KYBERNETIKA — VOLUME *49* (2013), NUMBER 2, PAGES 236–257

A NEW REGULAR MULTIPLIER EMBEDDING

Gemayqzel Bouza Allende and Jürgen Guddat

Embedding approaches can be used for solving non linear programs P. The idea is to define a one-parametric problem such that for some value of the parameter the corresponding problem is equivalent to P. A particular case is the multipliers embedding, where the solutions of the corresponding parametric problem can be interpreted as the points computed by the multipliers method on P. However, in the known cases, either path-following methods can not be applied or the necessary conditions for its convergence are fulfilled under very restrictive hypothesis. In this paper, we present a new multipliers embedding such that the objective function and the constraints of $P(t)$ are C^3 differentiable functions. We prove that the parametric problem satisfies the JJT-regularity generically, a necessary condition for the success of the path-following method.

Keywords: Jongen–Jonker–Twilt regularity, multipliers method, embedding

Classification: 90C31, 49M30

1. INTRODUCTION

Let us consider the nonlinear optimization problem:

$$(P) \qquad \min\{f(x) \mid x \in M\},$$

$$M = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} h_i(x) = 0, & i = 1, \dots, m, \\ g_j(x) \geq 0, & j = 1, \dots, s. \end{array} \right\}$$

An appealing solution approach is to embed P into a one-parametric problem, $\Phi(P)$, which for all $t \in [0,1]$ defines the optimization problem

$$\Phi(P)(t) \qquad \min\{\overline{f}(y,t) |\; y \in M(t)\}$$

$$M(t) = \left\{ y \in \mathbb{R}^{\overline{n}} \;\middle|\; \begin{array}{ll} \overline{h}_i(y,t) = 0, & i = 1, \dots, \overline{m}, \\ \overline{g}_j(y,t) \geq 0, & j = 1, \dots, \overline{s}, \end{array} \right\}$$

where $\overline{f}, \overline{h}_i,\ i = 1, \dots, \overline{m},\ \overline{g}_j,\ j = 1, \dots, \overline{s} : \mathbb{R}^{\overline{n}} \times \mathbb{R} \to \mathbb{R}$ are directly determined by $f, h_i,\ i = 1, \dots, m,\ g_j,\ j = 1, \dots, s$. Moreover, $\Phi(P)(0)$ has an evident solution and problem $\Phi(P)(1)$ is equivalent to P.

Let us define $\Sigma(\Phi(P)) = \{(y,t) : y \text{ solves } \Phi(P)(t)\}$ as the set of solutions of the one-parametric problem $\Phi(P)$. An approach for obtaining $(y,1) \in \Sigma(\Phi(P))$, and hence a solution of P, is the path-following method. The idea is the following, starting with $(y, 0) \in \Sigma(\Phi(P))$, the goal is to try to reach $t = 1$ using a procedure that, given $(y, t) \in \Sigma(\Phi(P))$, compute points near to (y, t) applying continuation methods at the paths of solutions containing (y, t). If $(y, t) \in \Sigma(\Phi(P)(t))$ is of one of the five types of solutions studied in Jongen et al. [17, 18], these paths are well defined and provide a local description of $\Sigma(\Phi(P))$. So, the approach will be able to recover the local structure of the set of solutions $(\Phi(P)$. That is why it is important to study how strong is to assume that the elements of $\Sigma(\Phi(P))$ are of one of these types, i. e. $\Phi(P)$ is JJT-regular, i. e., if the set of functions $f, h_i,\ i = 1, \ldots, m, g_j,\ j = 1, \ldots, s$, defining P such that $\Sigma(\Phi(P))$ is JJT-regular, is large or not. This analysis is done from a generic viewpoint, recall that a subset of a topological space, is generic if it is the countable intersection of open and dense subsets of the space. Due to for the JJT-regularity, three times differentiability is needed. So, in this framework, we will consider the set of the three-times continuously differentiable functions endowed with the strong topology, see [15] as the topological space. Therefore, we will study if the JJT-regularity hold for a generic set of three-times continuously differentiable functions $f, h_i,\ i = 1, \ldots, m, g_j,\ j = 1, \ldots, s$.

Some embeddings are inspired in other algorithms, such as penalty methods and modified Newton algorithm, see Gómez et al. [12] and Schmidt [22] respectively. Indeed, at least locally around y, evident solution of $\Phi(P)(0)$, the elements of $\Sigma(\Phi(P))$ can be interpreted as the points computed by the method applied to P. For the multipliers method, two embeddings have been defined in Dentcheva et al. [9]. They fulfill that the objective function includes the Lagrange function plus a quadratic penalty term and that, locally around $t = 0$, a curve of saddle points can be numerically tracked. However the corresponding parametric problem is not JJT-regular for generic non linear programs with equality and inequality constraints, see Bouza [7] and Bouza et al. [8].

In this paper a new multipliers embedding is presented. It is shown that the JJT-regularity is fulfilled for a generic non-linear problem with equality and inequality constraints. Moreover it is proven that, for almost every quadratic perturbation of f and linear perturbation of $(h_1, \ldots, h_m, g_1, \ldots, g_s)$, the embedding constructs a regular problem.

The paper is organized as follows. Now, a brief introduction to multipliers' method and the theory of one-parameter optimization is given. The third section is dedicated to the analysis of some properties of the proposed embedding and two examples illustrate its numerical behavior. The main results are proven in Section 4.

2. PRELIMINARY ASPECTS AND NOTATIONS

In this part we present the notations as well as some definitions and results of parametric optimization and the multipliers method.

Notations

Let us define the following optimization problem

$$(P) \quad \min\{f(x) \mid x \in M\}, \tag{1}$$

where the set of feasible points is

$$M = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} h_i(x) = 0, & i = 1, \dots, m, \\ g_j(x) \geq 0, & j = 1, \dots, s, \end{array} \right\} \tag{2}$$

and the parametric problem:

$$(P(t)) \quad \min\{f(x,t) \mid x \in M(t)\}, \tag{3}$$

where

$$M(t) = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} h_i(x,t) = 0, & i = 1, \dots, m, \\ g_j(x,t) \geq 0, & j = 1, \dots, s, \end{array} \right\} \tag{4}$$

shortly denoted as $P(t)_{f,H,G}$ where $H = (h_1(x,t), \dots, h_m(x,t))$ and $G = (g_1(x,t), \dots, g_s(x,t))$.

$J_+(\mu) = \{j : \mu_j \neq 0\}$. I_m represents the identity matrix of dimension m and the space of symmetric $n \times n$-matrices be identified by $\mathbb{R}^{n(n+1)/2}$. $0_{n\times m}$ is a zero matrix of $n \times m$. $\Pi_x : \mathbb{R}^{n+l} \to \mathbb{R}^n$, $\Pi_x(x,y) = x$, the projecting function onto $\mathbb{R}^n$. Analogously, if $A \subset \mathbb{R}^{n+l}$, $\Pi_x(A)$ is the projection of A in $\mathbb{R}^n$. The indicator function $1_A(x)$ of the set A is $1_A(x) = \begin{cases} 1, & x \in A; \\ 0, & x \notin A. \end{cases}$

We will denote as $C^k(\mathbb{R}^n, \mathbb{R}^m)$, the set of k-times continuously differentiable functions $F : \mathbb{R}^n \to R^m$. So, a problem P can be identified as $(f, h_1, \dots, h_m, g_1, \dots, g_s)$, an element of $C^k(\mathbb{R}^n, \mathbb{R}^{1+m+s})$. We will assume that this space is endowed with the strong topology, that is a topology defined by the open neighborhoods of 0:

$$V_{\varepsilon(x)}(0) = \{F \in C^k(\mathbb{R}^n, \mathbb{R}^{1+m+s}) : |\partial^r F(x)| < \varepsilon(x), \text{ for all } x \in \mathbb{R}^n, 0 \leq r \leq k\},$$

where $\varepsilon(x)$ is a continuous positive function, for more details see Hirsch [15]. Here $\partial^r F(x)$ denotes the derivative of order r of F.

Given a topological space, a set is generic if it is the intersection of countable many open and dense sets. A property $\mathbb{P}$ is generically fulfilled if there is a generic set such that all its elements satisfy $\mathbb{P}$. Taking into account the characteristics of $C^k(\mathbb{R}^{n+1}, \mathbb{R}^{1+m+s})$ endowed with the strong topology, a generic property, is not too restrictive.

On JJT-regularity for one parametric programs

Now let us present some definitions and results of parametric optimization, for more details see Gómez et al. [13] and Guddat et al. [14]. Classical definitions of non-linear programming are easily extended to the one-parametric case:

$J_0(x,t) = \{j \mid g_j(x,t) = 0\}$ is the active index set of (x,t).
$L(x, \lambda, \mu, t) = f(x,t) - \sum_{i=1}^m \lambda_i h_i(x,t) - \sum_{j \in J_0} \mu_j g_j(x,t)$ is the Lagrangian function of $P(t)$, and λ_i, $i = 1, \dots, m$, μ_j, $j \in J_0(x,t)$, are the associated Lagrange multipliers.

Now let us present the constraints qualifications:

The Linear Independence Constraint Qualification (*LICQ*) holds at (x,t), if the vectors $\{\nabla_x h_i(x,t),\ i = 1, \dots, m,\ \nabla_x g_j(x,t),\ j \in J_0(x,t)\}$ are linearly independent.

The Mangasarian Fromovitz Constraint Qualification (MFCQ), is satisfied if

(MF1) the vectors $\nabla_x h_1(x,t),\dots,\nabla_x h_m(x,t)$ are linearly independent.

(MF2) there is $\xi \in \mathbb{R}^n$ such that $\nabla_x h_i(x,t)\xi = 0$, $i=1,\dots,m$ and $\nabla_x g_j(x,t)\xi > 0$, $j \in J_0(x,t)$.

In general, the solutions of non-linear optimization problems are difficult to compute. So, weaker conditions which allow a practical description are considered as solutions. That is why in this paper, the set of solutions of $P(t)$ is considered as $\Sigma_{gc}(P(t))$, the set of generalized critical points of $P(t)$. We say that $(x,t) \in \Sigma_{gc}(P(t))$ is a generalized critical point (g.c. point), if $x \in M(t)$ and there exists $(\mu_0,\lambda,\mu) \neq 0$, $\mu_0 \geq 0$, such that

$$\mu_0 \nabla_x f(x,t) + \sum_{i=1}^{m} \lambda_i \nabla_x h_i(x,t) + \sum_{j \in J_0(x,t)} \mu_j \nabla_x g_j(x,t) = 0. \tag{5}$$

Note that if $\mu_i \geq 0$, the Fritz John necessary optimality condition is recovered, see Bazaraa et al. [4]. So, in particular, local minimizers are generalized critical points. The study of this larger set is also very interesting because Σ_{gc} includes the closure of the set of local minimizers.

In case $(x,t) \in \Sigma_{gc}$ and LICQ holds, we will say that (x,t) is a critical point. This kind of points are very important because we can guarantee that $\mu_0 \neq 0$ at (5), and, hence, the objective function plays a role at the optimality condition. Without loss of generality μ_0 is taken as 1.

For the local structure of Σ_{gc}, we use the five Types of g.c. points defined by Jongen, Jonker and Twilt [18], [17]. First we present the definition of non-degenerated critical points.

Definition 2.1. *Type 1* — $(\overline{x},\overline{t})$ is a non-degenerated critical point, or a point of Type 1 if

(1a) LICQ holds.

(1b) $J_0(\overline{x},\overline{t}) = J_+(\overline{\mu})$.

(1c) $\nabla_x^2(\overline{x},\overline{t})L(\overline{x},\overline{\lambda},\overline{\mu},\overline{t})|_{T_xM(t)}$ is non-singular.

Here $T(x,t) = \left\{ \xi \in \mathbb{R}^n \,\middle|\, \begin{array}{lcll} \nabla_x h_i(x,t)^T\xi & = & 0, & i=1,\dots,m, \\ \nabla_x g_j(x,t)^T\xi & = & 0, & j \in J_0(x,t) \end{array} \right\}$ is the tangent space at (x,t) and $\nabla_x^2 L(x,t)|_{T(x,t)}$ represents $V^T \nabla_x^2 L V$, where V is a matrix whose columns are a basis of $T(x,t)$.

In this case Σ_{gc} is locally described as a curve $(x(t),t)$. The points of the Types 2–5 represent four basic degeneracies based on the failure of the conditions (1a), (1b) and (1c). For details on their definition and properties, we refer to Jongen et al. [17], Gómez et al. [13], and Guddat et al. [14].

Type 2 — There exists a unique $j^* \in J_0(\overline{x},\overline{t})$ such that $\mu_{j^*} = 0$, (1a) and (1c) hold. In this case, locally around $(\overline{x},\overline{t})$, the active indexes combinations of the elements of Σ_{gc} are $J_0(\overline{x},\overline{t})$ and $J_0(\overline{x},\overline{t}) \setminus \{j^*\}$. The definition is completed with two conditions. One is the non degeneracy of $(\overline{x},\overline{t})$ as g.c. point of two problems P^1 and P^2, where locally around $(\overline{x}, \overline{t})$ their g.c. points include $\Sigma_{gc}(P(t))$. The other guarantee that the sign of $g_{j^*}(x,t)$ changes around $(\overline{x}, \overline{t})$ at the curve of g.c. points of P^2, whose set of active index is $J_0(\overline{x}, \overline{t}) \setminus \{j^*\}$. From this, it follows that exactly in one case, either for $t > \overline{t}$ or for $t < \overline{t}$, this curve contains g.c. points of the original problem. So, locally Σ_{gc} is the union of a curve and a branch.

Type 3 — At this point only (1c) fails in the sense that $\nabla_x^2(\overline{x}, \overline{t})L(\overline{x}, \overline{\lambda}, \overline{\mu}, \overline{t})|_{T_x M(t)}$ has exactly a zero eigenvalue. $(\overline{x}, \overline{\lambda}, \overline{\mu}, \overline{t})$ is also a non-degenerated critical point of problem P^3, which is defined as the minimization of t on (x, λ, μ, t), where $(x,t) \in \Sigma_{gc}$ and (λ, μ) are the associated multipliers. This allows to describe, locally, Σ_{gc} as a curve in which either $t > \overline{t}$ or $t < \overline{t}$ for all its elements. In fact $(\overline{x}, \overline{t})$ is a quadratic turning point in Σ_{gc}.

Type 4 — Here LICQ fails, $|m| + |J_0(\bar{x}, \bar{t})| \leq n$ and again extra assumptions, based on the non degeneracy of a certain g.c. point of a related problem (P^4) are added. Again $(\overline{x}, \overline{t})$ is a quadratic turning point. The degeneracy in the set of feasible solutions leads to bad cases such as the instability of $M(t)$, which, in this case means that, locally around $\overline{t}$, either for $t > \overline{t}$ or for $t < \overline{t}$, it holds that $M(t) = \emptyset$.

Type 5 — LICQ fail and $m + |J_0(\bar{x}, \bar{t})| = n + 1$. The other conditions of the definition implies that Σ_{gc} is locally equal to the union of $|J_0(\bar{x}, \bar{t})|$ of the feasible branches of g.c. points of problems

$$P(j^*) \quad \min f(x,t) \text{ s.t. } h_i(x,t) = 0,\ i = 1, \ldots, m,\ g_j(x,t) = 0,\ j \in J_0(\overline{x}, \overline{t}), \setminus \{j^*\},$$

where $j^* \in J_0(\overline{x}, \overline{t})$.

Let Σ_{gc}^{ν}, $\nu \in \{1, \ldots, 5\}$ be the set of g.c. points of type ν. Figure 1 illustrates the local structure of Σ_{gc} in the neighborhood of these singularities.

We say that a one-parametric problem is JJT-regular if its g.c. points are of type 1, 2, 3, 4 or 5. The class $\mathcal{F}$ is the set of functions defining JJT-regular problems with m equalities and s inequalities. That is

$$\mathcal{F} = \Big\{ (f, H, G) \in C^3(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}^{1+m+s}) \mid \Sigma_{gc}(P(t)_{(f,H,G)}) \subset \bigcup_{\nu=1}^{5} \Sigma_{gc}^{\nu} \Big\}.$$

The following two theorems show that it is not too strong to assume that a general parametric problem is regular.

Theorem 2.2. (Perturbation Theorem, cf. Gómez et al. [13]) Fix $(f, H, G) \in C^3(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}^{1+m+s})$. Let $q = (b, A, c, D, e, F) \in \mathbb{R}^n \times \mathbb{R}^{n(n+1)/2} \times \mathbb{R}^m \times \mathbb{R}^{mn} \times \mathbb{R}^s \times \mathbb{R}^{sn}$ and $Q = \{q : (f(x,t) + b^T x + x^T A x, H(x,t) + c + Dx, G(x,t) + e + Fx) \notin \mathcal{F}\}$. Then, each measurable subset of Q has zero Lebesgue-measure.

Theorem 2.3. (Genericity Theorem, cf. Gómez et al. [13]) $\mathcal{F}$ is open and dense with respect to the strong topology on $C^3(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}^{1+m+s})$.

We want to point out that the perturbation theorem is proven by means of the following lemma, cf. [13].

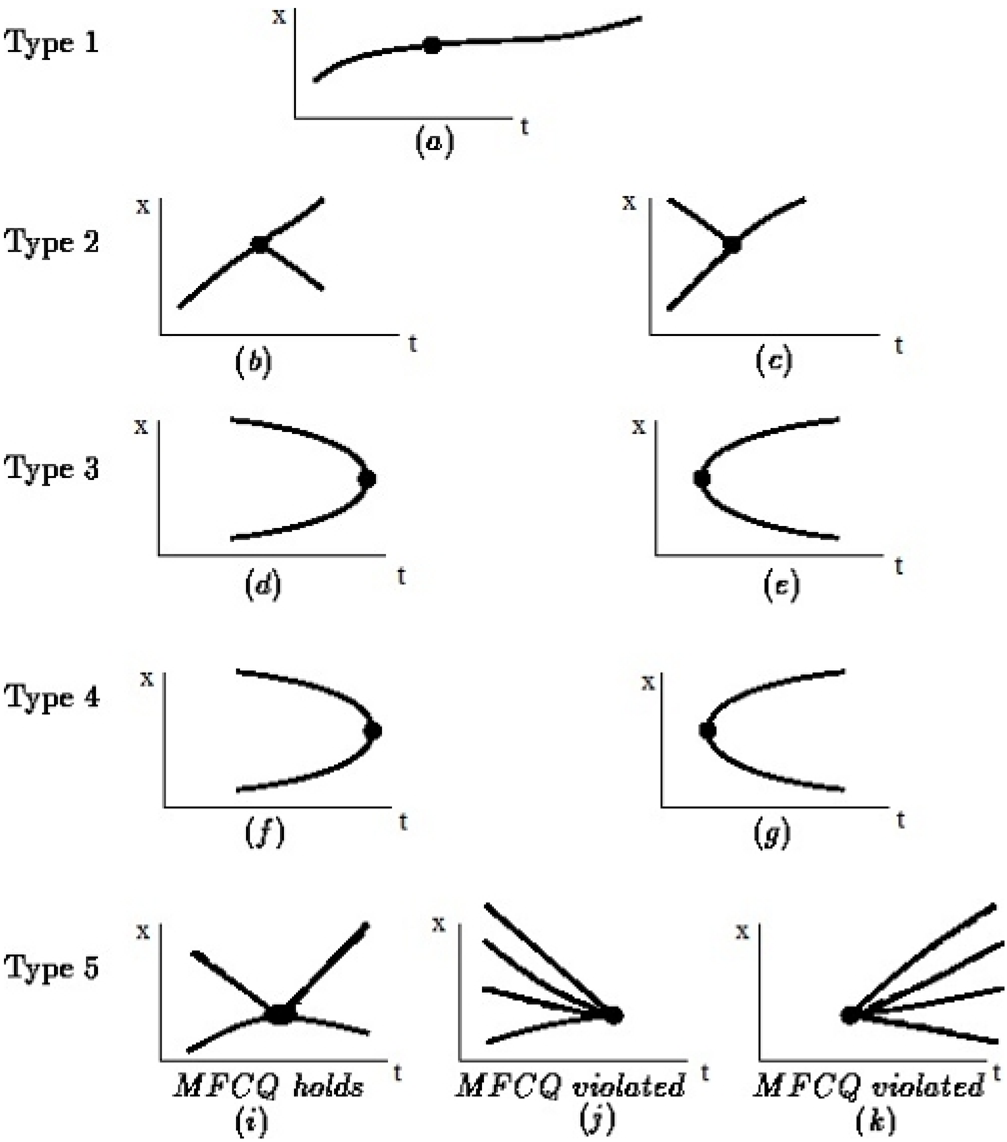

Fig. 1. Local structure of Σ_{gc}.

Lemma 2.4. (Parametric Sard's Lemma) Let us consider $\phi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^r$ be a C^k–function, $k > \max\{0, n-r\}$, $x \in \mathbb{R}^n$, $z \in \mathbb{R}^p$. If 0 is a regular value of ϕ, (i.e. $rank(\nabla\phi(x,z)) = r\ \forall(x,z) : \phi(x,z) = 0$), then, for almost all $z \in \mathbb{R}^p$, 0 is a regular value of $\hat{\phi}_z(x) = \phi(x,z)$.

On the other hand the genericity result follows after the Perturbation Theorem and continuity arguments.

Multipliers method

This algorithm, also known as Augmented Lagrangian Method, appears in Bertsekas [5]. First an optimization problem with non-negative constraints, whose objective function includes the Lagrange function of P and a quadratic penalty term is defined. This problem is solved iteratively via a penalty strategy and the multipliers are also computed, providing a saddle point of the Lagrange function (for more details see Bazaraa et al. [4] and Luenberger [21]).

This algorithm has been adapted to the solution of problems such as variational inequality problems, quadratic programs, video restoration and production planning and scheduling, see Iusem [16], Dostál et al. [10], Afonso et al. [1] and Li et al. [20]. Improvements of the method via convexifications and new updates can be found in Li et al. [19], Avelino-Vicente [3], Birgin-Martínez [6] and Andreani et al. [2].

3. PROPOSED EMBEDDING FOR MULTIPLIERS METHOD

In this section we present a new embedding for the multipliers method. That is, $\overline{f}(y,t)$, the objective function of the parametric problem defined by the embedding, includes the Lagrangian of problem P augmented with a quadratic penalty term. Moreover saddle points of $\overline{f}(y,t)$ can be computed by a path-following strategy at least for all t in a certain interval. These properties are also fulfilled by the others multipliers embeddings proposed in Dentcheva et al. [9]. However, as discussed in Bouza et al. [8] and Bouza [7], the genericity of the JJT-regularity is only shown if problem P has not inequality constraints. In this proposal, the JJT-regularity holds for generic problems P with equality and inequality constraints.

First, inequality constraints are written as equality restrictions by means of C^∞-penalty functions. That is, we consider functions $\chi(u):\mathbb{R}\to\mathbb{R}$, $\chi\in C^\infty$ such that $\chi(u)\geq 0$, for all $u\in\mathbb{R}$ and $\chi(u)=0$ if and only if $u\geq 0$. So, $g(x)\geq 0 \Leftrightarrow \chi(g(x))=0$. Given $\hat{L}(x,\lambda,\mu)=f(x)-\sum_{i=1}^m \lambda_i h_i(x)-\sum_{j=1}^s \mu_j\chi(g_j(x))$ and $p>0$, the embedding is described by the application, $\Phi: C^3(\mathbb{R}^n, R^{1+m+s})\to C^3(\mathbb{R}^{n+m+s}\times\mathbb{R},\mathbb{R}^{1+(m+s)+1})$

$$\Phi\begin{pmatrix} f\\ h_1\\ \vdots\\ h_m\\ g_1\\ \vdots\\ g_s\end{pmatrix}(x,\lambda,\mu,t)=\begin{pmatrix} t\hat{L}(x,\lambda,\mu)+(1-t)[\|x\|^2-\|\lambda\|^2-\|\mu\|^2]+\|v\|^2+\|w\|^2\\ th_1(x)+(1-t)v_1\\ \vdots\\ th_m(x)+(1-t)v_m\\ t\chi(g_1(x))+(1-t)w_1\\ \vdots\\ t\chi(g_s(x))+(1-t)w_s\\ p-\|x\|^2-\|\lambda\|^2-\|\mu\|^2-\|v\|^2-\|w\|^2\end{pmatrix}. \tag{6}$$

So, given the functions $(f,h_1,\ldots,h_m,g_1,\ldots,g_s)$, the multipliers embedding defined by $\Phi(f,h_1,\ldots,h_m,g_1,\ldots,g_s)$ is:

$$(P_M(t))\quad \min\begin{Bmatrix} t\hat{L}(x,\lambda,\mu)+(1-t)[\|x\|^2-\|\lambda\|^2-\|\mu\|^2]+\|v\|^2+\|w\|^2\\ (x,\lambda,\mu,v,w)\in M_M(t)\end{Bmatrix} \tag{7}$$

$$M_M(t) := \left\{ (x,\lambda,\mu,v,w) \;\middle|\; \begin{array}{rcll} th_i(x) + (1-t)v & = & 0, & i = 1\dots m, \\ t\chi(g_j(x)) + (1-t)w & = & 0, & j = 1\dots s, \\ p - \|(x,\lambda,\mu,v,w)\|^2 & \geq & 0. & \end{array} \right\}. \tag{8}$$

Two points shall be remarked at the definition of $M_M(t)$. First note that constraint $p - \|(x,\lambda,\mu,v,w)\|^2 \geq 0$ has been added, here p is a positive constant which does not depend on problem P. This inequality provides a global bound for the set of feasible solutions which is very useful for the existence of solutions of $P_M(t)$. Parameter p can be taken around the order of the maximum representable number at the computer because, although the set $M_M(t)$ is shrunk, larger numbers will lead to huge numerical errors and this is not interesting from a practical viewpoint. According to the experience of the expert, smaller values can be also considered. With respect to $\chi(u)$, different C^∞-penalty functions can be used. Some examples are:

- $\chi(x) = 1_{(-\infty,0)}(x)e^{\frac{1}{x}}$, see Hirsch[15].
- $\chi(x) = -1_{(-\infty,0)}(x)xe^{\frac{1}{x}}$.

From now on, $y = (x,\lambda,\mu,v,w)$ and $J_0(y,t)$ is the set of active indexes of (y,t). If (y,t) is a g.c. point, (α,β,δ) are the multipliers associated to the constraints $th_i(x) + (1-t)v = 0$, $i = 1,\dots,m$, $t\chi(g_j(x)) + (1-t)w = 0$, $j = 1,\dots,s$ and $p - \|y\|^2 \geq 0$, respectively. As usual, for $(y,t) \in \Sigma_{gc}(\Phi(P)(t))$, if $J_0(y,t) = \varnothing$, i.e. $\|y\|^2 < p$, then δ is 0. $L_M(y,\alpha,\beta,\delta,t)$ denotes the Lagrangian of the parametric problem $\Phi(f,h_1,\dots,h_m,g_1,\dots,g_s)(t)$.

Now we will present some simple properties of the embedding, we denote by :

Proposition 3.1. Let us assume that M, (cf. (2)), the set of feasible solution of P, is non-empty. Then the parametric problem $P_M(t)$ (cf. (7)) satisfies that:

(a) If M is compact, for p large enough, it holds that $\Pi_x(M_M(1)) = M$.

(b) If p is large enough, for all $t \in [0,1]$, $P_M(t)$ has a generalized solution.

(c) $y = 0_{n+2m+2s\times 1}$ is a $\min_{x,v,w}$-$\max_{\lambda,\mu}$ of P(0), and is a g.c. point of type 1.

Proof. (a) $M_M(1)$ is described as $h_i(x) = 0, i = 1,\dots,m$, $\chi(g_j(x)) = 0$, $j = 1,\dots,s$ and $p - \|y\|^2 \geq 0$. By the properties of χ, this means that $h_i(x) = 0$, $i = 1,\dots,m$, $g_j(x)) \geq 0$, $j = 1,\dots,s$. So, $\Pi_x(M_M(1)) \subset M$.
On the other hand, as M is compact, for p large enough it holds that $M \subset \{x : \|x\|^2 \leq p\}$. Then for all $x \in M$ $(x, 0_{2m+2s\times 1}) \in M_M(1)$. So, $x \in \Pi_x(M(1))$. Hence $M \subset \Pi_x(M_M(1))$ and $M = \Pi_x(M_M(1))$.

(b) Let x be a feasible point and p be such that $\|x\|^2 \leq p$. Then $(x,0,0,0,0) \in M_M(t)$ for all t. Hence $M_M(t) \neq \varnothing$ is a closed bounded set. So, $M_M(t)$ is compact and $P_M(t)$ has a minimum for all t. The result follows directly after noting that minimizers are generalized critical points.

(c) $P_M(0)$ is $\min \|(x,v,w)\|^2 - \|(\lambda,\mu)\|^2$ s.t. $v=w=0$, $p-\|y\|^2 \geq 0$. After some easy calculations it can be seen that the LICQ holds for all feasible point. At $(0,0,0,0,0) \in M_M(0)$, the set of active index set is empty. On the other hand,

$$T_xM = span \left\{ \begin{matrix} I_n & 0_{n\times(m+s)} \\ 0_{(m+s)\times n} & I_{m+s} \\ 0_{(m+s)\times n} & 0_{(m+s)\times(m+s)} \end{matrix} \right\}$$

and

$$D_x^2 L_M = \begin{pmatrix} I_n & 0_{n\times(m+s)} & 0_{n\times(m+s)} \\ 0_{(m+s)\times n} & -I_{m+s} & 0_{(m+s)\times n} \\ 0_{(m+s)\times n} & 0_{(m+s)\times(m+s)} & I_{m+s} \end{pmatrix}.$$

So,

$$(D_x^2 L_M)|_{T_xM} = \begin{pmatrix} I_n & 0_{n\times(m+s)} \\ 0_{(m+s)\times n} & -I_{m+s} \end{pmatrix}$$

is a non singular matrix and $y=0$ is a point of Type 1.

Evidently $y=0$ is a saddle point of $P(0)$ because it is a saddle point of the equivalent problem $\min \|x\|^2 - \|(\lambda,\mu)\|^2$ s.t. $p - \|(x,\lambda,\mu,v,w)\|^2 \geq 0$. □

Remark 3.2. (v,w) allows us to obtain the quadratic penalty term.
By (c), locally around $y=0$, the set of g.c. points is described as a curve of saddle points.

Now we will present some properties of the generalized critical points.

Proposition 3.3.

(a) Assume that problem $P(t)$ defined in (7) is JJT-regular. Then only points of type 1,2, 3 or 4 may appear.

(b) For this kind of embedding the MFCQ is satisfied if and only if the LICQ holds.

(c) If LICQ fails at $y \in M_M(t)$, $t<1$, then $J_0(y,t)=\{1\}$, $(\lambda,\mu)=0$ and

$$\begin{aligned} -\tfrac{t^2}{(1-t)^2}\left[\sum_{i=1}^m h_i(x)\nabla_x h_i(x) + \sum_{j=1}^s \chi(g_j(x))\chi'(g_j(x))\nabla_x g_j(x)\right] - x &= 0 \\ \|x\|^2 + \tfrac{t^2}{(1-t)^2}\left[\sum_{i=1}^m h_i^2(x) + \sum_{j=1}^s \chi^2(g(x)_j^2(x))\right] &= p. \end{aligned}$$

P r o o f. (a) The total of constraints $m+s+1$ is always smaller than or equal to the total of variables, $n+2m+2s$. So, no points of type 5 may appear.

(b) It is well known that the fulfillment of LICQ implies that MFCQ holds.

If MFCQ holds, then by MF1, the gradients of the equality constraints are linearly independent. As there is only one inequality restriction, $p-\|y\|^2 \geq 0$, LICQ fails if and only if its gradient is a linear combination of the gradients of the equality constraints. Taking ξ as in MF2, $y\xi > 0$, while by the linear dependence $y\xi = 0$, a contradiction.

(c) Writing the gradients of the $m+s$ equalities it is clear that they are linearly independent. So, if the LICQ fails at the point (y,t), then $J_0(y,t)=\{1\}$. Moreover, the (active) inequality is a linear combination of the gradients of the equalities. Taking the linear combination,

$$\begin{array}{rcll} t\sum_{i=1}^{m}\alpha_i\nabla_x h_i+\sum_{j=1}^{s}\beta_i\nabla_x[\chi(g_j)]-2x & = & 0, & \\ -2\lambda_i & = & 0, & i=1,\dots,m, \\ -2\mu_j & = & 0, & j=1,\dots,s, \\ (1-t)\alpha-2v & = & 0, & \\ (1-t)\beta-2w & = & 0, & \end{array}$$

it follows that $(\lambda,\mu)=0$ and $(\alpha,\beta)=\frac{2}{1-t}(v,w)$. Using the constraints

$$\begin{array}{rcll} th_i(x)+(1-t)v & = & 0, & i=1,\dots,m, \\ t\chi(g_j(x))+(1-t)w & = & 0, & j=1,\dots,s, \end{array}$$

the values of (v,w) are $-\frac{t}{1-t}(h_1(x),\dots,h_m(x),\chi(g_1(x)),\dots,\chi(g_s(x)))$. So, $(\alpha,\beta)=-\frac{2t}{(1-t)^2}(h_1(x),\dots,h_m(x),\chi(g_1(x)),\dots,\ \chi(g_s(x))$ and the result now follows after substituting these values at the equalities given by the linear combination and $\|y\|^2=p$. □

Now we will illustrate the behavior of the embedding approach when it is applied to solve two non-linear programs. The parametric problem (7) is solved by PAFO package. PAFO is a path-following and jumps routine for solving JJT-regular parametric problems, see Gollmer et al. [11] and Guddat et al. [14]. As already remarked, locally around $(y,t)\in\cup_{i=1}^{5}\Sigma_{gc}$, the set of g.c. points can be described as the solutions of (finitely many) well known nonlinear systems. Given a starting solution and a closed interval T, under JJT regularity, PAFO solves those systems by a predictor-corrector scheme and hence computes the g.c. points around (y,t). The step size of parameter t is dynamically managed in order to ensure the fulfillment of certain errors bounds. Although, as reported in ch.5, [14], this package also includes a subroutine for jumping to another connected component, in this paper we will only use the path-following strategy.

For this embedding, we created a sub-routine whose inputs are the objective function and the equality constraints of P. The resulting parametric functions define the parametric problem $\Phi(P)$, which will be solved by PAFO. The solution computed by PAFO will be compared with the result obtained by MatLab 7.6 and by the software KNITRO available at GAMS 22.2, applied to problem P with the extra constraint $\|x\|^2\leq p$. In all cases, $p=50$.

Let us present the first example

$$\begin{array}{lrcl} \min & -x_1 & & \\ s.t & 1+x_1(1-x_1)^2-x_2 & \geq & 0. \end{array}$$

We have to solve the parametric problem:

$$\begin{array}{c} \min\left[\begin{array}{c} t(-x_1-\mu(1+x_1(1-x_1)^2-x_2))+ \\ +(1-t)(\|x-x_0\|^2-(\mu-\mu_0)^2)+(w-w_0)^2 \end{array}\right] \\ s.t.\ \ t\chi(1+x_1(1-x_1)^2-x_2)+(1-t)(w-w_0)=0 \\ \|x-x_0\|^2+(\mu-\mu_0)^2+(w-w_0)^2\leq p. \end{array} \tag{9}$$

We started with point $(x_0, \mu_0, w_0) = (1, 0, 0, 0)$. Figure 2 shows that there are singularities of type 2 and 3. In fact $t \leq .96$ for all computed g.c. point, so $t = 1$ is never reached. Although we do not obtain a g.c. point of $P(1)$, for $t = .95$, (7,0) is at least a feasible point close to (7.071, 0), the solution computed by KNITRO and by MatLab 7.6 using in both cases (1,0) as starting point.

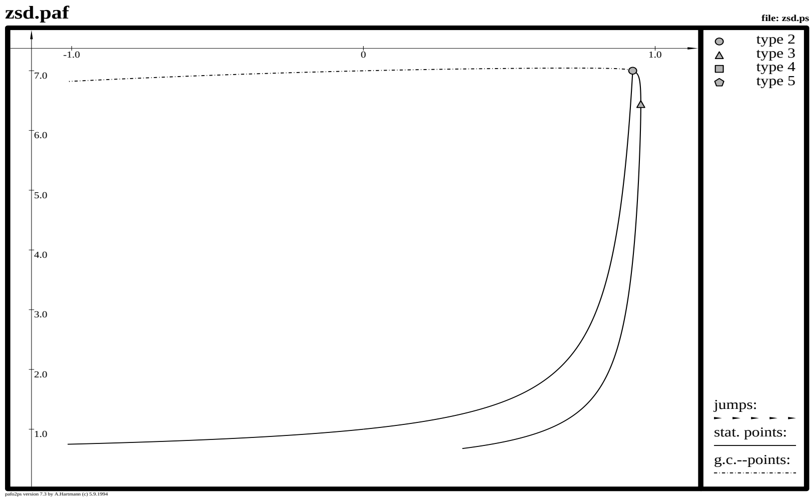

Fig. 2. x_1 vs. t.

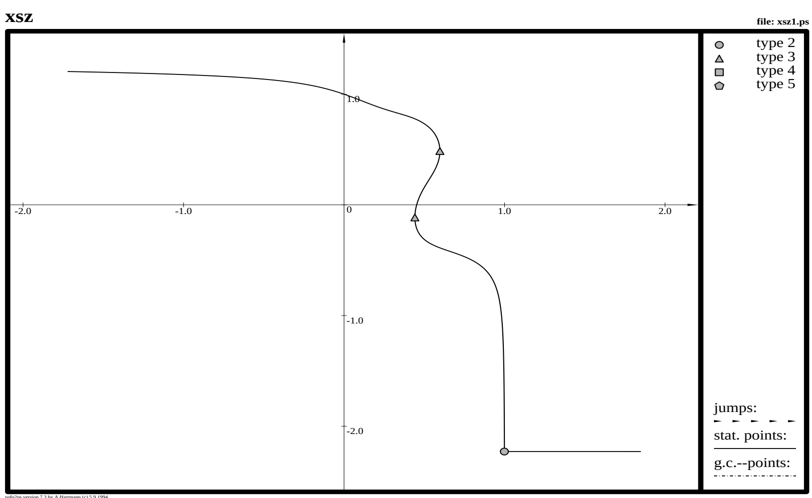

Fig. 3. x_1 vs. t.

The second example is:

$$\begin{array}{rrcl} & \min x_1 x_2 & & \\ \text{s.t.} & 1 + x_1(1 - x_1 + x_1^2) - x_2 & \geq & 0. \end{array} \tag{10}$$

Now, as can be seen in Figure 3, we are able to obtain a solution of the original problem, although some singularities appeared. Using the same initial point in the three cases, (0,0) is the solution given by the professional softwares while our approach calculated the point (4.99 –4.98). As $f(0,0) = 0 > f(4.99, -4.98) = -24.85$, our approach calculated the best solution.

As we have seen, in these two illustrative examples, the embedding computed good solutions compared to the results obtained via KNITRO and MatLab, even if there are singularities where the min-max property is lost. In both cases the problems are JJT-regular. In the next section we will analyze, from a theoretical viewpoint, what can we expect in the generic case.

4. MAIN RESULTS

We begin with the presentation of these two properties

(A1) $m > 1$.

(A2) Let δ be the multiplier associated to the compactification constraint at a critical point, then $1 - t + \delta \neq 0$.

Now we prove two lemmas

Lemma 4.1. Let us define $q = (A, b, c, d) \in \mathbb{R}^{\frac{n(n+1)}{2}} \times \mathbb{R}^n \times \mathbb{R}^{n\times(m+s)} \times \mathbb{R}^{m+s}$ and assume that $(A1)$ holds. Then for almost every (b, d), there is not (y, t) critical point of $\Phi(f(x) + x^T Ax + b^T x, (h_1(x), \ldots, h_m(x), g_1(x), \ldots, g_s(x)) + (c^T x + d))$, $t \in (0, 1)$, such that $\|y\|^2 = p$ and $(\lambda, \mu) = 0$.

Proof. The critical points of $\Phi(f, h_1(x), \ldots, h_m(x), g_1(x), \ldots, g_s(x))$ can be described as zeros of

$$\mathcal{H}(y, \alpha, \beta, \delta, t) = 0, \tag{11}$$

where $\mathcal{H}(y, \alpha, \beta, \delta, t)$ is the function whose components are the gradient of the Lagrangian function of the parametric problem, with respect to y, the equality constraints and $\|y\|^2 - p$ if it is active. Let us suppose that there is a critical point with $J_0(y, t) \neq \varnothing$ and $(\lambda, \mu) = 0$, then $(v, w, \alpha, \beta) = 0$. It means that the following system is satisfied:

$$\begin{array}{rcl} t[\nabla_x f(x) + 2Ax + b] + 2(1 - t - \delta)x & = & 0, \\ h_i(x) + c_i^T x + d_i & = & 0, \\ \|x\|^2 - p & = & 0. \end{array} \tag{12}$$

Now taking the Jacobian of the system with respect to $(y,\ t,\ b,\ d)$, we obtain that it has full row rank. Applying Lemma 2.4, it holds that for almost every (b, d) at (y, t) solution of (12), the Jacobian with respect to (y, t) has full(row) rank. But this is impossible because this matrix has more rows than columns. So, there is no solution of system (12) for almost every (b, d). □

Remark 4.2. Together with Proposition 3.3(c), this means that for almost every (b,d), LICQ fails at a feasible point if and only if $(\lambda,\mu)=0$ and $\|y\|^2=p$.

The second lemma provides the desired structure of the Hessian matrix of the Lagrange function

Lemma 4.3. Suppose that $(A1)$ and $(A2)$ hold. Let $t\in(0,1)$ and denote as $\mathcal{M}=\nabla^2_{y,\alpha,\beta,\delta}L_M(y,\alpha,\beta,\delta,t)$ the Hessian of the Lagrangian function at the g.c. point (y,t). If LICQ holds at (y,t), maybe after permutations of rows and columns, $\mathcal{M}=\begin{pmatrix}B & C^T\\ C & D\end{pmatrix}$, where $B=C^TD^{-1}C$, B is a symmetric sub-matrix of $\nabla^2_x L_M(y,\alpha,\beta,\delta,t)$ and D is a symmetric matrix, whose rank is equal to the rank of $\mathcal{M}$.

P r o o f. For simplicity we define τ and $\bar{t}$ as $1-t$ and $1-t+\delta$, respectively. If $|J_0|\neq 0$, then

$$\mathcal{M}=\begin{pmatrix}\otimes+2A & \otimes & \otimes & 0 & 0 & \otimes & \otimes & -2x\\ \otimes & 2\bar{t}I_m & 0 & 0 & 0 & 0 & 0 & -2\lambda\\ \otimes & 0 & 2\bar{t}I_s & 0 & 0 & 0 & 0 & -2\mu\\ 0 & 0 & 0 & 2(1-\delta)I_m & 0 & \tau I_m & 0 & -2v\\ 0 & 0 & 0 & 0 & 2(1-\delta)I_s & 0 & \tau I_s & -2w\\ \otimes & 0 & 0 & \tau I_m & 0 & 0 & 0 & 0\\ \otimes & 0 & 0 & 0 & \tau I_s & 0 & 0 & 0\\ -2x & -2\lambda & -2\mu & -2v & -2w & 0 & 0 & 0\end{pmatrix}. \tag{13}$$

As $t\in(0,1)$ and (A2) holds, $\tau\neq 0$ and $\bar{t}\neq 0$. Moreover by Lemma 4.1 for almost every (b,d) at $(y,t)=(x,\lambda,\mu,v,w,t)$ critical point $(\lambda,\mu)\neq 0$. Using these fact, it can be seen that matrix

$$\nabla^2_{\lambda,\mu,v,w,\alpha,\beta,\delta}L_M(y,\alpha,\beta,\delta,t)$$
$$=\begin{pmatrix}-2\bar{t}I_m & 0 & 0 & 0 & 0 & 0 & -2\lambda\\ 0 & -2\bar{t}I_s & 0 & 0 & 0 & 0 & -2\mu\\ 0 & 0 & 2(1-\delta)I_m & 0 & \tau I_m & 0 & -2v\\ 0 & 0 & 0 & 2(1-\delta)I_s & 0 & \tau I_s & -2w\\ 0 & 0 & \tau I_m & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \tau I_s & 0 & 0 & 0\\ -2\lambda & -2\mu & -2v & -2w & 0 & 0 & 0\end{pmatrix}$$

is non singular.

If $J_0=\varnothing$, $\mathcal{M}$ is obtained by deleting the last row and the last column, and making $\delta=0$ in (13). As in this case $\tau=\bar{t}\neq 0$, $\nabla^2_{\lambda,\mu,v,w,\alpha,\beta}L_M(y,\alpha,\beta,\delta,t)$ is non singular in this case too.

Define D_1 as $\nabla^2_{\lambda,\mu,v,w,\alpha,\beta}L_M(y,\alpha,\beta,t)$, if $J_0=\varnothing$ and $\nabla^2_{\lambda,\mu,v,w,\alpha,\beta,\delta}L_M(y,\alpha,\beta,\delta,t)$, otherwise. Now let D be a square symmetric sub-matrix of $\mathcal{M}$ such that $rank(D)=rank(\mathcal{M})$ and D_1 is contained in D. After some rows and columns permutations, it follows that $\mathcal{M} = \begin{pmatrix} B & C^T \\ C & D \end{pmatrix}$. Using the equality of the ranks of D and $\mathcal{M}$, for some matrix Λ, it holds that

$$\begin{pmatrix} B \\ C \end{pmatrix} = \begin{pmatrix} C^T \\ D \end{pmatrix} \Lambda.$$

As D_1 is a sub-matrix of D, B is a square symmetric sub-matrix of $\nabla_x^2 L(y, \alpha, \beta, \delta, t)$. Moreover, $\Lambda = D^{-1}C$. Taking into account that $B = C^T\Lambda$, then $B = C^T D^{-1} C$ and the desired result evidently holds. $\square$

Remark 4.4. The property (A1) is not very strong. (A2) fails only at critical points of the original problem in which $\nabla_x f$ is a combination of x and $\nabla_x h_1(x), \ldots, \nabla_x h_m(x)$. This kind of singularity leads us to a g.c. of P. So, when it is detected, the algorithm succeeded in computing a solution of our problem.

Let us now present the Perturbation Theorem. We will perturb the objective function by a quadratic function $\frac{x^T Ax}{2} + b^T x$, and the constraints linearly by $[c_H^T x + d_H, c_G^T x + d_G]$, where $A \in \mathbb{R}^{\frac{n\times(n+1)}{2}}, b \in \mathbb{R}^n, c \in \mathbb{R}^{n\times(m+s)}, d \in \mathbb{R}^{m+s}$. Here $q = (A, b, c, d)$ is the perturbation parameter.

Theorem 4.5. Suppose that $(A1)$ holds and $\chi(x) = 1_{(-\infty,0)}(x)e^{\frac{1}{x}}$. For almost every q, the g.c. points of $\Phi(f(x) + x^T Ax + b^T x, [h_1, \ldots, h_m, g_1, \ldots, g_s](x) + [c_H^T x + d_H, c_G^T x + d_G])$ with $t < 1$ either satisfy $(A2)$ or are points of type 1, 2, 3 or 4.

P r o o f. The proof is divided in two parts. First we will show which types of g.c. points may appear if LICQ holds. $\square$

Lemma 4.6. Supposed $(A1)$ holds and let us consider the parametric problem defined as $\Phi(f(x) + x^T Ax + b^T x, [h_1, \ldots, h_m, g_1, \ldots, g_s](x) + [c_H^T x + d_H, c_G^T x + d_G])$. Then, for almost every q, the g.c. points with $t < 1$ in which LICQ hold, either satisfy $(A2)$ or are points of type 1, 2 or 3.

P r o o f. First note that for the JJT characterization of critical points, we need to describe the critical point condition, the number of zero multipliers and the rank of $\mathcal{M} = \nabla^2_{y,\alpha,\beta,\delta} L_M(y, \alpha, \beta, \delta, t)$, as a system of equations. For the rank condition, by Lemma 4.3, B is a sub-matrix of $\nabla_x^2 L_M(y, \alpha, \beta, \delta, t)$. So, the possible indexes of the rows of $\mathcal{M}$ forming matrix B are subsets of $\{1, \ldots, n\}$. Fix $\hat{J} \subset \{1\}$ and $I \subset \{1, \ldots, n\}$.

By using $\mathcal{H}(y, \alpha, \beta, \delta, t)$ (cf. (11)), the critical points such that the active index set is $\hat{J}$ and B is the sub-matrix of $\mathcal{M}$ whose rows and columns are in I, solves system:

$$\begin{aligned} \mathcal{H}(y, \alpha, \beta, \delta, t) &= 0, && (14)\\ \mathcal{M}(y, \alpha, \beta, \delta, t) &= g, && (15)\\ g_B &= (g_C)^T (g_D)^{-1} g_C \ (B = C^T D^{-1} C), && (16)\\ \delta &= 0, \text{ if } J_0(y, t) = \{1\},\ \delta = 0. && (17) \end{aligned}$$

Using the C^3 differentiability of the $(f, h_1, \dots, h_m, g_1, \dots, g_s)$ the system is described by a C^1 function. Assume that $(A2)$ holds. We want to prove that the Jacobean with respect to variables, multipliers and parameters of the system has full row rank. First we focus in the columns corresponding to $\partial_{g,A,b,\alpha,\beta}$:

$$\begin{array}{c|cccc}
 & \partial_g & \partial_A & \partial_b & \partial_{\alpha,\beta} \\
 & 0 & \bigotimes & tI_n & \bigotimes \\
 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & \tau I_{m+s} \\
\mathcal{H}(y,\alpha,\beta,\delta,t)=0 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & \\
\hline
\mathcal{M}(y,\alpha,\beta,\delta,t)=g & I_\star & \begin{pmatrix} tI_{n(n+1)/2} \\ 0 \end{pmatrix} & 0 & 0 \\
\hline
g_B=(g_C)^T(g_D)^{-1}g_C & I_{\star\star}|\bigotimes & 0 & 0 & 0 \\
\hline
\delta=0 & 0 & 0 & 0 & 0
\end{array} \tag{18}$$

where $\tau = 1-t$, $\star = \frac{(n+2m+2s)(n+2m+2s+1)}{2}$, $\star\star = n+2m+2s-rank(M)$. As $t \in (0,1)$, the derivatives of the first n equations of $\mathcal{H}(y,\alpha,\beta,\delta,t)=0$, are linearly independent(l.i.) with respect to the others rows of (18), see the columns corresponding to ∂_b. Moreover, using columns $\partial_{\alpha,\beta}$, the linear independence of the rows $(0 \quad 0 \quad 0 \quad \tau I_{m+s})$ is also direct. By Lemma 4.3, g_B is a sub-matrix of $D_x^2 L_M$, which is described at the first $n(n+1)/2$ equations. So, $\star\star \leq n(n+1)/2$ and the rows corresponding to the derivatives of $\mathcal{M}(y,\alpha,\beta,\delta,t)=g$ and $g_B=(g_C)^T(g_D)^{-1}g_C$ are also l.i.

Now we consider the different combinations of the set of active indexes and zero multipliers.

Case $J_0 = \{1\}, \delta = 0$: we only need to prove that the following matrix has full row rank.

$$\begin{array}{ccccc}
\partial_x & \partial_{\lambda,\mu} & \partial_{v,w} & \partial_\delta & \partial_t \\
\bigotimes & -2\tau I_{m+s} & 0 & -2\begin{pmatrix}\lambda \\ \mu\end{pmatrix} & 2\frac{1}{t}\begin{pmatrix}\lambda \\ \mu\end{pmatrix} \\
\bigotimes & 0 & \tau I_{m+s} & 0 & -\frac{1}{t}\begin{pmatrix}v \\ w\end{pmatrix} \\
-2x & -2(\lambda,\mu) & -2(v,w) & 0 & 0 \\
\hline
0 & 0 & 0 & 1 & 0.
\end{array} \tag{19}$$

We take the linear combination of the rows. Recalling that $\tau \neq 0$, because $t \in (0,1)$, after some calculations it follows that $(\lambda,\mu) = 0$ and hence $(v,w) = 0$. So, the only possibility is that either the combination is identically zero or $x = 0$, and hence $0 = \|y\|^2$, contradicting the fact that $J_0 = \{1\}$. So, the Jacobean matrix has full row rank.

Case $J_0 = \{1\}$ and $\delta \neq 0$: Equation $\delta = 0$ is not taken into account. Now we need to show that the following matrix has full row rank

$$\begin{array}{ccccc}
\partial_x & \partial_{\lambda,\mu} & \partial_{v,w} & \partial_\delta & \partial_t \\
\bigotimes & -2(\tau+\delta)I_{m+s} & 0 & -2\begin{pmatrix}\lambda\\ \mu\end{pmatrix} & 2\frac{1-\delta}{t}\begin{pmatrix}\lambda\\ \mu\end{pmatrix} \\
\bigotimes & 0 & \tau I_{m+s} & 0 & -\frac{1}{t}\begin{pmatrix}v\\ w\end{pmatrix} \\
-2x & -2(\lambda,\mu) & -2(v,w) & 0 & 0.
\end{array} \tag{20}$$

Note that as $(A2)$ does not hold $\tau + \delta \neq 0$. Taking the linear combination of the rows and following the same analysis done in the last case the independence is obtained.

Case $J_0 = \varnothing$: The corresponding matrix is

$$\begin{array}{cccccc}
\partial_x & \partial_{\lambda,\mu} & \partial_{v,w} & \partial_{\alpha,\beta} & \partial_\delta & \partial_t \\
\bigotimes & -2\tau I_{m+s} & 0 & 0 & -2\begin{pmatrix}\lambda\\ \mu\end{pmatrix} & 2\frac{1}{t}\begin{pmatrix}\lambda\\ \mu\end{pmatrix} \\
\bigotimes & 0 & \tau I_{m+s} & 0 & 0 & -\frac{1}{t}\begin{pmatrix}v\\ w\end{pmatrix}
\end{array}$$

and the independence of the rows is evident.

So, in all cases, the Jacobean of system (18), with respect to variables and parameters, has full row rank. Taking into account that the system is described by a C^1 function and the number of equations and of variables, the parametric Sard's Lemma can be applied. So, the Jacobean of system (18), with respect to variables, has full row rank for almost every q. Hence, in this case, the number of equations is smaller than or equal to the number of variables. As a result for almost every q we obtain these three possibilities

- $\star\star = 0$ and $J_0 = J_+(\delta)$, that is (y,t) is a point of type 1.
- $J_0 = \{1\}$, $\delta = 0$ and $D^2_{y,\alpha,\beta,\delta}L_M$ is regular: the last condition leads to the non degeneracy of (y,t) as g.c. point of problem P^1. On the other hand, as $J_0 = \{1\}$ and $\delta = 0$, (y,t) also solves the system $\mathcal{H}(y,\alpha,\beta,t) = 0$, $\|y\|^2 = p$. Now we consider the previous system with the addition of the equations describing the rank condition of matrix $D^2_{y,\alpha,\beta}L_M(y,\alpha,\beta,0,t)$. A similar analysis, using Sard Lemma, implies that for almost every q, the Jacobean of the system, with respect to (y,α,β,t) has full row rank. As the number of equations is smaller than or equal to the number of involved variables, we obtain that $D^2_{y,\alpha,\beta,}L_M$ is non singular. Moreover, as the Jacobean matrix of the system is non singular, it holds that $ya_y \neq 0$ if $\nabla_{y,\alpha,\beta,t}\mathcal{H}(y,\alpha,\beta,t)\begin{pmatrix}a_y\\ a_\alpha\\ a_\beta\\ 1\end{pmatrix} = 0$. So, (y,t) is a point of type 2, see [14] for a more detailed explanation.
- $D_{y,\alpha,\beta,\delta}L_M$ has exactly a zero eigenvalue and $J_0 = J_+(\delta)$: Taking the eigenvector corresponding to 0 at $D_{y,\alpha,\beta,\delta}L_M$, after some algebraic manipulations and recalling that the Jacobean matrix of system (18), with respect to $(y,\alpha,\beta,\delta,t)$ is non

singular, we obtain the non-degeneracy of the point at the problem corresponding to the definition of points of type 3. So, $(y,t) \in \Sigma^3_{gc}$.

So, we can conclude that for almost every q, the critical points of the perturbed problem fulfilling $(A2)$, are points of type 1, 2 or 3. $\square$

Let us consider the points in which the LICQ is not satisfied. As already shown in Lemma 3.3(c), linear dependency implies that the gradient of $p - \|y\|^2$ is a linear combination of the gradients of the other constraints, which are linearly independent. As $r \neq 0$, without loss of generality we assume that $r = 1$. That is, the points (y,t) in which LICQ fails are such that for some (α,β), $\delta_0 = 0$, $(y,\alpha,\beta,\delta_0,t)$ solves the following system:

$$\begin{array}{rcll} D_y\mathcal{L}(y,\alpha,\beta,\delta_0,t) & = & 0, & \\ th_i(x) + (1-t)v & = & 0, & i = 1,\dots,m, \\ t\chi(g_j(x)) + (1-t)w & = & 0, & j = 1,\dots,s, \\ \|y\|^2 & = & p, & \end{array} \tag{21}$$

where $\mathcal{L}(y,\alpha,\beta,\delta_0,t) = \delta_0[t\hat{L}(x,\lambda,\mu) + (1-t)\left(\|x\|^2 - \|\lambda\|^2 - \|\mu\|^2\right) + \|v\|^2 + \|w\|^2] + \sum_{i=1}^m \alpha_i h_i + \sum_{j=1}^s \beta_j \chi(g(x)) - \|y\| + p$. We will prove that, for almost every q, (y,t) is a g.c. point of type 4.

Lemma 4.7. If $\chi(x) = 1_{(-\infty,0)}(x)e^{\frac{1}{x}}$, then for almost every q, if (y,t) is a g.c. point of the perturbed problem $\Phi(f(x) + x^T Ax + b^T x, (h_1,\dots,h_m,g_1,\dots,g_s)(x) + (c^T x + d))$ such that LICQ fails, then $(y,t) \in \Sigma^4_{gc}$.

Proof. By Lemma 3.3(c), $(\lambda,\mu) = 0$. Let us define $(\alpha,\beta,1)$ such that:

$$\begin{array}{rcl} t\sum_{i=1}^m \alpha_i[\nabla_x h_i(x) + c_h] + t\sum_{j=1}^s \nabla_x[\chi(g_j(x) + c_g^T x + d_g)]\beta - 2x & = & 0, \\ (1-t)\alpha - 2v & = & 0, \\ (1-t)\beta - 2w & = & 0. \end{array}$$

We need to prove that $(y,\alpha,\beta,0,t)$ is a point of type 1 of (P^4), where

$$(P^4) \quad \min_{y,t,\alpha,\beta} t \text{ s.t. } (21).$$

i.e. that (1a) and (1c) holds, because as there is not inequality constraints, (1b) is satisfied. First we will show that the LICQ holds for almost every (c,d).

Let us consider the system describing the linear dependency at the original problem and two cases: $v_i \neq 0$ and $v = 0$. In the first case, for $i:\ v_i \neq 0$, the Jacobian matrix is:

$$\begin{array}{ccccccccc} \partial_x & \partial_\lambda & \partial_\mu & \partial_v & \partial_w & \partial_\alpha & \partial_\beta & \partial_{c_h^i} & \partial_{d_h^i} \\ \otimes & 0 & 0 & 0 & 0 & \otimes & \otimes & \alpha I_n & 0 \\ 0 & I_m & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_s & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2I_m & 0 & tI_m & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2I_s & 0 & tI_s & 0 & 0 \\ \otimes & 0 & 0 & (1-t)I_m & 0 & 0 & 0 & \otimes & tI_m \\ \otimes & 0 & 0 & 0 & (1-t)I_s & 0 & 0 & 0 & 0 \\ 2x & 0 & 0 & 2v & 2w & 0 & 0 & 0 & 0 \end{array}$$

and has full row rank.

In the second case, the linear dependency implies that $w \neq 0$. We assume that $w_1 \neq 0$. So the Jacobean is

$$\begin{array}{ccccccccc}
\partial_x & \partial_\lambda & \partial_\mu & \partial_v & \partial_w & \partial_\alpha & \partial_\beta & \partial_{d_{g_1}} & \partial_{d_h^i} \\
\bigotimes & 0 & 0 & 0 & 0 & \bigotimes & \bigotimes & \beta_1 diag(t\chi(g_j(x)+c_{m+j}^T x+d_g)) & 0 \\
0 & I_m & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & I_s & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2I_m & 0 & tI_m & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2I_s & 0 & tI_s & 0 & 0 \\
\bigotimes & 0 & 0 & (1-t)I_m & 0 & 0 & 0 & 0 & tI_m \\
\bigotimes & 0 & 0 & 0 & (1-t)I_s & 0 & 0 & 0 & 0 \\
2x & 0 & 0 & 2v & 2w & 0 & 0 & 0 & 0.
\end{array}$$

The previous matrix has full row rank because $\beta_1 \neq 0$, recall that $w + \beta(1-t) = 0$, and $diag(t\chi(g_j(x) + c_{m+j}^T x + d_g)) = tdiag(e^{g_j(x)+c_{m+j}^T x + d_{g_j}})$ is a diagonal regular matrix. So, by Lemma 2.4, the Jacobean matrix with respect to (y, α, β, t) has full row rank. As the rows coincide with the gradients of the constraints of (P^4), LICQ is satisfied. Moreover, by Lemma 6.1, cf. [12], it can be seen that $(W^T \nabla_y^2 \mathcal{L} W)$ is non singular, where the columns of W form a basis of the subspace orthogonal to $\nabla_y[th_1, \ldots, th_m, t\chi(g_1(x)) \ldots, t\chi(g_s(x)) + (1-t)(v, w)]$.

Define $Q_1 = \{(c, d) :$ (1a) fails at $(P^4)\}$. Note that the number of equations of the system is equal to the number of variables $(y, \alpha, \beta, \delta_0, t)$. So, for almost every (c, d), the cardinality of the set $Y(c, d) = \{(y, t) : \exists(\alpha, \beta) \in \mathbb{R}^{m+s},\ (y, \alpha, \beta, 0, t)$ satisfies (21), $\}$ is at most a numerable. This means that, it is enough to prove that

$$Q_2(y, t, c, d) = \{(A, b) : \text{ (1c) fails at } (y, t) \in Y(c, d)\}$$

has Lebesgue measure equal to 0. Indeed, by the cardinality of $Y(c, d)$, the Lebesgue measure of $\cup_{(y,t)\in Y(c,d)} Q_2(y, t, c, d)$ is 0. On the other hand, (1a) and (1c) are fulfilled if q is an element of the complement of the union of $Q_1 \times \mathbb{R}^{\frac{n(n+1)}{2}+n}$ and $\{q : (c, d) \in Q_1^c, (A, b) \in \cup_{(y,t)\in Y(c,d)} Q_2(y, t, c, d)$. By the Theorem of Fubbini, these sets, and hence their union, also has 0-Lebesgue measure. So, we fix, without loss of generality (y, α, β, t) and $(d_h,\ c_g,\ d_g) = 0$, and we will prove that condition (1c) holds at (y, α, β, t) g.c. point of P^4 for almost every (A, b).

First we show that for almost every (A, b), the gradient of the objective function of $\Phi(f + \frac{x^T Ax}{2} + b^T x, (h_1, \ldots, h_m, g_1, \ldots, g_s)(x))$ is not generated by the gradients of the functions describing the constraints. In other case, taking the combination, in particular it holds that

$$\begin{aligned}
& \nabla_x f(x) + Ax + b + \sum_{i=1}^{m} \lambda_i \nabla_x h_i(x) + \sum_{j=1}^{s} \mu_j \chi'(g_j(x)) \nabla_x g_j(x) + 2(1-t)x \\
= \quad & \sum_{i=1}^{m} \alpha_i \nabla_x h_i + \sum_{j=1}^{s} \beta_j \chi'(g_j(x)) \nabla_x g_j(x).
\end{aligned}$$

As the LICQ fails, x is a linear combination of $\nabla_x[h_1,\dots,h_m,\chi(g_1(x))\dots,\chi(g_s(x))]$. So, rearranging terms it holds that $\nabla_x f(x)+Ax+b$ belongs to the subspace E generated by the $m+s$ vectors $\nabla[h_1,\dots,h_m,\chi(g_1),\dots,\chi(g_s))]$. As $m+s<n$, E is a proper subspace of $\mathbb{R}^n$ and, hence has 0 measure. Recalling that x is fixed, for all A, for almost every b, $\nabla f(x)+Ax+b\notin E$. The result now follows, again, as a consequence of the theorem of Fubbini.

The one-dimensional tangent subspace of (P^4) is obtained as $w=(Wb,w_2,1,0)$, $w_2\in R^{m+s+1}$ and $b=(W^T\nabla_y^2\mathcal{L}W)^{-1}W^T\nabla_y\overline{f}(y,t)$, recall that $\overline{f}$ is the objective function of (7). Writing down the Hessian matrix of the Lagrangian function of (P^4) restricted to w, after some algebraic manipulations, we obtain that it is non singular if and only if $[\nabla\overline{f}(y,t)]^TW(W^T\nabla_y^2\mathcal{L}W)^{-1}W^T[\nabla\overline{f}(y,t)]\neq 0$. However, as $(W^T\nabla_y^2\mathcal{L}W)^{-1}$ is non singular, the contrary holds if and only if $[\nabla\overline{f}(y,t)]^TW=0$ which contradicts the fact that $\overline{f}(y,t)\notin E$. Hence we can conclude that if LICQ fails (y,t) is a point of type 4 for almost every q.

Combining the results of Lemma 4.6 and Lemma 4.7, Theorem 4.5 is shown. □

Now we will prove the genericity theorem

Theorem 4.8. Assume that

$$A=\{(f,h_1,\dots,h_m,g_1\dots,g_s):1-t+\delta\neq 0,\forall(y,t)\in\Sigma_{gc}\Phi(f,h_1,\dots,h_m,g_1\dots,g_s)\}.$$

If $\chi(x)=1_{(-\infty,0)}(x)e^{\frac{1}{x}}$, then the set $\Phi^{-1}(\mathcal{F}\cap\mathcal{A})$ is generic in A with respect to the strong topology.

Proof. As the main ideas can be found in [13], we only give the particularities of this case. First we claim that for $r=3,4,\dots$, the sets

$$I_r=\left\{(f,H,G)\in C^3(\mathbb{R}^n,\mathbb{R}^{1+m+s}):\Sigma_{gc}(\Phi(f,H,G))\cap\left[\mathbb{R}^{n+2m+2s}\times[\frac{1}{r},\frac{r-1}{r}]\right]\subset\bigcup_{i=1}^{5}\Sigma_{gc}^i\right\}$$

are open and dense sets in A. Here $B^p=\{x:\|x\|\le p\}$.

I_r is open: Take $(f^*,H^*,G^*)\in I_r$. $\mathcal{F}$ is open as well as the set of problems such that their g.c. points (y,t), $t\in[\frac{1}{r},1-\frac{1}{r}]$ are of Type 1-5. So, there is $V_{\varepsilon_0(y,t)}$, a neighborhood of $\Phi(f^*,H^*,G^*)$ such that $\Sigma_{gc}(\Phi(f,H,G))\cap\left[\mathbb{R}^{n+2m+2s}\times[\frac{1}{r},1-\frac{1}{r}]\right]\subset\cup_{i=1}^5\Sigma_{gc}^i$. As the g.c. points are always included in the compact set B^p, the definition of $\varepsilon_0(y,t)$ outside B^p is not important. So, we will reduce our analysis to B^p. Take

$$\varepsilon(x)=\min\varepsilon_0(x,\lambda,\mu,v,w,t)\quad \text{s.t } \|x,\lambda,\mu,v,w\|^2\le p,\quad t\in[\frac{1}{r},1-\frac{1}{r}]$$

As the set of feasible solutions is compact, $\varepsilon(x)$ is a positive continuous function on B^p.

By the C^∞-differentiability of χ and the compactness of B^p, there is $\delta>0$ such that if $\|a-b\|<\delta$ and $x\in B^p$, then

$$|\chi(a)-\chi(b)|,|\chi'(a)-\chi'(b)|,|\chi''(a)-\chi''(b)|,|\chi'''(a)-\chi'''(b)|<\varepsilon(x)/(p+1).$$

Define $\varepsilon^*(x) = \min\left\{\frac{\varepsilon(x)}{p+1}, \delta\right\}$, $x \in B^p$. Evidently it is a continuous and positive function on B^p. It will be completed as a continuous function to $\mathbb{R}^n$. We will show that if $(f, H, G) \in V_{\varepsilon^*(x)}(f^*, H^*, G^*)$, then $\Phi(f, H, G) \in V_{\varepsilon(x)}\Phi(f^*, H^*, G^*)$.

First note that for $x \in B^p$, as p large ($p \gg 1$), and $0 < t < 1$,

$$|t[h_i(x) - h_i^*(x)]| < t\frac{\varepsilon(x)}{p+1} \leq \varepsilon(x).$$

Moreover if $\|g_j(x) - g_j^*(x)\| < \delta$ recalling that χ is a monotone increasing function

$$|t[\chi(g_j(x)) - \chi(g_j^*(x))]| < t\frac{\varepsilon(x)}{p+1} \leq \varepsilon(x).$$

Finally

$$\begin{aligned} & |t[f(x) - f^*(x) - \sum_{i=1}^{m} \lambda_i[h_i(x) - h_i * (x)] - \sum_{j=1}^{s} \mu[\chi(g_1(x)) - \chi(g_1(x))]]| \\ < \quad & t\frac{\varepsilon(x)[1 + \|\lambda\| + \|\mu\|]}{p+1} \leq \varepsilon(x) \end{aligned}$$

because $\|\lambda\| + \|\mu\| \leq \|(\lambda, \mu)\| \leq \|y\| \leq p$.

As

$$\Phi(f, H, G) - \Phi(f^*, H^*, G^*) = \begin{pmatrix} t[f(x) - f^*(x) - \sum_{i=1}^{m} \lambda_i[h_i(x) - h_i * (x)]- \\ -t\sum_{j=1}^{s} \mu[\chi(g_1(x)) - \chi(g_1(x))]] \\ t[h_1(x) - h_1 * (x)], \\ \vdots \\ t[h_m(x) - h_m * (x)], \\ t[\chi(g_1(x)) - \chi(g_1^*(x))], \\ \vdots \\ t[\chi(g_s(x)) - \chi(g_s^*(x))], \\ 0 \end{pmatrix}$$

each component is bounded by $\varepsilon^*(x)$. Analogous inequalities fulfill the derivatives up to the $k = 3$ of the functions. So, for all $(f, H, G) \in V_{\varepsilon^*(x)}(f^*, H^*, G^*)$, their g.c. points (y, t), $t \in [\frac{1}{r}, 1 - \frac{1}{r}]$ are of Type 1–5. Hence $V_{\varepsilon(x)}\Phi(f^*, H^*, G^*) \subset I_r$.

But $A \cap I_r$ is an open subset of A. As a consequence,

$$\Phi^{-1}(F \cap A) = A \cap [\cap_{r=2}^{\infty} I_r] = \cap_{r=2}^{\infty}[A \cap I_r]$$

is the intersection of a countable collection of open sets.

I_r is dense: Fix (f, H, G), and C_S^3-neighborhood V defined by $\varepsilon(x)$. Again as B^p contains the feasible sets of $\Phi(f, H, G)$, the analysis will be reduced to the set B^p. Using the Perturbation Theorem (4.5), there exists a (A, b, c, d) such that all the generalized critical points of $\Phi(f(x) + x^T Ax + b^T x; (H(x); G(x)) + (c^T x + d))$ are of type 1, 2, 3, 4 or solutions of $P(1)$ and $\|x^T Ax + b^T x; (c^T x + d)\| < \min_{x \in B^p} \varepsilon(x)$, . Hence $f(x) + x^T Ax + b^T x; (H(x); G(x)) + (c^T x + d) \in V$ and regular. $\square$

5. CONCLUSIONS

In this paper we have presented a new multipliers embedding $\Phi(f,H,G)$ such that if $(f,H,G) \in C^k$, $\Phi(f,H,G) \in C^k$ and generically the solutions of $\Phi(f,H,G)$ can be characterized.

Roughly speaking, for almost every quadratic perturbation of $f(x)$ and linear perturbation of $(H(x),G(x))$, the parametric problem is in the class $\mathcal{F}$ or the g.c. points are solutions of the non parametric problem P, identified by (f,H,G). Moreover, if $\Phi(f,H,G) \notin \mathcal{F}$, we can find variations, as small as desired, such that the regularity of $P(t)$ holds. This result is stronger than the classical theorem because the changes are only made on the original problem (f,H,G).

We considered two illustrative examples and compared the solution of P computed by this embedding and other softwares. We observed that the value of the objective function of P was smaller in our case. Although more numerical experience is needed, at least our proposal behaves well for two non-convex problems.

ACKNOWLEDGEMENT

First of all we would like to thanks the anonymous referees for their careful reading of the manuscript and their valuable comments. The research of the first author was supported by Project CAPES-MES-CUBA 226/2012 “Modelos de Otimização e Aplicações”.

(Received January 28, 2012)

REFERENCES

[1] M. Afonso, J. Bioucas-Dias, and M. Figueiredo: Fast image recovery using variable splitting and constrained optimization. IEEE Trans. Signal Process. *19* (2010), 2345–2356.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. *28* (2008), 1286–1309.

[3] C Avelino and L. N. Vicente: Updating the multipliers associated with inequality constraints in an augmented Lagrangian multiplier method. J. Optim. Theory Appl. *199* (2003), 215–233.

[4] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty: Non Linear Programming Theory and Algorithms. John Willey and Sons, 1993.

[5] D. P. Bertsekas: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York 1982.

[6] E. G. Birgin and J. M. Martínez: Augmented lagrangian method with non-monotone penalty parameters for constrained. Optimization Online, `http://www.optimization-online.org/DB_FILE/2010/06/2662.pdf` 2010.

[7] G. Bouza: A new embedding for the augmented Lagrangean method. Investigación Oper. *22* (2001), 145–153.

[8] G. Bouza and J. Guddat: A note on embeddings for the augmented Lagrange method. Yugosl. J. Oper. Res. *20* (2010), 183–196.

[9] D. Dentcheva, R. Gollmer, J. Guddat, and J. Rückmann: Pathfollowing methods in non linear optimization, multipliers embedding. ZOR *41* (1995), 127–152.

[10] Z. Dostal, A. Friedlander, and A. Santos: Augmented Lagrangians with adaptative precision control for quadratic programming problems with equality constrains. Comput. Optim. Appl. *14* (1999), 37–53.

[11] R. Gollmer, U. Kausmann, D. Nowack, K. Wendler, and J. Bacallao Estrada: Computerprogramm PAFO. Humboldt–Universitaet, Institut fuer Mathematik 2004.

[12] W. Gómez: On generic quadratic penalty embeddings for non linear optimization problems. Optimization *50* (2001), 279–295.

[13] W. Gómez, J. Guddat, H. Th. Jongen, J. J. Rückmann, and C. Solano: Curvas criticas y saltos en la optimizacion no lineal. `http://www.emis.de/monographs/curvas/index.html` 2000.

[14] J. Guddat, F. Guerra, and H. Th. Jongen: Parametric Optimization: Singularities, Pathfollowing and Jumps. Teubner and John Wiley, Chichester 1990.

[15] M. Hirsch: Differential Topology. Springer Verlag, New York 1976.

[16] A. N. Iusem: Augmented Lagrangean methods and proximal point methods for convex optimization. Investigación Oper. *8* (1999), 11–49.

[17] H. Th. Jongen, P. Jonker, and F. Twilt: Critical sets in parametric optimization. Math. Programming *34* (1986), 333–353.

[18] H. Th. Jongen, P. Jonker, and F. Twilt: On one-parametrer families of optimization problems: Equality constrains. J. Optim. Theory Appl. *48* (1986), 141–161.

[19] D. Li and X. L. Sun: Local convexification of the Lagrangian function in non-convex poptimization. J. Optim. Theory Appl. *104* (2000), 109–120.

[20] Z. Li and M. G. Ierapetritou: Production planning and scheduling integration through augmented lagrangian optimization. Comput. and Chemical Engrg. *34* (2010), 996–1006.

[21] D. G. Luenberger and Yinyu Ye: Linear and Nonlinear Programming. Third edition. Internat. Ser. Oper. Res. Management Sci. Springer, New York 2008.

[22] R. Schmidt: Eine modifizierte standard Einbettung zur Behandlung von Gleichungs und Ungleichungs Restriktionen. Master's Thesis, Humboldt Universitaet zu Berlin 2000.

Gemayqzel Bouza Allende, Universidad de La Habana, San Lázaro y L, Vedado, La Habana. Cuba.

e-mail: gema@matcom.uh.cu

Jürgen Guddat, Humboldt–Universität zu Berlin, Unter den Linden 6 D-10099 Berlin. Germany.

e-mail: guddat@mathematik.hu-berlin.de