Alfred Auslender · Héctor Ramírez C.

Penalty and barrier methods for convex semidefinite programming

Abstract In this paper we present penalty and barrier methods for solving general convex semidefinite programming problems. More precisely, the constraint set is described by a convex operator that takes its values in the cone of negative semidefinite symmetric matrices. This class of methods is an extension of penalty and barrier methods for convex optimization to this setting. We provide implementable stopping rules and prove the convergence of the primal and dual paths obtained by these methods under minimal assumptions. The two parameters approach for penalty methods is also extended. As for usual convex programming, we prove that after a finite number of steps all iterates will be feasible.

Keywords Semidefinite programming · Penalty and barrier methods · Asymptotic functions · Recession functions · Convex analysis

1 Introduction

Let S^m be the space of symmetric real $m \times m$ matrices endowed with the inner product $A \cdot B = \text{trace}(AB)$ denoting the trace of the matrix product AB, and let S^m_+ be the cone of positive semidefinite symmetric matrices. Related to S^m_+ we define the partial ordering $\succeq$ via

$$A \succeq B \quad \Leftrightarrow \quad B \preceq A \quad \Leftrightarrow \quad A - B \in S^m_+, \quad \forall A, B \in S^m.$$

A. Auslender (✉)
Institut Girard Desargues, Université Claude Bernard – Lyon 43, boulevard du 11 Novembre 1918, 69622 Villeurbane Cedex, France
E-mail: aauslen@cegetel.fr

H. Ramírez C.
Department of Mathematical Engineering, Universidad de Chile & Centre for Mathematical Modelling, UMR 2071, Universidad de Chile-CNRS, Casilla 170-3 Santiago 3, Chile
E-mail: hramirez@dim.uchile.cl

We denote $A \succ 0$ or $0 \prec A$ if $A \in S^m_{++}$, the cone of positive definite symmetric $m \times m$ matrices.

Similar relations can be established for S^m_- and S^m_{--}, the cones of negative semidefinite and definite symmetric $m \times m$ matrices, respectively.

Throughout the general development, we denote by $\mathbb{R}^n$ an arbitrary finite real-dimensional space, and by $\langle \cdot, \cdot \rangle$ an arbitrary inner product on $\mathbb{R}^n$.

This paper is focused on convex optimization with constraint sets described mainly by $\succeq$ convex maps, which are defined as follows: let X be a convex set in $\mathbb{R}^n$, a map $G \colon X \to S^m$ is said to be $\succeq$ convex if

$$G(\lambda x + (1-\lambda)y) \preceq \lambda G(x) + (1-\lambda)G(y), \quad \forall x, y \in X, \ \forall \lambda \in [0, 1].$$

Simple examples of $\succeq$ convex maps that show the interest of this notion are affine maps as $G(x) = B + \sum_{j=1}^{n} x_j A_j$ with $B, \ A_j \in S^m$, or functions of the form $G(x) = B + \sum_{j=1}^{p} g_j(x) A_j$ where the $g_j(\cdot)$'s are convex functions while the A_j's are positive semidefinite matrices. Similarly, matrix convex functions, for instance $x^2 \colon S^m \to S^m$ and $-\log x \colon S^m_{++} \to S^m_{++}$, are $\succeq$ convex maps defined on a matrix space. Other examples, properties and applications of such maps can be found in the books of Bhatia (1997, chapter 5), Bonnans and Shapiro (2000, chapter 5), and Ben-Tal and Nemirowskii (2002, chapter 4).

Throughout this paper, we suppose that G is a $\succeq$ convex map, continuously differentiable (C^1) on $\mathbb{R}^n$, and $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $i = 0, 1, \ldots, p$, are convex, lower semicontinuous(lsc) functions. Thus, we define

$$D = \{x \in \mathbb{R}^n \colon \ f_i(x) \leq 0, \qquad \forall i = 1, \ldots, p\},$$
$$E = \{x \in \mathbb{R}^n : \ G(x) \preceq 0\}, \qquad C = D \cap E,$$

and consider the optimization problem

$$(\mathrm{P}) \qquad v = \inf\{f_0(x) \mid x \in C\}.$$

The aim of this paper is to propose penalty and barrier methods for solving (P). Methods of this kind have been widely developed in nonlinear optimization (i.e. $C = D$). In this context, Auslender et al. (1997) have proposed a unified framework containing most of the methods given in the literature. The article (Auslender et al. 1997) also provides a systematic way to generate penalty and barrier methods.

In the case when $C = D \cap E$ and G is an affine map into S^m, Auslender (1999) proposed a general framework for solving (P). Roughly speaking, a systematic way for building penalty and barrier functions ϕ_r with parameter $r > 0$ going ultimately to 0 was presented. These functions are defined in order to solve a family of unconstrained minimization problems of the form

$$(\mathrm{P}_r) \qquad v_r = \inf\{f_0(x) + \phi_r(x) \mid x \in \mathbb{R}^n\}.$$

In the work by Auslender (1999), the existence of optimal solutions x_r of (P_r) is guaranteed by supposing Slater's condition and the usual hypothesis that the optimal set S of (P) is nonempty and compact. Then, it was proven that the generalized sequence $\{x_r\}_{r>0}$ is bounded with all its limit points in S.

In the first part of this paper our objective is to improve the results established in Auslender (1999) in three directions. Firstly, we give an implementable stopping rule that ensures the obtainment of x_r in a finite number of steps by any usual unconstrained descent method. This avoids the exact minimization used in Auslender (1999) to obtain x_r.

Secondly, here G is no longer affine but $\succeq$ convex. Hence, the convergence analysis is now much more complicated than in the affine case. Indeed, the computation of the recession function of ϕ_r by a useful formula is actually no longer available when G is $\succeq$ convex, contrary to the case when G is affine. Unfortunately, the recession functional analysis is a key element in our approach. The only known result when G is $\succeq$ convex appears in Graña Drummond and Peterzil (2002), where they use the classical log-barrier function in semidefinite programming (SDP) composed with $G(x)$ instead of a more general penalty or barrier function $\phi_r(x)$. In the work of Graña Drummond and Peterzil (2002), convergence properties are obtained under a restrictive assumption (cf. Graña Drummond and Peterzil 2002, Assumption A2). Here, the convergence is proven for general penalty and barrier functions assuming the two usual hypotheses in constrained convex programming, that is, the optimal set of (P) is nonempty and compact, and Slater's condition holds.

A third direction is the improvement of the duality results given in Auslender (1999) and Auslender et al. (1997), where the exact solution of the Fenchel dual problem of (P_{r_k}) is supposed to be computed ($\{r_k\}$ is a sequence of positive real numbers going to 0). Obviously this is a theoretical result. Here we associate with x_{r_k} a multiplier Y_k given by an explicit formula. Then we prove that the sequence $\{Y_k\}$ is bounded and that each limit point of this sequence is an optimal solution of the usual Lagrangian dual of (P).

Penalty and barrier methods introduced in section 3 are based on a smoothing procedure and depend on a single parameter. This smoothing procedure involves two possible classes of penalty functions. The first class deals with the indicator function of $\mathbb{R}^p_- \times S^m_-$, while the second class concerns an exact penalty function. However, when $C = D$, i.e. when we only consider the classical convex constrained programming problem, a second approach can be used. This approach is only applied to functions of the second class mentioned above and its basic idea consists of distinguishing two parameters: the "smoothness parameter" r and the penalty weight β. This two-parameter approach has been firstly developed by Xavier (1992) for a specific hyperbolic function and has been also the base of a recent work of Gonzaga and Castillo (2003). Indeed, Gonzaga and Castillo (2003) introduce a method that uses a smooth approximation $\theta(\cdot)$ of the exact penalty function $t \to \max\{0, t\}$ and two parameters, r and β, so that the penalized function $\psi_{r,\beta}(x) := f_0(x) + \beta r \sum_{i=1}^{m} \theta\left(f_i(x)/r\right)$ is minimized at each iteration. The parameters play different roles: r always decreases in order to improve the precision of the approximation, and β increases to penalize an infeasible iteration. Thus, the aim of the second part of this article is to extend this approach to more general feasible sets C. Particularly, we consider $C = D \cap E$ instead of $C = D$, that is, a feasible set that involves semidefinite constraints. Nevertheless, our results are an improvement of those in Gonzaga and Castillo (2003) even in the nonlinear programming case where $C = D$. Indeed firstly, we only work in the convex case which allows us to give an implementable stopping rule [this is not the case in

Gonzaga and Castillo (2003)]. Secondly, we do not suppose neither the assumption named "Hypothesis" in Gonzaga and Castillo (2003) nor the compactness of the feasible set. Finally, we associate with the primal sequence a dual sequence of multipliers given by an explicit formula. Hence we prove that this dual sequence is bounded with each limit point being an optimal solution of the usual Lagrangian dual of (P). Such a result is not given in the work by Gonzaga and Castillo (2003).

The outline of this paper is as follows. In the next section we recall material concerning recession functions, convex analysis in SDP and matrix properties which will be needed in the sequel. In section 3 we present the penalty and barrier methods, including the convergence analysis concerning the primal path. Section 4 deals primarily with the dual path. Finally in section 5 we consider the penalty approach with two parameters.

2 Preliminaries

2.1 Asymptotic cones and functions

We recall some basic notions about asymptotic cones and functions [see for more details the books of Auslender and Teboulle (2003) and of Rockafellar (1970)].

The asymptotic cone of a set $Q \subseteq \mathbb{R}^n$ is defined to be

$$Q_\infty = \left\{ y : \exists t_k \to +\infty,\ x_k \in Q \quad \text{with } y = \lim_{k\to\infty} \frac{x_k}{t_k} \right\}. \tag{2.1}$$

When Q is convex and closed, it coincides with its recession cone

$$0^+(Q) := \{y : x + \lambda y \in Q \ \ \forall \lambda > 0, \quad \forall x \in Q\}. \tag{2.2}$$

Let f: $\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be lower-semicontinuous (lsc) and proper (i.e., $\exists x \in \operatorname{dom} f := \{x : f(x) < +\infty\}$). We recall that the asymptotic function f_∞ of f is defined by the relation

$$\operatorname{epi} f_\infty = (\operatorname{epi} f)_\infty,$$

where $\operatorname{epi} f := \{(x, r) :\ f(x) \le r\}$. As a straightforward consequence, we get (cf. Auslender and Teboulle 2003, Theorem 2.5.1)

$$f_\infty(y) = \inf \left\{ \liminf_{k\to+\infty} \frac{f(x_k t_k)}{t_k} :\ t_k \to +\infty,\ x_k \to y \right\} \tag{2.3}$$

where the sequences $\{t_k\}$ and $\{x_k\}$ belong to $\mathbb{R}$ and $\mathbb{R}^n$, respectively.

Remark 2.1 This formula is fundamental in the convergence analysis of unbounded sequences and is often used in the following way: let $\{x_k\}$ be an unbounded sequence satisfying

$$\lim_{k\to\infty} \|x_k\| = +\infty, \quad \lim_{k\to\infty} \frac{x_k}{\|x_k\|} = d \neq 0.$$

Suppose that $f_\infty(d) > -\infty$, and let $\alpha \in \mathbb{R}$ so that $f_\infty(d) > \alpha$. Then it follows from (2.3) that for all k sufficiently large we have

$$f(x_k) = f\left(\frac{x_k}{\|x_k\|} \|x_k\| \right) \ge \alpha \|x_k\|.$$

Note also that f_∞ is positively homogeneous, that is

$$f_\infty(\lambda d) = \lambda f_\infty(d) \quad \forall d,\ \forall \lambda > 0. \tag{2.4}$$

When f is a convex, lsc, proper function its asymptotic function coincides with its recession function

$$0^+ f(y) = \lim_{\lambda \to +\infty} \frac{f(x+\lambda y) - f(x)}{\lambda}, \quad \forall x \in \operatorname{dom} f, \tag{2.5}$$

deducing immediately that

$$f_\infty(y) = \lim_{t \to +\infty} \frac{f(ty)}{t}, \quad \forall y \in \operatorname{dom} f. \tag{2.6}$$

Furthermore, if $\partial f(x)$ denotes the (convex) subdifferential of f at x, we also have

$$f_\infty(y) = \sup\{\langle c, y\rangle \mid c \in \partial f(x), \quad x \in \operatorname{dom} \partial f\}. \tag{2.7}$$

Now consider the lsc functions $f,\ g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ satisfying $f_\infty(d) > -\infty$ and $g_\infty(d) > -\infty$. Then

$$(f+g)_\infty(d) \geq f_\infty(d) + g_\infty(d), \tag{2.8}$$

with equality in the convex case. Recall that $f_\infty(d) > -\infty$ always holds when f is convex, lsc and proper.

When f is convex, a useful consequence of (2.2) and (2.5) is the following

$$\{x :\ f(x) \leq \lambda\}_\infty = \{d :\ f_\infty(d) \leq 0\}, \tag{2.9}$$

for any λ such that $\{x :\ f(x) \leq \lambda\}$ is nonempty.

The following proposition is crucial in the convergence analysis. The reader can see a proof in Auslender and Teboulle (2003, chapter 3).

Proposition 2.1 *Let C be a closed convex set in $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex, lsc, proper function such that* dom $f \cap C$ *is nonempty. Consider the optimization problem*

$$\text{(P)} \quad \alpha = \inf\{f(x) \mid x \in C\},$$

and let S be the optimal set of (P)*. Then a necessary and sufficient condition for S to be nonempty and compact is given by*

$$f_\infty(d) > 0 \quad \forall d \in C_\infty,\ d \neq 0,$$

or equivalently

$$\lim_{||x|| \to \infty,\ x \in C} f(x) = +\infty.$$

In this case (P) *is said to be coercive.*

In our analysis, the asymptotic function of a composite function is of a particular interest. More precisely, we will consider the composition between a penalty or barrier function θ and the $\succeq$ convex function $G(\cdot)$.

Let us consider the following class of functions F introduced by Auslender et al. (1997)

$$F = \Big\{\theta : \mathbb{R} \to \mathbb{R} \cup \{+\infty\},\ \text{lsc, convex, proper and non-decreasing with}$$
$$\theta_\infty(1) > 0,\ \lim_{t\to\eta^-} \theta(t) = +\infty,\ \text{and}\ \operatorname{dom}\theta =]-\infty, \eta[\ \text{where}\ \eta \in [0, +\infty]\Big\}.$$

In the remainder of this paper, we consider two subclasses of F, namely F_1 and F_2 [cf. Auslender et al. (1997) and Chen and Mangasarian (1996), respectively] defined by

$$F_1 = \{\theta \in F:\ \theta \text{ is } C^1 \text{ on } \operatorname{dom}\theta,\ \theta_\infty(1) = +\infty,\ \theta_\infty(-1) = 0\},$$
$$F_2 = \Big\{\theta \in F:\ \operatorname{dom}\theta = \mathbb{R},\ \theta \text{ is } C^1,\ \theta_\infty(1) = 1,\ \lim_{t\to-\infty} \theta(t) = 0\Big\}.$$

For example, the functions

$\theta_1(u) = \exp(u)$,	$\operatorname{dom}\theta = \mathbb{R}$	$\to$ exponential penalty (Cominetti and Dussault 1994),
$\theta_2(u) = -\log(1-u)$,	$\operatorname{dom}\theta =]-\infty, 1[$,	$\to$ modified log barrier (Polyak 1992),
$\theta_3(u) = \frac{u}{1-u}$,	$\operatorname{dom}\theta =]-\infty, 1[$,	$\to$ hyperbolic modified barrier (Ben-Tal and Zibulevsky 1997),
$\theta_4(u) = -\log(-u)$,	$\operatorname{dom}\theta =]-\infty, 0[$,	$\to$ log barrier (Frisch 1995),
$\theta_5(u) = -u^{-1}$,	$\operatorname{dom}\theta =]-\infty, 0[$	$\to$ inverse barrier method (Den Hertzog et al. 1991),

belong to the class F_1, while the functions

$$\theta_6(u) = \log(1 + \exp(u)),\quad \theta_7(u) = 2^{-1}(u + \sqrt{u^2 + 4})$$

belong to F_2. Furthermore, systematic ways to generate classes of functions θ belonging either to F_1 or to F_2 are described in Auslender et al. (1997) and Chen and Mangasarian (1996).

The following result was proven in Auslender et al. (1997).

Proposition 2.2 *Let $\theta \in F$, f be a convex, lsc, proper function with $\operatorname{dom}\theta \cap f(\mathbb{R}^n) \neq \emptyset$ and consider the composite function*

$$g(x) = \theta(f(x)) \quad \text{if } x \in \operatorname{dom} f, \quad +\infty \ \text{otherwise}.$$

Then the function g is a convex, lsc, proper function and we have

$$g_\infty(d) = \theta_\infty(f_\infty(d)) \quad \text{if } d \in \operatorname{dom} f_\infty, \quad +\infty \ \text{otherwise}.$$

2.2 Convex analysis over the cone of symmetric semidefinite positive matrices

Let S^m be equipped with the inner product $A \cdot B := \mathrm{trace}(AB)$ where $\mathrm{trace}(A)$ denotes the trace of the matrix A. Let $A \in S^m$ with the eigenvalue decomposition $A = Q \Lambda Q^{\mathrm{t}}$. Thus Q is an orthogonal matrix whose columns $q_i,\ i = 1, \ldots, m$, are the orthonormalized eigenvectors of A, and Λ is a diagonal matrix whose entries $\lambda_i(A),\ i = 1, \ldots, m$, are the eigenvalues of A in nonincreasing order.

Let $c_i(A) := q_i q_i^{\mathrm{t}}$. The spectral decomposition of A can be written as

$$A = \sum_{i=1}^{m} \lambda_i(A) c_i(A).$$

Now, let $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$. For any $A \in S^m$ such that $\lambda_i(A) \in \mathrm{dom}\, g$ for each i, we set

$$g^\circ(A) := \sum_{i=1}^{m} g(\lambda_i(A)) c_i(A), \tag{2.10}$$

the usual matrix function associated with g. We are particularly interested here in the function Ψ_g: $S^m \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\Psi_g(A) = \sum_{i=1}^{m} g(\lambda_i(A)) \ \text{ if } \lambda_i(A) \in \mathrm{dom}\, g \text{ for each } i, \quad +\infty \ \text{ otherwise}, \tag{2.11}$$

or equivalently

$$\Psi_g(A) = \mathrm{trace}(g^\circ(A)) \ \text{ if } \lambda_i(A) \in \mathrm{dom}\, g \text{ for each } i, \quad +\infty \ \text{ otherwise.}$$

The function Ψ_g is a spectrally defined function and the following properties hold (see e.g., Auslender 2003, Proposition 2.2)

Proposition 2.3 *Suppose that $g \in F$. Then*

(i) *Ψ_g is a proper, lsc, convex function.*
(ii) *dom Ψ_g is open.*
(iii) *If g is C^1 on dom g, then Ψ_g is C^1 on dom Ψ_g with $\nabla \Psi_g(A) = (g')^\circ(A)$, for all $A \in$ dom Ψ_g.*
(iv) *$(\Psi_g)_\infty(D) = \Psi_{g_\infty}(D)$, for all D.*
(v) *For $g \in F$, Ψ_g is isotone, i.e., $A \succeq B \Rightarrow \Psi_g(A) \geq \Psi_g(B)$.*
(vi) *For all $D \in S^m$ it holds that*

$$\begin{aligned}(\Psi_\theta)_\infty(D) &= \delta_{S_-^m}(D), \quad \textit{if } \theta \in F_1, &(2.12)\\ &= \Psi_{a^+}(D), \quad \textit{if } \theta \in F_2, &(2.13)\end{aligned}$$

where $\delta_{S_-^m}$ is the indicator function of $S_-^m = -S_+^m$ and where $a^+ = \max(0, a)$ with $a \in \mathbb{R}$.

Consider the functions $\theta \in F$ given in section 2.1 and set $L := \Psi_\theta$. For $\theta \in F_1$, we have the following examples from the work of Auslender (1999):

$$L_1(D) = \operatorname{trace}(\exp D),$$

$$L_2(D) = \begin{cases} -\log(\det(I - D)) & \text{if } D \prec I, \\ +\infty & \text{otherwise}, \end{cases}$$

$$L_3(D) = \begin{cases} \operatorname{trace}((I - D)^{-1} D) & \text{if } D \prec I, \\ +\infty & \text{otherwise}, \end{cases}$$

$$L_4(D) = \begin{cases} -\log(\det(-D)) & \text{if } D \prec 0, \\ +\infty & \text{otherwise}, \end{cases}$$

$$L_5(D) = \begin{cases} \operatorname{trace}(-D^{-1}) & \text{if } D \prec 0, \\ +\infty & \text{otherwise}. \end{cases}$$

And for $\theta \in F_2$ we get

$$L_6(D) = \log(\det(I + \exp D)), \quad L_7(D) = \operatorname{trace}\left(\frac{D + \sqrt{D^2 + 4I}}{2}\right).$$

It is worthwhile to note that L_4 is the classical log-barrier function used in semi-definite programming (see Nesterov and Nemirovski 1994).

To end this subsection we recall two characterizations of $\succeq$ convexity. First, it is easy to show that $G\colon \mathbb{R}^n \to S^m$ is $\succeq$ convex iff for each $u \in \mathbb{R}^m$ the map $x \to u^{\mathrm{t}} G(x) u$ is convex. Then, if in addition G is continuously differentiable (C^1), these last assertions are also equivalent to

$$u^{\mathrm{t}} G(y) u \geq u^{\mathrm{t}} G(x) u + u^{\mathrm{t}} DG(x)(y - x) u, \quad \forall x, y \in \mathbb{R}^n, \ \forall u \in \mathbb{R}^m. \quad (2.14)$$

2.3 Matrix properties review

We start this section recalling the well-known Debreu's lemma.

Lemma 2.1 (Debreu's lemma). *Let $A \preceq 0$, we have that $v^t B v < 0$, for all $v \in \operatorname{Ker} A \setminus \{0\}$ if and only if there exists $r > 0$ such that $B + rA \prec 0$.*

Consider a symmetric matrix $A \in S^m$. Let $l_0(A)$ and $l_+(A)$ be the number of their null and nonnegative eigenvalues, respectively, and let $E(A) \in \mathbb{R}^{m \times l_0(A)}$ and $E^+(A) \in \mathbb{R}^{m \times l_+(A)}$ be matrices whose columns are orthonormalized eigenvectors of A associated with their null and nonnegative eigenvalues, respectively.

The following relations are directly established

$$\operatorname{Im} E(A) = \operatorname{Ker} A \subseteq \operatorname{Im} E^+(A) = \operatorname{Im} A^+ + \operatorname{Ker} A = \operatorname{Ker} A^-,$$

and hence

$$l_0(A) = \dim(\operatorname{Ker} A) \leq l_+(A) = \dim(\operatorname{Im} A^+) + \dim(\operatorname{Ker} A) = \dim(\operatorname{Ker} A^-),$$

where A^+ (A^-) denotes the orthogonal projection of $A \in S^m$ onto the cone S^m_+ (S^m_-) of $m \times m$ positive (negative) semidefinite symmetric matrices. This is given by

$$A^+ := Q \operatorname{diag}(\lambda_1(A)^+, \ldots, \lambda_m(A)^+) Q^t,$$

where Q is an orthogonal matrix such that its ith column is an eigenvector of A associated with $\lambda_i(A)$. Matrix A^- is similarly stated.

So, if $A \preceq 0$, then $A = A^-$ obtaining that $\operatorname{Im} E(A) = \operatorname{Im} E^+(A)$ and $l_0(A) = l_+(A) = \dim(\operatorname{Ker} A)$.

When $x \in I\!R^n$, similar relations hold for $E(G(x))$ and $E^+(G(x))$.

The following lemma is a direct consequence of the continuity of the eigenvalue function $\lambda_i(\cdot)$.

Lemma 2.2 *Consider a matrix $\tilde{A} \preceq 0$. If $A_k \to \tilde{A}$, then for all k sufficiently large, we have that $l_+(A_k) \leq l_0(\tilde{A})$.*

The next lemma will be very useful in the rest of this article. Its proof appears in Bonnans and Shapiro (2000, Example 3.140) and is included here in order to make this work as selfcontained as possible.

Lemma 2.3 *Consider a matrix $\tilde{A} \preceq 0$. If $A_k \to \tilde{A}$, then we can construct a matrix $E_k \in \mathbb{R}^{m \times l_0(\tilde{A})}$ whose columns are an orthonormal basis of the space spanned by the eigenvectors associated with the $l_0(\tilde{A})$ biggest eigenvalues of A_k, such that $E_k \to E(\tilde{A})$.*

Proof Consider $\tilde{E} := E(\tilde{A})$ and $\tilde{l} := l_0(\tilde{A}) = l_+(\tilde{A})$ (because $\tilde{A} \preceq 0$). For a given A, let $e_1(A), \ldots, e_{\tilde{l}}(A)$ be a set of orthonormal eigenvectors of A associated with their $\tilde{l}$ biggest eigenvalues $\lambda_1(A) \geq \cdots \geq \lambda_{\tilde{l}}(A)$. Denote by $L(A)$ the space spanned by the eigenvectors $e_1(A), \ldots, e_{\tilde{l}}(A)$ and let $P(A)$ be the orthogonal projection matrix onto $L(A)$. Note that $L(\tilde{A}) = \operatorname{Im} \tilde{E} = \operatorname{Ker} \tilde{A}$.

It is known that the projection matrix $P(A)$ is a continuous (and even analytic) function of A in a sufficiently small neighborhood of $\tilde{A}$ [see, for example, Kato (1970, Theorem 1.8) and Golub and Van Loan (1996, Corollary 8.1.11)]. Consequently the function $F(A) := P(A)\tilde{E}$ is also a continuous function of A in a neighborhood of $\tilde{A}$, and moreover $F(\tilde{A}) = \tilde{E}$. It follows that for all A sufficiently close to $\tilde{A}$, the rank of $F(A)$ is equal to the rank of $F(\tilde{A}) = \tilde{E}$, i.e., $\operatorname{rank} F(A) = \tilde{l}$. It means that the $\tilde{l}$ columns of $F(A)$ are linearly independent when A is sufficiently close to $\tilde{A}$. Now let $U(A)$ be a matrix whose columns are obtained by applying the Gram–Schmidt orthonormalization process to the columns of $F(A)$. The matrix $U(A)$ is well defined and continuous in a neighborhood of $\tilde{A}$. Even more, the matrices $U(A)$ satisfy that their columns are orthonormalized, i.e. $U(A)^t U(A) = I_{\tilde{l}}$, and $\operatorname{Im} U(A) = L(A)$, for all A sufficiently close to $\tilde{A}$. We also have that $U(\tilde{A}) = F(\tilde{A}) = \tilde{E}$. Hence the theorem follows by setting $E_k := U(A_k)$. □

From Lemmas 2.2 and 2.3 we get directly the following corollary concerning a feasible set $C = \{x\colon\ G(x) \preceq 0\}$ where $G\colon \mathbb{R}^n \to S^m$ is $\succeq$ convex and continuous.

Corollary 2.1 *Consider a point $\bar{x}$ such that $G(\bar{x}) \preceq 0$. If $x_k \to \bar{x}$, then for all k sufficiently large, we have that $l_+(G(x_k)) \leq l_0(G(\bar{x}))$. Furthermore, we can construct a matrix $E_k \in \mathbb{R}^{m \times l_0(G(\bar{x}))}$ whose columns are an orthonormal basis of the space spanned by the eigenvectors associated with the $l_0(G(\bar{x}))$ biggest eigenvalues of $G(x_k)$, such that $E_k \to E(G(\bar{x}))$.*

The notions introduced in this subsection allow us to characterize Slater's condition: there exists x^0 such that $G(x^0) \prec 0$, as follows.

Proposition 2.4 *Suppose that G is a $\succeq$ convex map C^1 on $\mathbb{R}^n$. Then Slater's condition is equivalent to* Robinson's constraint qualification *condition*

$$\text{for all } \bar{x} \text{ such that } G(\bar{x}) \preceq 0 \text{ there exists } \bar{h} \in \mathbb{R}^n \text{ such that } G(\bar{x}) + DG(\bar{x})\bar{h} \prec 0. \tag{2.15}$$

Moreover, Robinson's condition (2.15) *is always equivalent to*

$$\begin{aligned} &\text{for all } \bar{x} \text{ such that } G(\bar{x}) \preceq 0 \text{ there exists} \\ &\bar{h} \in \mathbb{R}^n \text{ such that } E(G(\bar{x}))^t DG(\bar{x})\bar{h} E(G(\bar{x})) \prec 0. \end{aligned} \tag{2.16}$$

Proof That Robinson's condition (2.15) implies Slater's condition is well-known and follows directly from the differentiability of G and the convexity of the set S^m_-. This is true even when G is not $\succeq$ convex. Conversely, Slater's condition and inequality (2.14) implies in a straightforward way condition (2.15). Finally, the equivalence between conditions (2.15) and (2.16) is due to Debreu's lemma (Lemma 2.1). □

3 Penalty and barrier methods: description and convergence analysis

For the sake of simplicity, we consider here the optimization problem (P) described in the introduction when $C = E$, i.e., problem (P) only contains semidefinite constraints. Then throughout this paper $G: \mathbb{R}^n \to S^m$ is a $\succeq$ convex map C^1 on $\mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}$ is a C^1 convex function, and we consider the optimization problem

$$\text{(P)} \qquad v = \inf\{f(x) \mid x \in C\},$$

where $C = \{x \in \mathbb{R}^n : \; G(x) \preceq 0\}$.

Indeed, if we define $D = \{x \in \mathbb{R}^n : \; F(x) \preceq 0\}$ when $F(x)$ is the diagonal matrix whose entries are given by the functions f_i's (obviously $F(\cdot)$ is a $\preceq$ convex map), then the constraint set $C = D \cap E$ is given by a convex operator that takes its values in S^m_-.

From now on we assume

- (A_1) The optimal set of (P), denoted by S, is nonempty and compact,
- (A_2) Slater's condition holds, i.e. there exists x^0 such that $G(x^0) \prec 0$.

Let $r > 0$ be a penalty parameter which will ultimately go to 0 and $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{r \to 0^+} \alpha(r) = 0 \quad \text{and} \quad \liminf_{r \to 0^+} \frac{\alpha(r)}{r} > 0. \tag{3.1}$$

We associate with each $\theta \in F$ the function Ψ_θ: $S^m \to \mathbb{R} \cup \{+\infty\}$ given by formula (2.11), and define the function $H^r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ by

$$H^r(x) = \Psi_\theta\left(\frac{G(x)}{r}\right) = \sum_{i=1}^{m} \theta\left(\frac{\lambda_i(G(x))}{r}\right), \tag{3.2}$$

where $\lambda_i(A)$ denotes the ith eigenvalue in nonincreasing order of A ($\lambda_1(A)$ is the largest eigenvalue of A).

In this section, we study methods that consist of solving "approximatively" the unconstrained minimization problems

$$(\mathrm{P}_r) \qquad v_r = \inf\{\phi_r(x) \mid x \in \mathbb{R}^n\}, \quad \text{where } \phi_r(x) = f(x) + \alpha(r)H^r(x). \tag{3.3}$$

It is worthwhile to note that when $C = D$ we recover the methods introduced in Auslender et al. (1997).

As in Auslender (1999) and Auslender et al. (1997), we consider two classes of methods; $\theta \in F_1$ and $\theta \in F_2$.

Throughout, we denote by S_r the optimal set of (P_r) and assume that

$$\alpha(r) = r, \quad \text{if } \theta \in F_1 \quad \text{and} \quad \lim_{r \to 0^+} \frac{\alpha(r)}{r} = +\infty, \quad \text{if } \theta \in F_2. \tag{3.4}$$

More precisely, we set

$$r_k > 0, \ \epsilon_k \geq 0, \ \gamma_k > 0 \ \text{ with } \ \lim_{k\to\infty} \epsilon_k = \lim_{k\to\infty} \gamma_k = \lim_{k\to\infty} r_k = 0. \tag{3.5}$$

Solving approximatively (P_{r_k}) means to compute $x_k \in \mathbb{R}^n$ such that if we set $\eta_k := \nabla\phi_{r_k}(x_k) = \nabla f(x_k) + \alpha(r_k)\nabla H^{r_k}(x_k)$ then

$$||\eta_k|| \leq \epsilon_k, \quad ||\eta_k|| \cdot ||x_k|| \leq \gamma_k. \tag{3.6}$$

Note that if the optimal set S_k of (P_{r_k}) is nonempty and compact, then any usual descent method (gradient type, Newton or quasi-Newton type method) provides in a finite number of steps a point x_k satisfying the stopping rule (3.6). Consequently, we will prove first that S_k is nonempty and compact. Indeed, it is true for all k when $\theta \in F_1$, and for k sufficiently large when $\theta \in F_2$. The next proposition will be a key result on this subject, and also for other purposes.

Proposition 3.1 *For $i = 1, \ldots, m$ and $r > 0$, let $\tilde{\lambda}_i(x) = \lambda_i(G(x))$, $h_i^r(x) = \theta\left(\frac{\tilde{\lambda}_i(x)}{r}\right)$. Then*

(i) *$\tilde{\lambda}_i(\cdot)$ and $h_i^r(\cdot)$ are continuous functions on $\mathbb{R}^n$.*
(ii) *$(\tilde{\lambda}_i)_\infty(d) > -\infty$, for all d.*
(iii) *$(h_i^r)_\infty(d) \geq 0$, for all d.*

(iv) $\tilde{\lambda}_1$ *is a convex continuous function on* $\mathbb{R}^n$ *and*

$$(\tilde{\lambda}_1)_\infty(d) \le 0 \quad \textit{iff} \;\; d \in C_\infty. \tag{3.7}$$

Furthermore, h_1^r *is an lsc proper convex function, and for each* $d \in \mathbb{R}^n$ *we have*

$$(h_1^r)_\infty(d) = \frac{(h_1^1)_\infty(d)}{r} = \begin{cases} \delta_{\mathbb{R}_-}((\tilde{\lambda}_1)_\infty(d)) & \textit{if } \theta \in F_1, \\ \frac{((\tilde{\lambda}_1)_\infty(d))^+}{r} & \textit{if } \theta \in F_2, \end{cases} \tag{3.8}$$

where $\delta_{\mathbb{R}_-}(y) = \{0, \;\; \textit{if } y \le 0\,; \; +\infty, \;\; \textit{if } y > 0\}$.

Proof (i) Since $\lambda_i(\cdot)$ and $G(\cdot)$ are continuous, their composition $\tilde{\lambda}_i(\cdot)$ is also continuous. In order to prove that $h_1^r(\cdot)$ is continuous, let $y = \lim_{k\to\infty} y_k$, then since $\tilde{\lambda}_i(\cdot)$ is continuous we have $\tilde{\lambda}_i(y_k)/r \to \tilde{\lambda}_i(y)/r$. If $(\tilde{\lambda}_i)(y)/r \notin \operatorname{bd}\operatorname{dom}\theta$ then, by continuity of θ on $\operatorname{int}\operatorname{dom}\theta$, we have $h_i^r(y_k) \to h_i^r(y)$. If $\tilde{\lambda}_i(y)/r \in \operatorname{bd}\operatorname{dom}\theta$, that is, $\tilde{\lambda}_i(y)/r = \eta$, the same limit holds thanks to the property $\lim_{u\to\eta^-} \theta(u) = +\infty$.

(ii) Let $d' \to d,\ t \to +\infty$, and let x^0 satisfy Slater's condition (A_2). Since G is $\succeq$ convex, for each $u \in \mathbb{R}^n$ we get (cf. (2.14))

$$u^t G(td')u \ge u^t G(x^0)u + u^t DG(x^0)(td' - x^0)u.$$

Taking $u = u_i$ such that $||u_i|| = 1$ and $G(td')u_i = \lambda_i(G(td'))u_i$, this last inequality yields

$$\frac{\lambda_i(G(td'))}{t} \ge -\frac{||G(x^0)||}{t} - ||DG(x^0)|| \cdot \left\| d' - \frac{x^0}{t} \right\|. \tag{3.9}$$

Passing to the liminf in (3.9), we obtain

$$(\tilde{\lambda}_i)_\infty(d) = \liminf_{t\to\infty,\ d'\to d} \frac{\lambda_i(G(td'))}{t} \ge -||DG(x^0)|| \cdot ||d||.$$

(iii) Since θ is nondecreasing we have from (3.9) with $G(\cdot)/r$ instead of $G(\cdot)$ that

$$\frac{1}{t} h_i^r(td') \ge \frac{1}{t}\theta\left(\frac{t}{r}\left[-\frac{||G(x^0)||}{t} - ||DG(x^0)|| \cdot \left\| d' - \frac{x^0}{t} \right\|\right]\right).$$

Passing to the liminf in this last inequality and using formula (2.3) we get

$$\begin{aligned}(h_i^r)_\infty(d) &= \liminf_{t\to\infty,\ d'\to d} \frac{h_i^r(td')}{t} \ge \liminf_{t\to\infty,\ u\to -\frac{1}{r}||DG(x^0)||\cdot||d||} \frac{\theta(tu)}{t} \\ &= \theta_\infty\left(-\frac{1}{r}||DG(x^0)|| \cdot ||d||\right),\end{aligned}$$

and, by virtue of the inequality $\theta_\infty \ge 0$, it follows that $(h_i^r)_\infty(d) \ge 0$.

(iv) Since $\tilde{\lambda}_1(x) = \max\{u^t G(x) u;\ \|u\| = 1,\ u \in \mathbb{R}^m\}$ and since G is $\succeq$ convex, we have that $\tilde{\lambda}_1(\cdot)$ is convex as a supremum of convex functions. Furthermore, since $C = \{x\colon\ \tilde{\lambda}_1(x) \leq 0\}$, it follows from (2.9) that $C_\infty = \{d\colon (\tilde{\lambda}_1)_\infty(d) \leq 0\}$ and then equivalence (3.7) holds.
So, by Proposition 2.2 we get that $h_i^r(\cdot)$ is lsc, convex and proper. Moreover, since θ_∞ is positively homogeneous, and $\operatorname{dom}\theta_\infty$ is either equal to $\mathbb{R}_-$ or $\mathbb{R}$, using again Proposition 2.2 we obtain

$$(h_1^r)_\infty(d) = \begin{cases} \frac{1}{r}\theta_\infty((\tilde{\lambda}_1)_\infty(d)) & \text{if } (\tilde{\lambda}_1)_\infty(d) \in \operatorname{dom}\theta_\infty, \\ +\infty & \text{otherwise,} \end{cases}$$

so that

$$(h_1^r)_\infty(d) = \frac{(h_1^1)_\infty(d)}{r}.$$

Finally, equality (3.8) is a immediate consequence of these formulas and the definition of θ_∞. □

Now we proceed to prove that the optimal set S_r is nonempty and compact. As we mentioned before, this condition is enough to show that the rule defining the point x_k is implementable.

Theorem 3.1 (i) *Suppose that either $\theta \in F_1$, or $\theta \in F_2$ and $f_\infty(d) \geq 0$ for all d. Then S_r is nonempty and compact for all $r > 0$.*
(ii) *If $\theta \in F_2$ then S_r is nonempty and compact for all $r > 0$ sufficiently small.*

Proof (i) By Proposition 3.1, we have $(h_i^r)_\infty(d) \geq 0$, for all d, $i = 1, \ldots, m$ and $r > 0$, and since $\phi_r(x) = f(x) + \alpha(r)\sum_{i=1}^m h_i^r(x)$ we have from inequality (2.8) and formula (3.8) that

$$(\phi_r)_\infty(d) \geq f_\infty(d) + \frac{\alpha(r)}{r}(h_1^1)_\infty(d) \quad \forall d. \tag{3.10}$$

Suppose that $\theta \in F_1$. We get from (3.10) and Proposition 3.1, part (iv) that

$$(\phi_r)_\infty(d) \geq \begin{cases} f_\infty(d) & \text{if } d \in C_\infty, \\ (\phi_r)_\infty(d) = +\infty & \text{otherwise.} \end{cases}$$

Hence, since S is nonempty and compact it follows from Proposition 2.1 that $(\phi_r)_\infty(d) > 0$, for all $d \neq 0$, which is equivalent to saying that S_r is nonempty and compact.
Now suppose that $\theta \in F_2$ and $f_\infty(d) \geq 0$, for all d. Inequality (3.10) and Proposition 3.1, part (iv) imply again that $(\phi_r)_\infty(d) > 0$, for all $d \neq 0$, and the same conclusion holds.

(ii) Assume that $\theta \in F_2$. We shall prove that S_r is nonempty and compact for $r > 0$ sufficiently small. By contradiction, suppose the existence of sequences $r_k \to 0^+$, $d_k \to d \neq 0$ such that

$$f_\infty(d_k) + \frac{\alpha(r_k)}{r_k}(h_1^1)_\infty(d_k) \leq 0.$$

Due to the lower semicontinuity of f_∞ and $(h_1^1)_\infty$, and the fact that $\liminf_{k\to\infty} \frac{\alpha(r_k)}{r_k} = +\infty$, we apply liminf to the last inequality to obtain $(h_1^1)_\infty(d) = 0$ and $f_\infty(d) \le 0$. However, Proposition 3.1 tells us that $(h_1^1)_\infty(d) = 0$ is equivalent to $d \in C_\infty$ implying that $f_\infty(d) \le 0$ for some $d \in C_\infty$, $d \ne 0$, which is impossible because S is nonempty and compact. □

Remark 3.1 (i) Note that if f is an extended lsc function satisfying that $\inf\{f(x)|x \in \mathbb{R}^n\} > -\infty$, then condition $f_\infty(d) \ge 0$, for all d, always holds.

(ii) When $\theta \in F_2$ and is strictly increasing (which is the case of all the current examples), we can suppose, without loss of generality, that $f_\infty(d) \ge 0$ for all d. Indeed, if we set $g(x) := \theta(f(x))$, then problem (P) is equivalent to convex problem

$$(\mathrm{P}_s)\ \alpha = \inf\{g(x) \mid x \in C\}$$

in the sense that problems (P) and (P_s) share the same optimal set. This is due to the strict monotonicity of function θ. Hence condition $g_\infty(d) \ge 0$ for all d, follows from the fact that θ is nonnegative.

Theorem 3.2 *Let $\{x_k\}$ be a sequence satisfying relations (3.6). Then, this sequence is bounded and each limit point of this sequence is an optimal solution of* (P).

Proof Let x^0 be an arbitrary interior point of C (i.e. x^0 satisfies Slater's condition (A_2)). Since function $x \to \phi_r(x) = f(x) + \alpha(r)H^r(x)$ is convex, it follows from the definition of x_k and $\eta_k = \nabla\phi_{r_k}(x_k)$ (cf. (3.6)) that

$$f(x_k) + \alpha(r_k)H^{r_k}(x_k) \le f(x^0) + \alpha(r_k)H^{r_k}(x^0) + \langle \eta_k, x_k - x^0\rangle,$$

Hence, as a consequence of the monotonicity of θ we get for k sufficiently large

$$f(x_k) + \frac{\alpha(r_k)}{r_k}\sum_{i=1}^{m} r_k\theta\left(\frac{\lambda_i(G(x_k))}{r_k}\right) \le f(x^0) + m\alpha(r_k)\theta(\lambda_1(G(x^0))) + \langle \eta_k, x_k - x^0\rangle. \tag{3.11}$$

First, we proceed to prove that the sequence $\{x_k\}$ is bounded. We argue by contradiction. Without loss of generality we can assume that

$$||x_k|| \to +\infty, \quad \lim_{k\to\infty} \frac{x_k}{||x_k||} = d \ne 0,$$

Proposition 3.1, part (ii) says that $(\tilde{\lambda}_i)_\infty(d) > -\infty$. So, we define $\epsilon_i < (\tilde{\lambda}_i)_\infty(d)$. By formula (2.3) (see Remark 2.1) we have for all k sufficiently large

$$\tilde{\lambda}_i(x_k) = \tilde{\lambda}_i\left(\frac{x_k}{||x_k||}||x_k||\right) \ge \epsilon_i||x_k||.$$

By dividing (3.11) by $||x_k||$ we obtain from the last inequality

$$\frac{1}{||x_k||}f\left(\frac{x_k}{||x_k||}||x_k||\right) + \frac{\alpha(r_k)}{r_k}\sum_{i=1}^{m}\frac{r_k}{||x_k||}\theta\left(\frac{\epsilon_i||x_k||}{r_k}\right)$$
$$\le \frac{f(x^0)}{||x_k||} + m\frac{\alpha(r_k)}{||x_k||}\theta(\lambda_1(G(x^0)) + \frac{\langle \eta_k, x_k - x^0\rangle}{||x_k||}.$$

Taking the limit when $k \to +\infty$ and using relations (3.5) and (3.6) and formula (2.3), we get

$$f_\infty(d) + \lim_{k\to\infty} \frac{\alpha(r_k)}{r_k} \sum_{i=1}^{m} \theta_\infty(\epsilon_i) \le 0. \tag{3.12}$$

Now recall that if $\theta \in F_1$, then $\alpha(r) = r$, $\theta_\infty(-1) = \theta_\infty(0) = 0$ and $\theta_\infty(1) = +\infty$. Then we obtain from (3.12) that

$$\theta_\infty(\epsilon_i) = 0. \tag{3.13}$$

In the case when $\theta \in F_2$, we have $\lim_{k\to\infty} \frac{\alpha(r_k)}{r_k} = +\infty$, $\theta_\infty(-1) = \theta_\infty(0) = 0$ and $\theta_\infty(1) = 1$, and therefore (3.13) also holds. Thus, inequality (3.12) implies that $f_\infty(d) \le 0$. Furthermore, since θ_∞ is positively homogeneous it follows from (3.13) that $\epsilon_i \le 0$. Hence, letting $\epsilon_1 \to (\tilde{\lambda}_1)_\infty(d)$ we get that $(\tilde{\lambda}_1)_\infty(d) \le 0$, which is equivalent to $d \in C_\infty$ (cf. Proposition 3.1). This together with $f_\infty(d) \le 0$ and $d \ne 0$ implies a contradiction with the fact that the optimal solution set S is nonempty and compact.

We have proven that the sequence $\{x_k\}$ is bounded. Now let x be an limit point of the sequence $\{x_k\}$. For simplicity of notation, we suppose that $x = \lim_{k\to\infty} x_k$. We shall show that x is an optimal solution of (P).

Let $\delta < f(x)$, $\delta_i < \lambda_i(G(x))$ for all $i = 1, \ldots, m$. By continuity of functions f and $\lambda_i(G(\cdot))$, we have for all k sufficiently large that

$$\delta < f(x_k), \quad \delta_i < \lambda_i(G(x_k)), \;\; \forall i = 1, \ldots, m.$$

Then, from inequalities (3.6) and (3.11), and the monotonicity of θ it follows

$$\delta + \frac{\alpha(r_k)}{r_k} \sum_{i=1}^{m} r_k \theta\left(\frac{\delta_i}{r_k}\right) \le f(x^0) + m\alpha(r_k)\theta(\lambda_1(G(x^0))) + (\epsilon_k ||x^0|| + \gamma_k). \tag{3.14}$$

On the other hand, the following relations are satisfied (cf. (3.1) and (3.5))

$$\lim_{k\to\infty} \epsilon_k = \lim_{k\to\infty} \gamma_k = \lim_{k\to\infty} \alpha(r_k) = \lim_{k\to\infty} r_k = 0.$$

So, passing to the liminf in (3.14) we get

$$\delta + \lim_{k\to\infty} \frac{\alpha(r_k)}{r_k} \sum_{i=1}^{m} \theta_\infty(\delta_i) \le f(x^0),$$

which implies that $\theta_\infty(\delta_i) = 0$, for all i, and also $\delta \le f(x^0)$. In particular, $\theta_\infty(\delta_1) = 0$ which means that $\delta_1 \le 0$. Hence, by letting $\delta \to f(x)$ and $\delta_1 \to \lambda_1(G(x))$ we deduce that

$$x \in C \quad \text{and} \;\; f(x) \le f(x^0) \;\; \forall x^0 \in \operatorname{int} C.$$

Finally, continuity of f implies that $f(x) \le f(u)$ for all $u \in C$, that is, x is an optimal solution of (P). We thus obtain the desired result. □

4 Duality results

We associate with problem (P), defined in section 3, the following Lagrangian functional

$$L(x, Y) = f(x) + Y \cdot G(x), \quad \forall x \in \mathbb{R}^n, \ \forall Y \in S^m,$$

as well as the following dual functional

$$p(Y) = \begin{cases} -\inf\{L(x, Y) \mid x \in \mathbb{R}^n\} & \text{if } Y \succeq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Thus the (Lagrangian) dual problem of (P) is given by

$$\text{(D)} \quad \gamma = \inf\{p(Y) \mid Y \in S^m\}.$$

As in section 3, we suppose that f is a C^1 convex function, G is $\succeq$ convex and that Assumptions (A_1) and (A_2) and (3.4) still hold. Thus, if the primal path $\{x_k\}$ satisfies the stopping rule (3.6), the convergence Theorem 3.2 still tells us that the sequence $\{x_k\}$ is bounded and that each of its limit points is an optimal solution of (P). It is also well known that there is no duality gap between (P) and (D), and that the set T of optimal solutions of (D) is nonempty and compact under these assumptions (see e.g., Bonnans and Shapiro 2000, Theorem 5.81). Furthermore, the matrix $\bar{Y} \succeq 0$ will be an optimal solution of (D) iff there exists $\bar{x} \in C$ such that

$$\nabla_x L(\bar{x}, \bar{Y}) = \nabla f(\bar{x}) + DG(\bar{x})^t \bar{Y} = 0 \quad \text{and} \quad G(\bar{x}) \cdot \bar{Y} = 0. \tag{4.1}$$

Note that, for a linear operator $Ay := \sum_{i=1}^{n} y_i A_i$ with $A_i \in S^m$, as $DG(x)$, we have for its adjoint operator A^{t} the formula:

$$A^{\mathrm{t}} Z = (A_1 \cdot Z, \ldots, A_n \cdot Z)^{\mathrm{t}}, \quad \forall Z \in S^m. \tag{4.2}$$

Let

$$Y_k = \frac{\alpha(r_k)}{r_k} (\theta')^{\circ} \left(\frac{G(x_k)}{r_k} \right) = \frac{\alpha(r_k)}{r_k} \sum_{i=1}^{m} \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right) e_i^k (e_i^k)^t, \tag{4.3}$$

where $(\theta')^{\circ}$ is the matrix function associated with θ', defined in (2.10), and e_i^k's are orthonormal eigenvectors of $G(x_k)$ associated with the eigenvalues $\lambda_i(G(x_k))$.

Using the derivation rule given in Proposition 2.3, part (iii), we get

$$\eta_k = \nabla f(x_k) + DG(x_k)^t Y_k. \tag{4.4}$$

The aim of this section is to prove that the sequence $\{Y_k\}$ is bounded and that each limit point of this sequence is an optimal solution of the dual problem (D).

Theorem 4.1 *Consider a sequence $\{x_k\}$ satisfying relations (3.6), and let $\{Y_k\}$ be the sequence defined by formula* (4.3). *Then, $\{Y_k\}$ is bounded and each of its limit points is an optimal solution of* (D).

Proof It was proven in Theorem 3.2 that the sequence $\{x_k\}$ is bounded and that each of its limit points is an optimal solution of (P). Let $\bar{x}$ be a limit point of $\{x_k\}$ and $\bar{l} := l_0(\bar{x})$ be the number of the null eigenvalues of $G(\bar{x})$. For simplicity we suppose without loss of generality that $\lim_{k\to+\infty} x_k = \bar{x}$.

Now by Lemma 2.3 there exist sequences of orthonormal vectors $\{e_i^k\}$, $i = 1, \ldots, m$, which are eigenvectors of $G(x_k)$ associated with $\lambda_i(G(x_k))$, converging toward $\bar{e}_i$ such that the set $\{\bar{e}_i\colon i = 1, \ldots, m\}$ is an orthonormal eigenbasis of the matrix $G(\bar{x})$.

In order to prove that the sequence $\{Y_k\}$ is bounded, we will show that each sequence $\left\{\frac{\alpha(r_k)}{r_k}\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right)\right\}$ Particularly, we will show that, for all $i = \bar{l} + 1, \ldots, m$, these sequences converge to 0. This will be very useful to conclude that any limit point of $\{Y_k\}$ is a solution of (D).

First let us prove that

$$\lim_{t\to-\infty} \theta'(t) = 0. \tag{4.5}$$

Indeed, since θ' is nonnegative and nondecreasing it follows that $\lim_{t\to-\infty} \theta'(t) = \epsilon \geq 0$ and $\theta'(u) \geq 0$, for all $u \in \operatorname{dom}\theta$. Now formula (2.7) implies

$$\theta_\infty(-1) = \sup\{\langle -1, \theta'(t)\rangle\colon\ t \in \operatorname{dom}\theta\} = -\epsilon,$$

which together with the equality $\theta_\infty(-1) = 0$ allows us to conclude (4.5).

Now we proceed to show that

$$\frac{\alpha(r_k)}{r_k}\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right) \to 0, \quad \forall i = \bar{l} + 1, \ldots, m. \tag{4.6}$$

Lemma 2.2 tells us that $l_+(G(x_k)) \leq l_0(G(\bar{x})) =: \bar{l}$. This implies that for k sufficiently large we have

$$\lambda_i(G(x_k)) \leq \frac{\lambda_i(G(\bar{x}))}{2} < 0, \quad \forall i = \bar{l} + 1, \ldots, m.$$

In the case when $\theta \in F_1$, we know that $\alpha(r) = r$ and limit (4.6) follows directly from (4.5). Suppose then that $\theta \in F_2$. Since θ' is nonnegative and nondecreasing the last inequality yields to

$$0 \leq \frac{\alpha(r_k)}{r_k}\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right) \leq \frac{\alpha(r_k)}{r_k}\theta'\left(\frac{\lambda_i(G(\bar{x}))}{2r_k}\right). \tag{4.7}$$

Also from the fact that θ is nonnegative and convex, we get

$$\begin{aligned} 0 \leq \frac{\alpha(r_k)}{2r_k}\theta'\left(\frac{\lambda_i(G(\bar{x}))}{2r_k}\right)(-\lambda_i(G(\bar{x}))) &\leq \alpha(r_k)\left[\theta(0) - \theta\left(\frac{\lambda_i(G(\bar{x}))}{2r_k}\right)\right] \\ &\leq \alpha(r_k)\theta(0), \end{aligned}$$

which together with $\lim_{k\to\infty} \alpha(r_k) = 0$ and inequality (4.7) implies condition (4.6). Now let us prove that, for all $i = 1, \ldots, \bar{l}$, the sequences $\left\{\frac{\alpha(r_k)}{r_k}\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right)\right\}$

are bounded. We argue by contradiction. Since $\theta'(\cdot) \geq 0$ we can suppose without loss of generality that

$$\lim_{k\to\infty} \mu_k = +\infty \text{ with } \mu_k := \sum_{i=1}^{\bar{l}} \frac{\alpha(r_k)}{r_k}\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right).$$

Then set

$$\hat{\eta}_k = \frac{1}{\mu_k}\nabla f(x_k) + DG(x_k)^t\left(\sum_{i=1}^{\bar{l}} \xi_k^i e_i^k (e_i^k)^t\right), \quad \text{with}$$

$$\xi_k^i := \frac{\alpha(r_k)}{\mu_k r_k}\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right) \in [0, 1]. \tag{4.8}$$

Dividing (4.4) by μ_k and using (4.6) we get

$$\lim_{k\to\infty} \hat{\eta}_k = 0. \tag{4.9}$$

We can consider, passing to a subsequence if necessary, that each sequence $\{\xi_k^i\}$ converges to some $\bar{\xi}^i \in [0, 1]$. Moreover, since $\sum_{i=1}^{\bar{l}} \xi_k^i = 1$ for all k, it follows that $\sum_{i=1}^{\bar{l}} \bar{\xi}^i = 1$.

Letting $k \to +\infty$ in (4.8) and using that $e_i^k \to \bar{e}_i$, condition (4.9) implies that

$$DG(\bar{x})^t\left(\sum_{i=1}^{\bar{l}} \bar{\xi}^i \bar{e}_i(\bar{e}_i)^t\right) = 0, \tag{4.10}$$

with $\bar{\xi}^i \geq 0$ satisfying that $\sum_{i=1}^{\bar{l}} \bar{\xi}^i = 1$. We will verify that (4.10) contradicts Robinson's condition (2.16) (which is equivalent to Slater's condition). Indeed, by definition of the adjoint operator, condition (4.10) can be written as

$$\sum_{i=1}^{\bar{l}} \bar{\xi}^i \left(\bar{e}_i(\bar{e}_i)^t\right) \cdot DG(\bar{x})h = \sum_{i=1}^{\bar{l}} \bar{\xi}^i (\bar{e}_i)^t \left[DG(\bar{x})h\right] \bar{e}_i = 0, \quad \forall h \in I\!R^n. \tag{4.11}$$

Let $\bar{h}$ be the direction appearing in Robinson's condition (2.16). Since $\bar{\xi}^i \geq 0$ and $(\bar{e}_i)^t \left[DG(\bar{x})\bar{h}\right] \bar{e}_i < 0$ for all $i = 1, \ldots, \bar{l}$, we immediately get that every term of the sum in (4.11) is equal to 0, and consequently $\bar{\xi}^i = 0$ for all $i = 1, \ldots, \bar{l}$. This contradicts the equality $\sum_{i=1}^{\bar{l}} \bar{\xi}^i = 1$. Hence, we have proven that the sequences $\left\{\frac{\alpha(r_k)}{r_k}\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right)\right\}$ are bounded for all $i = 1, \ldots, \bar{l}$. This together with (4.6) implies the boundedness of $\{Y_k\}$.

Finally, let $\bar{Y}$ be a limit point of $\{Y_k\}$. Since $Y_k \succeq 0$ (because θ is nondecreasing), it directly follows that $\bar{Y} \succeq 0$. On the other hand, condition (4.4) implies that $\nabla_x L(\bar{x}, \bar{Y}) = \nabla f(\bar{x}) + DG(\bar{x})^t \bar{Y} = 0$. Furthermore, from (4.6) and since the sequences $\left\{\frac{\alpha(r_k)}{r_k}\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right)\right\}$ are bounded for all $i = 1, \ldots, \bar{l}$, it follows that $\bar{Y} = \sum_{i=1}^{\bar{l}} \bar{\delta}_i \bar{e}_i \bar{e}_i^t$ with $\bar{\delta}_i \geq 0$, which implies

$$G(\bar{x}) \cdot \bar{Y} = 0.$$

Hence $\bar{Y}$ satisfies optimality conditions (4.1). We thus conclude that $\bar{Y}$ is an optimal solution of (D). □

5 Penalty methods with two parameters

We consider again in this section the convex optimization problem (P) defined in section 3 and suppose assumptions (A_1) and (A_2). Additionally, we will also suppose

$$(A_3) \quad f_\infty(d) \geq 0, \quad \forall d.$$

It was noted in Remark 3.1 that there is no loss of generality to make such an assumption. In this section, we will only work with penalty functions θ that belong to F_2. In this way, for any real $r_k,\ \beta_k > 0$ we consider

$$p^{r_k}(x) = r_k H^{r_k}(x) = r_k \sum_{i=1}^{m} \theta\left(\frac{\lambda_i(G(x))}{r_k}\right),$$

and we define

$$\psi_k(x) = f(x) + \beta_k p^{r_k}(x).$$

The main computation of the forthcoming algorithm will be to solve approximatively, at each iteration k, the unconstrained optimization problem

$$(P_k) \quad v_k = \inf\{\psi_k(x) \mid x \in \mathbb{R}^n\}.$$

Let S_k be the optimal set of (P_k), and let $\{\epsilon_k\}$ and $\{\gamma_k\}$ be sequences such that

$$\forall k\colon\ \epsilon_k > 0,\ \gamma_k > 0, \quad \lim_{k\to\infty} \epsilon_k = \lim_{k\to\infty} \gamma_k = 0. \tag{5.1}$$

As in Theorem 3.1, we can show that S_k is nonempty and compact for each k. Hence, following the discussion of section 3, we can compute for each k a point x_k satisfying

$$||\eta_k|| \leq \epsilon_k, \quad ||\eta_k||||x_k|| \leq \gamma_k, \quad \text{where}\ \ \eta_k = \nabla\psi_k(x_k). \tag{5.2}$$

As we have seen before this can be done in a finite number of steps with any usual descent method.

Now we proceed similarly to Gonzaga and Castillo (2003). The parameters r_k and β_k play two different roles: r_k always decreases in order to improve the approximation of the function $t \to t^+$ by the mapping $t \to r_k\theta(t/r_k)$, while β_k is a penalty weight that increases only at an infeasible iteration point x_k.

The algorithm proposed in this article is the following:

1. Let $\beta_0 = r_0 = 1$ and $k = 0$.
2. Compute x_k satisfying (5.2).
3. Update $r_{k+1} = \frac{r_k}{2}$, and if x_k is feasible then set $\beta_{k+1} = \beta_k$, otherwise set $\beta_{k+1} = 2\beta_k$. Finally set $k = k + 1$.

When $C = D$ and $\epsilon_k = 0$ (that is, x_k is an exact minimizer of (P_k)), our algorithm coincides with the proposed one by Gonzaga and Castillo (2003). We refer the reader to this article for a detailed discussion of this scheme.

In addition to the hypothesis made in this section, we denote by $\{x_k\}$, $\{r_k\}$ and $\{\beta_k\}$ the sequences generated by our algorithm. In this context, the following convergence result holds.

Theorem 5.1 *The sequence $\{x_k\}$ is bounded and all its limit points are optimal solutions of* (P).

Proof We start this proof establishing five conditions that will be important in the sequel. First, by construction of the algorithm we have

$$1 \le \beta_k, \quad \beta_k r_k \le 1, \ \forall k. \tag{5.3}$$

Second, since $\lim_{t\to-\infty} \theta(t) = 0$ we obtain

$$\lim_{k\to\infty} \theta\left(\frac{\lambda_i(G(x^0))}{r_k}\right) = 0, \quad \forall i = 1, 2, \ldots, m, \ \forall x^0 \in \operatorname{int} C.$$

Consequently,

$$\lim_{k\to\infty} \beta_k p^{r_k}(x^0) = \lim_{k\to\infty} \beta_k r_k \sum_{i=1}^{m} \theta\left(\frac{\lambda_i(G(x^0))}{r_k}\right) = 0, \quad \forall x^0 \in \operatorname{int} C. \tag{5.4}$$

Third, since for all $t > 0$ the function $r \to r(\theta(t/r) - \theta(0))$ is nondecreasing on $I\!R_{++}$, and since $\theta(0) \ge 0$, we deduce that

$$r\theta(t/r) \ge \theta(t) - \theta(0), \quad \forall t \in \mathbb{R}, \ \forall r \in (0, 1]. \tag{5.5}$$

Fourth, convexity of the function ψ_k and the definition of $\eta_k := \nabla\psi_k(x_k)$ imply that

$$f(x_k) + \beta_k p^{r_k}(x_k) \le f(x^0) + \beta_k p^{r_k}(x^0) + \langle \eta_k, x_k - x^0\rangle, \quad \forall x^0 \in \operatorname{int} C. \tag{5.6}$$

Finally, since θ is nonnegative we get from (5.2)–(5.4) and (5.6)

$$f(x_k) + r_k\theta\left(\frac{\lambda_1(G(x_k))}{r_k}\right) \le f(x^0) + \mu_k(x^0),$$
$$\text{with} \quad \lim_{k\to\infty} \mu_k(x^0) = 0, \ \forall x^0 \in \operatorname{int} C. \tag{5.7}$$

Now let us show that the sequence $\{x_k\}$ is bounded. By contradiction, we can suppose, passing to a subsequence if necessary, that

$$||x_k|| \to +\infty, \quad \lim_{k\to\infty} \frac{x_k}{||x_k||} = d \ne 0.$$

By Proposition 3.1, part (ii) it follows that $(\tilde{\lambda}_1)_\infty(d) > -\infty$. Set $\alpha_1 < (\tilde{\lambda}_1)_\infty(d)$. From formula (2.3) (see Remark 2.1) we have for all k sufficiently large

$$\tilde{\lambda}_1(x_k) = \tilde{\lambda}_1\left(\frac{x_k}{||x_k||}||x_k||\right) \geq \alpha_1||x_k||.$$

This together with the monotonicity of θ and inequality (5.7) yields to

$$\frac{f(x_k)}{||x_k||} + \frac{r_k}{||x_k||}\theta\left(\frac{\alpha_1||x_k||}{r_k}\right) \leq \frac{f(x^0)}{||x_k||} + \frac{\mu_k(x^0)}{||x_k||}.$$

By passing to the liminf in this last inequality we get

$$f_\infty(d) + \theta_\infty(\alpha_1) \leq 0. \tag{5.8}$$

Since f_∞ and θ_∞ are nonnegative we obtain that $\theta_\infty(\alpha_1) = 0$, and consequently $f_\infty(d) = 0$. Furthermore, due to relations $\theta_\infty(-1) = 0$ and $\theta_\infty(1) = 1$ it follows that $\alpha_1 \leq 0$. Then letting $\alpha_1 \uparrow (\tilde{\lambda}_1)_\infty(d)$ it follows that $(\tilde{\lambda}_1)_\infty(d) \leq 0$, or equivalently, $d \in C_\infty$. This together with $f_\infty(d) \leq 0,\ d \neq 0$ contradicts the fact that S is nonempty and compact.

Let $\bar{x}$ be a limit point of the sequence $\{x_k\}$. For the sake of simplicity, we can suppose that $\bar{x} = \lim_{k\to\infty} x_k$.

Firstly, we proceed to prove that $\bar{x}$ is feasible. This is obviously true if for all k sufficiently large the iteration point x_k is feasible for problem (P), i.e. $x_k \in C$. If this is not the case, we have from the construction of the algorithm

$$\lim_{k\to\infty} \beta_k = +\infty. \tag{5.9}$$

At the first iteration, the convexity of function ψ_0 implies

$$f(x_0) + p^1(x_0) \leq f(x_k) + p^1(x_k) + \langle \eta_0, x_0 - x_k\rangle. \tag{5.10}$$

Using inequality (5.5) we get

$$r_k\theta(\lambda_i(G(x_k))/r_k) \geq \theta(\lambda_i(G(x_k))) - \theta(0), \quad \forall i = 1, \ldots, m,$$

which yields to

$$p^{r_k}(x_k) \geq p^1(x_k) - m\theta(0).$$

Adding this last inequality to (5.10) we obtain

$$f(x_0) + p^1(x_0) - m\theta(0) \leq f(x_k) + p^{r_k}(x_k) + \langle \eta_0, x_0 - x_k\rangle,$$

deducing from relation (5.6) that

$$(\beta_k - 1)p^{r_k}(x_k) \leq \beta_k r_k \sum_{i=1}^{m} \theta\left(\frac{\lambda_i(G(x^0))}{r_k}\right) + \|\eta_0\|\|x_k\| + \|\eta_k\|\|x_k - x^0\| + K,$$

where K is a constant. Hence, from the boundedness of $\{x_k\}$ and relations (5.1), (5.2) and (5.4) we can give an upper bound $\hat{K}$ for the right-hand side of the last inequality. Thus, from the fact that θ is nonnegative it follows for all k sufficiently large that

$$r_k\theta\left(\frac{\lambda_1(G(x_k))}{r_k}\right) \le \frac{\hat{K}}{(\beta_k - 1)},$$

Passing to the liminf and using formula (2.3) and (5.9) we get $\theta_\infty(\lambda_1(G(\bar{x}))) \le 0$. As a consequence we conclude that $\lambda_1(G(\bar{x})) \le 0$, that is, $\bar{x}$ is feasible for problem (P).

Secondly, we shall prove that $\bar{x}$ is an optimal solution of (P). Since $\theta(\cdot) \ge 0$ and inequality (5.7) we have

$$f(x_k) \le f(x^0) + \mu_k(x^0) \quad \forall x^0 \in \operatorname{int} C.$$

We thus obtain at the limit that $f(\bar{x}) \le f(x^0)$ for all $x^0 \in \operatorname{int} C$. Hence, continuity of function f implies that $\bar{x}$ is an optimal solution of (P). □

In the next theorem we extend to our semidefinite framework the main result of the article (cf. Gonzaga and Castillo 2003, Theorem 1). For this purpose, we denote by F_2^* the subset of functions $\theta \in F_2$ satisfying the inequality $\theta'(0) > 0$. We remark that θ_6 and θ_7 belong to F_2^*.

The following theorem says that for $\theta \in F_2^*$ and k sufficiently large, the point x_k will be feasible. This result is important for optimization problems where feasibility is a key issue. Of course, there are some examples of $\theta \in F_1$ ($-\log(x)$, $1/x, \ldots$) for which x_k is strictly feasible, but in these cases the starting point of the numerical methods used to obtain x_k must also be strictly feasible, which can be a difficult task for some problems. Thanks to the next theorem this difficulty is avoided when $\theta \in F_2^*$.

Theorem 5.2 *Suppose in addition to hypothesis of Theorem 5.1 that $\theta \in F_2^*$. Then, there exists k_0 such that for all $k \ge k_0$, x_k is feasible.*

Proof We argue by contradiction. So, since $\{x_k\}$ is bounded, we can assume the existence of a convergent but infeasible subsequence of $\{x_k\}$ (which for simplicity will be also called $\{x_k\}$). Hence, by construction of our algorithm, $\beta_k \to +\infty$. Let $\bar{x} := \lim_{k\to\infty} x_k$. It follows from Theorem 5.1 that $\bar{x}$ is an optimal solution of (P).

In the rest of this proof, we consider that k is large enough. If $G(\bar{x}) \prec 0$ then by smoothness of the function G we get $G(x_k) \prec 0$, obtaining directly a contradiction. We then suppose $\operatorname{Im} E(G(\bar{x})) = \operatorname{Ker} G(\bar{x}) \ne \{0\}$, that is, $G(\bar{x})$ is singular. By Proposition 2.4, Slater's condition (A_2) is equivalent to Robinson's condition (2.16), which can be written at $\bar{x}$ as follows

There exists $\bar{h} \in I\!R^n$ and $\rho > 0$ such that $E(G(\bar{x}))^t DG(\bar{x})\bar{h}E(G(\bar{x})) \prec -\rho I_m$,

where I_m is the identity matrix in S^m.

Hence, continuity of $DG(\cdot)$ implies that

$$E_k^t DG(x_k)\bar{h}E_k \prec -\tfrac{1}{2}\rho I_m, \tag{5.11}$$

where $E_k \in I\!R^{m\times l_0(G(\bar{x}))}$ are the matrices given by Corollary 2.1, i.e., the columns of matrices E_k are the orthonormalized eigenvectors of $G(x_k)$ associated with their

$l_0(G(\bar{x}))$ largest eigenvalues, and $E_k \to E(G(\bar{x}))$. Corollary 2.1 also tells us that $l_+(G(x_k)) \leq l_0(G(\bar{x}))$. Actually we have

$$\lambda_i(G(x_k)) \leq \mu < 0, \quad \forall\, i = l_0(G(\bar{x})) + 1, \ldots, m, \tag{5.12}$$

where $\mu > \bar{\mu} := \max\{\lambda_i\,;\, \lambda_i = \lambda_i(G(\bar{x})) < 0\}$.

We proceed to compute the inner product $\langle \eta_k, \bar{h}\rangle = \eta_k^t \bar{h}$, where $\eta_k = \nabla \psi_k(x_k)$ and $\bar{h}$ is the vector appearing in (5.11).

From the derivation rule given in Proposition 2.3, part (iii) we get

$$\begin{aligned} \nabla p^{r_k}(x_k)^t \bar{h} &= \sum_{i=1}^{m} \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right) c_i(x_k) \cdot DG(x_k)\bar{h} \\ &= \sum_{i=1}^{m} \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right) (e_i^k)^t \left(DG(x_k)\bar{h} \right) e_i^k, \end{aligned} \tag{5.13}$$

where $c_i(x_k) := e_i^k (e_i^k)^t$ and vectors e_i^k's are the columns of E_k such that each e_i^k corresponds to the eigenvector of $G(x_k)$ associated with $\lambda_i(G(x_k))$.

Condition (5.13) implies that $\langle \eta_k, \bar{h}\rangle = \nabla f(x_k)^t \bar{h} + \beta_k \sum_{i=1}^m \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right)$ $(e_i^k)^t (DG(x_k)\bar{h}) e_i^k$, which can be rewritten as

$$\begin{aligned} &-\frac{\langle \eta_k, \bar{h}\rangle}{\beta_k} + \frac{\nabla f(x_k)^t \bar{h}}{\beta_k} + \sum_{i=l_0(G(\bar{x}))+1}^{m} \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right) (e_i^k)^T \left(DG(x_k)\bar{h} \right) e_i^k \\ &= - \sum_{i=1}^{l_0(G(\bar{x}))} \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right) (e_i^k)^T \left(DG(x_k)\bar{h} \right) e_i^k. \end{aligned} \tag{5.14}$$

Taking the limit when $k \to +\infty$ we have that the terms $-\frac{\langle \eta_k, \bar{h}\rangle}{\beta_k}$ and $\frac{\nabla f(x_k)^t \bar{h}}{\beta_k}$ converge toward 0 due to relations (5.1) and (5.2), and $\beta_k \to +\infty$. By (5.12), we obtain $\lambda_i(G(x_k)/r_k) \to -\infty$ for all $i = l_0(G(\bar{x})) + 1, \ldots, m$. This together with the limit $\lim_{t \to -\infty} \theta'(t) = 0$ implies that $\theta'(\lambda_i(G(x_k)/r_k)) \to 0$ for all $i = l_0(G(\bar{x})) + 1, \ldots, m$. Then we deduce that the entire left hand side of (5.14) converges toward 0.

We will obtain a contradiction by showing that the right hand side of (5.14) is strictly positive. Indeed, condition (5.11) implies that $(e_i^k)^t (DG(x_k)\bar{h}) e_i^k < -\rho/2$ for $i = 1, \ldots, l_0(G(\bar{x}))$, and since θ is nondecreasing, $\theta'(\cdot) \geq 0$ and $l_+(G(x_k)) \leq l_0(G(\bar{x}))$ it follows that

$$\begin{aligned} &- \sum_{i=1}^{l_0(G(\bar{x}))} \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right) (e_i^k)^t \left(DG(x_k)\bar{h} \right) e_i^k \geq \frac{\rho}{2} \sum_{i=1}^{l_0(G(\bar{x}))} \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right) \\ &\geq \frac{\rho}{2} \sum_{i=1}^{l_+(G(x_k))} \theta' \left(\frac{\lambda_i(G(x_k))}{r_k} \right) \geq \frac{\rho}{2} \theta'(0) l_+(x_k). \end{aligned}$$

But $\theta'(0) > 0$ (because $\theta \in F_2^*$) and x_k is infeasible, i.e. $l_+(x_k) \geq 1$. Hence the right-hand side of (5.14) has a strictly positive lower bound. The theorem follows. □

As for penalty and barrier methods with one parameter we can associate with the sequence $\{x_k\}$ a sequence $\{Y_k\}$ of dual multipliers defined by

$$Y_k = \beta_k(\theta')^0\left(\frac{G(x_k)}{r_k}\right) = \beta_k \sum_{i=1}^{m} \theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right) e_i^k (e_i^k)^t, \tag{5.15}$$

where $(\theta')^0$ is the matrix function associated with θ', defined in (2.10), and e_i^k's are orthonormal eigenvectors of $G(x_k)$ associated with the eigenvalues $\lambda_i(G(x_k))$. Then we have

$$\eta_k = \nabla\psi_k(x_k) = \nabla f(x_k) + DG(x_k)^t Y_k. \tag{5.16}$$

As in section 4, we prove in the next thorem that the sequence $\{Y_k\}$ is bounded with each of its limit points being an optimal solution of (D).

Theorem 5.3 *Suppose that the assumptions of Theorem 5.2 are satisfied. Consider a sequence $\{x_k\}$ satisfying relations* (5.2)*, and let $\{Y_k\}$ be the sequence defined by formula* (5.15)*. Then, $\{Y_k\}$ is bounded and each of its limit points is an optimal solution of* (D).

Proof By Theorems 5.1 and 5.2 we can assume, without loss of generality, that the sequence $\{x_k\}$ converges to an optimal solution $\bar{x}$ of (P) and for k sufficiently large x_k is feasible and $\beta_k = \beta \geq 1$. Since x_k is feasible and by the monotonicity of $\theta'(\cdot)$, we have that $\theta'(\lambda_i(G(x_k))/r_k) \leq \theta'(0)$ for all i, which proves that the sequence $\{Y_k\}$ is bounded.
Let $\bar{Y}$ be a limit point of $\{Y_k\}$. The proof is now similar to the one given in Theorem 4.1. Since $Y_k \succeq 0$, it directly follows that $\bar{Y} \succeq 0$. On the other hand, condition (5.16) implies that $\nabla_x L(\bar{x}, \bar{Y}) = \nabla f(\bar{x}) + DG(\bar{x})^t \bar{Y} = 0$.

Let $\bar{l} := l_0(G(\bar{x}))$ be the number of null eigenvalues of $G(\bar{x})$. Since $\lim_{t\to-\infty} \theta'(t) = 0$ (cf. (4.5)), we get

$$\beta_k\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right) \to 0 \quad \forall i = \bar{l}+1, \ldots, m, \tag{5.17}$$

and since the sequences $\left\{\beta_k\theta'\left(\frac{\lambda_i(G(x_k))}{r_k}\right)\right\}$ are bounded for all $i = 1, \ldots, \bar{l}$, it follows that $\bar{Y} = \sum_{i=1}^{\bar{l}} \bar{\delta}_i \bar{e}_i \bar{e}_i^t$ with $\bar{\delta}_i \geq 0$, which implies that $G(\bar{x}) \cdot \bar{Y} = 0$. Hence $\bar{Y}$ satisfies optimality conditions (4.1). We thus conclude that $\bar{Y}$ is an optimal solution of (D). □

Acknowledgements Héctor Ramírez was Partially supported by Ecos-Conicyt C00E05.

References

Auslender A (1999) Penalty and barrier methods: a unified framework. SIAM J Optim 10:211–230

Auslender A (2003) Variational inequalities over the cone of semidefinite positive matrices and over the Lorentz cone. Optimization methods and software, pp 1–18

Auslender A, Teboulle M (2003) Asymptotic cones and functions in optimization and variational inequalities. Springer monographs in mathematics, Springer, Berlin Heidelberg New York
Auslender A, Cominetti R, Haddou M (1997) Asymptotic analysis of penalty and barrier methods in convex and linear programming. Math Opers Res 22:43–62
Bhatia R (1997) Matrix analysis. Springer graduate texts in mathematics, Springer, Berlin Heidelberg New York
Bonnans JF, Shapiro A (2000) Perturbation analysis of optimization problems. Springer series in operations research, Springer, Berlin Heidelberg New York
Ben-Tal A, Nemirowskii A (2002) Lectures on modern convex optimization, analysis, algorithms, and engineering applications. MPS-SIAM series on optimization, SIAM, Philadelphia
Ben-Tal A, Zibulevsky M (1997) Penalty-barrier methods for convex programming problems. SIAM J Optim 7:347–366
Chen C, Mangasarian OL (1996) A class of smoothing functions for nonlinear and mixed complementary problems. Compt Optim Appl 5:97–138
Cominetti R, Dussault JP (1994) A stable exponential penalty method with superlinear convergence. JOTA 83:285–309
Den Hertzog D, Roos C, Terlaky T (1991) Inverse barrier method for linear programming. Report 91–27, Faculty of Technical Mathematics and Informatics, Delft University of Technology, Netherlands
Frisch KR (1995) The logarithmic potential method of convex programming. Memorandum of May 13, 1995, University Institute of Economics, Oslo, Norway
Golub GH, Van Loan CF (1996) Matrix computations, 3rd edn, The Johns Hopkins University Press, Baltimore
Gonzaga C, Castillo RA (2003) A nonlinear programming algorithm based on non-coercive penalty functions. Math Program Ser A 96:87–101
Graña Drummond LM, Peterzil Y (2002) The central path in smooth convex semidefinite programming. Optimization 51:207–233
Kato T (1970) Perturbation theory for linear operators. Springer, Berlin Heidelberg New York
Lewis AS (1996) Convex analysis on the Hermitian matrices. SIAM J Optim 6:164–177
Nesterov YN, Nemirovski AS (1994) Interior point polynomial algorithms in convex programing. SIAM, Philapdelphia
Polyak RA (1992) Modified barrier functions (theory and methods). Math Prog 54:177–222
Rockafellar RT (1970) Convex analysis. Princeton University Press, Princeton
Seeger A (1997) Convex analysis of spectrally defined matrix functions. SIAM J Optim 7:679–696
Vandenbergue L, Boyd S (1995) Semidefinite programming. SIAM Rev 38:49–95
Xavier AE (1992) Hyperbolic penalization, PhD thesis, COPPE - Federal University of Rio de Janeiro, Rio de Janeiro, Brazil (in Portuguese)