A Global Algorithm for Nonlinear Semidefinite Programming

Rafael Correa*, Héctor Ramírez C.†

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet SYDOCO

Rapport de recherche n° 4672 — Decembre 2002 — 21 pages

Abstract: In this paper we propose a global algorithm for solving nonlinear semidefinite programming problems. This algorithm, inspired in the classic SQP (Sequentially Quadratic Programming) method, modifies the S-SDP (Sequentially Semidefinite Programming) local method by using a nondifferentiable merit function combined with a linear search strategy.

Key-words: Nonlinear semidefinite programming, sequentally programming, global convergence

* Centre for Mathematical Modelling, UMR 2071, University of Chile-CNRS. Casilla 170-3 Santiago 3 Chile. Email - rcorrea@dim.uchile.cl

† Department of Mathematical Engineering, University of Chile. Casilla 170-3 Santiago 3 Chile, and Projet SYDOCO, INRIA-Rocquencourt, France. Email - Hector.Ramirez_Cabrera@inria.fr

Un algorithme global pour la programmation semidéfinie nonlinéaire

Résumé : Dans cette article on propose un algorithme pour résoudre des problèmes semidéfinis nonlinéaires. Cet algorithme, qui est inspiré de l'algorithme classique SQP (Programmation Séquentielle Quadratique), modifie l'algorithme local S-SDP (Programmation Séquentielle Semidéfinie) en utilisant une fonction de pénalisation non dérivable et une recherche linéaire.

Mots-clés : Programmation semidéfinie nonlinéaire, programmation séquentielle, convergence globale

AMS Subject Classification: Primary 90C22.

1 Introduction

We consider the nonlinear programming problem

$$(P) \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \mathcal{A}(x) \preceq 0 \\ & h(x) = 0, \end{array}$$

where $x \in I\!R^n$, $\mathcal{A}$ is a smooth function whose values are symmetric matrices; $\preceq$ denotes the negative semi-definite order (that is, $A \preceq B$ if and only if $A - B$ is a negative semi-definite matrix); h is a vector smooth function with values in $I\!R^m$; and f is the smooth objective function. The smoothness of all these functions is specified at each statement.

This problem becomes interesting when the linear matrix formulation [21]:

$$(LMI) \qquad \begin{array}{ll} \text{minimize} & f(x) = c^T x \\ \text{subject to} & \mathcal{A}(x) = A_0 + \sum_{i=1}^m x_i A_i \preceq 0 \end{array}$$

does not give a satisfactory model for certain problems, particularly those from control theory [1, 5, 8].

This paper is organized as follows. In section 2 the optimality and constraints qualification conditions for problem (P) are presented. The results contained in this section are adaptations of known results (see for example [15, 19]). Here only the optimality conditions that are useful in our context are discussed. Other conditions can be found in [2, 6]. In Section 3 we demostrate some exactness results associated with the Lagrangian, the Augmented Lagrangian and the Han Penalty function. The first one is well-known and we review it to make our exposition self-contained. The exactness of the Augmented Lagrangian and the Han Penalty function are extensions of the corresponding classical mathematical programming results [3, 14]. In Section 4 we propose a global S-SDP (Sequentially Semidefinite Programming) and prove its convergence. The convergence of a local S-SDP algorithm has been proved by Fares, Noll and Apkarian [8].

1.1 Notations

Throughout we denote by S_m the set of all symmetric matrices of dimension m, by S_m^+ the set of all symmetric positive semi-definite matrices, and by S_m^{++} the set of all symmetric positive definite matrices. The sets S_m^- and S_m^{--} are defined similarly. For all these sets of matrices we use the trace product $\langle A, B\rangle = Tr(AB)$, and the Frobenius norm $\|A\|_{Fr} = \sqrt{Tr(A^2)}$. For a given matrix A, $\lambda_j(A)$ denotes its j-th eigenvalue in non-increasing order and A_+ denotes the matrix defined by

$$A_+ := P\, diag((\lambda_1)_+, ..., (\lambda_m)_+)P^T. \tag{1}$$

Where $(\lambda)_+ = \max\{0, \lambda\}$ and P is the matrix in the spectral decomposition $A = P\,diag(\lambda_1, ..., \lambda_m)P^T$. It is easy to see that A_+ is the orthogonal projection of A on S_m^+.

Given a matrix-valued function $\mathcal{A}(\cdot)$ we will use the notation

$$D\mathcal{A}(x_*) = (\frac{\partial \mathcal{A}(x_*)}{\partial x_i})_{i=1}^n = (\frac{\partial \mathcal{A}(x_*)}{\partial x_1}, ..., \frac{\partial \mathcal{A}(x_*)}{\partial x_n})^T$$

for its differential operator evaluated at x_*. This notation comes from the fact that

$$D\mathcal{A}(x_*)y = \sum_{i=1}^n y_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} \quad \forall y \in I\!R^n. \tag{2}$$

Finally, if $V = (V_1, ..., V_n)^T$ is a linear operator from $I\!R^n$ to S_m, as $D\mathcal{A}(x_*)$, we have for the adjoint operator V^* the formula

$$V^*Z = (Tr(V_1Z), ..., Tr(V_nZ))^T \quad \forall Z \in S_m. \tag{3}$$

2 Optimality Conditions Review

In this section we state first and second-order optimality conditions for (P) and discuss their implications. To this end, consider the Lagrangian $L : I\!R^n \times S_m \times I\!R^p \to I\!R$ of problem (P) defined by

$$L(x, Z, \lambda) = f(x) + Tr(Z\mathcal{A}(x)) + \lambda^T h(x). \tag{4}$$

2.1 First order optimality condition

The **Karush-Kuhn-Tucker** necessary optimality conditions for a feasible point x_* of (P) are given by the existence of $Z_* \in S_m$ and $\lambda_* = (\lambda_{*1}, ..., \lambda_{*p})^T \in I\!R^p$ such that

$$(KKT) \qquad \begin{array}{rcl} \nabla f(x_*) + D\mathcal{A}(x_*)^* Z_* + \sum_{j=1}^p \lambda_{*j} \nabla h_j(x_*) & = & 0 \\ Tr(Z_*\mathcal{A}(x_*)) & = & 0 \\ Z_* & \succeq & 0. \end{array}$$

The pair (Z_*, λ_*) is called the (KKT)-multiplier associated with x_*. The complementarity condition $Tr(Z_*\mathcal{A}(x_*)) = 0$ has the following useful equivalent forms:

$$\lambda_j(Z_*) = 0 \quad \text{or} \quad \lambda_j(\mathcal{A}(x_*)) = 0, \quad \forall\, j \in \{1, ..., m\}, \tag{5}$$

or

$$Z_*\mathcal{A}(x_*) = 0. \tag{6}$$

Both forms are easily obtained from the *trace (or Frobenius) inequality*:

$$Tr(AB) \leq \sum_{j=1}^{m} \lambda_j(A)\lambda_j(B), \tag{7}$$

where the equality holds if and only if there is a matrix P such that $P^{-1}AP$ and $P^{-1}BP$ are diagonal (see, for example, [20]).

Condition (5) allows us to define the *strict complementarity condition* in (KKT) as follows

$$\lambda_j(Z_*) = 0 \quad \text{if and only if} \quad \lambda_j(\mathcal{A}(x_*)) < 0, \quad \forall\, j \in \{1, ..., m\}. \tag{8}$$

As is well-known, the (KKT) conditions are not a consequence of the optimality of x_*, and to ensure this consequence, we must suppose an extra condition. In this paper, we will use *Robinson's constraint qualification condition* [16]

$$0 \in int \left\{ \begin{pmatrix} \mathcal{A}(x_*) \\ h(x_*) \end{pmatrix} + \begin{pmatrix} D\mathcal{A}(x_*) \\ \nabla h(x_*) \end{pmatrix} I\!R^n - \begin{pmatrix} S_m^- \\ \{0\} \end{pmatrix} \right\}, \tag{9}$$

where $int\, C$ denotes the topological interior of the set C. A direct consequence of [11, Chapter 3, prop. 2.1.12] is the equivalence between condition (9) and

$$\{\nabla h_j(x_*)\} \qquad \text{are linearly independents, and} \tag{10a}$$

$$\exists \bar{d} \in I\!R^n \text{ s. t.} \qquad \begin{cases} \nabla h(x_*)\bar{d} = 0 \\ \text{and } \mathcal{A}_*(\bar{d}) \prec 0, \end{cases} \tag{10b}$$

where $\mathcal{A}_* : I\!R^n \to S_m$ is the linear affine function defined by $\mathcal{A}_*(y) := \mathcal{A}(x_*) + \sum_{i=1}^n y_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i}$. It can be shown that under (9) the set of (KKT) Lagrange multipliers is nonempty and also bounded [13].

We will also consider the Transversality Condition: the function $\psi : I\!R^n \to I\!R^p \times S_r$ defined by

$$\psi(d) := ((\nabla h(x_*)d)^T, N^T \sum_{i=1}^{n} d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N)^T \tag{11}$$

is surjective. Where

$$N = [v_1 ... v_r] \tag{12}$$

is the matrix whose columns v_i are an orthonormal basis of $Ker\ \mathcal{A}(x_*)$. We set $N = 0$ if $Ker\ \mathcal{A}(x_*) = \{0\}$. This condition has originally been defined in the context of smooth manifolds [9] and implies the Robinson's constraint qualification condition (10), moreover, (11) guarantees the uniqueness of the (KKT)-multiplier. Unfortunately, this condition can be very strong, because it forces $n \geq p + r(r+1)/2$, where $r = dim[Ker\, \mathcal{A}(x_*)]$.

It is clear that the Transversality Condition (11) is not verified when the matrix $\mathcal{A}(x_*)$ has a block-submatrix structure. Indeed, in this case the multiplier Z_* is not unique and therefore the transversality condition does not hold. This problem can be easily avoided if we assume this condition for each submatrix of $\mathcal{A}(x_*)$, that is, the surjectivity of the function ψ associated with each submatrix. For simplicity of notation we only consider the case where $\mathcal{A}(x_*)$ is a one-block matrix. More details about the transversality condition in the semidefinite programming context can be seen in [19] and the references within.

2.2 Second-Order sufficient condition

In this section we introduce only the second-order sufficient conditions that will be used in this paper as well as results that involve transversality condition (11). We assume that f, h and $\mathcal{A}$ are twice differentiable at x_*.

Given a set $B \subseteq I\!R^m$ we define

$$S_m^-(B) := \{M \in S_m : w^T M w \leq 0, \quad \forall w \in B\}. \tag{13}$$

Proposition 1 *A sufficient condition to obtain the isolated optimality of x_* for problem (P), is the existence of $(Z_*, \lambda_*) \in S_m \times I\!R^p$ such that (x_*, Z_*, λ_*) satisfies (KKT) and*

$$d^T \nabla^2_{xx} L(x_*, Z_*, \lambda_*) d > 0, \tag{14}$$

for all non zero vector $d \in C(x_*)$. Where

$$C(x_*) = \{d \in I\!R^n : \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} \in S_m^-(Ker\, \mathcal{A}(x_*)),\ \nabla h(x_*) d = 0 \text{ and } \nabla f(x_*)^T d = 0\} \tag{15}$$

is a cone of critical directions for problem (P) at the point x_*.

Proof. See for example [17, Theorem 2.2] and note that $T_{S_m^-}(\mathcal{A}(x_*)) = S_m^-(Ker\, \mathcal{A}(x_*))$. ■

Remark 1 *Condition (14) can be far from necessary. For instance, in the problem (LMI) we always have $\nabla^2_{xx} L = 0$, thus, if $C(x_*) \neq \{0\}$ condition (14) never holds. This is because condition (14) does not consider the geometry of S_m^-. This kind of problem was the motivation for works such as [2, 6, 19] in the nineties. We will just consider the nonlinear problem (P), where the algorithm S-SDP makes sense.*

Let us define now a larger cone of critical directions $C'(x_*, Z_*)$, which considers the (KKT)-multiplier Z_* associated with the matrix inequality $\mathcal{A}(x) \preceq 0$, as follows:

$$\begin{aligned} C'(x_*, Z_*) := \{d \in I\!R^n : & \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} \in S_m^-(Ker\ \mathcal{A}(x_*)), \\ & Im\, Z_* \subseteq Ker\ Pr \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} \text{ and } \nabla h(x_*) d = 0\}, \end{aligned} \tag{16}$$

where Pr is the orthogonal projection in $I\!R^m$ over $Ker\ \mathcal{A}(x_*)$. Note that $Pr = NN^T$ with N defined in (12).

The next proposition relates both cones of critical directions and the function ψ.

Proposition 2 *Let x_* be a solution of (P) and (Z_*, λ_*) a (KKT)-multiplier. Let us consider also the function ψ, defined in (11), and the cones of critical directions defined above. Then*

$$Ker\ \psi \subseteq C(x_*) \subseteq C'(x_*, Z_*), \tag{17}$$

with equality when the strict complementarity condition (8) holds.

Proof. First, note that we can write (6) in the equivalent form

$$Z_* = N\phi_* N^T, \tag{18}$$

with $\phi_* \in S_r^+$ and $r = dim\ Ker\ \mathcal{A}(x_*)$. Then, to prove the first inclusion in (17), it is sufficient to show that $\nabla f(x_*)^T d = 0$. This comes from the first equation in (KKT) and the equality

$$Tr(Z_* \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i}) = Tr(\phi_* N^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N). \tag{19}$$

For the second inclusion, if $d \in C(x_*)$ then $\nabla f(x_*)^T d = 0$ and $\nabla h(x_*) d = 0$, and we obtain from (19) and the first equation in (KKT) that $Tr(\phi_* N^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N) = 0$. Since $\sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} \in S_m^-(Ker\ \mathcal{A}(x_*))$, we see that $N^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N \in S_r^-$, and using (7), we deduce from the last equality that

$$N^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N \phi_* = 0, \tag{20}$$

which is equivalent to

$$Pr \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} Z_* = NN^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N \phi_* N^T = 0, \tag{21}$$

and we conclude that $d \in C'(x_*, Z_*)$.

If in addition we assume the strict complementarity condition (8), we have that ϕ_* is nonsingular, and from the equivalence between (21) and (20) we deduce the converse inclusion $C'(x_*, Z_*) \subseteq Ker\ \psi$. ■

A direct consequence of proposition 1 is the following sufficient condition:

Proposition 3 *Under the hypotheses of proposition 1 with $C(x_*)$ replaced by $C'(x_*, Z_*)$, the point x_* is an isolated local minimum of (P).*

This sufficient condition is of course stronger then (14). This condition is used in the local convergence result of the S-SDP algorithm (Theorem 11).

3 Exact Penalty Functions

A couple (x_*, y_*) in the product set $X \times Y$ is said to be a *saddle-point* of the function $\varphi : X \times Y \to I\!R$ on $X \times Y$ if

$$\varphi(x_*, y) \leq \varphi(x_*, y_*) \leq \varphi(x, y_*), \quad \forall x \in X, \forall y \in Y.$$

We say that a function $\Phi : I\!R^n \to I\!R$ is an *exact penalty function* in a local minimum x_* of (P), if x_* is a local minimum of Φ too.

In this section we study different penalty functions associated with problem (P) and state necessary and sufficient conditions for exactness. A general approach for the study of exact penalty functions can be found in [4, Section 3.4.2].

3.1 The Lagrangian Function in the Convex Case

Let us consider the particular case of (P) when $h : I\!R^n \to I\!R^p$ is an affine function $h(x) = h_0 + Hx$, with $H \in I\!R^{p\times n}$ and $h_0 \in I\!R^p$, $f : I\!R^n \to I\!R$ is convex and $\mathcal{A}(\cdot)$ is convex in the sense of semidefinite order, that is

$$\mathcal{A}(tx + (1-t)y) \preceq t\mathcal{A}(x) + (1-t)\mathcal{A}(y), \quad \forall t \in [0,1], \forall x, y \in I\!R^n.$$

With these assumptions, (P) will be denoted by (P_C) (*convex problem*). For the (KKT)-multipliers (Z_*, λ_*) associated with the solution x_* of (P_C), it can be shown that the function $L(\cdot, Z_*, \lambda_*)$, defined in (4), is an exact penalty function for (P_C). This is an immediate consequence of the fact that (x_*, Z_*, λ_*) is a saddle-point of the Lagrangian function on $I\!R^n \times S_m^+ \times I\!R^p$ (see, for example, [22, theorem 4.1.3]). However, it is known that the Lagrangian is no longer an exact penalty function in the nonconvex case, which is the reason other penalty functions are introduced to obtain exactness results for our general problem (P).

3.2 The Augmented Lagrangian

We define the augmented Lagrangian function L_σ associated with problem (P) as

$$L_\sigma(x, Z, \lambda) = f(x) + \lambda^T h(x) + \frac{\sigma}{2}\|h(x)\|^2 + \\ Tr(Z[\mathcal{A}(x) + (-\sigma^{-1}Z - \mathcal{A}(x))_+]) + \frac{\sigma}{2}\|\mathcal{A}(x) + (-\sigma^{-1}Z - \mathcal{A}(x))_+\|_{Fr}^2, \tag{22}$$

where $\sigma > 0$ is the penalty parameter. In [18], L_σ is called the *proximal augmented Lagrangian.*

If (x_*, Z_*, λ_*) is any point satisfying (KKT), from (6) it can be shown that

$$L_\sigma(x_*, Z_*, \lambda_*) = f(x_*). \tag{23}$$

In the next theorem we prove that $L_\sigma(\cdot, Z_*, \lambda_*)$ is an exact penalty function when σ is sufficiently large.

Theorem 4 *Let us assume that* f, h *and* $\mathcal{A}$ *are twice differentiable at* x_* *and that* (x_*, Z_*, λ_*) *satisfies* (KKT) *conditions and the second-order sufficient condition* (14). *Then, there are a neighborhood* V *of* x_* *and a real* $\bar{\sigma} > 0$ *such that for all* $\sigma \geq \bar{\sigma}$, (x_*, Z_*, λ_*) *is a saddle-point of* L_σ *on* $V \times (S_m \times I\!R^p)$. *Moreover,*

$$L_\sigma(x, Z_*, \lambda_*) > L_\sigma(x_*, Z_*, \lambda_*) \geq L_\sigma(x_*, Z, \lambda),$$

for all $(x, Z, \lambda) \in V \times S_m \times I\!R^p$ *with* $x \neq x_*$.

Proof. Since the operator $(\cdot)_+$, defined in (1), is the projection on S_m^+, we have that

$$\|(-\sigma^{-1}Z - \mathcal{A}(x))_+ - (-\sigma^{-1}Z - \mathcal{A}(x))\|_{Fr}^2 \leq \|W - (-\sigma^{-1}Z - \mathcal{A}(x))\|_{Fr}^2 \tag{24}$$

for all $W \in S_m^+$, and then

$$\begin{aligned} \frac{\sigma}{2}\|(-\sigma^{-1}Z - \mathcal{A}(x))_+ + \mathcal{A}(x)\|_{Fr}^2 &+ Tr(Z[(-\sigma^{-1}Z - \mathcal{A}(x))_+ + \mathcal{A}(x)]) \\ &\leq Tr(Z[W + \mathcal{A}(x)]) + \frac{\sigma}{2}\|W + \mathcal{A}(x)\|_{Fr}^2, \end{aligned} \tag{25}$$

for all $W \in S_m^+$. Taking $x = x_*$ and $W = -\mathcal{A}(x_*)$ (which belongs to S_m^+), we get

$$\frac{\sigma}{2}\|(-\sigma^{-1}Z - \mathcal{A}(x_*))_+ + \mathcal{A}(x_*)\|_{Fr}^2 + Tr(Z[(-\sigma^{-1}Z - \mathcal{A}(x_*))_+ + \mathcal{A}(x_*)]) \leq 0,$$

hence

$$L_\sigma(x_*, Z, \lambda) \leq f(x_*) = L_\sigma(x_*, Z_*, \lambda_*), \quad \forall Z \in S_m, \forall \lambda \in I\!R^p, \forall \sigma > 0.$$

Let us prove now the second inequality. Let $\bar{B}_\varepsilon(x_*)$ be a closed ball with center x_* and radius ε such that $f(x) > f(x_*)$ for all feasible points $x \in \bar{B}_\varepsilon(x_*)$, $x \neq x_*$. We prove that for all $\sigma > 0$ sufficiently large, x_* is the unique point satisfying $\inf_{x \in \bar{B}_\varepsilon(x_*)} L_\sigma(x, Z_*, \lambda_*) = f(x_*)$. For this purpose, we define the problem:

$$\psi_\sigma := \inf_{\substack{(x,\,W) \,\in\, \bar{B}_\varepsilon(x_*) \,\times\, S_m \\ \mathcal{A}(x) \,\preceq\, W}} \{f(x) + Tr(Z_*W) + \lambda_*^T h(x) + \frac{\sigma}{2}(\|h(x)\|^2 + \|W\|_{Fr}^2)\}, \tag{26}$$

and from inequality (25) we can deduce that

$$\psi_\sigma = \inf_{x \in \bar{B}_\varepsilon(x_*)} L_\sigma(x, Z_*, \lambda_*). \tag{27}$$

To conclude, we show that $(x_*, 0, Z_*, 0)$ is a point that satisfies the Karush-Kuhn-Tucker and the second-order sufficient conditions for the optimization problem (26). The Lagrangian associated with minimization problem (26) is

$$\begin{aligned} \tilde{L}(x, W, \Omega, \alpha) := f(x) &+ Tr(Z_*W) + \lambda_*^T h(x) + \frac{\sigma}{2}\|h(x)\|^2 \\ &+ \frac{\sigma}{2}\|W\|_{Fr}^2 + \frac{\alpha}{2}(\|x - x_*\|^2 - \varepsilon^2) + Tr(\Omega(\mathcal{A}(x) - W)), \end{aligned} \tag{28}$$

and the (KKT) conditions are

$$\begin{aligned}\nabla f(x) + \nabla h(x)^T \lambda_* + \sigma \nabla h(x)^T h(x) + \alpha(x - x_*) + D\mathcal{A}(x)^* \Omega = 0\\ Z_* + \sigma W = \Omega\\ \tfrac{\alpha}{2}(\|x - x_*\|^2 - \varepsilon^2) = 0\\ Tr(\Omega(\mathcal{A}(x) - W)) = 0\\ \alpha \geq 0, \quad \Omega \succeq 0\\ \|x - x_*\| - \varepsilon \leq 0\\ \mathcal{A}(x) - W \preceq 0.\end{aligned}$$

It can be easily seen that $(x, W, \Omega, \alpha) = (x_*, 0, Z_*, 0)$ satisfies all these conditions.

The Hessian of $\tilde{L}$ with respect to the variables (x, W) at $(x_*, 0, Z_*, 0)$ is given by

$$\nabla^2_{(x,W)} \tilde{L}(x_*, 0, Z_*, 0) = \left(\begin{array}{c|c} \begin{array}{c} \nabla^2 f(x_*) + \sum_{j=1}^{p} \lambda_{*j} \nabla^2 h_j(x_*) + \mathcal{H} \\ +\sigma \sum_{j=1}^{p} h_j(x_*) \nabla^2 h_j(x_*) + \sigma \nabla h(x_*)^T \nabla h(x_*) \end{array} & 0 \\ \hline 0 & \sigma I_m \end{array} \right), \tag{29}$$

with $\mathcal{H} := [Tr(Z_* \frac{\partial^2 \mathcal{A}(x_*)}{\partial x_i \partial x_j})]_{i,j}$, and the cone of critical directions for problem (26) is

$$\begin{aligned}\tilde{C}(x_*, 0) = \{(d, U) \in I\!R^n \times S_m : \ \nabla f(x_*)^T d + Tr(Z_* U) + \lambda_*^T \nabla h(x_*) d = 0,\\ \sum_{i=1}^{n} d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} - U \in S_m^-(Ker\ \mathcal{A}(x_*))\}.\end{aligned}$$

In what follows we will state the second-order sufficient condition

$$(d^T, U) \nabla^2 \tilde{L}(x_*, 0, Z_*, 0) \begin{pmatrix} d \\ U \end{pmatrix} > 0,$$

for any nonzero vector $(d, U) \in \tilde{C}(x_*, 0)$. This condition can be written as

$$d^T \nabla^2 f(x_*) d + \sum_{j=1}^{p} \lambda_{*j} d^T \nabla^2 h_j(x_*) d + d^T \mathcal{H} d + \sigma \|\nabla h(x_*) d\|^2 + \sigma \|U\|^2_{Fr} > 0, \tag{30}$$

for any nonzero vector $(d, U) \in \tilde{C}(x_*, 0)$.

The case when $d = 0$ and $U \neq 0$ is trivial. Another easy case is when we take $(d, U) \in \tilde{C}(x_*, 0)$ with $\|U\|_{Fr} > \delta\|d\|$ or $\|\nabla h(x_*)d\| > \delta\|d\|$ for some fixed $\delta > 0$, indeed

$$\begin{aligned} d^T\nabla^2 f(x_*)d + &\sum_{j=1}^{p}\lambda_{*j}d^T\nabla^2 h_j(x_*)d + d^T\mathcal{H}d + \sigma(\|\nabla h(x_*)d\|^2 + \|U\|_{Fr}^2) \\ &> d^T\nabla^2 f(x_*)d + \sum_{j=1}^{p}\lambda_{*j}d^T\nabla^2 h_j(x_*)d + d^T\mathcal{H}d + \sigma\delta^2\|d\|^2 \\ &\geq -\|\nabla^2 f(x_*) + \sum_{j=1}^{p}\lambda_{*j}\nabla^2 h_j(x_*) + \mathcal{H}\|\|d\|^2 + \sigma\delta^2\|d\|^2, \end{aligned}$$

and (30) is verified by taking $\sigma \geq \sigma_\delta := \frac{1}{\delta^2}\|\nabla^2 f(x_*) + \sum_{j=1}^{p}\lambda_{*j}\nabla^2 h_j(x_*) + \mathcal{H}\|$.

Finally we show that such a $\delta > 0$ always exists. We proceed by contradiction. Let us suppose that there is a sequence $\{(d_k, U_k)\}$ in $\tilde{C}(x_*, 0)$ such that

$$\|U_k\|_{Fr} \leq \frac{1}{k}\|d_k\|, \tag{31}$$

$$\|\nabla h(x_*)d_k\| \leq \frac{1}{k}\|d_k\| \tag{32}$$

and

$$d_k^T\nabla^2 f(x_*)d_k + \sum_{j=1}^{p}\lambda_{*j}d_k^T\nabla^2 h_j(x_*)d_k + d_k^T\mathcal{H}d_k \leq 0, \quad \forall k. \tag{33}$$

If we divide (33) by $\|d_k\|^2$ and suppose that $\frac{d_k}{\|d_k\|} \to \hat{d}$, by taking limit in this inequality we get

$$\hat{d}^T\nabla^2_{xx}L(x_*, Z_*, \lambda_*)\hat{d} \leq 0, \tag{34}$$

which means, by proposition 1, that $\hat{d} \notin C(x_*)$.

On the other hand, since $(d_k, U_k) \in \tilde{C}(x_*, 0)$, we have that

$$v^T\sum_{i=1}^{n}d_{ki}\frac{\partial\mathcal{A}(x_*)}{\partial x_i}v \leq v^TU_kv, \quad \text{for all } v \in Ker\ \mathcal{A}(x_*) \text{ and all } k,$$

and using the fact that $\|U_k\|_{Fr} \geq \frac{v^TU_kv}{\|v\|^2}$, for all $v \neq 0$, together with (31) we obtain

$$v^T\sum_{i=1}^{n}d_{ki}\frac{\partial\mathcal{A}(x_*)}{\partial x_i}v \leq \frac{1}{k}\|d_k\|\|v\|^2, \quad \text{for all } v \in Ker\ \mathcal{A}(x_*) \text{ and all } k,$$

which implies that

$$\sum_{i=1}^{n}\hat{d}_i\frac{\partial\mathcal{A}(x_*)}{\partial x_i} \in S_m^-(Ker\ \mathcal{A}(x_*)).$$

The equality $\nabla h(x_*)\hat{d} = 0$ follows directly from (32) and since $\nabla f(x_*)^T d_k = -Tr(Z_* U_k) - \lambda_*^T \nabla h(x_*) d_k$, from (31) and (32) we obtain that $\nabla f(x_*)^T \hat{d} = 0$. In this way we conclude that $\hat{d} \in C(x_*)$, which contradicts (34). ■

3.3 The Han penalty function

We define now another penalty function associated with problem (P), which will be a key issue in the global algorithm that we will describe in section 4. For $\sigma > 0$ we define

$$\theta_\sigma(x) = f(x) + \sigma(\lambda_1(\mathcal{A}(x))_+ + \|h(x)\|). \tag{35}$$

This function comes from the Han penalty function in mathematical programming [3, 10]. In the rest of this section we will prove some properties of θ_σ and its exactness.

In order to compute the directional derivative $\theta'_\sigma(x; d)$, we start by recalling a particular chain rule.

Lemma 5 *Let $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ be a function with directional derivative $\varphi'(x; d) = \lim_{t\to 0^+} t^{-1}(\varphi(x+td) - \varphi(x))$, and let $\phi : \mathbb{R}^m \to \mathbb{R}^p$ be a Lipschitz function in a neighborhood of $\varphi(x)$ with directional derivative $\phi'(\varphi(x); \varphi'(x; d))$. Then, the function $(\phi \circ \varphi)$ has a directional derivative at x in the direction d given by*

$$(\phi \circ \varphi)'(x; d) = \phi'(\varphi(x); \varphi'(x; d)). \tag{36}$$

Proof. By using the usual notation $o(t)$ for a function verifying $\lim_{t\to 0} t^{-1} o(t) = 0$, we can write for $t > 0$

$$\begin{aligned} t^{-1}[(\phi\circ\varphi)(x+td) - (\phi\circ\varphi)(x)] &= t^{-1}[\phi(\varphi(x) + t\varphi'(x; d) + o(t)) - (\phi\circ\varphi)(x)] \\ &= t^{-1}[\phi(\varphi(x) + t\varphi'(x; d)) - \phi(\varphi(x))] + t^{-1}o(t), \end{aligned}$$

and we can conclude by taking the limit when $t \to 0^+$. ■

As a consequence of this result we give in the next lemma the directional derivative of the penalty function θ_σ.

Lemma 6 *If f, h and $\mathcal{A}$ in (35) have directional derivatives at x in the direction d, where x is a feasible point for (P), then θ_σ also has a directional derivative that can be characterized by*

$$\theta'_\sigma(x; d) = f'(x; d) + \sigma(\lambda_1(N^T \mathcal{A}'(x; d) N)_+ + \|h'(x; d)\|),$$

where N is the matrix defined in (12).

Proof. Let x be a feasible point. From lemma 5, we have that

$$\begin{aligned} \theta'_\sigma(x; d) &= f'(x; d) + \sigma([\lambda_1(\mathcal{A}(\cdot))_+]'(x; d) + \|h'(x; d)\|) \\ &= f'(x; d) + \sigma([\lambda_1(\cdot)_+]'(\mathcal{A}(x); \mathcal{A}'(x; d)) + \|h'(x; d)\|). \end{aligned}$$

For $A \in S_m^-$ and $B \in S_m$, we easily check that

$$[\lambda_1(\cdot)_+]'(A;B) = \begin{cases} 0 & \text{, if } \lambda_1(A) < 0 \\ \lambda_1'(A;B)_+ & \text{, if } \lambda_1(A) = 0, \end{cases}$$

and using formula (2.8) in [7] for the calculus of the directional derivative of $\lambda_1(A) = \max\{x^T Ax : \|x\| = 1\}$, we obtain that $\lambda_1'(A;B) = \max\{x^T Bx : \|x\| = 1 \text{ and } x^T Ax = \lambda_1(A)\}$. Then, if $\lambda_1(A) = 0$, we can write $\lambda_1'(A;B) = \lambda_1(N^T BN)$ with N a matrix whose columns are an orthonormal base of $Ker\ A$.

We conclude replacing A by $\mathcal{A}(x)$, B by $\mathcal{A}'(x;d)$ and recalling that N is the matrix 0 when $\lambda_1(A) < 0$. ■

Remark 2 *If f, h and $\mathcal{A}$ are differentiable at x, then*

$$\theta_\sigma'(x;d) = \nabla f(x)d + \sigma(\lambda_1(N^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x)}{\partial x_i} N)_+ + \|\nabla h(x)d\|).$$

In the following proposition we give a lower bound for the parameter σ in order to obtain the exactness of θ_σ.

Proposition 7 *If x_* is a feasible point of (P) and θ_σ has a (strict) local minimum at x_*, then x_* is a (strict) local minimum of (P). Furthermore, if f, h and $\mathcal{A}$ are differentiable at x_* and if the transversality condition* (11) *is verified, then $\sigma \geq \max\{Tr(Z_*), \|\lambda_*\|\}$.*

Proof. If x_* is a local minimum of θ_σ, there is a neighborhood V of x_* such that for all $x \in V$ we have that $\theta_\sigma(x_*) \leq \theta_\sigma(x)$, and since x_* is feasible we obtain

$$\begin{aligned} f(x_*) = \theta_\sigma(x_*) &\leq \theta_\sigma(x) \quad \forall x \in V \\ &= f(x) \quad \forall x \in V,\ x \text{ feasible}, \end{aligned}$$

which means that x_* is a local minimum of (P). When the minimum x_* is strict, the proof is identical.

Now, let us assume that f, h and $\mathcal{A}$ are differentiable at x_* and that the transversality condition holds. Since x_* is a local minimum of θ_σ, we have that $\theta_\sigma'(x_*;d) \geq 0$ for all directions d and using lemma 6 we can write

$$0 \leq \nabla f(x_*)^T d + \sigma(\lambda_1(N^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N)_+ + \|\nabla h(x_*)d\|), \quad \forall d \in I\!R^n. \tag{37}$$

Let us first consider the case when $Ker\ \mathcal{A}(x_*) = \{0\}$. This implies $Z_* = 0$ and $N = 0$, hence from inequality (37) and the first equation in (KKT), we see that $\sigma \geq \frac{\lambda_*^T \nabla h(x_*)d}{\|\nabla h(x_*)d\|}$ for all nonzero $d \in I\!R^n$. The surjectivity of $\nabla h(x_*)$ shows that $\sigma \geq \|\lambda_*\|$.

Let us suppose now that $Ker\ \mathcal{A}(x_*) \neq \{0\}$. From the first equation in (KKT), inequality (37) and equality (18), we can write for all $d \in I\!R^n$

$$\begin{aligned}\sigma(\|\nabla h(x_*)d\| + \lambda_1(N^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N)_+) &\geq \lambda_*^T \nabla h(x_*)d + Tr(Z_* \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i}) \\ &= \lambda_*^T \nabla h(x_*)d + Tr(\phi_* N^T \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x_*)}{\partial x_i} N),\end{aligned}$$

Then, the transversality condition (surjectivity of ψ) allows us to say that

$$\sigma(\|v\| + \lambda_1(W)_+) \geq \lambda_*^T v + Tr(\phi_* W), \tag{38}$$

for all $(v, W) \in I\!R^p \times S_r$, and we can conclude from the inequality

$$\|W\|_2 := \sqrt{\max\{\lambda^2 : \lambda \text{ is an eigenvalue of } W\}} \geq \lambda_1(W)_+$$

and the equality

$$\|(\lambda_*, \phi_*)\|_D := \sup\{|\lambda_*^T v + Tr(\phi_* W)| : \|v\| + \|W\|_2 = 1\} = \max\{\|\lambda_*\|, Tr(Z_*)\}.$$

■

We conclude this section establishing sufficient conditions for exactness of the Han penalty function θ_σ. In proposition 9 we consider the convex case and in theorem 10 the general one.

The following useful lemma is a direct consequence of Frobenius inequality (7).

Lemma 8 *If $Z \succeq 0$ and $\sigma \geq \max\{Tr(Z), \|\lambda\|\}$, then $L(\cdot, Z, \lambda) \leq \theta_\sigma(\cdot)$.*

Proposition 9 *Let us consider the convex problem (P_C) and let us suppose that f, h and $\mathcal{A}$ are differentiable at a solution x_* of (P_C). Then, if (Z_*, λ_*) are (KKT)-multipliers associated with x_* and $\sigma \geq \max\{Tr(Z_*), \|\lambda_*\|\}$, we have that θ_σ has a global minimum in x_*.*

Proof. Let us suppose that (x_*, Z_*, λ_*) satisfies the (KKT) conditions. For the convex problem (P_C), it can be easily seen that $L(\cdot, Z_*, \lambda_*)$ has a global minimum at x_*, that is, $\theta_\sigma(x_*) = f(x_*) = L(x_*, Z_*, \lambda_*) \leq L(x, Z_*, \lambda_*)$ for all x. If $\sigma \geq \max\{Tr(Z_*), \|\lambda_*\|\}$, from lemma 8 we have $L(\cdot, Z_*, \lambda_*) \leq \theta_\sigma(\cdot)$, which leads to the desired result. ■

Theorem 10 *Let us suppose that f, h and $\mathcal{A}$ are differentiable at x_*. Let (x_*, Z_*, λ_*) be a point that satisfies the (KKT) conditions and the second-order sufficient condition (14). If $\sigma > \max\{Tr(Z_*), \|\lambda_*\|\}$, then θ_σ has a strict local minimum in x_*.*

Proof. Taking $Z = Z_*$ and $W = (-\mathcal{A}(x))_+ = -(\mathcal{A}(x))_-$ in (25) we have that

$$\begin{aligned}\frac{r}{2}\|(-\frac{Z_*}{r} - \mathcal{A}(x))_+ + \mathcal{A}(x)\|_{Fr}^2 + Tr(Z_*[(-\frac{Z_*}{r} - \mathcal{A}(x))_+ + \mathcal{A}(x)]) \\ \leq Tr(Z_*\mathcal{A}(x)_+) + \frac{r}{2}\|\mathcal{A}(x)_+\|_{Fr}^2.\end{aligned} \tag{39}$$

Hence, using Cauchy-Schwarz and Frobenius (7) inequalities, we obtain

$$\begin{aligned} L_r(x, Z_*, \lambda_*) &\leq f(x) + \lambda_*^T h(x) + \frac{r}{2}\|h(x)\|^2 + Tr(Z_*\mathcal{A}(x)_+) + \frac{r}{2}\|\mathcal{A}(x)_+\|_{Fr}^2 \\ &\leq f(x) + \|h(x)\|(\|\lambda_*\| + \frac{r}{2}\|h(x)\|) + \lambda_1(\mathcal{A}(x))_+\big(Tr(Z_*) + \frac{r}{2}\sum_{j=1}^{m}\lambda_j(\mathcal{A}(x)_+)\big).\end{aligned}$$

The last inequality follows from $\lambda_1(\mathcal{A}(x)_+) = \lambda_1(\mathcal{A}(x))_+$. Since $\sigma > \max\{Tr(Z_*), \|\lambda_*\|\}$, for any fixed $r > 0$ there is a neighborhood V_r of x_* such that

$$L_r(x, Z_*, \lambda_*) \leq f(x) + \sigma(\|h(x)\| + \lambda_1(\mathcal{A}(x))_+) = \theta_\sigma(x), \quad \forall x \in V_r.$$

From theorem 4, we know that there is an $\bar{r} > 0$ and a neighborhood $\bar{V}$ of x_* where x_* is a strict minimum of $L_{\bar{r}}(\cdot, Z_*, \lambda_*)$. This implies that x_* is a strict minimum of θ_σ on $\bar{V} \bigcap V_{\bar{r}}$
■

4 Sequentially Semidefinite Programming

In this section we propose a global S-SDP algorithm for solving problem (P). This algorithm is inspired by the classical Sequentially Quadratic Programming (SQP). We begin by recalling the local S-SDP algorithm proposed in [8] and its convergence theorem.

Given an initial point (x_0, Z_0, λ_0) close to a point (x_*, Z_*, λ_*) that satisfies the (KKT) conditions, we generate a sequence (x_k, Z_k, λ_k) by solving the tangent problem:

$$(T_k) \quad \begin{array}{ll} \text{minimize}_{d\in \mathbb{R}^n} & \nabla f(x_k)^T d + \frac{1}{2} d^T M_k d \\ \text{subject to} & \mathcal{A}_k(d) \preceq 0 \\ & h(x_k) + \nabla h(x_k) d = 0. \end{array}$$

Where $\mathcal{A}_k(d) := \mathcal{A}(x_k) + \sum_{i=1}^{n} d_i \frac{\partial \mathcal{A}(x_k)}{\partial x_i}$, and the matrix M_k replaces the Hessian $\nabla_{xx}^2 L(x_k, Z_k, \lambda_k)$, emulating the so-called Quasi-Newton methods.

If d_k is the solution of (T_k), we define $x_{k+1} = x_k + d_k$. The point $(d_k, Z_{k+1}, \lambda_{k+1})$ is obtained from the (KKT) conditions for the minimization problem (T_k), that is

$$(KKT_k) \quad \begin{aligned} \nabla f(x_k) + D\mathcal{A}(x_k)^* Z_{k+1} + \nabla h(x_k)^T \lambda_{k+1} + M_k d_k &= 0 \quad (40a) \\ \mathcal{A}_k(d_k) &\preceq 0 \quad (40b) \\ h(x_k) + \nabla h(x_k) d_k &= 0 \quad (40c) \\ Z_{k+1} &\succeq 0 \quad (40d) \\ Tr(Z_{k+1}\mathcal{A}_k(d_k)) &= 0 \quad (40e) \end{aligned}$$

Theorem 11 *Let (x_*, Z_*, λ_*) be a point satisfying the (KKT) conditions and the second-order sufficient condition given in proposition 3. Suppose that $(D\mathcal{A}(x_*), \nabla h(x_*)^T)^T$ has full rank, and that $M_k \to \nabla^2_{xx} L(x_*, Z_*, \lambda_*)$. Then, there is $\delta > 0$ such that if $\|x_0 - x_*\| < \delta$, $\|(Z_0, \lambda_0) - (Z_*, \lambda_*)\| < \delta$ and $\|M_k - \nabla^2_{xx} L(x_*, Z_*, \lambda_*)\| < \delta$ for all k, the sequence (x_k, Z_k, λ_k) generated by the algorithm S-SDP is well defined and converges superlinearly to (x_*, Z_*, λ_*). The convergence is even quadratic if $M_k - \nabla^2_{xx} L(x_*, Z_*, \lambda_*) = O(\|x_k - x_*\| + \|(Z_k, \lambda_k) - (Z_*, \lambda_*))\|$.*

Our purpose here is to extend the S-SDP algorithm to obtain the global convergence. For this, we consider the Han penalty function, defined in (35), and an *Armijo linear search.*

In the following proposition we prove that the solution d_k of (T_k) is a descent direction for θ_σ at the point x_k when M_k is positive definite and σ is sufficiently large.

Proposition 12 *Suppose that f, h and $\mathcal{A}$ are C^1 functions and that their derivatives are locally Lipschitz at x_k. Using the penalty function θ_σ, defined in (35), if $(d_k, Z_{k+1}, \lambda_{k+1})$ verifies the (KKT_k) conditions, written in (40), then*

$$\begin{aligned} \theta'_\sigma(x_k; d_k) &\leq \nabla f(x_k)^T d_k - \sigma(\lambda_1(\mathcal{A}(x_k))_+ + \|h(x_k)\|) \\ &= -d_k^T M_k d_k + Tr(Z_{k+1}\mathcal{A}(x_k)) + \lambda_{k+1}^T h(x_k) - \sigma(\lambda_1(\mathcal{A}(x_k))_+ + \|h(x_k)\|). \end{aligned} \tag{41}$$

Furthermore, if $\sigma \geq \max\{Tr(Z_{k+1}), \|\lambda_{k+1}\|\}$ we obtain

$$\theta'_\sigma(x_k; d_k) \leq -d_k^T M_k d_k. \tag{42}$$

Proof. Let us fix $t \in [0,1]$. By (40c) we have that

$$\|h(x_k) + t\nabla h(x_k) d_k\| = (1-t)\|h(x_k)\|.$$

and by the convexity of $\lambda_1(\cdot)_+$ and (40b) we obtain

$$\begin{aligned} \lambda_1(\mathcal{A}_k(td_k))_+ &= \lambda_1(\mathcal{A}(x_k) + t\sum_{i=1}^n d_{ki}\frac{\partial \mathcal{A}(x_k)}{\partial x_i})_+ \\ &\leq (1-t)\lambda_1(\mathcal{A}(x_k))_+ + t\lambda_1(\mathcal{A}_k(d_k))_+ = (1-t)\lambda_1(\mathcal{A}(x_k))_+. \end{aligned}$$

From these relations we have that

$$\|\cdot\|'(h(x_k); \nabla h(x_k) d_k) = -\|h(x_k)\|, \tag{43a}$$

$$(\lambda_1(\cdot)_+)'(\mathcal{A}(x_k); \sum_{i=1}^n d_{ki}\frac{\partial \mathcal{A}(x_k)}{\partial x_i}) \leq -\lambda_1(\mathcal{A}(x_k))_+. \tag{43b}$$

and applying lemma 5 we get

$$\theta'_\sigma(x_k; d_k) \leq \nabla f(x_k)^T d_k - \sigma(\lambda_1(\mathcal{A}(x_k))_+ + \|h(x_k)\|),$$

and from (40a), (40c) and (40e) we obtain

$$\theta'_\sigma(x_k;d_k) \leq Tr(Z_{k+1}\mathcal{A}(x_k)) + h(x_k)^T\lambda_{k+1} - \sigma(\|h(x_k)\| + \lambda_1(\mathcal{A}(x_k))_+) - d_k^T M_k d_k.$$

Finally, if $\sigma \geq \max\{Tr(Z_{k+1}), \|\lambda_{k+1}\|\}$, the Cauchy-Schwarz and Frobenius (7) inequalities lead to the result. ■

We are now ready to describe the iteration k of the global algorithm for solving problem (P). We suppose that x_k is known.

Step 1 Compute a point $(d_k, Z_{k+1}, \lambda_{k+1})$ satisfying (KKT_k) in (40).

Step 2 Compute σ_k satisfying $\sigma_k \geq \max\{Tr(Z_{k+1}), \|\lambda_{k+1}\|\}$, in such a way that the sequence $\{\sigma_k\}$ satisfies the following properties:

$$\begin{array}{ll} (a) & \sigma_k \geq \max\{Tr(Z_{k+1}), \|\lambda_{k+1}\|\} + \bar{\sigma}. \\ (b) & \text{For all } k \geq k_1, \\ & \text{if } \sigma_{k-1} \geq \max\{Tr(Z_{k+1}), \|\lambda_{k+1}\|\} + \bar{\sigma}, \text{ then } \sigma_k = \sigma_{k-1}. \\ (c) & \text{If } \{\sigma_k\} \text{ is bounded, then } \sigma_k \text{ is modified just finitely many times,} \end{array} \tag{44}$$

where $k_1 \in I\!N$ and $\bar{\sigma} > 0$ are fixed parameters. A simple way to update σ_k verifying (44) is defining $\sigma_k = \max\{1.5\sigma_{k-1}, \max\{Tr(Z_{k+1}), \|\lambda_{k+1}\|\} + \bar{\sigma}\}$ when (b) fails.

Step 3 The step length α_k is computed by using an *Armijo search rule*, that is, α_k is an approximation of the maximum $\alpha \in (0,1]$ which verifies

$$\theta_{\sigma_k}(x_k + \alpha d_k) \leq \theta_{\sigma_k}(x_k) + w\alpha\Delta_k. \tag{45}$$

Where $0 < w < 1$ and Δ_k is the upper bound of $\theta'_{\sigma_k}(x_k;d_k)$ given in (41). The existence of α_k is guaranteed by inequality (45) and the fact that M_k is positive definite. More precisely, α_k can be computed as follows:

Step 0 $j := 0, \quad r_j := 1.$

Step 1 If (45) is satisfied with $\alpha = r_j$ then $\alpha_k = r_j$ and stop the linear search.

Step 2 If not, take $r_{j+1} = \beta r_j$, increase j by one and go to Step 1. with $\beta \in (0,1)$ a fixed constant.

Step 4 Define $x_{k+1} = x_k + \alpha_k d_k$.

Theorem 13 *Let us suppose that f, h and $\mathcal{A}$ are C^1 functions and that their derivatives are Lipschitz. If we consider the global algorithm described in the steps 1 to 4 and suppose that the matrices M_k are chosen positive definite such that the sequence $\{M_k\}$ is bounded together with the sequence $\{M_k^{-1}\}$, then, one of the following situations occurs for the sequence $(x_k, Z_{k+1}.\lambda_{k+1})$:*

1. *The sequences $\{\sigma_k\}$ and $\{(Z_{k+1}, \lambda_{k+1})\}$ are unbounded.*
2. *There is an index k_2 such that σ_k is constant for all $k \geq k_2$. In this case one of the following situations occurs:*

(a) $\theta_{\sigma_k}(x_k) \to -\infty$, *or*

(b) $\nabla_x L(x_k, Z_{k+1}, \lambda_{k+1}) \to 0$, $h(x_k) \to 0$, $\lambda_1(\mathcal{A}(x_k))_+ \to 0$ *and* $Tr(Z_{k+1}\mathcal{A}(x_k)) \to 0$

Proof. 1) The equivalence between the unboundedness of $\{\sigma_k\}$ and $\{(Z_{k+1}, \lambda_{k+1})\}$ is direct from (44)(a)-(b).

2) Let us suppose that $\{\sigma_k\}$ is bounded. By (44)(c) we know that there is an index k_2 such that $\sigma_k = \sigma := \sigma_{k_2}$ for all $k \geq k_2$.

To conclude we prove that if (a) is not true then (b) holds. From (45), with $\alpha = \alpha_k$, we know that the sequence $\{\theta_{\sigma_k}(x_k)\}$ is decreasing for all $k \geq k_2$, and then $\theta_{\sigma_k}(x_k) \geq C$ for some constant C, obtaining again from (45) the limit $\alpha_k \Delta_k \to 0$.

All limits in (b) are consequences of the existence of $\bar{\alpha} > 0$ such that $\alpha_k \geq \bar{\alpha}$ for all $k \geq k_2$, which implies from the limit above that $\Delta_k \to 0$. Indeed, inequalities

$$\begin{aligned}\Delta_k &\leq -d_k^T M_k d_k + \lambda_1(\mathcal{A}(x_k))_+ Tr(Z_{k+1}) + \|\lambda_{k+1}\| \|h(x_k)\| - \sigma(\lambda_1(\mathcal{A}(x_k))_+ + \|h(x_k)\|) \\ &\leq -d_k^T M_k d_k + (\sigma - \bar{\sigma})(\lambda_1(\mathcal{A}(x_k))_+ + \|h(x_k)\|) - \sigma(\lambda_1(\mathcal{A}(x_k))_+ + \|h(x_k)\|)\end{aligned}$$

prove that

$$\Delta_k \leq -d_k^T M_k d_k - \bar{\sigma}(\lambda_1(\mathcal{A}(x_k))_+ + \|h(x_k)\|) \leq 0, \tag{46}$$

which implies that if $\Delta_k \to 0$ then

$$\lambda_1(\mathcal{A}(x_k))_+ \to 0 \quad \text{and} \quad h(x_k) \to 0.$$

Inequality (46) also shows that $-d_k^T M_k d_k \to 0$, and together with the fact that $\{M_k^{-1}\}$ is bounded, it is easy to conclude that $d_k \to 0$. This, together with (40a) and the boundedness of $\{M_k\}$, implies that

$$\nabla_x L(x_k, Z_{k+1}, \lambda_{k+1}) \to 0.$$

Finally, by definition of Δ_k we know that

$$\nabla f(x_k)^T d_k = \Delta_k + \sigma_k(\lambda_1(\mathcal{A}(x_k))_+ + \|h(x_k)\|) \to 0,$$

and from (40a) we see that

$$Tr(Z_{k+1} \sum_{i=1}^n d_{ki} \frac{\partial \mathcal{A}(x_k)}{\partial x_i}) + \lambda_{k+1}^T \nabla h(x_k) d_k = -d_k^T M_k d_k - \nabla f(x_k)^T d_k \to 0.$$

This limit and the equalities (40e) and (40c), together with the boundedness of $\{\lambda_{k+1}\}$, allow us to write

$$\begin{aligned}\lim_{k\to+\infty} Tr(Z_{k+1}\mathcal{A}(x_k)) &= \lim_{k\to+\infty} -Tr(Z_{k+1} \sum_{i=1}^n d_{ki} \frac{\partial \mathcal{A}(x_k)}{\partial x_i}) \\ &= \lim_{k\to+\infty} \lambda_{k+1}^T \nabla h(x_k) d_k = \lim_{k\to+\infty} -\lambda_{k+1}^T h(x_k) = 0.\end{aligned}$$

Let us prove now that $\alpha_k \geq \bar{\alpha} \geq 0$. If $\alpha_k < 1$, by the Armijo search rule, there is an $r_j \in (0,1]$ such that $\alpha_k = \beta r_j$ and

$$\theta_{\sigma_k}(x_k + r_j d_k) > \theta_{\sigma_k}(x_k) + w r_j \Delta_k. \tag{47}$$

Let us consider the first order Taylor expansion

$$\begin{aligned} f(x_k + r_j d_k) &= f(x_k) + r_j \nabla f(x_k) d_k + O(r_j^2 \|d_k\|^2), \\ h(x_k + r_j d_k) &= h(x_k) + r_j \nabla h(x_k) d_k + O(r_j^2 \|d_k\|^2), \\ &= (1 - r_j) h(x_k) + r_j (h(x_k) + \nabla h(x_k) d_k) + O(r_j^2 \|d_k\|^2), \\ \mathcal{A}(x_k + r_j d_k) &= \mathcal{A}(x_k) + r_j \sum_{i=1}^{n} d_{ki} \frac{\partial \mathcal{A}(x_k)}{\partial x_i} + O(r_j^2 \|d_k\|^2) \\ &= (1 - r_j)\mathcal{A}(x_k) + r_j (\mathcal{A}(x_k) + \sum_{i=1}^{n} d_{ki} \frac{\partial \mathcal{A}(x_k)}{\partial x_i}) + O(r_j^2 \|d_k\|^2). \end{aligned}$$

Since $r_j \leq 1$, from the convexity of $\lambda_1(\cdot)_+$, and the relations (40b) and (40c) we obtain

$$\begin{aligned} \|h(x_k + r_j d_k)\| &= (1 - r_j)\|h(x_k)\| + O(r_j^2 \|d_k\|^2), \\ \lambda_1(\mathcal{A}(x_k + r_j d_k))_+ &\leq (1 - r_j)\lambda_1(\mathcal{A}(x_k))_+ + O(r_j^2 \|d_k\|^2), \end{aligned}$$

which imply from (47) the inequality

$$\theta_{\sigma_k}(x_k) + r_j \Delta_k + C_1 r_j^2 \|d_k\|^2 > \theta_{\sigma_k}(x_k) + w r_j \Delta_k,$$

that is, $-(1-w)\bar{\alpha}_k \Delta_k < C_1 r_j^2 \|d_k\|^2$, for some constant $C_1 > 0$. Due to inequality (46), the boundedness of $\{M_k^{-1}\}$ and the fact that M_k is positive definite, we see that $\Delta_k \leq -C_2 \|d_k\|^2$ for some constant $C_2 > 0$. The last two inequalities show that

$$r_j \geq \frac{C_2}{C_1}(1 - w) > 0,$$

and the proof is complete. ■

Acknowledgments

The authors are grateful to Frédéric Bonnans and Dominik Noll for fruitful discussions concerning the topic of this paper.

References

[1] P. Apkarian and D. Noll. A prototype primal-dual LMI-interior algorithm for nonconvex robust control problems. Submitted, 2001.

[2] J. F. Bonnans, R. Cominetti, and A. Shapiro. Second-order optimality conditions based on parabolic second-order tangent sets. *SIAM J. on Optimization*, 9(2):pp. 466–492, 1999.

[3] J. F. Bonnans, J-C. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Numerical Optimisation.* Springer-Verlag, Berlin, 2002.

[4] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems.* Springer-Verlag, New York, 2000.

[5] B. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in Systems and Control Theory.* vol. 15 of SIAM Studies in Applied Mathematics, SIAM, Philadelphia, 1994.

[6] R. Cominetti. Metric regularity, tangent sets and second-order optimality conditions. *Applied Mathematics and Optimization*, 21:pp. 265–287, 1990.

[7] R. Correa and A. Seeger. Directional derivatives of minimax functions. *Nonlinear Analysis*, 9:pp. 13–22, 1985.

[8] B. Fares, D. Noll, and P. Apkarian. Robust control via sequential semidefinite programming. *SIAM J. Control Optim.*, 40(6):pp. 1791–1820, 2002.

[9] M. Golubitsky and V. Guillemin. *Stable mappings and their singularities.* Springer, New York, 1973.

[10] S. P. Han. A globally convergent method for nonlinear programming. *J. Optim. Theory Appl.*, 22:pp. 297–309, 1977.

[11] J-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I.* Springer-Verlag, Berlin, 1993.

[12] R. Horn and C. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[13] S. Kurcyusz. On the existence and nonexistence of Lagrange multipliers in banach spaces. *Journal of Optimization Theory and Applications*, 20(1):pp. 81–110, 1976.

[14] J. Nocedal and S. J. Wright. *Numerical optimization.* Springer series in operations research, New York, 1999.

[15] S. M. Robinson. First-order conditions for general nonlinear optimization. *SIAM Journal on Applied Mathematics*, 30(4):pp. 597–607, 1976.

[16] S. M. Robinson. Stability theorems for systems of inequalities, Part II: differentiable nonlinear systems. *SIAM J. Numerical Analysis*, 13:pp. 497–513, 1976.

[17] S. M. Robinson. Generalized equations and their solutions, part II: Applications to nonlinear programming. *Math Programming Stud.*, 19:pp. 200–221, 1982.

[18] R. T. Rockafellar and R. J-B. Wets. *Variational analysis.* Springer-Verlag, Berlin, 1998.

[19] A. Shapiro. First and second-order analysis of nonlinear semidefinite programs. *Mathematical Programming*, 77(2):pp. 301–320, 1997.

[20] C. M. Theobald. An inequality for the trace of the product of two symmetric matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, pages pp. 77–265, 1975.

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

[22] H. Wolkowitz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming: Theory, Algorithms and Applications.* Kluwer's International Series in Operations Research and Management Science, 2000.