nonlinear equality and inequality constraints and to nonlinear objective functions.

The Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathcal{S}^m \to \mathbb{R}$ of (5) is defined as follows:

$$\mathcal{L}(x, Y) := b^T x + G(x) \bullet Y. \tag{6}$$

Its gradient with respect to x is given by

$$g(x, Y) := \nabla_x \mathcal{L}(x, Y) = b + \nabla_x (G(x) \bullet Y) \tag{7}$$

and its Hessian by

$$H(x, Y) := \nabla_x^2 \mathcal{L}(x, Y) = \nabla_x^2 (G(x) \bullet Y). \tag{8}$$

3.1 Assumptions

In the following we state second order sufficient conditions due to [17] that are weaker than the ones used in [6, 8].

A1 We assume that $\bar{x}$ is a unique local minimizer of (5) that satisfies the Mangasarian-Fromovitz constraint qualification, i.e., there exists a vector $\Delta x \neq 0$ such that $G(\bar{x}) + DG(\bar{x})[\Delta x] \prec 0$, where by definition $DG(x)[s] = \sum_{i=1}^{n} s_i D_{x_i} G(x)$.

Assumption **A1** implies that the first-order optimality condition is satisfied, i.e., there exist matrices $\bar{Y}$, $\bar{S} \in \mathcal{S}^m$ such that

$$\begin{aligned} G(\bar{x}) + \bar{S} &= 0, \\ g(\bar{x}, \bar{Y}) &= 0, \\ \bar{Y}\bar{S} &= 0, \\ \bar{Y},\, \bar{S} &\succeq 0. \end{aligned} \tag{9}$$

A triple $(\bar{x}, \bar{Y}, \bar{S})$ satisfying (9), will be called a sationary point of (5).

Due to Lemma 1 (a) the third equation in (9) can be substituted by $\bar{Y}\bar{S} + \bar{S}\bar{Y} = 0$. This reformulation does not change the set of sationary points, but it reduces the underlying system of equations (via a symmetrization of YS) in the variables (x, Y, S), such that it has now the same number of equations and variables. This is a useful step in order to apply the implicit function theorem.

A2 We also assume that $\bar{Y}$ is unique and that $\bar{S}, \bar{Y}$ are strictly complementary, i.e. $(\bar{Y}, \bar{S}) \in \mathcal{C}$.

According to Lemma 1 (c), there exists a unitary matrix $U = [U_1, U_2]$ that simultaneously diagonalizes $\bar{Y}$ and $\bar{S}$. Here, U_2 has $r := \operatorname{rank}(\bar{S})$ columns and U_1 has $m - r$ columns. Moreover the first $m - r$ diagonal entries of $U^T \bar{S} U$ are zero, and the last r diagonal entries of $U^T \bar{Y} U$ are zero. In particular, we obtain

$$U_1^T G(\bar{x}) U_1 = 0 \quad \text{and} \quad U_2^T \bar{Y} U_2 = 0. \tag{10}$$

A vector $h \in \mathbb{R}^n$ is called a critical direction at $\bar{x}$ if $b^T h = 0$ and it is the limit of feasible directions of (5), i.e. if there exist $h^k \in \mathbb{R}^n$ and $\epsilon_k > 0$ with $\lim_{k\to\infty} h^k = h$, $\lim_{k\to\infty} \epsilon_k = 0$, and $G(\bar{x} + \epsilon_k h^k) \preceq 0$ for all k. As shown in [1] the cone of critical directions at a strictly complementary local solution $\bar{x}$ is given by

$$C(\bar{x}) := \{h \mid U_1^T DG(\bar{x})[h] U_1 = 0\}. \tag{11}$$

A3 We further assume that $\bar{x}, \bar{Y}$ satisfies the second order sufficient condition:

$$h^T(\nabla_x^2 \mathcal{L}(\bar{x}, \bar{Y}) + \mathcal{H}(\bar{x}, \bar{Y}))h > 0 \qquad \forall\, h \in C(\bar{x}) \setminus \{0\} \tag{12}$$

Here $\mathcal{H}$ is a nonnegative matrix related to the curvature of the semidefinite cone in $G(\bar{x})$ along direction $\bar{Y}$ (see [17]) and is given by its matrix entries

$$\mathcal{H}_{i,j} := -2\bar{Y} \bullet G_i(\bar{x}) G(\bar{x})^\dagger G_j(\bar{x}),$$

where $G_i(\bar{x}) := DG(\bar{x})[e_i]$ with e_i denoting the i-th unit vector. Furthermore, $G(\bar{x})^\dagger$ denotes the Moore-Penrose pseudo-inverse of $G(\bar{x})$, i.e.

$$G(\bar{x})^\dagger = \sum \lambda_i^{-1} u_i u_i^T,$$

where λ_i are the nonzero eigenvalues of $G(\bar{x})$ and u_i corresponding orthonormal eigenvectors.

Remark 2 *The Moore-Penrose inverse* $M^\dagger$ *is a continuous function of* M*, when the perturbations of* M *do not change its rank, see [3].*

The second order sufficient condition can equivalently be stated as follows: There exists a number $\mu > 0$ such that

$$h^T(\nabla_x^2 \mathcal{L}(\bar{x}, \bar{Y}) + \mathcal{H}(\bar{x}, \bar{Y}))h > \mu \qquad \forall\, h \in C_1(\bar{x}) := \{h \in C(\bar{x}) \mid \|h\| = 1\}. \tag{13}$$

Since the set $C_1(\bar{x})$ is compact the strict positivity of the quadratic form still remains valid for vectors h in a tiny ρ-enlargement of $C_1(\bar{x})$ (for some smaller value of $\mu > 0$). Taking the conic hull of the ρ-enlargement of $C_1(\bar{x})$ leads to the following result:

Lemma 2 *If the second order sufficient condition is satisfied at a critical point* $(\bar{x}, \bar{Y})$*, then there exists some* $\mu > 0$ *and* $\rho > 0$ *such that:*

$$h^T(\nabla_x^2 \mathcal{L}(\bar{x}, \bar{Y}) + \mathcal{H}(\bar{x}, \bar{Y}))h > \mu \|h\|^2, \ \forall h \in C^\rho(\bar{x}),$$

where

$$C^\rho(\bar{x}) := \{\lambda h \in \mathbb{R}^n \mid \lambda \geq 0, \quad \|h\| = 1, \ \exists y \in C_1(\bar{x}), \|y - h\| \leq \rho\}. \tag{14}$$

The curvature term can be rewritten as follows:

$$\begin{aligned} h^T \mathcal{H}(\bar{x}, \bar{Y})h &= \sum_{i,j} h_i h_j (-2\bar{Y} \bullet G_i(\bar{x}) G(\bar{x})^\dagger G_j(\bar{x})), \\ &= -2\bar{Y} \bullet (\sum_{i,j} h_i h_j G_i(\bar{x}) G(\bar{x})^\dagger G_j(\bar{x})), \\ &= -2\bar{Y} \bullet (\sum_{i=1}^n h_i G_i(\bar{x}) G(\bar{x})^\dagger \sum_{j=1}^n h_j G_j(\bar{x})), \\ &= -2\bar{Y} \bullet DG(\bar{x})[h] G(\bar{x})^\dagger DG(\bar{x})[h]. \end{aligned} \tag{15}$$

Note that in the particular case where G is affine (i.e. $G(x) = \mathcal{A}(x) + C$, with a linear map $\mathcal{A}$ and $C \in \mathbb{S}^m$), the curvature term is given by

$$h^T \mathcal{H}(\bar{x}, \bar{Y})h := -2\bar{Y} \bullet (\mathcal{A}(h)(\mathcal{A}(\bar{x}) + C)^\dagger \mathcal{A}(h)). \tag{16}$$

4 Sensitivity result

Let us now consider the following quadratic semidefinite programming problem

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & b^T x + \frac{1}{2} x^T H x \\ s.t. & \\ & \mathcal{A}(x) + C \preceq 0. \end{array} \tag{17}$$

Here, $\mathcal{A} : \mathbb{R}^n \longrightarrow \mathbb{S}^m$ is a linear function, $b \in \mathbb{R}^n$, and $C, H \in \mathbb{S}^m$. The data to this problem is

$$\mathscr{D} := [\mathcal{A}, b, C, H]. \tag{18}$$

In the next theorem, we present a sensitivity result for the solutions of (17), when the data $\mathscr{D}$ is changed to $\mathscr{D} + \triangle\mathscr{D}$ where

$$\triangle\mathscr{D} := [\triangle\mathcal{A}, \triangle b, \triangle C, \triangle H] \tag{19}$$

is a sufficiently small perturbation.

The triple $(\bar{x}, \bar{Y}, \bar{S}) \in \mathbb{R}^n \times \mathbb{S}^m \times \mathbb{S}^m$ is a stationary point for (17), if

$$\begin{array}{rcl} \mathcal{A}(\bar{x}) + C + \bar{S} &=& 0, \\ b + H\bar{x} + \mathcal{A}^*(\bar{Y}) &=& 0, \\ \bar{Y}\bar{S} + \bar{S}\bar{Y} &=& 0, \\ \bar{Y}, \bar{S} &\succeq& 0 \,. \end{array} \tag{20}$$

Remark 3 *Below, we consider tiny perturbations $\triangle\mathscr{D}$ such that there is an associated strictly complementary solution $(x, Y, S)(\triangle\mathscr{D})$ of (20). For such x there exists $U_1 = U_1(x)$ and an associated cone of critical directions $C(x)$. The basis $U_1(x)$ generally is not continuous with respect to x. However, the above characterization (11) of $C(\bar{x})$ under strict complementarity can be stated using any basis of the orthogonal space of $G(\bar{x})$. Since such basis can be locally parameterized in a smooth way over the set $\mathcal{C}$ (4) it follows that locally, the set $C(x)$ forms a closed point to set mapping.*

The following is a slight generalization of Theorem 1 in [8].

Theorem 1 *Let the point $(\bar{x}, \bar{Y}, \bar{S})$ be a stationary point satisfying the assumptions **A1-A3** for the problem (17) with data $\mathscr{D}$. Then, for all sufficiently small perturbations $\triangle\mathscr{D}$ as in (19), there exists a locally unique stationary point $(\bar{x}(\mathscr{D} + \triangle\mathscr{D}), \bar{Y}(\mathscr{D} + \triangle\mathscr{D}), \bar{S}(\mathscr{D} + \triangle\mathscr{D}))$ of the perturbed program (17) with data $\mathscr{D} + \triangle\mathscr{D}$. Moreover, the point $(\bar{x}(\mathscr{D}), \bar{Y}(\mathscr{D}), \bar{S}(\mathscr{D}))$ is a differentiable function of the perturbation (19), and for $\triangle\mathscr{D} = 0$, we have $(\bar{x}(\mathscr{D}), \bar{Y}(\mathscr{D}), \bar{S}(\mathscr{D})) = (\bar{x}, \bar{Y}, \bar{S})$. The derivative $D_{\mathscr{D}}(\bar{x}(\mathscr{D}), \bar{Y}(\mathscr{D}), \bar{S}(\mathscr{D}))$ of $(\bar{x}(\mathscr{D}), \bar{Y}(\mathscr{D}), \bar{S}(\mathscr{D}))$ with respect to $\mathscr{D}$ evaluated at $(\bar{x}, \bar{Y}, \bar{S})$ is characterized by the directional derivatives*

$$(\dot{x}, \dot{Y}, \dot{S}) := D_{\mathscr{D}}(\bar{x}(\mathscr{D}), \bar{Y}(\mathscr{D}), \bar{S}(\mathscr{D}))[\triangle\mathscr{D}]$$

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

$$\begin{array}{rcl} \mathcal{A}(\dot{x}) + \dot{S} &=& -\triangle C - \triangle\mathcal{A}(\bar{x}), \\ H\dot{x} + \mathcal{A}^*(\dot{Y}) &=& -\triangle b - \triangle H\bar{x} - \triangle\mathcal{A}^*(\bar{Y}), \\ \bar{Y}\dot{S} + \dot{Y}\bar{S} + \dot{S}\bar{Y} + \bar{S}\dot{Y} &=& 0, \end{array} \tag{21}$$

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

This theorem is related to other sensitivity results for semidefinite programming problems (see, for instance, [2, 7, 20]). Local Lipschitz properties under strict complementarity can be found in [12]. In [17] the directional derivative $\dot{x}$ is given as solution of a quadratic problem.

Proof:

Following the outline in [8] this proof is based on the application of the implicit function theorem to the system of equations (20). In order to apply this result we show that the matrix of partial derivatives of system (20) with respect to the variables (x, Y, S) is regular. To this end it suffices to prove that the system

$$\begin{aligned} \mathcal{A}(\dot{x}) + \dot{S} &= 0, \\ H\dot{x} + \mathcal{A}^*(\dot{Y}) &= 0, \\ \bar{Y}\dot{S} + \dot{Y}\bar{S} + \dot{S}\bar{Y} + \bar{S}\dot{Y} &= 0, \end{aligned} \tag{22}$$

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

Let $(\dot{x}, \dot{Y}, \dot{S}) \in \mathbb{R}^n \times \mathbb{S}^m \times \mathbb{S}^m$ be a solution of (22). Since $\bar{Y}$ and $\bar{S}$ are strictly complementary, it follows from part (c) of Lemma 1, the existence of an orthonormal matrix U such that:

$$\bar{Y} = U\tilde{Y}U^T, \quad \bar{S} = U\tilde{S}U^T \tag{23}$$

where

$$\tilde{Y} = \begin{bmatrix} \bar{Y}_1 & 0 \\ 0 & 0 \end{bmatrix} \quad , \quad \tilde{S} = \begin{bmatrix} 0 & 0 \\ 0 & \bar{S}_2 \end{bmatrix}, \tag{24}$$

with $\bar{Y}_1, \bar{S}_2$ diagonal and positive definite. Furthermore, the matrices $\dot{Y}, \dot{S} \in \mathbb{S}^m$ satisfying (22) fulfill the relations

$$\dot{Y} = U\check{Y}U^T, \quad \dot{S} = U\check{S}U^T \tag{25}$$

where

$$\check{Y} = \begin{bmatrix} \dot{Y}_1 & \dot{Y}_3 \\ \dot{Y}_3^T & 0 \end{bmatrix}, \quad \check{S} = \begin{bmatrix} 0 & \dot{S}_3 \\ \dot{S}_3^T & \dot{S}_2 \end{bmatrix}, \quad \text{and} \quad \dot{Y}_3\bar{S}_2 + \bar{Y}_1\dot{S}_3 = 0. \tag{26}$$

Using the decomposition given in (10), the first equation of (22) and (25) we have

$$U_1^T \mathcal{A}(\dot{x}) U_1^T = -U_1^T U \check{S} U^T U_1 = 0.$$

It follows that $\dot{x} \in C(\bar{x})$. Now using (16), (23-26), the first equation in (20) and the first equation in (22), we obtain

$$\begin{aligned} \dot{x}^T \mathcal{H}(\bar{x}, \bar{Y})\dot{x} &= -2\bar{Y} \bullet \mathcal{A}(\dot{x})(\mathcal{A}(\bar{x}) + C)^\dagger \mathcal{A}(\dot{x}), \\ &= -2\bar{Y} \bullet \dot{S}(-\bar{S})^\dagger \dot{S}, \\ &= -2\tilde{Y} \bullet \check{S}(-\tilde{S})^\dagger \check{S}, \quad \text{since} \quad \bar{S}_2 \succ 0, \quad \tilde{S}^\dagger = \begin{bmatrix} 0 & 0 \\ 0 & \bar{S}_2^{-1} \end{bmatrix}. \\ &= -2\dot{Y}_3 \bullet \dot{S}_3. \end{aligned}$$

By the same way, using the first two relations in (25) and the first two equations of (22), one readily verifies that

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

Consequently

$$\dot{x}^T (H + \mathcal{H}(\bar{x}, \bar{Y}))\dot{x} = 0. \tag{27}$$

This implies that $\dot{x} = 0$, since $\dot{x} \in C(\bar{x})$. Using Remark 1 it follows also that $\dot{Y}_3 = \dot{S}_3 = 0$.

By the first equation of (22), we obtain

$$\dot{S} = -\mathcal{A}(\dot{x}) = -\mathcal{A}(0) = 0. \tag{28}$$

Thus, it only remains to show that $\dot{Y} = 0$. In view of (26) we have

$$\check{Y} = \begin{bmatrix} \dot{Y}_1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{29}$$

Now suppose that $\dot{Y}_1 \neq 0$. Since $\bar{Y}_1 \succ 0$, it is clear that there exists some $\bar{\tau} > 0$ such that $\bar{Y}_1 + \tau\dot{Y}_1 \succ 0 \quad \forall \tau \in (0, \bar{\tau}]$. If we define $\bar{Y}_\tau := \bar{Y} + \tau\dot{Y}$ it follows that

$$\bar{Y}_\tau = U(\tilde{Y} + \tau\check{Y})U^T \succeq 0 \wedge \bar{Y}_\tau \neq \bar{Y}, \quad \forall \tau \in (0, \bar{\tau}].$$

Moreover, using (20),(22), and (28), one readily verifies that $(\bar{x}, \bar{Y}_\tau, \bar{S})$ also satisfies (20) for all $\tau \in (0, \bar{\tau}]$. This contradicts the assumption that $(\bar{x}, \bar{Y}, \bar{S})$ is a locally unique stationary point. Hence $\dot{Y}_1 = 0$ and by (29), $\dot{Y} = 0$.

We can now apply the implicit function theorem to the system

$$\begin{aligned} \mathcal{A}(x) + C + S &= 0, \\ Hx + b + \mathcal{A}^*(Y) &= 0, \\ YS + SY &= 0. \end{aligned} \tag{30}$$

As we have just seen, the linearization of (30) at the point $(\bar{x}, \bar{Y}, \bar{S})$ is nonsingular. Therefore the system (30) has a differentiable and locally unique solution $(\bar{x}(\triangle\mathscr{D}), \bar{Y}(\triangle\mathscr{D}), \bar{S}(\triangle\mathscr{D}))$. By the continuity of $\bar{Y}(\triangle\mathscr{D}), \bar{S}(\triangle\mathscr{D})$ with respect to $\triangle\mathscr{D}$ it follows that for $\| \triangle\mathscr{D} \|$ sufficiently small $\bar{Y}(\triangle\mathscr{D}) + \bar{S}(\triangle\mathscr{D}) \succ 0$, i.e. $(\bar{Y}(\triangle\mathscr{D}) + \bar{S}(\triangle\mathscr{D})) \in \mathcal{C}$.

Consequently, part (b) of Lemma 1 we have $\bar{Y}(\triangle\mathscr{D}), \bar{S}(\triangle\mathscr{D}) \succeq 0$. This implies that the locally solutions of the system (30) are actually stationary points.

Note that the dimension of the image space of $\bar{S}(\triangle\mathscr{D})$ is constant for all $\| \triangle\mathscr{D} \|$ sufficiently small. According to Remark 2 it holds that $\bar{S}(\triangle\mathscr{D})^\dagger \to \bar{S}^\dagger$ when $\triangle\mathscr{D} \to 0$.

Finally we prove that the second-order sufficient condition is invariant under small perturbations $\triangle\mathscr{D}$ of the problem data $\mathscr{D}$. We just need to show that there exists $\bar{\varepsilon} > 0$ such that for all $\triangle\mathscr{D}$ with $\| \triangle\mathscr{D} \| \leq \bar{\varepsilon}$ it holds:

$$h^T((H + \triangle H) + \mathcal{H}(\bar{x}(\triangle\mathscr{D}), \bar{Y}(\triangle\mathscr{D})))h > 0 \qquad \forall h \in C(\bar{x}(\triangle\mathscr{D}))/\{0\}. \tag{31}$$

Since $C(\bar{x}(\triangle\mathscr{D}))/\{0\}$ is a cone, it suffices to consider unitary vectors, i.e. $\| h \| = 1$. We assume by contradiction that there exists $\varepsilon_k \to 0$, $\{\triangle\mathscr{D}_k\}$ with $\| \triangle\mathscr{D}_k \| \leq \varepsilon_k$, and $\{h_k\}$ with $h_k \in C(\bar{x}(\triangle\mathscr{D}_k))/\{0\}$ such that

$$h_k^T((H + \triangle H_k) + \mathcal{H}(\bar{x}(\triangle\mathscr{D}), \bar{Y}(\triangle\mathscr{D}_k)))h_k \leq 0. \tag{32}$$

We may assume that h_k converges to h with $\| h \| = 1$, when $k \to \infty$. Since $\triangle\mathscr{D}_k \to 0$, we obtain from the already mentioned convergence $\bar{S}(\triangle\mathscr{D})^\dagger \to \bar{S}^\dagger$ and simple continuity arguments that:

$$0 < h^T Hh + h^T\mathcal{H}(\bar{x}, \bar{Y})h \leq 0. \tag{33}$$

The left inequality of this contradiction follows from the second order sufficient condition since $h \in C(\bar{x}(0))/\{0\}$ due to Remark 3. ■

The sensitivity result of Theorem 1 was used in [8] to establish local quadratic convergence of the SSP method. By extending this result to the weaker form of second order sufficient condition, the analysis in [8] can be applied in a straightforward way to this more general class of nonlinear semidefinite programs. In fact, the analysis in [8] only used local Lipschitz continuity of the solution with respect to small changes of the data. Of course the differentiability established in Theorem 1 implies local Lipschitz continuity. We believe, however, that local Lipschitz continuity is also valid if the strict complementarity assumption is not satisfied.

5 Nonconvex quadratic SDP subproblems

The quadratic approximation to (5) at some primal point $x \in \mathbb{R}^n$ and at an associated dual estimate $Y \in \mathcal{S}^m$ is given by

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \quad q(s) \quad \text{subject to} \quad G(x) + DG(x)[s] \preceq 0, \tag{34}$$

where $q(s) := b^T s + \frac{1}{2} s^T H s$ and $H = \nabla_x^2 \mathcal{L}(x, Y)$.

In the following we analyze quadratic approximations of the form (34) for (x, Y) sufficiently close to $(\bar{x}, \bar{Y})$, where $(\bar{x}, \bar{Y}, \bar{S})$ is a fixed stationary point of (5) satisfying the general assumptions **A1-A3** of Section 3. It can be easily verified, that the same assumtions also hold true at the solution $(0, \bar{Y})$ of the above problem (34) for the data given by $(\bar{x}, \bar{Y})$. By Theorem 1, there exists a unique local optimal solution $\bar{s}_q$ of (34) with norm $\|\bar{s}_q\| = O(\|(x, Y) - (\bar{x}, \bar{Y})\|)$. Moreover, the second order sufficient condition (12) also holds for (34) at $\bar{s}_q$ with associated multiplyer $\bar{Y}_q$. i.e. by Lemma 2, there exists $\epsilon > 0$ such that

$$-u^T H u \leq 2(1 - \epsilon) \bar{Y}_q \bullet DG(x)[u] \bar{S}_q^\dagger DG(x)[u] \qquad \forall \;\; u \in C^\rho(\bar{s}_q), \tag{35}$$

where $C^\rho(\bar{s}_q)$ is a ρ-enlargement as in (14) of the critical cone to the problem (34) at the point $\bar{s}_q$, i.e. $C(\bar{s}_q)$; and $\bar{S}_q = -G(x) - DG(x)[\bar{s}_q]$. Note that $C(\bar{s}_q)$ coincides with $C(x)$ at the data set given by $(\bar{x}, \bar{Y})$.

5.1 Barrier calculus

We extend the vector $s \in \mathbb{R}^n$ to $\tilde{s} := (s_0; s) \in \mathbb{R}^{n+1}$ and consider a reformulation of (34) in $\mathbb{R}^{n+1}$ by adding a slack variable for the objective function,

$$\underset{\tilde{s} \in \mathbb{R}^{n+1}}{\text{minimize}} \quad s_0 \quad \text{subject to} \quad q(s) - s_0 \leq 0, \quad G(x) + DG(x)[s] \preceq 0. \tag{36}$$

The first constraint in (36) will be denoted by $\tilde{q}(\tilde{s}) := q(s) - s_0 \leq 0$.

For problem (36) we consider the barrier functions

$$\tilde{\Phi}(\tilde{s}) := \Phi(s) := -\log(\det(-G(x) - DG(x)[s]))$$

and

$$\tilde{\Psi}(\tilde{s}) := -\log(-\tilde{q}(\tilde{s})) + \tilde{\Phi}(\tilde{s}).$$

Let $\tilde{s}$ be strictly feasible for (36) and $\tilde{h} := (h_0; h) \in \mathbb{R}^{n+1}$ be arbitrary. For small $|t|$ we consider the "line function"

$$l_{\tilde{s}, \tilde{h}, H}(t) := l(t) := \tilde{\Psi}(\tilde{s} + t\tilde{h}). \tag{37}$$

This function depends on the parameters $\tilde{s}, \tilde{h}$ and on the data of (34), in particular, also on $H = \nabla_x^2 \mathcal{L}(x, Y)$. Subsequently we will also consider the function $l_{\tilde{s}, \tilde{h}, 0}$, where H of (37) is replaced with the zero matrix.

The first derivative of l at $t = 0$ is given by

$$l'(0) = \frac{-h_0 + (b + Hs)^T h}{-\tilde{q}(\tilde{s})} - (G(x) + DG(x)[s])^{-1} \bullet DG(x)[h],$$

To simplify the notation we write $b_s := b + Hs$ in the sequel. Since the x is fixed in (36), let us also introduce the notations

$$DG[s] := DG(x)[s] \quad \text{and} \quad \hat{G}[s] := G(x) + DG(x)[s] \tag{38}$$

We obtain a slightly more compact formula for the first derivative,

$$l'(0) = \frac{b_s^T h - h_0}{-\tilde{q}(\tilde{s})} - (\hat{G}[s])^{-1} \bullet DG[h],$$

Likewise, the second and third derivatives are given by

$$l''(0) = \frac{(b_s^T h - h_0)^2}{\tilde{q}(\tilde{s})^2} + \frac{h^T H h}{-\tilde{q}(\tilde{s})} + \|(\hat{G}[s])^{-\frac{1}{2}} DG[h] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2$$

and

$$l'''(0) = 2\frac{(b_s^T h - h_0)^3}{-\tilde{q}(\tilde{s})^3} + 3\frac{h^T H h}{-\tilde{q}(\tilde{s})} \frac{(b_s^T h - h_0)}{-\tilde{q}(\tilde{s})} + D^3 \tilde{\Phi}(\tilde{s})[\tilde{h}, \tilde{h}, \tilde{h}].$$

The function $\tilde{\Psi}$ is self-concordant [13], iff for any $\tilde{s}$, $\tilde{h}$ as above, the first and the third derivatives of l at $t = 0$ can be bounded by suitable fixed multiples of the second derivative, i.e. if there exist $\alpha \geq 1$, $\beta \geq 1$ such that

$$l'''(0) \leq 2\alpha (l''(0))^{\frac{3}{2}} \quad \text{and} \quad (l'(0))^2 \leq \beta l''(0). \tag{39}$$

If H is a positive semidefinite matrix, then it is known that $\tilde{\Psi}$ is self-concordant with $\alpha = 1$ and $\beta = m + 1$, see [13], Proposition 5.4.5.

To guarantee quadratic convergence of Newton's method it is sufficient, in fact, to establish the first condition for any fixed number α, see e.g. [13], Thm 2.2.3. Likewise, the second condition can be relaxed with some other constant $\beta < \infty$ in place of "$m + 1$".

Below we do not show self-concordance but a weaker property that establishes these inequalities not for all strictly feasible $\tilde{s}$ but only for $\tilde{s}$ near the central path – i.e. near a solution of $\min_{\tilde{s}} \frac{s_0}{\mu} + \tilde{\Psi}(\tilde{s})$ – *and* near the optimal solution $\bar{s}_q$.

For a theoretical analysis of an interior point method it suffices to use the self-concordance conditions for the convergence of Newtons method only in a neighborhood of the central path, and to bound the distance from optimality only along straight lines connecting the current iterate with an optimal solution, see e.g. Lemma 2.11 and Section 3.1 in [9]. Thus, the local self-concordance provides the theoretical basis for a polynomial rate of convergence – provided a starting point is given that lies close to the central path and to the optimal solution.

The following lemma is a helpful consequence of the self-concordance properties.

Lemma 3 *Let $\delta \in (0, 1)$ be given and let $\tilde{s}$ be a stricly feasible point of problem (36) such that the condition*

$$-\frac{h^T H h}{-\tilde{q}(\tilde{s})} \leq (1 - \delta) \|(\hat{G}[s])^{-\frac{1}{2}} DG[h] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2, \tag{40}$$

holds for all $h \in \mathbb{R}^n$. Then there exist $\alpha \geq 1$, $\beta \geq 1$ depending only on δ such that the self-concordance condition (39) holds true.

Proof:

Let us fix a point $\tilde{s} = (s_0, s)$ satisfying (40) and certain arbitrary $\tilde{h} = (h_0, h) \in \mathbb{R}^{n+1}$. By [13], Proposition 5.4.5, the inequalities (39) are satisfied with $\alpha = 1$ and $\beta = m+1$ for all feasible points $\tilde{s}$ and all vectors $\tilde{h} \in \mathbb{R}^{n+1}$ when $H = 0$. If we take the particular selection $\check{s} := (s_0 + \frac{1}{2}s^T Hs; s)$ (as strictly feasible point of the problem (36) for $H = 0$) and $\check{h} := (h_0 - s^T Hh; h) \in \mathbb{R}^{n+1}$ as the direction and apply the self-concordance for the case $H = 0$, i.e.

$$l'''_{\check{s},\check{h},0}(0) \leq 2(l''_{\check{s},\check{h},0}(0))^{\frac{3}{2}} \qquad \text{and} \qquad (l'_{\check{s},\check{h},0}(0))^2 \leq (m+1)l''_{\check{s},\check{h},0}(0), \tag{41}$$

then we obtain the inequalities

$$\left(\frac{b_s^T h - h_0}{-\tilde{q}(\tilde{s})} - (\hat{G}[s])^{-1} \bullet DG[h]\right)^2 \leq (m+1)\left(\frac{(b_s^T h - h_0)^2}{(\tilde{q}(\tilde{s}))^2} + \|(\hat{G}[s])^{-\frac{1}{2}} \bullet DG[h](\hat{G}[s])^{-\frac{1}{2}}\|_F^2\right) \tag{42}$$

and

$$2\frac{(b_s^T h - h_0)^3}{(-\tilde{q}(\tilde{s}))^3} + D^3\tilde{\Phi}(\tilde{s})[\tilde{h}, \tilde{h}, \tilde{h}] \leq 2\left(\frac{(b_s^T h - h_0)^2}{(\tilde{q}^2(\tilde{s}))^2} + \|(\hat{G}[s])^{-\frac{1}{2}} \bullet DG[h](\hat{G}[s])^{-\frac{1}{2}}\|_F^2\right)^{\frac{3}{2}}. \tag{43}$$

If we write down the above condition (40) as follows

$$\|(\hat{G}[s])^{-\frac{1}{2}} DG[h](\hat{G}[s])^{-\frac{1}{2}}\|_F^2 \leq \frac{1}{\delta}\left(\frac{h^T Hh}{-\tilde{q}(\tilde{s})} + \|(\hat{G}[s])^{-\frac{1}{2}} DG[h](\hat{G}[s])^{-\frac{1}{2}}\|_F^2\right), \tag{44}$$

it can be easily noted that the right-hand sides of (42) and (43) can be bounded by $\beta l''(0)$ and $2\alpha(l''(0))^{\frac{3}{2}}$, respectively, for suitable selections of $\beta, \alpha > 0$. The second inequality of (39) follows straightforwardly, since $(l'(0))^2$ coincides with the left-hand side of (42).

Let us also note that $l'''(0)$ is equal to the left hand side of (43), except from the term $\frac{h^T Hh}{-\tilde{q}(\tilde{s})} \frac{(b_s^T h - h_0)}{-\tilde{q}(\tilde{s})}$. To obtain the first equation in (39) it suffices to show that

$$\left|\frac{h^T Hh}{-\tilde{q}(\tilde{s})} \frac{(b_s^T h - h_0)}{-\tilde{q}(\tilde{s})}\right| \leq 2\alpha(l''(0))^{\frac{3}{2}}.$$

This condition is implied by the following inequalities

$$\left|\frac{h^T Hh}{-\tilde{q}(\tilde{s})}\right| \leq 2\alpha\left(\frac{(b_s^T h - h_0)^2}{(\tilde{q}(\tilde{s}))^2} + \frac{h^T Hh}{-\tilde{q}(\tilde{s})} + \|(\hat{G}[s])^{-\frac{1}{2}} DG[h](\hat{G}[s])^{-\frac{1}{2}}\|_F^2\right)$$

and

$$\left|\frac{(b_s^T h - h_0)}{-\tilde{q}(\tilde{s})}\right| \leq 2\alpha\left(\frac{(b_s^T h - h_0)^2}{(\tilde{q}(\tilde{s}))^2} + \frac{h^T Hh}{-\tilde{q}(\tilde{s})} + \|(\hat{G}[s])^{-\frac{1}{2}} DG[h](\hat{G}[s])^{-\frac{1}{2}}\|_F^2\right)^{1/2}$$

which are both easily obtained from (44) by selecting again a suitable $\alpha > 0$. ∎

5.2 Local self-concordance

The aim of this section is to establish that $\tilde{\Psi}$ satisfies a certain form of self-concordance conditions when $\tilde{s}$ lies near the central path and near the optimal solution $\bar{s}$ of (36). Let us first consider the self-concordance exactly on the central path.

Theorem 2 *Let $(\bar{x}, \bar{Y}, \bar{S})$ be a fixed stationary point of (5) satisfying the general assumptions* ***A1-A3****. Consider some fixed (x, Y) sufficiently close to $(\bar{x}, \bar{Y})$ such that the stationary solution $(\bar{s}_q, \bar{Y}_q, \bar{S}_q)$ of the corresponding problem (34) also fulfills the general assumptions. Then, the self-concordance conditions (39) are satisfied on the points of the central path of (36) sufficiently close to the solution $\bar{s} = (q(\bar{s}_q), \bar{s}_q)$.*

Proof:

To simplify the notation we do not distinguish between $\bar{Y}$ and $\bar{Y}_q$; $\bar{S}$ and $\bar{S}_q$ or between $C(\bar{s}_q)$, $C(x)$ and $C(\bar{x})$. Note that for any point $\tilde{s}$ on the central path and near the optimal solution $\bar{s}$ of (36), the first component of $\nabla\tilde{\Psi}(\tilde{s})$ is $\frac{-1}{-\tilde{q}(\tilde{s})}$ and therefore, $-\tilde{q}(\tilde{s}) \equiv \mu$.

In analogy to the decomposition (10) let the matrix U be given such that

$$U_1^T(G(x) + DG(x)[\bar{s}_q])U_1 = 0 \quad \text{and} \quad U_2^T\bar{Y}U_2 = 0. \tag{45}$$

The following equations are a straightforward consequence of this decomposition and will be used in the sequel:

Let $P := U_1U_1^T$, then P is a projection, $P^2 = P$, and $\bar{P} := I - P = U_2U_2^T$ is a projection as well satisfying the equations

$$P\bar{S} = \bar{S}P = 0 = \bar{P}\bar{Y} = \bar{Y}\bar{P}, \qquad \bar{P}\bar{S} = \bar{S}\bar{P} = \bar{S}, \qquad P\bar{Y} = \bar{Y}P = \bar{Y}.$$

Using these relations the critical cone can be written as

$$C(x) = \{u \mid PDG[u]P = 0\},$$

where $DG[u] = D(x)G[u]$ according to the notation (38). The second order sufficient condition (35) is given then by the inequality

$$-u^THu \leq 2(1-\epsilon)\,\mathrm{trace}(\bar{Y}PDG[u]\bar{P}\bar{S}^\dagger\bar{P}DG[u]P)$$

for all $u \in C^\rho(\bar{x})$.

On the central path we have $YS = \mu I$ where $S = -G(x) - DG(x)[s]$, and the matrices Y, S converge to $\bar{Y}, \bar{S}$ for $\mu \to 0$,

$$Y = \bar{Y} + \Delta Y, \qquad S = \bar{S} + \Delta S \qquad \text{with} \quad \|\Delta Y\|, \|\Delta S\| = O(\mu).$$

Using again the notation (38) we denote $\hat{G}[s] := G(x) + DG(x)[s] = -S$. The above equations imply

$$(\hat{G}[s])^{-1} = -\frac{\bar{Y} + \Delta Y}{\mu}. \tag{46}$$

According to Lemma 3, it suffices to show the condition (40). Let $\mathcal{V} := \overline{\mathbb{R}^n \backslash C^\rho(x)}$ where ρ is as in Lemma 2. $\mathcal{V}$ is a closed cone such that $\mathcal{V} \cap C(x) = \{0\}$. We establish (40) separately for $h = u$ with $u \in C^\rho(x)$ and $h = v$ with $v \in \mathcal{V}$.

For $v \in \mathcal{V}$ with $\|v\| = 1$, it follows that $PDG[v]P \neq 0$, and hence

$$\|PDG[v]P\|_F^2 \geq \tilde{\epsilon}\|v\|^2$$

for some $\tilde{\epsilon} > 0$ independent of v. Since $\mathcal{V}$ is a cone, the above relation is true for all $v \in \mathcal{V}$. As $\bar{Y}$ is positive definite on the range space of P, it follows that

$$\|\bar{Y}^{\frac{1}{2}}PDG[v]P\bar{Y}^{\frac{1}{2}}\|_F^2 \geq \epsilon\|v\|^2$$

for some $\epsilon > 0$ independent of v. From (46) we obtain

$$\|\bar{Y}^{\frac{1}{2}} DG[v] \bar{Y}^{\frac{1}{2}}\|_F^2 \approx \mu^2 \|(\hat{G}[s])^{-\frac{1}{2}} DG[v] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2.$$

In particular,

$$\epsilon \|v\|^2 \leq 2\mu^2 \|(\hat{G}[s])^{-\frac{1}{2}} DG[v] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2$$

for small values of $\mu > 0$. Since $-\tilde{q}(\tilde{s}) = \mu$ on the central path we obtain

$$-\frac{v^T H v}{-\tilde{q}(\tilde{s})} \leq \frac{\|H\|}{\mu} \frac{2\mu^2}{\epsilon} \|(\hat{G}[s])^{-\frac{1}{2}} DG[v] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2 \leq (1-\delta) \|(\hat{G}[s])^{-\frac{1}{2}} DG[v] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2$$

whenever $\mu < \epsilon(1-\delta)/(2\|H\|)$.

We now turn to $u \in C^\rho(x)$. For Y near $\bar{Y}$ the eigenspace of the nonzero eigenvalues of $\bar{Y}$ depends continuously on ΔY, so that $U^T Y U$ is approximately a 2×2 block diagonal matrix with blocks conforming the partition in (10), and likewise for $U^T S U$. More precisely, the norm of the off-diagonal blocks of $U^T Y U$ and $U^T S U$ is of the order $\|\Delta Y\| + \|\Delta S\|$.

Denote the diagonal blocks of $U^T Y U$ by D_1 and D_2, and the diagonal blocks of $U^T S U$ by Λ_1 and Λ_2. By continuity, Λ_2 and D_1 contain the large elements, and the relation $YS = \mu I$ implies $D_2 \approx \mu \Lambda_2^{-1}$. From $-(\hat{G}[s])^{-1} = \frac{Y}{\mu}$ on the central path we therefore obtain for $u \in C^\rho(x)$ that

$$\begin{aligned} \|(\hat{G}[s])^{-\frac{1}{2}} D[u] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2 &= \|U^T (\hat{G}[s])^{-\frac{1}{2}} U U^T DG[u] U U^T (\hat{G}[s])^{-\frac{1}{2}} U\|_F^2 \\ &\approx \|\operatorname{Diag}(D_1/\mu, \Lambda_2^{-1})^{\frac{1}{2}} U^T DG[u] U \operatorname{Diag}(D_1/\mu, \Lambda_2^{-1})^{\frac{1}{2}}\|_F^2. \end{aligned}$$

The matrix $U^T DG[u] U$ can also be partitioned into a 2×2 block matrix with blocks conforming the partition in (45),

$$U^T DG[u] U = \begin{bmatrix} DG_{11} & DG_{12} \\ DG_{12}^T & DG_{22} \end{bmatrix}.$$

Straightforward reformulations then show

$$\begin{aligned} \|(\hat{G}[s])^{-\frac{1}{2}} DG[u] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2 &\geq 2 \|(D_1/\mu)^{\frac{1}{2}} DG_{12} \Lambda_2^{-\frac{1}{2}}\|_F^2 \\ &\approx \frac{2}{\mu} \bar{Y} \bullet DG[u] \bar{S}^\dagger DG[u] \geq -\frac{u^T H u}{-\tilde{q}(\tilde{s})(1-\epsilon)} \end{aligned}$$

where we use (35) and the relation $-\tilde{q}(\tilde{s}) = \mu$. For sufficiently small $\mu > 0$ (say $\mu \leq \hat{\mu}$) we may therefore bound

$$-\frac{u^T H u}{-\tilde{q}(\tilde{s})} < (1 - \frac{\epsilon}{2}) \|(\hat{G}[s])^{-\frac{1}{2}} DG[u] (\hat{G}[s])^{-\frac{1}{2}}\|_F^2.$$

Without loss of generality we assume that $\hat{\mu} \leq \epsilon(1-\delta)/(2\|H\|)$. Defining $\delta = \epsilon/2$ we have thus established (40) for $h = u$ and $h = v$ separately whenever $\mu \leq \hat{\mu}$. ■

Remark 4 *Theorem 2 established the self-concordance relations (39) for points $\tilde{s}$ on the central path for small values of $\mu > 0$. Replacing δ in (40) with $\delta/2$, Relation (40) can still be maintained for points near the central path, more precisely for all points $\tilde{s} + \Delta\tilde{s}$ where $\tilde{s}$ lies on the central path and*

$$\|(\hat{G}[\tilde{s}])^{-\frac{1}{2}} DG[\Delta\tilde{s}] (\hat{G}[\tilde{s}])^{-\frac{1}{2}}\|_F \leq \hat{\epsilon} \tag{47}$$

for some fixed (small) $\hat{\epsilon} \in (0, 1)$.

Proof: For $-\tilde{q}(\tilde{s}) = \mu$, relation (40) is established in Theorem 2. The relaxation $\delta \to \delta/2$ compensates for small relative changes in both sides of (40). We recall that the function $\tilde{\Psi}$ is the sum of the self-concordant function $\tilde{\Phi}$ and the logarithmic barrier term $-\log(-\tilde{q}(\ .\ ))$. Note that the right hand side of (40) is just $D^2\tilde{\Phi}(\tilde{s})[\tilde{h}, \tilde{h}]$ multiplied by $(1 - \delta)$. The second derivative of $\tilde{\Phi}$ satisfies a relative Lipschitz condition ([13], Theorem 2.1.1) where the norm of the perturbation is given by the left hand side of (47). The relative change of the right hand side of (40) can thus be bounded (arbitrarily close to 1) when $\hat{\epsilon} \in (0, 1)$ is sufficiently small. Likewise, within a sufficiently small inner Dikin ellipsoid defined by (47) the relative change of the left hand side of (40) is bounded as well. ■

As a phase-1-problem to find a starting point near the central path the objective function $\tilde{q}(\tilde{s})$ can be replaced with 0, and the condition $\tilde{q}(0) + D\tilde{q}(0)\tilde{s} \leq r$ for some $r \geq 0$ can be added.

The analysis of the computation of starting points satisfying the assumptions of Remark 4 based on convex SDP-subproblems is beyond the scope of this paper. Numerical results in [8, 21] however showed good global convergence of such an approach. The present paper aims at local properties but does not provide a complete analysis.

6 Discussion

The above analysis established that the nonconvex semidefinite program locally possesses some form of convexity property that allows for a concept of local self-concordance. It is known [13] that for any closed convex set (with nonempty interior) there exists a barrier function whose self-concordance parameter depends only on the dimension of the set. This independence of any other problem specific parameters makes the concept of self-concordance so strong, and it contrasts the result established in Section 5.2, where the self-concordance-parameters do depend on the problem (i.e. on $\epsilon \in (0, 1]$), and not merely on the dimension.

Here, we apply the self-concordance to a "genuinely" nonconvex problem, where the parameter ϵ of the second order condition specifies in how far the "nonconvexity" of the problem is locally dominated by the curvature of the semidefinite cone. For $\epsilon = 0$, the solution may no longer be unique; in fact when $\epsilon = 0$, the set of optimal solutions may be nonconvex, and therefore any barrier function will not be convex, in particular, not self-concordant, even locally. The domain of parameters ϵ which—in principle—permit self-concordant barriers is an open set ($\epsilon > 0$), and it is therefore not surprising that the self-concordance-parameters depend on $1/\epsilon$.

Acknowledgement

The third author thanks A. Shapiro for pointing to the weaker second order condition in [8]. This work has been supported by Conicyt Chile under the grants Nr. 1050696 (Fondecyt regular) and Nr. 7070326 (Incentivo a la Cooperación Internacional).

References

[1] Bonnans, F., Ramírez C., H. (2005): Strong regularity of semidefinite programming problems. Informe Técnico, DIM-CMM, Universidad de Chile, Santiago, Departamento de Ingeniería Matemática, No CMM-B-05/06-137

[2] Bonnans, F., Shapiro, A. (2000): *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research

[3] Bjork, A. (1996): Numerical methods for least squares problems, SIAM, Society for Industrial and Applied Mathematics, Philadelphia

[4] Correa, R., Ramírez C., H. (2004): A global algorithm for nonlinear semidefinite programming. SIAM J. Optim. **15**, 303–318

[5] Diehl, M., Jarre, F., Vogelbusch, C.H., (2006): Loss of Superlinear Convergence for an SQP-type method with Conic Constraints, SIAM Journal on Optimization, 1201–1210

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

[7] Freund, R.W., Jarre, F. (2004): A sensitivity result for semidefinite programs. Oper. Res. Lett. **32**, 126–132

[8] Freund, R.W., Jarre, F., Vogelbusch, C. (2007): Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. Math. Program. 109(2-3), 581–611.

[9] Jarre, F. (1994): Interior point methods via self-concordance or relative Lipschitz condition, Habiltitationsschrift, http://www.opt.uni-duesseldorf.de/en/forschung-fs.html

[10] Kocvara, M.; Stingl, M. (2004): Solving nonconvex SDP problems of structural optimization with stability control. Optim. Methods Softw. 19(5), 595–609

[11] Kocvara, M.; Stingl, M. (2003): Pennon: A code for convex nonlinear and semidefinite programming. Optim. Methods Softw. 18(3), 317–333

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

[13] Nesterov, Y., Nemirovskii, A. (1994): Interior-point polynomial algorithms in convex programming, SIAM Studies in Applied Mathematics, SIAM.

[14] Noll, D. (2007): Local convergence of an augmented Lagrangian method for matrix inequality constrained programming, Optimization Methods and Software, 22(5), 777-802

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

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

[17] Shapiro, A. (1997): First and second order analysis of nonlinear semidefinite programs. Math. Program. **77**, 301–320

[18] Shapiro, A. (2006): personal communication.

[19] M. Stingl (2006): On the Solution of Nonlinear Semidefinite Programs by Augmented Lagrangian Methods. Dissertation, Shaker Verlag, Aachen.

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

[21] Vogelbusch, C. (2007): Numerical Treatment of Nonlinear Semidefinite Programs, Dissertation, Heinrich-Heine University, Düsseldorf, Germany
http://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=3652