ESAIM: Control, Optimisation and Calculus of Variations February 2000, Vol. 5, p. 139–156
URL: http://www.emath.fr/cocv/

COMPUTATION OF THE DISTANCE TO SEMI-ALGEBRAIC SETS *

Christophe Ferrier[1]

Abstract. This paper is devoted to the computation of distance to set, called S, defined by polynomial equations. First we consider the case of quadratic systems. Then, application of results stated for quadratic systems to the quadratic equivalent of polynomial systems (see [5]), allows us to compute distance to semi-algebraic sets. Problem of computing distance can be viewed as non convex minimization problem: $d(u,S) = \inf_{x\in S} \|x-u\|^2$, where u is in $\mathbb{R}^n$. To have, at least, lower approximation of distance, we consider the dual bound $m(u)$ associated with the dual problem and give sufficient conditions to guarantee $m(u) = d(u,S)$. The second part deal with numerical computation of $m(u)$ using an interior point method in semidefinite programming. Last, various examples, namely from chemistry and robotic, are given.

Résumé. Dans cet article nous nous intéressons au calcul de la distance à un ensemble S défini par des équations polynomiales. Nous considérons d'abord le cas quadratique. Le passage au cas polynomial se fait ensuite grâce aux équivalents quadratiques des systèmes polynomiaux développés dans [5]. Le calcul de la distance peut être vu comme un problème de minimisation non convexe $d(u,S) = /inf_{x\in S}\|x-u\|^2$ où $u \in \mathbb{R}^n$. Pour obtenir, au moins un minorant de cette distance, nous considérons la borne duale $m(u)$ issue de la résolution du problème dual. De plus, nous donnons des conditions suffisantes pour avoir $m(u) = d(u,S)$. La seconde partie de cet article est consacrée au calcul de $m(u)$ en utilisant une méthode de points intérieurs en programmation semi-définie positive. Pour finir, nous donnons des exemples d'applications issus notamment de la chimie et de la robotique.

AMS Subject Classification. 90C46.

Received September 23, 1998. Revised October 1, 1999.

1. Introduction

Computing distance to a set defined by polynomial equations has various applications. In motion planning, Robot and obstacles are modelised in the configuration space, constructed taking robot's characteristics into account (orientation, speed, etc). So, it is necessary to compute the distance between a point and a set modelised by complex equations. Another application is the localization of set defined by polynomial equations, using bissection/exclusion technic. This can be used as a first step in the resolution of systems of polynomials equations, before using locally convergent technics.

Keywords and phrases: Distance, dual bond, optimality conditions, polynomial systems, interior point methods, semidefinite programming, location of zeros.

* *This work was realised during the Phd thesis of the author at Laboratoire Approximation and Optimization of the Université Paul Sabatier, France.*

[1] Département Mathématiques Appliquées et Analyse Numérique, Centre National d'Études Spatiales, 18 avenue E. Belin, 31401 Toulouse Cedex 4, France; e-mail: Christophe.Ferrier@cnes.fr

First, we will present a method which computes the distance between fixed point $u \in \mathbb{R}^n$ and set S defined by quadratic equations:

$$S = \{x \in \mathbb{R}^n \mid f_i(x) = 0, 1 \leq i \leq p,\ g_i(x) \leq 0, p+1 \leq i \leq p+q\}.$$

Where

$$\begin{array}{rcl} f_i(x) & = & x^T A_i x + b_i^T x + c_i,\ 1 \leq i \leq p, \\ g_i(x) & = & x^T A_i x + b_i^T x + c_i,\ p+1 \leq i \leq p+q, \end{array}$$

with A_i real symmetric matrices, $b_i \in \mathbb{R}^n$ and $c_i \in \mathbb{R}$. In order to compute the distance we consider the optimization problem:

$$(\mathcal{P}) \left\{ \begin{array}{l} \inf \|x-u\|^2 = d(u,S) \\ x \in S. \end{array} \right. \tag{1.1}$$

Since there is no assumption on S, it may be non-convex, non-connected and unbounded. Thus this problem is, in general, hard to solve. That's why we consider its dual:

$$m(u) = \sup_{(\lambda,\mu)\in\mathbb{R}^p\times\mathbb{R}^q_+} \inf_{x\in\mathbb{R}^n} \left\{ \|x-u\|^2 + \sum_{i=1}^{p} \lambda_i f_i(x) + \sum_{i=p+1}^{p+q} \mu_i g_i(x) \right\}.$$

We always have:

$$d(u,S) \geq m(u) \geq 0.$$

A simple calculus gives that $h(\lambda,\mu) = \inf_{x\in\mathbb{R}^n}\{\|x-u\|^2 + \sum_{i=1}^{p} \lambda_i f_i(x) + \sum_{i=p+1}^{p+q} \mu_i g_i(x)\}$ is a concave function in (λ,μ). So the dual problem is the maximization of concave function. By the way, it is easier to solve than the original problem. The price to pay is the possible difference between $d(u,S)$ and $m(u)$, called the *duality gap*.

One can show that the dual problem can be written as:

$$m(u) = \sup_{(\lambda,\mu)\in\bar{\Omega}} h(\lambda,\mu,u),$$

where

$$\bar{\Omega} = \left\{ (\lambda,\mu) \in \mathbb{R}^p \times \mathbb{R}^q_+, I + \sum_{i=1}^{p} \lambda_i A_i + \sum_{i=p+1}^{p+q} \mu_i A_i,\ \text{positive semi-definite} \right\}.$$

Moreover h is C^2 on the interior of $\bar{\Omega}$:

$$\Omega = \left\{ (\lambda,\mu) \in \mathbb{R}^p \times \mathbb{R}^q_+, I + \sum_{i=1}^{p} \lambda_i A_i + \sum_{i=p+1}^{q} \mu_i A_i,\ \text{positive definite} \right\}.$$

Then we give an explicit formulae for $h(\cdot,u)$, its gradient and its Hessian on Ω. The formula for the gradient allows a low cost computation.

It is easy to see that $\bar{\Omega}$ is convex. Then the dual problem is the maximization of concave function over convex set of semidefinite matrices. We show that this problem can be efficiently solved by interior point method. One

main advantage of such methods is avoiding, as far as possible, instability problems, which arise when the boundary of $\bar{\Omega}$ is approached.

One key question is when does $d(u,S) = m(u)$? We establish the following theorem:

Theorem 1.1. *Let $u \in \mathbb{R}^n$ such that $h(\cdot, u)$ reaches its maximum at a point (λ, μ) in $\Omega \cap \mathbb{R}^p \times \mathbb{R}^q_+$. We have:*

- *$h(\cdot, u)$ is differentiable in (λ, μ) and*
  - *$\nabla_\lambda h(\lambda, \mu, u) = (f_1(x(\lambda, \mu, u)), \dots, f_p(x(\lambda, \mu, u)))^T = 0_{\mathbb{R}^p}$,*
  - *$\frac{\partial}{\partial \mu_i} h(\lambda, \mu, u) = g_i(x(\lambda, \mu, u)) \leq 0$, if $\mu_i = 0$,*
  - *$\frac{\partial}{\partial \mu_i} h(\lambda, \mu, u) = g_i(x(\lambda, \mu, u)) = 0$, if $\mu_i > 0$,*
- *$m^*(u) = \|x(\lambda, \mu, u) - u\|^2 = \min_{x \in S} \|x - u\|^2$,*

Where $x(\lambda, \mu, u)$ is the minimizer of the Lagrangian.

One interesting fact is that hypothesis of the above theorem can be numerically checked. Also note that, when this hypothesis is true, we have the global optimum of the non-convex problem $(\mathcal{P})$. That is, we have the distance between a point u and S and the point of S where it is reached.

We then obtain other useful properties on the dual bound $m^*(u)$, namely:

Proposition 1.2.

$$m^*(u) = 0 \iff u \in S.$$

When the set S is only defined by equalities we have stability result:

Theorem 1.3. *Let $u \in \mathbb{R}^n$ such that $h(\cdot, u)$ reach its maximum in a point $\lambda \in \Omega \cap \mathbb{R}^p$. If S is regular in $x(\lambda, u)$, that is, if the gradient $\nabla f_i(x(\lambda, u))$ are linearly independent, then m^* is continuously differentiable on a neighborhood of u.*

Last, all those results stated for the quadratic case can be applied to the polynomial one, by the way of the quadratic equivalent system of polynomial system (see [5] which is devoted to the symbolic construction of such equivalent system in order to have the best dual bound for the quadratic equivalent system). So we are able to compute, at least, one minimizer of the distance to a set defined by polynomial equations. We give such examples in Section 4.2.

2. Dual problem

As we said above, in order to compute the distance, we consider the optimization problem:

$$(\mathcal{Q}_u) \begin{cases} \inf \|x - u\|^2 = d(u,S)^2 \\ x^T A_i x + b_i^T x + c_i = 0, 1 \leq i \leq p \\ x^T A_i x + b_i^T x + c_i \leq 0, p+1 \leq i \leq p+q \\ x \in \mathbb{R}^n. \end{cases}$$

As the function $\|. - u\|^2$ is coercive and by the Weistrass theorem, $\mathcal{Q}_u$ always has, at least, one solution. But due to non-convexity, this problem is hard to solve directly. So we will consider the dual problem. First we state some definitions and notations.

2.1. Definitions and notations

Definition 2.1. Let:

- Lagrangian of $\mathcal{Q}_u$:

$$\begin{array}{rcll} \mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^n & \longrightarrow & \mathbb{R} & \\ (x, \lambda, \mu, u) & \longrightarrow & \mathcal{L}(x, \lambda, \mu, u) = & \|x - u\|^2 + \sum_{i=1}^{p} \lambda_i f_i(x) \\ & & & + \sum_{i=p+1}^{p+q} \mu_i g_i(x) \end{array}$$

- h:

$$\begin{array}{rccl} h: & \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^n & \longrightarrow & \mathbb{R} \cup \{-\infty\} \\ & (\lambda, \mu, u) & \longrightarrow & h(\lambda, \mu, u) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu, u) \end{array}$$

- m^*:

$$\begin{array}{rccl} m^*: & \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ & u & \longrightarrow & m^*(u) = \sup_{(\lambda,\mu) \in \mathbb{R}^p \times \mathbb{R}^q_+} h(\lambda, \mu, u) \end{array}$$

The dual problem of $\mathcal{Q}_u$ is the calculus of $m(u)$. We need some notations:

- $\mathcal{A}(\lambda, \mu) = I + \sum_{i=1}^{p} \lambda_i A_i + \sum_{i=p+1}^{p+q} \mu_i A_i$;
- $\mathcal{B}(\lambda, \mu, u) = -2u + \sum_{i=1}^{p} \lambda_i b_i + \sum_{i=p+1}^{p+q} \mu_i b_i$;
- $\mathcal{C}(\lambda, \mu, u) = \|u\|^2 + \sum_{i=1}^{p} \lambda_i c_i + \sum_{i=p+1}^{p+q} \mu_i c_i$.

With those notations, Lagrangian becomes:

$$\mathcal{L}(x, \lambda, \mu, u) = x^T \mathcal{A}(\lambda, \mu) x + \mathcal{B}(\lambda, \mu, u)^T x + \mathcal{C}(\lambda, \mu, u).$$

Let

$$\Omega = \{(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q, \mathcal{A}(\lambda, \mu) \text{ positive definite}\}$$

and

$$\bar{\Omega}_+ = \{(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q_+, \mathcal{A}(\lambda, \mu) \text{ positive semi-definite}\}\cdot$$

Remark 2.2. Ω et $\bar{\Omega}_+$ are always non-empty. This is due to the fact that Ω always contains a neighborhood of origin.

We are now able to study the dual problem.

2.2. Properties of the dual bound

First let us give explicit formulae for h, its gradient and its Hessian.

Proposition 2.3. *For all $(\lambda, \mu) \in \Omega$,*

$$x(\lambda, \mu, u) = -\frac{1}{2} \mathcal{A}(\lambda, \mu)^{-1} \mathcal{B}(\lambda, \mu, u)$$

is unique minimum of $\mathcal{L}(., \lambda, \mu, u)$. Then we have:

$$h(\lambda, \mu, u) = -\frac{1}{4} \mathcal{B}(\lambda, \mu, u)^T \mathcal{A}(\lambda, \mu)^{-1} \mathcal{B}(\lambda, \mu, u) + \mathcal{C}(\lambda, \mu, u).$$

Proof. For all couple (x, u) the function $\mathcal{L}(x, \lambda, \mu, u)$, from Ω to $\mathbb{R}$, is strictly convex and coercive. This imply existence and uniqueness of $x(\lambda, \mu, u)$. Moreover, $\mathcal{L}(x, \lambda, \mu, u)$ reaches its minimum at $x(\lambda, \mu, u)$ if and only if $\nabla_x \mathcal{L}(x(\lambda, \mu, u), \lambda, \mu, u) = 0$. But

$$\nabla_x \mathcal{L}(x(\lambda, \mu, u), \lambda, \mu, u) = 2\mathcal{A}(\lambda, \mu) x(\lambda, \mu, u) + \mathcal{B}(\lambda, \mu, u)$$

hence:

$$x(\lambda,\mu,u) = -\frac{1}{2}\mathcal{A}(\lambda,\mu)^{-1}\mathcal{B}(\lambda,\mu,u). \tag{2.1}$$

So:

$$h(\lambda,\mu,u) = \mathcal{L}(x(\lambda,\mu,u),\lambda,\mu,u). \tag{2.2}$$

Replacing (2.1) in (2.2) we obtain:

$$h(\lambda,\mu,u) = -\frac{1}{4}\mathcal{B}(\lambda,\mu,u)^T\mathcal{A}(\lambda,\mu)^{-1}\mathcal{B}(\lambda,\mu,u) + \mathcal{C}(\lambda,\mu,u).$$

□

As the partial derivative of h, with respect to λ or to μ, has the same expression, we will write, in the next proposition, $\xi_i = \lambda_i$ if $1 \le i \le p$ and $\xi_i = \mu_i$ if $p+1 \le i \le p+q$.

Proposition 2.4. *h is twice continuously differentiable on Ω and*

$$\begin{aligned}
\frac{\partial h(\lambda,\mu,u))}{\partial \xi_i} &= \frac{1}{4}\mathcal{B}(\lambda,\mu,u)^T\mathcal{A}(\lambda,\mu)^{-1}A_i\mathcal{A}(\lambda,\mu)^{-1}\mathcal{B}(\lambda,\mu,u) \\
&\quad -\frac{1}{2}\mathcal{B}(\lambda,\mu,u)^T\mathcal{A}(\lambda,\mu)^{-1}\mathcal{B}_i + c_i, \quad 1 \le i \le p+q \qquad (2.3)\\
\frac{\partial^2 h(\lambda,\mu,u)}{\partial \xi_i \partial \xi_j} &= -\frac{1}{2}\mathcal{B}(\lambda,\mu,u)^T\mathcal{A}(\lambda,\mu)^{-1}A_j\mathcal{A}(\lambda,\mu)^{-1}A_i\mathcal{A}(\lambda,\mu)^{-1}\mathcal{B}(\lambda,\mu,u) \\
&\quad +\frac{1}{2}\mathcal{B}(\lambda,\mu,u)^T\mathcal{A}(\lambda,\mu)^{-1}A_j\mathcal{A}(\lambda,\mu)^{-1}b_i - \frac{1}{2}b_i^T\mathcal{A}(\lambda,\mu)^{-1}b_j \qquad (2.4)\\
&\quad +\frac{1}{2}\mathcal{B}(\lambda,\mu,u)^T\mathcal{A}(\lambda,\mu)^{-1}A_i\mathcal{A}(\lambda,\mu)^{-1}b_j, \quad 1 \le i,j \le p+q.
\end{aligned}$$

Proof. Straightforward by chain rule. □

Let us state some main properties of the function h.

Proposition 2.5. *Let $\delta\Omega$ be the boundary of Ω, $\delta\Omega = \bar{\Omega}\backslash\Omega$*

(i) *For all $(\lambda,\mu) \notin \bar{\Omega}$ we have:*

$$h(\lambda,\mu,u) = -\infty.$$

(ii) *For all $(\lambda,\mu) \in \delta\Omega$:*

$$h(\lambda,\mu,u) > -\infty \text{ if and only if } \mathcal{B}(\lambda,\mu,u) \text{ is orthogonal to } Ker\mathcal{A}(\lambda,\mu).$$

(iii) *For all $(\lambda,\mu) \in \delta\Omega$ such that $\mathcal{B}(\lambda,\mu,u)$ is orthogonal to $Ker\mathcal{A}(\lambda,\mu)$ we have:*

$$h(\lambda,\mu,u) = -\frac{1}{4}\Pi_{Im}(\mathcal{B}(\lambda,\mu,u))^T\mathcal{A}_{Im}(\lambda,\mu)^{-1}\Pi_{Im}(\mathcal{B}(\lambda,\mu,u)) + \mathcal{C}(\lambda,\mu,u).$$

Where Π_{Im} is the orthogonal projection on $Im\mathcal{A}(\lambda,\mu)$.

Proof. *(i)* Suppose $\mathcal{A}(\lambda,\mu)$ is not positive semi-definite. By definition there exists, at least, one negative eigenvalue ν of $\mathcal{A}(\lambda,\mu)$ and write v_ν the associated eigenvector. So we have:

$$\mathcal{L}(\alpha v_\mu,\lambda,\mu,u)=(\alpha v_\nu)^T\mathcal{A}(\lambda,\mu)(\alpha v_\nu)+\mathcal{B}(\lambda,\mu,u)^T(\alpha v_\mu)+\mathcal{C}(\lambda,\mu,u).$$

Then $\mathcal{L}(\alpha v_\mu,\lambda,\mu,u)\to-\infty$ when $\alpha\to\infty$. So

$$\inf_{x\in\mathbb{R}^n}\mathcal{L}(x,\lambda,\mu,u)=-\infty.$$

(ii) We have $h(\lambda,\mu,u)=\inf_{x\in\mathbb{R}^n}\mathcal{L}(x,\lambda,\mu,u)$. For all $x\in\mathbb{R}^n$, there exists unique vector x_{Im} in $\mathrm{Im}\mathcal{A}(\lambda,\mu)$ and unique vector x_{K} in $\mathrm{Ker}\mathcal{A}(\lambda,\mu)$, such that $x=x_{\mathrm{Im}}+x_{\mathrm{K}}$, so:

$$\begin{aligned}h(\lambda,\mu,u) &= \inf_{x\in\mathbb{R}^n}\{(x_{\mathrm{Im}}+x_{\mathrm{K}})^T\mathcal{A}(\lambda,\mu)(x_{\mathrm{Im}}+x_{\mathrm{K}})\\ &\quad+\mathcal{B}(\lambda,\mu,u)^T(x_{\mathrm{Im}}+x_{\mathrm{K}})+\mathcal{C}(\lambda,\mu,u)\}\\ &= \inf_{x\in\mathbb{R}^n}\{x_{\mathrm{Im}}^T\mathcal{A}(\lambda,\mu)x_{\mathrm{Im}}+2x_{\mathrm{Im}}^T\mathcal{A}(\lambda,\mu)x_{\mathrm{K}}+x_{\mathrm{K}}^T\mathcal{A}(\lambda,\mu)x_{\mathrm{K}}\\ &\quad+\mathcal{B}(\lambda,\mu,u)^Tx_{\mathrm{Im}}+\mathcal{B}(\lambda,\mu,u)^Tx_{\mathrm{K}}\}+\mathcal{C}(\lambda,\mu,u)\\ &= \inf_{x\in\mathbb{R}^n}\{x_{\mathrm{Im}}^T\mathcal{A}(\lambda,\mu)x_{\mathrm{Im}}+\mathcal{B}(\lambda,\mu,u)^Tx_{\mathrm{Im}}+\mathcal{B}(\lambda,\mu,u)^Tx_{\mathrm{K}})\}\\ &\quad+\mathcal{C}(\lambda,\mu,u)\end{aligned}$$

then

$$\begin{aligned}h(\lambda,\mu,u) = &\inf_{x_{\mathrm{Im}}\in\mathrm{Im}\mathcal{A}(\lambda,\mu)}\{x_{\mathrm{Im}}^T\mathcal{A}_{\mathrm{Im}}(\lambda,\mu)x_{\mathrm{Im}}+\mathcal{B}(\lambda,\mu,u)^Tx_{\mathrm{Im}}\}+\mathcal{C}(\lambda,\mu,u)\\ &+\inf_{x_{\mathrm{K}}\in\mathrm{Ker}\mathcal{A}(\lambda,\mu)}\{\mathcal{B}(\lambda,\mu,u)^Tx_{\mathrm{K}}\}.\end{aligned}$$

But $\inf_{x_{\mathrm{K}}\in\mathrm{Ker}\mathcal{A}(\lambda,\mu)}\{\mathcal{B}(\lambda,\mu,u)^Tx_{\mathrm{K}}\}$ takes the value 0 when $\mathcal{B}(\lambda,\mu,u)$ belongs to the kernel $\mathrm{Ker}\mathcal{A}(\lambda,\mu)^\perp$ and $-\infty$ in other cases. Proposition follows, since $\mathcal{A}_{\mathrm{Im}}(\lambda,\mu)$ is positive definite.

(iii) We have:

$$h(\lambda,\mu,u)=\inf_{x\in\mathrm{Im}\mathcal{A}(\lambda,\mu)}\{x^T\mathcal{A}_{\mathrm{Im}}(\lambda,\mu)x+\Pi_{\mathrm{Im}}\mathcal{B}(\lambda,\mu,u)^Tx+\mathcal{C}(\lambda,\mu,u)\}$$

and $\mathcal{A}_{\mathrm{Im}}(\lambda,\mu)$ is positive definite, as we have shown in *(ii)*. By Proposition 2.3, we have:

$$x(\lambda,\mu,u)=-\frac{1}{2}\mathcal{A}_{\mathrm{Im}}(\lambda,\mu)^{-1}\Pi_{\mathrm{Im}}\mathcal{B}(\lambda,\mu,u)$$

formulae for $h(\lambda,\mu,u)$ follows.

□

Remark 2.6. Above propositions show that $h(\cdot,u)$ is concave function, twice differentiable on interior of its domain $\bar{\Omega}$. Moreover, in the non-degenerate case, $h(\cdot,u)$ takes finite value at only few points of the boundary of its domain and we always know feasible point. So, computation of its maximum seems to be an easy task: $h(\cdot,u)$ is quite a barrier function. However, the matrix $\mathcal{A}(\lambda,\mu)$ becomes ill-conditioned near and on the boundary points of $\bar{\Omega}$ were $h(\cdot,u)$ takes finite value. So the computation of the maximum of the function $h(\cdot,u)$ requires some refinements.

Now we can state the following theorem, which can be related to the one of [8] (Th. XII.2.3.4) even if, here, there is no assumption on the boundedness of the optimisation domain:

Theorem 2.7. *Let $u\in\mathbb{R}^n$ such that $h(\cdot,u)$ reaches its maximum at point (λ,μ) in $\Omega\cap\mathbb{R}^p\times\mathbb{R}_+^q$. We have:*

- *$h(\cdot,u)$ is differentiable in (λ,μ) and*
  - *$\nabla_\lambda h(\lambda,\mu,u)=(f_1(x(\lambda,\mu,u)),\dots,f_p(x(\lambda,\mu,u)))^T=0_{\mathbb{R}^p}$,*
  - *$\frac{\partial}{\partial\mu_i}h(\lambda,\mu,u)=g_i(x(\lambda,\mu,u))\le 0$, if $\mu_i=0$,*

$-\ \frac{\partial}{\partial \mu_i} h(\lambda, \mu, u) = g_i(x(\lambda, \mu), u) = 0$, *if* $\mu_i > 0$,

- $m^*(u) = \|x(\lambda, \mu, u) - u\|^2 = \min_{x \in S} \|x - u\|^2$,

where $x(\lambda, \mu, u)$ *is the minimizer of the Lagrangian.*

Proof. As λ_0 belongs to Ω, $\nabla^2 \mathcal{L}(x, \lambda_0, \mu, u)$ is positive definite and so invertible. As x_0 minimize $\mathcal{L}(x, \lambda_0, \mu, u)$, we have $\nabla \mathcal{L}(x_0, \lambda_0, \mu_0, u) = 0$. By the implicit function theorem, there exist neighborhood $\mathcal{V}(\lambda_0, \mu_0)$ of (λ_0, μ_0), neighborhood $\mathcal{V}(x_0)$ of x_0 and an unique continuously differentiable function X, from $\mathcal{V}(\lambda_0, \mu_0)$ to $\mathcal{V}(x_0)$, such that:

$$\nabla_x \mathcal{L}(X(\lambda, \mu), \lambda, \mu, u) = 0, \forall (\lambda, \mu) \in \mathcal{V}(\lambda_0, \mu_0). \tag{2.5}$$

The positive definiteness of $\nabla^2 \mathcal{L}(x, \lambda_0, \mu, u)$ imply also that $\mathcal{L}(x, \lambda_0, \mu, u)$ is strictly convex in x. So, $\forall (\lambda, \mu) \in \mathcal{V}(\lambda_0, \mu_0)$, $X(\lambda, \mu)$ is the unique minimizer of $\mathcal{L}(x, \lambda, \mu, u)$. By the way:

$$h(\lambda, \mu, u) = \mathcal{L}(X(\lambda, \mu), \lambda, \mu, u). \tag{2.6}$$

So differentiability of $X(\lambda, \mu)$ and (2.6) imply differentiability of $h(\lambda, \mu, u)$ on $\mathcal{V}(\lambda_0, \mu_0)$. By the chain rule, we have:

$$\nabla_{(\lambda,\mu)} h(\lambda, \mu, u) = \nabla_x \mathcal{L}(X(\lambda, \mu), \lambda, \mu, u). \nabla_x X(\lambda, \mu) + \nabla_{(\lambda,\mu)} \mathcal{L}(X(\lambda, \mu), \lambda, \mu, u) \cdot \tag{2.7}$$

According to (2.5, 2.7) becomes:

$$\nabla_{(\lambda,\mu)} h(\lambda, \mu, u) = (f_1(X(\lambda, \mu), \dots, f_p(X(\lambda, \mu), g_{p+1}(X(\lambda, \mu), \dots, g_{p+q}(X(\lambda, \mu))^T. \tag{2.8}$$

By hypothesis $h(\cdot, u)$ reaches its maximum in (λ_0, μ_0), so (2.8) imply:

$$\begin{array}{ll} f_i(X(\lambda_0, \mu_0)) = 0, & 1 \le i \le p, \\ g_j(X(\lambda_0, \mu_0)) = 0, \text{ if } \mu_{0j} \neq 0 & \\ g_j(X(\lambda_0, \mu_0)) < 0, \text{ if } \mu_{0j} = 0, & p+1 \le j \le p+q. \end{array}$$

This proves that $X(\lambda_0, \mu_0)$ belongs to S. On the other hand, $m^*(u) \le \inf_{x \in S} \|x - u\|^2$, but we have $m^*(u) = \mathcal{L}(X(\lambda_0, \mu_0), \lambda_0, \mu_0, u) = \|X(\lambda_0, \mu_0) - u\|^2$. So $m^*(u) = \inf_{x \in S} \|x - u\|^2$. □

Another property of the dual bound is the characterization of the set S:

Proposition 2.8.

$$m^*(u) = 0 \iff u \in S.$$

Proof. First let us show that $u \in S$ imply $m^*(u) = 0$. We have

$$0 \le m^*(u) \le \inf_{x \in S} \|x - u\|^2.$$

But, if $u \in S$ then $\inf_{x \in S} \|x - u\|^2 = 0$, so $m^*(u) = 0$.

Suppose now $m^*(u) = 0$. We have

$$\begin{array}{ll} m^*(u) & = \sup_{(\lambda,\mu) \in \bar{\Omega}} h(\lambda, \mu, u) \\ h(0, 0, u) = 0. & \end{array}$$

So the supremum of $h(.,.,u)$ is reached at $(0,0)$. Now $(0,0)$ belongs to Ω, therefore, hypothesis of Theorem 1.1 are verified. If we note that $x(0,0,u)=u$, then Theorem 1.1 gives:

$$\begin{array}{l} f_i(u)=0, \quad 1\leq i\leq p,\\ g_j(u)<0, \quad p+1\leq j\leq p+q. \end{array}$$

So $u\in S$. □

When the set S is only defined by equalities we have the next stability result:

Theorem 2.9. *Let $u\in\mathbb{R}^n$ such that $h(\cdot,u)$ reaches its maximum at a point $\lambda\in\Omega\cap\mathbb{R}^p$. If S is regular in $x(\lambda,u)$, that is, if the gradients $\nabla f_i(x(\lambda,u))$ are linearly independent, then m^* is continuously differentiable in a neighborhood of u.*

Proof. We write $f(x)$ for $(f_1(x),\dots,f_p(x))^T$. By hypothesis $\lambda_0\in\Omega$, so :

- $\mathcal{L}(.,\lambda_0,u_0)$ has an unique minimizer $x(\lambda_0,u_0)=x_\circ$;
- $\nabla_x\mathcal{L}(x_\circ,\lambda_0,u_0)=0$;
- $\nabla^2_{x,x}\mathcal{L}(x_\circ,\lambda_0,u_0)$ is definite positive.

By the way, we can apply the implicit function theorem to:

$$(x,\lambda,u)\longrightarrow\nabla_x\mathcal{L}(x,\lambda,u),$$

at point $(x_\circ,\lambda_0,u_0)$. So, there exist neighborhoods $\mathcal{V}'(\lambda_0,u_0)$, $\mathcal{V}'(x_\circ)$ and an unique continuous function $X:\mathcal{V}'(\lambda_0,u_0)\to\mathcal{V}'(x_\circ)$ satisfying:

$$X(\lambda_0,u_0)=x_\circ \text{ and } \nabla_x\mathcal{L}(X(\lambda,u),\lambda,u)=0. \tag{2.9}$$

Moreover, X is C^1 on $\mathcal{V}'(\lambda_0,u_0)$ and its gradients are:

$$\begin{array}{rcl} \nabla_\lambda X(\lambda,u) & = & \left[\nabla^2_{x,x}\mathcal{L}(x(\lambda_0,u_0),\lambda_0,u_0)\right]^{-1}\nabla f(x(\lambda_0,u_0)),\\ \nabla_u X(\lambda,u) & = & 2\left[\nabla^2_{x,x}\mathcal{L}(x(\lambda_0,u_0),\lambda_0,u_0)\right]^{-1}. \end{array}$$

By the strict convexity of $\mathcal{L}(x,\lambda,u)$ in x and with the help of (2.9), $X(\lambda,u)$ is the unique minimizer of $\mathcal{L}(x,\lambda,u)$ for all (λ,u) in $\mathcal{V}'(\lambda_0,u_0)$.

Now $X(\lambda,u)$ is differentiable, therefore we obtain:

$$\begin{array}{rcl} \nabla_\lambda h(\lambda,u) & = & \nabla_{(\lambda,u)}\mathcal{L}(x(\lambda,u),\lambda,u)=(f_1(x),\dots,f_p(x))^T_{x(\lambda,u)},\\ \nabla_u h(\lambda,u) & = & -2(x-u). \end{array} \tag{2.10}$$

By hypothesis, $h(\cdot,u_0)$ is maximal at λ_0, so (2.10) gives:

$$\nabla_\lambda h(\lambda_0,u_0)=(f_1(x(\lambda_0,u_0)),\dots,f_p(x(\lambda_0,u_0)))=0_{\mathbb{R}^p}. \tag{2.11}$$

The functions f_i are differentiables. So, by the chain rule in (2.11), h is twice differentiable and its Hessian at point (λ_0,u_0) is:

$$\begin{array}{rcl} \nabla^2_{\lambda,\lambda}h(\lambda_0,u_0) & = & (\nabla f(x(\lambda_0,u_0)))^T\nabla_\lambda x(\lambda_0,u_0)\\ & = & \nabla f(x(\lambda_0,u_0))^T\left[\nabla^2_{x,x}\mathcal{L}(x(\lambda_0,u_0),\lambda_0,u_0)\right]^{-1}\nabla f(x(\lambda_0,u_0)). \end{array}$$

Now we supposed $\nabla f_i(x)$ linearly independent, therefore $\nabla f(x(\lambda_0,u_0))$ has p rank. Moreover, $[\nabla^2_{x,x}\mathcal{L}(x(\lambda_0,u_0),\lambda_0,u_0)]^{-1}$ is definite positive. Combining both precedent fact and by a classical result of linear algebra,

we have that $\nabla^2_{\lambda,\lambda} h(\lambda_0, u_0)$ is also definite positive and so invertible. So, we can apply the implicit function theorem to:

$$(\lambda, u) \longrightarrow \nabla_\lambda h(\lambda, u)$$

at (λ_0, u_0). There exists neighborhood $\mathcal{V}''(u_0)$ of u_0, neighborhood $\mathcal{V}''(\lambda_0)$ of λ_0 and an unique function Γ, C^1 from $\mathcal{V}''(u_0)$ to $\mathcal{V}''(\lambda_0)$, such that:

i) $\lambda_0 = \Gamma(u_0)$,
ii) $\forall u \in \mathcal{V}'''(u_0)\ \ \nabla_\lambda h(\Gamma(u), u) = 0$.

So, as for all $u \in \mathcal{V}''(u_0)$, $h(\cdot, u)$ is concave, differentiable at $\Gamma(u)$ and its differential at $\Gamma(u)$ is zero, $h(\cdot, u)$ reaches its maximum at $\Gamma(u)$ and so

$$m^*(u) = h(\Gamma(u), u).$$

By the way m^* is C^1 on a neighborhood of u_0. □

In conclusion:

1. As we saw in Remark 2.6, computation of h is numerically tractable.
2. In all cases calculus of $m(u)$ gives a lower bound on the distance.
3. sufficient condition for zero duality gap is numerically checked. When this condition is fullfilled we have the global optimum of non-convex problem $(\mathcal{Q}_u)$, that is the distance to set S.

The first item answer to the question "do the calculus of $m(u)$ possible?", the second and the third answer to the question "is it interesting?". However, as we approach the boundary of Ω, matrix $\mathcal{A}$ becomes more and more ill-conditioned, accordingly, the computation of h becomes more and more unstable. So, in the numerical process, it is important to stay, as long as possible, far from this boundary. That's why we choose interior point method to lead this computation.

3. Computation of the dual bound

As we saw in the above section, stability problems arise in the computation of h when the boundary of Ω is approached. As avoiding the boundary of the feasible domain is one main feature of the interior point method, we use it to solve our problem. Principle of such method is rather old, see Fiacco and McCormick's book in 1968 [6]. Recent development of such method started with the work of Karmarkar [11] for linear problems. For the non-linear convex case see work of Sonnevend [15]. It is still an active research field see Jarre [9, 10], Alizadeh [1] and monograph of Nesterov and Nemirovsky [13]. In addition our problem constraints are of type LMI (Linear Matrix Inequality) see [3, 7, 16, 18]. So they can be handle by special barrier function $\ln\det(X)$, which can be called "the principal actor" of such optimization problem, see [13].

3.1. Barrier function

Let P defined from $\Omega_+ = \{(\lambda, \mu) \in \Omega, \mu > 0\}$, to $\mathbb{R}$:

$$\begin{array}{rccl} P: & \Omega_+ & \longrightarrow & \mathbb{R} \\ & (\lambda, \mu) & \longrightarrow & P(\lambda) = \ln\det(\mathcal{A}(\lambda, \mu)) + \sum_{i=p+1}^{p+q} \ln(\mu_i) \end{array} \tag{3.1}$$

Proposition 3.1. *Function P is concave. If matrices A_i are linearly independent, P is strictly concave.*

Proof. Function $\sum_{i=p+1}^{p+q} \ln(\mu_i)$ is strictly concave as sum of strictly concave functions. It is well known that $A \longrightarrow \ln\det(A)$ is also strictly concave over the cone of semi-definite matrix. As $(\lambda, \mu) \longrightarrow \mathcal{A}(\lambda, \mu)$ is affine,

$\ln\det(\mathcal{A}(\lambda,\mu))$ is concave. Moreover, as $\ln\det$ is strictly concave, we have the next equivalence: P is not strictly concave if and only if there exists distinct point (λ,μ) and (λ',μ') in Ω_+, such that:

$$\mathcal{A}(\lambda,\mu) = \mathcal{A}(\lambda',\mu') \tag{3.2}$$

which imply linear dependence of matrices A_i, $1 \leq i \leq p+q$. □

Proposition 3.2. *We have:*

(i) $\lim_{(\lambda,\mu)\to(\bar\lambda,\bar\mu)\in\bar\Omega_+\setminus\Omega_+} P(\lambda,\mu) = -\infty$,

(ii) *P is twice continuously differentiable on Ω_+, its gradient and Hessian are:*

$$\begin{aligned}
\frac{\partial P}{\partial\lambda_i}(\lambda,\mu) &= trace(\mathcal{A}(\lambda,\mu)^{-1}A_i),\\
\frac{\partial P}{\partial\mu_i}(\lambda,\mu) &= trace(\mathcal{A}(\lambda,\mu)^{-1}A_i) + \frac{1}{\mu_i},\\
\frac{\partial^2 P}{\partial\lambda_i\partial\lambda_j}(\lambda,\mu) &= -trace\left(\mathcal{A}(\lambda,\mu)^{-1}A_j\mathcal{A}(\lambda,\mu)^{-1}A_i\right),\\
\frac{\partial^2 P}{\partial\lambda_i\partial\mu_j}(\lambda,\mu) &= -trace\left(\mathcal{A}(\lambda,\mu)^{-1}A_j\mathcal{A}(\lambda,\mu)^{-1}A_i\right),\\
\frac{\partial^2 P}{\partial\mu_i\partial\mu_j}(\lambda,\mu) &= -trace\left(\mathcal{A}(\lambda,\mu)^{-1}A_j\mathcal{A}(\lambda,\mu)^{-1}A_i\right) + \delta_{i,j}\frac{1}{\mu_i},
\end{aligned}$$

where

$$\delta_{i,j} = \begin{cases} 1 \text{ if } i=j,\\ 0 \text{ otherwise.}\end{cases}$$

Proof. First assertion is obvious and the second rely on the twice differentiability of $B \to \ln\det(B)$ over the space of real symmetric definite positive matrices. □

Definition 3.3. Let G be the penalized function:

$$\begin{array}{rccl} G: & \Omega_+\times\mathbb{R}_+\times\mathbb{R}^n & \longrightarrow & \mathbb{R}^n\\ & (\lambda,\mu,u,\alpha) & \longrightarrow & G(\lambda,\mu,u,\alpha)=h(\lambda,\mu,u)+\alpha P(\lambda,\mu)\end{array} \tag{3.3}$$

Due to properties on h and P, G is twice continuously differentiable and we have explicit formulae for its derivatives.

Proposition 3.4. *If $\bar\Omega$ is bounded and non-empty. For all $\alpha>0$ function $G(.,\alpha)$ attains its maximum in Ω_+, at an unique point and*

$$\alpha \longrightarrow (\lambda(\alpha),\mu(\alpha)) = \mathrm{argmax}_{(\lambda,\mu)\in\bar\Omega_+} G(\lambda,\mu,\alpha)$$

is C^1.

Proof. Let $\alpha>0$ and $(\lambda_k,\mu_k)\in\bar\Omega_+$ such that

$$G(\lambda_k,\mu_k,\alpha) \to \sup_{(\lambda,\mu)\in\bar\Omega_+} G(\lambda,\mu,\alpha).$$

For $\beta>0$ great enough, there exists a compact set $\mathcal{K}$ such that $(\lambda_k,\mu_k)\in\mathcal{K}\subset\bar\Omega_+$, this is due to the fact that the boundary of Ω_+ is compact and $G(.,\alpha)$ goes to $-\infty$ when one approaches the boundary. So, sequence

(λ_k, μ_k) have cluster points $(\lambda_\infty, \mu_\infty)$. As the function $G(., \alpha)$ is upper semi-continuous, for all cluster point, we have:

$$G(\lambda_\infty, \mu_\infty, \alpha) = \sup_{(\lambda,\mu)\in\bar{\Omega}_+} G(\lambda, \mu, \alpha).$$

But as $\bar{\Omega}_+$ is a non-empty bounded set, matrices A_i, $1 \leq i \leq p+q$, are linearly independent. Suppose to the contrary that there exists (λ, μ) such that

$$\sum_{i=1}^{p} \lambda_i A_i + \sum_{j=p+1}^{p+q} \mu_j A_j = 0. \tag{3.4}$$

Then, for all $(\lambda', \mu') \in \Omega_+$ and for all $\gamma \in \mathbb{R}$, (3.4) imply

$$A_0 + \sum_{i=1}^{p} (\lambda'_i + \gamma\lambda_i) A_i + \sum_{j=p+1}^{p+q} (\mu'_j + \gamma\mu_j) A_j = A_0 + \sum_{i=1}^{p} \lambda'_i A_i + \sum_{j=p+1}^{p+q} \mu' A_j. \tag{3.5}$$

Now, the second member matrix is, by assumption, positive semi-definite, therefore, for all $\gamma \in \mathbb{R}$, $(\lambda' + \gamma\lambda, \mu' + \gamma\mu) \in \Omega_+$. Which contradict the boundedness of $\bar{\Omega}_+$.

So by Proposition 3.1 P is strictly concave and as h is concave, $G(., \alpha)$ is also strictly concave over Ω_+. Hence, the maximum of the function $G(., \alpha)$ is reached at an unique point and all the cluster points are equal. Moreover, strict concavity of $G(., \alpha)$ imply invertibility of its Hessian matrix. So, implicit function theorem can be applied to:

$$(\lambda, \mu, \alpha) \to \nabla G(\lambda, \mu, \alpha)$$

at point $(\lambda_\infty, \mu_\infty)$. So, there exists neighborhoods $\mathcal{V}(\alpha)$, $\mathcal{V}(\lambda_\infty, \mu_\infty)$ and unique application $(\lambda(\alpha), \mu(\alpha))$, C^1 from $\mathcal{V}(\alpha)$ to $\mathcal{V}(\lambda_\infty, \mu_\infty)$, such that:

$$G(\lambda(\alpha), \mu(\alpha), \alpha) = G(\lambda_\infty, \mu_\infty, \alpha).$$

So, for all $\alpha' \in \mathcal{V}(\alpha)$, we have:

$$\nabla G(\lambda(\alpha'), \mu(\alpha'), \alpha') = 0 \tag{3.6}$$

But together with the strict concavity of the function $G(., \alpha')$, (3.6) imply that $G(., \alpha')$ is maximal at $(\lambda(\alpha'), \mu(\alpha'))$ and this maximum is unique. So the function

$$\alpha \longrightarrow (\lambda(\alpha), \mu(\alpha)) = \text{argmax}_{(\lambda,\mu)\in\bar{\Omega}_+} G(\lambda, \mu, \alpha)$$

is C^1. □

3.2. **Convergence**

Now, we will show that our method converge under mild assumption.

First some notations. When such points exists we will write $(\lambda_\alpha, \mu_\alpha)$ for the points where the maximum of $G(., \alpha)$ is reached.

Theorem 3.5. *If $\bar{\Omega}_+$ has non-empty interior and if one of the conditions below is satisfied:*

(i) *$\bar{\Omega}_+$ is bounded,*

(ii) *there exists* $a > 0$ *such that, for all* $\alpha < a$, $\lim_{\|(\lambda,\mu)\|\to\infty} G(\lambda,\mu,\alpha) = -\infty$*;*
then, when penalty coefficient α *leads to 0:*

- *sequence* $(\lambda_\alpha, \mu_\alpha)$ *admits, at least, one cluster point and all cluster point of* $(\lambda_\alpha, \mu_\alpha)$ *is global minimum of* $(\mathcal{D})$*;*
- $\alpha P(\lambda, \mu) \to 0$.

Proof. If $\bar{\Omega}_+$ is bounded, by Weierstrass theorem, P admits upper bound $M \in \mathbb{R}$ and the function h attains its maximum at point (λ_0, μ_0). If $\bar{\Omega}_+$ is not bounded, by assumption, there exists a real $a > 0$ such that, for all $\alpha < a$

$$\lim_{\|(\lambda,\mu)\|\to\infty} G(\lambda,\mu,\alpha) = -\infty.$$

So considering a suitable compact convex set included in $\bar{\Omega}_+$, we can restrict our proof to the case $\bar{\Omega}_+$ bounded.

If $\bar{\Omega}_+$ is bounded, by Weierstrass theorem, P admits upper bound $M \in \mathbb{R}$ and the function h attains its maximum at point (λ_0, μ_0).

Let α_k be infinite decreasing sequence of positive values. Let the function P_M, from $\bar{\Omega}_+$ to $\mathbb{R}$, be $P_M(\lambda, \mu) = M - P(\lambda, \mu)$. Obviously, for all $(\lambda, \mu) \in \bar{\Omega}_+$, we have $P_M(\lambda, \mu) \leq 0$. Let $G_M(\lambda, \mu, \alpha) = h(\lambda, \mu) + \alpha P_M(\lambda, \mu)$. As G_M is translated from G, the function $G_M(\lambda, \mu, \alpha)$ reaches its maximum at the same points $(\lambda_\alpha, \mu_\alpha)$, as $G(\lambda, \mu, \alpha)$. So we have, for all $\alpha > 0$:

$$h(\lambda_0, \mu_0) \geq h(\lambda_\alpha, \mu_\alpha) \geq h(\lambda_\alpha, \mu_\alpha) + \alpha P_M(\lambda_\alpha, \mu_\alpha). \tag{3.7}$$

By continuity of h and as all points of $\bar{\Omega}_+$ are $\bar{\Omega}_+$-interior point sequence's limit ($\bar{\Omega}_+$ is convex), we have that there exists $(\tilde{\lambda}, \tilde{\mu}) \in \bar{\Omega}_+$ such that:

$$h(\tilde{\lambda}, \tilde{\mu}) \geq h(\lambda_0, \mu_0) - \varepsilon,$$

for all $\varepsilon > 0$. This imply, for all $0 \leq \alpha < a$

$$h(\lambda_\alpha, \mu_\alpha) + \alpha P_M(\lambda_\alpha, \mu_\alpha) \geq h(\tilde{\lambda}, \tilde{\mu}) + \alpha P_M(\tilde{\lambda}, \tilde{\mu}) \geq h(\lambda_0, \mu_0) - \varepsilon + \alpha P_M(\tilde{\lambda}, \tilde{\mu}). \tag{3.8}$$

But as for all $\alpha < a$, the function $G_M(., \alpha)$ is upper bounded and for all $(\lambda, \mu) \in \bar{\Omega}_+$, $G_M(\lambda_\alpha, \mu_\alpha, \alpha) \geq G_M(\lambda, \mu, \alpha)$, the sequence $G_M(\lambda_{\alpha_k}, \mu_{\alpha_k}, \alpha_k)$ is bounded and admits cluster points written $\bar{G}_M$. By (3.8) we have, for all $\varepsilon > 0$:

$$\bar{G}_M \geq h(\lambda_0, \mu_0) + \varepsilon$$

combine with (3.7) and we obtain:

$$h(\lambda_0, \mu_0) \geq \bar{G}_M \geq h(\lambda_0, \mu_0) + \varepsilon. \tag{3.9}$$

This is true for all $\varepsilon > 0$. So (3.9) gives $\bar{G}_M = h(\lambda_0, \mu_0)$. So

$$\lim_{k\to+\infty} G_M(\lambda_{\alpha_k}, \mu_{\alpha_k}, \alpha_k) = h(\lambda_0, \mu_0). \tag{3.10}$$

But as $\lim_{k\to+\infty} h(\lambda_{\alpha_k}, \mu_{\alpha_k}) = h(\lambda_0, \mu_0)$, equation (3.10) imply

$$\lim_{k\to+\infty} P_M(\lambda_{\alpha_k}, \mu_{\alpha_k}) = 0.$$

Hence

$$\lim_{k\to+\infty} G(\lambda_{\alpha_k}, \mu_{\alpha_k}, \alpha_k) = h(\lambda_0, \mu_0),$$
$$\lim_{k\to+\infty} P(\lambda_{\alpha_k}, \mu_{\alpha_k}) = 0.$$

As $(\lambda_{\alpha_k}, \mu_{\alpha_k})$ is in $\bar{\Omega}_+$ which is compact, the sequence haves cluster points $(\bar{\lambda}, \bar{\mu}) \in \bar{\Omega}_+$. By continuity of h, we have that $h(\bar{\lambda}, \bar{\mu}) = h(\lambda_0, \mu_0)$. So, all cluster point $(\lambda_{\alpha_k}, \mu_{\alpha_k})$ is a global optimum of h. □

4. Numerical experiments

In order to compute the dual bound we solve the sequence of problems $(\mathcal{D}_\alpha)$ below when α decrease to 0:

$$(\mathcal{D}_\alpha) \quad \sup_{(\lambda,\mu)\in\mathbb{R}^p\times\mathbb{R}^q} \{h(\lambda, \mu) + \alpha P(\lambda, \mu)\}.$$

For the first value of the penalty we choose the origin as starting point, because it is the only feasible point that we know. In the sequel, the starting point is the previous maximizer.

For each α, we compute the maximum using Newton's method. We adapt the penalty step size in function of how many newton steps have been needed to reach the previous maximizer. The linear algebraic routines used in our C-code are from the language C library Meschach (shareware on internet [17]).

We test our algorithm using "artificial" polynomial systems, specially designed to be hard to solve from the research program POSSO (POlynomial System SOlving) and from real problem as combustion chemistry and robotic (Stewart platform).

4.1. **Quadratic case**

4.1.1. *Big System: 100 unknowns, 100 equations.*

We consider the following system from [20]:

$$x_i^2 + \sum_{k=1}^{n} x_k - 2x_i - 10 = 0, 1 \le i \le n.$$

We took $n = 100$. In general, for the point u we have tested, we have distance and solution point where it is reached.

```
cycles CPU=113873, eps= 2.220450e-13,
mu: 10.000000 -> 0.100000
 distance = 1.019348
 from point u: dim:100 (0 ... 0 )

 to point x0: dim: 100
 0.101934789  ...  0.101934789
 system value at x0: < 5.4e-15
cycles CPU=230526, eps= 2.220450e-13,
mu: 10.000000 -> 0.100000
 distance = 23.980652
 from point u:  dim: 100 (2.5 ... 2.5)

 to point x0:  dim: 100
```

```
 0.101934789 ...  0.101934789

 system value at  x0: < 1.6e-14
```

4.1.2. *robotic:* 9×9.

Considered system is:

$$\begin{cases} f_1(x) &= {x_1}^2 + {x_2}^2 + {x_3}^2 - 12\,x_1 - 68 = 0, \\ f_2(x) &= {x_4}^2 + {x_5}^2 + {x_6}^2 - 12\,x_5 - 68 = 0, \\ f_3(x) &= {x_7}^2 + {x_8}^2 + {x_9}^2 - 24\,x_8 - 12\,x_9 + 100 = 0, \\ f_4(x) &= x_1x_4 + x_2x_5 + x_3x_6 - 6\,x_1 - 6\,x_5 - 52 = 0, \\ f_5(x) &= x_1x_7 + x_2x_8 + x_3x_9 - 6\,x_1 - 12\,x_8 - 6\,x_9 + 64 = 0, \\ f_6(x) &= x_4x_7 + x_5x_8 + x_6x_9 - 6\,x_5 - 12\,x_8 - 6\,x_9 + 32 = 0, \\ f_7(x) &= 2\,x_2 + 2\,x_3 - x_4 - x_5 - 2\,x_6 - x_7 - x_9 + 18 = 0, \\ f_8(x) &= x_1 + x_2 + 2\,x_3 + 2\,x_4 + 2\,x_6 - 2\,x_7 + x_8 - x_9 - 38 = 0, \\ f_9(x) &= x_1 + x_3 - 2\,x_4 + x_5 - x_6 + 2\,x_7 - 2\,x_8 + 8 = 0. \end{cases}$$

See [2] for details.

For this problem we only have, in general, lower approximation of distance.

```
cycles CPU=27 ns=9, eps= 2.220450e-13,
mu: 10.000000 -> 0.010000
 distance = 13.366627
 from point u:  dim: 9  (1  1  1  1  1  1  1  1  1)
 to point x0 :  dim: 9
  3.49273324  -3.63941286   9.19066306    4.308928   1.88292196
  8.275434     3.1542493    4.11066123    3.20553313
 system value at x0 :   dim: 8
2.27373675e-13 1.42108547e-14 3.12638804e-13 1.56319402e-13
2.7000624e-13 1.13686838e-13 4.26325641e-14 1.42108547e-14
```

Here we have distance and point in S where it is reached.

```
cycles CPU=25, eps= 2.220450e-13,
mu: 10.000000 -> 0.010000
 distance = 15.342724
 from point u:  dim: 9  (0  0  0  0  0  0  0  0  0)
 to point x0 :  dim: 9
  3.68391804  -4.15602597   8.92236933   4.14668808   1.35553913
  8.05542796   1.42738915   3.53910593   4.46967467
 system value at x0:  dim: 8
 -1.75453803 -0.34404164  -4.0339362   -0.720836513   3.03902923
  1.30092245 0.0225397674 -0.00848423342
```

We see that point $x_\circ$ is not system solution. From this computation, we know that there is no solution in hypercube centered in (0 0 0 0 0 0 0) and with radius 15.342724. This information can be used by bisection-exclusion algorithms to localize the roots of polynomial systems, see [4].

4.2. **Polynomial systems**

4.2.1. *Chemical combustion*

This example is from [19]

$$\begin{cases} p_1(z) &= z_2 + 2\,z_6 + z_9 + 2\,z_{10} - \frac{1}{100000} = 0, \\ p_2(z) &= z_3 + z_8 - \frac{3}{100000} = 0, \\ p_3(z) &= z_1 + z_3 + 2\,z_5 + 2\,z_8 + z_9 + z_{10} - \frac{1}{20000} = 0, \\ p_4(z) &= z_4 + 2\,z_7 - \frac{1}{100000} = 0, \\ p_5(z) &= 5.140437 \times 10^{-8} z_5 - {z_1}^2 = 0, \\ p_6(z) &= 1.006932 \times 10^{-7} z_6 - {z_2}^2 = 0, \\ p_7(z) &= 7.816278 \times 10^{-16} z_7 - {z_4}^2 = 0, \\ p_8(z) &= 1.496236 \times 10^{-7} z_8 - z_1 z_3 = 0, \\ p_9(z) &= 6.194411 \times 10^{-8} z_9 - z_1 z_2 = 0, \\ p_{10}(z) &= 2.089296 \times 10^{-15} z_{10} - z_1 z_2^2 = 0. \end{cases}$$

Quadratic equivalent system is:

$$\begin{cases} f_1(x) &= x_2 + 2\,x_6 + x_9 + 2\,x_{10} - \frac{1}{100000} = 0, \\ f_2(x) &= x_3 + x_8 - \frac{3}{100000} = 0, \\ f_3(x) &= x_1 + x_3 + 2\,x_5 + 2\,x_8 + x_9 + x_{10} - \frac{1}{20000} = 0, \\ f_4(x) &= x_4 + 2\,x_7 - \frac{1}{100000} = 0, \\ f_5(x) &= 5.140437 \times 10^{-8} x_5 - {x_1}^2 = 0, \\ f_6(x) &= 1.006932 \times 10^{-7} x_6 - {x_2}^2 = 0, \\ f_7(x) &= 7.816278 \times 10^{-16} x_7 - {x_4}^2 = 0, \\ f_8(x) &= 1.496236 \times 10^{-7} x_8 - x_1 x_3 = 0, \\ f_9(x) &= 6.194411 \times 10^{-8} x_9 - x_1 x_2 = 0, \\ f_{10}(x) &= 2.089296 \times 10^{-15} x_{10} - x_1 x_{11} = 0, \\ f_{11}(x) &= x_{11} - x_2^2 = 0. \end{cases}$$

In general we have distance and solution point where it is reached.

```
cycles CPU=78 ns=11, eps= 2.220450e-13,
alpha : 10.000000 -> 0.010000,
 distance = 0.000022
 from point u: dim: 11
  0  0  0  0  0  0  0  0  0  0
 to point x0:  dim: 11
1.7225278e-20 5.7417594e-20 1.4848484e-05 1.4210854e-19
6.0606060e-07 2.0202020e-06         5e-06 1.5151515e-05
1.3131313e-06 2.3232323e-06 3.6151826e-29

 system value at x0: dim: 11
 5.5904174e-20 -3.3881317e-21  1.3552527e-20  1.4230153e-19
-2.9671021e-40 -3.2967801e-39 -2.0194839e-38 -2.5576928e-25
-9.8903405e-40 -6.2272522e-49  3.6151823e-29
cycles CPU=110 ns=11, eps= 2.220450e-13,
mu: 2.000000 -> 0.010000,
 distance = 21.307277
 from point u:  dim: 11
  -6  -9   8  -6  -7  -8  -6  -4  -8  -6 -5
```

```
 to point x0:  dim: 11
-1.12673899e-15 -1.61070089e-13   4.00001485 -4.5238948e-14
 1.00000061     -2.66666465       5e-06      -3.99998485
-1.33333202      3.33333566      -1.62501599e-15
 system value at x0: dim: 11
-1.59362515e-13 -2.91208911e-15 -5.00151271e-15 -4.56175252e-14
-1.26954076e-30 -2.59435735e-26 -2.04656242e-27  4.5069727e-15
-1.8148395e-28  -1.83096888e-30 -1.62501599e-15
```

4.3. **Numerical bifurcation test**

We consider system below, from [12]:

$$\begin{cases} p_1(x) &= 5\,{z_1}^9 - 6\,{z_1}^5 z_2 + z_1 {z_2}^4 + 2\,z_1 z_3 = 0 \\ p_2(x) &= -2\,{z_1}^6 z_2 + 2\,{z_1}^2 {z_2}^3 + 2\,z_2 z_3 = 0 \\ p_3(x) &= {z_1}^2 + {z_2}^2 - 0.265625 = 0. \end{cases}$$

Quadratic equivalent system is:

$$\begin{cases} x_{12}x_{13} - 6\,x_9x_8 + x_{11}x_2 + 2\,x_1x_3 = 0, \\ -2\,x_6x_{12} + 2\,x_9x_2 + 2\,x_2x_3 = 0, \\ {x_1}^2 + {x_2}^2 - 0.265625 = 0, \\ x_2x_8 - x_1x_6 = 0, \\ -x_{10} + x_2x_8 = 0, \\ -x_8x_{11} + x_7x_{10} = 0, \\ -x_5x_7 + x_2x_{10} = 0, \\ x_4x_{11} - {x_7}^2 = 0, \\ -x_8 + x_1x_5 = 0, \\ x_1x_2 - x_4 = 0, \\ x_6x_7 - x_5x_{11} = 0, \\ {x_{10}}^2 - x_9x_{12} = 0, \\ -x_1x_{13} + {x_8}^2 = 0, \\ -x_4x_9 + x_6x_7 = 0, \\ x_4x_{13} - x_8x_{10} = 0, \\ -x_7x_9 + x_6x_{11} = 0, \\ x_{10}x_{11} - {x_9}^2 = 0, \\ -x_6x_{12} + x_4x_{13} = 0, \end{cases} \begin{cases} x_2x_4 - x_7 = 0, \\ -x_4x_5 + x_2x_8 = 0, \\ -x_5x_{12} + {x_8}^2 = 0, \\ -x_9x_{10} + x_{11}x_{12} = 0, \\ -x_7x_{13} + {x_{10}}^2 = 0, \\ {x_6}^2 - x_4x_{10} = 0, \\ -x_8x_{12} + x_5x_{13} = 0, \\ x_6x_{13} - x_{10}x_{12} = 0, \\ x_8x_{13} - {x_{12}}^2 = 0, \\ {x_5}^2 - x_{12} = 0, \\ x_6x_{10} - x_8x_9 = 0, \\ x_{11} - x_2x_7 = 0, \\ -x_7x_{12} + x_6x_{10} = 0, \\ {x_5}^2 - x_1x_8 = 0, \\ -x_6x_8 + x_4x_{12} = 0, \\ x_6 - x_1x_4 = 0, \\ -x_9 + {x_4}^2 = 0, \\ {x_4}^2 - x_1x_7 = 0, \end{cases} \begin{cases} x_6 - x_2x_5 = 0, \\ x_4x_8 - x_1x_{10} = 0, \\ x_4x_{12} - x_5x_{10} = 0, \\ x_4x_{12} - x_2x_{13} = 0, \\ x_4x_8 - x_5x_6 = 0, \\ -x_2x_6 + {x_4}^2 = 0, \\ -x_1x_{11} + x_2x_9 = 0, \\ x_2x_9 - x_4x_7 = 0, \\ x_2x_{10} - x_1x_9 = 0, \\ {x_6}^2 - x_7x_8 = 0, \\ -x_4x_6 + x_2x_{10} = 0, \\ x_7x_{10} - x_6x_9 = 0, \\ x_1x_{12} - x_{13} = 0, \\ -{x_1}^2 + x_5 = 0, \\ x_1x_{12} - x_5x_8 = 0, \\ -x_2x_{12} + x_4x_8 = 0, \\ -x_5x_9 + {x_6}^2 = 0. \end{cases}$$

For the different points u, we have tested, we only manage to have non trivial lower approximation of distance.

```
cycles CPU=5558, eps= 2.220450e-13,
mu: 100.000000 -> 0.000100
 distance = 0.925606
 from point u : dim: 13
   1  1  0   0   0  0  0  0  0  0  0  0  0
 to point x0: dim: 13
   0.329446583    0.394387662  - 0.0110276803
   0.129282427    0.109119827    0.042683305
   0.0509202463   0.0362028528   0.0167837212
   0.0141194878   0.0201264018   0.0121048743
   0.00402183489
 system value at x0: between 3.5e-3 and 1.5e-6
```

```
cycles CPU=4571, eps= 2.220450e-13,
mu: 100.000000 -> 0.010000,
 distance = 13.666166
 from point u: dim: 13
   10  -10  1   0   0  0  0  0  0  0  0  0  0
 to point x0:  dim: 13
   0.361958856   -0.364033061  -0.00382504396
  -0.131744449    0.131999456  -0.0479597274
   0.0480543883   0.0479694041  0.0182949269
  -0.0174346063  -0.0174637981  0.0185660341
   0.00673531879
 system value at x0: between 8.7e-3 and 7.2e-6
```

5. Conclusion

We can see on numerical examples studied in Section 4, that our method has good convergence properties in practice. Moreover, even for the case where we did not have the an exact solution, the lower bound on the distance is non trivial. Thus, this can be used to localise the solution set by bissection/exclusion technic. This good behaviour obviously comes from the capacity of our method to avoid singularities on the boundary of Ω. Of course, there may exists other methods sharing such a capacity. For example Shor's space dilatation in the direction of two successive subgradients (see [14]) has been tested on simple examples and avoids the singularities. In order to compare it with our method, it would be necessary to test it on real examples of same kind as those of Section 4.

References

[1] F. Alizadeth, Interior point methods in semidefinite programming with application to combinatorial optimisation. *SIAM J. Optim.* **5** (1995) 13–51.

[2] A. Bellido, Construction de fonctions d'itération pour le calcul simultané des solutions d'équations et de systèmes d'équations algébriques. Thèse de doctorat de l'Universté Paul Sabatier, Toulouse (1992).

[3] S. Boyd *et al.*, *Linear Matrix Inequalities Problem in Control Theory*. SIAM, Philadelphia, *Stud. Appl. Math.* **15** (1995).

[4] J.-P. Dedieu and J.-C. Yakoubsohn, Localization of an algebraic hypersurface by the exclusion algorithm. *Appl. Algebra Engrg. Comm. Comput.* **2** (1992) 239–256.

[5] Ch. Ferrier, Hilbert's 17^{th} problem and best dual bounds in quadratic minimization. *Cybernetics and System Analysis* **5** (1998) 76–91.

[6] A.V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley (1968). Reprinted SIAM, 1990.

[7] R. Fletcher, Semi-definite matrix constraints in optimization. *SIAM J. Control Optim.* **23** (1985) 493–513.

[8] C. Lemarechal and J.-B. Hiriart-Urruty, *Convex Analysis and Minimization Algorithms II*. Springer Verlag, *Comprehensive Studies in Mathematics* **306** (1991).

[9] F. Jarre, Interior-point methods for convex programming. *Appl. Math. Optim.* **26** (1992) 287–391.

[10] F. Jarre, An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices. *SIAM J. Control Optim.* **31** (1993) 1360–1377.

[11] N. Karmarkar, A new polynomial-time algorithm for linear programming. *Combinatorica* **4** (1984) 373–395.

[12] R.B. Kearfott, Some tests of generalized bisection. *ACM Trans. Math. Software* **13** (1987) 197–200.

[13] Yu. Nesterov and A. Nemirovsky, *Interior-point polynomial methods in convex programming*. SIAM, Philadelphia, *Stud. Appl. Math.* **13** (1994).

[14] N.Z. Shor, Dual estimate in multi-extremal problems. *J. Global Optim.* **2** (1992) 411–418.

[15] G. Sonnevend, An "analytical centre" for polyhedrons and a new classe of global algorithms for linear (smooth, convex) programming. Springer Verlag, *Lecture Notes in Control and Inform. Sci.* **84**, *System Modeling and Optimisation*. 12th IFIP Conference on system optimisation 1984 (1986) 866–878.

[16] G. Sonnevend and J. Stoer, Global ellipsoidal approximation and homotopy methods for solving convex analitic programs. *Appl. Math. Optim.* **21** (1990) 139–165.

[17] D.E. Stewart, *Matrix Computation in C*. University of Queensland, Australia (1992). ftp site: des@thrain.anu.edu.au. directory: pub/meschach

[18] L. Vandenberghe and S. Boyd, Semidefinite programming. *SIAM Rev.* **1** (1996) 49–95.
[19] J. Verschelde, P. Verlinden and R. Cools, Homotopy exploiting newton polytopes for solving sparse polynomials systems. *SIAM J. Numer. Anal.* **31** (1994) 915–930.
[20] A. Wright, Finding solutions to a system of polynomial equations. *Math. Comp.* **44** (1985) 125–133.