Elementary optimality conditions for nonlinear SDPs

Florian Jarre
University Düsseldorf, Germany

December 6, 2010

Abstract

An increasing number of recent applications rely on the solution of nonlinear semidefinite programs. First and second order optimality conditions for nonlinear programs are widely known today. This paper presents a self-contained generalization of these optimality conditions to nonlinear semidefinite programs, highlighting some parallels and some differences. It starts by discussing a constraint qualification for both programs. First order optimality conditions are presented for the case where this constraint qualification is satisfied. For the second order conditions, in addition, strict complementarity is assumed.

Keywords: Nonlinear semidefinite program, constraint qualification, first and second order optimality conditions.

1 Introduction

Recent applications in areas such as free material optimization [6], robust control [4], passive reduced order modeling [5], combinatorial optimization [9], and others rely on the solution of nonlinear semidefinite programs. An implementation for solving such problems is presented in [7], for example.

We derive a generalization of the optimality conditions for nonlinear programs to nonlinear semidefinite programs, highlighting some parallels and some differences. Aim of this paper is to condense and simplify the more general results in the book by Bonnans and Shapiro [2].

In Section 2.1 the problems and notation are introduced. Then, constraint qualifications for both programs are introduced. First order optimality conditions are presented for the case where a constraint qualification is satisfied. Finally, second order conditions are presented, where in addition, strict complimentarity is assumed.

2 First order conditions

2.1 NLSDPs and NLPs

This paper concerns nonlinear semidefinite optimization problems of the form

$$\text{minimize} \quad f(x) \quad | \quad \begin{array}{l} F(x) = 0, \\ G(x) \preceq 0. \end{array} \tag{1}$$

Throughout, we assume that $f : I\!R^n \to I\!R$, $F : I\!R^n \to I\!R^m$, and $G : I\!R^n \to \mathcal{S}^d$ are continuously differentiable. The inequality $G(x) \preceq 0$ means that the matrix $G(x)$ is required to be negative semidefinite. $\mathcal{S}^d$ stands for the space of real symmetric $d \times d$-matrices, and $\mathcal{S}^d_+$ denotes the cone of positive semidefinite matrices in $\mathcal{S}^d$. The scalar product of two matrices is denoted by $A \bullet B = \text{trace}(A^T B) = \sum_{i,j} A_{i,j} B_{i,j}$.

For comparison we also consider nonlinear optimization problems of the form

$$\text{minimize} \quad f(x) \quad | \quad \begin{array}{l} F(x) = 0, \\ G(x) \leq 0, \end{array} \tag{2}$$

with componentwise[1] inequalities $G(x) \leq 0$.

This section will emphasize the differences between NLSDPs of the form (1) and NLPs of the form (2). In particular, we emphasize two major points:

- First, the relation of tangential and linearized cones does not translate in a straightforward fashion from (2) to (1).
- And second, at an optimal solution, the curvature in the Lagrangian "differs substantially" in (2) and in (1).

The derivatives of f and F at a point x are, as usual, represented by a row vector $Df(x)$ and the $m \times n$ Jacobian matrix $DF(x)$. The entries of the vector $F(x)$ will be denoted by $F_\nu(x)$ for $1 \leq \nu \leq m$ and the first and second derivatives of F_ν by $DF_\nu(x)$ and $D^2F_\nu(x)$. Thus, $DF_\nu(x)$ is the ν-th row of $DF(x)$.

The derivative of G at a point x is a linear map $DG(x) : I\!R^n \to \mathcal{S}^d$. When applying the linear map $DG(x)$ to a vector Δx we use the notation $DG(x)[\Delta x] \in \mathcal{S}^d$, or

$$DG(x)[\Delta x] = \sum_{i=1}^{n} \Delta x_i G_i(x) \quad \text{where} \quad G_i(x) := \frac{\partial}{\partial x_i} G(x) \in \mathcal{S}^d.$$

[1]Assuming that $G(x)$ is a symmetric matrix, and that the "off-diagonal inequalities" are thus listed twice among the inequalities $G(x) \leq 0$, introduces some complication in the discussion of nondegeneracy for (2). Below, we will work with Lagrange multipliers being symmetric as well, thus "counting" the off-diagonal inequalities just once. To make things short, we may simply ignore the fact that formally, (2) contains redundant constraints.

$DG(x)$ satisfies the characteristic equation $G(x+\Delta x) \approx G(x) + DG(x)[\Delta x]$. The matrix entries of $G_i(x)$ will be referred to by $(G_i(x))_{k,l}$. As a short hand notation we will write

$$(DG(x)[\Delta x])_{k,l} = \sum_{i=1}^{n} (G_i(x))_{k,l} \Delta x_i =: (DG(x))_{k,l} \Delta x$$

where $(DG(x))_{k,l}$ is a row vector with entries $(G_i(x))_{k,l}$.

2.2 Linearized subproblems

Assume that some point $\bar{x} \in I\!R^n$ is given. Linearizing f, F, and G in (1) about $\bar{x}$ yields the *linearized semidefinite programming problem:*

$$\text{minimize} \quad f(\bar{x}) + Df(\bar{x})\Delta x \quad | \quad \begin{array}{l} F(\bar{x}) + DF(\bar{x})\Delta x = 0, \\ G(\bar{x}) + DG(\bar{x})[\Delta x] \preceq 0. \end{array} \tag{3}$$

Here, we maintain the cone constraint "$\preceq 0$" and do not use a linearization involving the tangential cone to the set of negative semidefinite matrices. This allows a direct analysis of SQP type methods and of optimality conditions as discussed below.

Likewise, for (2), we obtain the linearized problem

$$\text{minimize} \quad f(\bar{x}) + Df(\bar{x})\Delta x \quad | \quad \begin{array}{l} F(\bar{x}) + DF(\bar{x})\Delta x = 0, \\ G(\bar{x}) + DG(\bar{x})[\Delta x] \le 0. \end{array} \tag{4}$$

We recall that the first order optimality conditions of (1) or of (2) simply refer to the problems (3) or (4). It is therefore crucial to understand the relation of (1) and (3).

2.3 Tangential and linearized cones

We denote the feasible set of the NLSDP (1) by $\mathcal{F}_1$ and recall that the tangential cone of $\mathcal{F}_1$ at a point $\bar{x} \in \mathcal{F}_1$ is given by

$$\mathcal{T}_1 := \{\Delta x \mid \exists s^k \to \Delta x, \;\; \exists \alpha_k > 0, \; \alpha_k \to 0 \;\text{ such that }\; \bar{x} + \alpha_k s^k \in \mathcal{F}_1\}. \tag{5}$$

Loosely speaking, if the feasible set is smooth in a neighborhood of $\bar{x}$, the set $\mathcal{T}_1$ may be viewed as the set of all directions Δx such that $\bar{x} + \epsilon \Delta x$ is feasible for small $\epsilon > 0$ or almost feasible. It is straightforward to verify that $Df(\bar{x})\Delta x \ge 0$ must hold true for all $\Delta x \in \mathcal{T}_1$ if $\bar{x}$ is a local minimizer of (1).

We denote the tangential cone of (3) at $\overline{\Delta x} = 0$ by $\mathcal{L}_1$. $\mathcal{L}_1$ is the

linearized cone

of the NLSDP (1) at $\bar{x}$. As before, $Df(\bar{x})\Delta x \ge 0$ must hold true for all $\Delta x \in \mathcal{L}_1$ if $\bar{x}$ is a minimizer of (3).

Thus, the relation of the first order optimality conditions of (1) and (3) depends on the relation of $\mathcal{T}_1$ and $\mathcal{L}_1$.

For illustration we return to the NLP (2) and denote the feasible set of (2) by $\mathcal{F}_2$ and the tangential cone at a point $\bar{x} \in \mathcal{F}_2$ by $\mathcal{T}_2$. Likewise, let $\mathcal{L}_2$ be the linearized cone of the NLP (2) at a point $\bar{x}$. The linearized cone is thus given by

$$\begin{aligned}\mathcal{L}_2 = \{\Delta x \quad | \quad & F(\bar{x}) + DF(\bar{x})\Delta x = 0, \\ & G_{k,l}(\bar{x}) + (DG(\bar{x}))_{k,l}\Delta x \leq 0 \ \text{ for all } k,l \text{ with } G_{k,l}(\bar{x}) = 0\}.\end{aligned} \tag{6}$$

A key argument in nonlinear optimization is based on the fact that $\mathcal{T}_2 \subset \mathcal{L}_2$ always holds true. This fact can be established by a simple argument: For $\Delta x \in \mathcal{T}_2$ we have by definition (5)

$$0 = F(\bar{x} + \alpha_k s^k) = F(\bar{x}) + \alpha_k DF(\bar{x}) s^k + o(\alpha_k).$$

Dividing this by $\alpha_k > 0$ and using the fact that $F(\bar{x}) = 0$ yields

$$0 = DF(\bar{x}) s^k + o(1),$$

and taking the limit as $k \to \infty$ yields $0 = DF(\bar{x})\Delta x$, i.e. Δx satisfies the equality constraints in the definition of $\mathcal{L}_2$. Applying the same argument to the active inequalities $G_{k,l}(x) \leq 0$ for all k,l with $G_{k,l}(\bar{x}) = 0$, we obtain that

$$\mathcal{T}_2 \subset \mathcal{L}_2 \tag{7}$$

for NLPs of the form (2).

Constraint qualifications for NLP

- The converse direction $\mathcal{L}_2 \subset \mathcal{T}_2$ can be shown, for example, when the MFCQ-constraint qualification for (2),

$$DF(\bar{x}) \text{ has linearly independent rows} \tag{8}$$

$$\begin{aligned}&\exists \Delta x \text{ such that } DF(\bar{x})\Delta x = 0 \text{ and} \\ &\quad (DG(\bar{x}))_{k,l}\Delta x < 0 \quad \text{for all} \quad (k,l) \text{ with } (G(\bar{x}))_{k,l} = 0,\end{aligned} \tag{9}$$

  is satisfied (see [8, 12, 13] or [11] Definition 12.5). In this case, since the above characterization of $\mathcal{L}_2$ is polyhedral, also the set

$$\mathcal{T}_2 \text{ is polyhedral.}$$

- For later reference we also recall the linear independence constraint qualification LICQ for problem (2),

$$DF_\nu(\bar{x}) \text{ and } (DG(\bar{x}))_{k,l} \text{ are linearly independent} \tag{10}$$

  for $1 \leq \nu \leq m$ and (k,l) with $k \leq l$ and $(G(\bar{x}))_{k,l} = 0$. It is easy to see that (10) implies (8), (9).

Note that the semidefinite constraint $G(x) \preceq 0$ can be represented by an exponential (but finite!) number of smooth nonlinear inequality constraints, namely that the determinants of all principal submatrices of "$-G(x)$" be nonnegative. Thus, one might expect, that the tangential cone $\mathcal{T}_1$ is also a polyhedral cone. As we will see, this is not the case! (The determinant will not satisfy MFCQ (9) when zero is an eigenvalue of multiplicity more than one.)

We derive a representation of $\mathcal{T}_1$: Let $\bar{x} \in \mathcal{F}_1$ be given and, as in Definition (5), a sequence $\alpha_k > 0$ with $\alpha_k \to 0$, and a sequence $s^k \to \Delta x$, such that $F(\bar{x} + \alpha_k s^k) = 0$ and $G(\bar{x} + \alpha_k s^k) \preceq 0$ for all k.

As before it follows that $DF(\bar{x})\Delta x = 0$. Let

$$G(\bar{x}) = U \Lambda U^T = \begin{bmatrix} U^{(1)} & U^{(2)} \end{bmatrix} \begin{bmatrix} \Lambda^{(1)} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (U^{(1)})^T \\ (U^{(2)})^T \end{bmatrix}$$

be the eigenvalue decomposition of $G(\bar{x})$ where $\Lambda^{(1)}$ is negative definite. Using the Schur complement it is straightforward to verify that the tangential cone of "$-\mathcal{S}^d_+$" at $G(\bar{x})$ is given by all matrices of the form

$$W = \begin{bmatrix} U^{(1)} & U^{(2)} \end{bmatrix} \begin{bmatrix} * & * \\ * & \tilde{W}^{(2)} \end{bmatrix} \begin{bmatrix} (U^{(1)})^T \\ (U^{(2)})^T \end{bmatrix} \tag{11}$$

where $\tilde{W}^{(2)} \preceq 0$ and "$*$" can be anything. We keep U fixed and define $\tilde{G}(x) := U^T G(x) U$. (Then, $G(x) \preceq 0$ iff $\tilde{G}(x) \preceq 0$.) We partition

$$\tilde{G}(x) := \begin{bmatrix} \tilde{G}^{(1)}(x) & \tilde{G}^{(1,2)}(x) \\ \tilde{G}^{(1,2)}(x)^T & \tilde{G}^{(2)}(x) \end{bmatrix}. \tag{12}$$

conforming the partition in (11). (Thus, $\tilde{G}^{(1)}(\bar{x})$ and $\Lambda^{(1)}$ coincide and $\tilde{G}^{(1,2)}(\bar{x}) = 0$, $\tilde{G}^{(2)}(\bar{x}) = 0$.) Since $\tilde{G}(\bar{x} + \alpha_k s^k) \preceq 0$ it follows that

$$0 \succeq \tilde{G}^{(2)}(\bar{x} + \alpha_k s^k) - \underbrace{\tilde{G}^{(2)}(\bar{x})}_{=0} \approx \alpha_k D\tilde{G}^{(2)}(\bar{x})[s^k].$$

Dividing by $\alpha_k > 0$ and taking the limit as $k \to \infty$, we get

$$D\tilde{G}^{(2)}(\bar{x})[\Delta x] \preceq 0,$$

i.e., by (11), $G(\bar{x})[\Delta x]$ lies in the tangential cone of "$-\mathcal{S}^d_+$" at $G(\bar{x})$. The above implies that

$$\mathcal{T}_1 \subset \{\Delta x \mid DF(\bar{x})\Delta x = 0, \quad D\tilde{G}^{(2)}(\bar{x})[\Delta x] \preceq 0\} \tag{13}$$

in analogy to (7). By linearizing the equality constraints and the "active part" of the inequalities, the set on the right hand side appears to be a straightforward generalization of the definition of $\mathcal{L}_2$ in (6). It turns out, however, that the set on the right hand side is not necessarily $\mathcal{L}_1$:

The example $x \in I\!R$,

$$G(x) := \begin{bmatrix} -2 & x \\ x & -x^2 \end{bmatrix} \preceq 0$$

shows that $\mathcal{F}_1 = I\!R$, and thus, also $\mathcal{T}_1 = I\!R$, while the feasible set of (3) at $\bar{x} = 0$ is given by

$$\left\{ \Delta x \mid \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \Delta x \\ \Delta x & 0 \end{bmatrix} \preceq 0 \right\} = \{0\}. \tag{14}$$

Hence, also the linearized cone is given by $\mathcal{L}_1 = \{0\}$ and does not contain $\mathcal{T}_1$ in contrast to (7).

2.4 A constraint qualification for NLSDP

We return to the NLSDP (1). The MFCQ condition (8), (9) allows a straightforward extension: Problem (1) satisfies Robinson's constraint qualification [12] (or a generalized MFCQ) at a point $\bar{x} \in \mathcal{F}$, if

$$\begin{aligned} &DF(\bar{x}) \text{ has linearly independent rows, and there exists a vector } d \text{ such that} \\ &DF(\bar{x})d = 0 \text{ and} \\ &G(\bar{x}) + DG(\bar{x})[d] \prec 0. \end{aligned} \tag{15}$$

The last two lines of (15) coincide with Slater's condition for (3).

The modified example $x \in I\!R^2$, and

$$G(x) := \begin{bmatrix} -2 & x_1 \\ x_1 & x_2 - x_1^2 \end{bmatrix} \preceq 0$$

coincides with the above example (leading to (14)) when x_2 is set to 0. However, this example satisfies (15) at $\bar{x} = (0,0)^T$, and the feasible set of (3) at $\bar{x} = 0$ is given by

$$\left\{ \Delta x \mid \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \Delta x_1 \\ \Delta x_1 & \Delta x_2 \end{bmatrix} \preceq 0 \right\} = \{\Delta x \mid \Delta x_1^2 \le -2\Delta x_2\}.$$

Straightforward calculations show that the tangential cone of this set is given by $\mathcal{L}_1 = \{\Delta x \mid \Delta x_2 \le 0\}$ and that $\mathcal{T}_1 = \mathcal{L}_1$. This is true more generally, as is made explicit in the next lemma.

Lemma 1 *If the constraint qualification (15) is satisfied, then $\mathcal{L}_1 = \mathcal{T}_1$ coincides with the right hand side of (13).*

Lemma 1 establishes the earlier remark that the tangential cone of an NLSPD is not polyhedral, in general, in contrast to the tangential cone of an NLP. This concludes our discussions of the first "major point" referred to in Section 2.1.

Lemma 1 and the above discussions are known for a long time, see the results of Robinson, [12], but they may not be widely known. As conic optimization has returned to the focus of research following the generalization of interior-point methods to semidefinite programming, [10, 1, 15] these conditions now become of wider general interest.

2.5 KKT conditions

For the remainder of this paper we assume that the constraint qualification (15) at $\bar{x}$ is always satisfied. We assume that $\bar{x}$ is a local minimizer of (1). Lemma 1 implies that the KKT conditions of (3) for $\Delta x = 0$ can be translated to (1). This yields the following first order conditions:

Theorem 1 *Let $\bar{x}$ be a local minimizer of (1) and let (1) be regular at $\bar{x}$ in the sense of (15). Then there exists a matrix $\bar{Y} \in \mathcal{S}^d_+$ and a vector $\bar{y} \in I\!R^m$ such that*

$$Df(\bar{x})^T + DF(\bar{x})^T \bar{y} + \begin{bmatrix} G_1(\bar{x}) \bullet \bar{Y} \\ \vdots \\ G_m(\bar{x}) \bullet \bar{Y} \end{bmatrix} = 0 \quad \textit{and} \quad G(\bar{x}) \bullet \bar{Y} = 0. \tag{16}$$

□

The Lagrangian function underlying Theorem 1 is given by

$$L(x, y, Y) := f(x) + F(x)^T y + G(x) \bullet Y \tag{17}$$

with $Y \in \mathcal{S}^d_+$, and the first equation in (16) states that the derivative $D_x L(\bar{x}, \bar{y}, \bar{Y})$ of the Lagrangian with respect to x is zero at the optimal solution,

$$D_x L(\bar{x}, \bar{y}, \bar{Y}) = 0.$$

The Lagrangian for the NLP (2) is identical to (17) except that we have $Y \in \mathcal{S}^d$, $Y \geq 0$ for (2). The fact that the Lagrangian of (1) and of (2) is the same function (just with a different domain for Y) leads us to the second crucial difference of the NLSDP (1) and the NLP (2).

3 Second order conditions

In this section we assume that f, F, and G are all twice continuously differentiable. Here, the second derivative $D^2 G$ at a point $x \in I\!R^n$ is a symmetric bilinear mapping,

$$D^2 G(x) : I\!R^n \times I\!R^n \to \mathcal{S}^d, \qquad D^2 G(x)[\Delta x, \Delta x] = \sum_{i,j=1}^{n} \Delta x_i \Delta x_j G_{i,j}(x) \in \mathcal{S}^d$$

with $G_{i,j}(x) := \frac{\partial^2}{\partial x_i \partial x_j} G(x)$, and for $Y \in \mathcal{S}^d$ we have

$$Y \bullet D^2 G(x)[\Delta x, \Delta x] = \sum_{i,j=1}^{n} \Delta x_i \Delta x_j (Y \bullet G_{i,j}(x)) = \Delta x^T H \Delta x,$$

where $H = H(x, Y)$ is the symmetric $n \times n$-matrix with matrix entries $Y \bullet G_{i,j}(x)$.

3.1 Cone of critical directions

Let $\bar{x}$ be a local minimizer of the NLSDP (1) and let the generalized MFCQ condition (15) be satisfied at $\bar{x}$.

Then, the KKT conditions state that $Df(\bar{x})\Delta x \geq 0$ for all $\Delta x \in \mathcal{T}_1$. Clearly, if $Df(\bar{x})\Delta x > 0$, then Δx is an ascent direction for f. The cone of critical directions is given by

$$\mathcal{C}_1 := \{\Delta x \in \mathcal{T}_1 \mid Df(\bar{x})\Delta x = 0\}.$$

Let $\Delta x \in \mathcal{C}_1$ ($\Delta x \neq 0$) be given. By the generalized MFCQ (15) there exist differentiable curves $x(t) \in \mathcal{F}_1$ with $x(0) = \bar{x}$ and $\dot{x}(0) = \Delta x$. Since $Df(\bar{x})\dot{x}(0) = 0$ the question whether f is locally increasing or decreasing along such a curve depends on the second order terms in the constraints and in the objective function.

For the case of an NLP of the form (2) the cone of critical directions (short: critical cone) depends on three conditions:

$$DF(\bar{x})\Delta x = 0, \tag{18}$$

$$(DG(\bar{x}))_{k,l}\Delta x = 0 \quad \text{for all} \quad (k,l) \text{ with } \bar{Y}_{k,l} > 0 \tag{19}$$

$$(DG(\bar{x}))_{k,l}\Delta x \leq 0 \quad \text{for all} \quad (k,l) \text{ with } (G(\bar{x}))_{k,l} = 0,\ \bar{Y}_{k,l} = 0. \tag{20}$$

Note that the condition $Df(\bar{x})\Delta x = 0$ does not need to be included above since, by the KKT conditions, $Df(\bar{x})$ is spanned by the vectors in (18) and (19). The set of vectors Δx with (18) – (20) always contains the critical cone; and if LICQ (10) is satisfied, it coincides with the critical cone. If we assume LICQ and strict complementarity, then condition (20) will not apply, and the critical cone is in fact a linear subspace,

$$\mathcal{C}_2 = \{\Delta x \mid (18) \text{ and } (19) \text{ hold}\}.$$

Returning to the NLSDP (1) we also assume uniqueness of the multipliers, strict complementarity, $G(\bar{x}) - \bar{Y} \prec 0$, and a constraint qualification that is detailed in the next section. Proceeding in an analogous fashion, namely by replacing the inequalities $D\tilde{G}^{(2)}(\bar{x})[\Delta x] \preceq 0$ associated with the positive part of $\bar{Y}$ with equalities, the critical cone is given by

$$\mathcal{C}_1 = \{\Delta x \mid DF(\bar{x})\Delta x = 0, \quad (U^{(2)})^T DG(\bar{x})[\Delta x]U^{(2)} = 0\}. \tag{21}$$

Let $U^{(2)}$ have q columns, i.e. $U^{(2)}$ is a $d \times q$ matrix. It turns out that $\mathcal{C}_1$ is the linearized cone of the following boundary manifold of the feasible set

$$\mathcal{F}_1^{bd} := \{x \mid F(x) = 0, \quad \text{rank}(G(x)) = d - q\} \tag{22}$$

at $\bar{x}$. Locally, near $\bar{x}$, this is a subset of $\mathcal{F}_1$ (since the negative eigenvalues of $G(x)$ remain negative for small $\|x - \bar{x}\|$). In order to guarantee coincidence of the tangential cone of $\mathcal{F}_1^{bd}$ with $\mathcal{C}_1$ we require that $\mathcal{F}_1^{bd}$ satisfies some constraint qualification at $\bar{x}$. Yet, on first sight, it might not be quite obvious how to define a constraint qualification for the rank condition.

3.2 A second order constraint qualification

Recall the definition $\tilde{G}(x) := U^T G(x) U$ and its partition (12) and note that $\tilde{G}^{(1)}(x) \prec 0$ for small $\|x - \bar{x}\|$.

The condition that $G(x)$ (and thus also $\tilde{G}(x)$) has rank $d - q$ (see (22)) therefore translates to the condition that the Schur complement

$$\tilde{G}^{(2)}(x) - \tilde{G}^{(1,2)}(x)^T (\tilde{G}^{(1)}(x))^{-1} \tilde{G}^{(1,2)}(x) \equiv 0 \tag{23}$$

equals zero (for small $\|x - \bar{x}\|$).

Keeping U and the partition of $\tilde{G}$ fixed, the above representation of the constraint "rank$(G(x)) = d - q$" is simply a system of $q(q+1)/2$ nonlinear equations. Hence, MFCQ (8), (9) coincides with LICQ (10), and the tangential cone of $\mathcal{F}_1^{bd}$ at $\bar{x}$ always is a subset of $\mathcal{C}_1$. Since $\tilde{G}^{(1,2)}(\bar{x}) = 0$, the derivative of (23) at $x = \bar{x}$ is given by

$$D\tilde{G}^{(2)}(\bar{x}) = D_x \left((U^{(2)})^T G(x) U^{(2)} \right) \Big|_{x=\bar{x}}. \tag{24}$$

Regularity of $\mathcal{F}_1^{bd}$ follows, for example, if the $q(q+1)/2$ gradients of the constraints in (24) and the m gradients of $F_\nu(x)$ at $\bar{x}$, $(1 \le \nu \le m)$ are linearly independent.

This requirement, however, may be too strong; in particular, in the case that G is diagonal or block diagonal. In this case, also the matrix U can be chosen as a block diagonal matrix, and it suffices to check linear independence of the gradients associated with F_ν $(1 \le \nu \le m)$ and with the $U^{(2)}$-parts of the diagonal blocks of G.

Likewise, for example, when G is a congruence transformation of a block diagonal matrix, the requirement of linear independence can also be restricted to a smaller subset of active gradients, but if the transformation is not known explicitly, this subset may be difficult to identify. In order not to exclude such cases we generalize the requirement above and simply define that problem (1) satisfies the second order constraint qualification if the linearized cone of $\mathcal{F}_1^{bd}$ coincides with the right hand side of (21).

3.3 A second example

For problem (2), if $\bar{x}$ is a local minimizer satisfying LICQ (10), then the Hessian of the Lagrangian is positive semidefinite on $\mathcal{C}_2$, and conversely, if a KKT point $\bar{x}$ of (2) is given such that the Hessian of the Lagrangian is positive definite on $\mathcal{C}_2$, then $\bar{x}$ is a strict local minimizer of (2).

Again, this property does not translate to (1). As an example by Diehl et.al. [3] illustrates, the Hessian of the Lagrangian may be negative definite even if the cone of critical directions contains nonzero elements: Consider the problem of two variables

$$\text{minimize} \quad -x_1^2 - (1-x_2)^2 \quad \Big| \quad \begin{pmatrix} -1 & x_1 & x_2 \\ x_1 & -1 & 0 \\ x_2 & 0 & -1 \end{pmatrix} \preceq 0. \tag{25}$$

The semidefiniteness constraint is satisfied if, and only if, $x_1^2 + x_2^2 \le 1$. Thus, the global optimal solution is easily identified as $\bar{x} = (0, -1)^T$.

If we consider the equivalent NLP problem with the 1×1-matrix-constraint $G(x) := x_1^2 + x_2^2 - 1 \le 0$, this minimizer is "perfectly fine". (It satisfies strict complementarity, LICQ (10), and second order growth condition.)

Since the semidefiniteness constraint in (25) is linear, the Hessian of the Lagrangian of (25) is given by the Hessian of f and is thus negative definite everywhere, also the cone of critical directions which is given by all vectors of the form $(x_1, 0)^T$ with $x_1 \in \mathbb{R}$.

As pointed out in [3], this lack of semidefiniteness has implications: A sequential semidefinite programming algorithm that uses subproblems with a *convex* quadratic objective function and linearized semidefiniteness constraints generally will not converge superlinearly, no matter how the convex objective function is chosen. (And subproblems with a *nonconvex* quadratic objective function and linear semidefiniteness constraints are very hard to solve.)

This is in strong contrast to sequential quadratic programming algorithms for NLPs of the form (2) that do converge quadratically under standard assumptions – also when suitable convex quadratic objective functions are being used for the subproblems.

3.4 Second order necessary and sufficient conditions

The main result of this section is stated in the following theorems:

Theorem 2 *Let $\bar{x}$ be a local minimizer of (1) and let $\bar{x}, \bar{y}, \bar{Y}$ be a strictly complementary complementary KKT-point. Assume that problem (1) satisfies the second order constraint qualification of Section 3.2. Then*

$$h^T(D_x^2 L(\bar{x}, \bar{y}, \bar{Y}) + K(\bar{x}, \bar{Y}))h \ge 0 \qquad \forall\ h \in \mathcal{C}_1,$$

where $K(\bar{x}, \bar{Y}) \succeq 0$ is a matrix depending on the curvature of the semidefinite cone at $G(\bar{x})$ and the directional derivatives of G at $\bar{x}$, and is given by its matrix entries

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

The converse direction also holds without assuming regularity of $\mathcal{F}_1^{bd}$:

Theorem 3 *Let $\bar{x}$ be a strictly complementary KKT-point of (1) and assume that*

$$h^T(D_x^2 L(\bar{x}, \bar{y}, \bar{Y}) + K(\bar{x}, \bar{Y}))h > 0 \qquad \forall\ h \in \mathcal{C}_1.$$

Then, $\bar{x}$ is a strict local minimizer of (1) that satisfies the second order growth condition,

$$\exists \epsilon > 0, \quad \delta > 0: \qquad f(x+s) \ge f(x) + \epsilon\|s\|^2 \quad \forall\ \|s\| \le \delta \text{ with } x+s \in \mathcal{F}_1.$$

A stronger second order sufficient condition requiring that $D_x^2 L(\bar{x}, \bar{y}, \bar{Y})$ be positive definite on $\mathcal{C}_1$ was already given in [13, Definition 2.1], while the weaker form above is due to [14].

The weak form complementing the necessary condition includes the "extra term" $h^T K(\bar{x}, \bar{Y})h$ which explains the observation in (25) that the Hessian of the Lagrangian may be negative definite at the optimal solution, even if $\mathcal{C}_1$ contains nonzero elements. For completeness we present a self-contained proof of both results:

Proof of Theorem 2: Assume that $\bar{x}$ is a local minimizer of (1) and that (1) satisfies the second order constraint qualification of Section 3.2. Let $\Delta x \in \mathcal{C}_1$ be given. Then there exist $\alpha_k > 0$, $\alpha_k \to 0$, $s^k \to \Delta x$ with $\bar{x} + \alpha_k s^k \in \mathcal{F}_1^{bd}$. By assumption,

$$0 \le f(\bar{x} + \alpha_k s^k) - f(\bar{x})$$

for all sufficiently large k. Observe that the right hand side above equals

$$\alpha_k Df(\bar{x})s^k + \frac{\alpha_k^2}{2}(s^k)^T D^2 f(\bar{x})s^k + o(\alpha_k^2).$$

Inserting the KKT conditions, namely,

$$Df(\bar{x})s^k = -\bar{y}^T DF(\bar{x})s^k - \bar{Y} \bullet DG(\bar{x})[s^k]$$

into this expression, and canceling a factor α_k, we obtain

$$0 \le -\bar{y}^T DF(\bar{x})s^k - \bar{Y} \bullet DG(\bar{x})[s^k] + \frac{\alpha_k}{2}(s^k)^T D^2 f(\bar{x})s^k + o(\alpha_k). \tag{26}$$

From

$$0 = F_\nu(\bar{x} + \alpha_k s^k) = \underbrace{F_\nu(\bar{x})}_{=0} + \alpha_k DF_\nu(\bar{x})s^k + \frac{\alpha_k^2}{2}(s^k)^T D^2 F_\nu(\bar{x})s^k + o(\alpha_k^2)$$

it follows

$$-DF_\nu(\bar{x})s^k = \frac{\alpha_k}{2}(s^k)^T D^2 F_\nu(\bar{x})s^k + o(\alpha_k).$$

Inserting this in (26) yields

$$0 \le -\bar{Y} \bullet DG(\bar{x})[s^k] + \frac{\alpha_k}{2}(s^k)^T (D^2 f(\bar{x}) + \sum_{\nu=1}^m \bar{y}_\nu D^2 F_\nu(\bar{x}))s^k + o(\alpha_k). \tag{27}$$

In analogy to the transformation $\tilde{G}(x) = U^T G(x) U$, we define $\tilde{Y} = U^T \bar{Y} U$. Note that by complementarity, only $\tilde{Y}^{(2)}$ is nonzero. Using (23) (in line 5, below) it follows that

$$
\begin{aligned}
& \underbrace{\bar{Y} \bullet G(\bar{x})}_{=0} + \alpha_k \bar{Y} \bullet DG(\bar{x})[s^k] + \frac{\alpha_k^2}{2} \bar{Y} \bullet D^2 G(\bar{x})[s^k, s^k] + o(\alpha_k^2) \qquad (28) \\
= \;& \bar{Y} \bullet G(\bar{x} + \alpha_k s^k) \\
= \;& \tilde{Y} \bullet \tilde{G}(\bar{x} + \alpha_k s^k) \\
= \;& \tilde{Y}^{(2)} \bullet \tilde{G}^{(2)}(\bar{x} + \alpha_k s^k) \\
= \;& \tilde{Y}^{(2)} \bullet \Big(\tilde{G}^{(1,2)}(\bar{x} + \alpha_k s^k)^T (\tilde{G}^{(1)}(\bar{x} + \alpha_k s^k))^{-1} \tilde{G}^{(1,2)}(\bar{x} + \alpha_k s^k) \Big) \\
= \;& \alpha_k^2 \tilde{Y}^{(2)} \bullet \Big((D\tilde{G}^{(1,2)}(\bar{x})[s^k])^T (\tilde{G}^{(1)}(\bar{x}))^{-1} D\tilde{G}^{(1,2)}(\bar{x})[s^k] \Big) + o(\alpha_k^2).
\end{aligned}
$$

In line 6 above (the last line) we used that fact that $\tilde{G}^{(1,2)}(\bar{x}) = 0$, so that

$$
\tilde{G}^{(1,2)}(\bar{x} + \alpha_k s^k) = \alpha_k D\tilde{G}^{(1,2)}(\bar{x})[s^k] + o(\alpha^k).
$$

Dividing the first and the last term in the above chain of equalities by α_k we obtain.

$$
\begin{aligned}
& \bar{Y} \bullet DG(\bar{x})[s^k] + \frac{\alpha_k}{2} \bar{Y} \bullet D^2 G(\bar{x})[s^k, s^k] + o(\alpha_k) \\
= \;& \alpha_k \tilde{Y}^{(2)} \bullet \Big((D\tilde{G}^{(1,2)}(\bar{x})[s^k])^T (\tilde{G}^{(1)}(\bar{x}))^{-1} D\tilde{G}^{(1,2)}(\bar{x})[s^k] \Big) + o(\alpha_k) \\
= \;& \alpha_k \tilde{Y} \bullet \Big((D\tilde{G}(\bar{x})[s^k])^T (\tilde{G}(\bar{x}))^{\dagger} D\tilde{G}(\bar{x})[s^k] \Big) + o(\alpha_k),
\end{aligned}
$$

where $G^{\dagger}$ denotes the pseudo inverse of a matrix G. In the last line above, we use the fact that $\tilde{G}(\bar{x})$ is zero outside the $\tilde{G}^{(1)}(\bar{x})$-block and $\bar{Y}$ is zero outside the $\tilde{Y}^{(2)}$-block. Undoing the change of basis (i.e. multiplying all matrices $\tilde{G}$ and $\tilde{Y}$ by U and by U^T from left and right) yields

$$
\begin{aligned}
& -\bar{Y} \bullet DG(\bar{x})[s^k] \\
= \;& \frac{\alpha_k}{2} \bar{Y} \bullet \big(D^2 G(\bar{x})[s^k, s^k] - 2DG(\bar{x})[s^k]^T (G(\bar{x}))^{\dagger} DG(\bar{x})[s^k] \big) + o(\alpha_k).
\end{aligned}
$$

Inserting this in (27) finally yields

$$
\begin{aligned}
0 \;\leq\; & \frac{\alpha_k}{2} (s^k)^T (D^2 f(\bar{x}) + \sum_{\nu=1}^{m} \bar{y}_\nu D^2 F_\nu(\bar{x})) s^k \\
& + \frac{\alpha_k}{2} \bar{Y} \bullet \big(D^2 G(\bar{x})[s^k, s^k] - 2DG(\bar{x})[s^k]^T (G(\bar{x}))^{\dagger} DG(\bar{x})[s^k] \big) + o(\alpha_k).
\end{aligned}
$$

Note that the right hand side can be written in terms of the Hessian of the Lagrangian (17),

$$
0 \leq \frac{\alpha_k}{2} \big((s^k)^T D^2 L(\bar{x}, \bar{y}, \bar{Y}) s^k - 2\bar{Y} \bullet DG(\bar{x})[s^k]^T (G(\bar{x}))^{\dagger} DG(\bar{x})[s^k] \big) + o(\alpha_k).
$$

Dividing by α_k and taking the limit as $k \to \infty$, we arrive at:

$$
0 \leq \frac{1}{2} \big((\Delta x)^T D^2 L(\bar{x}, \bar{y}, \bar{Y}) \Delta x - 2\bar{Y} \bullet DG(\bar{x})[\Delta x]^T (G(\bar{x}))^{\dagger} DG(\bar{x})[\Delta x] \big) .
$$

Summarizing, $\bar{x}, \bar{y}, \bar{Y}$ satisfy the second order necessary condition of Theorem 2. #

Proof of Theorem 3: Let the assumptions of Theorem 3 be satisfied and assume that $\bar{x}$ is not a strict local minimizer as stated in the theorem, i.e. there exists a sequence of feasible points $x^{(k)}$ with

$$\|x^{(k)} - \bar{x}\| \leq \frac{1}{k} \quad \text{and} \quad f(x^{(k)}) < f(\bar{x}) + \frac{1}{k}\|x^{(k)} - \bar{x}\|^2.$$

In particular, $x^{(k)} \neq \bar{x}$ for all k. By considering a subsequence, one may assume without loss of generality that $s^{(k)} := \frac{x^{(k)} - \bar{x}}{\|x^{(k)} - \bar{x}\|} \to \bar{s}$. By construction, $\bar{s}$ lies in the tangential cone $\mathcal{T}_1$ of (1) at $\bar{x}$ and therefore satisfies (13).

We show that $\bar{s}$ also lies in the cone of critical directions (21) at $\bar{x}$. Let $g(x) := \bar{Y} \bullet G(x)$, so that for feasible points x we have $L(x, \bar{y}, \bar{Y}) = f(x) + g(x)$. Then,

$$\begin{aligned} \frac{1}{k}\|x^{(k)} - \bar{x}\|^2 &> f(x^{(k)}) - f(\bar{x}) \\ &= L(x^{(k)}, \bar{y}, \bar{Y}) - L(\bar{x}, \bar{y}, \bar{Y}) - (g(x^{(k)}) - g(\bar{x})) \end{aligned} \tag{29}$$

Using Taylors expansion and the fact that $D_x L(\bar{x}, \bar{y}, \bar{Y}) = 0$ we obtain

$$L(x^{(k)}, \bar{y}, \bar{Y}) = L(\bar{x}, \bar{y}, \bar{Y}) + \frac{1}{2}(x^{(k)} - \bar{x})^T D_{xx}^2 L(\bar{x}, \bar{y}, \bar{Y})(x^{(k)} - \bar{x}) + o(\|x^{(k)} - \bar{x}\|^2).$$

Dividing $L(x^{(k)}, \bar{y}, \bar{Y}) - L(\bar{x}, \bar{y}, \bar{Y})$ by $\|x^{(k)} - \bar{x}\|^2$ and using (29) we get

$$\frac{1}{k} > \frac{1}{2}(s^{(k)})^T D_{xx}^2 L(\bar{x}, \bar{y}, \bar{Y}) s^{(k)} - \frac{g(x^{(k)}) - g(\bar{x})}{\|x^{(k)} - \bar{x}\|^2} + o(1). \tag{30}$$

Since $s^{(k)}$ converges and $g(x^{(k)}) \leq 0 = g(\bar{x})$, there exists $M < \infty$ such that

$$0 \leq -\frac{g(x^{(k)}) - g(\bar{x})}{\|x^{(k)} - \bar{x}\|^2} \leq M.$$

Thus,

$$0 = \lim_{k \to \infty} \frac{g(x^{(k)}) - g(\bar{x})}{\|x^{(k)} - \bar{x}\|}. \tag{31}$$

Here,

$$\begin{aligned} g(x^{(k)}) - g(\bar{x}) &= Dg(\bar{x})(x^{(k)} - \bar{x}) + o(\|x^{(k)} - \bar{x}\|) \\ &= \bar{Y} \bullet DG(\bar{x})[x^{(k)} - \bar{x}] + o(\|x^{(k)} - \bar{x}\|). \end{aligned}$$

Inserting this in (31) and taking the limit yields

$$0 = Dg(\bar{x})\bar{s} = \bar{Y} \bullet DG(\bar{x})[\bar{s}].$$

Since $\bar{s} \in \mathcal{T}_1$ it follows $DF(\bar{x})\bar{s} = 0$ and from strict complementarity and the second relation in (13) it follows $(U^{(2)})^T DG(\bar{x})[\bar{s}]U^{(2)} = 0$ and thus, $\bar{s} \in \mathcal{C}_1$.

Starting from (29) we may now follow the chain of inequalities in the proof of Theorem (2) with reversed signs. With $\alpha_k := \|x^{(k)} - \bar{x}\|$ we obtain as in (27)

$$\frac{\alpha_k}{k} \geq -\bar{Y} \bullet DG(\bar{x})[s^k] + \frac{\alpha_k}{2}(s^k)^T(D^2 f(\bar{x}) + \sum_{\nu=1}^{m} \bar{y}_\nu D^2 F_\nu(\bar{x}))s^k + o(\alpha_k). \quad (32)$$

Since $G(x^{(k)}) \preceq 0$ we may continue as in (28),

$$\begin{aligned}
& \underbrace{\bar{Y} \bullet G(\bar{x})}_{=0} + \alpha_k \bar{Y} \bullet DG(\bar{x})[s^k] + \frac{\alpha_k^2}{2}\bar{Y} \bullet D^2 G(\bar{x})[s^k, s^k] + o(\alpha_k^2) \\
= \; & \tilde{Y}^{(2)} \bullet \tilde{G}^{(2)}(\bar{x} + \alpha_k s^k) \\
\leq \; & \tilde{Y}^{(2)} \bullet \Big(\tilde{G}^{(1,2)}(\bar{x} + \alpha_k s^k)^T (\tilde{G}^{(1)}(\bar{x} + \alpha_k s^k))^{-1} \tilde{G}^{(1,2)}(\bar{x} + \alpha_k s^k)\Big) \\
= \; & \alpha_k^2 \tilde{Y}^{(2)} \bullet \Big((D\tilde{G}^{(1,2)}(\bar{x})[s^k])^T (\tilde{G}^{(1)}(\bar{x}))^{-1} D\tilde{G}^{(1,2)}(\bar{x})[s^k]\Big) + o(\alpha_k^2).
\end{aligned}$$

Again, it follows

$$\begin{aligned}
& \bar{Y} \bullet DG(\bar{x})[s^k] + \frac{\alpha_k}{2}\bar{Y} \bullet D^2 G(\bar{x})[s^k, s^k] + o(\alpha_k) \\
\leq \; & \alpha_k \tilde{Y} \bullet \Big((D\tilde{G}(\bar{x})[s^k])^T (\tilde{G}(\bar{x}))^\dagger D\tilde{G}(\bar{x})[s^k]\Big) + o(\alpha_k) \\
= \; & \alpha_k \bar{Y} \bullet \big((DG(\bar{x})[s^k])^T (G(\bar{x}))^\dagger DG(\bar{x})[s^k]\big) + o(\alpha_k),
\end{aligned}$$

leading to

$$\begin{aligned}
\frac{\alpha_k}{k} \geq \; & \frac{\alpha_k}{2}(s^k)^T(D^2 f(\bar{x}) + \sum_{\nu=1}^{m} \bar{y}_\nu D^2 F_\nu(\bar{x}))s^k \\
& + \frac{\alpha_k}{2}\bar{Y} \bullet \big(D^2 G(\bar{x})[s^k, s^k] - 2DG(\bar{x})[s^k]^T (G(\bar{x}))^\dagger DG(\bar{x})[s^k]\big) + o(\alpha_k).
\end{aligned}$$

where the right hand side can be written in terms of the Hessian of the Lagrangian,

$$\frac{\alpha_k}{k} \geq \frac{\alpha_k}{2}\big((s^k)^T D^2 L(\bar{x}, \bar{y}, \bar{Y})s^k - 2\bar{Y} \bullet DG(\bar{x})[s^k]^T (G(\bar{x}))^\dagger DG(\bar{x})[s^k]\big) + o(\alpha_k).$$

Dividing by α_k and taking the limit as $k \to \infty$, we arrive at:

$$0 \geq \frac{1}{2}\left(\bar{s}^T D^2 L(\bar{x}, \bar{y}, \bar{Y})\bar{s} - 2\bar{Y} \bullet DG(\bar{x})[\bar{s}]^T (G(\bar{x}))^\dagger DG(\bar{x})[\bar{s}]\right).$$

in contradiction to the conditions in the theorem. #

Conclusion

The goal of this paper is an easy and self-contained presentation of optimality conditions for nonlinear semidefinite programs – but the nature of the second order conditions makes it difficult to live up to this goal. We have outlined regularity conditions and contrasted optimality conditions for NLP and NLSDP in nondegenerate cases. Under standard nondegeneracy assumptions, the tangential cone of an NLP is polyhedral, but the same is generally not true for the tangential cone of an NLSDP. At an optimal solution the Hessian of the Lagrangian of an NLP is positive semidefinite along directions in the critical cone, but this is generally not the case for NLSDPs. In particular, the latter observation has impact on the design of Sequential Semidefinite Programming approaches. A very detailed discussion (also in case that strict complementarity or some regularity conditions are violated) can be found in the monograph by Bonnans and Shapiro [2].

Acknowledgment

The author is thankful to two anonymous referees whose constructive criticism helped to improve the presentation of the paper.

References

[1] F. Alizadeh, "Optimization over the positive semi-definite cone: interior-point methods and combinatorial applications," in: P. Pardalos, ed., *Advances in Optimization and Parallel Computing* (North-Holland, Amsterdam, 1992) pp. 1–25.

[2] F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer-Verlag, New York, 2000.

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

[4] B. Fares, D. Noll, P. Apkarian, (2002): Robust control via sequential semidefinite programming, SIAM J. Contr. Opt. (40) 1791–1820.

[5] Freund, R. W, Jarre, F; Vogelbusch, C. H. Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling Math. Prog. **xx**, –

[6] S. Gaile, G.Leugering, M. Stingl. Free Material Optimization for Plates and Shells. To appear in Proceedings of 23rd IFIP TC 7 Conference on System Modelling and Optimization. Cracow, Poland, July 23-27, 2006

[7] M. Kocvara, M. Stingl, Pennon-AMPL User's Guide (Version 1.3) http://www2.am.uni-erlangen.de/˜kocvara/pennon/pennon1_3_ampl.pdf

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

[9] J.E. Nesterov, (1998): Semidefinite relaxation and nonconvex quadratic optimization. Optim. Meth. Software **9** 141–160.

[10] J.E. Nesterov and A.S. Nemirovsky, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications* (SIAM, Philadelphia, 1994).

[11] J. Nocedal and S.J. Wright: Numerical Optimization. Springer (Second edition, 2006).

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

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

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

[15] G. Sonnevend, “Applications of analytic centers to feedback control systems", Control of uncertain systems, Bremen, June 1989, D. Hinrichsen and B. Martensson (eds.), Birkhäuser (1990).