RAIRO-Oper. Res. 42 (2008) 123–139
DOI: 10.1051/ro:2008005

RAIRO Operations Research
www.rairo-ro.org

A LOGARITHM BARRIER METHOD FOR SEMI-DEFINITE PROGRAMMING

Jean-Pierre Crouzeix[1] and Bachir Merikhi[2]

Abstract. This paper presents a logarithmic barrier method for solving a semi-definite linear program. The descent direction is the classical Newton direction. We propose alternative ways to determine the step-size along the direction which are more efficient than classical line-searches.

Keywords. Linear semi-definite programming, barrier methods, line-search.

Mathematics Subject Classification. 90C22, 90C05, 90C51.

1. Introduction

In this paper we present an algorithm for solving the optimization problem:

$$m_d = \inf_y \left[b^t y \;:\; \sum_{i=1}^{m} y_i A_i - C \;\in K,\; y \in \mathbb{R}^m \right], \tag{D}$$

where K denotes the cone of $n \times n$ symmetric positive semi-definite matrices, the vector $b \in \mathbb{R}^m$ and the $n \times n$ symmetric matrices C and $A_i, i = 1, \ldots, m$, are given. The dual problem of (D) is:

$$m_p = \max_X [\langle C, X\rangle : X \in K,\; \langle A_i, X\rangle = b_i \;\forall\, i = 1, \ldots, m], \tag{P}$$

Received September 01, 2006. Accepted September 01, 2006.

[1] LIMOS, Université Blaise Pascal, Campus des Cézaux, 63174 Aubière, France; jp.crouzeix@isima.fr

[2] Laboratoire d'optimisation, Université Ferhat Abbas, Algérie; the research of this author has been made possible thanks to a PROFAS grant and the hospitality of Université Blaise Pascal. b_merikhi@yahoo.fr

Article published by EDP Sciences

where by $\langle C, X\rangle$ we denote the trace of the matrix (C^tX). It is recalled that $\langle\cdot,\cdot\rangle$ corresponds to an inner product on the space of $n \times n$ matrices.

These problems are linear. Their feasible sets involving the cone of positive semi-definite matrices, a non polyhedral convex cone, they are called linear semi-definite programs. Such problems are the object of a particular attention since the papers by Alizadeh [1,2], as well on a theoretical or an algorithmical aspect, see for instance the following references [1–4,6,7].

Under suitable conditions, solving (D) is equivalent to solving (P): the optimal solutions of one problem being easily obtained when one optimal solution of the other problem is known. In this paper, the problem (D) is approximated by the problem (D_r), $(r > 0)$,

$$m(r) = \inf\ [\ f_r(y)\ :\ y \in \mathbb{R}^m], \tag{D_r}$$

where the barrier function $f_r : \mathbb{R}^m \to (-\infty, +\infty]$ is defined by

$$f_r(y) = \begin{cases} b^t y + nr\ln r - r\ln[\det(\sum_{i=1}^m y_i A_i - C)] & \text{if} \quad y \in \widehat{Y}, \\ +\infty & \text{if not,} \end{cases}$$

with

$$\widehat{Y} = \left\{ y \in \mathbb{R}^m : \text{the matrix } \sum_{i=1}^m y_i A_i - C \in \widehat{K} \right\},$$

and $\widehat{K} = \operatorname{int}(K)$ is the cone of $n \times n$ symmetric positive definite matrices. This problem is solved via a classical Newton descent method. The difficulty is in the line-search: the presence of a determinant in the definition of f_r induces high computational costs in classical exact or approximate line-searches. Here, instead of minimizing f_r along the descent direction d at the current point x, we minimize a function $\widetilde{\theta}$ such that

$$\frac{1}{r}\,[f_r(x+td) - f_r(x)] = \theta(t) \le \widetilde{\theta}(t) \quad \forall\, t > 0, \quad \theta(0) = \widetilde{\theta}(0), \quad \theta'(0) = \widetilde{\theta}'(0) < 0.$$

This function $\widetilde{\theta}$ needs to be appropriately chosen so that the optimal t is easily obtained and to be close enough to θ in order to give a significant decrease of f_r in the iteration step. We propose in this paper functions θ for which the optimal solution t is explicitly obtained and a good quality of the approximation of θ by $\widetilde{\theta}$ is ensured by the condition $\theta''(0) = \widetilde{\theta}''(0)$.

In the next section, we briefly recall some results in linear semi-definite programming. Section 3 studies the problem (D_r), in particular the behavior of its optimal value and its optimal solutions when $r \to 0$. Section 4 shows how to compute the Newton descent direction. Section 6 is devoted to the determination of efficient approximations $\widetilde{\theta}$, these approximations are deduced from inequalities shown in Section 5. The algorithm is resumed in Section 7 and numerical experiments presented in Section 8 show the efficiency of the approximations when compared with classical line-searches.

2. A brief background in linear semi-definite programming

Throughout the paper, we use the following notation:

$$Y = \{y \in \mathbb{R}^m : \textstyle\sum_{i=1}^m y_i A_i - C \in K\}, \quad F = \{X \in K : \langle A_i, X\rangle = b_i \; \forall\, i\},$$
$$\widehat{Y} = \{y \in \mathbb{R}^m : \textstyle\sum_{i=1}^m y_i A_i - C \in \widehat{K}\}, \quad \widehat{F} = \{X \in F \,:\, X \in \widehat{K}\}.$$

It is easily seen that $-\infty \le m_p \le m_d \le +\infty$ (weak duality). In this paper we assume that the two following assumptions hold:

(H1) The system of equations $\langle A_i, X\rangle = b_i$, $i = 1, ..., m$ is of rank m.
(H2) The sets $\widehat{Y}$ and $\widehat{F}$ are non empty.

Then it is known that (see for instance [1,3]):

(a) $-\infty < m_p = m_d < +\infty$.
(b) The sets of optimal solutions of (P) and (D) are non empty convex compact sets.
(c) If $\bar{X}$ is an optimal solution of (P), then $\bar{y}$ is an optimal solution of (D) if and only if

$$\bar{y} \in Y \quad \text{and} \quad \left(\sum_{i=1}^m \bar{y}_i A_i - C\right)\bar{X} = 0.$$

(d) If $\bar{y}$ is an optimal solution of (D), then $\bar{X}$ is an optimal solution of (P) if and only if

$$\bar{X} \in F \quad \text{and} \quad \left(\sum_{i=1}^m \bar{y}_i A_i - C\right)\bar{X} = 0.$$

3. The problem (D_r): theoretical aspects

Recall that (D_r), $r > 0$, is the problem

$$m(r) = \inf\,[\, f_r(y) \,:\, y \in \mathbb{R}^m], \qquad (D_r)$$

with $f_r : \mathbb{R}^m \to (-\infty, +\infty]$ defined by

$$f_r(y) = \begin{cases} b^t y + nr\ln r - r\ln\left[\det(\sum_{i=1}^m y_i A_i - C)\right] & \text{if } y \in \widehat{Y}, \\ +\infty & \text{if not.} \end{cases}$$

We start with the study of this function.

3.1. f_r is a twice differentiable strictly convex function

The following notation will be used in the expressions of the gradient and the Hessian of f_r: given $y \in \widehat{Y}$, we introduce the $m \times m$ symmetric positive definite matrix $B(y)$ and the lower triangular $m \times m$ matrix $L(y)$ such that

$$B(y) = \sum_{i=1}^{m} y_i A_i - C = L(y) L^t(y).$$

Next, for $i, j = 1, 2, \cdots, m$, we define

$$\widehat{A}_i(y) = [L(y)]^{-1} A_i [L^t(y)]^{-1},$$

$$b_i(y) = \mathrm{trace}(\widehat{A}_i(y)) = \mathrm{trace}(A_i B^{-1}(y)),$$
$$\Delta_{ij}(y) = \mathrm{trace}(B^{-1}(y) A_i B^{-1}(y) A_j) = \mathrm{trace}(\widehat{A}_i(y)\widehat{A}_j(y)).$$

Thus $b(y)$ is a vector of $\mathbb{R}^m$ and $\Delta(y)$ is a symmetric $m \times m$ matrix

Theorem 1. *The function f_r is twice continuously differentiable on $\widehat{Y}$. Actually, for all $y \in \widehat{Y}$ we have:*
(a) $\nabla f_r(y) = b - r b(y)$;
(b) $\nabla^2 f_r(y) = r\Delta(y)$;
(c) *the matrix $\Delta(y)$ is definite positive.*

Proof. (a) Denote by $(e_1, e_2, \cdots, e_m)$ the canonical basis of $\mathbb{R}^m$. Let $i \in \{1, \cdots, m\}$ and $z_i \in \mathbb{R}$, $z_i \neq 0$. Then,

$$\begin{aligned}
\frac{f_r(y + z_i e_i) - f_r(y)}{z_i} &= b_i - \frac{r}{z_i}[\ln\det(B(y + z_i e_i)) - \ln\det(B(y))], \\
&= b_i - \frac{r}{z_i}[\ln\det(L(y)[I + z_i\widehat{A}_i]L^t(y)) - \ln\det(B(y))], \\
&= b_i - \frac{r}{z_i}\ln\det(I + z_i\widehat{A}_i(y)), \\
&= b_i - \frac{r}{z_i}\ln[1 + z_i \mathrm{trace}(\widehat{A}_i(y)) + z_i\varepsilon(z_i)]
\end{aligned}$$

where the function ε is such that $\varepsilon(z) \to 0$ when $z \to 0$. Pass to the limit when $z_i \to 0$.

(b) In the same manner, given $i, j \in \{1, \cdots, m\}$, let us consider

$$\frac{b_i(y + z_j e_j) - b_i(y)}{z_j} = \frac{-1}{z_j}[\mathrm{trace}(A_i[B^{-1}(y + z_j e_j) - B^{-1}(y)])].$$

But,

$$\begin{aligned}
B^{-1}(y + z_j e_j) - B^{-1}(y) &= [B(y) + z_j A_j]^{-1} - B^{-1}(y), \\
&= [B(y)(I + z_j B^{-1}(y) A_j)]^{-1} - B^{-1}(y), \\
&= [(I + z_j B^{-1}(y) A_j)^{-1} - I] B^{-1}(y).
\end{aligned}$$

Neglecting the second order terms in z_j, we obtain

$$\frac{b_i(y+z_je_j)-b_i(y)}{z_j} \sim \mathrm{trace}(A_iB^{-1}(y)A_jB^{-1}(y)).$$

Pass to the limit when $z_j \to 0$. On the other hand the equality

$$\mathrm{trace}(B^{-1}(y)A_iB^{-1}(y)A_j) = \mathrm{trace}(\widehat{A}_i(y)\widehat{A}_j(y))$$

is immediate.

(c) Let $d \neq 0$. Next, let $M = \sum_{i=1}^m d_i\widehat{A}_i(y)$. Then (H1) implies $M \neq 0$. On the other hand,

$$\langle \nabla^2 f_r(y)d, d\rangle = r\mathrm{trace}\left(\sum_{i,j} d_id_j\widehat{A}_i(y)\widehat{A}_j(y)\right) = r\mathrm{trace}(M^2) > 0,$$

from what we deduce that the matrix $\nabla^2 f_r(y)$ is positive definite. □

Since f_r is strictly convex, (D_r) has at most one optimal solution.

3.2. (D_r) has one unique optimal solution

Because the convex function f_r takes the value $+\infty$ on the boundary of its domain and is differentiable on the interior, it is lower semi-continuous. In order to prove that (D_r) has one optimal solution, it suffices to prove that the recession cone of f_r is reduced to the origin. Before that, we show the following result:

Proposition 1. *$d = 0$ whenever $b^td \leq 0$ and $\sum_{i=1}^m d_iA_i \in K$.*

Proof. Assume that $d \neq 0$, $b^td \leq 0$ and $C = \sum_{i=1}^m d_iA_i \in K$. Then (H1) implies $C \neq 0$. Let some $\widehat{X} \in \widehat{F} \subset \widehat{K}$, such $\widehat{X}$ exists in view of assumption (H2). Then,

$$0 < \langle C, \widehat{X}\rangle = \sum_{i=1}^m d_i\langle A_i, \widehat{X}\rangle = b^td.$$

The proposition is proved. □

Theorem 2. *$d = 0$ if $(f_r)_\infty(d) \leq 0$.*

Proof. Fix some $y \in \widehat{Y}$, such y exists in view of assumption (H2). The recession function $(f_r)_\infty$ of f_r is defined as

$$(f_r)_\infty(d) = \lim_{t\to+\infty}\left[\xi(t) = \frac{f_r(y+td)-f_r(y)}{t}\right].$$

Let $B = B(y) = \sum_{i=1}^m y_iA_i - C$, B is a positive definite symmetric matrix, there exists a non singular lower triangular matrix L such that $B = LL^t$. Given d, set

$H(d) = \sum_{i=1}^m d_i A_i$. Then, for any t such that the matrix $B + tH(d)$ is positive definite,

$$\begin{aligned} \xi(t) &= b^t d - rt^{-1}[\ln\det(B + tH(d)) - [\ln\det(B)), \\ &= b^t d - rt^{-1}[\ln\det(I + tE(d)] \end{aligned}$$

where $E(d) = L^{-1}H(d)(L^{-1})^t$. We deduce that ,

$$\xi(t) = \begin{cases} b^t d - rt^{-1}\ln\det(I + tE(d)) & \text{if } I + tE(d) \in \widehat{K}, \\ +\infty & \text{otherwise.} \end{cases}$$

The condition $[f_r]_\infty(d) \le 0$ is therefore equivalent to say that $H(d)$ is positive semi-definite (hence $E(d)$ is also positive definite) and

$$b^t d \le r \lim_{t\to\infty} \frac{1}{t}\ln\det(I + tE(d)) = r \lim_{t\to\infty} \sum_{i=1}^n \frac{1}{t}\ln(1 + t\lambda_i(d)) = 0,$$

where by $\lambda_i(d)$ we denote the eigenvalues of $E(d)$. Pass to the limit and apply Proposition 1. □

We denote by $y(r)$ or y_r the unique optimal solution of (D_r).

3.3. When $r \to 0$

Next, we turn our interest in the behavior of the optimal value $m(r)$ and the optimal solution $y(r)$ of (D_r) for $r \to 0$. For that, let us introduce the function $h : \mathbb{R}^m \times \mathbb{R} \to (-\infty, +\infty]$ defined by

$$h(y,t) = \begin{cases} b^t y - \ln\det\left[\sum_{i=1}^m y_i A_i - tC\right] & \text{if } \sum_{i=1}^m y_i A_i - tC \in \widehat{K}, \\ +\infty & \text{otherwise.} \end{cases}$$

It is easily shown that h is convex and lower semi-continuous. Next, consider the function $\phi : \mathbb{R}^m \times \mathbb{R} \times \mathbb{R} \to (-\infty, +\infty]$ defined by

$$\phi(y,t,r) = \begin{cases} rh(r^{-1}y, r^{-1}t) & \text{if } r > 0, \\ h_\infty(y,t) & \text{if } r = 0, \\ +\infty & \text{if } r < 0. \end{cases}$$

Then, ϕ is also lower semi-continuous and convex, see for instance Rockafellar [8]. Next, define $f : \mathbb{R}^m \times \mathbb{R} \to (-\infty, +\infty]$ by

$$f(y,r) = \phi(y,1,r)$$

f is also convex and lower semi-continuous. By construction,

$$f(y,r) = \begin{cases} f_r(y) & \text{if} \quad r > 0, \\ b^t y & \text{if} \quad r = 0,\ y \in Y, \\ +\infty & \text{otherwise.} \end{cases} \tag{2}$$

Define $m : \mathbb{R} \to (-\infty, +\infty]$ by $m(r) = \inf_y [f(y,r) : y \in \mathbb{R}^m]$. This function is convex. Furthermore $m(0) = m_d$ and $m(r)$ is the optimal value of (D_r) when $r > 0$. It is clear that for $r > 0$

$$m(r) = f_r(y(r)) = f(y(r), r)$$

and

$$0 = \nabla f_r(y(r)) = \nabla_y f(y(r), r) = b - rb(y_r).$$

Theorem 3. *The functions m and y are continuously differentiable on $(0, +\infty)$. We have, for all $r > 0$,*

$$r\Delta(y_r)y'(r) - b(y_r) = 0,$$
$$m'(r) = n + n\ln(r) - \ln(\det(B(y_r)).$$

Moreover,

$$m_d = m(0) \le b^t y(r) \le m_d + nr. \tag{3}$$

Proof. Let $\bar{r} > 0$, $\nabla_y f(y(\bar{r}), \bar{r}) = 0$ because $y(\bar{r})$ is an optimal solution of $(D_{\bar{r}})$. The function f is twice continuously differentiable on $\widehat{Y} \times]0, +\infty[$ and the matrix $\nabla^2_{yy} f(y(\bar{r}), \bar{r})$ is positive definite. Applying the implicit function theorem to the equation $0 = T(y,r) = \nabla_y f(y,r)$ at the point $(y(\bar{r}), \bar{r})$ we deduce that in a neighborhood of $\bar{r}$ the function y is continuously differentiable and

$$\nabla^2_{yy} f(y_r, r) y'(r) - b(y_r) = 0.$$

Since $m(r) = f(y(r), r)$ and y is continuously differentiable on $(0, \infty)$, m is also continuously differentiable and

$$\begin{aligned} m'(r) &= f'_y(y(r), r)y'(r) + f'_r(y(r), r) \\ &= n + n\ln(r) - \ln[\det(B(y_r)] \end{aligned}$$

because $f'_y(y(r), r) = [\nabla_y f(y(r), r)]^t = 0$. Next, because the function m is convex,

$$m(0) \ge m(r) + (0 - r)m'(r)$$

from what we obtain,

$$+\infty > m_d = m(0) \ge b^t y(r) - nr > -\infty.$$

On the other hand $y_r \in \widehat{Y} \subset Y$ and therefore $b^t y(r) \ge m_d$. □

Let us denote by S_d the set of optimal solutions of (D), we know that this set is closed convex bounded and not empty. The distance of a point y to the set S_D is defined as usual by

$$d(y, S_d) = \inf_z [\, \|y - z\| \,:\, z \in S_d \,].$$

The following result concerns the behavior of y_r and $m(r)$ when $r \to 0$.

Theorem 4. *Assume that $r \to 0$, then $d(y_r, S_d) \to 0$ and $m(r) \to m_d$.*

Proof. Let us consider the multivalued map S defined on $\mathbb{R}$ by

$$S(r) = \{y \in Y \,:\, b^t y \leq m_d + nr\}.$$

Its graph is closed, $S(r) = \emptyset$ if $r < 0$. If $r > 0$, $\emptyset \neq S(0) = S_d \subset S(r)$, $y_r \in S(r)$ and $S(r)$ is a closed convex set. The recession cone of $S(r)$, $r > 0$, coincides with the recession cone of $S(0)$. Hence $S(r)$ is compact because $S(r)$ is so. We deduce that the multivalued map S is upper semi-continuous (USC) on $[0, +\infty)$ because compact-valued with a closed graph. Since $y_r \in S(r)$, then $d(y_r, S(0)) \to 0$ when $r \to 0$.

It remains to prove that $m(r) \to m_d = m(0)$ when $r \to 0$. Since the function m is convex, it is enough to prove that it is lower semi-continuous at 0. We proceed by contradiction, if not there exist $\lambda < m_d$ and a sequence $\{r_k\}$ of positive numbers converging to 0 such that $m(r_k) < \lambda$. Let $y_k = y(r_k)$. Then there exist $\bar{y} \in S_d$ and a sub-sequence $\{y_{k_l}\}$ converging to $\bar{y}$. Since the function f is lower semi-continuous on $\mathbb{R}^m \times \mathbb{R}$ and $f(\bar{y}, 0) = m_d > \lambda$, one has for l large enough

$$\lambda > m(r_{k_l}) = f(y_{k_l}, r_{k_l}) > \lambda$$

which is not possible. □

4. The Newton descent direction and the line-search

Due to the presence of the barrier function, the problem (D_r) can be considered as unconstrained. This problem will be solved via a classical descent algorithm. Because the function f_r takes the value ∞ on the boundary of Y, the iterates will stay in $\widehat{Y}$. Thus, the method that we propose is an interior point method.

Assume that our current iterate is $y \in \widehat{Y}$. For descent direction d at y, we take the solution of the linear system

$$[\nabla^2 f_r(y)]d = -\nabla f_r(y).$$

According to Theorem 1, the linear system is equivalent to the system

$$\Delta(y)d = b(y) - \frac{1}{r}b \qquad (1)$$

with $B(y)$, $b(y)$ and $\Delta(y)$ defined as in Section 3.1. The matrix $\Delta(y)$ being definite positive, the linear system (1) can be efficiently solved via a Cholewsky decomposition. Of course, we assume $\nabla f_r(y) \neq 0$ (if not the optimum is reached). It follows that $d \neq 0$.

The next step in the algorithm consists in the choice of $\bar{t} > 0$ giving a significant decrease of the function f_r on the half line $y + td$, $t > 0$. Then, the next iterate will be taken equal to $y + \bar{t}d$. To do that, we consider the function

$$\begin{aligned}\theta(t) &= \frac{1}{r}[f_r(y+td) - f_r(y)], \quad y + td \in \widehat{Y}, \\ \theta(t) &= \frac{1}{r}b^t d - \ln\det(B(y+td)) + \ln\det(B(y)).\end{aligned}$$

Since $\nabla^2 f_r(y)d = -\nabla f_r(y)$ one has

$$d^t\nabla^2 f_r(y)d = -d^t\nabla f_r(y) = d^t b(y) - rd^t b.$$

In order to simplify the notation, y and d staying fixed in the following, we set

$$B = B(y) = \sum_{i=1}^{m} y_i A_i - C \quad \text{and} \quad H = \sum_{i=1}^{m} d_i A_i.$$

Since B is symmetric and positive definite, there exists a lower triangular matrix L such that $B = LL^t$. Next, we set

$$E = L^{-1}H[L^{-1}]^t.$$

Since $d \neq 0$, assumption (H1) implies $H \neq 0$ from what we have $E \neq 0$.

With this notation, for all $t > 0$ such that $I + tE$ is positive definite,

$$\theta(t) = t[\text{trace}(E) - \text{trace}(E^2)] - \ln\det(I + tE). \tag{2}$$

Denote by λ_i the eigenvalues of the symmetric matrix E, then

$$\theta(t) = \sum_{i=1}^{n}[t(\lambda_i - \lambda_i^2) - \ln(1 + t\lambda_i)], \quad t \in [0, \widehat{t}) \tag{3}$$

where

$$\widehat{t} = \sup[\, t \,:\, 1 + t\lambda_i > 0 \text{ for all } i\,] = \sup[\, t \,:\, y + td \in \widehat{Y}]. \tag{4}$$

Observe that $\widehat{t} = +\infty$ if E is positive semi-definite and $0 < \widehat{t} < +\infty$ if not. It is clear that θ is convex on $[0, \widehat{t}[$, $\theta(0) = 0$ and

$$0 < \sum \lambda_i^2 = \theta''(0) = -\theta'(0).$$

Also $\theta(t) \to +\infty$ when $t \to \widehat{t}$. It follows that there exists one unique t_{opt} such that $\theta'(t_{opt}) = 0$, θ reaches its minimum in this point.

Unfortunately, there is no explicit formula giving t_{opt} and solving the equation by iterative methods needs successive computations of the functions θ and θ'. These computations have a high numerical cost because the expression of θ in (2) contains a determinant not easily handled and (3) needs the knowledge of the eigenvalues of E, a difficult numerical problem. This leads to think of alternative approaches.

Once E is computed, it is easy to compute the two following quantities

$$\operatorname{trace}(E) = \sum_i e_{ii} = \sum_i \lambda_i \quad \text{and} \quad \operatorname{trace}(E^2) = \sum_{i,j} e_{ij}^2 = \sum_i \lambda_i^2.$$

In Section 6, we take advantage of these data to propose lower bounds of $\widehat{t}$ and functions bounded from below by θ. Before, we look at some useful inequalities on a sample of numbers when the sum of the numbers and the sum of their squares are known.

5. Some useful inequalities

As usual in statistics, given a sample of n real numbers $x_1, x_2, \ldots, x_n$, we consider their arithmetic mean $\bar{x}$ and their standard deviation σ_x. These quantities are defined as follows:

$$\bar{x} = \frac{1}{n}\sum x_i \quad \text{and} \quad \sigma_x^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 = \frac{1}{n}\sum (x_i - \bar{x})^2.$$

The following result is due to Wolkowicz-Styan [10], see also Crouzeix-Seeger [5] for additional results.

Proposition 2.

$$\bar{x} - \sigma_x\sqrt{n-1} \le \min_i x_i \le \bar{x} - \frac{\sigma_x}{\sqrt{n-1}},$$

$$\bar{x} + \frac{\sigma_x}{\sqrt{n-1}} \le \max_i x_i \le \bar{x} + \sigma_x\sqrt{n-1}.$$

In the particular case where all x_i are positive, one deduces

$$n\ln(\bar{x} - \sigma_x\sqrt{n-1}) \le \sum_{i=1}^n \ln(x_i) \le n\ln(\bar{x} + \sigma_x\sqrt{n-1}),$$

where, by convention, $\ln(t) = -\infty$ if $t \le 0$. The next result is still better.

Theorem 5. *Assume that $x_i > 0$ for $i = 1, 2, \cdots, n$. Then*

$$n\ln(\bar{x} - \sigma_x\sqrt{n-1}) \le A \le \sum_{i=1}^n \ln(x_i) \le B \le n\ln(\bar{x}),$$

with

$$A = (n-1)\ln\left(\bar{x} + \frac{\sigma_x}{\sqrt{n-1}}\right) + \ln(\bar{x} - \sigma_x\sqrt{n-1}),$$

and

$$B = \ln(\bar{x} + \sigma_x\sqrt{n-1}) + (n-1)\ln\left(\bar{x} - \frac{\sigma_x}{\sqrt{n-1}}\right).$$

Proof. If $\sigma_x = 0$, then $x_i = \bar{x}$ for all i and the inequalities hold. Assume $\sigma_x > 0$. Let us consider the two following problems where $\bar{x}$ and σ_x are fixed,

$$A = \inf_x \left[\sum_{i=1}^{n} \ln(x_i) \,:\, \sum_{i=1}^{n}(x_i - \bar{x}) = 0 \quad \text{and} \quad \sum_{i=1}^{n}(x_i - \bar{x})^2 = n\sigma_x^2\right],$$

$$B = \sup_x \left[\sum_{i=1}^{n} \ln(x_i) \,:\, \sum_{i=1}^{n}(x_i - \bar{x}) = 0 \quad \text{and} \quad \sum_{i=1}^{n}(x_i - \bar{x})^2 = n\sigma_x^2\right].$$

The second problem has always optimal solutions, the first problem has optimal solutions if $\bar{x} - \sigma_x\sqrt{n-1} > 0$ because Proposition 2 and in the other case $-\infty = A < \sum \ln(x_i)$.

Apply the first order necessary optimality condition: if x is optimal solution of one problem or the other one, there exist α and β such for all i

$$(x_i - \bar{x})^2 - \alpha(x_i - \bar{x}) + \beta = 0.$$

Thus each $(x_i - \bar{x})$ is a root of the equation

$$w^2 - \alpha w + \beta = 0.$$

Denote by a and b the two roots of this equation. The quantities $(x_i - \bar{x})$ divide into two parts, p equal to a, $n - p$ equal to b. From $\sigma_x \neq 0$ we deduce that $1 \leq p \leq n-1$ and $a \neq b$. Hence,

$$0 = \sum_i (x_i - \bar{x}) = pa + (n-p)b,$$

and

$$n\sigma_x^2 = \sum_i (x_i - \bar{x})^2 = pa^2 + (n-p)b^2.$$

From what we deduce that either

$$a = \bar{x} + \sigma_x\sqrt{\frac{n-p}{p}} \quad \text{and} \quad b = \bar{x} - \sigma_x\sqrt{\frac{p}{n-p}}$$

or

$$a = \bar{x} - \sigma_x\sqrt{\frac{n-p}{p}} \quad \text{and} \quad b = \bar{x} + \sigma_x\sqrt{\frac{p}{n-p}}.$$

Denote by $h(p)$ and $k(p)$ the following quantities

$$h(p) = \frac{p}{n} \ln \left[\bar{x} + \sigma_x \sqrt{\frac{n-p}{p}} \right] + \frac{n-p}{n} \ln \left[\bar{x} - \sigma_x \sqrt{\frac{p}{n-p}} \right],$$

$$k(p) = \frac{p}{n} \ln \left[\bar{x} - \sigma_x \sqrt{\frac{n-p}{p}} \right] + \frac{n-p}{n} \ln \left[\bar{x} + \sigma_x \sqrt{\frac{p}{n-p}} \right].$$

Then,

$$\frac{A}{n} = \min_{p=1,\cdots,n-1} [\min[h(p), k(p)]] \quad \text{and} \quad \frac{B}{n} = \max_{p=1,\cdots,n-1} [\max[h(p), k(p)]].$$

But $h(p) = k(n-p)$ for any $p = 1, ..., n-1$ and therefore

$$\frac{A}{n} = \min_{p=1,\cdots,n-1} [h(p)] \quad \text{and} \quad \frac{B}{n} = \max_{p=1,\cdots,n-1} [h(p)]. \tag{5}$$

It is interesting to set $t(p) = \sqrt{\frac{p}{n-p}}$ and to consider the function

$$\gamma(t) = \frac{t^2}{t^2+1} \ln[\bar{x} + \sigma_x t^{-1}] + \frac{1}{t^2+1} \ln[\bar{x} - \sigma_x t].$$

Then,

$$\gamma'(t) = \frac{2t}{(1+t^2)^2} \left[\ln \frac{\bar{x} + \sigma_x t^{-1}}{\bar{x} - \sigma_x t} \right] - \frac{\sigma_x}{(1+t^2)} \left[\frac{1}{\bar{x} + \sigma_x t^{-1}} + \frac{1}{\bar{x} - \sigma_x t} \right].$$

The concavity of the function $t \to t^{-1}$ implies that

$$\ln(t+\delta) - \ln(t) < \frac{\delta}{2} \left[\frac{1}{t+\delta} + \frac{1}{t} \right] \qquad \forall\, t, \delta > 0,$$

from what we deduce that γ' is negative on $(0, \infty)$ and therefore γ is decreasing on this interval. It follows that

$$A = n\gamma(n-1) < B = n\gamma \left(\frac{1}{n-1} \right) < n \lim_{t \downarrow 0} \gamma(t) = n \ln(\bar{x}).$$

The remaining inequality is a straight consequence of the definition of A. $\square$

6. Back to the step-size procedure

Let us go back to Equations (3) and (4). We denote by $\bar{\lambda}$ and σ_λ the arithmetic mean and the standard deviation of the λ_i and by $\|\lambda\|$ the euclidean norm of the vector λ. Then, $\|\lambda\|^2 = n(\bar{\lambda}^2 + \sigma_\lambda^2) = \theta''(0) = -\theta'(0)$ and

$$\theta(t) = nt\bar{\lambda} - t\|\lambda\|^2 - \sum_{i=1}^{n} \ln(1 + t\lambda_i).$$

Our problem consists to find some $\bar{t} \in (0, \widehat{t})$ giving a significant decrease of the convex function θ.

We have said that the most natural choice, $\bar{t} = t_{\rm opt}$ where $\theta'(t_{\rm opt}) = 0$, presents numerical complications. It can be thought of a line-search by a method of Armijo-Goldstein-Price type but this line-search needs also several computations of functions θ and θ'. Nevertheless, if we decide for such a line-search, it is convenient to know, for lack of the upper-bound $\widehat{t}$ of the domain of θ which is numerically difficult to obtain, a lower-bound of $\widehat{t}$. Such a bound is issued from Proposition 2

$$\widehat{t_1} = \sup\,[\,t\,:\,1 + t\beta_1 > 0\,] \quad \text{with} \;\; \beta_1 = \bar{\lambda} - \sigma_\lambda\sqrt{n-1}.$$

Another bound $\widehat{t_2}$ is due to the fact that $|\lambda_i| \leq \|\lambda\|$ for all i

$$\widehat{t_2} = \sup\,[\,t\,:\,1 + t\beta_2 > 0\,] \quad \text{with} \;\; \beta_2 = -\|\lambda\|.$$

Then, $0 < \widehat{t_2} \leq \widehat{t_1} \leq \widehat{t} \leq +\infty$. As already said, the inequality $\widehat{t} \geq \widehat{t_1}$ is a consequence of Proposition 2. To prove that $\widehat{t_1} \geq \widehat{t_2}$ it is enough to prove that $\|\lambda\|^2 \geq \beta_1^2$. This inequality is equivalent to

$$0 \leq (n-1)\bar{\lambda}^2 + \sigma_\lambda^2 + 2\sigma_\lambda\bar{\lambda}\sqrt{n-1} = (\bar{\lambda}\sqrt{n-1} + \sigma_\lambda)^2.$$

Another strategy consists in minimizing an upper-approximation $\widetilde{\theta}$ of θ. To be efficient, this approximation must be simple and close enough to θ. Here we require

$$0 = \widetilde{\theta}(0), \quad \|\lambda\|^2 = \widetilde{\theta}''(0) = -\widetilde{\theta}'(0).$$

Theorem 5 provides such an approximation: set $x_i = 1 + t\lambda_i$, then $\bar{x} = 1 + t\bar{\lambda}$ and $\sigma_x = t\sigma_\lambda$. Next, define

$$\theta_0(t) = \gamma_0 t - (n-1)\ln(1 + \alpha_0 t) - \ln(1 + \beta_0 t),$$

with

$$\gamma_0 = n\bar{\lambda} - \|\lambda\|^2, \;\; \alpha_0 = \bar{\lambda} + \frac{\sigma_\lambda}{\sqrt{n-1}} \text{ and } \;\; \beta_0 = \beta_1 = \bar{\lambda} - \sigma_\lambda\sqrt{n-1}.$$

It is clear that θ_0 is convex, its domain is $[0, \widehat{t}_0)$ with $\widehat{t}_0 = \widehat{t}_1$ and

$$\theta(t) \le \theta_0(t) \ \ \forall\, t \ge 0, \ \ \theta_0(0) = 0 \quad \text{and} \ \ \theta_0''(0) = -\theta_0'(0) = \|\lambda\|^2.$$

One can also thought of simpler functions than θ_0 involving only one logarithm. We consider functions of the following type

$$\widetilde{\theta}(t) = \widetilde{\gamma} t - \widetilde{\delta} \ln(1 + \widetilde{\beta} t), \ \ t \in [0, \widetilde{t})$$

where in order to fulfill the requirements

$$\|\lambda\|^2 = \widetilde{\delta}\widetilde{\beta}^2 = \delta\widetilde{\beta} - \widetilde{\gamma}, \qquad \widetilde{t} = \sup\,[\, t \,:\, 1 + t\widetilde{\beta} > 0\,].$$

Such functions are convex.

Of course $\widetilde{t} \le \widehat{t}$ is required. In line with the lower-bounds $\widehat{t}_1$ and $\widehat{t}_2$, we consider the two functions θ_1 and θ_2 corresponding to β_1 and β_2. In the following result, we compare θ_0, θ_1 and θ_2 . As in other parts of the paper $\ln(r) = -\infty$ if $r \le 0$.

Proposition 3. θ_i, $i = 0, 1, 2$, *is strictly convex on* $[0, \widehat{t}_i)$, $\theta_i(t) \to +\infty$ *when* $t \to \widehat{t}_i$. *Furthermore,* $\theta(t) \le \theta_0(t) \le \theta_1(t) \le \theta_2(t) \le +\infty$ *for all* $t > 0$.

Proof. The first part is immediate. The inequality $\theta(t) \le \theta_0(t)$ is a straight consequence of Theorem 5. Set $\nu(t) = \theta_1 - \theta_0$. Because $\beta_0 = \beta_1$ and $\alpha_0 \ge \beta_0$ one has for $t > 0$

$$\nu''(t) = \frac{\delta_1\beta_1^2 - \beta_0^2}{(1 + \beta_0 t)^2} - \frac{(n-1)\alpha_0^2}{(1 + \alpha_0 t)^2} = \frac{(n-1)\alpha_0^2}{(1 + \beta_0 t)^2} - \frac{(n-1)\alpha_0^2}{(1 + \alpha_0 t)^2} \ge 0.$$

Because $\nu(0) = \nu'(0) = 0$, one deduces that $\nu(t) \ge 0$ for $t > 0$.

Next, set $\mu(t) = \theta_2 - \theta_1$. Then, $\mu(0) = \mu'(0) = 0$ and

$$\mu''(t) = \|\lambda\|^2 \left[\frac{1}{(1 + \beta_2 t)^2} - \frac{1}{(1 + \beta_1 t)^2} \right] \ge 0.$$

Here again $\mu(t) \ge 0$ for all $t > 0$. □

We deduce that the function θ_i reaches its minimum in one unique value $\bar{t}_i$ which is the root of the equation $\theta_i'(t) = 0$. For $i = 1, 2$ one has

$$\bar{t}_i = \frac{\delta_i}{\gamma_i} - \frac{1}{\beta_i} \quad \text{and} \quad \theta_i(\bar{t}_i) = \frac{\|\lambda\|^2}{\beta_i} + \frac{\|\lambda\|^2}{\beta_i^2} \ln(1 - \beta).$$

In particular,

$$\bar{t}_2 = \frac{1}{1 + \|\lambda\|} \quad \text{and} \quad \theta_2(\bar{t}_2) = -\|\lambda\| + \ln(1 + \|\lambda\|).$$

The solution of $\theta_0'(t) = 0$ leads to the equation $t^2 - 2bt + ct = 0$ with

$$b = \frac{1}{2}\left(\frac{n}{\gamma_0} - \frac{1}{\alpha_0} - \frac{1}{\beta_0}\right) \quad \text{and} \quad c = -\frac{\|\lambda\|^2}{\alpha_0\beta_0\gamma_0},$$

whose the two roots are $t = b \pm \sqrt{b^2 - c}$. For $\bar{t}_0$ we take the root which belongs to the interval $(0, \widehat{t}_0)$ (there is only one).

Thus, the three values $\bar{t}_0$, $\bar{t}_1$ and $\bar{t}_2$ are explicitly computed. It is clear that

$$\theta(\bar{t}_2) \le \theta_2(\bar{t}_2), \quad \theta(\bar{t}_1) \le \theta_1(\bar{t}_1) \le \theta_1(\bar{t}_2) \le \theta_2(\bar{t}_2)$$

and

$$\theta(\bar{t}_0) \le \theta_0(\bar{t}_0) \le \theta_0(\bar{t}_1) \le \theta_1(\bar{t}_1) \le \theta_2(\bar{t}_2).$$

7. Description of the algorithm

Initialization: One decides for a step-size strategy and we choose the parameters $\varepsilon > 0$, $r > 0$, $\rho > 0$, $\sigma \in (0,1)$. We start with some $y \in \widehat{Y}$.

Main step: **(a)** Compute $B = B(y)$ and L such that $LL^t = B$.

- **(b)** Compute $g = b - rb(y)$ and $H = r\Delta(y)$.
- **(c)** Solve the equation $Hd = -g$. Compute E, trace(E) and trace(E^2).
- **(d)** Compute $\bar{\lambda}$ and $\bar{\sigma}_\lambda$.
- **(e)** Obtain $\bar{t}$ using the step-size strategy. Take $\bar{y} = y + \bar{t}d$.
- **(f)** If $|b^ty - b^t\bar{y}| > \rho nr$, do $y = \bar{y}$ and go to **(a)**.
- **(g)** If $nr > \varepsilon$, do $y = \bar{y}$, $r = \sigma r$ and go to **(a)**.
- **(h)** **Stop:** $\bar{y}$ is an approximate solution of the problem (D).

As said previously, the optimal solution of problem (D_r) is only one approximate solution of problem (D), more r is close to 0, more the approximation is good. Unfortunately, more r is close to 0, more (D_r) is badly conditioned. It is the reason why we use in the first iterations of the algorithm large r instead of dealing directly with a value of r such that $nr < \varepsilon$. The reason for the updating of r is the following: if $y(r)$ is the exact solution of problem (D_r), then $b^ty(r) \in [m_d, m_d{+}nr]$, it is wasting time to continue iterations on (D_r) when $|b^ty - b^t\bar{y}| \le \rho nr$, with ρ near 1. For ρ one can consider for instance the values 0.5, 1, 2, 3. For σ, we can take for instance the values 0.1, 0.25, 0.5. We describe now four different strategies for the step-size:

- **Strategy Ls**: A classical line-search of Armijo-Goldstein-Price type.
- **Strategy S**$_i$, $i = 0, 1, 2$: $\bar{t} = \bar{t}_i$ with $\bar{t}_i$ defined as in the last section.

8. Numerical experiments

The computations have been performed on a D 810 station with Delphi 5.

8.1. Example cube

$n = 2m$, C is the $n \times n$ identity matrix, $b = (2, ..., 2)^t \in \mathbb{R}^m$ and the entries of the $n \times n$ matrix A_k, $k = 1, \cdots m$, are given by:

$$A_k[i,j] = \begin{cases} 1 & \text{if } i = j = k & \text{or } i = j = k + m, \\ a^2 & \text{if } i = j = k + 1 & \text{or } i = j = k + m + 1, \\ -a & \text{if } i = k, j = k + 1 & \text{or } i = k + m, j = k + m + 1, \\ -a & \text{if } i = k + 1, j = k & \text{or } i = k + m + 1, j = k + m, \\ 0 & \text{otherwise.} \end{cases}$$

$a \in \mathbb{R}$ is given.

Test 1. $(m, n) = (50, 100)$ and $a = 0$. Then, it is known that the vector $y = (1, ..., 1)^t \in \mathbb{R}^m$ is the optimal solution and $y_0 = (1.5, ..., 1.5)^t \in \mathbb{R}^m$ is feasible. We take for parameters in the algorithm $\rho = 1$, $\sigma = 0.125$, $r_0 = 0.3$, $\varepsilon = 0.1$ and for initial point y_0 . The following array describes the results.

Strategy	Computational time	Number of iterations
S_0	34 s	3
S_1	34 s	4
S_2	660 s	25
Ls	dvg	dvg

dvg means that the algorithm does not terminate within a finite time.

Test 2. In this test, the data are the same as in the first test, except $\rho = 2$ in place of $\rho = 1$.

Strategy	Computational time	Number of iterations
S_0	33 s	3
S_1	34 s	4
S_2	480 s	18
Ls	dvg	dvg

The results of these two tests show that the strategies S_2 and Ls do not compete with S_0 and S_1 and $a = 2$ or 5. In the next experiments, we continue only with S_0 and S_1.

Test 3. Same data as in test 1 except $C = -2I$ in place of $C = I$. We start the algorithm with the feasible point $y_0 = (0, ..., 0)^t$ and we take $\rho = 1$.

Strategy	S_0		S_1	
a	2	5	2	5
computational time	165 s	80 s	180 s	120 s
number of iterations	10	5	12	13

Test 4. Same data as in test 3 except $\rho = 2$ in place of $\rho = 1$.

Strategy	S_0		S_1	
a	2	5	2	5
computational time	130 s	50 s	135 s	56 s
number of iterations	7	3	10	5

References

[1] F. Alizadeh, Interior point methods in semi-definite programming with application to combinatorial optimization. *SIAM J. Optim.* **5** (1995) 13–55.
[2] F. Alizadeh, J.-P. Haberly, and M.-L. Overton, Primal-dual interior-point methods for semidefinite programming, convergence rates, stability and numerical results. *SIAM J. Optim.* **8** (1998) 746–768.
[3] D. Benterki, J.-P. Crouzeix, and B. Merikhi, A numerical implementation of an interior point method for semi-definite programming. *Pesquisa Operacional* **23–1** (2003) 49–59.
[4] J.-F. Bonnans, J.-C. Gilbert, C. Lemaréchal, and C. Sagastizàbal, Numerical optimization, theoretical and practical aspects. *Mathematics and Applications* **27**, Springer-Verlag, Berlin (2003).
[5] J.-P. Crouzeix and A. Seeger, New bounds for the extreme values of a finite sample of real numbers. *J. Math. Anal. Appl.* **197** (1996) 411–426.
[6] M. Kojima, S. Shindoh, and S. Hara, Interior point methods for the monotone semi-definite linear complementarity problem in symmetric matrices. *SIAM J. Optim.* **7** (1997) 86–125.
[7] M. Overton and H. Wolkowicz, Semi-definite programming. *Math. program. Serie B* **77** (1997) 105–109.
[8] R.T. Rockafellar, *Convex analysis.* Princeton University Press, New Jerzey (1970).
[9] L. Vanderberghe and S. Boyd, Positive definite programming. *SIAM Review* **38** (1996) 49–95.
[10] H. Wolkowicz and G.-P.-H. Styan, Bounds for eigenvalues using traces. *Linear Algebra Appl.* **29** (1980) 471–506.