Julio López; Héctor Ramírez C.

On the central paths and Cauchy trajectories in semidefinite programming

Kybernetika, Vol. 46 (2010), No. 3, 524--535

Persistent URL: http://dml.cz/dmlcz/140766

KYBERNETIKA — VOLUME *46* (2010), NUMBER 3, PAGES 524–535

ON THE CENTRAL PATHS AND CAUCHY TRAJECTORIES IN SEMIDEFINITE PROGRAMMING

Julio López and Héctor Ramírez C.

In this work, we study the properties of central paths, defined with respect to a large class of penalty and barrier functions, for convex semidefinite programs. The type of programs studied here is characterized by the minimization of a smooth and convex objective function subject to a linear matrix inequality constraint. So, it is a particular case of convex programming with conic constraints. The studied class of functions consists of spectrally defined functions induced by penalty or barrier maps defined over the real nonnegative numbers. We prove the convergence of the (primal, dual and primal-dual) central path toward a (primal, dual, primal-dual, respectively) solution of our problem. Finally, we prove the global existence of Cauchy trajectories in our context and we recall its relation with primal central path when linear semidefinite programs are considered. Some illustrative examples are shown at the end of this paper.

Keywords: semidefinite programming, central paths, penalty/barrier functions, Riemannian geometry, Cauchy trajectories

Classification: 90C22, 53C25.

1. INTRODUCTION

Semidefinite programming (SDP) has been one of the most developed subjects in optimization during the last decades. The tremendous research activity in this area was spurred by the discovery of important applications in structural design, support vector machines (e. g. [2]), combinatorial optimization (e. g. [6]), among others areas, as well as the development of efficient numerical algorithms. For more details, see the surveys of Todd [19] and of Vanderberghe and Boyd [20] and references therein.

Most of the applications mentioned above are modeled as linear SDP problems, that is, the minimization of a linear function subject to a linear matrix inequality. In this article, we consider a slight generalization allowing the objective function to be convex (but smooth), see problem (P) in Section 2.2. This convex SDP problem is obviously a convex program. So, it can be solved via convex programming methods. Among those methods, interior-point algorithms seem to be the most efficient ones in practice. Roughly speaking, they are based on the good properties exhibited for the central path with respect to the logarithm barrier function (e. g. Todd [19]). This explains the large literature existing on the subject of this paper. A very brief bibliographic summary on central paths for SDP problems is the following. In the case of linear SDP, the convergence of the central path towards a primal-dual optimal solution has been first established by Kojima et al. [12]. Some years later Halická et al. [9] characterized its limit as the solution (called *the analytic center*) of a related optimization problem when the strict complementarity condition holds. They have also shown that this characterization is not longer true when this conditions does not hold. For general convex (smooth) SDP problems, only Graña-Drummond and Peterzil [8] have established the convergence of the primal-dual central path but under a restrictive assumption (cf. [8, Assumption A2]). Further details on this subject can be found in E. de Klerk [11].

The relation between primal central path and the Cauchy trajectories (defined in a related Riemannian manifold) has also attired the attention of many researchers. For instance, Iusem et al. [10] study it for classical convex programs constrained only over the positive orthant, while Cruz Neto et al. [5] do it for linear SDP.

The aim of our paper is to extend some of the previous result to our convex SDP setting, allowing also the use of a larger class of penalty/barrier functions (not only limited to standard barrier functions). Indeed, this larger class consists of spectrally defined functions induced by a quite general class of functions defined over the real positive numbers. So, we prove that the (primal, dual and primal-dual) central path, with respect to a function in this class, is well-defined and converges toward a (primal, dual and primal-dual, resp.) solution of our problem. We also show that the Cauchy trajectories are also well-defined in our context. These results thus extend those from [1, 5, 10] mentioned above and constitute the core of our paper.

The outline of this paper is the following. In Section 2, we introduce the basic notation, our problem, its dual, and main assumptions of this paper, and the class of functions used in our analysis. In Section 3, we establish the divers notions of central paths, as well as their main properties. For this, we split this section into three subsections, devoted to primal, dual and primal-dual central path, respectively.

Finally, in Section 4, we introduce some basic concepts about Riemannian geometry, including the definition of the Cauchy trajectories for our convex SDP framework. We then prove they are well-defined and recall their existing connection with primal central path for linear SDP. Some illustrative examples are given at the end.

2. PRELIMINARIES

2.1. Notation

The following notation is used throughout this paper. The space of $n \times n$ symmetric real matrices is denoted by $\mathcal{S}^n$. This space is endowed with the trace inner product $\langle X, Y\rangle = \text{trace}(XY)$, for all $X, Y \in \mathcal{S}^n$, where the trace of a matrix $X = (X_{ij}) \in \mathcal{S}^n$ is defined as $\text{trace}(X) = \sum_{i=1}^n X_{ii}$. Also, $\mathcal{S}^n_+$ (resp. $\mathcal{S}^n_{++}$) denotes the cone of symmetric positive semidefinite (resp. definite) matrices. Its Lowner order $\succeq$ is defined by $X \succeq Y$ iff $X - Y \in \mathcal{S}^n_+$ (order $\succ$ is similarly defined). We denote the spectral decomposition of X in $\mathcal{S}^n$ by $\sum_{i=1}^n \lambda_i(X) e_i(X) e_i(X)^\top$, where $e_i(X)$'s are the orthonormal eigenvectors associated with the eigenvalues $\lambda_i(X)$'s of X. Finally, for any set $\mathcal{S}$, $\partial\mathcal{S}$ stands for its boundary and $\Gamma_0(\mathcal{S})$ denotes the set of extended real-valued, proper, closed and convex functions defined on $\mathcal{S}$.

2.2. Problem statement, basic duality notions and main assumptions

In this paper, we consider the following convex semidefinite programming problem:

$$(P) \qquad \min f(X) \,:\, X \in \mathcal{F}_P := \{X \in \mathcal{S}^n : \mathcal{A}X = b,\ X \succeq 0\},$$

where $f : \mathcal{S}^n \to \mathbb{R}$ is convex and twice continuously differentiable, $b \in \mathbb{R}^m$ and $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ is a linear operator, defined by $\mathcal{A}X := (\langle A_i, X\rangle)_{i=1}^m \in \mathbb{R}^m$, with $A_i \in \mathcal{S}^n$. It is easy to see that the adjoint operator of $\mathcal{A}$ is given by $\mathcal{A}^*y = \sum_{i=1}^m y_i A_i$. The classical Lagrangian dual problem associated with problem (P) is

$$(D) \qquad \max_{(S,y)\in\mathcal{S}^n\times\mathbb{R}^m} p(S,y) \,:\, S \succeq 0,$$

where $p : \mathcal{S}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$ is defined as $p(S,y) = b^\top y + \inf_{X\in\mathcal{S}^n}\{f(X) - \langle X, S + \mathcal{A}^*y\rangle\}$. Note that $(S^*, y^*) \in \mathcal{S}^n_+ \times \mathbb{R}^m$ is a solution of (D) if and only if there exists $X^* \in \mathcal{F}_P$ such that

$$S^* + \mathcal{A}^*y^* = \nabla f(X^*), \quad \langle X^*, S^*\rangle = 0. \tag{1}$$

Throughout this paper, we assume that the following standard assumptions hold without explicitly mentioning them in the statements of our results.

(A1) The optimal set of (P), denoted by $S(P)$, is nonempty and compact.

(A2) $\mathcal{F}_P^0 := \mathcal{F}_P \cap \mathcal{S}^n_{++} = \{X \in \mathcal{S}^n : \mathcal{A}X = b,\ X \succ 0\} \neq \emptyset$ (Slater's condition).

(A3) The matrices A_i, $i = 1, \dots, m$, are linearly independent.

Assumption (A3) is not a restrictive hypothesis and ensures the one-to-one correspondence between the dual variables y and S. Assumption (A2) ensures that the set of optimal solutions of (D), denoted by $S(D)$, is nonempty and compact. Finally, (A1)-(A2) guarantee that there is no duality gap between (P) and (D). These consequences are necessary for ensuring the existence of the (primal and dual) central path defined in Section 3 (e. g. [19]).

2.3. Penalty and barrier functions for positive semidefinite matrices

In this paper we are interested in functions υ satisfying the following properties:

$$\begin{cases} \text{(i)} & \upsilon \in \Gamma_0(\mathbb{R}), \\ \text{(ii)} & (0,\infty) \subset \operatorname{dom} \upsilon \subset [0,\infty), \\ \text{(iii)} & \upsilon \in C^3(0,\infty) \text{ and } \lim_{s\to 0^+} \upsilon'(s) = -\infty, \\ \text{(iv)} & \forall s > 0,\ \upsilon''(s) > 0. \end{cases} \tag{2}$$

This class of functions is known as *Legendre type functions* and is denoted by $\mathcal{L}$.

We divide $\mathcal{L}$ into two subclasses $\mathcal{L}_1$ and $\mathcal{L}_2$ defined by

$$\mathcal{L}_1 = \{\upsilon \in \mathcal{L} : \upsilon(0) = +\infty\} \ \text{ and } \ \mathcal{L}_2 = \{\upsilon \in \mathcal{L} : \upsilon(0) < +\infty\}.$$

For example, the functions $\upsilon_1(s) = -\log(s)$ and $\upsilon_2(s) = s^{-1}$ belong to the class $\mathcal{L}_1$ (barrier type functions), while the functions $\upsilon_3(s) = s\log(s) - s$ (with the convention $0\log 0 = 0$) and $\upsilon_4(s) = -\frac{1}{r}s^r$ (where $r \in (0,1)$) belong to the class $\mathcal{L}_2$.

Additionally, throughout this paper, when $\upsilon \in \mathcal{L}_1$ we assume that

$$\text{(v) } \upsilon \text{ is nonincreasing and } \lim_{s\to+\infty} \upsilon'(s) = 0,$$

while when $\upsilon \in \mathcal{L}_2$ we assume that

$$\text{(vi) for all } \alpha \in \mathbb{R},\ \{x \in \mathbb{R}_+ : \upsilon(x) \le \alpha\} \text{ is bounded.}$$

This larger class will be denoted by $\overline{\mathcal{L}}$. Note that υ_i, $i = 1, \dots, 4$, belong to $\overline{\mathcal{L}}$. This class provides us penalty and barrier functions for $\mathcal{S}^n_+$ via the next definition.

Definition 2.1. A function $\Psi : \mathcal{S}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be spectrally defined if there is a symmetric function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ such that

$$\Psi(X) = \Psi_g(X) := g(\lambda(X)), \qquad \forall X \in \mathcal{S}^n,$$

where $\lambda(X)$ is a vector whose components are the eigenvalues of X. Recall that a function $g : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is called symmetric if $g(x) = g(Px)$, for all permutation matrix $P \in \mathbb{R}^{n\times n}$.

For $\upsilon \in \overline{\mathcal{L}}$ we define $g(x) = \sum_{i=1}^n \upsilon(x_i)$. In this context, υ is called the *Legendre Kernel* of g. Clearly, it holds that $g \in \Gamma_0(\mathbb{R}^n)$. Moreover, function g is symmetric and is strictly convex on int(dom(g)), so this function *induces a spectrally defined function* given by $\Psi_\upsilon(X) := \Psi_g(X) = \sum_{i=1}^n \upsilon(\lambda_i(X))$.

For instance, functions υ_i, $i = 1, \dots, 4$, given above, induce the following spectrally defined functions: $\Psi_{\upsilon_1}(X) = -\text{trace}(\log(X)) = -\sum_i \log(\lambda_i(X))$, $X \succ 0$, $\Psi_{\upsilon_2}(X) = \text{trace}(X^{-1}) = \sum_i \lambda_i(X)^{-1}$, $X \succ 0$, $\Psi_{\upsilon_3}(X) = \text{trace}(X\log(X) - X) = \sum_i [\lambda_i(X)\log(\lambda_i(X)) - \lambda_i(X)]$, $X \succeq 0$ (setting $0\log 0 = 0$), $\Psi_{\upsilon_4}(X) = -\frac{1}{r}\text{trace}(X^r) = -\frac{1}{r}\sum_i \lambda_i(X)^r$, $X \succeq 0$, with $r \in (0,1)$. Here the logarithm function of a positive definite matrix X has been defined as $\log(X) = \sum_{i=1}^n \log(\lambda_i(X))e_i(X)e_i(X)^\top$, where $X = \sum_{i=1}^n \lambda_i(X)e_i(X)e_i(X)^\top$ denotes the spectral decomposition of X.

Remark 2.2. A function $\upsilon \in \mathcal{L}_1$ induces a spectrally defined function Ψ_υ which is a barrier function for $\mathcal{S}^n_+$, that is, for all $X_0 \in \partial\mathcal{S}^n_+$, $\lim_{X\to X_0} \Psi_\upsilon(X) = +\infty$. (cf. [17, Theorem 9.1])

Definition 2.3. A function $h \in \Gamma_0(\mathbb{R}^n)$ is essentially smooth if it is differentiable on $\operatorname{int}\operatorname{dom}(h)$ and for every sequence $(x^k) \subset \operatorname{int}\operatorname{dom}(h)$ converging to a boundary point of dom(h) we have $|\nabla h(x^k)| \to +\infty$ as $k \to +\infty$.

The following results follow from [13] and [14].

Proposition 2.4. Let $\upsilon \in \overline{\mathcal{L}}$ and Ψ_υ be its induced spectrally defined function. It holds that Ψ_υ belong to $\Gamma_0(\mathcal{S}^n_+) \cap C^3(\mathcal{S}^n_{++})$, is essentially smooth with $\text{int}(\text{dom}\Psi_\upsilon) = \mathcal{S}^n_{++}$ and is strictly convex on $\mathcal{S}^n_{++}$. Moreover, its gradient is given by

$$\nabla\Psi_\upsilon(X) = \sum_{i=1}^n \upsilon'(\lambda_i(X))e_i(X)e_i(X)^\top,$$

where $e_i(X)$ denotes the orthonormal eigenvector of X associated with the eigenvalue $\lambda_i(X)$, and $\nabla^2\Psi_\upsilon(X)$ is positive definite, for all $X \in \mathcal{S}^n_{++}$.

We end this section by characterizing the analycity of function Ψ_υ. This result is a direct consequence of [18, Section 2].

Lemma 2.5. The function $\upsilon : \mathbb{R} \to \mathbb{R}$ is analytic if and only if its induced spectrally defined function Ψ_υ is analytic.

3. CENTRAL PATHS IN SEMIDEFINITE PROGRAMMING

3.1. Properties of primal central path

From now on we assume that υ belong to $\overline{\mathcal{L}}$. Let Ψ_υ be its induced spectrally defined function. We define the primal penalty/barrier problem (parameterized by $\mu > 0$) by means of

$$(P_\mu) \qquad\qquad \min\{f(X) + \mu\Psi_\upsilon(X) : \ \mathcal{A}X = b \ , \ X \in \mathrm{dom}(\Psi_\upsilon)\}.$$

Let us denote by $X(\mu)$ the solution of (P_μ). The curve $\{X(\mu) : \mu > 0\}$ is called the *primal central path* with respect to Ψ_υ. Since $\upsilon \in \overline{\mathcal{L}}$, function Ψ_υ satisfies the assumptions given in [5, Section 3]. Consequently, the results therein can be directly extended to our context. We thus obtain the following proposition which proof is omitted.

Proposition 3.1. The following assertions hold true:
(i) The primal central path $\{X(\mu) : \mu > 0\}$ is well defined and belong to $\mathcal{F}^0_P$.
(ii) The function $\mu \mapsto \Psi_\upsilon(X(\mu))$ is nonincreasing.
(iii) The set $\{X(\mu) : 0 < \mu < \bar{\mu}\}$ is bounded, for each $\bar{\mu} > 0$.
(iv) All limits points of the primal central path belong to $S(P)$.

The next result shows the convergence of the primal central path with respect to functions in $\mathcal{L}_2$. It slightly generalizes [5, Theorem 3.2] from a linear SDP setting to our convex one. For the sake of space, we omit its proof.

Theorem 3.2. Suppose that $\upsilon \in \mathcal{L}_2$. Then the primal central path with respect to Ψ_υ converges, when $\mu \to 0$, toward the unique minimizer of Ψ_υ over $S(P)$.

3.2. Properties of dual central path

The dual central path $\{(S(\mu), y(\mu)) : \mu > 0\}$ with respect to Ψ_υ is defined as follows

$$S(\mu) = -\mu\nabla\Psi_\upsilon(X(\mu)) \quad \text{ and } \quad y(\mu) = (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(\nabla f(X(\mu)) - S(\mu)). \tag{3}$$

Notice that Proposition 3.1(i) ensures that dual central path is well-defined.

In the next result, we prove the boundedness of the dual central path and that its limit points are solution of the problem (D).

Proposition 3.3. The dual central path defined in (3) is bounded and each of its limit points belongs to $S(D)$.

Proof. In the case when $\upsilon \in \mathcal{L}_1$, the result was proven in [3, Theorem 4.1]. Otherwise, when $\upsilon \in \mathcal{L}_2$, we will first prove the boundedness of the dual central path by verifying that the eigenvalues of $S(\mu)$ are bounded (indeed, the boundedness of $y(\mu)$ follows directly from its definition in (3)). Then, we proceed to prove that every limit point satisfies optimality conditions (1). Let us demonstrate that the eigenvalues of $S(\mu)$ are bounded. Theorem 3.2 ensures the convergence of $X(\mu)$, when $\mu \to 0$, toward a particular solution in $S(P)$. We denote this solution by $\bar{X}$. It follows that the eigenvalues $\lambda_i(X(\mu))$ of $X(\mu)$ converge toward the eigenvalues $\lambda_i(\bar{X})$ of $\bar{X}$. Define the index sets $I_0 = \{i : \lambda_i(\bar{X}) = 0\}$ and $I_+ = \{1, \dots, n\} \setminus I_0 = \{i : \lambda_i(\bar{X}) > 0\}$. We will study separately the cases $i \in I_0$ and $i \in I_+$. From the definition of $S(\mu)$, given in (3), we know that its eigenvalues are given by

$$\lambda_i(S(\mu)) = -\mu\upsilon'(\lambda_i(X(\mu))), \quad \text{for all } i = 1, \dots, n. \tag{4}$$

Then, continuity of υ' implies that $\upsilon'(\lambda_i(X(\mu))$ converge to the finite value $\upsilon'(\lambda_i(\bar{X}))$, for all $i \in I_+$. Consequently, $\lambda_i(S(\mu))$ converge to 0, for all $i \in I_+$. This will be also useful when we prove the optimality of the limit points.

On the other hand, relation (4) and hypothesis (2), Part (iii), implies that $\lambda_i(S(\mu))> 0$, for all $i \in I_0$, when μ is small enough. So, for these indexes $i \in I_0$, we can argue by contradiction and suppose the existence of a sequence of positive numbers $\{\mu_k\}$ satisfying that $\mu_k \to 0$ and $\zeta_k = \sum_{i\in I_0} \lambda_i(S(\mu_k)) \to +\infty$, when $k \to +\infty$. Without loss of generality we can assume that there exist positive values $\hat{\xi}^i$, not all equal to zero, such that $\xi_k^i = \zeta_k^{-1}\lambda_i(S(\mu_k)) \to \hat{\xi}^i$, for all $i \in I_0$. Consider the spectral decompositions $X(\mu) = \sum_{i=1}^n \lambda_i(X(\mu))e_i(X(\mu))e_i(X(\mu))^\top$ and $\bar{X} = \sum_{i=1}^n \lambda_i(\bar{X})e_i(\bar{X})e_i(\bar{X})^\top$, so that their orthonormal eigenvectors are chosen in order to satisfy $e_i(X(\mu_k)) \to e_i(\bar{X})$, for all $i \in I_0$, when $k \to +\infty$ (see, for instance, [3, Lemma 2.3]). From the definitions of $S(\mu)$ and $y(\mu)$, given in (3), we obtain

$$\mathcal{A}^*y(\mu_k)+S(\mu_k) = \mathcal{A}^*[(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(\nabla f(X(\mu_k))-S(\mu_k))]+S(\mu_k) = \nabla f(X(\mu_k)). \tag{5}$$

Dividing the above expression by ζ_k and taking $k \to +\infty$, we get

$$(I - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}) \sum_{i\in I_0} \hat{\xi}^i e_i(\bar{X})e_i(\bar{X})^\top = 0,$$

that is, matrix $\hat{S} = \sum_{i\in I_0} \hat{\xi}^i e_i(\bar{X})e_i(\bar{X})^\top$ belongs to $(\operatorname{Ker}\mathcal{A})^\perp = \operatorname{Im}\mathcal{A}^*$, or equivalently, there exists $\hat{y} \in \mathbb{R}^m$ such that $\mathcal{A}^*\hat{y} + \hat{S} = 0$. Moreover, since $\hat{S} \succeq 0$ and $\langle \hat{S}, \bar{X}\rangle = 0$, it follows that $S(D)$ is unbounded (it is enough to check that $S_\alpha = S^* + \alpha\hat{S}$ and $y_\alpha = y^* + \alpha\hat{y}$ satisfy (1), for all $\alpha \geq 0$ and $(S^*, y^*) \in S(D)$). This contradicts Slater's condition (A2). The boundedness of the dual central path follows.

Consider now a sequence $(S(\mu_k), y(\mu_k))$ converging to $(\bar{S}, \bar{y})$. Since $\lambda_i(\bar{S}) = 0$ for all $i \in I_+$, complementarity condition in(1) is trivially satisfied. Moreover, it follows from the analysis above that $\bar{S} \succeq 0$. Finally, passing to the limit $k \to +\infty$ in (5), we conclude that $(\bar{X}, \bar{S}, \bar{y})$ satisfies (1). The optimality of $(\bar{S}, \bar{y})$ is thus verified. □

3.3. Convergence of the primal-dual central path

The *primal-dual central path* with respect to Ψ_υ is the curve $\{(X(\mu), S(\mu), y(\mu)) : \mu > 0\}$, where $X(\mu)$ and $(S(\mu), y(\mu))$ were defined in the previous sections. Hence, this curve is well defined as the unique solution of the optimality conditions of (P_μ):

$$\mathcal{A}X = b, \quad X \succ 0, \quad \mathcal{A}^*y + S = \nabla f(X), \quad S + \mu\nabla\Psi_\upsilon(X) = 0. \tag{6}$$

In the following result we prove that the primal-dual central path is an analytic curve. It generalizes [11, Theorem 3.3] from a linear SDP setting to our convex one.

Proposition 3.4. Suppose that f and υ are analytic functions. Then the primal-dual central path with respect to Ψ_υ is an analytic curve contained in $\mathcal{S}^n_{++} \times \mathcal{S}^n \times \mathbb{R}^m$.

Proof. Define the function $\Xi : \mathcal{S}^n_{++} \times \mathbb{R}^m \times \mathcal{S}^n \times \mathbb{R}_{++} \to \mathbb{R}^m \times \mathcal{S}^n \times \mathcal{S}^n$ as $\Xi(X, y, S, \mu) = (\mathcal{A}X - b, \mathcal{A}^*y + S - \nabla f(X), S + \mu\nabla\Psi_\upsilon(X))$. Clearly, $\Xi(X, y, S, \mu) = 0$ is equivalent to (6). Then, we have that $\Xi(X(\mu), y(\mu), S(\mu), \mu) = 0$, for all $\mu > 0$. Moreover, it follows from Lemma 2.5 that Ξ is an analytic function. Hence, the result is obtained when we applied the implicit function theorem for analytic functions (see [7, Theorem 10.2.4]) to the function Ξ. For this, we only need to prove that its derivative with respect to (X, y, S) is nonsingular everywhere, or equivalently, that the kernel of

$$\nabla_{(X,y,S)}\Xi(X, y, S, \mu) = \begin{pmatrix} \mathcal{A} & 0 & 0 \\ -\nabla^2 f(X) & \mathcal{A}^* & I \\ \mu\nabla^2\Psi_\upsilon(X) & 0 & I \end{pmatrix}$$

is reduced to 0. We proceed to prove this claim. Let $U, W \in \mathcal{S}^n$ and $z \in \mathbb{R}^m$ s.t.

$$\mathcal{A}U = 0, \quad -\nabla^2 f(X)U + \mathcal{A}^*z + W = 0, \quad \mu\nabla^2\Psi_\upsilon(X)U + W = 0. \tag{7}$$

First and second equation in (7) implies that $\langle W, U\rangle = \langle\nabla^2 f(X)U, U\rangle$, obtaining from the third one that

$$\langle[\nabla^2 f(X) + \mu\nabla^2\Psi_\upsilon(X)]U, U\rangle = 0. \tag{8}$$

Since f is convex and $\nabla^2\Psi_\upsilon(X)$ is positive definite (cf. Proposition 2.4), it follows from (8) that $U = 0$. Finally, substituting this in (7) and using that A_i are linearly independent matrices, we conclude that $W = 0$ and $z = 0$. The result follows. □

We state here below the convergence of primal-dual central path. Its proof follows word by word the article [9] (see also [11]), where the result was shown only for the logarithmic barrier function in a linear SDP setting. Thus, it is omitted. The only important difference is that here we need to work with semi-analytic sets (and its corresponding curve selection lemma) instead of algebraic ones. We address the reader to [15] for the definitions.

Theorem 3.5. Suppose that υ and f are analytic functions. Then the primal-dual central path with respect to Ψ_υ converges, when $\mu \to 0$, toward a point in $S(P) \times S(D)$.

4. CAUCHY TRAJECTORIES FOR SEMIDEFINITE PROGRAMMING

In this section, we prove that the Cauchy trajectories (defined in a suitable chosen Riemannian manifold) are well-defined even for our convex SDP problem (P), and we recall the existing connection between primal central path and the Cauchy trajectories when the (P) is a linear SDP problem (that is, when f is a linear mapping).

4.1. Some basic concepts on Riemannian geometry

Given $X \in \mathcal{S}^n_{++}$ we introduce a new inner product on $\mathcal{S}^n_{++}$ as follows:

$$(U, V)_X = \langle \nabla^2 \Psi_\upsilon(X) U, V \rangle, \quad \forall\, U, V \in \mathcal{S}^n. \tag{9}$$

Now, since $\mathcal{S}^n_{++}$ is an open set contained in an Euclidean space, it can be seen as a smooth manifold. Moreover, for every $X \in \mathcal{S}^n_{++}$, the tangent space $T_X \mathcal{S}^n_{++}$ can be identify with $\mathcal{S}^n$. Hence, $M := (S_{++}, (\cdot,\cdot)_X)$ can be seen as a Riemannian manifold. The corresponding gradient vector field of f restricted to M is thus given by

$$\text{grad } f|_M(X) = [\nabla^2 \Psi_\upsilon(X)]^{-1} \nabla f(X), \tag{10}$$

where $\nabla f(X)$ denotes the Euclidean gradient in $\mathcal{S}^n$. See [16] for more details.

Consider $N = \mathcal{F}^0_P$ as a smooth submanifold of $\mathcal{S}^n_{++}$. Its tangent space is given by $T_X N = \operatorname{Ker} \mathcal{A} = \{V \in \mathcal{S}^n : \ \mathcal{A}V = 0\}$, for all $X \in \mathcal{F}^0_P$. On the other hand, the Riemannian gradient of the restriction of f to N at $X \in N$ can be computed as

$$\text{grad } f|_N(X) = \Pi_X \left([\nabla^2 \Psi_\upsilon(X)]^{-1} \nabla f(X) \right), \tag{11}$$

where $\Pi_X : \mathcal{S}^n \to \operatorname{Ker} \mathcal{A}$ is the $(\cdot,\cdot)_X$−orthogonal projection onto the linear subspace $\operatorname{Ker} \mathcal{A}$. Since $\mathcal{A}$ is onto, this projection can be explicitly stated for all $X \in N$, obtaining the following expression:

$$\text{grad } f|_N(X) = [\nabla^2 \Psi_\upsilon(X)]^{-1} [I - \mathcal{A}^* (\mathcal{A} [\nabla^2 \Psi_\upsilon(X)]^{-1} \mathcal{A}^*)^{-1} \mathcal{A} [\nabla^2 \Psi_\upsilon(X)]^{-1}] \nabla f(X). \tag{12}$$

Now, we are able to define the *Cauchy problem* as follows

$$\begin{cases} Z'(t) &= -\text{grad } f|_N(Z(t)), \\ Z(0) &= Z^0 \in \mathcal{F}^0_P. \end{cases} \tag{13}$$

The solutions of (13) are called *Cauchy trajectories*. We study their properties in the next section.

4.2. Global existence of the Cauchy trajectories in convex SDP

Under the assumptions on υ and f, the Cauchy-Lipschitz theorem ensures the existence and uniqueness (at least locally) of a classical solution of (13). Then

$$T_{\max} = \sup\{T > 0 : \exists!\text{ solution } Z \text{ of (13) on } [0, T) \text{ s.t. } Z([0, T)) \subset \mathcal{F}^0_P\} \tag{14}$$

is always positive. In this section, we prove that the Cauchy trajectories are defined globally (i. e. $T_{\max} = +\infty$). This analysis extends the result proven by Alvarez et al. [1] from a convex programming framework to our convex SDP one.

Since Ψ_υ is a strictly convex and essentially smooth (cf. Proposition 2.4), we can easily adapt the following technical results from [1]. Their proofs are then omitted.

Lemma 4.1. (i) Let $\{X_k\}$ be a sequence of matrices in $\mathcal{S}^n_{++}$ converging toward X^* in $\partial\mathcal{S}^n_{++}$. Then, every limit point of $\frac{\nabla\Psi_\upsilon(X_k)}{\|\nabla\Psi_\upsilon(X_k)\|}$ belong to the normal cone $\mathcal{N}_{\mathcal{S}^n_+}(X^*)$.
(ii) If $X^* \in \partial\mathcal{F}^0_P = \partial\mathcal{S}^n_{++} \cap \mathcal{A}^{-1}(b)$, then $\mathcal{N}_{\mathcal{S}^n_+}(X^*) \cap \operatorname{Im}\mathcal{A}^* = \{0\}$.

Now we state the main result of this section.

Theorem 4.2. The following statements hold:
(i) The trajectory $Z(t)$ is well-defined for all $t \geq 0$, that is, $T_{\max} = +\infty$.
(ii) The mapping $t \mapsto f(Z(t))$ is nonincreasing.
(iii) The mapping $t \mapsto (Z'(t), Z'(t))_{Z(t)}$ belongs to $L^1([0,+\infty),\mathbb{R})$.
(iv) The curve $f(Z(t))$ converges, when $t \to +\infty$, toward optimal value of (P).

Proof. Let $Z : [0, T_{\max}) \to \mathcal{F}^0_P$ be a maximal solution of (13). The schedule of the proof is the following. We first prove that the properties stated in Parts (ii) and (iii) holds true for Z on $[0, T_{\max})$. Then, we demonstrate Part (i) (i. e. $T_{\max} = +\infty$). It implies that Parts (ii) and (iii) are true. Finally, the convergence of $f(Z(t))$ is directly verified and, in order to conclude Part (iv), we only need to prove that its limit coincides with the optimal value of (P), which can be done using the same arguments (related to Bregman distances) of [1, Proposition 4.4]. This part of the proof is then omitted.

The equation (12) directly implies that $Z' \in \operatorname{Ker}\mathcal{A}$. Moreover, for all $Y \in \operatorname{Ker}\mathcal{A}$, we also obtain from (12) that $(Z' + [\nabla^2\Psi_\upsilon(Z)]^{-1}\nabla f(Z), Y + Z')_Z = 0$, or equivalently

$$\langle \nabla f(Z) + \nabla^2\Psi_\upsilon(Z)Z', Y + Z'\rangle = 0, \quad Y \in \operatorname{Ker}\mathcal{A}. \tag{15}$$

Replacing $Y = 0$ in (15) yields

$$\frac{d}{dt}f(Z) + \langle \nabla^2\Psi_\upsilon(Z)Z', Z'\rangle = \langle \nabla f(Z) + \nabla^2\Psi_\upsilon(Z)Z', Z'\rangle = 0. \tag{16}$$

Since $\nabla^2\Psi_\upsilon(Z) \succ 0$, we obtain from (16) that $t \mapsto f(Z(t))$ is nonincreasing on $[0, T_{\max})$. Moreover, since f is bounded from below on $\mathcal{F}_P$ (cf. (A1)), we get that $f(Z(t))$ converges when $t \nearrow T_{\max}$. So, if we integrate (16) from 0 to $t < T_{\max}$, the resulting expression

$$\int_0^t \langle \nabla^2\Psi_\upsilon(Z)Z', Z'\rangle ds = -\int_0^t \frac{d}{ds}f(Z)ds = -f(Z(t)) + f(Z^0)$$

shows us that $\langle \nabla^2\Psi_\upsilon(Z(\cdot))Z'(\cdot), Z'(\cdot)\rangle \in L^1([0, T_{\max}); \mathbb{R})$.

Now, let us prove that $T_{\max} = +\infty$ by contradiction, i. e. we assume that $T_{\max} < +\infty$. Since $f(Z(t))$ is nonincreasing, it follows that $Z(t) \in \Gamma_{f(Z^0)} := \{Y \in \mathcal{F}_P : f(Y) \leq f(Z^0)\}$, for all $t \in [0, T_{\max})$. So, boundedness of $\Gamma_{f(Z^0)}$ (cf. (A1)) implies that the trajectory $\{Z(t) : t \in [0, T_{\max})\}$ is bounded. Consequently, its set of limit points, denoted by Ω, is nonempty. Define $K = \{Z(t) : t \in [0, T_{\max})\} \cup \Omega$. This set is clearly compact. Then, if we show that $K \subset \mathcal{S}^n_{++}$, we contradict the maximality of $T_{\max}$. Once again we argue by contradiction, that is, we assume the existence of $X^* \in K \cap \partial\mathcal{S}^n_{++}$. Denote by $\{Z(t_k)\}$ the sequence in K converging toward X^* when $k \to +\infty$ (or $t_k \nearrow T_{\max}$). Since Ψ_υ is essentially smooth (cf. Proposition 2.4), we have that $\|\nabla\Psi_\upsilon(Z(t_k))\| \to +\infty$, when $k \to +\infty$. So, with out loss of generality we can assume that $\frac{\nabla\Psi_\upsilon(Z(t_k))}{\|\nabla\Psi_\upsilon(Z(t_k))\|} \to V$, for some $V \in \mathcal{S}^n \setminus \{0\}$. Furthermore, Lemma 4.1, Part (i), implies that V belongs to the normal cone $\mathcal{N}_{\mathcal{S}^n_+}(X^*)$. Let V_0 be the orthogonal projection of V onto $\operatorname{Ker}\mathcal{A}$. Notice that $V_0 \neq 0$. Indeed, otherwise $V \in (\operatorname{Ker}\mathcal{A})^\perp \cap \mathcal{N}_{\mathcal{S}^n_+}(X^*)$ and Lemma 4.1, Part (ii), implies that $V = 0$. By replacing $Y = V_0$ in (15) and using (16), we get

$$\langle \nabla f(Z), V_0\rangle + \langle \nabla^2\Psi_\upsilon(Z)Z', V_0\rangle = 0.$$

By integrating the expression above from 0 to t_k, we obtain

$$\langle \nabla\Psi_\upsilon(Z(t_k)), V_0\rangle = \langle \nabla\Psi_\upsilon(Z^0), V_0\rangle - \int_0^{t_k} \langle \nabla f(Z(s)), V_0\rangle ds. \tag{17}$$

Note that the right-hand side of (17) is bounded (because f is twice continuously differentiable and the curve $\{Z(t) : t \in [0, T_{\max})\}$ is bounded). Then, if we divide (17) by $\|\nabla\Psi_\upsilon(Z(t_k))\|$ and take limit $k \to +\infty$, we deduce that $\langle V, V_0\rangle = 0$, and consequently $V_0 = 0$, which is a contradiction. Hence, the theorem follows. □

Remark 4.3. Following the arguments given in [1, Theorem 4.1], the previous theorem can be extended to the case when f is a nonconvex function. For this, some additional (but standard) hypotheses on υ are needed[1]. For the sake of space we have focused here only on the convex case.

Once we have verified that the Cauchy trajectories are well-defined, we would like to relate them with the central path defined in Section 3. Unfortunately, this cannot be done for our convex (nonlinear) problem (P) (a counterexample in a convex programming framework is shown in [10]). However, it is indeed possible when f is linear as it was shown in Cruz Neto et al. [5]. We recall this result here below.

Lemma 4.4. Suppose that f is linear. If $Z^0 \in \mathcal{F}^0_P$ satisfies that $\nabla\Psi_\upsilon(Z^0) \in \operatorname{Im}\mathcal{A}^*$, then the Cauchy trajectory $Z(t)$ coincides with the curve $X(1/t)$, which is just a different parameterization of the primal central path with respect to Ψ_υ.

We end this article illustrating the main result of this section for a particular instance of problem (P). That is, we establish the Cauchy problem, given by (13), for which its solution converges to the optimum of (P).

Example 4.5. Consider problem (P) where the linear function $\mathcal{A}$ is given by $\mathcal{A}X = \operatorname{trace}(X)$ and $b = 1$. So, the feasible set of (P) is given by $\mathcal{F}_P = \{X \in \mathcal{S}^n_+ : \operatorname{trace}(X) = 1\}$. We chose the penalty function $\upsilon(s) = \upsilon_3(s) = s\log(s) - s$. In this case, $[\nabla^2\Psi_\upsilon(Z)]^{-1}Y = Z^{1/2}YZ^{1/2}$ for all $Y \in \mathcal{S}^n$. Then, the Cauchy problem (13) is given by

$$Z' + Z^{1/2}\nabla f(Z)Z^{1/2} - \operatorname{trace}(Z\nabla f(Z))Z = 0. \tag{18}$$

[1]Actually, these hypotheses are satisfied by functions υ_i, $i = 1, \ldots, 4$, given in Section 2.3.

In the particular case when f is spectrally defined, that is, when $f(X) = g(\lambda(X))$ for some symmetric function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, problem (P) ca be solved as the convex program where function g is minimized over the simplex in $\mathbb{R}^n$. On the other hand, our Cauchy problem (18) can be simplify by considering that the orthonormal eigenvectors $e_i(Z)$ are constant over time, obtaining the following differential equations on the eigenvalues $\lambda_i(\cdot) = \lambda_i(Z(\cdot))$ of Z:

$$\lambda_i'(t) + \lambda_i(t)\left[\frac{\partial g}{\partial x_i}(\lambda(Z(t))) - \sum_{j=1}^{n} \lambda_j(t)\frac{\partial g}{\partial x_j}(\lambda(Z(t)))\right] = 0, \quad i = 1, \dots, n.$$

For suitable choices of g, this is a Lotka-Volterra-type system that naturally arises in population dynamics theory. We thus recover the result obtained in Alvarez et al.[1].

Example 4.6. Consider problem (P) defined in the previous example and chose the logarithmic barrier function $\upsilon(s) = \upsilon_1(s) = -\log(s)$. In this case, $[\nabla^2\Psi_\upsilon(Z)]^{-1}Y = ZYZ$ for all $Y \in \mathcal{S}^n$. Then, the Cauchy problem (13) is given by

$$Z' + Z\nabla f(Z)Z - \frac{\operatorname{trace}(Z^2\nabla f(Z))}{\operatorname{trace}(Z^2)}Z^2 = 0. \tag{19}$$

In the particular case when f is spectrally defined, the Cauchy problem (19) can be simplify as before, obtaining the following differential equations on the eigenvalues $\lambda_i(\cdot)$ of Z:

$$\lambda_i'(t)+\lambda_i^2(t)\left[\frac{\partial g}{\partial x_i}(\lambda(Z(t))) - \frac{1}{\sum_{j=1}^n \lambda_j^2(t)}\sum_{j=1}^{n} \lambda_j^2(t)\frac{\partial g}{\partial x_j}(\lambda(Z(t)))\right] = 0, \quad i = 1, \dots, n.$$

This equation was considered by Bayer and Lagarias [4] for linear programs. For more details, see [1, Section 4.4] and references therein.

ACKNOWLEDGEMENT

The authors wish to thank to Prof. Felipe Alvarez for stimulating conversations about Section 4 of this paper. This work was supported by FONDECYT grant 1070297, FONDAP in Applied Mathematics and BASAL project.

(Received April 14, 2010)

REFERENCES

[1] F. Alvarez, J. Bolte, and O. Brahic: Hessian Riemannian flows in convex programming. SIAM J. Control Optim. *43* (2004), 2, 477–501.

[2] F. Alvarez, J. López and H. Ramírez C.: Interior proximal algorithm with variable metric for second-order cone programming: applications to structural optimization and classification by hyperplanes. To appear in Optimization Methods and Software.

[3] A. Auslender and H. Ramírez C.: Penalty and barrier methods for convex semidefinite programming. Math. Methods Oper. Res. *43* (2006), 2, 195–219.

[4] D. A. Bayer and J. C. Lagarias: The nonlinear geometry of linear programming I. Affine and projective scaling trajectories. Trans. Amer. Math. Soc. *314* (1989), 499–526.

[5] J. X. Cruz Neto, O. P. Ferreira, P. R. Oliveira, and R. C. M. Silva: Central paths in semidefinite programming, generalized proximal point method and Cauchy trajectories in Riemannian manifolds. J. Optim. Theory Appl. *1* (2008), 1–16.

[6] M. X. Goemans: Semidefinite programming in combinatorial optimization. Lectures on Mathematical Programming (ismp97). Math. Programming Ser. B *79* (1997), 1–3, 143–161.

[7] J. A. Dieudonné: Foundations of Modern Analysis, Academic Press, New York 1969.

[8] L. M. Graña Drummond and Y. Peterzil: The central path in smooth convex semidefinite programming. Optimization *51* (2002), 207–233.

[9] E. Halická, E. de Klerk, and C. Roos: On the convergence of the central path in semidefinite optimization. SIAM J. Optim. *12* (2002), 4, 1090–1099.

[10] A. N. Iusem, B. F. Svaiter, and J. X. da Cruz Neto: Central paths, generalized proximal point methods and Cauchy trajectories in Riemannian manifolds. SIAM J. Control Optim. *37* (1999), 2, 566–588.

[11] E. de Klerk: Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications. (Applied Optimization *65*.) Kluwer Academic Publishers, Dordrecht 2002.

[12] M. Kojima, S. Shindoh, and S. Hara: Interior-point methods for the monotone semidefinite linear complementary problem in symmetic matrices. SIAM J. Optim. *7* (1997), 86–125.

[13] A. S. Lewis: Convex Analysis on the Hermitian Matrices. SIAM J. Optim., *6* (1996), 1, 164–177.

[14] A. S. Lewis and H. S. Sendov: Twice differentiable spectral functions. SIAM J. Matrix Anal. Appl. *23* (2001), 2, 368–386.

[15] S. Lojasiewicz: Ensembles Semi-analitiques. Inst. Hautes Études Sci., Bures-sur-Yvette 1965.

[16] P. Petersen: Riemannian Geometry. Springer-Verlag, New York 1998.

[17] A. Seeger: Convex analysis of spectrally defined matrix functions. SIAM J. Optim. *7* (1997), 679–696.

[18] A. Shapiro: On Differentiability of Symmetric Matrix Valued Functions. Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, 2002.

[19] M. Todd: Semidefinite optimization. Acta Numer. *10* (2001), 515–560.

[20] L. Vandenberghe and S. Boyd: Semidefinite programming. SIAM Rev. *38* (1996), 1, 49–95.

Julio López, Departamento de Ingeniería Matemática, FCFM, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile.

e-mail: jclopez@dim.uchile.cl

Héctor Ramírez C., Departamento de Ingeniería Matemática, Centro de Modelamiento Matemático(CNRS UMI 2807), FCFM, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile.

e-mail: hramirez@dim.uchile.cl