Convexity of vector-valued mappings

L. M. Graña Drummond* A. N. Iusem†

May 21, 2001

Abstract

The aim of this paper is to study the convexity of vector-valued mappings, related to the partial order induced in a finite dimensional linear space by a closed convex cone. We review some of the main properties of such mappings and we give special emphasis to the study of first order optimality conditions for minimization problems in this partial order, where the data functions are differentiable, and convex in the sense just mentioned. For the unconstrained case, we derive a first order necessary and sufficient condition similar to the one which holds for real-valued optimization. We study constrained problems only for the case of symmetric matrix-valued data with the partial order given by the cone of positive semidefinite matrices, for which 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.

Key words: Cone constraints, vector optimization, first order optimality conditions, convex programming, duality.

*Department of Mathematics of the University of Haifa, Haifa, 31905, Israel (**mauricio@mathcs2.haifa.ac.il**). The work of this author was supported by a post-doctoral fellowship within the Department of Mathematics of the University of Haifa.

†Instituto de Matemática Pura e Aplicada (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil (**iusp@impa.br**). The work of this author was partially supported by CNPq grant no. 301280/86

1. Introduction

In this paper we discuss convexity of vector valued mappings, related to the partial order induced in a vector space by a closed and convex cone K. This partial order is given by $x \succeq y$ if and only if $x - y \in K$. Mappings between finite dimensional vector spaces which satisfy the usual inequality defining convexity, but with "$\succeq$" substituting for "$\geq$", are called K-convex.

After introducing several properties of K-convex mappings in Section 2, we consider in Section 3 first order optimality conditions for minimization problems in this partial order. For unconstrained problems with a differentiable and K-convex objective f, we prove that, if the positive dual cone of K has nonempty interior, then, as in the case of real-valued functions, x minimizes f if and only if $df(x) = 0$. In Section 4 we consider the particular case arising when the mappings take values in the space of symmetric matrices of order m, and the cone K is S^m_+, i.e. the set of positive semidefinite symmetric matrices or order m. When $K = S^m_+$, K-convex mappings are called psd-convex. A few additional properties specific of psd-convex mappings are established. In Section 5 we discuss first order optimality conditions for inequality constrained problems where both the objective and the constraint mappings are psd-convex, and both the minimization and the inequalities are understood in the sense of the partial order induced by S^m_+. 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.

To our knowledge, features of K-convex mappings were not systematically studied before. In [2] the author extends to vector valued-mappings defined on convex subsets of $\mathbb{R}^n$ the notions of concavity, quasiconcavity and pseudoconcavity by means of three different order relations generated by a closed cone with nonempty interior in $\mathbb{R}^m$. The first one, generated by the cone itself, gives rise to the same notion we are dealing with in our work. The second and the third are analogous but they are generated by the cone without the origin and by the interior of the cone, respectively. The author focuses on first order characterizations and relationships among these classes. The only point in common between this work and ours is the generalized gradient inequality (Theorem 6.1 in [2] and Proposition 10 in ours) which we prove in a different way.

The 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 [5], 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 [6] 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 [5] 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 [5] nor [6] consider the case of vector-valued objectives, i.e. situations where the minimization is performed with respect to the partial order induced by a cone, which are the main subject of our work.

A couple of elementary properties of psd-mappings are presented in [3], which also considers the problem of minimizing a real-valued convex function subject to a single psd-constraint. This paper is devoted to the issues of welldefinedness and convergence of the central path associated to the primal-dual pair of problems, which are unrelated to our current subject.

2. K-convex mappings

Let V a finite dimensional real vectorial space with an inner product $\langle \cdot, \cdot \rangle$, and let $K \subset V$ be a closed convex cone. We consider the partial order in V, given by $x \succeq y$ if and only if $x - y \in K$, and $y \preceq x$ if and only if $x \succeq y$.

The two most significant examples are the cases of $V = \mathbb{R}^m$, with K being the nonnegative orthant of $\mathbb{R}^m$, in which case "$\succeq$" is the natural partial order in $\mathbb{R}^m$, and $V = S^m$, the subspace of symmetric matrices in $\mathbb{R}^{m \times m}$, with $K = S^m_+$, the cone of symmetric positive semidefinite matrices in $\mathbb{R}^{m \times m}$. Sections 4 and 5 are devoted to this second particular case.

We say that the mapping $f : \mathbb{R}^n \to V$ is *K-convex*, if $f(\lambda x + (1-\lambda)y) \preceq \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0, 1]$. Clearly, this extends the classical notion of convexity for real-valued functions (take $V = \mathbb{R}$ and $K = \mathbb{R}_+$, the nonnegative real line).

First of all, let us consider the following example of a whole collection of K-convex mappings.

Example 1. Take arbitrary V and K as above and let

$$f(x) = z_0 + \sum_{i=1}^{n} x_i z_i + \sum_{i=n+1}^{r} f_i(x) z_i \ ,$$

where $z_i \succeq 0$ and $f_i : \mathbb{R}^n \to \mathbb{R}$ is a convex function for $i = n+1, \ldots, r$. If, for instance, $f_i = 0$, for all $i = n+1, \ldots, r$, then f is affine. Observe also that a whole family of quadratic forms can be obtained by choosing the proper f_i's .

We define the *positive dual of K* as $K^* = \{z \in V : \ \langle z, x \rangle \geq 0 \text{ for all } x \in K\}$. We have the following result.

Proposition 1. $(K^*)^* = K$.

Proof. See Theorem 14.1 of [4]. □

In order to study some general features of K-convex mappings, we will consider, for each $z \in K^*$, the scalar function $\psi_z : \mathbb{R}^n \to \mathbb{R}$, given by $\psi_z(x) = \langle z, f(x) \rangle$. We begin with a very useful characterization of K-convexity.

Proposition 2. *The mapping f is K-convex if and only if the numerical function $\psi_z(x) = \langle z, f(x) \rangle$ is convex for all $z \in K^*$.*

Proof. Take $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$. We must establish that $\Delta := \lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda) y) \succeq 0$, i.e., that $\Delta \in K$. In view of Proposition 1, $\Delta \in K$ if and only if $\Delta \in (K^*)^*$, which, in turn, is equivalent to the convexity of $\psi_z(x)$ for all $z \in K^*$. □

The following criterion can be seen as an extension of the well known fact that a real-valued function is convex if and only if its epigraph is a convex set. We will call the set $\{(x, y) \in \mathbb{R}^n \times V : f(x) \preceq y\}$ *K-epigraph* of f.

Proposition 3. *The mapping f is K-convex if and only if its K-epigraph is a convex subset of $\mathbb{R}^n \times V$.*

Proof. Assume that the K-epigraph is a convex set. Take $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$. Clearly, the pairs $(x, f(x))$ and $(y, f(y))$ are on the K-epigraph and, by the convexity assumption, the point $(\lambda x + (1-\lambda) y, \lambda f(x) + (1-\lambda) f(y))$ is also in the K-epigraph, which means that f is K-convex. The converse is immediate.

□

Now we study the K-versions of some very well known properties of real-valued convex functions. The first one is the K-Jensen's inequality.

Proposition 4. *The mapping f is K-convex if and only if $f(\sum_{i=1}^r \lambda_i x^i) \preceq \sum_{i=1}^r \lambda_i f(x^i)$, for all $x^1, \ldots, x^r \in \mathbb{R}^n$ and all $\lambda_1, \ldots, \lambda_r \geq 0$ such that $\sum_{i=1}^r \lambda_i = 1$, with $r \geq 2$.*

Proof. Since f is K-convex, in view of Proposition 3, its K-epigraph is a convex subset of $\mathbb{R}^n \times V$. The pairs $(x^1, f(x^1)), \ldots, (x^r, f(x^r))$ are in the K-epigraph of f, and thus the convex combination $\sum_{i=1}^r \lambda_i (x^i, f(x^i))$ too, so that the result follows.

□

Another elementary property of convex functions that also holds in its K -version is the following.

Proposition 5. *The class of K-convex mappings is closed under nonnegative linear combinations.*

Proof. The proof is immediate using the fact that K is a convex cone. □

Given $z \in V$ and $f : \mathbb{R}^n \to V$, we call the set $\{x \in \mathbb{R}^n : f(x) \preceq z\}$ *K-level set* (of level z) of f. We have that a necessary but not sufficient condition for the K-convexity of f is the convexity of this subset of $\mathbb{R}^n$.

Proposition 6. *If f is a K-convex mapping, then its K-level sets are convex.*

Proof. The result follows easily from the definition of K-convexity.

□

Since "$\preceq$" is not a total order, it may not always make sense to speak of the K-supremum of a V-subset. Nevertheless, sometimes we may be able to consider the K-supremum of a family of mappings $\{f_i : \mathbb{R}^n \to V : i \in I\}$ as being the application $f(x) = sup\{f_i(x) : i \in I\}$. We then have the following result.

Proposition 7. *Given a finite or infinite family $\{f_i : i \in I\}$ of K-convex mappings, then f, the pointwise K-supremum of the collection (whenever it exists), is also a K-convex mapping.*

Proof. In view of Proposition 3, it is enough to show that the K-epigraph of f is a convex subset of $\mathbb{R}^n \times V$. Clearly, the K-epigraph of f coincides with the intersection of the K-epigraphs of all the f_i's and the result follows from the K-convexity of these mappings.

□

Our next proposition also generalizes a property of scalar convex functions; it states that a mapping is K-convex iff is its restriction to closed segments is also K-convex.

Proposition 8. *The mapping f is K-convex if and only if for every $x, y \in \mathbb{R}^n$ its restriction to the segment $[x, y]$, $\phi : [0, 1] \to V$, given by $\phi(\lambda) = f(\lambda x + (1 - \lambda)y)$ is K-convex.*

Proof. By Proposition 2, the result holds if and only if it holds for $\psi_z(x) = \langle z, f(x)\rangle (z \in K^*)$. The result follows since this property characterizes numerical convex functions.

□

The following proposition furnishes a tool for constructing new K-convex mappings from existing ones.

Proposition 9. *Assume that f is a K-convex mapping. Then*
i) If $\Psi : V \to V$ is K-convex and nondecreasing in the K-order, then the composition $x \mapsto \Psi(f(x))$ is K-convex.
ii) If $A : \mathbb{R}^r \to \mathbb{R}^n$ is linear and $b \in \mathbb{R}^n$, then $x \mapsto f(A(x) + b)$ is a K-convex mapping from $\mathbb{R}^r$ to V.
iii) If $T : V \to W$ is linear, where W is a finite dimensional linear and $\hat{K} \subset W$ is a closed convex cone, then $x \mapsto T(f(x))$ is a $\hat{K}$-convex mapping from $\mathbb{R}^n$ to W.

Proof. Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. By the K-convexity of f, we have $f(\lambda x + (1 - \lambda)y) \preceq \lambda f(x)+(1-\lambda)f(y)$. Therefore, from the K-monotonicity of Ψ, it follows that $\Psi(f(\lambda x+(1-\lambda)y)) \preceq \Psi(\lambda f(x) + (1 - \lambda)f(y))$. Finally, using the K-convexity of Ψ, we get $\Psi(f(\lambda x + (1 - \lambda)y)) \preceq \lambda\Psi(f(x)) + (1 - \lambda)\Psi(f(y))$, and the proof of i) is complete.

The proofs of ii) and iii) follow immediately from the linearity of A and T, respectively.

□

We end the section by studying a couple of differential properties of K-convex mappings. From now on, and until the end of the section, we assume that f is differentiable. We recall that the differential of f at x is the mapping $df(x) : \mathbb{R}^n \to V$, defined by $df(x)y = \sum_{i=1}^n y_i \frac{\partial f}{\partial x_i}(x)$. Without loss of generality, we may suppose that the vectors in V are expressed in terms of an orthonormal basis, so that, if $\dim(V) = m$, we have $\langle x, y\rangle = \sum_{i=1}^m x_i y_i$.

We begin by what can be called the *generalized gradient inequality*, which, as we will see in our next proposition, charaterizes differentiable K-convex mappings.

Proposition 10. *The mapping f is K-convex if and only if $f(y) \succeq f(x) + df(x)(y - x)$ for all $x, y \in \mathbb{R}^n$.*

Proof. The result follows from Proposition 2, since the gradient inequality characterizes scalar convex functions.

□

Observe that, as in the scalar case, if f is differentiable just at $x \in \mathbb{R}^n$, then the generalized gradient inequality holds at x. Observe also that the above result suggests what can be an adequate definition of the K-subdifferential of f at a point $x \in \mathbb{R}^n$, i.e., the set of arrays of vectors $(z^1, \ldots, z^n)$, such that $z^i \in V$, for all $i = 1, \ldots, n$ and $f(y) \succeq f(x) + \sum_{j=1}^n (y_j - x_j) z_j^i$ for all $y \in V$.

As in the numerical case, we also have a differential-monotonicity criterion, as the next proposition shows.

Proposition 11. *The mapping f is K-convex if and only if $(df(x) - df(y))(x - y) \succeq 0$ for all $x, y \in \mathbb{R}^n$.*

Proof. Once again, the result follows from Proposition 2, since the convexity of ψ_z for all $z \in K^*$ is equivalent to the monotonicity of $\nabla\psi_z$ for all $z \in K^*$ which means that

$$0 \leq \langle \nabla\psi_z(x) - \nabla\psi_z(y), x - y\rangle = \langle z, (df(x) - df(y))(x - y)\rangle$$

for all $z \in K^*$, which in turn is equivalent, in view of Proposition 1, to $(df(x) - df(y))(x - y) \in (K^*)^* = K$, which occurs if and only if $(df(x) - df(y))(x - y) \succeq 0$.

□

3. First order optimality conditions for K-unconstrained problems

Assume now that $f : \mathbb{R}^n \to V$ is a differentiable K-convex mapping and consider problem (P):

$$\begin{aligned} &\min f(x) \\ &\text{s.t.} \quad x \in \mathbb{R}^n. \end{aligned}$$

The problem consists of finding $x^* \in \mathbb{R}^n$ such that $f(x^*) \preceq f(x)$, for all $x \in \mathbb{R}^n$ (or for all x in a neighbourhood of x^*, in the case of local minima), i.e., we do not seek a minimal element in the K-order but a global (or local) minimizer.

We will see that most of the main properties of the real valued unconstrained problem have its corresponding K-versions. We begin by showing that the optimal set of problem (P) is convex.

Proposition 12. *Assume that the cone K^* has nonempty topological interior. Then the set of solutions of problem (P), when nonempty, is a convex subset of $\mathbb{R}^n$.*

Proof. Let x^* be an optimal solution of problem (P). Note that $f(\bar{x}) \preceq f(x^*)$ is equivalent to $\psi_z(\bar{x}) \leq \psi_z(x^*)$, for all $z \in K^*$, where $\psi(z) = \langle z, f(x)\rangle$. So, from the optimality of x^*, it follows that $f(\bar{x}) \preceq f(x^*)$ iff $\psi_z(\bar{x}) = \psi_z(x^*)$, for all $z \in K^*$, which, in turn, is equivalent to the fact that $0 = \langle z, f(\bar{x}) - f(x^*)\rangle$, for all $z \in K^*$. Since the interior of K^* is nonempty, it contains a basis of V and the last equality implies that $f(\bar{x}) = f(x^*)$. Therefore, the set of optima of problem (P) is given by $\{x \in \mathbb{R}^n : f(x) \preceq f(x^*)\}$, which, by Proposition 6, is a convex subset of $\mathbb{R}^n$. □

Our next result shows that every local minimizer of problem (P) is indeed a global minimizer of the same problem.

Proposition 13. *Let x^* be a local minimizer of problem (P). Then x^* is a global minimizer of problem (P).*

Proof. Let $B(x^*, \varepsilon)$ be an open ball centered at x^*, with radius $\varepsilon > 0$ and such that $f(x^*) \preceq f(x)$, for all $x \in B(x^*, \varepsilon)$. Let y be an arbitrary point in $\mathbb{R}^n$. For $\lambda \in [0, 1]$ sufficiently small, we will have that $(1 - \lambda)x^* + \lambda y \in B(x^*, \varepsilon)$, so

$$f(x^*) \preceq f((1 - \lambda)x^* + \lambda y).$$

Then, the result follows from the above inequality combined with the K-convexity of f. □

Now we will see that, as in the numerical case, a necessary and sufficient condition for optimality of a point is the fact that the partial derivatives of the objective vanish at that point.

Proposition 14. *Assume that the cone $K^* \subset V$ has nonempty topological interior and that f is differentiable at $x^* \in \mathbb{R}^n$. Then x^* is a minimizer of problem (P) if and only if $\frac{\partial f}{\partial x_i}(x^*) = 0$, for $i = 1, \ldots, n$.*

Proof. If the partial derivatives of f vanish at x^*, then, by Proposition 10, x^* is a minimizer of problem (P). Conversely, if $f(x^*) \preceq f(x)$, for all $x \in \mathbb{R}^n$, then $\psi_z(x^*) \leq \psi_z(x)$, for all $z \in K^*$ and all $x \in \mathbb{R}^n$. Hence, $x^* \in \operatorname{argmin}\{\psi_z(x) : x \in \mathbb{R}^n\}$, for all $z \in K^*$. So,

$$0 = \frac{\partial \psi_z}{\partial x_i}(x^*) = \langle z, \frac{\partial f}{\partial x_i}(x^*)\rangle,$$

for all $i = 1, \ldots, n$ and all $z \in K^*$. Since K^* has nonempty interior, the above equation implies that $\frac{\partial f}{\partial x_i}(x^*) = 0$, for all $i = 1, \ldots, n$.

□

Note that, as in the real-valued function case, we do not need the K-convexity of f for proving the necessity.

4. Psd-convex mappings

Take now V as S^m, the subspace of $\mathbb{R}^{m\times m}$ consisting of all the symmetric matrices, with $\langle X, Y\rangle = \mathrm{tr}[XY] = \sum_{i,j=1}^m X_{ij}Y_{ij}$ and let K be the cone of positive semidefinite symmetric matrices of order m (denoted by S_+^m). Observe that S_+^m is a convex closed cone with nonempty interior, given by the set of positive definite symmetric matrices. Furthermore, we have that $(S_+^m)^* = S_+^m$. In this context, a S_+^m-convex mapping $f : \mathbb{R}^n \to S^m$ will be called *positive semidefinite convex* (psd-convex henceforth). Clearly, f is psd-convex iff $\Delta = \lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y) \succeq 0$, for any $x, y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$, i.e., if $u^t \Delta u \geq 0$ for every $u \in \mathbb{R}^m$, $x, y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$. Therefore, we have that f is psd-convex iff the scalar function $\phi_u : \mathbb{R}^n \to \mathbb{R}$, given by $\phi_u(x) = u^t f(x) u$, is convex for all $u \in \mathbb{R}^m$. In particular, taking $u = e^i$ $(i = 1, \ldots, m)$, where e^i is the i-th canonical vector, we see that f_{ii} $(i = 1, \ldots, m)$ is a convex function if f is a psd-convex mapping.

Now we will study some elementary properties of psd-convex mapping. First of all, observe that since S_+^m is a closed convex cone with nonempty interior and $(S_+^m)^* = S_+^m$, all properties established in Sections 2 and 3 for general K-convex mappings also hold for the psd-convex ones. Next we give a criterion of psd-convexity for mappings with simultaneously diagonalizable image.

Proposition 15. *Suppose that $f : \mathbb{R}^n \to S^m$ is such that $f(x)f(y) = f(y)f(x)$, for all $x, y \in \mathbb{R}^n$. Then f is psd-convex if and only if its eigenvalues are convex functions of x.*

Proof. We know that the conmutativity assumption is equivalent to the fact that the set $\{f(x) : x \in \mathbb{R}^m\}$ is simultaneously diagonalizable. Hence, there exists an orthogonal matrix Q such that $f(x) = Q^t diag(\lambda_1(f(x)), \ldots, \lambda_m(f(x)))Q$, for all $x \in \mathbb{R}^n$, where $\lambda_i(f(x))$ is the i-th eigenvalue of $f(x)$. We know that f is psd-convex iff $\phi_u(x) = u^t f(x) u$ is a convex function for all $u \in \mathbb{R}^m$, then the "if" part follows from the fact that $\phi_u(x) = \sum_{i=1}^m (Qu)_{ii}^2 \lambda_i(f(x))$; the "only if" part follows from the same inequality by taking $u = Q^{-1}e^i$ $(i = 1, \ldots, m)$, where e^i is the i-th canonical vector of $\mathbb{R}^m$.

□

We point out that, in general, the convexity of the eigenvalues of a mapping is not a necessary and sufficient condition for its psd-convexity. In Section 2 of [5] there is a 2×2 example of a non-psd-convex mapping for wich both eigenvalues are actually convex functions of x. What is always true, and will be shown next, is that the largest eigenvalue of a psd-convex mapping is a

convex function of x; once again, the converse is not true (see Section 2 of [5] for an example of a non-psd-convex mapping such that its largest eigenvalue is convex).

Proposition 16. *Assume that f is a psd-convex mapping. Then its largest eigenvalue is a convex function of $x \in \mathbb{R}^n$. If, furthermore, we assume that $f(x) \succeq 0$ for all $x \in \mathbb{R}^n$, then the Euclidean operator norm of f, $||f(x)||$, is a convex function of x. Moreover, if $\{f_i : i \in I\}$ is a collection of psd-convex mappings such that $f_i(x) \succeq 0$ for all $x \in \mathbb{R}^n$ and all $i \in I$, then $x \mapsto sup\{||f_i(x)|| : i \in I\}$ is a convex function.*

Proof. The first assertion follows from the fact that the largest eigenvalue of f is given by $\lambda_{max}(f(x))$ $= \max\{v^t f(x)v : v \in \mathbb{R}^m, ||v|| = 1\} = \max\{\phi_v(x) : v \in \mathbb{R}^m, ||v|| = 1\}$, because the maximum of convex functions is convex too. The second assertion follows from the first one because $f \succeq 0$ implies $||f(x)|| = \lambda_{max}(f(x))$. Finally, the third assertion follows from the second one in view of the fact that the pointwise supremum of convex functions is convex.

□

We point out that the first assertion of Proposition 16 has also been proved in Section 2 of [5]. A very useful fact that will be needed in the sequel is the following

$$f(x)_{ij} = (e^i)^t f(x) e^j = (\phi_{e^i}(x) + \phi_{e^j}(x) - \phi_{e^i - e^j}(x))/2, \tag{1}$$

where e^i, as usual, stands for the i-th canonical vector of $\mathbb{R}^m$ and $\phi_u(x) = u^t f(x) u$. Using this characterization of f's components, we can extend some well known properties of usual convex functions. For instance, we have the following result.

Proposition 17. *If f is a psd-convex mapping such that $f(x) \preceq M$ for some $M \in S^m$ and for all $x \in L \subset \mathbb{R}^n$, where L is in an affine manifold, then f is constant on L.*

Proof. If $f(x) \preceq M$, for all $x \in S$, then $\phi_u(x) \leq u^t M u \leq ||M||$, for all $u \in \mathbb{R}^m$ with $||u|| = 1$ and all $x \in L$. Hence, all the ϕ_u's are constant on S (see Corollary 8.6.2 in [4]). The result follows then from (1).

□

Using again (1) we can see that psd-convex mappings inherit some important continuity properties from the real-valued convex functions ϕ_u's.

Proposition 18. *Every psd-convex mapping f is continuous. Moreover, f is Lipschitz continuous on any compact subset of $\mathbb{R}^n$.*

Proof. Both results follow from (1), since they hold for scalar convex functions.

□

To end up with our list of properties of psd-convex mappings, we mention some properties related to differentability. The first one is the psd-version of a well known fact that holds for convex functions.

Proposition 19. *Let $f : \mathbb{R}^n \to S^m$ be a psd-convex mapping. Then f is differentiable almost everywhere. Moreover, the differential df is continuous at all points where it exists.*

Proof. These facts hold for numerical convex functions (see Theorem 7.8 of [7]), so the results are a consequence of (1).

□

An immediate consequence of Proposition 19 is that a differentiable psd-convex mapping is automatically continuously differentiable. Finally, we give a psd-convexity criterion for twice continuosly differentiable mapppings. Recall that the second differential of f at $x \in \mathbb{R}^n$ is given by the bilinear mapping $d^2f(x) : \mathbb{R}^n \times \mathbb{R}^n \to S^m$, defined by $d^2f(x)(u,v) = \sum_{i,j=1}^n u_i v_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$.

Proposition 20. *The mapping f is psd-convex if and only if $d^2f(x)(y,y)$ is positive semidefinite for all $x, y \in \mathbb{R}^n$.*

Proof. In order to prove the result, we just need to notice that $u^t d^2f(x)(y,y)u = y^t \nabla^2 \phi_u(x) y$ and use the fact that, for all $u \in \mathbb{R}^m$, due to the convexity of ϕ_u, its Hessian is positive semidefinite.

□

5. First order optimality conditions for psd-constrained problems

Given X, $Y \in S^m$, we say that $X \prec Y$ if and only if $Y - X \in \mathrm{int}(S^m_+)$, i.e. $Y - X$ belongs to the open cone S^m_{++} of symmetric positive definite matrices in $\mathbb{R}^{m \times m}$.

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):

$$\min f(x)$$
$$\text{s.t.} \quad g_i(x) \preceq 0 \quad (1 \leq i \leq p)$$

Let $F \subset \mathbb{R}^n$ be the feasible set for problem (CP), i.e. $F = \{x \in \mathbb{R}^n : g_i(x) \preceq 0 \ (1 \leq i \leq p)\}$. We will discuss in this paper the case in which a solution x^* of (CP) is such that $f(x^*)$ is a minimizer, and not just a minimal element in the partial order $\succeq$, i.e. x^* solves (CP) when $f(x) - f(x^*)$ is positive semidefinite for all $x \in F$. The Pareto optimality interpretation of (CP), meaning that x^* solves (CP) when $f(x^*) - f(x)$ is *not* positive semidefinite for any $x \in F$, will be discussed in a forthcoming paper.

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. We say that $\mathbf{Z} \in L(S^{r_i}, S^m)$ is nonnegative (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 $\mathbf{Z}^i \in L(S^m, S^{r_i})$ $(1 \leq i \leq p)$ is given by:

$$df(x) + \sum_{i=1}^{p} \mathbf{Z}^i dg_i(x) = 0 \qquad \text{Lagrangian condition,} \tag{2}$$

$$g_i(x) \preceq 0 \quad (1 \leq i \leq p) \qquad \text{primal feasibility,} \tag{3}$$

$$\mathbf{Z}^i \succeq 0 \quad (1 \leq i \leq p), \qquad \text{dual feasibility,} \tag{4}$$

$$\mathbf{Z}^i(g_i(x)) = 0 \quad (1 \leq i \leq p) \qquad \text{complementarity.} \tag{5}$$

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 (2) 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.

The analysis of the first order optimality conditions requires the following elementary result.

Proposition 21. *If A and B belong to S^m_+ then $tr[AB] \geq 0$.*

Proof. Since B belongs to S^m_+, there exists an orthogonal matrix Q and a nonnegative and diagonal matrix D with diagonal entries d_i $(1 \leq i \leq m)$ such that $B = Q^t D Q$. Since $\mathrm{tr}[XYZ] = \mathrm{tr}[Y^t X^t Z^t]$ for any matrices X, Y and Z, we have

$$\mathrm{tr}[AB] = \mathrm{tr}[AQ^t DQ] = \mathrm{tr}[QAQ^t D] = \sum_{i=1}^{m} (QAQ^t)_{ii} d_i. \tag{6}$$

Since A belongs to S^m_+, we have that QAQ^t also belongs to S^m_+, and thus it has nonnegative diagonal entries. The result follows from (6).

□

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 (2)–(5), under an additional and much stronger constraint qualification.

Theorem 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 $g_i(\hat{x}) \prec 0$, $(1 \leq i \leq p)$.
Then $x^* \in \mathbb{R}^n$ solves (CP) if and only if there exist mappings $Z^i : \mathbb{R}^m \to S^{r_i}$ $(1 \leq i \leq p)$ such that, for all $u \in \mathbb{R}^m$,

$$u^t df(x^*)u + \sum_{i=1}^{p} tr[Z^i(u) dg_i(x^*)] = 0 \qquad \textit{Lagrangian condition,} \tag{7}$$

$$g_i(x^*) \preceq 0 \quad (1 \leq i \leq p) \qquad \textit{primal feasibility,} \tag{8}$$

$$Z^i(u) \succeq 0 \quad (1 \leq i \leq p) \qquad \textit{dual feasibility,} \tag{9}$$

$$tr[Z^i(u) g_i(x^*)] = 0 \quad (1 \leq i \leq p) \qquad \textit{complementarity.} \tag{10}$$

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 $\sum_{k=1}^{r_i} \sum_{\ell=1}^{r_i} Z^i_{k\ell}(u) \frac{\partial g_i(x^*)_{k\ell}}{\partial x_j} = \mathrm{tr}[Z^i(u) \frac{\partial g_i(x^*)}{\partial x_j}]$.

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

$$h(Z^1, \ldots, Z^p)_j = -\sum_{i=1}^{p} \mathrm{tr}[Z^i A^{ij}], \tag{11}$$

$$C = \{(Z^1, \ldots, Z^p) \in S_+^{r_1} \times \cdots \times S_+^{r_p} : \mathrm{tr}[Z^i g_i(x^*)] = 0 \ (1 \leq i \leq p)\}. \tag{12}$$

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 (11), with $A^{ij} = \frac{\partial g_i(x^*)}{\partial x_j}$, 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

$$u^t[df(x^*)w]u = \langle u^t df(x^*)u, w\rangle < 0, \tag{13}$$

$$-\sum_{i=1}^{p} \mathrm{tr}[Z^i dg_i(x^*)w] = \langle h(Z^1, \ldots, Z^p), w\rangle \geq 0, \tag{14}$$

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

$$\theta = -u^t[df(x^*)w]u. \tag{15}$$

By (13), $\theta > 0$. Take $\beta > 0$ such that

$$\beta |u^t[df(x^*)(\hat{x} - x^*)]u| \leq \frac{\theta}{2}, \tag{16}$$

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

$$\hat{w} = w + \beta(\hat{x} - x^*). \tag{17}$$

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

$$(v^k)^t g_{i(k)}(x^* + \alpha_k\hat{w})v^k > 0 \tag{18}$$

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 (18), 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

$$\bar{v}^t g_\ell(x^*)\bar{v} = 0. \tag{19}$$

Let $\gamma = \bar{v}^t g_\ell(\hat{x})\bar{v}$. Since $||\bar{v}|| = 1$, $\gamma < 0$ by assumption (b). Then, using the generalized gradient inequality (Proposition 10) for the psd-convex mapping g_ℓ, we have

$$0 > \gamma = \bar{v}^t g_\ell(\hat{x})\bar{v} \geq \bar{v}^t g_\ell(x^*)\bar{v} + \bar{v}^t[dg_\ell(x^*)(\hat{x} - x^*)]\bar{v} = \bar{v}^t[dg_\ell(x^*)(\hat{x} - x^*)]\bar{v}, \tag{20}$$

using (19) 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 (19), $(Z^1, \ldots, Z^p) \in C$, as given by (12) and it follows from (14) that

$$0 \geq \sum_{i=1}^{p} \mathrm{tr}[Z^i dg_i(x^*)w] = \bar{v}^t[dg_\ell(x^*)w]\bar{v}. \tag{21}$$

Multiplying (20) by β and adding (21), we get in view of (17)

$$0 > \beta\gamma \geq \bar{v}^t[dg_\ell(x^*)(w + \beta(\hat{x} - x^*))]\bar{v} = \bar{v}^t[dg_\ell(x^*)\hat{w}]\bar{v}. \tag{22}$$

By (22) and continuity of dg_ℓ, which follows from Proposition 19, there exists $\varepsilon > 0$ such that, for all $v \in B(\bar{v}, \varepsilon)$ and all $\alpha \in [0, \varepsilon]$,

$$v^t[dg_\ell(x^* + \alpha\hat{w})\hat{w}]v \leq \frac{\beta\gamma}{2} < 0. \tag{23}$$

Note that

$$v^t g_\ell(x^* + \alpha\hat{w})v = v^t g_\ell(x^*)v + \alpha \int_0^1 v^t[dg_\ell(x^* + t\alpha\hat{w})\hat{w}]v dt. \tag{24}$$

Since $v^t g_\ell(x^*)v \leq 0$ for all $v \in \mathbb{R}^{r_\ell}$ because $x^* \in F$, it follows from (23) and (24) that

$$v^t g_\ell(x^* + \alpha\hat{w})v \leq \frac{\alpha\beta\gamma}{2} < 0 \tag{25}$$

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 (25) contradicts (18). 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

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

$$-\theta + \beta u^t[df(x^*)(\hat{x} - x^*)]u \leq -\theta + \frac{\theta}{2} = -\frac{\theta}{2} < 0, \tag{26}$$

using (17) in the first equality, (15) in the second one and (16) in the inequality. It follows from (26) that there exists $\hat{\alpha} \in [0, \bar{\alpha}]$ such that

$$0 > u^t f(x^* + \hat{\alpha}\hat{w})u - u^t f(x^*)u = u^t[f(x^* + \hat{\alpha}\hat{w}) - f(x^*)]u. \tag{27}$$

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 (27). 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 (7). Condition (8) holds because x^*, being a solution of (CP), belongs to F. Condition (9) and (10) follow from (12).

($\Leftarrow$) Assume that (7)–(10) hold. By Proposition 10 and psd-convexity of f, we have, for all $x \in F$,

$$f(x) \succeq f(x^*) + df(x^*)(x - x^*). \tag{28}$$

Take any $u \in \mathbb{R}^m$. By (28) and (7)

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

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

$$\sum_{i=1}^p \mathrm{tr}[Z^i(u)(g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*))] + \sum_{i=1}^p \mathrm{tr}[Z^i(u) g_i(x^*)] + \sum_{i=1}^p \mathrm{tr}[Z^i(u)(-g_i(x))]. \tag{29}$$

Since all the g_i's are psd-convex, by Proposition 10, $g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*) \in S_+^{r_i}$. Since $x \in F$, $-g_i(x) \in S_+^{r_i}$. By (9) and Proposition 21, both the second and the fourth term in the right hand side of (29) are nonnegative. The third term vanishes by (10). 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 (8), we conclude that x^* solves (CP). □

Note that, as could be expected, psd-convexity of f is not needed for establishing necessity of (7)–(10). 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., [1]). Nevertheless, we found some technical obstacles when trying to establish necessity of (7)–(10) 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 (7)–(10).

Now we can 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

$$\phi_u(x) = u^t f(x)u, \tag{30}$$

$$\mathcal{L}_u(x, Z) = \mathcal{L}_u(x, Z^1, \ldots, Z^p) = \phi_u(x) + \sum_{i=1}^{p} \mathrm{tr}[Z^i g_i(x)], \tag{31}$$

$$\psi_u(Z) = \inf_{x \in \mathbb{R}^n} \mathcal{L}_u(x, Z). \tag{32}$$

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 2. *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 $\psi_u(Z) \leq \phi_u(x)$.
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 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 (31), (32),

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

where the second inequality follows from Proposition 21 and the facts that $x \in F$, $Z \succeq 0$.
ii) Immediate from (i), noting that, in view of (30), 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 1 holds, and therefore there exist $Z^i(u) \in S^{r_i}$ $(1 \leq i \leq p)$ such that (7)–(10) hold. Take $Z^*(u) = (Z^1(u), \ldots, Z^p(u))$. By (10), $\mathrm{tr}[Z^i(u) g_i(x^*)] = 0$, so that

$$\mathcal{L}_u(x^*, Z^*(u)) = \phi_u(x^*). \tag{33}$$

By (31) and (7),

$$\nabla_x \mathcal{L}_u(x^*, Z^*(u)) = u^t df(x^*) u + \sum_{i=1}^{p} \mathrm{tr}[Z^i(u) dg_i(x^*)] = 0. \tag{34}$$

Since $\mathcal{L}_u(\cdot, Z^*(u))$ is convex, it follows from (34) that x^* minimizes $\mathcal{L}_u(\cdot, Z^*(u))$, i.e.,

$$\psi_u(Z^*(u)) = \mathcal{L}_u(x^*, Z^*(u)). \tag{35}$$

By (33) and (35), $\psi_u(Z^*(u)) = \phi_u(x^*)$, and thus $Z^*(u)$ solves (D_u) by (ii). ∎

Before establishing sufficiency and necessity of (2)–(5), we present an example for which (7)–(10) hold, but (2)–(5) do not hold, showing that an additional constraint qualification is indeed needed.

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

$$f(x) = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \tag{36}$$

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

$$g_1(x) = -x_1 - x_2 + 1, \tag{37}$$

$$g_2(x) = -x_2 - x_3 + 1, \tag{38}$$

$$g_3(x) = -x_1 + x_2 + 1, \tag{39}$$

$$g_4(x) = x_2 - x_3 + 1. \tag{40}$$

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

$$g_i(x^*) = 0 \qquad (1 \leq i \leq 4), \tag{41}$$

so that x^* is feasible, and that the constraints associated with (37) and (39) are equivalent to $|x_2| \leq x_1 - 1$, while those associated with (38) and (40) are equivalent to $|x_2| \leq x_3 - 1$. It follows that $x_1 - 1 \geq 0$, $x_3 - 1 \geq 0$, $(x_1 - 1)(x_3 - 1) \geq x_2^2$, 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 $g_i(\hat{x}) \prec 0$ $(1 \leq i \leq 4)$. We check next that (7)–(10) hold at x^*. Conditions (8) and (10) hold trivially because of (41), independently of the choice of $Z^i(u)$. Before presenting the appropriate $Z^i(u)$'s, note that, for all $x \in \mathbb{R}^3$, $u \in \mathbb{R}^2$,

$$u^t df(x^*)u = (u_1^2, 2u_1u_2, u_2^2), \tag{42}$$

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

$$\nabla g_1(x) = (-1, -1, 0)^t, \tag{43}$$

$$\nabla g_2(x) = (0, -1, -1)^t, \tag{44}$$

$$\nabla g_3(x) = (-1, 1, 0)^t, \tag{45}$$

$$\nabla g_4(x) = (0, 1, -1)^t. \tag{46}$$

It follows from (42)–(46) that in this case (7), written component-wise, becomes

$$u_1^2 = Z^1(u) + Z^3(u), \tag{47}$$

$$2u_1u_2 = Z^1(u) + Z^2(u) - Z^3(u) - Z^4(u), \tag{48}$$

$$u_2^2 = Z^2(u) + Z^4(u), \tag{49}$$

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

$$\big(Z^1(u), Z^2(u), Z^3(u), Z^4(u)\big) = \begin{cases} \big(\frac{1}{2}(u_1+u_2)^2 - u_2^2, u_2^2, \frac{1}{2}(u_1-u_2)^2, 0\big) & \text{if } \sqrt{2}|u_2| \leq |u_1+u_2| \\ \big(0, \frac{1}{2}(u_1+u_2)^2, u_1^2, u_2^2 - \frac{1}{2}(u_1+u_2)^2\big) & \text{if } \sqrt{2}|u_2| > |u_1+u_2|, \end{cases} \tag{50}$$

it is easy to check that x^* and these $Z^i(u)$'s satisfy (47)–(49) and (9).

On the other hand, we show next that there exist no $\mathbf{Z}^1, \ldots, \mathbf{Z}^4$ such that (2)–(5) hold. In particular, the system given by (2) and (4) 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 (2) component-wise (i.e. taking the partial derivatives with respect to x_1, x_2 and x_3), we get

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = Z^1 + Z^3, \tag{51}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = Z^1 + Z^2 - Z^3 - Z^4, \tag{52}$$

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = Z^2 + Z^4. \tag{53}$$

Conditions (51)–(53) can be rewritten, after partial elimination, as

$$Z^1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} + Z^4, \tag{54}$$

$$Z^2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} - Z^4, \tag{55}$$

$$Z^3 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} - Z^4. \tag{56}$$

Condition (55), 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 (54), $Z^1 \succeq 0$ implies that $\lambda = 1$, i.e. $Z^4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Replacing this value of Z^4 in (56) we get $Z^3 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$, in contradiction with $Z^3 \succeq 0$. Thus (2) and (4) make an infeasible system.

We introduce next the constraint qualification required for necessity of (2)–(5). Take h as in (11), i.e. $h : S^{r_1} \times \cdots \times S^{r_p} \to \mathbb{R}^n$ defined as

$$h(Z^1, \ldots, Z^p) = -\sum_{i=1}^{p} \mathrm{tr}[Z^i dg_i(x^*)] = -\sum_{i=1}^{p}\sum_{k=1}^{r_i}\sum_{\ell=1}^{r_i} Z^i_{k\ell} \nabla g_i(x^*)_{k\ell}, \tag{57}$$

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

$$V = \{(Z^1, \ldots, Z^p) \in S^{r_1} \times \cdots \times S^{r_p} : \mathrm{tr}[Z^i g_i(x^*)] = 0 \ (1 \le i \le p)\}. \tag{58}$$

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 $\ell \ge k$ at $x = x^*$ are linearly independent.

Proposition 22. *If the set $\{\nabla g_i(x^*)_{k\ell} : 1 \le i \le p, 1 \le k \le r_i, k \le \ell \le r_i\}$ is linearly independent then $Ker(h) = \{0\}$, and thus (CQ) holds.*

Proof. Take $(Z^1, \ldots, Z^p) \in \mathrm{Ker}(h)$. By (57) and the fact that both the Z^i's and the $g_i(x)$'s are symmetric, we have

$$0 = h(Z^1, \ldots, Z^p) = 2\sum_{i=1}^{p}\sum_{k=1}^{r_i}\sum_{\ell=k}^{r_i} Z^i_{k\ell} \nabla g_i(x^*)_{k\ell}. \tag{59}$$

In view of the linear independence of the $\nabla g_i(x^*)_{k\ell}$'s, we conclude from (59) 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 (58) and Proposition 22, 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 22 requires $n \geq \frac{1}{2}\sum_{i=1}^{p} r_i(r_i+1)$.

In order to prove necessity of (2)–(5) under Slater's condition and (CQ), we need the following lemma.

Lemma 1. *Under the hypothesis of Theorem 1, if x^* solves (CP) and the matrices $Z^i(u)$ satisfying (7)–(10) can be chosen so that they are component-wise quadratic functions of u, then (2)–(5) 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

$$Z^i(u)_{k\ell} = u^t M^{ik\ell} u. \tag{60}$$

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

$$[\mathbf{Z}^i(A)]_{qs} = \sum_{k=1}^{r_i}\sum_{\ell=1}^{r_i} M^{ik\ell}_{qs} A_{k\ell} \ \ (1 \leq q, s \leq m). \tag{61}$$

We check first that the $\mathbf{Z}^i$'s defined by (61) satisfy (4). 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

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

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

using (61) in the second equality, (60) in the fourth one, and (9) and Proposition 21 in the inequality. Thus (4) holds for $\mathbf{Z}^i$ as defined by (61).

Now we look at the Lagrangian condition (2). Replacing (60) and (61) in (7), we get easily

$$u^t\left[df(x^*) + \sum_{i=1}^{p} \mathbf{Z}^i dg_i(x^*)\right] u = u^t df(x^*) u + u^t\left[\sum_{i=1}^{p} \mathbf{Z}^i dg_i(x^*)\right] u = 0 \tag{62}$$

for all $u \in \mathbb{R}^m$. Let $\{e^1, \ldots, e^m\}$ be the canonical basis of $\mathbb{R}^m$. Substracting (62) 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 (2). Replacing now (60) and (61) in (10), and substracting the resulting equation with $u = e^q - e^s$ from the same equation with $u = e^q + e^s$, we get (5). Condition (3) holds because x^*, being a solution of (CP), belongs to F. □

Now we present our proof of necessity and sufficience of (2)–(5) under Slater's condition and (CQ) above.

Theorem 3. *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 (57) and (58).
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 (2)–(5) hold for $x = x^*$.

Proof. ($\Rightarrow$) Since we are within the hypotheses of Theorem 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 (7)–(10). In view of Lemma 1, 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 (7) and (10) has a unique solution given by

$$(\bar{Z}^1(u), \ldots, \bar{Z}^p(u)) = h_{/V}^{-1}(u^t df(x^*)u). \tag{63}$$

By Theorem 1, this unique solution satisfies (9) and (10). By linearity of $h_{/V}^{-1}$ there exist $b^i_{k\ell j} \in \mathbb{R}$ $(1 \leq i \leq p,\ 1 \leq k, \ell \leq r_i,\ 1 \leq j \leq n)$ such that, for all $x \in \mathbb{R}^n$, $[h_{/V}^{-1}(x)]^i_{k\ell} = \sum_{j=1}^n b^i_{k\ell j} x_j$. In view of (63), we have

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

$$\sum_{j=1}^n b^i_{k\ell j} \sum_{q=1}^m \sum_{s=1}^m u_q u_s \frac{\partial f(x^*)_{qs}}{\partial x_j} = \sum_{q=1}^m \sum_{s=1}^m u_q u_s \sum_{j=1}^n b^i_{k\ell j} \frac{\partial f(x^*)_{qs}}{\partial x_j}. \tag{64}$$

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

$$M^{ik\ell}_{qs} = \sum_{j=1}^n b^i_{k\ell j} \frac{\partial f(x^*)_{qs}}{\partial x_j}. \tag{65}$$

It follows from (64) and (65) 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 1.
($\Leftarrow$) Assume that (2)–(5) hold. By Proposition 10 and psd-convexity of f, we have, for all $x \in F$,

$$f(x) \succeq f(x^*) + df(x^*)(x - x^*). \tag{66}$$

Take any $u \in \mathbb{R}^m$. By (66) and (2)

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

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

$$f(x^*) + \sum_{i=1}^{p} \mathbf{Z}^i[g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*)] + \sum_{i=1}^{p} \mathbf{Z}^i[g_i(x^*)] + \sum_{i=1}^{p} \mathbf{Z}^i[-g_i(x)]. \tag{67}$$

Since all the g_i's are psd-convex, by Proposition 10, $g_i(x) - g_i(x^*) - dg_i(x^*)(x - x^*) \in S_+^{r_i}$. Since $x \in F$, $-g_i(x) \in S_+^{r_i}$. By (4) 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 (67) are nonnegative. The third term vanishes by (5). It follows that $f(x) \succeq f(x^*)$ for all $x \in F$. Since x^* belongs to F by (3), we conclude that x^* solves (CP). □

We remark that (CQ) implies that the dual solution $(\mathbf{Z}^1, \ldots, \mathbf{Z}^p)$ of (2)–(5) is unique, as the following result shows.

Proposition 23. *Under the hypotheses of Theorem 1, if (CQ) holds and (CP) has optimal solutions, then there exist unique* $\mathbf{Z}^1, \ldots, \mathbf{Z}^p \in L(S^{r_1}, S^m) \times \cdots \times L(S^{r_p}, S^m)$ *which satisfy* (2)–(5).

Proof. Fix $u \in \mathbb{R}^m$. By Theorem 1, if (CP) has optimal solutions, then there exist $x^* \in \mathbb{R}^n$, $Z^1(u), \ldots, Z^p(u) \in S^{r_1} \times \cdots \times S^{r_p}$ which satisfy (7)–(10). Under (CQ), the system of linear equations in $Z^1(u), \ldots, Z^p(u)$ given by (7) and (10) has a unique solution $\bar{Z}^1(u), \ldots, \bar{Z}^p(u)$. By Theorem 3, there exist $(\mathbf{Z}^1, \ldots, \mathbf{Z}^p) \in L(S^{r_1}, S^m) \times \cdots \times L(S^{r_p}, S^m)$ such that (2)–(5) are satisfied with $x = x^*$. Since $\mathbf{Z}^i$ is linear there exist $z^i_{qsk\ell} \in \mathbb{R}$ $(1 \leq k, \ell \leq r_i,\ 1 \leq q, s \leq m)$ such that

$$[\mathbf{Z}^i(A)]_{qs} = \sum_{k=1}^{r_i} \sum_{\ell=1}^{r_i} z^i_{qsk\ell} A_{k\ell}. \tag{68}$$

Pre and postmultiplying (2) and (5) by u, it is easy to check that the matrices $\hat{Z}^i(u)$ with entries

$$\hat{Z}^i(u)_{k\ell} = \sum_{q=1}^{m} \sum_{s=1}^{m} z^i_{qsk\ell} u_q u_s \tag{69}$$

also satisfy (7) and (10). We conclude that $\hat{Z}^i(u) = \bar{Z}^i(u)$ for all $i \in \{1, \ldots, p\}$, i.e., in view of (69),

$$\sum_{q=1}^{m} \sum_{s=1}^{m} z^i_{qsk\ell} u_q u_s = \bar{Z}^i(u)_{k\ell}. \tag{70}$$

Substracting (70) 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

$$z^i_{qsk\ell} = \frac{1}{4}\big[\bar{Z}^i(e^q + e^s)_{k\ell} - \bar{Z}^i(e^q - e^s)_{k\ell}\big]. \tag{71}$$

Since the $\bar{Z}^i(u)$ are unique, we get from (71) that the $z^i_{qsk\ell}$'s are uniquely determined, and then it follows from (68) 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 (2) and (5) has a unique solution. A fortiori, the full system (2)–(5) also has a unique solution. □

In connection with Theorem 3, we mention that the problem in Example 2 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 1, note that the $Z_i(u)$'s given by (50) 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, (2)–(5), and for the second one, (7)–(10).

Example 3. Take $n = 3$, $m = 2$, $p = 3$, $r_1 = r_2 = r_3 = 1$. We consider f as in (36) of Example 2, i.e. $f(x) = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix}$, and

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

Slater's condition is satisfied, e.g., by $\hat{x} = (2, 1, 3)^t$, and (CQ) holds by virtue of Proposition 22, 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 (2)–(5) 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, (7)–(10) 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.

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

$$\hat{\mathcal{L}}\big(x, \mathbf{Z}^1, \ldots, \mathbf{Z}^p\big) = f(x) + \sum_{i=1}^{p} \mathbf{Z}^i\big(g_i(x)\big). \tag{72}$$

Observe that when f and all the g_i's are psd-convex, $\hat{\mathcal{L}}\big(\cdot, \mathbf{Z}^1, \ldots, \mathbf{Z}^p\big)$ is also psd-convex for all $\mathbf{Z}^1, \ldots, \mathbf{Z}^p$, by Proposition 9, iii). For those $\mathbf{Z}^1, \ldots, \mathbf{Z}^p$ such that $\min_{x \in \mathbb{R}^n} \hat{\mathcal{L}}\big(x, \mathbf{Z}^1, \ldots, \mathbf{Z}^p\big)$, in the partial order $\preceq$ of S^m, exists, we define

$$\Psi\big(\mathbf{Z}^1, \ldots, \mathbf{Z}^p\big) = \min_{x \in \mathbb{R}^n} \hat{\mathcal{L}}\big(x, \mathbf{Z}^1, \ldots, \mathbf{Z}^p\big). \tag{73}$$

The duality theorem for this formulation is

Theorem 4. *Consider problem (CP) with f and all the g_i's psd-convex.*
i) (weak duality) For all $(\mathbf{Z}^1, \ldots, \mathbf{Z}^p) \in dom(\Psi)$ satisfying $\mathbf{Z}^1 \succeq 0, \ldots, \mathbf{Z}^p \succeq 0$, and all $x \in F$, it holds that $\Psi(\mathbf{Z}^1, \ldots, \mathbf{Z}^p) \preceq f(x)$.
ii) If there exist $\mathbf{Z}^* \in L(S^{r_1}, S^m) \times \cdots \times L(S^{r_p}, S^m)$ 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 $\mathbf{Z}^* \in L(S^{r_1}, S^m) \times \cdots \times L(S^{r_p}, S^m)$ such that $\Psi(\mathbf{Z}^*) = f(x^*)$ and $\mathbf{Z}^*$ solves max $\Psi(\mathbf{Z})$ subject to $\mathbf{Z} \succeq 0$.

Proof. i) By (72), (73),

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

where the second inequality follows from the facts that $x \in F$, so that $-g_i(x) \succeq 0$, and $\mathbf{Z}^1 \succeq 0, \ldots, \mathbf{Z}^p \succeq 0$.
ii) Immediate from (i).
iii) Under the assumptions of this item Theorem 3 holds, and therefore there exist $\mathbf{Z}^i \in L(S^{r_i}, S^m)$ $(1 \leq i \leq p)$ such that (2)–(5) hold. Take $\mathbf{Z}^* = (\mathbf{Z}^1, \ldots, \mathbf{Z}^p)$. By (5), $\mathbf{Z}^i(g_i(x^*)) = 0$, so that

$$\hat{\mathcal{L}}(x^*, \mathbf{Z}^*) = f(x^*). \tag{74}$$

By (72) and (2),

$$d_x \hat{\mathcal{L}}_u(x^*, \mathbf{Z}^*) = df(x^*) + \sum_{i=1}^{p} \mathbf{Z}^i dg_i(x^*) = 0. \tag{75}$$

Since $\hat{\mathcal{L}}_u(\cdot, \mathbf{Z}^*)$ is psd-convex, as we have already observed, it follows from (75) and Proposition 14 that x^* minimizes $\hat{\mathcal{L}}_u(\cdot, \mathbf{Z}^*)$, i.e. $\mathbf{Z}^* \in \mathrm{dom}(\Psi)$ and

$$\Psi(\mathbf{Z}^*) = \hat{\mathcal{L}}_u(x^*, \mathbf{Z}^*). \tag{76}$$

By (74) and (76), $\Psi(\mathbf{Z}^*) = f(x^*)$, 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 Dr. R.S. Burachik on the subject of this paper.

References

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

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

[3] Graña Drummond, L.M. and Peterzil, Y. The central path in smooth convex semidefinite programs, preprint.

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

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

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

[7] Sundaram, R.K. (1996). *A First Course in Optimization Theory*, Cambridge University Press.