J Glob Optim (2009) 45:211–227
DOI 10.1007/s10898-008-9367-x

On the convergence of the entropy-exponential penalty trajectories and generalized proximal point methods in semidefinite optimization

O. P. Ferreira · P. R. Oliveira · R. C. M. Silva

Received: 13 June 2007 / Accepted: 6 October 2008 / Published online: 21 October 2008

Abstract The convergence of primal and dual central paths associated to entropy and exponential functions, respectively, for semidefinite programming problem are studied in this paper. It is proved that the primal path converges to the analytic center of the primal optimal set with respect to the entropy function, the dual path converges to a point in the dual optimal set and the primal-dual path associated to this paths converges to a point in the primal-dual optimal set. As an application, the generalized proximal point method with the Kullback-Leibler distance applied to semidefinite programming problems is considered. The convergence of the primal proximal sequence to the analytic center of the primal optimal set with respect to the entropy function is established and the convergence of a particular weighted dual proximal sequence to a point in the dual optimal set is obtained.

Keywords Generalized proximal point methods · Bregman distances · Central path · Semidefinite programming

1 Introduction

The first purpose of this paper is to analyze the convergence of primal and dual central paths associated to entropy and exponential functions, respectively, for semidefinite programming (SDP) problem. To be more precise, let us consider $I\!R^n$ the n-dimensional Euclidean space, S^n the set of all symmetric $n \times n$ matrices, S^n_+ the cone of positive semidefinite $n \times n$

O. P. Ferreira
IME, Universidade Federal de Goiás, Goiania, GO 74001-970, Brazil
e-mail: orizon@mat.ufg.br

P. R. Oliveira
COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, Brazil
e-mail: poliveir@cos.ufrj.br

R. C. M. Silva (✉)
DM/ICE, Universidade Federal do Amazonas, Manaus, AM 69077-000, Brazil
e-mail: rmesquita@ufam.edu.br

symmetric matrices. Let denote $X \succeq 0$ to mean that $X \in S^n_+$, tr to mean the trace of $n \times n$ matrices and set $X \bullet Y = \operatorname{tr} XY$ for all $X, Y \in S^n$. The primal SDP problem becomes

$$(P) \quad \min\{C \bullet X : \mathcal{A}X = b, X \succeq 0\},$$

where the data consist of $C \in S^n$, $b \in I\!R^m$ and a linear operator $\mathcal{A} : S^n \to I\!R^m$, the primal variable is $X \in S^n$. Adding the entropy penalty function in the objective function of (P), we obtain its penalized version

$$(P_\mu) \quad \min\{C \bullet X + \mu X \bullet \ln(X) : \mathcal{A}X = b, X \succ 0\}, \quad \mu > 0,$$

where $X \succ 0$ means that $X \in S^n_{++}$. The associated dual problem to (P) is

$$(D) \quad \max\{b^T y : \mathcal{A}^* y + S = C, S \succeq 0\},$$

where $\mathcal{A}^* : I\!R^m \to S^n$ denotes the adjoint application associated to $\mathcal{A}$ and $(S, y) \in S^n \times I\!R^m$ are the dual variables. Adding the exponential penalty function in the objective function of (D) we obtain its penalized version

$$(D_\mu) \quad \max\{b^T y - \mu \operatorname{tr} \mathrm{e}^{-S/\mu - I} : \mathcal{A}^* y + S = C\}, \quad \mu > 0.$$

The feasible primal and dual sets are denoted by $\mathcal{F}(P) = \{X \in S^n : \mathcal{A}X = b, X \succeq 0\}$ and $\mathcal{F}(D) = \{(S, y) \in S^n \times I\!R^m : \mathcal{A}^* y + S = C, S \succeq 0\}$, respectively. The interior of primal and dual feasible sets are denoted by $\mathcal{F}^0(P) = \{X \in S^n : \mathcal{A}X = b, X \succ 0\}$ and $\mathcal{F}^0(D) = \{(S, y) \in S^n \times I\!R^m : \mathcal{A}^* y + S = C, S \succ 0\}$ respectively. We also write $\mathcal{F}^*(P)$ and $\mathcal{F}^*(D)$ for the sets of optimal solutions of (P) and (D) respectively.

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

(**A1**) $\mathcal{A} : S^n \to I\!R^m$ is a surjective linear operator;
(**A2**) $\mathcal{F}^0(P) \neq \varnothing$ and $\mathcal{F}^0(D) \neq \varnothing$.

Assumption **A1** is not really crucial for our analysis but it is convenient to ensure that the dual variables S and y are in one-to-one correspondence. Assumption **A2** ensures that both (P) and (D) have optimal solutions, the optimal values of (P) and (D) are equal and the sets of their optimal solutions $\mathcal{F}^*(P)$ and $\mathcal{F}^*(D)$ are bounded (see, for example Todd [28]). It is also important to ensure the existence of the central path. Indeed, our first goal is to prove that assumption **A2** implies that the problems (P_μ) and (D_μ) have unique solution $X(\mu)$ and $(S(\mu), y(\mu))$, respectively. As a consequence, it is easy to see that $X(\mu)$ and $S(\mu)$ satisfy the equality

$$S(\mu) = -\mu \ln(X(\mu)) - \mu I, \quad \mu > 0.$$

The sets of points $\{X(\mu) : \mu > 0\}$ and $\{S(\mu) : \mu > 0\}$ denote the primal and dual central paths associated to entropy and exponential function, respectively.

It is worthwhile to mention that, the only divergence which is also a Bregman distance arises from entropy function, see Iusem and Monteiro [14]. Although Doljansky and Teboulle [8] has been studied Bregman distance associated to entropy function in SDP, the convergence of the central path associated to it was lacking. Most interior point methods "follows" the central path approximately to reach the optimal set and as we will show, the primal proximal sequence rests on central path associated to Bregman distance. Also, we will prove that the primal and dual central paths converge to a solution of (P) and (D), respectively, as μ goes to 0, i.e. the primal path converges to the analytic center of the primal optimal set with respect to the entropy function, the dual path converges to a point in the dual optimal set and the primal-dual path associated to this paths converges to a point in the primal-dual optimal set. So, we can think (P_μ) and (D_μ) as entropy and exponential penalty methods, respectively, for solving SDP problems. Cominetti and San Martín [4] have investigated the asymptotic behavior of the primal and dual trajectories associated to entropy and exponential penalty functions, respectively, in linear program. In particular, they have obtained a characterization of its limit points. More generally, Iusem and Monteiro [14] have given a characterization of the limit point of the dual central path associated to a large class of penalty functions, including exponential penalty function, for linearly constrained convex programming problems. Study on central path associated to convex SDP problems with more general restrictions can be found in Aulender and Héctor Ramírez [2].

Ours second goal is to apply the results obtained about the primal and dual central paths to study the generalized proximal point method to solve the problem (P). This method generates a sequence $\{X_k\} \subset S^n_{++}$ with starting point $X_0 \in \mathcal{F}^0(P)$ according to the iteration

$$X_{k+1} = \arg\min_{X \in S^n_{++}} \{C \bullet X + \lambda_k D(X, X_k) : \mathcal{A}X = b\}, \tag{1}$$

where the sequence $\{\lambda_k\} \subset I\!R_{++}$ satisfies $\sum_{k=0}^{\infty} {\lambda_k}^{-1} = +\infty$ and $D : S^n_{++} \times S^n_{++} \to I\!R$ is the Kullback-Leibler distance (which is also a Bregman distance) defined by

$$D(X, Y) = X \bullet \ln(X) - X \bullet \ln(Y) + \operatorname{tr} Y - \operatorname{tr} X.$$

We will prove that the sequence $\{X_k\}$ is contained in the primal central path. As a consequence, both converge to the same specific optimal solution, namely, the analytic center of the primal optimal set with respect to the generalized distance. This idea has, at first, appeared in Iusem et al. [15], they proved this connection between the central path and the generalized proximal point sequence in some special cases, including linear programming. On the other hand, Doljansky and Teboulle [8] have introduced a generalized proximal method for convex SDP problems and established its convergence properties. Besides, they study the correspondent dual augmented Lagrangian method. Several works dealing with this issue include Aulender and Teboulle [1] and Mosheyev and Zibulevski [22]. So, we are bringing together the ideas of both Iusem et al. [15] and Doljansky and Teboulle [8].

The optimality condition for (1) determines the dual sequence $\{S_k\}$ defined as

$$S_k = \lambda_k(\ln(X_k) - \ln(X_{k+1})), \quad k = 0, 1, 2, \ldots.$$

From the dual sequence $\{S_k\}$ we define the weighed dual sequence $\{\bar{S_k}\}$ constructed as

$$\bar{S_k} = \sum_{j=0}^{k} {\lambda_j}^{-1} \mu_k S_j, \qquad \mu_k = \left(\sum_{j=0}^{k} {\lambda_j}^{-1} \right)^{-1}, \quad k = 0, 1, 2, \ldots.$$

We will prove that the sequence $\{\bar{S_k}\}$ is contained in the dual central path. As a consequence, it converges to a solution. Partial results regarding the behavior of the weighed dual sequence in linear programming have been obtained in several papers including [6,16,24,25,29]. The full convergence of the weighed dual sequence, for Bregman distances including the Kullback–Leibler distance, for linearly constrained convex programming problems has been obtained by Iusem and Monteiro [14].

At this point, it is important to emphasize that, though our convergence analysis of the dual sequence is limited to the case of convex SDP, by no means we advocate the use of the proximal point method for solving SDP problems, be it with the classical quadratic regularization or with Bregman functions. The method is intended rather for the general nonlinear problem, and if we restrict our analysis to the case of a SDP problems it is just because our analytical tools do not allow us to go further. We expect that the results in this paper will be the first step toward a convergence analysis of the dual sequence in the general nonlinear case.

This said, it is worthwhile to make some comments on the proximal method with Bregman barriers:

Firs of all, we point out one advantage of the use of barriers appears in a very important application of the proximal method: when it is applied to the dual of a constrained convex optimization problem, the proximal point method gives rise to primal-dual methods, called Augmented Lagrangian algorithms, whose subproblems are always unconstrained. The subproblems of the Augmented Lagrangian methods resulting from proximal method with Bregman barriers have objective functions which are as smooth as the original constraints, and which can be minimized with Newton's method, see Doljansky and Teboulle [8].

If we compare now the proximal method with Bregman barriers for linearly constrained convex optimization with the so called *interior point* algorithms, we observe that both share unconstrained subproblems (after dealing in an appropriate way with the linear constraint) which can be solved with fast second order methods. The difference lies in the fact that the specific logarithmic barrier, typical of interior point methods, has a property, namely self-concordance, which allows estimates of the number of iterations needed to achieve a given accuracy, ensuring that the running time of the algorithm is bounded by a polynomial function of an adequate measures of the size of the problem. This feature is not shared by Bregman barriers, which in general are not self-concordant. Nevertheless, the proximal point method with Bregman functions, and the resulting smooth Augmented Lagrangians, have proved to be efficient tools in several specific instances, justifying thus the study of its convergence properties, as has been done in the many papers mentioned above. A survey on these augmented Lagrangian methods and its connection with the proximal point method with Bregman distances can be found in Iusem [13].

The organization of our paper is as follows. In Subsect. 1.1, we list some basic notation and terminology used in our presentation. In Sect. 2, we present the well definedness of the primal-dual central path and establish some results about it. In Sect. 3, we describe the proximal point method and establish its connection with the central path. As a consequence, we prove the convergence of the weighed dual sequence. We end the paper by giving in Sect. 4 some remarks and open problems.

1.1 Notation and terminology

The following notations and results of matrix analysis are used throughout our presentation, they can be found in Horn and Johnson [12]. $I\!R^n$ denotes the n-dimensional Euclidean space. $I\!R^n_+ = \{(x_1, \ldots, x_n) \in I\!R^n;\ x_i \geq 0\ \forall\ i = 1, \ldots, n\}$ and $I\!R^n_{++} = \{(x_1, \ldots, x_n) \in I\!R^n;\ x_i > 0\ \forall\ i = 1, \ldots, n\}$ denote nonnegative and positive orthant, respectively. The set of all $n \times m$ matrices is denoted by $I\!R^{n\times m}$. The (i, j)th entry of a matrix $X \in I\!R^{n\times m}$ is denoted by X_{ij} and the jth column is denoted by X_j. The transpose of $X \in I\!R^{n\times m}$ is denoted by X^T. The set of all symmetric $n \times n$ matrices is denoted by S^n. The cone of positive semidefinite (resp., definite) $n \times n$ symmetric matrices is denoted by S^n_+ (resp., S^n_{++}) and ∂S^n_+ denotes the boundary of S^n_+. $X \succeq 0$ means that $X \in S^n_+$ and $X \succ 0$ means that $X \in S^n_{++}$. The trace of a matrix $X \in I\!R^{n\times n}$ is denoted by $\operatorname{tr} X \equiv \sum_{i=1}^{p} X_{ii}$. Given X and Y in $I\!R^{n\times m}$, the inner product between them is defined as $X \bullet Y \equiv \operatorname{tr} X^T Y = \sum_{i=1, j=1}^{n,m} X_{ij} Y_{ij}$. The Frobenius norm of the matrix X is defined as $\|X\| \equiv (X \bullet X)^{1/2}$. The submatrix X_{JK}

of X is the matrix whose entries lie in the rows of X indexed by the set J and the columns indexed by the set K where J and K are two subsets of $\{1, \ldots, n\}$. For square matrices X, X_{JJ} is called a principal submatrix of M, which is denoted simply by X_J.

For a linear operator $\mathcal{A} : S^n \to I\!R^m$, its *adjoint* is the unique linear operator $\mathcal{A}^* : I\!R^m \to S^n$ satisfying $\langle \mathcal{A}X, y\rangle = \langle X, \mathcal{A}^* y\rangle$ for all X and y. The image and null spaces of a linear operator $\mathcal{A}$ will be denoted by Im $(\mathcal{A})$ and Null$(\mathcal{A})$, respectively.

The vector of eigenvalues of a $n \times n$ matrix X will be denoted by $\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))^T$, where the eigenvalues are ordered as $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$.

Lemma 1.1 *For any* $X, Y \in S^n$, $X \bullet Y \leqslant \lambda(X)^T \lambda(Y)$.

Proof See, for example, Dym [9], Lemma 23.16, p. 507. □

2 Primal, dual and primal-dual central paths

In this section we study the convergence of primal and dual central paths associated to the entropy and exponential penalty functions, respectively, for SDP problems. We are going to prove that the central path is well defined, is an analytic curve, bounded and that all its cluster points are solutions of the primal and dual problems, respectively.

The *primal central path* to the Problem (P), with respect to the entropy penalty function $S^n_{++} \ni X \mapsto X \bullet \ln(X)$, is the set of points $\{X(\mu) : \mu > 0\}$ where $X(\mu)$ is defined as

$$X(\mu) = \text{argmin}\{C \bullet X + \mu X \bullet \ln(X) : \mathcal{A}X = b,\ X \succ 0\},\ \mu > 0, \tag{2}$$

i.e. $X(\mu)$ is the solution of the problem (P_μ).

Theorem 2.1 *The primal central path is well defined and is in* $\mathcal{F}^0(P)$.

Proof For each $\mu > 0$ we define $\phi_\mu : S^n_{++} \to I\!R$ by $\phi_\mu(X) = C \bullet X + \mu X \bullet \ln(X)$. The function $\phi_\mu(\cdot)$ is strictly convex and extends continuously to S^n_+ with the convention that $0 \ln 0 = 0$. Its gradient is given by $\nabla\phi_\mu(X) = C + \mu \ln(X) + \mu I$ and $\mathrm{e}^{-(C+\mu I)/\mu} \in S^n_{++}$ is the unique minimizer. Take $\tilde{X} \in \mathcal{F}^0(P)$, thus $\mathcal{L} = \{X \in S^n_+ : \phi_\mu(X) \leqslant \phi_\mu(\tilde{X})\}$ is bounded and nonempty and as $\phi_\mu(\cdot)$ is continuous in S^n_+ we conclude that $\mathcal{L}$ is compact and nonempty. Because $\mathcal{F}(P)$ is closed and nonempty we have that $\mathcal{L} \cap \mathcal{F}(P)$ is also compact and nonempty. Therefore, the strictly convexity of $\phi_\mu(\cdot)$ implies that it has a unique minimizer $X(\mu) \in \mathcal{F}(P)$, which implies that the primal central path is well defined.

It remains to show that $X(\mu) \in \mathcal{F}^0(P)$. Assume by contradiction that $X(\mu) \in \partial\mathcal{F}(P) = \{X \in S^n : AX = b, X \succeq 0, \det X = 0\}$, where $\det X$ denotes the determinant of the matrix X. Define

$$Z_\varepsilon = (1 - \varepsilon)X(\mu) + \varepsilon\tilde{X},$$

where $\varepsilon \in (0, 1)$. Then, as $\tilde{X} \in \mathcal{F}^0(P)$, $X(\mu) \in \partial F(P)$, $\varepsilon \in (0, 1)$ and $F^0(P)$ is convex, we conclude that $Z_\varepsilon \in F^0(P)$ for all $\varepsilon \in (0, 1)$. Now combining definitions of $X(\mu)$ and Z_ε with convexity of $\phi_\mu(\cdot)$ after some algebraic manipulation we obtain

$$0 \leqslant \phi_\mu(Z_\varepsilon) - \phi_\mu(X(\mu)) \leqslant \nabla\phi_\mu(Z_\varepsilon) \bullet (Z_\varepsilon - X(\mu)) = \frac{\varepsilon}{1-\varepsilon}\nabla\phi_\mu(Z_\varepsilon) \bullet (\tilde{X} - Z_\varepsilon),$$

which implies $0 \leqslant \nabla\phi_\mu(Z_\varepsilon) \bullet (\tilde{X} - Z_\varepsilon)$. So, from Lemma 1.1

$$\begin{aligned} 0 \leqslant \nabla\phi_\mu(Z_\varepsilon) \bullet (\tilde{X} - Z_\varepsilon) &= (C + \mu\ln(Z_\varepsilon) + \mu I) \bullet (\tilde{X} - Z_\varepsilon) \\ &= \mu\ln(Z_\varepsilon) \bullet \tilde{X} - C \bullet Z_\varepsilon - \mu\ln(Z_\varepsilon) \bullet Z_\varepsilon - \mu I \bullet Z_\varepsilon \\ &\quad + (C + \mu I) \bullet \tilde{X} \\ &\leqslant \mu \sum_{i=1}^{n} \lambda_i(\tilde{X})\lambda_i(\ln(Z_\varepsilon)) - \phi_\mu(Z_\varepsilon) - \mu I \bullet Z_\varepsilon + (C + \mu I) \bullet \tilde{X}. \end{aligned}$$

Since above inequality holds for all $\varepsilon \in (0, 1)$, letting ε goes to 0 we obtain an absurd. Indeed, as we are under the hypothesis $X(\mu) \in \partial\mathcal{F}(P)$, using the fact that Z_ε goes to $X(\mu)$, $\mu > 0$, $\tilde{X} \succ 0$ and the function ϕ_μ is continuous, the right side of the above inequality goes to $-\infty$. Therefore, this absurd implies the desired result. □

Applying Lagrange theorem to (P_μ) we obtain that $X(\mu)$, as defined in (2), satisfies the system

$$\begin{aligned} \mathcal{A}X = b, \quad X \succ 0, \\ \mathcal{A}^*y + S = C, \\ S + \mu\ln(X) + \mu I = 0, \quad \mu > 0. \end{aligned} \tag{3}$$

for some $(S(\mu), y(\mu)) \in S^n \times I\!R^m$. Note that Theorem 2.1 implies that (3) has unique solution. Moreover, (3) also gives necessary and sufficient condition for optimality in the dual. So,

$$S(\mu) = -\mu\ln(X(\mu)) - \mu I, \quad \mu > 0, \tag{4}$$

is the unique solution of (D_μ), for some $y(\mu) \in I\!R^m$, i.e.,

$$S(\mu) = \operatorname{argmax}\{b^T y - \mu \operatorname{tr} \mathrm{e}^{-S/\mu - I} : \mathcal{A}^*y + S = C\}, \ \ \mu > 0. \tag{5}$$

The *dual central path* associated to (P) is the set of points $\{S(\mu) : \mu > 0\}$, where $S(\mu)$ satisfies (5), or equivalently (4), and the set of points $\{(X(\mu), y(\mu), S(\mu)) : \mu > 0\}$ denotes the *primal-dual central path* which is the unique solution of (4). Now, we are going to prove that the primal-dual central path is an analytic curve. It will follows from a straightforward application of the implicit function theorem that deals with analytic functions, as given, e.g., in Dieudonné [7], Theorem 10.2.4, p. 268.

Theorem 2.2 *The primal-dual central path is an analytic curve contained in* $S^n_{++} \times I\!R^m \times S^n$.

Proof First of all, we introduce the map $\Psi : S^n_{++} \times I\!R^m \times S^n \times I\!R_{++} \to I\!R^m \times S^n \times S^n$ defined by

$$\Psi(X, y, S, \mu) = \big(\mathcal{A}X - b, \mathcal{A}^*y + S - C, \mu\nabla\varphi(X) + S\big),$$

where $\varphi : S^n_{++} \to I\!R$ is given by $\varphi(X) = X \bullet \ln(X)$. Note that $\Psi(X, y, S, \mu) = 0$ is equivalent to the system (3). Since the central path is the unique solution of the system (3) we have that $\Psi(X(\mu), y(\mu), S(\mu), \mu) = 0$, for all $\mu > 0$. So, as Ψ is an analytic function the statement follows from the implicit function theorem by showing that its derivative with respect to (X, y, S) is nonsingular everywhere. To show that the derivative of Ψ is nonsingular it is sufficient to prove that its null-space is the trivial one. Assume that

$$\nabla_{(X,y,S)}\Psi(X, y, S, \mu)(U, v, W) = 0,$$

equivalently,

$$\begin{aligned} \mathcal{A}U &= 0, \\ \mathcal{A}^*v + W &= 0, \\ \mu\nabla^2\varphi(X(\mu))U + W &= 0. \end{aligned} \tag{6}$$

Last equation of (6) implies that $W = -\mu\nabla^2\varphi(X(\mu))U$. Substituting in the second equation of (6) we obtain $\mu U = (\nabla^2\varphi(X(\mu)))^{-1}\mathcal{A}^*v$ and in view of first equation $\mathcal{A}(\nabla^2\varphi(X(\mu)))^{-1}\mathcal{A}^*v = 0$. Finally, as $\nabla^2\varphi(X(\mu))$ is positive definite and $\mathcal{A}$ is surjective we have that $\mathcal{A}(\nabla^2\varphi(X(\mu)))^{-1}\mathcal{A}^*$ is nonsingular, thus latter equality implies that $v = 0$ and consequently $W = U = 0$. Therefore, the derivative of Ψ is nonsingular and the statement follows. □

The Theorem 2.1 implies that the primal central path is well defined and is in $\mathcal{F}^0(P)$. So, for all $\mu > 0$, we have from (3) that

$$\mu\ln(X(\mu)) + \mu I = -C + \mathcal{A}^*y(\mu), \tag{7}$$

for some $y(\mu) \in \mathbb{R}^m$.

Proposition 2.1 *The the following statements hold:*

(i) *the function* $0 < \mu \mapsto X(\mu) \bullet \ln(X(\mu))$ *is non-increasing*;
(ii) *the set* $\{X(\mu) : 0 < \mu \leq \bar{\mu}\}$ *is bounded, for each* $\bar{\mu} > 0$;
(iii) *all cluster points of the primal central path are solutions of the Problem* (P).

Proof To simplify the notations let $\varphi : S^n_{++} \to \mathbb{R}$ be given by $\varphi(X) = X \bullet \ln(X)$. So, (7) is equivalent to

$$\mu\nabla\varphi(X(\mu)) = -C + \mathcal{A}^*y(\mu). \tag{8}$$

Take $\mu_1, \mu_2 > 0$ with $\mu_1 < \mu_2$. Since φ is convex, see the Appendix, and $X(\mu_1) - X(\mu_2) \in$ Null $\mathcal{A}$ we have from (8) that

$$\mu_1(\varphi(X(\mu_1) - \varphi(X(\mu_2)) \leqslant \mu_1\nabla\varphi(X(\mu_1)) \bullet (X(\mu_1) - X(\mu_2)) = -C \bullet (X(\mu_1) - X(\mu_2))$$

and

$$\mu_2(\varphi(X(\mu_2) - \varphi(X(\mu_1)) \leqslant \mu_2\nabla\varphi(X(\mu_2)) \bullet (X(\mu_2) - X(\mu_1)) = -C \bullet (X(\mu_2) - X(\mu_1)).$$

Now, combining the latter two equations we obtain that $(\mu_1 - \mu_2)(\varphi(X(\mu_1) - \varphi(X(\mu_2)) \leqslant 0$ and as $\mu_1 < \mu_2$ we have that $\varphi(X(\mu_2)) \leqslant \varphi(X(\mu_1))$. So, the statement (i) is established.

Now, fix $\bar{\mu} > 0$. Similar argument used to prove item (i) implies that

$$\mu(\varphi(X(\mu) - \varphi(X(\bar{\mu})) \leqslant -C \bullet (X(\mu) - X(\bar{\mu})),$$

for all $0 < \mu < \bar{\mu}$. From item (i) we have that $0 \leqslant \varphi(X(\mu)) - \varphi(X(\bar{\mu}))$, for all $0<\mu<\bar{\mu}$, then above equation implies that $C \bullet X(\mu) \leqslant C \bullet X(\bar{\mu})$, for all $0 < \mu < \bar{\mu}$. So,

$$\{X(\mu) : 0 < \mu < \bar{\mu}\} \subset \{X \in \mathcal{F}(P) : C \bullet X \leqslant C \bullet X(\bar{\mu})\}.$$

Since the function $\mathcal{F}(P) \ni X \mapsto C \bullet X$ is convex and has a sub-level $\mathcal{F}^*(P)$ non-empty and bounded, all its sub-level are bounded. So, the sub-level set $\{X \in \mathcal{F}(P) : C \bullet X \leqslant C \bullet X(\bar{\mu})\}$ is bounded. Therefore, the statement (ii) follows from the last inclusion.

Let $\bar{X}$ be a cluster point of $\{X(\mu) : \mu > 0\}$. First, note that $\mathcal{A}\bar{X} = b$ and $\bar{X} \succeq 0$, i.e., $\bar{X} \in \mathcal{F}(P)$. Let $\{\mu_k\}$ be a sequence of positive numbers such that $\lim_{k\to+\infty}\mu_k = 0$ and

$\lim_{k\to+\infty} X(\mu_k) = \bar{X}$. Take X^* a solution of the Problem (P) and $X \in \mathcal{F}^0(P)$. For $\epsilon > 0$, define

$$Y(\epsilon) = (1-\epsilon)X^* + \epsilon X.$$

Due the fact that $X^* \in \partial\mathcal{F}^0(P)$, $X \in \mathcal{F}^0(P)$ and $\mathcal{F}^0(P)$ is convex we have $Y(\epsilon) \in \mathcal{F}^0(P)$, for $\epsilon \in (0, 1]$. From (2) we have

$$C \bullet X(\mu_k) + \mu_k\varphi(X(\mu_k)) \leqslant C \bullet Y(\epsilon) + \mu_k\varphi(Y(\epsilon)),$$

or,

$$\mu_k(\varphi(X(\mu_k)) - \varphi(Y(\epsilon))) \leqslant C \bullet (Y(\epsilon) - X(\mu_k)).$$

Now, since φ is convex and $Y(\epsilon) \in \mathcal{F}^0(P)$, it easy to conclude from above inequality that

$$\mu_k\nabla\varphi(Y(\epsilon)) \bullet (X(\mu_k) - Y(\epsilon)) \leqslant C \bullet (Y(\epsilon) - X(\mu_k)).$$

Thus, taking limits in the latter inequality as k goes to $+\infty$ we obtain $0 \leqslant C \bullet (Y(\epsilon) - \bar{X})$. In this inequality, if ϵ tends to 0, it gives

$$0 \leqslant C \bullet (X^* - \bar{X}), \quad \text{or equivalently,} \quad C \bullet \bar{X} \leqslant C \bullet X^*.$$

Therefore, as X^* is a solution of the Problem (P) and $\bar{X} \in \mathcal{F}(P)$, we have from above equation that $\bar{X}$ is also solution of the Problem (P) and the proof of the statement *(iii)* is concluded. □

Theorem 2.3 *Let $X^c \in S^n_+$ be the analytic center of $\mathcal{F}^*(P)$, i.e., the unique point satisfying*

$$X^c = \operatorname{argmin}\{X \bullet \ln(X) : X \in \mathcal{F}^*(P)\}. \tag{9}$$

Then $\lim_{\mu\to 0} X(\mu) = X^c$.

Proof To simplify the notations let $\varphi : S^n_{++} \to I\!R$ the function defined in the proof of the above proposition. Using the convention $0\ln 0 = 0$, it is not hard to see that φ is continuous in S^n_+. Take $\bar{X}$ a cluster point of the primal central path and a sequence of positive numbers $\{\mu_k\}$ such that $\lim_{k\to+\infty}\mu_k = 0$ and $\lim_{k\to+\infty} X(\mu_k) = \bar{X}$. Note that, from Proposition 2.1(iii), implies that $\bar{X} \in \mathcal{F}^*(P)$. So, it is feasible for the problem in (9). Now, we are going to prove that $\bar{X}$ is a solution to the problem in (9). From (3) we have $C + \mu_k\nabla\varphi(X(\mu_k)) = \mathcal{A}^*y(\mu_k)$, for some $y(\mu_k) \in I\!R^m$. So,

$$\mu_k\nabla\varphi(X(\mu_k)) \bullet (X - X(\mu_k)) = (\mathcal{A}^*y(\mu_k) - C) \bullet (X - X(\mu_k)),$$

for all $X \in \mathcal{F}^*(P)$. Using the convexity of φ and the fact that $X - X(\mu_k) \in \text{Null}(\mathcal{A})$ the latter equation becomes

$$\mu_k(\varphi(X(\mu_k)) - \varphi(X)) \leqslant C \bullet X - C \bullet X(\mu_k).$$

Because $X \in \mathcal{F}^*(P)$ and $\mu_k > 0$, it follows from the latter inequality that $\varphi(X(\mu_k)) \leqslant \varphi(X)$. Now, as φ is continuous we can take limits, as k goes to $+\infty$, in this inequality to conclude that $\varphi(\bar{X}) \leqslant \varphi(X)$, i.e., $\bar{X} \bullet \ln(\bar{X}) \leqslant X \bullet \ln(X)$, for all $X \in \mathcal{F}^*(P)$. Thus, any cluster point of the primal central path satisfies (9). Now, since $\mathcal{F}^*(P)$ is compact and the function $S^n_{++} \ni X \mapsto X \bullet \ln(X)$ is strictly convex, see the Appendix, the problem in (9) has unique solution X^c. So, the primal central path has unique cluster point. Therefore, the primal central path converges to X^c and the theorem is proved. □

In the next proposition our goal is to state and prove that the dual central path is bounded, as μ goes to 0, and all its cluster points are solutions of the problem (D).

Proposition 2.2 *The following statements hold*:

(i) *the set* $\{S(\mu) : 0 < \mu \leq \bar{\mu}\}$ *is bounded, for each* $\bar{\mu} > 0$;
(ii) *all cluster points of the dual central path are solutions of the problem* (D).

Proof To prove item (i), let X^0 and S^0 be strictly feasible for (P) and (D), respectively. Orthogonality relation implies that

$$(X(\mu) - X^0) \bullet (S(\mu) - S^0) = 0, \quad \mu > 0.$$

Since $X(\mu) \succ 0$ and $S^0 \succ 0$, simple algebraic manipulations in above equation yield $X^0 \bullet S(\mu) \leq X(\mu) \bullet S(\mu) + X^0 \bullet S^0$. Now, combining this inequality with (4) we obtain

$$X^0 \bullet S(\mu) \leq -\mu X(\mu) \bullet \ln(X(\mu)) - \mu \,\mathrm{tr}(X(\mu)) + X^0 \bullet S^0.$$

Then, as $X(\mu) \succ 0$, use Proposition 2.1(i) and $\mu > 0$ in the last inequality to get

$$X^0 \bullet S(\mu) \leq -\mu X(\bar{\mu}) \bullet \ln(X(\bar{\mu})) + X^0 \bullet S^0, \tag{10}$$

for all $0 < \mu \leq \bar{\mu}$. We remark that if $S(\mu)$ is positive semidefinite for all $\mu \in (0, \bar{\mu}]$ then we are done, but we cannot ensure it, so, we have to go further. First, as $X(\mu) \in S^n_{++}$ there exists an orthogonal matrix $Q(\mu)$ such that

$$X(\mu) = Q^T(\mu)\Lambda(\mu)Q(\mu)$$

where $\Lambda(\mu) \in S^n_{++}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $X(\mu)$. From (4), we obtain

$$S(\mu) = Q^T(\mu)(-\mu(\ln(\Lambda(\mu)) + I))Q(\mu), \tag{11}$$

where $-\mu(\ln(\Lambda(\mu)) + I) = \mathrm{diag}(-\mu(\ln(\lambda_1(X(\mu)) + 1), \ldots, -\mu(\ln(\lambda_n(X(\mu)) + 1))$. Let $X^c \in S^n_+$ be the analytic center of $\mathcal{F}^*(P)$ and let

$$B := \left\{j : \lambda_j(X^c) > 0\right\} \quad \text{and} \quad \tilde{B} := \{j : \lambda_j(X^c) = 0\}.$$

From Theorem 2.3 we have that $X^c = \lim_{\mu \to 0} X(\mu)$. So, it is easy to show that

$$\lim_{\mu \to 0} -\mu(\ln(\Lambda(\mu)_B) + I_B) = 0, \tag{12}$$

and there exists $\tilde{\mu} > 0$ such that for all $0 < \mu < \tilde{\mu} \leq \bar{\mu}$ there holds

$$-\mu(\ln(\Lambda(\mu)_{\tilde{B}}) + I_{\tilde{B}}) \succ 0. \tag{13}$$

Combining Eqs. (11), (12) and (13) it simple to conclude that for all $0 < \mu < \tilde{\mu} \leq \bar{\mu}$

$$S(\mu) = Q^T(\mu)\left(-\mu(\ln(\Lambda(\mu)) + I)\right) Q(\mu) \succeq 0. \tag{14}$$

Set $\Sigma(\mu) := -\mu(\ln(\Lambda(\mu)) + I)$. Thus, (12) implies that $\Sigma(\mu)_B$ goes to the null matrix, as μ goes to 0, and (13) implies that $\Sigma(\mu)_{\tilde{B}}$ is positive definite. It follows from (11) that $S(\mu) = Q^T(\mu)\Sigma(\mu)Q(\mu)$, hence

$$X^0 \bullet S(\mu) = (Q(\mu)X^0Q^T(\mu))_B \bullet \Sigma(\mu)_B + (Q(\mu)X^0Q^T(\mu))_{\tilde{B}} \bullet \Sigma(\mu)_{\tilde{B}}.$$

As $\Sigma(\mu)_{\tilde{B}} \succ 0$ and $(Q^T(\mu)X^0Q(\mu))_{\tilde{B}} \succ 0$ we have that

$$\lambda_{\min}((Q(\mu)X^0Q^T(\mu))_{\tilde{B}})\|\Sigma(\mu)_{\tilde{B}}\| \leqslant (Q(\mu)X^0Q^T(\mu))_{\tilde{B}} \bullet \Sigma(\mu)_{\tilde{B}}.$$

Thus, combining the above equation with $\lambda_{\min}(X^0) \leq \lambda_{\min}((Q(\mu)X^0Q^T(\mu))_{\tilde{B}})$ (see in Horn and Johnson [12], Theorem 4.3.15, p. 189) we obtain

$$\|\Sigma(\mu)_{\tilde{B}}\| \leq \frac{X^0 \bullet S(\mu) - (Q(\mu)X^0Q^T(\mu))_B \bullet \Sigma(\mu)_B}{\lambda_{\min}(X^0)}.$$

Because $\Sigma(\mu)_B$ goes to the null matrix as μ goes to 0 and $Q(\mu)$ is an orthogonal matrix, last inequality together with (10) imply that $\Sigma(\mu)_{\tilde{B}}$ is bounded as μ goes to 0. Therefore, as

$$\|S(\mu)\|^2 = \|\Sigma(\mu)_B\|^2 + \|\Sigma(\mu)_{\tilde{B}}\|^2,$$

$\Sigma(\mu)_B$ and $\Sigma(\mu)_{\tilde{B}}$ are bounded as μ goes to 0, we conclude that $\{S(\mu) : 0 < \mu \leq \bar{\mu}\}$ is bounded. So, the statement (i) is established.

For proving item (ii), let $\bar{S}$ be a cluster point of the dual central path. Note that it is sufficient to show that

$$\mathcal{A}^*\bar{y} + \bar{S} = C, \quad X^*\bar{S} = 0, \quad \bar{S} \succeq 0, \tag{15}$$

for some $\bar{y} \in I\!R^m$ and $X^* \in \mathcal{F}^*(P)$. Since the dual central path satisfies the second equation in (3) we just have to show that $\bar{S}$ satisfies the last two equations in (15). Let $\{\mu_k\}$ be a sequence such that $\lim_{k\to+\infty} \mu_k = 0$ and $\bar{S} = \lim_{k\to+\infty} S(\mu_k)$. First note that $X(\mu)\ln(X(\mu))$ is bounded as μ goes to 0 and from Theorem 2.3 we have that $X^c = \lim_{\mu\to 0} X(\mu)$. Thus it follows from (4) that

$$X^c\bar{S} = \lim_{k\to+\infty} X(\mu_k)S(\mu_k) = -\lim_{k\to+\infty} (\mu_k X(\mu_k)\ln(X(\mu_k)) + \mu_k X(\mu_k)) = 0.$$

As $X^c \in \mathcal{F}^*(P)$ the second relation in (15) holds. Finally, it remains to show the third relation in (15). Using the same notation to prove item (i), we have from (11) that

$$\bar{S} = \lim_{k\to+\infty} S(\mu_k) = \lim_{k\to+\infty} (Q^T(\mu_k)(-\mu_k(\ln(\Lambda(\mu_k)) + I)Q(\mu_k)). \tag{16}$$

Because $Q(\mu_k)$ is orthogonal for all k, we can assume without loss of generality that $\lim_{k\to+\infty} Q(\mu_k) = Q$. Since $-\mu_k(\ln(\Lambda(\mu_k)) + I)$ converges as k goes to $+\infty$, thus we conclude from (14) that

$$\bar{S}_B = \lim_{k\to+\infty} S(\mu_k)_B = 0, \quad \bar{S}_{\tilde{B}} = \lim_{k\to+\infty} S(\mu_k)_{\tilde{B}} \succeq 0.$$

Hence, from last equation and (16) we have that $\bar{S} \succeq 0$. Therefore, the third relation in (15) is proved and statement (ii) is established. □

The Proposition 2.2 extends to SDP the Proposition 3.1 of Cominetti and San Martín [4]. Now, we are going to prove the convergence of the primal-dual central path using result of the theory of semianalytic sets due to Lojasiewicz [21]. It is worth pointing out that in our proof the key arguments are the same of Halická et al. [10].

Definition 2.1 A subset $W \subseteq I\!R^n$ is called a semianalytic set if it is described by a finite union of sets

$$\{x \in I\!R^n : f_1(x) = 0, \ldots, f_m(x) = 0, g_1(x) > 0, \ldots, g_l(x) > 0\},$$

where $f_1, \ldots, f_m, g_1, \ldots, g_l$ are real analytic functions.

Lemma 2.1 (*Curve selection lemma*) *Let* $W \subseteq I\!R^n$ *be a semianalytic set. If* $0 \in \overline{W} - W$, *where* $\overline{W}$ *is the closure of* W, *then there exists some* $\varepsilon > 0$ *and a real analytic curve* $\alpha : [0, \varepsilon) \to \overline{W}$ *with* $\alpha(0) = 0$ *and* $\alpha(t) \in W$ *for* $t \in (0, \varepsilon)$.

Proof See, Lojasiewicz [21], Proposition 2, p. 103. □

A particular version of this lemma was used by Kojima et al. [18] and Halická et al. [10], to prove the convergence of the central path, in a different setting. Other applications of this lemma in mathematical programming can be found in Bolte et al. [3] and Papa Quiroz and Roberto Oliveira [23]. This lemma has been used in other contexts, see for example Kurdyka et al. [20] and its references. For a more general version of this lemma, see Shiota [26] , property I.2.1.7 on p. 42.

Lemma 2.2 *Let $f : I \to I\!R$ be an analytic function such that $f(x) = 0$ for all $x \in U$, where $U \subset I$, is a set with a cluster point $x_0 \in I$. Then $f(x) = 0$ for all $x \in I$.*

Proof See, for example, Krantz and Parks [19], Corollary 1.2.6, p. 14. □

Theorem 2.4 *The primal-dual central path converges.*

Proof From Propositions 2.1 and 2.2 we have that primal-dual central path is bounded. Take (X^*, y^*, S^*) a cluster point of the primal-dual central path and let $\{\mu_k\}$ be a sequence of positive numbers such that $\lim_{k\to+\infty} \mu_k = 0$ and $\lim_{k\to+\infty}(X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$. Let W be a semianalytic set defined by

$$W = \left\{ (\bar{X}, \bar{y}, \bar{S}, \mu) \in S^n_{++} \times I\!R^m \times S^n \times I\!R_{++} : \begin{array}{r} \mathcal{A}\bar{X} = 0 \\ \mathcal{A}^*\bar{y} + \bar{S} = 0 \\ (\bar{S} + S^*) + \mu \ln(\bar{X} + X^*) + \mu I = 0 \\ \bar{X} + X^* \succ 0 \\ \mu > 0 \end{array} \right\}.$$

Note that the zero element belongs to $\overline{W} - W$. Indeed, consider the sequence

$$(\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) := (X(\mu_k) - X^*,\ y(\mu_k) - y^*,\ S(\mu_k) - S^*,\ \mu_k).$$

Obviously, $(\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) \in W$. Thus, as $\lim_{k\to+\infty}(X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$ we have that

$$\lim_{k\to+\infty} (\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) = (0_{n\times n}, 0_m, 0_{n\times n}, 0).$$

So, Lemma 2.1 implies the existence of an $\varepsilon > 0$ and an analytic function $\alpha : [0, \varepsilon) \mapsto \overline{W}$ with $\alpha(0) = 0$ and $\alpha(t) = (\bar{X}(t), \bar{y}(t), \bar{S}(t), \mu(t)) \in W$ for $t \in (0, \varepsilon)$. Now, since the system that defines the central path has a unique solution, it easy to see that the system that defines W also has a unique solution given by

$$\bar{X}(t) = X(\mu(t)) - X^*, \quad \bar{y}(t) = y(\mu(t)) - y^*, \quad \bar{S}(t) = S(\mu(t)) - S^*, \quad \mu(t) > 0,$$

for $t>0$. As $\mu(0) = 0$, $\lim_{k\to+\infty} \mu_k = 0$ and $\lim_{k\to+\infty}(X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$ above equalities imply

$$\lim_{t\downarrow 0} X(\mu(t)) = X^*, \quad \lim_{t\downarrow 0} y(\mu(t)) = y^*, \quad \lim_{t\downarrow 0} S(\mu(t)) = S^*, \quad \lim_{t\downarrow 0} \mu(t) = 0.$$

Since $\mu : [0, \varepsilon) \mapsto I\!R$ is a real analytic function satisfying $\mu(t) > 0$ on $(0, \varepsilon)$ and $\mu(0) = 0$, we must have that $\mu'(0) \geqslant 0$. Thus, we have two possibilities:

(i) $\mu'(0) > 0$;
(ii) $\mu'(0) = 0$.

If $\mu'(0) > 0$, there exists an interval $(0, \delta)$ where $\mu'(t) > 0$. In this case, μ is increasing which implies that it is invertible in this interval. Now, if $\mu'(0) = 0$ we claim that there exists an interval $(0, \delta)$ where $\mu'(t) > 0$. Otherwise, there exists a sequence $\{t_k\}$ in $(0, \varepsilon)$ such that $\lim_{k\to\infty} t_k = 0$ and $\mu'(t_k) = 0$. As μ is a analytic we obtain from Lemma 2.2 that $\mu'(t) = 0$ for all $t \in [0, \varepsilon)$ or equivalently μ is constant in $[0, \varepsilon)$. Because, $\mu(0) = 0$ we conclude $\mu(t) = 0$ for all $t \in [0, \varepsilon)$, but this is an absurd. So, the claim is established. As a consequence μ is invertible in this interval. Therefore, in any of the two possibilities, there exists the inverse function $\mu^{-1} : [0, \mu(\delta)) \to [0, \delta)$ with $\mu^{-1}(0) = 0$. This implies that

$$\lim_{s\to 0^+} X(s) = \lim_{s\to 0^+} X(\mu(\mu^{-1}(s))) = \lim_{s\to 0^+} \bar{X}(\mu^{-1}(s)) + X^* = X^*.$$

Similarly, $\lim_{s\to 0^+} y(s) = y^*$, $\lim_{s\to 0^+} S(s) = S^*$ and the result follows. □

Cominetti and San Martín [4] have obtained the characterization of the limit point of the primal-dual central path associated to the entropy-exponential penalty in linear programming. The above theorem guarantees the convergence of the primal-dual central path to SDP. In Theorem 2.3 above the characterization of the limit point is obtained only with respect to primal central path. The characterization of the limit point for the dual central path is an open problem.

3 Central paths and generalized proximal point methods

In this section we study a generalized proximal point method to solve the problem (P) and present some convergence results for it. In particular, we are going to prove that the primal and weighed dual sequences are contained in the primal and dual central paths, respectively. Consequently, both converge. It is worthwhile to mention that our goal in this section is to bring to SDP context the ideas of Iusem et al. [15] and Iusem and Monteiro [14].

We begin with the *Kullback–Leibler distance* $D : S^n_{++} \times S^n_{++} \to I\!R$ given by

$$D(X, Y) = X \bullet \ln(X) - X \bullet \ln(Y) + \operatorname{tr} Y - \operatorname{tr} X.$$

The last function can also be seen as a Bregman distance associated to the entropy penalty function $\varphi(X) = X \bullet \ln(X)$ considered in Doljansky and Teboulle [8].

Remark 3.1 For each fixed $Y \in S^n_{++}$ it is to easy see that $D(., Y)$ is C^2, strictly convex and can be continuously extended to S^n_+ with the convention $0 \ln 0 = 0$. Also, note that $D(., Y)$ differ from the entropy function by affine term. So, is not had to see that all results from Sect. 2 can be proved to this penalty function by following the same pattern used to prove them for entropy penalty function.

The primal central path to the Problem (P), with respect to the function $D(., X_0)$, is the set of points $\{X(\mu) : \mu > 0\}$, where $X(\mu)$ is defined as

$$X(\mu) = \operatorname{argmin}_{X\in S^n_{++}} \{C \bullet X + \mu D(X, X_0) : \mathcal{A}X = b\}, \ \mu > 0. \tag{17}$$

Theorem 3.1 *The following statements hold*:

(i) *the primal central path with respect to the function $D(., X_0)$ is well defined and is in $\mathcal{F}^0(P)$;*

(i) *if* $\hat{X} \in S^n_+$ *is the analytic center of* $\mathcal{F}^*(P)$*, i.e., the unique point satisfying*

$$\hat{X} = \operatorname{argmin}\{D(X, X_0) : X \in \mathcal{F}^*(P)\},$$

then $\lim_{\mu \to 0} X(\mu) = \hat{X}$.

Proof (*i*) and (*ii*) follow from Remark 3.1 and similar arguments used to prove Theorems 2.1 and 2.3, respectively. □

Theorem 3.1 (*i*) guarantees that the primal central path to the Problem (P), with respect to the function $D(., X_0)$, is well defined and is in $\mathcal{F}^0(P)$. So, for all $\mu > 0$, we have from (17) that

$$C + \mu(\ln(X(\mu) - \ln(X_0)) = \mathcal{A}^* y(\mu),$$

for some $y(\mu) \in I\!R^m$.

The dual central path associated to the Problem (P), with respect to the function $D(., X_0)$, is the set of points $\{S(\mu) : \mu > 0\}$, where $S(\mu)$ satisfies

$$S(\mu) = -\mu(\ln(X(\mu)) - \ln(X_0)), \quad \mu > 0,$$

or equivalently, $(S(\mu), y(\mu))$ is the unique solution of the optimization problem

$$\max \left\{ b^T y - \mu \operatorname{tr} e^{-S/\mu + \ln(X_0)} : \mathcal{A}^* y + S = C \right\}, \ \mu > 0.$$

The set $\{(X(\mu), y(\mu), S(\mu)) : \mu > 0\}$ denotes the primal-dual central path with respect to the function $D(., X_0)$, and it is the unique solution of the following system of nonlinear equations

$$\begin{aligned} \mathcal{A}X = b, \quad X \succ 0,\\ \mathcal{A}^* y + S = C,\\ S + \mu \ln(X) - \mu \ln(X_0) = 0, \quad \mu > 0. \end{aligned} \tag{18}$$

Remark 3.2 Similarly to the proof of the Theorem 2.2 we can prove that the primal-dual central path, with respect to the function $D(., X_0)$, is an analytic curve contained in $S^n_{++} \times I\!R^m \times S^n$.

Theorem 3.2 *The primal-dual central path with respect to the function* $D(., X_0)$ *converges.*

Proof The proof follows similar arguments used to prove the Theorem 2.4. □

The proximal point method with the generalized distance D, for solving the problem (P), generates a sequence $\{X_k\} \subset S^n_{++}$ with starting point $X_0 \in \mathcal{F}^0(P)$ and

$$X_{k+1} = \arg \min_{X \in S^n_{++}} \{C \bullet X + \lambda_k D(X, X_k) : \mathcal{A}X = b\}, \tag{19}$$

where the sequence $\{\lambda_k\} \subset I\!R_{++}$ satisfies

$$\sum_{k=0}^{\infty} {\lambda_k}^{-1} = +\infty. \tag{20}$$

From now on we refer to the above sequence $\{X_k\}$ as primal proximal point sequence with respect to D, associated to $\{\lambda_k\}$ and starting point X_0. Remark 3.1 and a similar argument used in the proof of Theorem 2.1 allow to prove the well-definedness of the proximal point sequence. Moreover, (19) implies that $\{X_k\}$ satisfies

$$C + \lambda_k(\ln(X_{k+1}) - \ln(X_k)) = \mathcal{A}^* z_k, \tag{21}$$

for some sequence $\{z_k\}$ in $I\!R^m$ and $k = 0, 1, 2, \ldots$. Also, the optimality condition for (19) determines the dual sequence $\{S_k\}$ defined as

$$S_k = \lambda_k \left(\ln(X_k) - \ln(X_{k+1})\right), \quad k = 0, 1, 2, \ldots. \tag{22}$$

From the dual sequence $\{S_k\}$ we define the weighed dual proximal sequence $\{\bar{S}_k\}$ constructed as

$$\bar{S}_k = \sum_{j=0}^{k} \lambda_j{}^{-1} \mu_k S_j, \tag{23}$$

where

$$\mu_k = \left(\sum_{j=0}^{k} \lambda_j{}^{-1} \right)^{-1},$$

for $k = 0, 1, 2, \ldots$.

Theorem 3.3 *Let $\{X(\mu) : \mu > 0\}$ and $\{S(\mu) : \mu > 0\}$ be the primal and dual central paths associated to $D(., X_0)$, respectively. Suppose given a sequence $\{\lambda_k\} \subset I\!R_{++}$ satisfying (20), and the sequence $\{\mu_k\}$ defined as*

$$\mu_k = \left(\sum_{j=0}^{k} \lambda_j{}^{-1} \right)^{-1}, \quad \textit{for} \quad k = 0, 1, 2 \ldots. \tag{24}$$

Then $X_{k+1} = X(\mu_k)$ and $\bar{S}_k = S(\mu_k)$ for $k = 0, 1, 2 \ldots$, where $\{X_k\}$ and $\{\bar{S}_k\}$ are the primal and weighed dual sequences associated to $\{\lambda_k\}$, respectively. As a consequence,

$$\lim_{k \to +\infty} (X_k, \bar{S}_k) = (X^*, S^*),$$

where $(X^*, S^*) = \lim_{\mu \to 0}(X(\mu), S(\mu))$.

Proof Let $\{X_k\}$ and $\{S_k\}$ be the primal and dual sequences, respectively. Now, From (19), (21) and (22) we have that X_k and S_k satisfies

$$\begin{aligned} \mathcal{A}X_{k+1} &= b, \quad X_{k+1} \succ 0, \\ \mathcal{A}^* z_k + S_k &= C, \\ S_k &= \lambda_k(\ln(X_k) - \ln(X_{k+1})), \quad \lambda_k > 0 \end{aligned}$$

for some sequence $\{z_k\}$ in $I\!R^m$ and $k = 0, 1, 2, \ldots$. From the last equation of the previous system, it follows that $\sum_{j=0}^{k}(1/\lambda_j)S_j = (\ln(X_0) - \ln(X_{k+1}))$. Last expression together with (23) and (24) imply that

$$\bar{S}_k = -\mu_k(\ln(X_{k+1}) - \ln(X_0)).$$

So, it is easy to conclude that X_k and $\bar{S}_k$ satisfies

$$\begin{aligned} \mathcal{A}X_{k+1} &= b, \quad X_{k+1} \succ 0, \\ \mathcal{A}^* \bar{y}_k + \bar{S}_k &= C, \\ \bar{S}_k + \mu_k \ln(X_{k+1}) - \mu_k \ln(X_0) &= 0, \quad \mu_k > 0. \end{aligned}$$

for $\bar{y}_k = \mu_k \sum_{j=0}^{k}(1/\lambda_j)z_j$, $k = 0, 1, 2, \ldots$. So, the previous system and (18) imply that $X_{k+1} = X(\mu_k)$, $\bar{y}_k = y(\mu_k)$ and $\bar{S}_k = S(\mu_k)$. As $\{\lambda_k\}$ satisfies (20) we have that $\lim_{k\to+\infty} \mu_k = 0$. Now, use the fact that $\lim_{\mu\to 0}(X(\mu), S(\mu)) = (X^*, S^*)$ to conclude that $\lim_{k\to+\infty}(X_k, \bar{S}_k) = (X^*, S^*)$, and the proof is complete. □

With similar arguments used in the proof of Theorem 3 of Iusem et al. [15] we can prove that, for each positive decreasing sequence $\{\mu_k\}$, there exists a sequence $\{\lambda_k\} \subset I\!R_{++}$ satisfying (20) such that the primal sequence $\{X_k\}$ and the weighed dual sequence $\{\bar{S}_k\}$ associated to it satisfy $X_{k+1} = X(\mu_k)$ and $\bar{S}_k = S(\mu_k)$, where $\{X(\mu) : \mu > 0\}$ and $\{S(\mu) : \mu > 0\}$ are the primal and dual central paths associated to $D(., X_0)$, respectively.

4 Final remarks

In this paper we have studied the convergence of primal and dual central paths associated to the entropy and exponential functions, respectively, for SDP problems. Cominetti and San Martín [4] have investigated the asymptotic behavior of the primal and dual trajectories associated to the entropy and exponential penalty functions, respectively, in linear program. In particular, they have obtained a characterization of its limit points. More generally, Iusem and Monteiro [14] have given a characterization of the limit of the dual central path associated to a large class of penalty functions, including exponential penalty function, for linear constrained convex programming problems. Partial characterizations of the limit point of the central path with respect to the log-barrier function for SDP problems have been obtained by Sporre and Forsgren [27], Halická et al. [11], and da Cruz Neto et al. [5]. For more general functions, including the exponential penalty function, the characterization of the limit point of the dual central path associated to them is an open problem.

We do not have considered the behavior of the dual sequence in the generalized proximal method, its convergence with respect to entropy function is still a open problem even in linear programming. Since, we do not know how to prove the convergence of the dual sequence to the solution, we consider the weighted dual sequence and prove its convergence to the solution of the dual semidefinite problem. The proof of this fact is in Theorem 3.3, since the weighted dual sequence in on dual central path.

The characterization of the limit point of the primal sequence is obtained by combining Theorems 3.2 and 3.3, because primal sequence in on dual central path and both converges to $X^c \in S^n_+$ the analytic center of $\mathcal{F}^*(P)$. This technique does not works to study the dual sequence.

Appendix

Let $f : (0, +\infty) \to I\!R$ a analytical function having the following expansion by power series

$$f(x) = \sum_{i=0}^{\infty} a_i\, x^i. \tag{25}$$

So, we define the function of matrix $\varphi : S^n_{++} \to I\!R$ as follows $\varphi(X) = \operatorname{tr} f(X)$, or equivalently,

$$\varphi(X) = \sum_{i=0}^{\infty} a_i \ \operatorname{tr} X^i. \tag{26}$$

Therefore, the gradient of φ is given by $\nabla\varphi(X) = f'(X)$, i.e.,

$$\nabla\varphi(X) = \sum_{i=1}^{\infty} a_i \, \nabla \operatorname{tr} X^i = \sum_{i=1}^{\infty} i \, a_i X_i^{i-1}. \tag{27}$$

Indeed, as X is a symmetric matrix it easy to see that $\nabla \operatorname{tr} X^i = i X^{i-1}$, for all $i = 0, 1, 2, \ldots$. Hence, because φ is an analytical function, taking derivative in (26) we conclude that (27) holds.

Letting $f(x) = x \ln(x)$, in this case, $\varphi(X) = X \bullet \ln(X)$ and we obtain that

$$\nabla\varphi(X) = \ln(X) + I.$$

Let $X, Y \in S^n_{++}$ with $X \neq Y$. Using last equality we have, after simples manipulation, that

$$\varphi(X) - \varphi(Y) - \nabla\varphi(Y) \bullet (X - Y) = X \bullet \ln(X) - X \bullet \ln(Y) + I \bullet Y - I \bullet X. \tag{28}$$

On the other hand, as $X, Y \in S^n_{++}$, there exist Q and R orthonormal matrices and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, $\Omega = \operatorname{diag}(\omega_1, \ldots, \omega_n)$ diagonal matrices, satisfying $\lambda_1 \geq \cdots \geq \lambda_n$ and $\omega_1 \geq \cdots \geq \omega_n$, such that

$$X = Q^T \Lambda Q, \quad Y = R^T \Omega R. \tag{29}$$

Hence, Lemma 1.1 implies that $X \bullet \ln(Y) \leq \sum_{i=1}^n \lambda_i \ln(\omega_i)$. Thus, it follows from (28) and (29) that

$$\varphi(X) - \varphi(Y) - \nabla\varphi(Y) \bullet (X - Y) \geq \sum_{i=1}^{n} \lambda_i \ln(\lambda_i) - \sum_{i=1}^{n} \lambda_i \ln(\omega_i) + \sum_{i=1}^{n} \omega_i - \sum_{i=1}^{n} \lambda_i.$$

Now, let $h : \mathbb{R}^n_{++} \to \mathbb{R}^n$ the entropy function $h(x_1, \ldots, x_n) = \sum_{i=1}^n x_i \ln(x_i)$. As h is strictly convex, gradient inequality gives

$$0 < h(\lambda) - h(\omega) - \nabla h(\omega) \bullet (\lambda - \omega) = \sum_{i=1}^{n} \lambda_i \ln(\lambda_i) - \sum_{i=1}^{n} \lambda_i \ln(\omega_i) + \sum_{i=1}^{n} \omega_i - \sum_{i=1}^{n} \lambda_i,$$

where $\lambda = (\lambda_1, \ldots, \lambda_n)$ and $\omega = (\omega_1, \ldots, \omega_n)$. So, combining two above equation we conclude that

$$\varphi(X) > \varphi(Y) + \nabla\varphi(Y) \bullet (X - Y), \quad X, Y \in S^n_{++}, \quad X \neq Y.$$

and therefore we have that φ is strictly convex.

Acknowledgements The authors would like to thank the two anonymous referees and the Associate Editor for their constructive comments which have helped to substantially improve the presentation of the paper. O. P. Ferreira's work was supported in part by CNPq Grants 302618/2005-8 and 475647/2006-8, PRONEX–Optimization(FAPERJ/CNPq) and FUNAPE/UFG. P. R. Oliveira was supported in part by CNPq. R. C. M. Silva was supported in part by CAPES-PICDT.

References

1. Aulender, A., Teboulle, M.: Interior gradient and proximal methods for convex and conic optimization. SIAM J. Optim. **16**(3), 697–725 (2006)
2. Aulender, A., Héctor Ramírez, C.: Penalty and barrier methods for convex semidefinite programming. Math. Methods Oper. Res. **63**, 195–219 (2006)

3. Bolte, J., Daniilidis, A., Lewis, A.: The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2007)
4. Cominetti, R., San Martín, J.: Asymptotic analysis of the exponential penalty trajectory in linear programming. Math. Program. **67**(2), 169–187 (1994)
5. da Cruz Neto, J.X., Ferreira, O.P., Monteiro, R.D.C.: asymptotic behavior of the central path for a special class of degenerate SDP problems. Math. Program. **103**(3), 487–514 (2005)
6. da Cruz Neto, J.X., Iusem, A.N., Ferreira, O.P., Monteiro, R.D.C: Dual convergence of the proximal point method with Bregman distances for linear programming. Optim. Methods Softw. Engl. **22**, 339–360 (2007)
7. Dieudonné, J.A.: Foundations of Modern Analysis. AcademicPress, New York (1960)
8. Doljansky, M., Teboulle, M.: An interior proximal algorithm and the exponential multiplier method for semidefinite programming. SIAM J. Optim. **9**(1), 1–13 (1998)
9. Dym, H.: Linear algebra in action. Graduate Studies in Mathematics **78**, AMS 2007, Providence, RI
10. Halická, M., de Klerk, E., Roos, C.: On the convergence of the central path in semidefinite optimization. SIAM J. Optim. **12**(4), 1090–1099 (2002)
11. Halická, M., de Klerk, E., Roos, C.: Limiting behavior of the central path in semidefinite optimization. Optim. Methods Softw. **20**(1), 99–113 (2005)
12. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, New York (1985)
13. Iusem, A.: Augmented Lagrangian methods and proximal point methods for convex optimization. Investigación Operativa **8**, 11–49 (1999)
14. Iusem, A.N., Monteiro, R.D.C.: On dual convergence of the generalized proximal point method with Bregman distances. Math. Oper. Res. **25**(4), 606–624 (2000)
15. Iusem, A.N., Svaiter, B.F., da Cruz Neto, J.X.: Central paths, generalized proximal point methods, and cauchy trajectories in Riemannian manifolds. SIAM J. Control Optim. **37**(2), 566–588 (1999)
16. Jensen, D.L., Polyak, R.A.: The convergence of a modified barrier method for convex programming. IBM J. Res. Dev. **38**(3), 307–321 (1994)
17. de Klerk, E.: Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications, Applied Optimization, Vol. 65. Kluwer Academic Publishers, Dordrecht (2002)
18. Kojima, M., Meggido, N., Noma, T., Yoshise, A.: A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems, Lecture Notes in Computer Science, Vol. 538. Springer Verlag, Berlin, Germany (1991)
19. Krantz, S., Parks, H.R.: A Primer of Real Analytic Functions. Birkhäuser Verlag, Boston (1992)
20. Kurdyka, K., Mostowski, T., Parusinski, A.: Proof of the gradient conjecture of R.Thom. Annals Math. II Ser. **152**(3), 763–792 (2000)
21. Lojasiewicz, S.: Ensembles Semi-analitiques. I.H.E.S., Bures-sur-Yvette (1965)
22. Mosheyev, L., Zibulevski, M.: Penalty/barrier multiplier algorithm for semidefinite programming. Optim. Methods Softw. **13**, 235–261 (2000)
23. Papa Quiroz, E.A., Roberto Oliveira, P.: A new barrier for a class of semidefinte problems. RAIRO-Oper. Res. **40**, 303–323 (2006)
24. Polyak, R., Teboulle, M.: Nonlinear rescaling and proximal-like methods in convex optimization. Math. Program. **76**(2), 265–284 (1997)
25. Powell, M.J.D.: Some convergence properties of the modified log barrier method for linear programming. SIAM J. Optim. **5**(4), 695–739 (1995)
26. Shiota, M.: Geometry of Subanalytic and Semialgebraics Sets. Birkhäuser, Boston (1997)
27. Sporre, G., Forsgren, A.: Characterization of the limit point of the central path in semidefinite programming. Technical Report TRITA-MAT-2002-OS12, Departament of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden, June 2002
28. Todd, M.J.: Semidefinite optimization. Acta Numerica **10**, 515–560 (2001)
29. Tseng, P., Bertsekas, D.P.: On the convergence of the exponential multiplier method for convex programming. Math. Program. **60**(1), 1–19 (1993)