Journal of Convex Analysis
Volume 10 (2003), No. 1, 1–19

First Order Conditions for Ideal Minimization of Matrix-Valued Problems

L. M. Graña Drummond*
Programa de Engenharia de Sistemas de Computação, COPPE–UFRJ, CP 68511, Rio de Janeiro, RJ, 21945–970, Brazil
lmgd@cos.ufrj.br

A. N. Iusem†
Instituto de Matemática Pura e Aplicada (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil
iusp@impa.br

Received August 20, 1999
Revised manuscript received June 13, 2001

The aim of this paper is to study first order optimality conditions for ideal efficient points in the Löwner partial order, when the data functions of the minimization problem are differentiable and convex with respect to the cone of symmetric semidefinite matrices. We develop two sets of first order necessary and sufficient conditions. The first one, formally very similar to the classical Karush-Kuhn-Tucker conditions for optimization of real-valued functions, requires two constraint qualifications, while the second one holds just under a Slater-type one. We also develop duality schemes for both sets of optimality conditions.

Keywords: Vector optimization, Löwner order, ideal efficiency, first order optimality conditions, convex programming, duality

1. Introduction

In this paper we discuss first order optimality conditions for Löwner inequality constrained problems, where both the objective and the constraint mappings are positive semidefinite-convex (psd-convex, henceforth), i.e. convex with respect to the Löwner order, and the minimization is carried out in the sense of ideal efficiency. We derive two sets of optimality conditions, each of which is both sufficient and necessary for optimality, when the data are psd-convex. One of them, that holds under a Slater-type constraint qualification, gives rise to multipliers which do not belong to a finite dimensional vector space. The other one, formally very similar to the classical Karush-Kuhn-Tucker conditions for optimization of real-valued functions, produces multipliers which do belong to a finite dimensional space, but requires an additional and stronger constraint qualification, akin to linear independence of the gradients of the active constraints in the real-vualed case, which, in particular, implies uniqueness of the dual solution. We develop duality schemes (including strong duality results) for both sets of optimality conditions.

For vector optimization, i.e. when the underlying partial order is induced by an arbitrary pointed closed convex cone, the matter of first (and second) order conditions has been widely studied for the case of Pareto and weak Pareto minimization on finite dimensional

*The work of this author was supported by a post-doctoral fellowship within the Department of Mathematics of the University of Haifa.
†The work of this author was partially supported by CNPq grant no. 301280/86.

ISSN 0944-6532 / $ 2.50

linear spaces (see [18], [6] and [7]). Existence and uniqueness of Lagrange-Kuhn-Tucker multipliers for efficient points in Banach spaces are considered in [1] and [19]. For multiobjective optimization in finite dimensional vector spaces, that is to say, for efficiency in the componentwise partial order, first order conditions, together with the issue of constraint qualifications, are studied in [13] and [14]. For second order optimality conditions in multicriteria we refer to [4].

The particular case of first order necessary and sufficient conditions for minimization of real-valued functions under a single symmetric matrix valued-constraint has already been systematically studied in [16], where existence and uniqueness of Lagrange multipliers for problems with optimal solutions is established under a transversality assumption, which is an analogous of the condition of linear independence of the gradients of active constraints used in classical nonlinear optimization. In [17] the same author shows that a necessary condition for uniqueness of a matrix of Lagrange multipliers (satisfying optimality conditions) is precisely the transversality condition. In [16] it is shown that for the real-valued objective case, under a Slater-type constraint qualification, the duality gap between the primal and the dual problems is zero; second order conditions and a sensitivity analysis of such problems are also given. We emphasize that neither [16] nor [17] consider the case of symmetric matrix-valued objectives, i.e. situations where the minimization is performed with respect to the partial order induced by the positive semidefinite cone of symmetric matrices (see [9]), which are the main subject of our work.

In [8], the authors also consider the problem of minimizing a real-valued convex function subject to a single psd-constraint. Their paper is devoted to the issues of welldefinedness and convergence of the central path associated to the primal-dual pair of problems, and, as a by product, first order conditions for existence of optimal solutions and compactness of the optimal set are derived. However, these conditions are unrelated to our current subject.

2. First formulation of the optimality conditions

Consider S^m, the subspace of $\mathbb{R}^{m\times m}$ consisting of all the symmetric matrices, with the inner product given by $\langle X,Y\rangle = \mathrm{tr}[XY] = \sum_{i,j=1}^m X_{ij}Y_{ij}$, and let S_+^m be the cone of positive semidefinite symmetric matrices of order m. Observe that S_+^m is a pointed closed convex cone with nonempty interior, given by the set of positive definite symmetric matrices S_{++}^m. Furthermore, we have that $(S_+^m)^*$, the positive dual or polar cone of S_+^m, i.e. the cone of matrices which form an accute angle with every symmetric positive semidefinite matrix, coincides with S_+^m, so that S_+^m is self-polar; that is to say, $X \in S_+^M$ if and only if $\mathrm{tr}(XY) \geq 0$ for all $Y \in S_+^m$ (this result is known as Fejér's Theorem, see Proposition 2.1 below).

The Löwner partial order is defined in the following way: given $X, Y \in S^m$, we say that $X \preceq Y$ if and only if $Y - X \in S_+^m$, i.e. $Y - X$ belongs to the cone S_+^m of symmetric positive semidefinite definite matrices in $\mathbb{R}^{m\times m}$. We also say that $X \prec Y$ if and only if $Y - X \in int(S_+^m)$, i.e. $Y - X$ belongs to the cone S_{++}^m of symmetric positive definite matrices of order m.

Take now differentiable mappings $f : \mathbb{R}^n \to S^m$, $g_i : \mathbb{R}^n \to S^{r_i}$ $(1 \leq i \leq p)$ and consider problem (CP):

$$\min f(x)$$

$$\text{s.t.} \quad g_i(x) \preceq 0 \ \ (1 \le i \le p) \tag{1}$$

Let $F \subset \mathbb{R}^n$ be the feasible set for problem (CP), i.e. $F = \{x \in \mathbb{R}^n : g_i(x) \preceq 0 \ (1 \le i \le p)\}$. We will discuss in this paper the case of ideal minimal (or ideal efficient) solutions of (CP). In other words, (CP) consists of finding $x^* \in F$ such that $f(x) - f(x^*)$ is positive semidefinite for all $x \in F$, or, in compact notation, $f(x^*) \preceq f(x)$ for all $x \in F$.

We recall that a mapping $h : \mathbb{R}^n \to S^m$ is positive semidefinite-convex (psd-convex) if $h(\lambda x + (1-\lambda)y) \preceq \lambda h(x) + (1-\lambda)h(y)$ for all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$. Clearly, h is psd-convex iff $x \mapsto \psi_Z(x) := tr(Zh(x))$ is a convex scalar function for all $Z \in S^m_+$. In particular, x^* is a Löwner ideal minimizer of h in an open set if and only if

$$\frac{\partial \psi_Z}{\partial x_i}(x^*) = 0. \tag{2}$$

We also recall a couple of well known results on psd-convexity. For a differentiable mapping h, psd-convexity can be characterized by a sort of extension of the classical gradient inequality, namely

$$dh(x)(y-x) + h(x) \preceq h(y) \tag{3}$$

for all $x, y \in \mathbb{R}^n$, where $dh(x)y = \sum_{i=1}^n y_i \frac{\partial h}{\partial x_i}(x)$, see e.g. Theorem 6.1 of [3].

Another important property of differentiable psd-convex mappings is the fact that they are automatically continuously differentiable. In fact, if h is psd-convex and differentiable, clearly the real-valued function $\phi_u(x) := u^t h(x) u$ is convex in the usual sense, so ϕ_u will be continuously differentiable (in general, convex functions are differentiable almost everywhere). Therefore, h will be continuosly differentiable, since we have the following identity

$$h(x)_{ij} = (e^i)^t h(x) e^j = (\phi_{e^i}(x) + \phi_{e^j}(x) - \phi_{e^i - e^j}(x))/2$$

where e^i stands for the i-th canonical vector.

We will consider two formulations of first order optimality conditions for problem (CP). When f and all the g_i's are psd-convex, both of them generate sufficient conditions. Regarding necessity, and assuming psd-convexity of the g_i's, the second formulation holds under a standard constraint qualification, namely a Slater's condition, while the first one requires an additional and rather strong constraint qualification, akin to linear independence of the gradients of the active constraints in the real-valued case, which in particular implies uniqueness of the dual solution. On the other hand, in the first formulation the dual variables, or multipliers, belong to a finite dimensional vector space, which is not the case in the second formulation. Also, the conditions of the first formulation are formally simpler and have a closer ressemblance to the KKT conditions for the case of real-valued functions. We will give an example showing that the first formulation does not hold in the absence of the stronger constraint qualification.

Let $L(S^{r_i}, S^m)$ be the set of linear transformations from S^{r_i} to S^m. Following the terminology used in [12], p. 81, we say that $\mathbf{Z} \in L(S^{r_i}, S^m)$ is nondecreasing (denoted as $\mathbf{Z} \succeq 0$) if $\mathbf{Z}(A) \in S^m_+$ for all $A \in S^{r_i}_+$.

The first formulation of first order optimality conditions, for $x \in \mathbb{R}^n$ and $\mathbf{Z}^i \in L(S^{r_i}, S^m)$ $(1 \le i \le p)$ is given by:

$$df(x) + \sum_{i=1}^p \mathbf{Z}^i dg_i(x) = 0 \qquad \text{Lagrangian condition,} \tag{4}$$

$$g_i(x) \preceq 0 \quad (1 \leq i \leq p) \qquad \text{primal feasibility,} \tag{5}$$

$$\mathbf{Z}^i \succeq 0 \quad (1 \leq i \leq p) \qquad \text{dual feasibility,} \tag{6}$$

$$\mathbf{Z}^i(g_i(x)) = 0 \quad (1 \leq i \leq p) \qquad \text{complementarity.} \tag{7}$$

We observe that, under differentiability of f, $df(x)$ is a linear transformation from $\mathbb{R}^n$ to S^m which takes at the point $y \in \mathbb{R}^n$ the value $df(x)y \in S^m$, defined as the direccional derivative of $f(x)$ in the direccion y, i.e. $[df(x)y]_{k\ell} = \langle \nabla f(x)_{k\ell}, y \rangle$. By the same token, $dg_i(x)$ is a linear tranformation from $\mathbb{R}^n$ to S^{r_i}. Then, the expression $\mathbf{Z}^i dg_i(x)$ in (4) is understood as the linear transformation from $\mathbb{R}^n$ to S^m which takes at a vector $y \in \mathbb{R}^n$ the value $\mathbf{Z}^i(dg_i(x)y)$, a matrix in S^m.

We close this section with some elementary results, whose proof will be included to make the paper somewhat more self-contained. Item (i) in the following proposition is known as Fejér's Theorem (cf. [11]).

Proposition 2.1. *(i)* X *belongs to* S^m_+ *if and only if* $tr(XY) \geq 0$ *for all* $Y \in S^m_+$.
(ii) *If* X *and* Y *belong to* S^m_+ *and* $tr(XY) = 0$ *then* $XY = 0$.
(iii) *If* $g : S^p \to S^q$ *is psd-convex and* $\mathbf{Z} \in L(S^q, S^m)$ *is nondecreasing, then* $h : S^p \to S^m$ *defined as* $h(X) = \mathbf{Z}(g(X))$ *is psd-convex.*

Proof. (i) We prove first the "if" part. Take any $u \in \mathbb{R}^m$ and let $Y = uu^t \in S^m$. Then Y belongs to S^m_+ and so $0 \leq$tr$(XY) = u^t X u$. It follows that $X \in S^m_+$. For the "only if" part, we recall that any $Y \in S^m$ can be written as $Y = U^t D U$ where D is a diagonal matrix, with the eigenvalues of Y, say $\lambda_1, \ldots, \lambda_m$, in its diagonal, and U is an orthonormal matrix, whose k-th row, say u^k, is an eigenvector of Y with eigenvalue λ_k. With this notation, it is easy to check that

$$\mathrm{tr}(XY) = \sum_{i=1}^{m} \lambda_i \left[(u^i)^t X u^i\right] \tag{8}$$

for all $X \in \mathbb{R}^{m \times m}$. Then for $i \in \{1, \ldots, m\}$, we get $(u^i)^t X u^i \geq 0$, because $X \in S^m_+$, and $\lambda_i \geq 0$, because $Y \in S^m_+$. It follows from (8) that $\mathrm{tr}(XY) \geq 0$.

(ii) Write $Y = U^t D U$, with D and U as above. Since $\{u^1, \ldots, u^m\}$ is a basis of $\mathbb{R}^m$, it suffices to prove that $XYu^i = 0$ $(1 \leq i \leq m)$. This is certainly true if $u^i \in \mathrm{Ker}(Y)$. Otherwise, since $Y \in S^m_+$, we have $Yu^i = \lambda_i u^i$ with $\lambda_i > 0$. We get from $\mathrm{tr}(XY) = 0$ and (8) that, for $i \in \{1, \ldots, m\}$, $0 = \lambda_i (u^i)^t X u_i$, implying that $0 = (u^i)^t X u_i$, so that $Xu^i = 0$, because $X \in S^m_+$, and therefore $XYu^i = \lambda_i X u^i = 0$.

(iii) Since g is psd-convex, we have, for all $X, Y \in S^m$ and all $\lambda \in [0, 1]$,

$$\lambda g(X) + (1 - \lambda) g(Y) - g(\lambda X + (1 - \lambda) Y) \in S^m_+. \tag{9}$$

Since $\mathbf{Z}$ is nondecreasing and linear, we have

$$\lambda h(X) + (1 - \lambda) h(Y) - h(\lambda X + (1 - \lambda) Y) =$$
$$\lambda \mathbf{Z}(g(X)) + (1 - \lambda) \mathbf{Z}(g(Y)) - \mathbf{Z}[g(\lambda X + (1 - \lambda) Y)] =$$
$$\mathbf{Z}[\lambda g(X) + (1 - \lambda) g(Y) - g(\lambda X + (1 - \lambda) Y)] \in S^m_+, \tag{10}$$

using (9) in the last inequality. Therefore, psd-convexity of h follows immediately from (10).

□

3. Interpretations of the first formulation

We now consider a particular case of problem (CP), which will furnish a couple of interesting interpretations of conditions (4)-(7). Take $m = 1$, $p = 1$ and $r_1 = r$, i.e., we are considering the problem of minimizing a real-valued function subject to a single positive semidefinite constraint:

$$\min f(x) \tag{11}$$

$$\text{s.t.} \quad g(x) \preceq 0\,, \tag{12}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to S^r$. We mention that we have taken $p = 1$ just for the sake of a simpler notation; the discussion which follows works equally well with any p. Indeed, taking $p = 1$ in (CP) entails no loss of generality at all: the constraints in (1) can be replaced by a unique equivalent constraint $G(x) \preceq 0$, with $G : \mathbb{R}^n \to S^q$, where $q = \sum_{i=1}^p r_i$. We have kept, in our presentation, several constraints, instead of just one, in order to make (CP) look closer to the standard real-valued convex programming problem.

The Wolfe dual problem associated with (CP) is given by (DP)

$$\max f(x) + \operatorname{tr}(Zg(x))$$

$$\text{s.t.} \quad \frac{\partial f}{\partial x_i}(x) + \operatorname{tr}\left(Z\frac{\partial g}{\partial x_i}(x)\right) = 0 \quad (1 \leq i \leq n)\,,$$

where the variables are pairs $(x, Z) \in \mathbb{R}^n \times S^r$.

Let us now see how conditions (4)-(7) do look in this particular case. First of all, observe that, since $S^1 = \mathbb{R}$, we have that $L(S^r, S^1)$ is the dual space of S^r, so that the multiplier $\mathbf{Z}$ is a linear form and, therefore, there exists a symmetric r-matrix, wich we will also call Z, such that $\mathbf{Z}(X) = \operatorname{tr}(ZX)$. Moreover, the nondecreasing condition for $\mathbf{Z}$ simply means that $\operatorname{tr}(ZX) \geq 0$ for all $X \in S^r_+$, or, in view of Proposition 2.1(i), that Z belongs to S^r_+. Finally, note that, from the fact that Z and $-g(x)$ are positive semidefinite, condition (7) is equivalent, by Proposition 2.1(ii), to $Zg(x) = 0$. Therefore, conditions (4)-(7) become

$$\frac{\partial f}{\partial x_i}(x) + \operatorname{tr}\left[Z\frac{\partial g}{\partial x_i}(x)\right] = 0 \quad (1 \leq i \leq n), \tag{13}$$

$$g(x) \preceq 0, \tag{14}$$

$$Z \succeq 0, \tag{15}$$

$$Zg(x) = 0. \tag{16}$$

A weak duality result for the pair of primal-dual problems (CP) and (DP) has been presented in Proposition 1 in [8]. It implies that for all pair (x, Z) that satisfies conditions (13)-(14) it holds that x is optimal for (CP) and (x, Z) is optimal for (DP).

In order to obtain another interpretation for the multiplier Z, observe now that problem (CP) can be written as the following equivalent scalar-valued problem (SP)

$$\min f(x)$$

$$\text{s.t.} \quad \lambda_i(g(x)) \leq \ 0 \quad (1 \leq i \leq r)\,, \tag{17}$$

where $\lambda_i(g(x))$ stands for the i-th eigenvalue of $g(x)$.

Next we will see how the multiplier Z is related to problem (SP). Actually, we will see that the eigenvalues of Z are precisely the Karush-Kuhn-Tucker multipliers of problem (SP).

Proposition 3.1. *For the problem given by* (11)-(12), *conditions* (4)–(7) *are precisely the KKT conditions for problem (SP), with the i-th eigenvalue* $\lambda_i(Z)$ *of* Z *given by* (13)–(16) *as multiplier of the i-th constraint in* (17).

Proof. As we already observed, conditions (4)–(7) reduce, in this particular case, to conditions (13)–(16). We observe now that conditions (14) and (15) can respectively be written as

$$\lambda_i(g(x)) \leq 0 \ \ (1 \leq i \leq r) \ , \tag{18}$$

and

$$\lambda_i(Z) \geq 0 \ \ (1 \leq i \leq r) \ , \tag{19}$$

where $\lambda_i(Z)$ and $\lambda_i(g(x)$ are the i-th eigenvalues of Z and $g(x)$, respectively.

Let us now take a closer look at the complementarity condition (16); transposing on both sides of the equation yields, in view of the symmetry of Z and $g(x)$, $g(x)Z = 0$. Hence, Z and $g(x)$ commute, and, therefore, they are simultaneously diagonizable, which means that there exists an orthogonal matrix $P(x)$ of order r, such that $Z = P^t(x)D_Z P(x)$ and $g(x) = P^t(x)D_{g(x)}P(x)$, where D_Z and $D_{g(x)}$ are diagonal matrices, with the eigenvalues of Z and $g(x)$ in their diagonals, respectively. In view of this, condition (16) can be written as

$$\lambda_i(Z)\lambda_i(g(x)) = 0 \ \ (1 \leq i \leq r) \ . \tag{20}$$

Finally, let us focuse our attention on the Lagrangian condition (13). We claim that

$$\mathrm{tr}\left[Z\frac{\partial g}{\partial x_i}(x)\right] = \sum_{j=1}^{r}\left(P(x)\frac{\partial g}{\partial x_i}(x)P^t(x)\right)_{jj}\lambda_j(Z). \tag{21}$$

Indeed, we know that for three arbitrary square matrices A, B and C we have $\mathrm{tr}[ABC] = \mathrm{tr}[B^t A^t C^t]$; so, taking $A = \frac{\partial g}{\partial x_i}(x)$, $B = P^t(x)$ and $C = D_Z P(x)$, and using the fact that $Z = P^t(x)D_Z P(x)$, we obtain (21).

For simplicity, from now on we will drop the variable x in our formulae, so we have that $g = P^t D_g P$, or, equivalently, $D_g = PgP^t$. Hence,

$$\frac{\partial (PgP^t)_{jj}}{\partial x_i} = \frac{\partial \lambda_j(g)}{\partial x_i}.$$

So, we have

$$\frac{\partial \lambda_j(g)}{\partial x_i} = \frac{\partial(\sum_{k,l} P_{jl}\, g_{lk} P_{jk})}{\partial x_i} = \sum_{k,l}\left[\left(\frac{\partial P_{jl}}{\partial x_i}\, g_{lk} + P_{jl}\frac{\partial g_{lk}}{\partial x_i}\right)P_{jk} + P_{jl}\, g_{lk}\frac{\partial P_{jk}}{\partial x_i}\right]$$

$$= \left(\frac{\partial P}{\partial x_i}\, gP^t\right)_{jj} + \left(P\frac{\partial g}{\partial x_i}P^t\right)_{jj} + \left(P\, g\frac{\partial P^t}{\partial x_i}\right)_{jj}. \tag{22}$$

Since $P^tP = I$ and $PgP^t = D_g$, we get from (22)

$$\left(P\frac{\partial g}{\partial x_i}P^t\right)_{jj} = \frac{\partial \lambda_j(g)}{\partial x_i} - \left(\frac{\partial P}{\partial x_i}P^tD_g\right)_{jj} - \left(D_gP\frac{\partial P^t}{\partial x_i}\right)_{jj}. \tag{23}$$

Observe now that $PP^t = I$ implies that $\frac{\partial P}{\partial x_i}P^t = -P\frac{\partial P^t}{\partial x_i}$. Then, in a similar way, we obtain from (23)

$$\left(P\frac{\partial g}{\partial x_i}P^t\right)_{jj} = \frac{\partial \lambda_j(g)}{\partial x_i}, \tag{24}$$

using $(\frac{\partial P}{\partial x_i}P^tD_g)_{jj} = (D_gP\frac{\partial P^t}{\partial x_i})_{jj}$, which is a consequence of the fact that Dg is a diagonal matrix. Therefore, combining (21) and (24), we get the following expression for the Lagrangian condition (13):

$$\operatorname{tr}\left[Z\frac{\partial g}{\partial x_i}(x)\right] + \sum_{j=1}^{r}\frac{\partial \lambda_j(g(x))}{\partial x_i}\lambda_j(Z) = 0 \quad (1 \le i \le n). \tag{25}$$

Whence from (18), (19), (20) and (25), we have that conditions (4)–(7) (or equivalently (13)–(16)) are equivalent to

$$\operatorname{tr}\left[Z\frac{\partial g}{\partial x_i}(x)\right] + \sum_{j=1}^{r}\frac{\partial \lambda_j(g(x))}{\partial x_i}\lambda_j(Z) = 0 \quad (1 \le i \le n),$$

$$\lambda_i(g(x)) \le 0 \quad (1 \le i \le r),$$

$$\lambda_i(Z) \ge 0 \quad (1 \le i \le r),$$

$$\lambda_i(Z)\lambda_i(g(x)) = 0 \quad (1 \le i \le r),$$

which are preciselyly the Karush-Kuhn-Tucker conditions for problem (SP) with $\lambda_i(Z)$ as the KKT multiplier associated with the i-th constraint in (17). □

4. Second formulation of the optimality conditions

We present next the second formulation of first order optimality conditions for problem (CP), in which case the multipliers are mappings from $\mathbb{R}^m$ to S^m, nonlinear in general. The result on necessity (under a Slater's conditions) and sufficiency of these conditions in the psd-convex case, presented in the following theorem, will be used later on to establish necessity of (4)–(7), under an additional and much stronger constraint qualification.

Theorem 4.1. *Assume that*

i) Both f and the g_i's are psd-convex and differentiable.

ii) (Slater's condition) There exists $\hat{x} \in \mathbb{R}^n$ such that $g_i(\hat{x}) \prec 0$, $(1 \le i \le p)$.

Then $x^* \in \mathbb{R}^n$ solves (CP) if and only if there exist mappings $Z^i : \mathbb{R}^m \to S^{r_i}$ $(1 \le i \le p)$ such that, for all $u \in \mathbb{R}^m$,

$$u^t df(x^*)u + \sum_{i=1}^{p} tr[Z^i(u)dg_i(x^*)] = 0 \qquad \textit{Lagrangian condition}, \tag{26}$$

$$g_i(x^*) \preceq 0 \quad (1 \leq i \leq p) \qquad \textit{primal feasibility,} \tag{27}$$

$$Z^i(u) \succeq 0 \quad (1 \leq i \leq p) \qquad \textit{dual feasibility,} \tag{28}$$

$$tr[Z^i(u)g_i(x^*)] = 0 \quad (1 \leq i \leq p) \qquad \textit{complementarity.} \tag{29}$$

Proof. $(\Rightarrow)$ Observe that $u^t df(x^*)u$ is a linear functional defined on $\mathbb{R}^n$ which takes at the point $y \in \mathbb{R}^n$ the real value $u^t[df(x^*)y]u$. Alternatively, through the canonical identification of $\mathbb{R}^n$ with its dual, we can look at $u^t df(x^*)u$ as the vector in $\mathbb{R}^n$ whose j-th component is given by $\sum_{k=1}^m \sum_{\ell=1}^m u_k u_\ell \frac{\partial f(x^*)_{k\ell}}{\partial x_j}$. By the same token, $\mathrm{tr}[Z^i(u)dg_i(x^*)]$ can be seen as the linear functional defined in $\mathbb{R}^n$ which sends $y \in \mathbb{R}^n$ to $\mathrm{tr}[Z^i(u)dg_i(x^*)y] = \sum_{k=1}^{r_i}\sum_{\ell=1}^{r_i} Z^i_{k\ell}(u)\langle \nabla g_i(x^*)_{k\ell}, y\rangle$, or as the vector in $\mathbb{R}^n$ whose j-th component is $\sum_{k=1}^{r_i}\sum_{\ell=1}^{r_i} Z^i_{k\ell}(u)\frac{\partial g_i(x^*)_{k\ell}}{\partial x_j} = \mathrm{tr}[Z^i(u)\frac{\partial g_i}{\partial x_j}(x^*)]$.

Given matrices $A^{ij} \in S^{r_i}$ $(1 \leq i \leq p, 1 \leq j \leq n)$ define $h : S^{r_1} \times \cdots \times S^{r_p} \to \mathbb{R}^n$ and $C \subset S^{r_1} \times \cdots \times S^{r_p}$ as

$$h(Z^1, \ldots, Z^p)_j = -\sum_{i=1}^p \mathrm{tr}[Z^i A^{ij}], \tag{30}$$

$$C = \{(Z^1, \ldots, Z^p) \in S^{r_1}_+ \times \cdots \times S^{r_p}_+ : \mathrm{tr}[Z^i g_i(x^*)] = 0 \ (1 \leq i \leq p)\}. \tag{31}$$

Observe that C is a closed and convex cone by linearity of the trace, and h is a linear trasformation between finite dimensional spaces, so that $h(C)$ is also a closed and convex cone. Consider now h as in (30), with $A^{ij} = \frac{\partial g_i}{\partial x_j}(x^*)$, i.e.,

$$A^{ij}_{k\ell} = \frac{\partial g_i(x^*)_{k\ell}}{\partial x_j}.$$

Assume that $u^t df(x^*)u \notin h(C)$ for some $u \in \mathbb{R}^m$. Then, by the Convex Separation Theorem, there exists $w \in \mathbb{R}^n$ such that

$$u^t[df(x^*)w]u = \langle u^t df(x^*)u, w\rangle < 0, \tag{32}$$

$$-\sum_{i=1}^p \mathrm{tr}[Z^i dg_i(x^*)w] = \langle h(Z^1, \ldots, Z^p), w\rangle \geq 0, \tag{33}$$

for all $(Z^1, \ldots, Z^p) \in C$ (the fact that $h(C)$ is a cone ensures that the scalar σ of the separating hyperplane $\langle w, x\rangle = \sigma$ vanishes). Let

$$\theta = -u^t[df(x^*)w]u. \tag{34}$$

By (32), $\theta > 0$. Take $\beta > 0$ such that

$$\beta|u^t[df(x^*)(\hat{x} - x^*)]u| \leq \frac{\theta}{2}, \tag{35}$$

where $\hat{x}$ is the Slater's point whose existence is assumed in (ii), and define

$$\hat{w} = w + \beta(\hat{x} - x^*). \tag{36}$$

We claim that there exists $\bar{\alpha}$ such that $g_i(x^* + \alpha\hat{w}) \preceq 0$ for all $\alpha \in [0, \bar{\alpha}]$ and all $i \in \{1, \ldots, p\}$. Otherwise there exist $\{\alpha_k\} \subset \mathbb{R}_{++}$, $\{i(k)\} \subset \{1, \ldots, p\}$ and $v^k \in \mathbb{R}^{r_{i(k)}}$ and such that $\lim_{k\to\infty} \alpha_k = 0$, $||v^k|| = 1$ for all k and

$$(v^k)^t g_{i(k)}(x^* + \alpha_k \hat{w}) v^k > 0 \tag{37}$$

for all k. Clearly some index, say ℓ, repeats infinitely often in the sequence $\{i(k)\}$. Since the sequence $\{v^k\}$ is bounded, we may assume, without loss of generality, that $i(k) = \ell$ for all k and that that $\lim_{k\to\infty} v^k = \bar{v} \in \mathbb{R}^{r_\ell}$. Taking limits as k goes to ∞ in (37), we get $\bar{v}^t g_\ell(x^*)\bar{v} \geq 0$. Since $\bar{v}^t g_\ell(x^*)\bar{v} \leq 0$ because $x^* \in F$, we conclude that

$$\bar{v}^t g_\ell(x^*)\bar{v} = 0. \tag{38}$$

Let $\gamma = \bar{v}^t g_\ell(\hat{x})\bar{v}$. Since $||\bar{v}|| = 1$, $\gamma < 0$ by assumption (ii). Then, using (3) for the psd-convex mapping g_ℓ, we have

$$0 > \gamma = \bar{v}^t g_\ell(\hat{x})\bar{v} \geq \bar{v}^t g_\ell(x^*)\bar{v} + \bar{v}^t[dg_\ell(x^*)(\hat{x} - x^*)]\bar{v} = \bar{v}^t[dg_\ell(x^*)(\hat{x} - x^*)]\bar{v}, \tag{39}$$

using (38) in the last equality. Take now $Z^i = 0 \in \mathbb{R}^{r_i \times r_i}$ for $i \neq \ell$ and $Z^\ell = \bar{v}\bar{v}^t$. Clearly, $Z^i \in S_+^{r_i}$ for $1 \leq i \leq p$. By (38), $(Z^1, \ldots, Z^p) \in C$, as given by (31) and it follows from (33) that

$$0 \geq \sum_{i=1}^{p} \operatorname{tr}[Z^i dg_i(x^*)w] = \bar{v}^t[dg_\ell(x^*)w]\bar{v}. \tag{40}$$

Multiplying (39) by β and adding (40), we get in view of (36)

$$0 > \beta\gamma \geq \bar{v}^t[dg_\ell(x^*)(w + \beta(\hat{x} - x^*))]\bar{v} = \bar{v}^t[dg_\ell(x^*)\hat{w}]\bar{v}. \tag{41}$$

By (41) and continuity of dg_ℓ (see our comments at the beginning of this section), there exists $\varepsilon > 0$ such that, for all $v \in B(\bar{v}, \varepsilon)$ and all $\alpha \in [0, \varepsilon]$,

$$v^t[dg_\ell(x^* + \alpha\hat{w})\hat{w}]v \leq \frac{\beta\gamma}{2} < 0. \tag{42}$$

Note that

$$v^t g_\ell(x^* + \alpha\hat{w})v = v^t g_\ell(x^*)v + \alpha \int_0^1 v^t[dg_\ell(x^* + t\alpha\hat{w})\hat{w}]v dt. \tag{43}$$

Since $v^t g_\ell(x^*)v \leq 0$ for all $v \in \mathbb{R}^{r_\ell}$ because $x^* \in F$, it follows from (42) and (43) that

$$v^t g_\ell(x^* + \alpha\hat{w})v \leq \frac{\alpha\beta\gamma}{2} < 0 \tag{44}$$

for all $v \in B(\bar{v}, \varepsilon)$ and all $\alpha \in (0, \varepsilon]$. But, for large enough k, $v^k \in B(\bar{v}, \varepsilon)$ and $\alpha_k \in (0, \varepsilon]$, so that (44) contradicts (37). The claim holds, and thus there exists $\bar{\alpha} > 0$ such that $x^* + \alpha\hat{w} \in F$ for all $\alpha \in [0, \bar{\alpha}]$.

Now, the directional derivative of $u^t f(\cdot)u$ at x^* in the direction $\hat{w}$ is

$$u^t[df(x^*)\hat{w}]u = u^t[df(x^*)w]u + \beta u^t[df(x^*)(\hat{x} - x^*)]u =$$

$$-\theta + \beta u^t[df(x^*)(\hat{x} - x^*)]u \leq -\theta + \frac{\theta}{2} = -\frac{\theta}{2} < 0, \tag{45}$$

using (36) in the first equality, (34) in the second one and (35) in the inequality. It follows from (45) that there exists $\hat{\alpha} \in [0, \bar{\alpha}]$ such that

$$0 > u^t f(x^* + \hat{\alpha}\hat{w})u - u^t f(x^*)u = u^t[f(x^* + \hat{\alpha}\hat{w}) - f(x^*)]u. \tag{46}$$

On the other hand, since x^* solves (CP) and $x^* + \hat{\alpha}\hat{w} \in F$ because $\hat{\alpha} \in [0, \bar{\alpha}]$, we get $f(x^* + \hat{\alpha}\hat{w}) - f(x^*) \succeq 0$, in contradiction with (46). This contradiction arises from the assumption that $u^t df(x^*)u \notin h(C)$ for some $u \in \mathbb{R}^m$. It follows that $u^t df(x^*)u \in h(C)$ for all $u \in \mathbb{R}^m$, i.e., there exists $(Z^1(u), \ldots, Z^p(u)) \in C$ such that $u^t df(x^*)u + \sum_{i=1}^p \mathrm{tr}[Z^i(u) dg_i(x^*)] = 0$, which gives (26). Condition (27) holds because x^*, being a solution of (CP), belongs to F. Condition (28) and (29) follow from (31).

$(\Leftarrow)$ Assume that (26)–(29) hold. By (3) and psd-convexity of f, we have, for all $x \in F$,

$$f(x) \succeq f(x^*) + df(x^*)(x - x^*). \tag{47}$$

Take any $u \in \mathbb{R}^m$. By (47) and (26)

$$u^t f(x)u \geq u^t f(x^*)u + u^t[df(x^*)(x - x^*)]u = u^t f(x^*)u - \sum_{i=1}^p \mathrm{tr}[Z^i(u)(dg_i(x^*)(x - x^*))] =$$

$$u^t f(x^*)u + \sum_{i=1}^p \mathrm{tr}[Z^i(u)(g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*) + g_i(x^*) - g_i(x))] = u^t f(x^*)u +$$

$$\sum_{i=1}^p \mathrm{tr}[Z^i(u)(g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*))] + \sum_{i=1}^p \mathrm{tr}[Z^i(u) g_i(x^*)] + \sum_{i=1}^p \mathrm{tr}[Z^i(u)(-g_i(x))]. \tag{48}$$

Since all the g_i's are psd-convex, $g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*) \in S_+^{r_i}$, by (3). Since $x \in F$, $-g_i(x) \in S_+^{r_i}$. By (28) and Proposition 2.1(i), both the second and the fourth term in the right hand side of (48) are nonnegative. The third term vanishes by (29). It follows that $u^t f(x)u \geq u^t f(x^*)u$ for all $u \in \mathbb{R}^m$ and all $x \in F$, so that $f(x) \succeq f(x^*)$ for all $x \in F$. Since x^* belongs to F by (27), we conclude that x^* solves (CP). $\square$

Note that, as could be expected, psd-convexity of f is not needed for establishing necessity of (26)–(29). If we want a similar result without psd-convexity of the g_i's, we should replace Slater's condition by a stronger constraint qualification. Natural candidates are $h(C) = G(x^*, F)$ (where $G(x^*, F)$ is the cone of feasible directions at x^*, i.e. the intersection of all cones containing $F - \{x^*\}$), or perhaps the weaker condition $h(C)^* = G(x^*, F)^*$, i.e., the equality of the positive duals of both cones (see, e.g., [2]). Nevertheless, we found some technical obstacles when trying to establish necessity of (26)–(29) under these assumptions, so it's possible that the appropriate constraint qualification for this problem should in fact be stronger than those just mentioned. Of course, psd-convexity of f and of all the g_i's is essential for sufficiency of (26)–(29).

5. Duality scheme for the second formulation

Now we develop a duality scheme for problem (CP). For each $u \in \mathbb{R}^m$, consider $\phi_u : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{L}_u : \mathbb{R}^n \times S^{r_1} \times \cdots \times S^{r_p} \to \mathbb{R}$ and $\psi_u : S^{r_1} \times \cdots \times S^{r_p} \to \mathbb{R} \cup \{-\infty\}$ as

$$\phi_u(x) = u^t f(x)u, \tag{49}$$

$$\mathcal{L}_u(x,Z) = \mathcal{L}_u(x,Z^1,\ldots,Z^p) = \phi_u(x) + \sum_{i=1}^{p} \mathrm{tr}[Z^i g_i(x)], \tag{50}$$

$$\psi_u(Z) = \inf_{x\in\mathbb{R}^n} \mathcal{L}_u(x,Z). \tag{51}$$

Then for each $u \in \mathbb{R}^m$ we have a dual problem (D_u) defined as

$$\max \psi_u(Z)$$

$$\text{s.t.} \quad Z \succeq 0,$$

where $Z \succeq 0$ means $Z^1 \succeq 0, \ldots, Z^p \succeq 0$. We present next our duality theorem.

Theorem 5.1. *Consider problem (CP) with f and all the g_i's psd-convex.*

i) (weak duality) For all $u \in \mathbb{R}^m$, all $Z \in S_+^{r_1} \times \cdots \times S_+^{r_p}$ and all $x \in F$, it holds that $\psi_u(Z) \le \phi_u(x)$.

ii) If there exist $Z^*(u) \in S_+^{r_1} \times \cdots \times S_+^{r_p}$ and $x^* \in F$ such that $\psi_u(Z^*(u)) = \phi_u(x^*)$ for all $u \in \mathbb{R}^m$ then x^* solves (CP) and $Z^*(u)$ solves (D_u).

iii) (strong duality) Assume that f and all the g_i's are continuously differentiable and that Slater's condition (assumption (b) in Theorem 4.1) holds. Then, if x^* solves (CP) there exists $Z^*(u) \in S^{r_1} \times \cdots \times S^{r_p}$, such that, for all $u \in \mathbb{R}^m$, $\psi_u(Z^*(u)) = \phi_u(x^*)$ and $Z^*(u)$ solves (D_u).

Proof. i) By (50), (51),

$$\psi_u(Z) = \inf_{y\in\mathbb{R}^n} \mathcal{L}_u(y,Z) \le \mathcal{L}_u(x,Z) = \phi_u(x) - \sum_{i=1}^{p} \mathrm{tr}[Z^i(-g_i(x))] \le \phi_u(x),$$

where the second inequality follows from Proposition 2.1(i) and the facts that $x \in F$, $Z \succeq 0$.

ii) Immediate from (i), noting that, in view of (49), x^* solves (CP) if and only if it is a solution of min $\phi_u(x)$ subject to $x \in F$ for all $u \in \mathbb{R}^m$.

iii) Under the assumptions of this item, Theorem 4.1 holds, and therefore there exist $Z^i(u) \in S^{r_i}$ $(1 \le i \le p)$ such that (26)–(29) hold. Take $Z^*(u) = (Z^1(u), \ldots, Z^p(u))$. By (29), $\mathrm{tr}[Z^i(u) g_i(x^*)] = 0$, so that

$$\mathcal{L}_u(x^*, Z^*(u)) = \phi_u(x^*). \tag{52}$$

By (50) and (26),

$$\nabla_x \mathcal{L}_u(x^*, Z^*(u)) = u^t df(x^*) u + \sum_{i=1}^{p} \mathrm{tr}[Z^i(u) dg_i(x^*)] = 0. \tag{53}$$

Note that (50) can be rewritten as $\mathcal{L}_u(x,Z) = \mathrm{tr}[Uf(x) + \sum_{i=1}^{p} Z_i g_i(x)]$ with $U = uu^t \in S_+^m$. Observe that, for $V \in S_+^m$, the linear transformation $\mathbf{V}$: $S^m \to S^1$ given by $\mathbf{V}(X) = \mathrm{tr}(VX)$ is nondecreasing. Since $Z^1(u) \succeq 0, \ldots Z^p(u) \succeq 0$ by (28), we get from Proposition 2.1(iii) that $\mathcal{L}_u(\cdot, Z^*(u))$ is convex, and then it follows from (53) that x^* minimizes $\mathcal{L}_u(\cdot, Z^*(u))$, i.e.,

$$\psi_u(Z^*(u)) = \mathcal{L}_u(x^*, Z^*(u)). \tag{54}$$

By (52) and (54), $\psi_u(Z^*(u)) = \phi_u(x^*)$, and thus $Z^*(u)$ solves (D_u) by (ii). $\square$

6. Analysis of the first formulation

Before establishing sufficiency and necessity of (4)–(7), we present an example for which (26)–(29) hold, but (4)–(7) do not hold, showing that an additional constraint qualification is indeed needed.

Example 6.1. Take $n = 3$, $m = 2$, $p = 4$ and $r_i = 1$ $(1 \le i \le 4)$. Define $f : \mathbb{R}^3 \to S^2$ as

$$f(x) = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \tag{55}$$

and $g_i : \mathbb{R}^3 \to \mathbb{R}$ $(1 \le i \le 4)$ as

$$g_1(x) = -x_1 - x_2 + 1, \tag{56}$$

$$g_2(x) = -x_2 - x_3 + 1, \tag{57}$$

$$g_3(x) = -x_1 + x_2 + 1, \tag{58}$$

$$g_4(x) = x_2 - x_3 + 1. \tag{59}$$

We claim that $x^* = (1, 0, 1)^t$ is a solution of problem (CP) with these data. Note that

$$g_i(x^*) = 0 \qquad (1 \le i \le 4), \tag{60}$$

so that x^* is feasible, and that the constraints associated with (56) and (58) are equivalent to $|x_2| \le x_1 - 1$, while those associated with (57) and (59) are equivalent to $|x_2| \le x_3 - 1$. It follows that $x_1 - 1 \ge 0$, $x_3 - 1 \ge 0$, $(x_1 - 1)(x_3 - 1) \ge x_2^2$, and therefore, for all feasible $x \in \mathbb{R}^3$, it holds that

$$0 \preceq \begin{pmatrix} x_1 - 1 & x_2 \\ x_2 & x_3 - 1 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} - I = f(x) - f(x^*),$$

establishing optimality of x^*. Observe also that $\hat{x} = (2, 0, 2)^t$ satisfies the Slater condition $g_i(\hat{x}) \prec 0$ $(1 \le i \le 4)$. We check next that (26)–(29) hold at x^*. Conditions (27) and (29) hold trivially because of (60), independently of the choice of $Z^i(u)$. Before presenting the appropriate $Z^i(u)$'s, note that, for all $x \in \mathbb{R}^3$, $u \in \mathbb{R}^2$,

$$u^t df(x^*)u = (u_1^2, 2u_1u_2, u_2^2), \tag{61}$$

and that for all $x \in \mathbb{R}^3$,

$$\nabla g_1(x) = (-1, -1, 0)^t, \tag{62}$$

$$\nabla g_2(x) = (0, -1, -1)^t, \tag{63}$$

$$\nabla g_3(x) = (-1, 1, 0)^t, \tag{64}$$

$$\nabla g_4(x) = (0, 1, -1)^t. \tag{65}$$

It follows from (61)–(65) that in this case (26), written component-wise, becomes

$$u_1^2 = Z^1(u) + Z^3(u), \tag{66}$$

$$2u_1u_2 = Z^1(u) + Z^2(u) - Z^3(u) - Z^4(u), \tag{67}$$

$$u_2^2 = Z^2(u) + Z^4(u), \tag{68}$$

with $Z^i(u) \in \mathbb{R}$ $(1 \le i \le 4)$, since all the r_i's are equal to 1. If we take now

$$\begin{aligned} &\big(Z^1(u), Z^2(u), Z^3(u), Z^4(u)\big) \\ &= \begin{cases} \big(\frac{1}{2}(u_1+u_2)^2 - u_2^2, u_2^2, \frac{1}{2}(u_1-u_2)^2, 0\big) & \text{if } \sqrt{2}|u_2| \le |u_1+u_2| \\ \big(0, \frac{1}{2}(u_1+u_2)^2, u_1^2, u_2^2 - \frac{1}{2}(u_1+u_2)^2\big) & \text{if } \sqrt{2}|u_2| > |u_1+u_2|, \end{cases} \end{aligned} \tag{69}$$

it is easy to check that x^* and these $Z^i(u)$'s satisfy (66)–(68) and (28).

On the other hand, we show next that there exist no $\mathbf{Z}^1, \dots, \mathbf{Z}^4$ such that (4)–(7) hold. In particular, the system given by (4) and (6) has no solution. Indeed, note first that with $r_i = 1$, we can identify $L(S^{r_i}, S^m)$ with S^m, associating $\mathbf{Z} \in L(S^1, S^m)$ with $Z = \mathbf{Z}(1) \in S^m$. In this case, $\mathbf{Z} \succeq 0$ if and only if $Z \succeq 0$, i.e. Z is positive semidefinite in $\mathbb{R}^{m\times m}$. Writing (4) component-wise (i.e. taking the partial derivatives with respect to x_1, x_2 and x_3), we get

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = Z^1 + Z^3, \tag{70}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = Z^1 + Z^2 - Z^3 - Z^4, \tag{71}$$

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = Z^2 + Z^4. \tag{72}$$

Conditions (70)–(72) can be rewritten, after partial elimination, as

$$Z^1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} + Z^4, \tag{73}$$

$$Z^2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} - Z^4, \tag{74}$$

$$Z^3 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} - Z^4. \tag{75}$$

Condition (74), together with $Z^2 \succeq 0$, $Z^4 \succeq 0$, implies that $Z^4 = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}$ with $\lambda \in [0,1]$. Replacing this expression of Z^4 in (73), $Z^1 \succeq 0$ implies that $\lambda = 1$, i.e. $Z^4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Replacing this value of Z^4 in (75) we get $Z^3 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$, in contradiction with $Z^3 \succeq 0$. Thus (4) and (6) make an infeasible system.

We introduce next the constraint qualification required for necessity of (4)–(7). Take h as in (30), i.e. $h : S^{r_1} \times \cdots \times S^{r_p} \to \mathbb{R}^n$ defined as

$$h(Z^1, \dots, Z^p) = -\sum_{i=1}^{p} \operatorname{tr}[Z^i dg_i(x^*)] = -\sum_{i=1}^{p}\sum_{k=1}^{r_i}\sum_{\ell=1}^{r_i} Z^i_{k\ell} \nabla g_i(x^*)_{k\ell}, \tag{76}$$

and define the subspace $V \subset S^{r_1} \times \cdots \times S^{r_p}$ as

$$V = \{(Z^1, \ldots, Z^p) \in S^{r_1} \times \cdots \times S^{r_p} : \mathrm{tr}[Z^i g_i(x^*)] = 0 \ (1 \leq i \leq p)\}. \tag{77}$$

Our constraint qualification is

CQ) $\mathrm{Ker}(h) \cap V = \{0\}$.

We establish next that (CQ) holds if the gradients of all the $g_i(.)_{k\ell}$'s with $\ell \geq k$ at $x = x^*$ are linearly independent.

Proposition 6.2. *If the set $\{\nabla g_i(x^*)_{k\ell} : 1 \leq i \leq p, 1 \leq k \leq r_i, k \leq \ell \leq r_i\}$ is linearly independent then Ker$(h) = \{0\}$, and thus (CQ) holds.*

Proof. Take $(Z^1, \ldots, Z^p) \in \mathrm{Ker}(h)$. By (76) and the fact that both the Z^i's and the $g_i(x)$'s are symmetric, we have

$$0 = h(Z^1, \ldots, Z^p) = 2 \sum_{i=1}^{p} \sum_{k=1}^{r_i} \sum_{\ell=k}^{r_i} Z^i_{k\ell} \nabla g_i(x^*)_{k\ell}. \tag{78}$$

In view of the linear independence of the $\nabla g_i(x^*)_{k\ell}$'s, we conclude from (78) that $Z^i_{k\ell} = 0$ for all $i \in \{1, \ldots, p\}$, all $k \in \{1, \ldots, r_i\}$ and all $\ell \in \{k, \ldots, r_i\}$. It follows that all the Z^i's vanish, i.e. $\mathrm{Ker}(h) = \{0\}$, implying that $\mathrm{Ker}(h) \cap V = \{0\}$. □

Looking at (77) and Proposition 6.2, it is clear that (CQ) is the natural extension to the matrix-valued case of the requirement of linear independence of the gradients of the active constraints in the real-valued case. Note that the assumption of Proposition 6.2, much stronger that (CQ), requires $n \geq \frac{1}{2} \sum_{i=1}^{p} r_i(r_i + 1)$.

In order to prove necessity of (4)–(7) under Slater's condition and (CQ), we need the following lemma.

Lemma 6.3. *Under the hypothesis of Theorem 4.1, if x^* solves (CP) and the matrices $Z^i(u)$ satisfying (26)–(29) can be chosen so that they are component-wise quadratic functions of u, then (4)–(7) hold.*

Proof. The assumption of this lemma means that there exist matrices $M^{ik\ell} \in \mathbb{R}^{m \times m}$ $(1 \leq i \leq p, 1 \leq k \leq r_i, 1 \leq \ell \leq r_i)$ such that

$$Z^i(u)_{k\ell} = u^t M^{ik\ell} u. \tag{79}$$

Define $\mathbf{Z}^i \in L(S^{r_i}, S^m)$ $(1 \leq i \leq p)$ as

$$[\mathbf{Z}^i(A)]_{qs} = \sum_{k=1}^{r_i} \sum_{\ell=1}^{r_i} M^{ik\ell}_{qs} A_{k\ell} \ \ (1 \leq q, s \leq m). \tag{80}$$

We check first that the $\mathbf{Z}^i$'s defined by (80) satisfy (6). We must verify that for all $A \succeq 0$, it holds that $\mathbf{Z}^i(A) \succeq 0$, i.e. that $y^t[\mathbf{Z}^i(A)]y \geq 0$ for all $y \in \mathbb{R}^m$. Note that

$$y^t[\mathbf{Z}^i(A)]y = \sum_{q=1}^{m} \sum_{s=1}^{m} y_q y_s [\mathbf{Z}^i(A)]_{qs} = \sum_{q=1}^{m} \sum_{s=1}^{m} y_q y_s \sum_{k=1}^{r_i} \sum_{\ell=1}^{r_i} M^{ik\ell}_{qs} A_{k\ell} =$$

$$\sum_{k=1}^{r_i}\sum_{\ell=1}^{r_i} A_{k\ell}\sum_{q=1}^{m}\sum_{s=1}^{m} M_{qs}^{ik\ell} y_q y_s = \sum_{k=1}^{r_i}\sum_{\ell=1}^{r_i} A_{k\ell} Z^i(y)_{k\ell} = \mathrm{tr}[Z^i(y)A] \geq 0,$$

using (80) in the second equality, (79) in the fourth one, and (28) and Proposition 2.1(i) in the inequality. Thus (6) holds for $\mathbf{Z}^i$ as defined by (80).

Now we look at the Lagrangian condition (4). Replacing (79) and (80) in (26), we get easily

$$u^t\left[df(x^*) + \sum_{i=1}^{p} \mathbf{Z}^i dg_i(x^*)\right]u = u^t df(x^*)u + u^t\left[\sum_{i=1}^{p} \mathbf{Z}^i dg_i(x^*)\right]u = 0 \tag{81}$$

for all $u \in \mathbb{R}^m$. Let $\{e^1, \ldots, e^m\}$ be the canonical basis of $\mathbb{R}^m$. Substracting (81) with $u = e^q - e^s$ from the same equation with $u = e^q + e^s$, we get

$$4\left[df(x^*) + \sum_{i=1}^{p} \mathbf{Z}^i dg_i(x^*)\right]_{qs} = 0,$$

which, after dividing by 4, gives the component-wise expression of (4). Replacing now (79) and (80) in (29), and substracting the resulting equation with $u = e^q - e^s$ from the same equation with $u = e^q + e^s$, we get (7). Condition (5) holds because x^*, being a solution of (CP), belongs to F. □

Now we present our proof of necessity and sufficience of (4)–(7) under Slater's condition and (CQ) above.

Theorem 6.4. *Assume that*

i) Both f and the g_i's are psd-convex and differentiable.

ii) (Slater's condition) There exists $\hat{x} \in \mathbb{R}^n$ such that $g_i(\hat{x}) \prec 0$, $(1 \leq i \leq p)$.

iii) (CQ) Ker$(h) \cap V = \{0\}$, with h and V as given by (76) and (77).

Then $x^* \in \mathbb{R}^n$ solves (CP) if and only if there exist $\mathbf{Z}^i \in L(S^{r_i}, S^m)$ $(1 \leq i \leq p)$ such that (4)–(7) hold for $x = x^*$.

Proof. $(\Rightarrow)$ Since we are within the hypotheses of Theorem 4.1, we know that there exist $Z^i(u)$ $(1 \leq i \leq p)$ such that $x^*, Z^1(u), \ldots, Z^p(u)$ satisfy (26)–(29). In view of Lemma 6.3, it suffices to show that these $Z^i(u)$'s can be chosen so that they are quadratic functions of u component-wise. We proceed to do so.

Consider the restriction $h_{/V}$ of h to V. Under (CQ), $h_{/V}$ is one-to-one, and thus it has a left inverse $h_{/V}^{-1} : \mathrm{Im}(h) \to S^{r_1} \times \cdots \times S^{r_p}$. It follows that the system of linear equations in $Z^i(u)$ given by (26) and (29) has a unique solution given by

$$(\bar{Z}^1(u), \ldots, \bar{Z}^p(u)) = h_{/V}^{-1}(u^t df(x^*)u). \tag{82}$$

By Theorem 4.1, this unique solution satisfies (28) and (29). By linearity of $h_{/V}^{-1}$ there exist $b_{k\ell j}^i \in \mathbb{R}$ $(1 \leq i \leq p,\ 1 \leq k, \ell \leq r_i,\ 1 \leq j \leq n)$ such that, for all $x \in \mathbb{R}^n$, $[h_{/V}^{-1}(x)]_{k\ell}^i = \sum_{j=1}^n b_{k\ell j}^i x_j$. In view of (82), we have

$$\bar{Z}^i(u)_{k\ell} = [h_{/V}^{-1}(u^t df(x^*)u)]_{k\ell}^i = \sum_{j=1}^{n} b_{k\ell j}^i [u^t df(x^*)u]_j =$$

$$\sum_{j=1}^{n} b^i_{k\ell j} \sum_{q=1}^{m} \sum_{s=1}^{m} u_q u_s \frac{\partial f(x^*)_{qs}}{\partial x_j} = \sum_{q=1}^{m} \sum_{s=1}^{m} u_q u_s \sum_{j=1}^{n} b^i_{k\ell j} \frac{\partial f(x^*)_{qs}}{\partial x_j}. \tag{83}$$

Consider now the matrices $M^{ik\ell} \in S^m$ with entries

$$M^{ik\ell}_{qs} = \sum_{j=1}^{n} b^i_{k\ell j} \frac{\partial f(x^*)_{qs}}{\partial x_j}. \tag{84}$$

It follows from (83) and (84) that $\bar{Z}^i(u)_{k\ell} = u^t M^{ik\ell} u$, i.e. each entry of each $\bar{Z}^i(u)$ is a quadratic function of u. The result follows from Lemma 6.3.

($\Leftarrow$) Assume that (4)–(7) hold. By (3) and psd-convexity of f, we have, for all $x \in F$,

$$f(x) \succeq f(x^*) + df(x^*)(x - x^*). \tag{85}$$

Take any $u \in \mathbb{R}^m$. By (85) and (4)

$$f(x) \succeq f(x^*) + df(x^*)(x - x^*) = f(x^*) - \sum_{i=1}^{p} \mathbf{Z}^i[dg_i(x^*)(x - x^*)] =$$

$$f(x^*) + \sum_{i=1}^{p} \mathbf{Z}^i[g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*) + g_i(x^*) - g_i(x)] =$$

$$f(x^*) + \sum_{i=1}^{p} \mathbf{Z}^i[g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*)] + \sum_{i=1}^{p} \mathbf{Z}^i[g_i(x^*)] + \sum_{i=1}^{p} \mathbf{Z}^i[-g_i(x)]. \tag{86}$$

Since all the g_i's are psd-convex, $g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*) \in S^{r_i}_+$ by (3). Since $x \in F$, we have that $-g_i(x) \in S^{r_i}_+$. By (6) and the definition of nonnegativity for elements of $L(S^{r_i}, S^m)$, both the second and the fourth term in the right hand side of (86) are nonnegative. The third term vanishes by (7). It follows that $f(x) \succeq f(x^*)$ for all $x \in F$. Since x^* belongs to F by (5), we conclude that x^* solves (CP). □

We remark that (CQ) implies that the dual solution $(\mathbf{Z}^1, \dots, \mathbf{Z}^p)$ of (4)–(7) is unique, as the following result shows.

Proposition 6.5. *Under the hypotheses of Theorem 4.1, if (CQ) holds and (CP) has optimal solutions, then there exist unique* $\mathbf{Z}^1, \dots, \mathbf{Z}^p \in L(S^{r_1}, S^m) \times \cdots \times L(S^{r_p}, S^m)$ *which satisfy* (4)–(7).

Proof. Fix $u \in \mathbb{R}^m$. By Theorem 4.1, if (CP) has optimal solutions, then there exist $x^* \in \mathbb{R}^n$, $Z^1(u), \dots, Z^p(u) \in S^{r_1} \times \cdots \times S^{r_p}$ which satisfy (26)–(29). Under (CQ), the system of linear equations in $Z^1(u), \dots, Z^p(u)$ given by (26) and (29) has a unique solution $\bar{Z}^1(u), \dots, \bar{Z}^p(u)$. By Theorem 6.4, there exist $(\mathbf{Z}^1, \dots, \mathbf{Z}^p) \in L(S^{r_1}, S^m) \times \cdots \times L(S^{r_p}, S^m)$ such that (4)–(7) are satisfied with $x = x^*$. Since $\mathbf{Z}^i$ is linear there exist $z^i_{qsk\ell} \in \mathbb{R}$ $(1 \le k, \ell \le r_i,\ 1 \le q, s \le m)$ such that

$$[\mathbf{Z}^i(A)]_{qs} = \sum_{k=1}^{r_i} \sum_{\ell=1}^{r_i} z^i_{qsk\ell} A_{k\ell}. \tag{87}$$

Pre and postmultiplying (4) and (7) by u, it is easy to check that the matrices $\hat{Z}^i(u)$ with entries

$$\hat{Z}^i(u)_{k\ell} = \sum_{q=1}^{m}\sum_{s=1}^{m} z^i_{qsk\ell} u_q u_s \tag{88}$$

also satisfy (26) and (29). We conclude that $\hat{Z}^i(u) = \bar{Z}^i(u)$ for all $i \in \{1, \ldots, p\}$, i.e., in view of (88),

$$\sum_{q=1}^{m}\sum_{s=1}^{m} z^i_{qsk\ell} u_q u_s = \bar{Z}^i(u)_{k\ell}. \tag{89}$$

Substracting (89) with $u = e^q - e^s$ from the same equation with $u = e^q + e^s$, where $\{e^1, \ldots, e^m\}$ is the canonical basis of $\mathbb{R}^m$, we obtain that

$$z^i_{qsk\ell} = \frac{1}{4}\big[\bar{Z}^i(e^q + e^s)_{k\ell} - \bar{Z}^i(e^q - e^s)_{k\ell}\big]. \tag{90}$$

Since the $\bar{Z}^i(u)$ are unique, we get from (90) that the $z^i_{qsk\ell}$'s are uniquely determined, and then it follows from (87) that the $\mathbf{Z}^i$'s are also uniquely determined, i.e. that the system of linear equations in $\mathbf{Z}^1, \ldots, \mathbf{Z}^p$ given by (4) and (7) has a unique solution. A fortiori, the full system (4)–(7) also has a unique solution. □

In connection with Theorem 6.4, we mention that the problem in Example 6.1 above does not satisfy (CQ). In fact, it holds that $\mathrm{Ker}(h) \cap V = \mathrm{Ker}(h) = \{(t, -t, -t, t) : t \in \mathbb{R}\} \neq \{0\}$. Regarding Lemma 6.3, note that the $Z_i(u)$'s given by (69) are piece-wise quadratic, but not quadratic in u.

We present next an example for which both Slater's condition and (CQ) hold, and thus the multipliers exist both for the first formulation, (4)–(7), and for the second one, (26)–(29).

Example 6.6. Take $n = 3$, $m = 2$, $p = 3$, $r_1 = r_2 = r_3 = 1$. We consider f as in (55) of Example 6.1, i.e. $f(x) = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix}$, and

$$g_1(x) = -x_1 - x_2 + 2, \quad g_2(x) = -x_1 + x_2, \quad g_3(x) = x_1 - x_3.$$

Slater's condition is satisfied, e.g., by $\hat{x} = (2, 1, 3)^t$, and (CQ) holds by virtue of Proposition 6.2, because the gradients of the g_i's, namely $(-1, -1, 0)^t$, $(-1, 1, 0)^t$, $(1, 0, -1)^t$, are linearly independent. The optimality conditions (4)–(7) hold with $x^* = (1, 1, 1)^t$, $Z^1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $Z^2 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ and $Z^3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, identifying as before $L(S^1, S^2)$ with S^2. Thus x^* solves (CP). Regarding the second formulation of the optimality conditions, (26)–(29) hold with the same x^* and $Z^1(u) = \frac{1}{2}(u_1 + u_2)^2$, $Z^2(u) = \frac{1}{2}(u_1 - u_2)^2$ and $Z^3(u) = u_2^2$, which are all nonnegative and quadratic in u.

7. Duality scheme for the first formulation

We present next the duality scheme corresponding to the first formulation of the optimality conditions. Define $\hat{\mathcal{L}} : \mathbb{R}^n \times L(S^{r_1}, S^m) \times \cdots \times L(S^{r_p}, S^m) \to S^m$ as

$$\hat{\mathcal{L}}\big(x, \mathbf{Z}^1, \ldots, \mathbf{Z}^p\big) = f(x) + \sum_{i=1}^{p} \mathbf{Z}^i\big(g_i(x)\big). \tag{91}$$

Observe that when f and all the g_i's are psd-convex, $\hat{\mathcal{L}}\big(\cdot, \mathbf{Z}^1, \dots, \mathbf{Z}^p\big)$ is also psd-convex for all $\mathbf{Z}^1 \succeq 0, \dots, \mathbf{Z}^p \succeq 0$, by Proposition 2.1(iii).

For those $\mathbf{Z}^1, \dots, \mathbf{Z}^p$ such that $\min_{x\in\mathbb{R}^n} \hat{\mathcal{L}}\big(x, \mathbf{Z}^1, \dots, \mathbf{Z}^p\big)$, in the partial order $\preceq$ of S^m, exists, we define

$$\Psi\big(\mathbf{Z}^1, \dots, \mathbf{Z}^p\big) = \min_{x\in\mathbb{R}^n} \hat{\mathcal{L}}\big(x, \mathbf{Z}^1, \dots, \mathbf{Z}^p\big). \tag{92}$$

The duality theorem for this formulation is

Theorem 7.1. *Consider problem (CP) with f and all the g_i's psd-convex.*

i) (weak duality) For all $\big(\mathbf{Z}^1, \dots, \mathbf{Z}^p\big) \in dom(\Psi)$ satisfying $\mathbf{Z}^1 \succeq 0, \dots, \mathbf{Z}^p \succeq 0$, and all $x \in F$, it holds that $\Psi\big(\mathbf{Z}^1, \dots, \mathbf{Z}^p\big) \preceq f(x)$.

ii) If there exist $\mathbf{Z}^* \in L(S^{r_1}, S^m) \times \dots \times L(S^{r_p}, S^m)$ and $x^* \in F$ such that $\Psi(\mathbf{Z}^*) = f(x^*)$ then x^* solves (CP) and $\mathbf{Z}^*$ solves min $\Psi(\mathbf{Z})$ subject to $\mathbf{Z} \succeq 0$.

iii) (strong duality) Assume that f and all the g_i's are continuously differentiable and both Slater's condition and (CQ) hold.

In these conditions, if x^* solves (CP) then there exists $\mathbf{Z}^* \in L(S^{r_1}, S^m) \times \dots \times L(S^{r_p}, S^m)$ such that $\Psi(\mathbf{Z}^*) = f(x^*)$ and $\mathbf{Z}^*$ solves max $\Psi(\mathbf{Z})$ subject to $\mathbf{Z} \succeq 0$.

Proof. i) By (91), (92),

$$\begin{aligned}\Psi\big(\mathbf{Z}^1, \dots, \mathbf{Z}^p\big) &= \min_{y\in\mathbb{R}^n} \hat{\mathcal{L}}_u\big(y, \mathbf{Z}^1, \dots, \mathbf{Z}^p\big) \preceq \hat{\mathcal{L}}_u\big(x, \mathbf{Z}^1, \dots, \mathbf{Z}^p\big) \\ &= f(x) - \sum_{i=1}^p \mathbf{Z}^i(-g_i(x)) \preceq f(x),\end{aligned}$$

where the second inequality follows from the facts that $x \in F$, so that $-g_i(x) \succeq 0$, and $\mathbf{Z}^1 \succeq 0, \dots, \mathbf{Z}^p \succeq 0$.

ii) Immediate from (i).

iii) Under the assumptions of this item Theorem 6.4 holds, and therefore there exist $\mathbf{Z}^i \in L(S^{r_i}, S^m)$ $(1 \le i \le p)$ such that (4)–(7) hold. Take $\mathbf{Z}^* = \big(\mathbf{Z}^1, \dots, \mathbf{Z}^p\big)$. By (7), $\mathbf{Z}^i\big(g_i(x^*)\big) = 0$, so that

$$\hat{\mathcal{L}}(x^*, \mathbf{Z}^*) = f(x^*). \tag{93}$$

By (91) and (4),

$$d_x \hat{\mathcal{L}}_u(x^*, \mathbf{Z}^*) = df(x^*) + \sum_{i=1}^p \mathbf{Z}^i dg_i(x^*) = 0. \tag{94}$$

Since $\hat{\mathcal{L}}_u(\cdot, \mathbf{Z}^*)$ is psd-convex, as we have already observed, because $\mathbf{Z}^i \succeq 0$ by (6), it follows from (2) and (94) that X^* minimizes $\hat{\mathcal{L}}_u(\cdot, \mathbf{Z}^*)$, i.e. $\mathbf{Z}^* \in \mathrm{dom}(\Psi)$ and

$$\Psi(\mathbf{Z}^*) = \hat{\mathcal{L}}_u(x^*, \mathbf{Z}^*). \tag{95}$$

By (93) and (95), $\Psi(\mathbf{Z}^*) = f(x^*)$, and thus $\mathbf{Z}^*$ solves max $\Psi(\mathbf{Z})$ subject to $\mathbf{Z} \succeq 0$ by (ii). □

Acknowledgements. We gratefully acknowledge several enlightening discussions with R.S. Burachik on the subject of this paper, and the very useful suggestions made by two anonymous referees.

References

[1] T. Amaroq, A. Taa: On Lagrange-Kuhn-Tucker multipliers for multiobjective optimization problems, Optimization 41(1) (1997) 159–172.

[2] M. Avriel: Nonlinear Programming, Analysis and Methods, Prentice Hall (1976).

[3] R: Cambini: Some new classes of generalized concave vector valued functions, Optimization 36(1) (1996) 11–24.

[4] R. Cambini: Second order optimality conditions in multiobjective programming, Optimization 44 (1998) 139–160.

[5] C. Escalante, M. Raydan: Dijkstra's algorithm for a constrained least-squares matrix problem, Numerical Linear Algebra with Applications 3 (1996) 459–471.

[6] F. Flores-Bazán: Ideal, weakly efficient solutions for convex vector optimization problems, Technical Report 99-26, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Concepción, Chile (1990).

[7] L. M. Graña Drummond, A. N. Iusem, B. F. Svaiter: On first order optimality conditions for vector optimization, preprint.

[8] L. M. Graña Drummond, Y. Peterzil: The central path in smooth convex semidefinite programs, Optimization, to appear.

[9] R. A. Horn, C. R. Johnson: Matrix Analysis, Cambridge University Press, Cambridge (1985).

[10] T. Illes, G. Kassay: Theorems of the alternative and optimality conditions for convexlike and general convexlike programming, Journal of Optimization Theory and Applications 101 (1999) 243–257.

[11] F. Jarre: Convex analysis on symmetric matrices, in: Handbook of Semidefinite Programming: Theory, Algorithms and Applications, H. Wolkowicz, R. Saigal, L. Vandenberghe (eds.), Kluwer Academic Publishers, Norwell (2000) 13–27.

[12] T. D. Luc: Theory of Vector Optimization, Lecture Notes in Economy and Mathematic Systems 319, Springer-Verlag, Berlin (1989).

[13] T. Maeda: Constraint qualifications in multiobjective optimization problems: differentiable case, Journal of Optimization Theory and Applications 80 (1994) 483–500.

[14] V. Preda, I. Chiţescu: On constraint qualifications in multiobjective optimization problems: semidifferentiable case, Journal of Optimization Theory and Applications 100 (1999) 417–434.

[15] R. T. Rockafellar: Convex Analysis, Princeton University Press (1970).

[16] A. Shapiro: First and second order analysis of nonlinear semidefinite programs, Mathematical Programming 77 (1997) 301–320.

[17] A. Shapiro: On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints, SIAM Journal on Optimization, to appear.

[18] C. Singh: Optimality conditions in multiobjective differentiable programming, Journal of Optimization Theory and Applications 53 (1987) 115–123.

[19] M. C. Travain: On Lagrange-Kuhn-Tucker multipliers for Pareto optimization problems, Numerical Functional Analysis and Optimization 15(5-6) (1994) 689–693.