Mathematical Programming 77 (1997) 301–320

First and second order analysis of nonlinear semidefinite programs

Alexander Shapiro [1]

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA

Received 5 December 1994

Abstract

In this paper we study nonlinear semidefinite programming problems. Convexity, duality and first-order optimality conditions for such problems are presented. A second-order analysis is also given. Second-order necessary and sufficient optimality conditions are derived. Finally, sensitivity analysis of such programs is discussed. © 1997 The Mathematical Programming Society, Inc. Published by Elsevier Science B.V.

Keywords: Semidefinite programming; Cone constraints; Convex programming; Duality; Second-order optimality conditions; Tangent cones; Optimal value function; Sensitivity analysis

1. Introduction

In this paper we consider the following optimization problem

$$\text{(P)} \qquad \min_{x\in\mathbb{R}^m} f(x) \quad \text{subject to } G(x) \preceq 0.$$

Here $G : \mathbb{R}^m \to \mathcal{S}_n$ is a mapping from $\mathbb{R}^m$ into the space $\mathcal{S}_n$ of $n \times n$ symmetric matrices and, for $A, B \in \mathcal{S}_n$, the notation $A \succeq B$ (the notation $A \preceq B$) means that the matrix $A - B$ is positive semidefinite (negative semidefinite). Consider the cone $\mathcal{K} \subset \mathcal{S}_n$ of positive semidefinite matrices. Then the constraint $G(x) \preceq 0$ can be written in the form of the cone constraint $G(x) \in -\mathcal{K}$.

We study the above optimization problem by employing general techniques of nonlinear programming under cone constraints. The organization of this paper is as follows. In Section 2 we discuss convexity, duality and first-order optimality conditions of the

[1] E-mail: ashapiro@isye.gatech.edu.

Published by Elsevier Science B.V.
PII S0025-5610(96)00081-0

problem (P). The material of that section is mainly of a survey nature although it seems that general nonlinear semidefinite programs were not systematically studied before (cf. [18,30]). In Section 3 we study the geometry of the cone $\mathcal{K}$. The material of that section is a summary and a simplification of results which appeared in various publications. In that section we make use of the transversality concept borrowed from differential geometry (cf. [30]). Section 4 is devoted to a derivation of second-order optimality conditions for the program (P). The second-order analysis is employed in Section 5 to an investigation of parameterized semidefinite programs. In that section differentiability properties of the optimal value and optimal solutions of parameterized semidefinite programs are discussed.

We assume that $f(x)$ and $G(x)$ are sufficiently smooth on $\mathbb{R}^m$ and use the following notation and terminology throughout the paper. For matrices $A, B \in \mathcal{S}_n$ we use the scalar product $A \bullet B = \operatorname{tr} AB$. The symbol "$\otimes$" denotes the Kronecker product of matrices (see, e.g., [11] for a discussion of basic properties of the Kronecker product). By $A^\dagger$ we denote the Moore–Penrose inverse of a matrix A. For a matrix $A \in \mathcal{S}_n$ we denote by $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$ its eigenvalues arranged in the decreasing order. We also use notation $\lambda_{\min}(C)$ for the smallest eigenvalue of a symmetric matrix C. For an $n \times m$ matrix B we use $\operatorname{vec}(B)$ to denote the $nm \times 1$ vector obtained by stacking columns of B. The notation $\mathrm{d}G(x)$ is used for the differential of the mapping $G(\cdot)$ at x, i.e. $\mathrm{d}G(x)$ is a linear mapping from $\mathbb{R}^m$ into $\mathcal{S}_n$ defined by

$$[\mathrm{d}G(x)]y = \sum_{i=1}^{m} y_i G_i(x), \tag{1}$$

where $G_i(x) = \partial G(x)/\partial x_i$ are $n \times n$ partial derivative matrices. By $\partial G(x)/\partial x$ we denote the $n^2 \times m$ Jacobian matrix of the mapping $\operatorname{vec} G(\cdot)$, i.e. $\partial G(x)/\partial x = [\operatorname{vec} G_1(x), \dots, \operatorname{vec} G_m(x)]$. For a set $\mathcal{M} \subset \mathcal{S}_n$ we denote by $\sigma(\cdot, \mathcal{M})$ its support function

$$\sigma(\Omega, \mathcal{M}) = \sup_{M \in \mathcal{M}} \Omega \bullet M. \tag{2}$$

Note that here $\mathcal{K}^+ = \mathcal{K}$, where $\mathcal{K}^+$ denotes the positive dual of the cone $\mathcal{K}$,

$$\mathcal{K}^+ = \{A \in \mathcal{S}_n \mid A \bullet Y \geqslant 0, \text{ for all } Y \in \mathcal{K}\}. \tag{3}$$

For a real valued function $h(x)$, defined on a vector space X, we denote by $h'(x,d)$ its directional derivative

$$h'(x,d) = \lim_{t \to 0^+} \frac{h(x+td) - h(x)}{t}, \tag{4}$$

and by $h''(x,d,v)$ its second-order directional derivative

$$h''(x,d,v) = \lim_{t \to 0^+} \frac{h(x + td + \frac{1}{2}t^2 v) - h(x) - th'(x,d)}{\frac{1}{2}t^2}. \tag{5}$$

2. Convexity, duality and first-order optimality conditions

We say that the mapping $G(x)$ is *positive semidefinite convex* (psd-convex) if it is convex with respect to the order relation imposed by the cone $\mathcal{K}$. That is, the inequality

$$tG(x) + (1-t)G(y) \succeq G(tx + (1-t)y) \tag{6}$$

holds for any $x, y \in \mathbb{R}^m$ and all $t \in [0,1]$. Condition (6) can be written in the following equivalent form

$$tG(x) + (1-t)G(y) - G(tx + (1-t)y) \in \mathcal{K}. \tag{7}$$

In case the mapping G is defined on the set of symmetric matrices by a scalar function $\phi : \mathbb{R} \to \mathbb{R}$, i.e. $G : \mathcal{S}_n \to \mathcal{S}_n$ and $G(x) = F^{\mathrm{T}}\phi(D)F$ where $x = F^{\mathrm{T}}DF$ is the spectral decomposition of the matrix x and $\phi(D)$ applies componentwise to the diagonal entries of D, there exists a well developed theory relating convexity properties of ϕ to (6) (see [11, 15]). We deal here with a somewhat different situation when x can be an arbitrary vector.

Proposition 1. *The mapping $G(x)$ is psd-convex if and only if for any $v \in \mathbb{R}^n$ the function $\varphi(x) = v^{\mathrm{T}}G(x)v$ is convex.*

Proof. We have that (6) holds iff the inequality

$$tv^{\mathrm{T}}G(x)v + (1-t)v^{\mathrm{T}}G(y)v \geqslant v^{\mathrm{T}}G(tx + (1-t)y)v, \tag{8}$$

is satisfied for any $v \in \mathbb{R}^n$. This proves the required assertion. $\square$

It can be remarked that Proposition 1 is a particular case of a following result which holds for general convex mappings with respect to convex cones. If $\mathcal{K}$ is a closed convex cone in a Banach space X, then it follows from the dual relation $(\mathcal{K}^+)^+ = \mathcal{K}$, that (7) holds iff

$$\langle \alpha, tG(x) + (1-t)G(y) - G(tx + (1-t)y) \rangle \geqslant 0$$

for any $\alpha \in \mathcal{K}^+ \subset X^*$. Here X^* denotes the dual space of X and for $\alpha \in X^*$, $x \in X$, $\langle \alpha, x \rangle$ stands for the value $\alpha(x)$. This implies that $G(x)$ is convex with respect to $\mathcal{K}$ iff the real valued function $\langle \alpha, G(\cdot) \rangle$ is convex for any $\alpha \in \mathcal{K}^+$. What is specific here about the cone of positive semidefinite matrices is that $\mathcal{K}^+ = \mathcal{K}$ and that any matrix $A \in \mathcal{K}$ can be represented as a sum of rank-one symmetric matrices (of the form vv^{T}).

Let us consider the following example of a *matrix valued* quadratic form

$$G(x) = A_0 + \sum_{i=1}^{m} x_i A_i + \frac{1}{2} \sum_{i,j=1}^{m} x_i x_j A_{ij}, \tag{9}$$

where A_0, A_i and A_{ij} are given $n \times n$ symmetric matrices. It follows then from Proposition 1 that this mapping is psd-convex iff for any $v \in \mathbb{R}^n$ the $m \times m$ matrix Q with elements $[Q]_{ij} = v^{\mathrm{T}}A_{ij}v$, is positive semidefinite. Note that $Q = (I_m \otimes v)^{\mathrm{T}}[A_{ij}](I_m \otimes v)$, where $[A_{ij}]$ is the $mn \times mn$ block matrix and I_m is the $m \times m$ identity matrix. Consequently positive semidefiniteness of Q, for any $v \in \mathbb{R}^n$, is implied by positive semidefiniteness of the block matrix $[A_{ij}]$. It follows that positive semidefiniteness of the block matrix $[A_{ij}]$ is a sufficient condition for psd-convexity of the quadratic form $G(x)$ defined in (9).

In case the components $g_{ij}(x)$, $i, j = 1, \ldots, n$, of the mapping $G(x)$ are twice continuously differentiable functions, it is possible to give the following characterization of the positive semidefinite convexity in terms of the Hessian matrices $\nabla^2 g_{ij}(x)$.

Proposition 2. *Suppose that the mapping $G(x)$ is twice continuously differentiable. Then $G(x)$ is psd-convex if and only if the $m \times m$ matrix $\sum_{i,j=1}^{n} v_i v_j \nabla^2 g_{ij}(x)$ is positive semidefinite for any $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ and any $x \in \mathbb{R}^m$. A sufficient condition for psd-convexity of $G(x)$ is positive semidefiniteness of the $mn \times mn$ block matrix $H(x) = [\nabla^2 g_{ij}(x)]$ for any $x \in \mathbb{R}^m$.*

Proof. By Proposition 1 we have that $G(x)$ is psd-convex iff the function $\varphi(x) = \sum_{i,j=1}^{n} v_i v_j g_{ij}(x)$ is convex for any $v \in \mathbb{R}^n$. Since $\varphi(x)$ is twice continuously differentiable, it is convex iff its Hessian matrix $\nabla^2\varphi(x) = \sum_{i,j=1}^{n} v_i v_j \nabla^2 g_{ij}(x)$ is positive semidefinite for any $x \in \mathbb{R}^m$. We can write $\nabla^2\varphi(x)$ in the form $\nabla^2\varphi(x) = (v \otimes I_m)^{\mathrm{T}} H(x)(v \otimes I_m)$. Then positive semidefiniteness of $\nabla^2\varphi(x)$ follows from positive semidefiniteness of $H(x)$. □

Remarks. If the mapping $G(x)$ is given in the form (9), then $[A_{ij}] = P^{\mathrm{T}}H(x)P$, where P is a certain permutation matrix (see [11, Section 4.3.8] for a discussion of such permutation matrices). It follows from the Proposition 1 that if $G(x)$ is psd-convex and B is an $n \times k$ matrix, then the mapping $x \to B^{\mathrm{T}}G(x)B$ is also psd-convex. Also if $G(x)$ is diagonal valued, $G(x) = \mathrm{diag}(g_{11}(x), \ldots, g_{nn}(x))$, then $G(x)$ is psd-convex iff the functions $g_{ii}(x)$, $i = 1, \ldots, n$, are convex.

It is interesting to relate psd-convexity of $G(x)$ to convexity of its eigenvalues $\lambda_i(x) = \lambda_i(G(x))$, $i = 1, \ldots, n$. The largest eigenvalue $\lambda_1(x)$ of $G(x)$ can be represented in the form

$$\lambda_1(x) = \max\{v^{\mathrm{T}}G(x)v \mid v \in \mathbb{R}^n,\ v^{\mathrm{T}}v = 1\}.$$

Consequently, if $G(x)$ is psd-convex, then $\lambda_1(x)$ is representable as a maximum of convex functions and hence is convex, and thus the set $G^{-1}(-\mathcal{K}) = \{x \in \mathbb{R}^m \mid \lambda_1(x) \leqslant 0\}$ of feasible solutions of the program (P) is convex. (Convexity of $G^{-1}(-\mathcal{K})$ can be also derived directly from condition (7).) The converse is not true, i.e. convexity of $\lambda_1(x)$ does not imply psd-convexity of $G(x)$. Indeed, suppose for example that $G(x)$ is a two by two diagonal matrix, $G(x) = \mathrm{diag}(g_{11}(x), g_{22}(x))$, such that $g_{11}(x) > g_{22}(x)$ for all x and $g_{11}(x)$ is convex while $g_{22}(x)$ is not. Then $\lambda_1(x) = g_{11}(x)$ is convex and $G(x)$ is not psd-convex.

As an another example consider the mapping $G(x) = a(x)a(x)^{\mathrm{T}}$, where $a(x)$ is a 2×1 vector valued function of $x \in \mathbb{R}$ given by $a_1(x) = 1 - 2x^2 + 2x$ and $a_2(x) = 1 - 2x^2 - 2x$. Then $\lambda_1(x) = (1 - 2x^2 + 2x)^2 + (1 - 2x^2 - 2x)^2$ and $\lambda_2(x) = 0$ are both convex. On the other hand $u^{\mathrm{T}}G(x)u = 4(1 - 2x^2)^2$, where $u = (1,1)^{\mathrm{T}}$, is not convex and hence $G(x)$ is not psd-convex. In this example both eigenvalues of $G(x)$ are convex while $G(x)$ is not psd-convex.

Also psd-convexity of $G(x)$ does not imply convexity of *all* its eigenvalues. Consider the following example, $G(x)$ is 2×2, $x \in \mathbb{R}$, with $g_{11}(x) = x$, $g_{22}(x) = -x$ and $g_{12}(x) = g_{21}(x) = 1$. Here the mapping $G(x)$ is linear and hence is psd-convex. On the other hand we have that $\lambda_1(x) + \lambda_2(x) = \operatorname{tr} G(x) = 0$ and hence $\lambda_2(x) = -\lambda_1(x)$. Since $\lambda_1(x)$ is a nonlinear convex function, it follows that in this example $\lambda_2(x)$ is not convex.

Consider now the Lagrangian function,

$$L(x, \Omega) = f(x) + \Omega \bullet G(x), \tag{10}$$

$\Omega \in \mathcal{S}_n$. Note that if the program (P) is convex, i.e. $f(x)$ is convex and $G(x)$ is psd-convex, then $L(\cdot, \Omega)$ is convex on $\mathbb{R}^m$ for any $\Omega \succeq 0$. We can associate with the program (P) the following dual program

$$\text{(D)} \qquad \max_{\Omega \in \mathcal{S}_n} \left\{ \phi(\Omega) := \min_{x \in \mathbb{R}^m} L(x, \Omega) \right\} \quad \text{subject to } \Omega \succeq 0.$$

We say that the *Slater condition*, for the program (P), holds if there exists a point $\bar{x} \in \mathbb{R}^m$ such that $G(\bar{x}) \in \operatorname{int}(-\mathcal{K})$, i.e. $G(\bar{x})$ is a negative definite matrix. By standard arguments of convex analysis we have that if (P) is convex and the Slater condition holds, then there is no duality gap between the programs (P) and (D) and the set of optimal solutions of the dual program (D) is nonempty and compact (see [23, Theorem 17, p. 41, and p. 47]). In the considered finite-dimensional case the converse of this statement is also true.

Consider the linear case, with $f(x) = c^{\mathrm{T}}x$ and $G(x) = A_0 + \sum_{i=1}^m x_i A_i$. Then $\phi(\Omega) = \Omega \bullet A_0$ if $c_i + \Omega \bullet A_i = 0$, $i = 1, \dots, m$, and $\phi(\Omega) = -\infty$ otherwise. Therefore in that case the dual problem takes the form (cf. [1,27])

$$\begin{aligned} \max_{\Omega \in \mathcal{S}_n} \quad & \Omega \bullet A_0 \\ \text{subject to} \quad & c_i + \Omega \bullet A_i = 0, \quad i = 1, \dots, m, \\ & \Omega \succeq 0. \end{aligned} \tag{11}$$

Consider now the following quadratic case, $f(x) = c^{\mathrm{T}}x$ and the mapping $G(x)$ is given in the form (9). Then

$$\phi(\Omega) = \min_{x \in \mathbb{R}^m} \{ \Omega \bullet A_0 + b^{\mathrm{T}}x + \tfrac{1}{2}x^{\mathrm{T}}Qx \},$$

where the components of vector b are $b_i = c_i + \Omega \bullet A_i$, $i = 1, \dots, m$, and the elements of the matrix Q are given by $[Q]_{ij} = \Omega \bullet A_{ij}$, $i, j = 1, \dots, m$. Note that by Proposition 1 the quadratic mapping $G(x)$ is psd-convex iff the matrix Q is positive semidefinite for any $\Omega \succeq 0$. Suppose that the $mn \times mn$ block matrix $[A_{ij}]$ is positive definite. We have then that for $\Omega = vv^{\mathrm{T}}$ the matrix Q can be written in the form $Q = (I_m \otimes v)^{\mathrm{T}}[A_{ij}](I_m \otimes v)$. Therefore in that case the matrix Q is positive definite for any non zero $\Omega \succeq 0$ and moreover

$$\phi(\Omega) = \Omega \bullet A_0 - \tfrac{1}{2} b^{\mathrm{T}} Q^{-1} b.$$

Let us discuss now first-order optimality conditions for the program (P). We assume that $f(x)$ and $G(x)$ are continuously differentiable and use the following condition which is an extension of the Mangasarian–Fromovitz constraint qualification used in nonlinear programming.

MF-condition. We say that the MF-condition holds at a feasible point x_0 if there exists a vector $h \in \mathbb{R}^m$ such that

$$G(x_0) + [\mathrm{d}G(x_0)]h \in \mathrm{int}(-\mathcal{K}), \tag{12}$$

i.e. $G(x_0) + [\mathrm{d}G(x_0)]h$ is a negative definite matrix. Recall that $\mathrm{d}G(x_0) : \mathbb{R}^m \to \mathcal{S}_n$ is a linear mapping defined in (1).

Note that in the convex case the MF-condition is equivalent to the Slater condition. This equivalence is well known in nonlinear programming. For cone type constraints it is discussed in [31].

Under the MF-condition the following first-order necessary conditions apply to the program (P). Let x_0 be a locally optimal solution of the program (P). Then there exists a positive semidefinite matrix Ω of Lagrange multipliers such that

$$\nabla_x L(x_0, \Omega) = 0, \tag{13}$$

$$\Omega G(x_0) = 0. \tag{14}$$

The MF-condition implies that the set of positive semidefinite matrices $\Omega \in \mathcal{S}_n$, satisfying conditions (13) and (14), is *nonempty* and *bounded* (cf. [13,16,21]). Note that since $G(x_0) \preceq 0$ and $\Omega \succeq 0$, condition (14) is equivalent to the condition $\Omega \bullet G(x_0) = 0$, which is the standard complementarity condition under cone constraints. Note also that (14) implies that the matrices Ω and $G(x_0)$ commute and have the same system of eigenvectors and that $\sum_{i=1}^{n} \alpha_i \gamma_i = 0$, where α_i and γ_i are eigenvalues (not necessarily arranged in the decreasing order) of the matrices Ω and $G(x_0)$, respectively. Since $\alpha_i \geqslant 0$ and $\gamma_i \leqslant 0$, the last equation is equivalent to the condition: $\alpha_i \gamma_i = 0$, $i = 1, \dots, n$.

In the linear case, optimality conditions (13) and (14) take the form

$$c_i + \Omega \bullet A_i = 0, \quad i = 1, \dots, m, \tag{15}$$

and

$$\Omega A_0 + \sum_{i=1}^{m} x_i \Omega A_i = 0, \tag{16}$$

respectively. It follows that if for a matrix $\Omega \succeq 0$ satisfying optimality conditions (15) and (16), matrices ΩA_i, $i = 1, \dots, m$, are linearly independent, then the program (P) has a unique optimal solution. Such uniqueness of the optimal solution was established for minimum trace factor analysis and some of its extensions (see [27]).

Because of the complementarity condition (14) we have that

$$\operatorname{rank} G(x_0) + \operatorname{rank} \Omega \leqslant n. \tag{17}$$

We say that the *strict complementarity condition* holds if

$$\operatorname{rank} G(x_0) + \operatorname{rank} \Omega = n. \tag{18}$$

Consider now the barrier function

$$\eta(x) = \begin{cases} -\log\det(-G(x)) & \text{if } G(x) \prec 0, \\ +\infty & \text{otherwise,} \end{cases} \tag{19}$$

used in semidefinite programming [17]. It is well known that $\eta(x)$ is convex if the mapping $G(x)$ is affine. Convexity of $\eta(x)$ also holds for psd-convex mappings.

Proposition 3. *If the mapping $G(x)$ is psd-convex, then the barrier function $\eta(x)$ is convex.*

Proof. In order to show that $\eta(x)$ is convex it will be sufficient to verify convexity of the function $\alpha(t) = \eta(x + ty)$, for any $x, y \in \mathbb{R}^m$, and such $t \in \mathbb{R}$ that $F(t) = -G(x + ty)$ is positive definite. Also it will be sufficient to consider the case where the mapping $F(t) = -G(x + ty)$ is twice continuously differentiable. We have then that $F(t) = A + tB + \frac{1}{2}t^2C + o(t^2)$, where $A = F(0)$, $B = \mathrm{d}F(0)/\mathrm{d}t$ and $C = \mathrm{d}^2F(0)/\mathrm{d}t^2$. Note that since the mapping $-F(t)$ is psd-convex, it follows from Proposition 2 that the matrix C, of second-order derivatives of $F(t)$ at $t = 0$, is negative semidefinite. We have that $\alpha(t) = -\log\det F(t)$. It can be calculated then that

$$\frac{\mathrm{d}^2\alpha(t)}{\mathrm{d}t^2} = \operatorname{tr}(B + tC)F(t)^{-1}(B + tC)F(t)^{-1} - \operatorname{tr} CF(t)^{-1}. \tag{20}$$

Since C is negative semidefinite, it follows from (20) that $\mathrm{d}^2\alpha(t)/\mathrm{d}t^2 \geqslant 0$ for all $t \in \mathbb{R}$ such that $F(t)$ is positive definite. Convexity of $\alpha(t)$ then follows. □

3. Geometry of the cone of positive semidefinite matrices

In this section we discuss various geometrical properties of the cone $\mathcal{K}$ and their implications on the semidefinite program (P). We start with a discussion of the transversality condition. For a detailed study of the transversality concept and relevant references we refer to [10].

Let X and Y be two finite-dimensional vector spaces, $W \subset Y$ be a smooth manifold and $g : X \to Y$ be a smooth (differentiable) mapping.

Definition 4. It is said that g intersects W transversally at a point $x \in X$ (denoted by $g \pitchfork_x W$) if either (i) $g(x) \notin W$ or (ii) $g(x) \in W$ and

$$Y = T(W, g(x)) + [\mathrm{d}g(x)]X. \tag{21}$$

If $g \pitchfork_x W$ for all $x \in X$ we say that g intersects W transversally (denoted $g \pitchfork W$).

Here $T(W, y)$ denotes the tangent space to W at $y \in W$ and $\mathrm{d}g(x) : X \to Y$ is the linear mapping corresponding to the differential of g at x. Note that if $g(x) \in W$ and $g \pitchfork_x W$, then the following dimensionality condition holds

$$\dim W + \dim X \geqslant \dim Y. \tag{22}$$

Transversality is a *generic* property in the following sense. Let Π be a finite-dimensional vector space and let $\mathcal{G}(x, \pi)$ be a mapping $\mathcal{G} : X \times \Pi \to Y$. We can view Π as a space of parameters and for $\pi \in \Pi$ define the mapping $g_\pi(\cdot) = \mathcal{G}(\cdot, \pi)$. For a discussion of the following result see, e.g., [10].

Proposition 5. *Suppose that the mapping $\mathcal{G}$ is infinitely differentiable (jointly in x and π) and that $\mathcal{G} \pitchfork W$. Then for almost every $\pi \in \Pi$ the mapping g_π intersects W transversally. That is, those $\pi \in \Pi$ such that g_π does not intersects W transversally form a set of Lebesgue measure zero in Π.*

In particular it follows that if the dimensionality condition (22) is not satisfied, then for *almost every* π the set

$$g_\pi^{-1}(W) = \{x \in X \mid g_\pi(x) \in W\}$$

is empty. In that sense the dimensionality condition (22) is *generic*.

Consider now the set $\mathcal{W}_r$ of $n \times n$ symmetric matrices of rank r. It is well known that $\mathcal{W}_r$ forms a smooth manifold of dimension

$$\dim \mathcal{W}_r = \tfrac{1}{2}n(n+1) - \tfrac{1}{2}(n-r)(n-r+1)$$

in the linear space of $n \times n$ symmetric matrices. Moreover, if $A \in \mathcal{W}_r$, then the tangent space $T(\mathcal{W}_r, A)$ can be defined by linear equations (e.g. [4,30])

$$T(\mathcal{W}_r, A) = \{Z \in \mathcal{S}_n \mid e_i^{\mathrm{T}} Z e_j = 0,\ 1 \leqslant i \leqslant j \leqslant n-r\}, \tag{23}$$

where $e_1, \ldots, e_{n-r}$, is a basis of the null space of the matrix A. By using this characterization of the tangent space $T(\mathcal{W}_r, A)$ it is not difficult to prove the following result (cf. [30]).

Proposition 6. *Suppose that $G(x) \in \mathcal{W}_r$, let $e_1, \dots, e_{n-r}$, be a basis of the null space of the matrix $G(x)$ and $G_i = \partial G(x)/\partial x_i$, $i = 1, \dots, m$. Then $G \pitchfork_x \mathcal{W}_r$ if and only if the m-dimensional vectors $v_{ij} = (e_i^{\mathrm T} G_1 e_j, \dots, e_i^{\mathrm T} G_m e_j)$, $1 \leqslant i \leqslant j \leqslant n-r$, are linearly independent.*

It follows from the dimensionality condition (22) that the following lower bound on the rank r of $G(x)$ holds *generically* (in the sense explained above)

$$\tfrac{1}{2}(n-r)(n-r+1) \leqslant m. \tag{24}$$

In particular, if the mapping G is affine, i.e. $G(x) = A_0 + \sum_{i=1}^m x_i A_i$, then for *almost every* $A_0 \in \mathcal{S}_n$ the vectors v_{ij}, $1 \leqslant i \leqslant j \leqslant n-r$, defined in Proposition 6 (with $G_i = A_i$), are linearly independent and the rank r of $G(x)$ satisfies the inequality (24) for all $x \in \mathbb{R}^m$ (cf. [3,26,30]).

The transversality is an analogue of the condition of linear independence of the gradients of active constraints used in nonlinear programming (also called nondegeneracy condition in linear programming). It is a well-known phenomenon that for very large linear programs quite often an optimal solution tends to happen at *numerically* degenerate points. This indicates that the above generic statements, which hold "almost surely", should be taken with caution. It also should be remembered that the mathematical statement of Proposition 5 explicitly depends on the employed parameterization which is given by the corresponding mapping $\mathcal{G}$. In the above example of affine mapping G, the program is parameterized by the matrix A_0, i.e. the mapping $\mathcal{G} : \mathbb{R}^m \times \mathcal{S}_n \to \mathcal{S}_n$ is given by $\mathcal{G}(x, A) = A + \sum_{i=1}^n x_i A_i$. In this case we have that $\mathrm{d}\mathcal{G}$ maps $\mathbb{R}^m \times \mathcal{S}_n$ onto $\mathcal{S}_n$ and hence the transversality condition $\mathcal{G} \pitchfork \mathcal{W}_r$ holds for any $r = 0, \dots, n$.

Let us note that if $G(x) = \operatorname{diag}(g_{11}(x), \dots, g_{nn}(x))$ is diagonal, then $G(x) \preceq 0$ means that $g_{ii}(x) \leqslant 0$, $i = 1, \dots, n$. In that case program (P) becomes a nonlinear programming problem subject to a finite number of inequality constraints. However, here $v_{ij} = 0$ for $i \neq j$ and hence if $\operatorname{rank} G(x) \leqslant n-2$, then the transversality condition $G \pitchfork_x \mathcal{W}_r$ does not hold. In this case the transversality condition is not reduced to the condition of linear independence of gradients $\nabla g_{ii}(x)$ of active at x constraints. This shows that an analogy with nonlinear programming should be taken cautiously.

Consider now a positive semidefinite matrix A of rank r. This matrix belongs to a face $\mathcal{F}$ of $\mathcal{K}$ of dimension $\frac{1}{2}r(r+1)$ generated by matrices $a_i a_j^{\mathrm T} + a_j a_i^{\mathrm T}$, $1 \leqslant i \leqslant j \leqslant r$, where $a_1, \dots, a_r$, is a basis of the range space of A. Note that $\mathcal{F} \subset T(\mathcal{W}_r, A)$ and that

$$\dim T(\mathcal{W}_r, A) - \dim \mathcal{F} = r(n-r).$$

It follows that if $G(x)$ is an affine mapping and x is an extreme point of the feasible set $G^{-1}(-\mathcal{K})$ of the program (P), then the inequality

$$\tfrac{1}{2}r(r+1) \leqslant \tfrac{1}{2}n(n+1) - m \tag{25}$$

holds for the rank r of $G(x)$ (see [19] for details). If, in addition, the objective function $f(\cdot)$ is linear, then it attains its minimum value at an extreme point of the feasible set and hence the upper bound (25) for the corresponding rank r holds at that point. In particular, if $m \geqslant n+1$, then $r \leqslant n-2$ at an extreme optimal solution [2,19]. Note that if $m \geqslant n+1$, then the feasible set $G^{-1}(-\mathcal{K})$ has a face of dimension greater than or equal to one provided the Slater condition holds. In that case there exists a linear function $f(\cdot)$ such that the associated program (P) has more than one optimal solution. Note also that, in contrast to standard linear programming, the dimensionality condition $m \geqslant n+1$ is essential here.

Consider again a matrix $A \in \mathcal{K}$ of rank r and let E be an $n \times (n-r)$ matrix whose columns $e_1, \dots, e_{n-r}$, form a basis of the null space of A. Then the tangent cone $T(\mathcal{K}, A)$ to the convex cone $\mathcal{K}$ at A can be written in the form

$$T(\mathcal{K}, A) = \{Z \in \mathcal{S}_n \mid E^{\mathrm{T}} Z E \succeq 0\}. \tag{26}$$

This can be proved, for example, by the following arguments. First, the cone $\mathcal{K}$ can be written in the form $\mathcal{K} = \{B \in \mathcal{S}_n \mid \lambda_n(B) \geqslant 0\}$. Second, the directional derivative of $\lambda_n(\cdot)$ at A in a direction Z is given by the smallest eigenvalue of the $(n-r) \times (n-r)$ matrix $E^{\mathrm{T}}ZE$, provided the basis $e_1, \dots, e_{n-r}$ is orthonormal [14]. Finally, the cone $T(\mathcal{K}, A)$ is formed by those Z for which this directional derivative is nonnegative.

Note that it follows from (23) and (26) that $T(\mathcal{W}_r, A)$ is the lineality space of $T(\mathcal{K}, A)$. This demonstrates a difference between polyhedral cones and the cone $\mathcal{K}$. Here the lineality space of $T(\mathcal{K}, A)$ is larger than the linear space generated by the corresponding face $\mathcal{F}$ of dimension $\frac{1}{2}r(r+1)$, except in trivial cases when $r = 0$ or $r = n$.

Now let x_0 be an optimal solution of the program (P), with $\operatorname{rank} G(x_0) = r$, and suppose that the transversality condition $G \pitchfork_{x_0} \mathcal{W}_r$ holds. By (21) and since $T(\mathcal{W}_r, G(x_0)) \subset T(-\mathcal{K}, G(x_0))$ we obtain that

$$\mathcal{S}_n = T(-\mathcal{K}, G(x_0)) + [\mathrm{d}G(x_0)]\mathbb{R}^m. \tag{27}$$

Since the cone $\mathcal{K}$ (the cone $-\mathcal{K}$) has a nonempty interior, condition (27) is equivalent to the MF-condition (12) (e.g., [20, Proposition 3.9]). Therefore we obtain that the transversality condition implies the MF-condition and hence existence of a matrix $\Omega \succeq 0$ of Lagrange multipliers satisfying optimality conditions (13) and (14).

Let us show that under the transversality condition the Lagrange multipliers matrix is unique. Indeed, since $\Omega \succeq 0$ and because of (14), the matrix Ω can be represented in the form $\Omega = E\Psi E^{\mathrm{T}}$, where E is an $n \times (n-r)$ matrix whose columns form a basis of the null space of the matrix $G(x_0)$ and Ψ is an $(n-r) \times (n-r)$ symmetric positive semidefinite matrix. We have then that

$$\frac{\partial L(x_0, \Omega)}{\partial x_i} = \frac{\partial f(x_0)}{\partial x_i} + \Psi \bullet (E^{\mathrm{T}} G_i E), \quad i = 1, \dots, m,$$

where $G_i = \partial G(x_0)/\partial x_i$. It follows then from the linear independence condition of proposition 6 that, for a given matrix E, the matrix Ψ is defined uniquely by the equations (13). Thus we have:

Proposition 7. *Let x_0 be an optimal solution of the program* (P), $\operatorname{rank} G(x_0) = r$, *and suppose that the transversality condition $G \pitchfork_{x_0} \mathcal{W}_r$ holds. Then there exists a unique matrix $\Omega_0 \succeq 0$ of Lagrange multipliers satisfying the first-order optimality conditions.*

In the convex case the Lagrange multipliers matrices form the set of optimal solutions of the dual program (D). Therefore, in the convex case, if the primal problem (P) has an optimal solution at which the transversality condition holds, then the associated dual problem has a unique optimal solution. Since the transversality is a generic property, in the sense which was explained earlier, we obtain that generically the dual problem possesses a unique optimal solution.

The transversality condition is a sufficient condition for uniqueness of the Lagrange multipliers but in general is not necessary. For a derivation of sufficient and "almost" necessary conditions for uniqueness of Lagrange multipliers see [32]. In particular it is shown in [32] that if there exists a matrix $\Omega \succeq 0$ of Lagrange multipliers, satisfying optimality conditions (13) and (14), and the strict complementarity condition (18) holds, then the transversality condition is also necessary for uniqueness of these Lagrange multipliers.

4. Second-order analysis

In this section we discuss second-order optimality conditions for the program (P). For that purpose we employ the concept of second-order tangent set to $\mathcal{K}$ at a point A in a direction B, which can be defined as follows

$$T^2(\mathcal{K}, A, B) = \{Z \in \mathcal{S}_n \mid \operatorname{dist}(A + tB + \tfrac{1}{2}t^2 Z, \mathcal{K}) = o(t^2),\ t \geqslant 0\}. \tag{28}$$

It is clear from the above definition that $T^2(\mathcal{K}, A, B)$ can be nonempty only if $A \in \mathcal{K}$ and $B \in T(\mathcal{K}, A)$. In a sense $T^2(\mathcal{K}, A, B)$ provides a second-order approximation of the set $\mathcal{K}$ in a way similar to the (first-order) tangent cone $T(\mathcal{K}, A)$, which can be defined as a collection of $X \in \mathcal{K}$ such that $\operatorname{dist}(A + tX, \mathcal{K}) = o(t)$, $t \geqslant 0$.

Consider now the smallest-eigenvalue function $\lambda_n(\cdot)$. Recall that the cone $\mathcal{K}$ can be written in the form $\mathcal{K} = \{X \in \mathcal{S}_n \mid \lambda_n(X) \geqslant 0\}$. Let $A \in \mathcal{S}_n$ be a positive semidefinite matrix of rank $r < n$ and let E be an $n \times (n-r)$ matrix of rank $n - r$ such that $AE = 0$ and $E^{\mathrm{T}}E = I_{n-r}$, i.e. the column vectors $e_1, \ldots, e_{n-r}$, of E form an orthonormal basis of the null space of A. Consider an (approximate) second-order curve of the form $X(t) = A + tB + \frac{1}{2}t^2 Q + o(t^2)$, $t \geqslant 0$, where $B, Q \in \mathcal{S}_n$. It is known that

$$\lambda_n(X(t)) = t\lambda_n'(A, B) + \tfrac{1}{2}t^2 \lambda_n''(A, B, Q) + o(t^2), \tag{29}$$

where the first-order directional derivative $\lambda_n'(A, B)$ is given by (e.g., [14])

$$\lambda_n'(A, B) = \lambda_{\min}(E^{\mathrm{T}}BE), \tag{30}$$

and the second-order directional derivative $\lambda_n''(A, B, Q)$ is given by (e.g., [14, Theorem 10] and [28, pp. 227–228])

$$\lambda_n''(A,B,Q) = \min_{u\in\mathcal{U}}\left\{u^{\mathrm{T}}Qu - 2\sum_{i=1}^{r}\frac{(u^{\mathrm{T}}Ba_i)^2}{\lambda_i(A)}\right\}. \tag{31}$$

Here a_i, $i = 1,\dots,r$, are orthonormal eigenvectors of A corresponding to its nonzero eigenvalues, $\mathcal{U} = \{u \mid u = Ev, v \in \mathcal{V}\}$ and $\mathcal{V}$ is the set of orthonormal eigenvectors of the $(n-r)\times(n-r)$ matrix $E^{\mathrm{T}}BE$ corresponding to its smallest eigenvalue $\lambda_{\min}(E^{\mathrm{T}}BE)$.

Note that

$$\sum_{i=1}^{r}\frac{(u^{\mathrm{T}}Ba_i)^2}{\lambda_i(A)} = u^{\mathrm{T}}BA^{\dagger}Bu, \tag{32}$$

where $A^{\dagger} = \sum_{i=1}^{r}\lambda_i(A)^{-1}a_ia_i^{\mathrm{T}}$ is the Moore–Penrose inverse of A. Let s be the multiplicity of the smallest eigenvalue of the matrix $E^{\mathrm{T}}BE$ and let $v_1,\dots,v_s$ be a corresponding set of orthonormal eigenvectors and $V = [v_1,\dots,v_s]$ be the corresponding $(n-r)\times s$ matrix. It follows then from (31) and (32) that

$$\lambda_n''(A,B,Q) = \lambda_{\min}(V^{\mathrm{T}}E^{\mathrm{T}}(Q-2BA^{\dagger}B)EV). \tag{33}$$

Suppose now that $\lambda_{\min}(E^{\mathrm{T}}BE) = 0$, which implies that $B \in T(\mathcal{K},A)$. It is clear from (33) that there exists $Q_0 \in \mathcal{S}_n$ such that $\lambda_n''(A,B,Q_0) > 0$. It follows then that the second-order tangent set to $\mathcal{K}$ at A in the direction B can be written in the form [7, Proposition 4.1]

$$T^2(\mathcal{K},A,B) = \{Q \in \mathcal{S}_n \mid \lambda_n''(A,B,Q) \geqslant 0\}. \tag{34}$$

Together with (33) this implies that

$$T^2(\mathcal{K},A,B) = \{Q \in \mathcal{S}_n \mid V^{\mathrm{T}}E^{\mathrm{T}}QEV \succeq 2V^{\mathrm{T}}E^{\mathrm{T}}BA^{\dagger}BEV\}. \tag{35}$$

For a point $x_0 \in \mathbb{R}^m$, satisfying the first-order optimality conditions, consider the cone of critical directions

$$C(x_0) = \{y \in \mathbb{R}^m \mid y^{\mathrm{T}}\nabla f(x_0) = 0,\ [\mathrm{d}G(x_0)]y \in T(-\mathcal{K},G(x_0))\}. \tag{36}$$

This cone represents directions for which first-order approximations of the objective function and the constraint mapping do not ensure local optimality of the point x_0. Note that because of the optimality condition (13), the cone $C(x_0)$ can be written in the following equivalent form

$$C(x_0) = \{y \in \mathbb{R}^m \mid \Omega\bullet[\mathrm{d}G(x_0)]y = 0,\ [\mathrm{d}G(x_0)]y \in T(-\mathcal{K},G(x_0))\}, \tag{37}$$

where Ω is a matrix of Lagrange multipliers satisfying the first-order optimality conditions.

We can write now second-order necessary optimality conditions for the program (P) as follows (see [7, Theorem 4.2] and [12]). Let x_0 be a locally optimal solution of the program (P) and suppose that the MF-condition holds. Then to each $y \in C(x_0)$

corresponds a matrix $\Omega \succeq 0$ of Lagrange multipliers, satisfying the first-order optimality conditions (13) and (14), such that

$$y^{\mathrm{T}}\nabla^2_{xx}L(x_0,\Omega)y \geqslant \sigma(\Omega, T^2(-\mathcal{K}, G(x_0), [\mathrm{d}G(x_0)]y)), \tag{38}$$

where $\sigma(\cdot,\cdot)$ is the support function of the second-order tangent set $T^2(-\mathcal{K}, G(x_0), [\mathrm{d}G(x_0)]y)$ (see Eq. (2) for a definition of the support function). Note that the right-hand side term in (38) is always less than or equal to zero [7]. Therefore the inequality (38) is weaker than the inequality $y^{\mathrm{T}}\nabla^2_{xx}L(x_0,\Omega)y \geqslant 0$.

Let us calculate now the support function, given in the right hand side of (38), for a Lagrange multipliers matrix Ω and $y \in C(x_0)$. Denote $B = [\mathrm{d}G(x_0)]y$ and let $\operatorname{rank} G(x_0) = r$ with $r < n$. Let $e_1, \dots, e_n$ be a set of orthonormal eigenvectors of the matrix $G(x_0)$ corresponding to its eigenvalues arranged in the decreasing order and consider the $n \times (n-r)$ matrix $E = [e_1, \dots, e_{n-r}]$ formed from the eigenvectors of $G(x_0)$ corresponding to its zero eigenvalue. Note that E has full column rank $n-r$ and is such that $G(x_0)E = 0$ and $E^{\mathrm{T}}E = I_{n-r}$. Because of the complementarity condition (14), the matrix Ω can be represented in the from $\Omega = E\Psi E^{\mathrm{T}}$ with Ψ being an $(n-r)\times(n-r)$ symmetric positive semidefinite matrix. Moreover, since $y \in C(x_0)$ we have (see (37)) that $\Omega \bullet B = 0$ and hence $\Psi \bullet (E^{\mathrm{T}}BE) = 0$. Therefore $\Psi = V\Xi V^{\mathrm{T}}$, where Ξ is a positive semidefinite matrix. Using the characterization (35) of the corresponding second-order tangent set we obtain

$$\begin{aligned}
&\sigma(\Omega, T^2(-\mathcal{K}, G(x_0), B)) \\
&\quad = \sup\{\Omega \bullet Q \mid V^{\mathrm{T}}E^{\mathrm{T}}QEV \preceq 2V^{\mathrm{T}}E^{\mathrm{T}}B[G(x_0)]^{\dagger}BEV\}, \\
&\quad = \sup\{\Xi \bullet (V^{\mathrm{T}}E^{\mathrm{T}}QEV) \mid V^{\mathrm{T}}E^{\mathrm{T}}QEV \preceq 2V^{\mathrm{T}}E^{\mathrm{T}}B[G(x_0)]^{\dagger}BEV\}.
\end{aligned}$$

Consequently

$$\begin{aligned}
&\sigma(\Omega, T^2(-\mathcal{K}, G(x_0), B)) \\
&\quad = 2\Xi \bullet (V^{\mathrm{T}}E^{\mathrm{T}}B[G(x_0)]^{\dagger}BEV) = 2\Omega \bullet (B[G(x_0)]^{\dagger}B).
\end{aligned}$$

We have that $[\mathrm{d}G(x_0)]y = \sum_{i=1}^{m} y_i G_i(x_0)$, where $G_i(x_0) = \partial G(x_0)/\partial x_i$, $i = 1, \dots, m$, are the partial derivative matrices, and thus

$$\begin{aligned}
&\sigma(\Omega, T^2(-\mathcal{K}, G(x_0), [\mathrm{d}G(x_0)]y)) \\
&\quad = 2\sum_{i,j=1}^{m} y_i y_j \Omega \bullet (G_i(x_0)[G(x_0)]^{\dagger}G_j(x_0)).
\end{aligned} \tag{39}$$

Consider the $m \times m$ matrix $H(x_0,\Omega)$ whose elements are given by

$$[H(x_0,\Omega)]_{ij} = -2\Omega \bullet (G_i(x_0)[G(x_0)]^{\dagger}G_j(x_0)), \quad i,j = 1, \dots, m. \tag{40}$$

Note that the matrix $H(x_0,\Omega)$ can be written in the following equivalent form

$$H(x_0,\Omega) = -2\Big(\frac{\partial G(x_0)}{\partial x}\Big)^{\mathrm{T}}(\Omega \otimes [G(x_0)]^{\dagger})\Big(\frac{\partial G(x_0)}{\partial x}\Big), \tag{41}$$

and hence has rank less than or equal to $r(n-r)$. We can write now the second-order necessary conditions (38) in the following form.

Theorem 8. *Let x_0 be a locally optimal solution of the program* (P) *and suppose that the MF-condition holds. Then to each $y \in \mathcal{C}(x_0)$ corresponds a matrix $\Omega \succeq 0$ of Lagrange multipliers, satisfying the first-order optimality conditions* (13) *and* (14), *such that*

$$y^{\mathrm{T}}\nabla^2_{xx}L(x_0,\Omega)y + y^{\mathrm{T}}H(x_0,\Omega)y \geqslant 0. \tag{42}$$

Note that since $G(x_0)$ is negative semidefinite, it follows from the representation (41) that the matrix $H(x_0,\Omega)$ is positive semidefinite. Therefore, as it was mentioned earlier, the inequality (42) is weaker than the inequality $y^{\mathrm{T}}\nabla^2_{xx}L(x_0,\Omega)y \geqslant 0$.

Let us consider now the following particular case. Suppose that $G(x_0) \in \mathcal{W}_r$, the transversality condition $G \pitchfork_{x_0} \mathcal{W}_r$ holds and let $\Omega \succeq 0$ be a matrix of Lagrange multipliers, satisfying the first-order optimality conditions (13) and (14). Recall that it follows from the transversality condition that the Lagrange multipliers matrix Ω is unique. Since $T(\mathcal{W}_r, G(x_0))$ is the lineality space of $T(-\mathcal{K}, G(x_0))$ and by (37), it is not difficult to see that if the strict complementarity condition (18) holds, then the critical cone $\mathcal{C}(x_0) = \mathcal{L}(x_0)$, where

$$\mathcal{L}(x_0) = \{y \in \mathbb{R}^m \mid [\mathrm{d}G(x_0)]y \in T(\mathcal{W}_r, G(x_0))\}. \tag{43}$$

By using the characterization (23) of the tangent space $T(\mathcal{W}_r, G(x_0))$, the linear space $\mathcal{L}(x_0)$ can be also written in the form

$$\mathcal{L}(x_0) = \{y \in \mathbb{R}^m \mid \sum_{i=1}^{m} y_i E^{\mathrm{T}} G_i(x_0) E = 0\}. \tag{44}$$

We obtain the following result (cf. [18,30]).

Theorem 9. *Let x_0 be a feasible point of the program* (P) *and suppose that* rank $G(x_0) = r$ *and that the transversality condition $G \pitchfork_{x_0} \mathcal{W}_r$ holds. Let $\Omega_0 \succeq 0$ be the corresponding matrix of Lagrange multipliers, satisfying the first-order optimality conditions* (13) *and* (14), *and suppose that the strict complementarity condition* (18) *holds. Then second-order necessary conditions for x_0 to be a locally optimal solution of* (P), *are*

$$y^{\mathrm{T}}\nabla^2_{xx}L(x_0,\Omega_0)y + y^{\mathrm{T}}H(x_0,\Omega_0)y \geqslant 0, \quad \textit{for all } y \in \mathcal{L}(x_0). \tag{45}$$

Conversely, second-order sufficient conditions for x_0 to be a locally optimal solution of (P), *are that the strict version of the inequality* (45) *holds, i.e.*

$$y^{\mathrm{T}}\nabla^2_{xx}L(x_0,\Omega_0)y + y^{\mathrm{T}}H(x_0,\Omega_0)y > 0, \quad \textit{for all nonzero } y \in \mathcal{L}(x_0). \tag{46}$$

Proof. Necessity of the second-order conditions (45) follows from the more general result of Theorem 8. Sufficiency can be proved exactly in the same way as in [30, Theorem 3.1]. That is, first it can be shown that these sufficient conditions are the standard second-order sufficient conditions for x_0 to be a local minimizer of $f(x)$ over the manifold $G^{-1}(\mathcal{W}_r)$, which is smooth near x_0 because of the transversality condition. Second, because of the strict complementarity condition (18), first-order conditions can be employed to verify local optimality of x_0 in directions which are orthogonal to the tangent space to $G^{-1}(\mathcal{W}_r)$ at x_0 (see [30] for details). □

If the program (P) is convex, then the second-order sufficient conditions (46) imply that x_0 is the unique optimal solution of (P). Note that if the program (P) is linear, then the Lagrangian $L(x,\Omega)$ is linear in x and hence the first term in the left-hand side of the inequalities (45) and (46) vanishes. Nevertheless, even in the linear case the second term in those inequalities does not vanish and, in a sense, represents the curvature of the cone $\mathcal{K}$.

It is recently shown in [5] that the corresponding second-order conditions are also sufficient in the general setting of Theorem 8.

Let us finally remark the following. We have that the Lagrange multipliers matrix can be represented in the form $\Omega_0 = E\Psi_0 E^{\mathrm{T}}$ and that

$$[G(x_0)]^{\dagger} = \sum_{k=n-r+1}^{n} \lambda_k^{-1} e_k e_k^{\mathrm{T}}, \tag{47}$$

where $\lambda_k = \lambda_k(G(x_0))$. It follows then from (40) that the elements of the matrix $H(x_0,\Omega_0)$ can be written in the form

$$[H(x_0,\Omega_0)]_{ij} = -2\sum_{s,t=1}^{n-r} \psi_{st}\Big\{ \sum_{k=n-r+1}^{n} \lambda_k^{-1}(e_s^{\mathrm{T}} G_i e_k)(e_k^{\mathrm{T}} G_j e_t)\Big\}, \tag{48}$$

where ψ_{st} are elements of the matrix Ψ_0 and $G_i = G_i(x_0)$. Note that ψ_{ij} correspond to Lagrange multipliers which were used in [30, equations (3.4) and (3.11)]. Note also that since $G(x_0) \preceq 0$, the eigenvalues λ_k, $k = n-r+1,\dots,n$, are negative.

5. Sensitivity analysis

In this section we discuss differential properties of the optimal value and an optimal solution of a semidefinite program depending on a parameter vector. Let us start by considering the following semidefinite program

$$(\mathrm{P}_A) \qquad \min_{x\in\mathbb{R}^m} f(x) \quad \text{subject to } A + G(x) \preceq 0,$$

parameterized by $A \in \mathcal{S}_n$. Denote by $\psi(A)$ the optimal value of (P_A). By the definition, $\psi(A) = +\infty$ if the feasible set of the program (P_A) is empty. Suppose that the unperturbed program (P) is *convex*, i.e. $f(x)$ is convex and $G(x)$ is psd-convex. The optimal value function $\psi(A)$ can be analyzed then by a straightforward extension of a corresponding analysis used for nonlinear convex programs [22,23].

It is not difficult to verify that the function $\psi(\cdot)$ is convex on $\mathcal{S}_n$. Suppose that the optimal value of (P) is finite and that the Slater condition for the program (P) holds. Because of the Slater condition, it follows that the feasible set of the program (P_A) is nonempty, and hence $\psi(A) < +\infty$, for all A in a neighborhood of the zero matrix in $\mathcal{S}_n$. Since $\psi(0)$ is finite, we obtain then by convexity, that $\psi(A)$ is finite for all A in a neighborhood of zero. Consequently the subdifferential $\partial\psi(0)$ is nonempty and compact. Moreover, it is possible to show that $\partial\psi(0)$ coincides with the set of optimal solutions of the dual program (D) (see [23, Theorem 16]). Therefore we obtain that, in the present situation, the set $\mathcal{Z}$ of the optimal solutions of (D) is nonempty and compact, that the optimal value function $\psi(\cdot)$ is directionally differentiable at $A = 0$ and its directional derivative in a direction $D \in \mathcal{S}_n$ is given by [23, Theorem 17]

$$\psi'(0, D) = \max_{\Omega\in \mathcal{Z}} \Omega \bullet D. \tag{49}$$

It follows from (49) that the optimal value function is differentiable at $A = 0$ iff $\mathcal{Z} = \{\Omega_0\}$ is a singleton, i.e. the dual program (D) has a unique optimal solution. Since a finite valued convex function is differentiable almost everywhere [22], it follows that generically the dual program has a unique optimal solution. We already obtained this result by means of the transversality theory in Section 3.

Let us consider now a general case of the following semidefinite program

$$(\mathrm{P}_u) \qquad \min_{x\in\mathbb{R}^m} f(x,u) \quad \text{subject to } G(x,u) \preceq 0,$$

parameterized by vector $u \in \mathbb{R}^k$. By $L(x, \Omega, u)$ we denote the corresponding Lagrangian function,

$$L(x, \Omega, u) = f(x,u) + \Omega \bullet G(x,u).$$

We assume that the functions $f : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ and $G : \mathbb{R}^m \times \mathbb{R}^k \to \mathcal{S}_n$ are continuously differentiable. We also assume that $f(\cdot, u_0) = f(\cdot)$ and $G(\cdot, u_0) = G(\cdot)$, i.e. for $u = u_0$ the parameterized program coincides with the unperturbed program (P).

By $\psi(u)$ and $\bar{x}(u)$ we denote the optimal value and an optimal solution of the program (P_u), respectively. We assume the following, so-called *inf-compactness*, condition:

there exists a number $\alpha > \psi(u_0)$ and a compact set $S \subset \mathbb{R}^m$ such that

$$\{x \mid f(x,u) \leqslant \alpha,\ G(x,u) \preceq 0\} \subset S \tag{50}$$

for all u in a neighborhood of u_0. Note that if for all u the program (P_u) is convex and the set of optimal solutions of the unperturbed program (P) is nonempty and bounded, then the inf-compactness condition holds automatically.

In the following theorem we give a description of the first-order differential behavior of the optimal value function $\psi(u)$ in the convex case. This result can be viewed as an extension of (49) and is basically due to Gol'shtein [9] (see [31] for the required extension of Gol'shtein's result to the case of cone constraints).

Theorem 10. *Suppose that the unperturbed program* (P) *is convex, that the set* $\mathcal{M}$ *of optimal solutions of* (P) *is nonempty and bounded, that the Slater condition for the program* (P) *holds and that the inf-compactness condition is satisfied. Then the optimal value function* $\psi(\cdot)$ *is directionally differentiable at* u_0 *and its directional derivative at* u_0 *in a direction* $d \in \mathbb{R}^k$ *is given by*

$$\psi'(u_0, d) = \min_{x \in \mathcal{M}} \max_{\Omega \in \mathcal{Z}} d^{\mathrm{T}} \nabla_u L(x, \Omega, u_0), \tag{51}$$

where $\mathcal{Z}$ *is the set of optimal solutions of the dual program* (D).

Another interpretation of the set $\mathcal{Z}$ is that it is the set of Lagrange multipliers of the program (P). Because of the Slater condition this set is nonempty and compact. Note that if (P_u) is convex for all u, then the inf-compactness condition follows from the nonemptiness and compactness of $\mathcal{M}$ and can be omitted from the above proposition.

Suppose now that $f(x,u)$ and $G(x,u)$ are twice continuously differentiable and that the assumptions of Theorem 9 are satisfied. That is, $G \,\overline{\pitchfork}_{x_0}\, \mathcal{W}_r$ and the strict complementarity condition (18) holds. Then by continuity arguments, assuming that $\bar{x}(u)$ tends to x_0 as $u \longrightarrow u_0$, we obtain that the strict complementarity condition holds at $\bar{x}(u)$ and $G(\bar{x}(u)) \in \mathcal{W}_r$ for all u sufficiently close to u_0. The following characterization of the differentiability properties of $\bar{x}(u)$ is then a consequence of the Implicit Function Theorem (cf. [8,29]). For given $x_0 \in \mathbb{R}^m$ and $\Omega_0 \in \mathcal{S}_n$ consider the quadratic function

$$\begin{aligned}\kappa(y,d) = {} & y^{\mathrm{T}} \nabla^2_{xx} L(z_0) y + 2 y^{\mathrm{T}} \nabla^2_{xu} L(z_0) d + d^{\mathrm{T}} \nabla^2_{uu} L(z_0) d \\ & + y^{\mathrm{T}} H_{xx}(z_0) y + 2 y^{\mathrm{T}} H_{xu}(z_0) d + d^{\mathrm{T}} H_{uu}(z_0) d,\end{aligned} \tag{52}$$

where $z_0 = (x_0, \Omega_0, u_0)$ and $H_{xx}(z_0)$, $H_{xu}(z_0)$, $H_{uu}(z_0)$ are matrices of order $m \times m$, $m \times k$, $k \times k$, respectively, whose elements are defined by

$$\begin{aligned} [H_{xx}(z_0)]_{ij} &= -2\Omega_0 \bullet ([\partial G(x_0,u_0)/\partial x_i][G(x_0)]^{\dagger}[\partial G(x_0,u_0)/\partial x_j]), \\ [H_{xu}(z_0)]_{ij} &= -2\Omega_0 \bullet ([\partial G(x_0,u_0)/\partial x_i][G(x_0)]^{\dagger}[\partial G(x_0,u_0)/\partial u_j]), \\ [H_{uu}(z_0)]_{ij} &= -2\Omega_0 \bullet ([\partial G(x_0,u_0)/\partial u_i][G(x_0)]^{\dagger}[\partial G(x_0,u_0)/\partial u_j]). \end{aligned} \tag{53}$$

Note that the matrix $H_{xx}(z_0)$, defined above, coincides with the matrix $H(x_0, \Omega_0)$ defined in (40) and used in (45) and (46). Consider also the following linear space depending on $d \in \mathbb{R}^k$,

$$\mathcal{Y}(d) = \{y \in \mathbb{R}^m \mid [\mathrm{d}G(x_0,u_0)](y,d) \in T(\mathcal{W}_r, G(x_0,u_0))\}, \tag{54}$$

which can be written in the equivalent form

$$\mathcal{Y}(d) = \left\{ y \;\Big|\; \sum_{i=1}^{m} y_i E^{\mathrm{T}} \left[\frac{\partial G(x_0,u_0)}{\partial x_i} \right] E + \sum_{j=1}^{k} d_j E^{\mathrm{T}} \left[\frac{\partial G(x_0,u_0)}{\partial u_j} \right] E = 0 \right\}. \tag{55}$$

Note that $\mathcal{Y}(0) = \mathcal{L}(x_0)$, where the linear space $\mathcal{L}(x_0)$ is defined in (43) and (44).

Theorem 11. *Let x_0 be an optimal solution of the program* (P). *Suppose that the assumptions of Theorem* 9 *hold, that the second-order sufficient conditions* (46) *are satisfied and that $\bar{x}(u)$ tends to x_0 as $u \to u_0$. Then:*

(i) *the optimal value function is twice continuously differentiable at u_0 and its second-order Taylor expansion is given by*

$$\psi(u_0 + d) = \psi(u_0) + d^{\mathrm{T}}\nabla_u L(x_0, \Omega_0, u_0) + \frac{1}{2} \min_{y \in \mathcal{Y}(d)} \kappa(y, d) + o(\|d\|^2), \tag{56}$$

(ii) *the optimal solution $\bar{x}(u)$ is continuously differentiable at u_0 and*

$$\bar{x}(u_0 + d) = \bar{x}(u_0) + \arg\min_{y \in \mathcal{Y}(d)} \kappa(y, d) + o(\|d\|). \tag{57}$$

There is an interesting application of the above sensitivity results to deriving asymptotics in statistical theory. Consider the parameterized program (P_A) and suppose that the parameter matrix $A = A_N$ is *random*. Suppose further that A_N converges in probability, as $N \to \infty$, to a matrix A_0 and that $N^{1/2}(a_N - a_0) \Rightarrow \mathrm{N}(0, \Gamma)$. That is, A_N has an asymptotically normal distribution with the covariance matrix Γ. Here $a_N = \mathrm{vec}\, A_N$, $a_0 = \mathrm{vec}\, A_0$ and the symbol "$\Rightarrow$" denotes convergence in distribution. Note that since a_N has only $\frac{1}{2}n(n+1)$ different elements, the rank of the $n^2 \times n^2$ covariance matrix Γ is $\frac{1}{2}n(n+1)$ at most. For example, A_N can be a sample covariance matrix calculated from a sample of size N. In that case, if the sample is drawn from a normally distributed population with the population covariance matrix A_0, then $\Gamma = 2M_n(A_0 \otimes A_0)M_n$, where M_n is a symmetric idempotent pattern matrix of rank $\frac{1}{2}n(n+1)$ (see, e.g., [6]).

Suppose further that the program (P) is convex, that the Slater condition for the program (P_{A_0}) holds and that the optimal value $\psi(A_0)$ is finite. It follows then from (49) that

$$\psi(A_N) - \psi(A_0) = \max_{\Omega \in \mathcal{Z}(A_0)} \Omega \bullet (A_N - A_0) + o(\|A_N - A_0\|), \tag{58}$$

where $\mathcal{Z}(A_0)$ denotes the set of optimal solutions of the dual of the program (P_{A_0}). Consequently, by the Delta Theorem (see, e.g., [24, pp. 259–260]),

$$N^{1/2}(\psi(A_N) - \psi(A_0)) \Rightarrow \max_{\Omega \in \mathcal{Z}(A_0)} \Omega \bullet Y, \tag{59}$$

with $y \sim \mathrm{N}(0, \Gamma)$, $y = \mathrm{vec}\, Y$. It follows that $\psi(A_N)$ has an asymptotically normal distribution iff $\mathcal{Z}(A_0) = \{\Omega_0\}$ is a singleton, in which case

$$N^{1/2}(\psi(A_N) - \psi(A_0)) \Rightarrow \mathrm{N}(0, \sigma^2), \tag{60}$$

where $\sigma^2 = \omega_0^{\mathrm{T}} \Gamma \omega_0$ and $\omega_0 = \mathrm{vec}\, \Omega_0$. In particular, if $\Gamma = 2M_n(A_0 \otimes A_0)M_n$, then $\sigma^2 = 2\,\mathrm{tr}[(\Omega_0 A_0)^2]$. For the minimum trace factor analysis this result was derived in [25].

Consider now the optimal solutions $\hat{x}_N = \bar{x}(A_N)$ and $x_0 = \bar{x}(A_0)$. Suppose that the regularity conditions (specified in Theorem 11) ensuring differentiability of the optimal solution $\bar{x}(A)$ at $A = A_0$, are satisfied. It follows then by the Delta Theorem that

$N^{1/2}(\hat{x}_N - x_0) \Rightarrow \mathrm{N}(0, \Psi)$. The covariance matrix Ψ can be calculated by using (57) although its explicit expression is messy.

References

[1] F. Alizadeh, Combinatorial optimization with interior point methods and semi-definite matrices, Ph.D. Thesis, University of Minnesota, 1991.

[2] F. Alizadeh, J.-P.A. Haeberly and M.L. Overton, Primal-dual interior-point methods for semidefinite programming, preprint, presented at the *XV Symposium on Mathematical Programming*, Ann Arbor, 1994.

[3] F. Alizadeh, J.-P.A. Haeberly and M.L. Overton, Complementarity and nondegeneracy in semidefinite programming, *Mathematical Programming* 77 (1997) 111-128 (this issue).

[4] V.I. Arnold, On matrices depending on parameters, *Russian Mathematical Surveys* 26 (1971) 29-43.

[5] J.F. Bonnans, R. Cominetti and A. Shapiro, Second order necessary and sufficient optimality conditions under abstract constraints, Preprint.

[6] M.W. Browne, Generalized least squares estimators in the analysis of covariance structures, *South African Statistical Journal* 8 (1974) 1-24.

[7] R. Cominetti, Metric regularity, tangent sets and second order optimality conditions, *Applied Mathematics and Optimization* 21 (1990) 265-287.

[8] A.V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming* (Academic Press, New York, 1983).

[9] E.G. Gol'shtein, *Theory of Convex Programming*, Translations of Mathematical Monographs, Vol. 36 (American Mathematical Society, Providence, RI, 1972).

[10] M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities* (Springer, New York, 1973).

[11] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, 1991).

[12] H. Kawasaki, An envelope-like effect of infinitely many inequality constraints on second order necessary conditions for minimization problems, *Mathematical Programming* 41 (1988) 73-96.

[13] S. Kurcyusz, On the existence and nonexistence of Lagrange multipliers in Banach spaces, *Journal of Optimization Theory and Applications* 20 (1976) 81-110.

[14] P. Lancaster, On eigenvalues of matrices dependent on a parameter, *Numerische Mathematik* 6 (1964) 377-387.

[15] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications* (Academic Press, New York, 1979).

[16] H. Maurer and J. Zowe, First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems, *Mathematical Programming* 16 (1979) 98-110.

[17] Y.E. Nesterov and A.S. Nemirovskii, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM Studies in Applied Mathematics (SIAM, Philadelphia, PA, 1994).

[18] M.L. Overton and R.S. Womersley, Second derivatives for optimizing eigenvalues of symmetric matrices, *SIAM Journal on Matrix Analysis and Applications* 16 (1995) 697-718.

[19] G. Pataki, On the multiplicity of optimal eigenvalues, Management Science Research Report #MSRR-604, GSIA, Carnegie Mellon University.

[20] J.P. Penot, On regularity conditions in mathematical programming, *Mathematical Programming Study* 19 (1982) 167-199.

[21] S.M. Robinson, First order conditions for general nonlinear optimization, *SIAM Journal on Applied Mathematics* 30 (1976) 597-607.

[22] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[23] R.T. Rockafellar, *Conjugate Duality and Optimization*, Regional Conference Series in Applied Mathematics (SIAM, Philadelphia, PA. 1974).

[24] R.Y. Rubinstein and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method* (John Wiley and Sons, New York, NY, 1993).

[25] A. Shapiro, Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis, *Psychometrika* 47 (1982) 187-199.

[26] A. Shapiro, On the unsolvability of inverse eigenvalue problems almost everywhere, *Linear Algebra and Its Applications* 49 (1983) 27-31.

[27] A. Shapiro, Extremal problems on the set of nonnegative definite matrices, *Linear Algebra and Its Applications* 67 (1985) 7–18.

[28] A. Shapiro, Perturbation theory of nonlinear programs when the set of optimal solutions is not a singleton, *Applied Mathematics and Optimization* 18 (1988) 215–229.

[29] A. Shapiro, Sensitivity analysis of nonlinear programs and differentiability properties of metric projections, *SIAM J. Control and Optimization* 26 (1988) 628–645.

[30] A. Shapiro and M.K.H. Fan, On eigenvalue optimization, *SIAM J. on Optimization* 5 (1995) 552–569.

[31] A. Shapiro, Directional differentiability of the optimal value function in convex semi-infinite programming, *Mathematical Programming* 70 (1995) 149–157.

[32] A. Shapiro, On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints, *SIAM J. on Optimization*, to appear.