

A NOTE ON EMBEDDINGS FOR THE AUGMENTED LAGRANGE METHOD

Gemayqzel BOUZA-ALLENDE

Universidad de La Habana
gema@matcom.uh.cu

Jurgen GUDDAT

Humboldt-Universitat zu Berlin
guddat@mathematik.hu-berlin.de

Received: November 2007 / Accepted: December 2010

Abstract: Nonlinear programs (P) can be solved by embedding problem P into one parametric problem $P(t)$, where $P(1)$ and P are equivalent and $P(0)$, has an evident solution. Some embeddings fulfill that the solutions of the corresponding problem $P(t)$ can be interpreted as the points computed by the Augmented Lagrange Method on P . In this paper we study the Augmented Lagrangian embedding proposed in [6]. Roughly speaking, we investigated the properties of the solutions of $P(t)$ for generic nonlinear programs P with equality constraints and the characterization of $P(t)$ for almost every quadratic perturbation on the objective function of P and linear on the functions defining the equality constraints.

Keywords: Augmented Lagrangian Method, JJT-regular, generalized critical points, generic set.

AMS Subject Classification: 90C31, 49M30.

1. INTRODUCTION

We consider the well known nonlinear optimization problem:

$$(P) \quad \min f(x), s.t., x \in M \quad (1)$$

$$M = \left\{ x \in R^n \mid \begin{array}{l} h_i(x)=0, i=1, \dots, m, \\ g_j(x) \geq 0, j=1, \dots, s. \end{array} \right\}$$

$$f, h_1, \dots, h_m, g_1, \dots, g_s \in C^k(R^n, R)$$

Problem P can be solved by algorithms such as the barrier, the penalty and the Augmented Lagrangian method. However, the convergence can be guaranteed under strong assumptions.

Since 1980, embedding methods have been proposed for solving nonlinear programming problems. This approach embeds P into one-parametric problem $P(t)$ and applies a path-following on the set of solutions of $P(t)$ for obtaining a solution of P . In order to have at least a local characterization of this curve, Jongen *et al.* have defined 5 types of points, see [15]-[16]. A parametric problem, such that all its solutions are of some of these types, is considered JJT-regular.

Of course we need to check if this regularity is a strong assumption or not. In the case of the embedding approach, the regularity is a mild hypothesis if the set of problems P such that the embedding defines a JJT-regular problem, is large. Nonlinear problem can be identified by $f, h_1, \dots, h_m, g_1, \dots, g_s \in [C^k(R^n, R)]^{+m+s}$. So, a large set of problems will be understood as a generic set of $[C^k(R^n, R)]^{+m+s}$ endowed with the strong topology, that is, a set equals to the countable intersection open and dense sets.

The Augmented Lagrangian Method, also known as Multipliers approach, combines the penalty and the Lagrange method. In particular, this method solves optimization problem \hat{P}_c , for increasing values of parameter c . By their parametric character, it is natural to relate Multiplier and embedding methods.

In this work we study in detail an embedding for the Augmented Lagrangian method proposed in [6] which calculates saddle points as in the multiplier's method. The two main results of the paper are the perturbation and the genericity theorems for problem with equality constraints ($s=0$). These results prove that for almost every perturbations quadratic of f and linear of h_1, \dots, h_m the embedding constructs a JJT-regular problem and that for a generic problem f, h_1, \dots, h_m , the parametric problem obtained via the embedding, is generic.

The paper has been organized as follows. In the next section we present the main definitions and results of one-parametric optimization and multiplier's method. In Section 3 we introduce the embedding and its relations with the classical Augmented Lagrange Method. After that, we present two numerical examples to illustrate the numerical behavior of the embedding under regularity. Finally, in the last section we prove the genericity and the perturbation theorems.

2. PRELIMINARY ASPECTS AND NOTATIONS

We will consider the optimization problem:

$$\min f(x), \quad s.t. x \in M \tag{2}$$

and the parametric problem $P(t)$ where, for all $t \in T \subset R$ we solve the problem:

$$M = \left\{ x \in R^n \mid \begin{matrix} h_i(x)=0, i=1, \dots, m, \\ g_j(x) \geq 0, j=1, \dots, s. \end{matrix} \right\} \tag{3}$$

$$\min f^t(x, t), \text{ s.t. } x \in M(t) \tag{4}$$

$$M(t) = \left\{ x \in R^n \mid \begin{matrix} h_i(x, t)=0, i=1, \dots, m, \\ g_j(x, t) \geq 0, j=1, \dots, s. \end{matrix} \right\} \tag{5}$$

Here $f^t, h_1^t, \dots, h_m^t, g_1^t, \dots, g_s^t \in C^k(R^n \times R, R), k \geq 3$ are the functions defining the objective function and the constraints of $P(t)$. For simplicity $h^\bullet(x, t) = (h_1^t(x, t), \dots, h_m^t(x, t))$ and $g^\bullet(x, t) = (g_1^t(x, t), \dots, g_s^t(x, t))$.

First we introduce the following notations.

The indicator function of set A is $1_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise} \end{cases}$

$\Pi_x : R^{n+m} \times R^n, \Pi_x(x, y) = x$ is the projecting function onto R^n . $R_+^n = \{x \in R^n : x \geq 0\}$, I_m denotes the identity matrix of dimension m and the space of symmetric $n \times n$ -matrices is identified by $R^{n(n+1)/2}$.

Classical definitions and results for nonlinear programs can be easily extended to parametric optimization, for more details, see [10] and [11]. Following their notations, the set of active index of (x, t) is $J_0(x, t) = \{j : g_j^t(x, t) = 0\}$ and the Lagrangian of $P(t)$ is denoted by $L(x, \lambda, \mu) = f^t(x, t) - \sum_{i=1}^m \lambda_i h_i^t(x, t) - \sum_{j \in J_0(x, t)} \mu_j g_j^t(x, t)$, where, $\lambda_1, \dots, \lambda_m, \mu_j, J \in_0(x, t)$ are the associated Lagrange multipliers.

Analogously, a Fritz John necessary condition can be formulated. That is, if x^* is a local minimizer of $P(t^*)$, then (x^*, t^*) is a FJ point, *i.e.* there is a non zero vector

$$(\mu_0, \lambda, \mu) \in R_+ \times R^m \times R_+^{|J_0(x, t)|} \text{ such that} \tag{6}$$

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

Note that if LICQ holds at (x^*, t^*) FJ point, *i.e.*

$$\nabla_x h_1^t(x^*, t^*), \dots, \nabla_x h_m^t(x^*, t^*), \nabla_x g_j^t(x^*, t^*), j \in J^0(x, t)$$

are linearly independent, then, in system (6), $\mu_0 \neq 0$. Without losing of generality (wlog), under LICQ, we take $\mu_0 = 1$.

Let us consider the feasible points (x, t) such that $J_0(x, t) = J_0$. Then (x, t) solves the following system:

$$\begin{aligned} h_i^t(x, t) &= 0, i = 1, \dots, m, \\ g_j^t(x, t) &= 0, j \in J_0 \end{aligned} \tag{7}$$

In particular, FJ points with $J_0(x, t) = J_0$ can be computed by solving system (6-7) and checking if $g_j^t(x, t) \geq 0, j \notin J_0$ and $(\mu^0, \mu) \geq 0$. If $(x, t, \mu_0, \lambda, \mu)$ fulfils (6-7), but $(\mu_0, \mu \geq 0)$ then (x, t) is a generalized critical (g.c.) point. Those g.c. points where LICQ holds are called critical points. The set of critical points is denoted by Σ_{crit} and the set of g.c. points, by Σ_{gc} .

Around $(x^0, t^0) \in \Sigma_{gc}$, the set J^0 can change or $(\mu_0, \mu_j, j \in J_0, g_k(x, t), k \in J_0)$ may become strictly negative. This means that, locally, even the feasibility may be lost. In order to determine which possibilities may appear, JJT-regularity defines 5 types of g.c. points, which allow the local characterization of Σ_{gc} . Let us begin with the definition of g.c. points of Type 1, also called non-degenerate critical point:

Type 1 $(\bar{x}, \bar{t}) \in \Sigma_{gc}^1$, if

(1a) LICQ holds.

(1b) $\bar{\mu}_j \neq 0$, for all $j \in J_0(\bar{x}, \bar{t})$.

(1c) $\nabla_x^2 L(\bar{x}, \bar{\lambda}, \bar{\mu}, \bar{t})|_{T_x M(t)(\bar{x}, \bar{t})}$ is non-singular, where $T_x M(t)(\bar{x}, \bar{t})$ is the tangent space of $M(\bar{t})$ at \bar{x} .

A matrix A over a subspace T is denoted by $A|_T$. It is non singular $V^T A V$ is regular for some (and hence all) matrix V , whose columns form a basis of T . As the LICQ holds at $\bar{x} \in M(\bar{t})$ the tangent subspace is

$$T_x M(t)(\bar{x}, \bar{t}) = \left\{ \xi \in \mathbb{R}^n : \nabla_x h_i^t(x^*, t^*) \xi = 0, i = 1, \dots, m, \nabla_x g_j^t(x^*, t^*) \xi = 0, j \in J_0(x, t) \right\}$$

Locally around a point of type 1, $J_0(\bar{x}, \bar{t})$ is constant as well as the number of positive multipliers $\bar{\mu}$ and positive eigenvalues $\nabla_x^2 L(\bar{x}, \bar{\lambda}, \bar{\mu}, \bar{t})|_{T_x M(t)(\bar{x}, \bar{t})}$. If $(x, t) \notin \Sigma_{gc}^1$, then it is a singular (or a degenerated) g.c. point. The points of the types 2-5 are singular points representing the four basic singularities (for the detailed definition, we refer to [10], [11], [14] and [15]). Here we only present the points of type 2 and 3, because the other two types, corresponding with the violation of LICQ, will not appear in our case.

Type 2: violation of (1b).

Type 3: violation of (1c).

Around points of type 2, $J_0(\bar{x}, \bar{t})$ changes, however the two possible sets and the corresponding g.c. points are easy to compute. The changes on the signs of $\bar{\mu}$ and of the eigenvalues of $\nabla_x^2 L(\bar{x}, \bar{\lambda}, \bar{\mu}, \bar{t})|_{T_x M(t)(\bar{x}, \bar{t})}$ are also known. For points of type 3, $J_0(\bar{x}, \bar{t})$ and the sign of $\bar{\mu}$ remain unchanged, however there is an eigenvalue of $\nabla_x^2 L(\bar{x}, \bar{\lambda}, \bar{\mu}, \bar{t})|_{T_x M(t)(\bar{x}, \bar{t})}$ whose sign changes.

Figure 1, see [11], shows the local structure of Σ_{gc} around each of these 3 types. The dotted line represents a change in the sign of the associated multipliers $\mu \geq 0$. Here $\bar{z} = (\bar{x}, \bar{t})$.

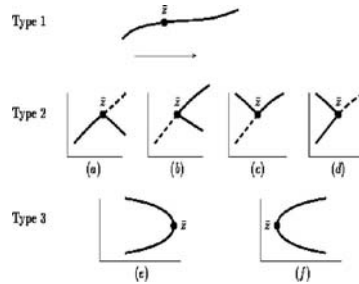


Figure 1: Local structure of Σ_{gc}

Let us now define the JJT-regularity.

Definition 1: Let $\Sigma_{gc}^v, v = 1, \dots, 5$, be the set of g.c. points of type v . The class F is defined by $F = \{f^t, h^\bullet, g^\bullet \in C^3(R^n \times R, R)^{1+m+s} \mid \Sigma_{gc} \subset \bigcup_{v=1}^5 \Sigma_{gc}^v\}$ and its elements are called regular in the sense of Jongen-Jonker-Twilt, or JJT-regular

In [10] it is shown that the set F is open and dense with respect to the strong topology on $C^3(R^n \times R, R)^{1+m+s}$, see [12] for more details on these topological aspects. On the other hand, the following perturbation result holds. Given $f^t, h^\bullet, g^\bullet \in C^3(R^n \times R, R)^{1+m+s}$, for almost every $Q = (b, A, c, D, e, F) \in R^{n+n(n+1)/2+m+mn+s+sn}$ we have

$$(f^t(x, t) + b^t x + x^T A x, h^\bullet(x, t) + c + D_x, g^\bullet(x, t) + e + F_x) \in F$$

"Almost every" means: the Lebesgue-measure of each measurable subset of Q such that

$$(f^t(x, t) + b^t x + x^T A x, h^\bullet(x, t) + c + D_x, g^\bullet(x, t) + e + F_x) \notin F, \text{ is zero.}$$

The main tool used in these proofs is the following lemma:

Lemma 1: (Parametric Sard's Lemma), (cf. [10]) Let us consider $\varphi: R^n \times R^p \rightarrow R^r$ be a C^k -function with $k > \max\{0, n-r\}$. If 0 is a regular value of φ , (i.e. $\text{rank}(\nabla_{x,z}\varphi(x, z)) = r$ for all (x, z) such that $\varphi(x, z) = 0$), then, for almost every $z \in R^p$, 0 is a regular value of $\varphi_z(x) = \varphi(x, z)$.

Now we present the main ideas of the embedding approach.

Embedding approach

An embedding can be seen as an application Φ such that for each nonlinear problem P , $\Phi(P)(t) = (t)$ fulfils

- $P(1)$ and P are equivalent
- There is a solution point of $P(t)$ for all $t \in [0,1]$.
- $P(0)$ has an evident solution

Starting with $(x(0),0)$ where $x(0)$ is an evident solution of $P(0)$, we try to follow a path $(x(t),t)$ where $x(t)$ is a solution of $P(t)$ for all $t \in [0,1]$.

The main difficulty of this approach is that we need to trace the path $(x(t),t)$ where $x(t)$ solves $P(t)$. From a numerical viewpoint, the solution is understood as a g.c. point. If $P(t)$ is JTT-regular, the types of g.c. points that may appear are known. That is why it is important to investigate if this hypothesis holds for generic problem P . Examples of genericity analysis of embeddings can be found in [9] and [19].

As we are going to deal with an embedding for the multipliers method, we need to establish the links between this embedding and the method. So, we will present the properties of the Augmented Lagrangian method.

Augmented Lagrangian Method

This algorithm, also known as the Multiplier's Method, appears in [3]. It constructs an optimization problem, whose objective function includes the Lagrange function and a quadratic penalty term. Roughly speaking, for problem P described in (2), the Lagrange method considers the parametric problem:

$$P_c(\lambda, \mu) \quad \min_x F_L(x, \lambda, \mu, c)$$

where

$$F_L(x, \lambda, \mu, c) = f(x) - \sum_{i=1}^m \lambda_i h_i(x) + c \sum_{i=1}^m h_i^2(x) + c \sum_{j=1}^s \left[\left(\frac{\mu_j}{2c} \right)^2 g_j(x) + \left(\min \left\{ 0, g_j(x) + \frac{\mu_j}{2c} \right\} \right)^2 \right] \quad (8)$$

Problem $P_c(\lambda, \mu)$ is solved for fixed (λ, μ, c) . Given $x_c(\lambda, \mu)$ its solution, the multipliers (λ, μ) and the penalty parameter c are updated in order to fulfill that c is large enough and $\mu \geq 0$. Then the process is repeated for the new values of the parameters. This approach computes saddle points of the Lagrange function minimum in x and maximum in the multipliers (λ, μ) , (for more details see [2] and [18])

This class of algorithms was applied to the solution of variational inequality problems in [13] and solves quadratic programming problems in [7]. Improvements of it via convexifications are discussed in [16]. New ways of updates can also be found in [1].

Based on these properties, given a nonlinear program P , the multipliers embedding shall define a parametric problem $\Phi(P)(t)$ such that the objective function

includes the Lagrangian function of P and a quadratic penalty term. On the other hand, as the multipliers method generates a sequence of saddle points of the Lagrange function of P , saddle points of the objective function of $\Phi(P)(t)$ shall be easy to compute at least locally around $t = 0$. In the next section we are going to present the embedding which fulfills these properties.

3. EMBEDDINGS FOR THE MULTIPLIERS METHOD

As already sketched, given problem P , the parametric program $\Phi(P)(t)$, defined by the multipliers embedding shall satisfy:

- There is (x_0, λ_0, μ_0) , an evident saddle point of the objective function of $\Phi(P)(0)$, minimum in x and maximum in (λ, μ) .
- $(x(t), \lambda(t), \mu(t))$, is a $\min_x - \max_{\lambda, \mu}$ point of the objective function of $\Phi(P)(t)$ and coincides with the set of g.c. points around $(x_0, \lambda_0, \mu_0, 0)$.
- The objective function includes the Augmented Lagrangian function of the original problem P .

The second condition can be guaranteed if $(x_0, \lambda_0, \mu_0, 0)$, is a point of type 1.

So, at least locally around, $(x_0, \lambda_0, \mu_0, 0)$ the set Σ_{gc} is a curve of saddle points of the objective function of $\Phi(P)(t)$, $\min_x - \max_{\lambda, \mu}$.

Let us present the multiplier embedding. Proposed in [6], the embedding defines, for each problem P (2), the parametric problem $\Phi(P)(t)$:

$$P(t) = \min_{x, \lambda, \mu \geq 0} tF_L(x, \lambda, \mu, t) + (1-t) \left[(x - x_0)^T A(x - x_0) - \|\lambda - \lambda_0\|^2 - \|\mu - \mu_0\|^2 \right] \quad (9)$$

$$F_L(x, \lambda, \mu, c) = f(x) - \sum_{i=1}^m \lambda_i h_i(x) - \sum_{i=1}^s \mu_i g_i(x) + \left(\frac{t}{1-t} \right)^2 \left(\sum_{i=1}^m h_i^2(x) + \sum_{i=1}^s [\min\{0, g_j(x)\}]^2 \right) \quad (10)$$

where A is a positive definite matrix.

Note that the objective function is not in C^3 if $s > 0$. As this differentiability condition is needed for our analysis, from now on we will assume that $s=0$. As the problem (9) is not defined at $t=1$, the study of Σ_{gc} is done in the closed subintervals of

$I=[0, 1)$ such as $\left[0, 1 - \frac{1}{n}\right]$, $n \rightarrow \infty$. The solutions for $t=1$ are understood as limits points of $(x, t) \in \Sigma_{gc}$, $t \rightarrow 1^-$.

Note that this embedding is a multiplier embedding. Evidently the objective function of (9), contains the Augmented Lagrangian of P . On the other hand, (x_0, λ_0, μ_0) is a saddle point of the objective function at $t=0$ of type 1.

Another important question is what may happen under JJT-regularity. As there are no constraints in problem (9), recall $s=0$, the following result is satisfied.

Proposition 3.1 *If $\Phi(P)(t) \in F$ then the g.c. points of $\Phi(P)(t)$ are of type 1 or 3.*

If all g.c. points of the parametric problem $\Phi(P)(t)$ are of type 1, it can be seen that there is a curve of g.c. points starting at (x_0, λ_0) parameterized by $(x(t), \lambda(t), \mu(t), t)$, $t \in [0,1]$. Moreover, $(x(t), \lambda(t))$, are saddle points of the objective function of $\Phi(P)(t)$, but, this good situation is not always possible. On the other hand, can we expect to find the solution if there are singular g.c. points. In the next section we will illustrate these two cases with two numerical examples.

4. TWO ILLUSTRATIVE EXAMPLES

In this section, we solve the two non-convex optimization problems P_1, P_2 by the multiplier embedding (9) with parameters $A=I_n$ and $(x_0, \lambda_0) = 0$. The parametric problem is solved by the path-following routine PAFO, see [8].

Path-following methods are widely used, for example in nonlinear optimization problems ([17], [20], [22] and [21]) and in variational inequality's problems ([5]). PAFO is a path-following and jumps routine for solving JJT-regular parametric problems. Locally around $(x, t) \in \bigcup_{v=1}^5 \Sigma_{gc}^v$, the set of g.c. points can be described as the solutions of (finitely many) well known nonlinear systems. So, given a starting solution and under JJT regularity, PAFO solves those systems by a predictor-corrector scheme and hence computes the g.c. points around (x, t) . Although it is possible to jump to another connected component (see ch.5, [11]), 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 $\Phi(P)(t)$, which will be solved by PAFO. To guarantee the existence of a solution of $\Phi(P)(t)$ for all $t \in [0,1]$, we add the restriction $\|x\|^2 + \|\lambda\|^2 \leq p$. Here $p=500$.

Let us begin with the first example. The problem is

$$(P_1) \quad \min x_2 \quad s.t. \quad 1 + x_1(1 - x_1 + x_1^2) = 0 \quad (11)$$

So, the parametric problem is

$$\begin{aligned} \min x_2 + \lambda \left[1 + x_1(1 - x_1 + x_1^2) - x_2 \right] + \left(\frac{t}{1-t} \right)^2 \left[1 + x_1(1 - x_1 + x_1^2) \right]^2 \\ + (1-t) \left[\|x - (1,1)\|^2 - \lambda^2 \right] \quad s.t. \quad x_1^2 + x_2^2 + \lambda^2 \leq 500 \end{aligned} \quad (12)$$

The point $(x_1^*, x_2^*, \lambda^*) = (-0.3040, 0.5754, 0)$ was the computed solution. As can be seen in Figure 2 only appeared points of type 1. In fact, we obtained the ideal situation

of computing saddle points of the objective function for $t \in [0,1)$. So, it shows that the computation of saddle points for all t is possible for non-convex problems.

In the second example, we changed one coefficient of $h(x)$ and applied the embedding to the problem

$$(P_2) \quad \min x_2 \quad s.t. \quad 1 + x_1(1 - x_1)^2 - x_2 = 0 \tag{13}$$

The resulting parametric problem is

$$\begin{aligned} \min x_2 + \lambda [1 + x_1(1 - x_1)^2 - x_2] + \left(\frac{t}{1-t}\right)^2 [1 + x_1(1 - x_1)^2 - x_2]^2 \\ + (1-t) [\|x - (1,1)\|^2 - \lambda^2] \quad s.t. \quad x_1^2 + x_2^2 + \lambda^2 \leq 500 \end{aligned} \tag{14}$$

PAFO found three singular points (two of Type 3 and one of Type 2) while solving problem (14), see Figure 3. Here the computed solution was $(x_1^*, x_2^*, \lambda^*) = (-2.229, -22.249, 0)$, very close to the solution of problem (11) with the extra constraint $x_1^2 + x_2^2 \leq 500$ computed by GAMS 22.2.

The classical multipliers embedding, see [4] did not even find a feasible point. This example shows that, although singular points may appear, the embedding (9) may find the solution of the original problem.

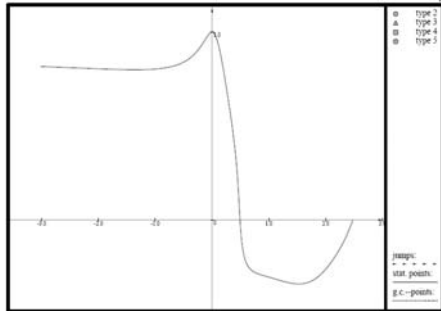


Figure 2: x_1 vs. t

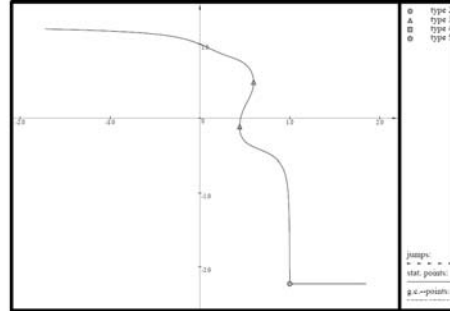


Figure 3: x_1 vs. t

As we could observe, in both examples PAFO was able to solve the parametric problems. In particular, this means that $\Phi(P_1)$ and $\Phi(P_2)$, were JJT-regular. But, what can we say about this hypothesis. Is it too strong, *i.e.* is $\Phi(P) \in F$ for generic P ?

In the next section we will study this problem.

5. MAIN RESULTS

In this section we will analyze which critical points may appear in the generic case. We will prove that generically $\Phi(P)(t)$ is JJT-regular. As $s=0$, $\Phi(P)(t)$ will be:

$$P(t) = \min_{x,\lambda,\mu \geq 0} tF_L(x, \lambda, \mu, t) + (1-t) \left[(x - x_0)^T A(x - x_0) - \|\lambda - \lambda_0\|^2 \right] \quad (15)$$

$$F_L(x, \lambda, \mu, c) = f(x) - \sum_{i=1}^m \lambda_i h_i(x) \left(\frac{t}{1-t} \right)^2 \left(\sum_{i=1}^m h_i^2(x) \right) \quad (16)$$

We begin studying how to take the parameters in order to have JJT-regularity for problem (15) fixed functions $f(x), h_1(x), \dots, h_m(x)$. We begin with a more general result, which do not take into account the positive definiteness of A and the term $-2x_0^T Ax$ is substituted by $y_0^T x$.

Theorem 1 Fix functions $f(x), h_1(x), \dots, h_m(x)$ and parameter λ_0 . For almost every $Q=(A, y_0)$ the generalized critical points of

$$P(t) = \min_{x,\lambda,\mu \geq 0} tF_L(x, \lambda, \mu, t) + (1-t) \left[x^T A x - 2y_0 x - \|\lambda - \lambda_0\|^2 - \|\mu - \mu_0\|^2 \right] \quad (17)$$

with $0 < t < 1$ are of type 1 or 3. Here $F_L(x, \lambda, \mu, t)$ is defined as in (16).

Proof: Using the same ideas described in [10], we write Σ_{gc} as solutions of one of the following systems of equations (one for each possible combination of (z_B, z_D, z_C)):

$$\begin{aligned} H(x, \lambda, \mu, t, A, y_0) &= 0 \\ \nabla_{x\lambda} H(x, \lambda, \mu, t, A, y_0) &= z \\ z_B &= z_C z_D^{-1} Z_C^T \end{aligned} \quad (18)$$

Here $H(x, \lambda, \mu, t, A, y_0)$ is the gradient of the objective function of (17) (with respect to (x, λ) z is an auxiliary variable, $\nabla_{x\lambda} H(x, \lambda, \mu, t, A, y_0) = \begin{pmatrix} B & C \\ C^T & D \end{pmatrix}$, D is a symmetric square matrix such that, $rank(D) = rank(\nabla_{x\lambda} H(x, \lambda, \mu, t, A, y_0)) = n - k$. Precisely the last two blocks of equations represent this rank condition. As $\nabla_{x\lambda} H(x, \lambda, \mu, t, A, y_0) = (\otimes I_m)$, D can be chosen in such a way that I_m is a sub-matrix of it. So B is a symmetric sub-matrix of $\otimes + 2(1-t)A$.

Then writing the derivatives of the functions describing system (18) with respect to the variables and parameters, we obtain:

	D_x	D_λ	D_z	D_{y_0}	D_A
$H = 0$	\otimes	\otimes	\otimes	$(1-t)I_n$	\otimes
$\nabla_{x\lambda} H(x, \lambda, \mu, t, A, y_0)$	\otimes	$(1-t)I_m$	0	0	0
$z_B = z_C z_D^{-1} Z_C^T$	0	0	$I_{\bullet\bullet} \otimes$	0	$\frac{(1-t)I_n}{0}$

(19)

Here $** = k(k+1)/2$

As B is a sub-matrix of $\nabla_x H(x, \lambda, \mu, t, A, y_0)$ and $* \leq n(n+1)/2$, this matrix has row full rank. Using the parametric Sard's lemma 1, it follows that for almost every Q , the rows corresponding with the derivatives with respect to (x, λ, t) of (18) has also full row rank. The rest of the proof uses the same arguments as in Theorem 6.18 [10].

Now we consider the case of the embedding where y_0 has a particular structure and A is positive definite.

Corollary 5.2 For almost every $Q=(A, x_0)$, A definite positive, the generalized critical points of the parametric problem (15) with $0 < t < 1$ are of type 1 or 3.

Proof: The set of symmetric positive definite matrices M is an open subset of the symmetric matrix set. From Theorem 1, the problem (17) is JTT-regular almost everywhere respect to the Lebesgue measure restricted to $M \times R^n$. As A is a non singular matrix, using the linear isomorphism $x_0 = 2Ay_0$, the structure of the embedding (15) is recovered.

Remark 1 This result shows that fixed problem P and parameter λ_0 , for almost every of (A, x_0) problem $\Phi(P) \in F$. As a consequence for all open set U we can find parameters $(A, x_0) \in U$ such that the corresponding embedding defines a regular problem.

Now we are going to prove a general genericity result for the embedding (9). As already remarked P can be identified by the functions defining its objective function $f(x)$ and constraints $h_1(x), \dots, h_m(x)$. So, $\Phi(f, h_1, \dots, h_m)$ represents the parametric problem $\Phi(P)$.

Theorem 2 The set $I = \left\{ f, h_1, \dots, h_m : \Phi(f, h_1, \dots, h_m) \Big|_{(0,1)} \in F \right\}$ is generic, i.e. is the intersection of a numerable collection of open and dense sets.

Proof: Let us first note that $I = \bigcup_{n=1}^{\infty} \left\{ f, h_1, \dots, h_m : \Phi(f, h_1, \dots, h_m) \Big|_{\left(\frac{1}{n}, 1 - \frac{1}{n}\right)} \in F \right\}$. If

we prove that these sets are open and dense, the result will be obtained.

First step: **Density.**

Lemma 2 [Perturbation Lemma]: Let $r = n(n-1)/2 + n + mn$. Every measurable subset of

$\left\{ (A, b, c) \in R^r : \Phi(f, h_1, \dots, h_m) \Big|_{\left(\frac{1}{n}, 1 - \frac{1}{n}\right)} \notin F \right\}$ has Lebesgue measure equal to zero.

Proof of Lemma 2: For the proof, we are going to use the same technique of the Perturbation Theorem (cf [10]).

Considering the manifold in variables (x, λ, A, b, c) described by:

$$\begin{aligned} H(x, \lambda, \mu, t, A, y_0) &= 0 \\ \nabla_{x\lambda} H(x, \lambda, \mu, t, A, y_0) &= z \\ z_B &= z_C z_D^{-1} Z_C^T \end{aligned}$$

The derivatives with respect to the variables and parameters are:

	D_x	D_λ	D_z	D_b	D_A
$H = 0$	\otimes	\otimes	\otimes	tI_n	\otimes
$\nabla_{x\lambda} H(x, \lambda, \mu, t, A, y_0)$	\otimes	$(1-t)I_m$	0	0	0
$z_B = z_C z_D^{-1} z_C^T$	0	0	$I_{\bullet\bullet} \otimes$	0	0

and it has obviously full row rank. So, the desired property follows as in Theorem 1.

Second step: Openness

Lemma 3 Let us assume that $\Phi(f, h_1, \dots, h_m) \Big|_{\left(\frac{1}{n}, 1, \frac{1}{n}\right)} \in F$. Then for every (x, λ, t) there are two neighborhoods U of (x, λ, t) and V of (f, h_1, \dots, h_m) such that:

- If $(x, \lambda, t) \in \Sigma_{gc}$, then the g.c. points of every element of $\Phi(V)$ are of type 1 or 3.
- If $(x, \lambda, t) \notin \Sigma_{gc}$ the elements of $\Phi(V)$ have not g.c. points in U .

Proof of Lemma 3: As $\Phi(f, h_1, \dots, h_m) \Big|_{\left(\frac{1}{n}, 1, \frac{1}{n}\right)} \in F$, if $(x, \lambda, t) \in \Sigma_{gc}$ then it is of type 1 or 3, see Proposition 3.1. If the neighborhoods U and V do not exist, then we can find sequences $(f, h_1, \dots, h_m)_k \rightarrow (f, h_1, \dots, h_m)$ and $(x, \lambda, t)_k \rightarrow (x, \lambda, t)$, such that

$$(x, \lambda, t)_k \in \Sigma_{gc} \Phi(f, h_1, \dots, h_m)_k / \left[\Sigma_{gc}^1 \cup \Sigma_{gc}^3 \right]$$

If $(x, \lambda, t) \in \Sigma_{gc}^1$ and $(x, \lambda, t)_k \notin \Sigma_{gc}^1$, (1c) is violated. But

$(f, h_1, \dots, h_m)_k \rightarrow (f, h_1, \dots, h_m)$ So for n large enough $(x, \lambda, t)_k$ are in a compact set K and

$$\|(f, h_1, \dots, h_m)_k - (f, h_1, \dots, h_m)\|_K < \max_{x \in K} \epsilon(x)$$

This means that there is a uniform convergence in the compact K and

$$\frac{\partial^r f_k}{\partial x} (x_k, \lambda_k, t_k) \rightarrow \frac{\partial^r f}{\partial x} (x, \lambda, t), r \leq 3, \text{ as well as for } (h_1, \dots, h_m).$$

If the Hessian of the Lagrange function of $\Phi(f, h_1, \dots, h_m)_k$ is denoted by $D^2 L_k$, by continuity, we have that $D^2 L_k(x_k, \lambda_k, t_k) \rightarrow D^2 L(x, \lambda, t)$, the Hessian matrix corresponding to $\Phi(f, h_1, \dots, h_m)$.

But $D^2L(x, \lambda, t)$ is not singular, so for k large enough, $D^2L_k(x_k, \lambda_k, t_k)$ must be regular. Then the sequence has only points of type 1 for k large enough. If $(x, \lambda, t) \in \Sigma_{gc}^3$, the result is analogous. In this case the sequence has points of type 1 or 3.

Now, the rest of the proof of Theorem 2, follows as in Theorem 6.22 [10].

The genericity results means that, given the parameters $(A_0, x_0, \lambda_0,)$, the embedding will define regular problems on a large set of nonlinear programs. Moreover, by Lemma 2, fixed P , there are perturbations as small as desired, such that the resulting parametric problem (15) is JJT-regular. As a consequence, we can say that JJT-regularity is not a strong hypothesis.

6. CONCLUSIONS

In this paper we have considered an embedding for the Lagrange multiplier's method, proposed in [6]. In order to guarantee differentiability, we have regarded the case in which P has no inequality constraints and the parametric problem was considered for $t \in [0,1)$. For this case, we have proved that it is not too strong to assume that the defined problem is JJT regular. Indeed, fixed P problem (respectively the parameters $(A, x_0, \lambda_0,)$ defining embedding $\Phi(P)(t)$ there is a large set of parameters $(A, x_0, \lambda_0,)$ (respectively problems P such that corresponding parametric problem $\Phi(P)(t)$ is JJT-regular.

However this embedding is not suitable for nonlinear programs with inequality constraints. That is why future work will be devoted to the construction of an embedding without this drawback.

Acknowledgements: We want to thank Prof Georg Still and the anonymous referees for their helpful comments and Jutta Kerger for revising a preliminary version of this manuscript.

REFERENCES

- [1] Avelino, C., and Vicente, L.N., "Updating the multipliers associated with inequality constraints in an augmented Lagrangian multiplier method", *Journal of Optimization Theory and Applications*, 119 (2003) 215–233.
- [2] Bazaraa, M.S., Sherali, H.D., and Shetty, C.M., *Non Linear Programming Theory and Algorithms*, John Willey and Sons, 1993.
- [3] Bertsekas, D.P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [4] Bouza, G., "A new embedding for the augmented Lagrangean method", *Investigacion operacional*, 22 (3) (2001) 145–153.
- [5] Chen, B., and Chen, X., "A global linear and local quadratic continuation smoothing method for variational inequalities with box constraints", *Computational Optimization*, 17(2000) 131–158.
- [6] Dentcheva, D., Gollmer, R., Guddat, J., and Ruckmann, J., "Pathfollowing methods in non linear optimization, multipliers embedding", *ZOR*, 41 (1995) 127–152.
- [7] Dostal, Z., Friedlander, A., and Santos, A., "Augmented Lagrangians with adaptative precision control for quadratic programming problems with equality constrains", *Computer Optimization Applications*, 14 (1) (1999) 37–53.

- [8] Gollmer, R., Kausmann, U., Nowack D., Wendler, K., and Bacallao Estrada J., *Computer Program PAFO*, Humboldt-Universitaet, Institut fuer Mathematik, 2004.
- [9] Gomez, W., "On generic quadratic penalty embeddings for non linear optimization problems", *Optimization*, 50 (3-4) (2001) 279–295.
- [10] Gomez, W., Guddat, J., Jongen, H. Th., Ruckmann, J.J., and Solano, C., *Curvas criticas y saltos en la optimización no lineal*. <http://www.emis.de/monographs/curvas/index.html>, 2000.
- [11] Guddat, J., Guerra-Vasquez, F., and Jongen, H. Th., *Parametric Optimization: Singularities, Pathfollowing and Jumps*, Teubner and John Wiley, Chichester, 1990.
- [12] Hirsch, M., *Differential Topology*. Springer Verlag, New York, 1976.
- [13] Iusem, A.N., "Augmented Lagrangean methods and proximal point methods for convex optimization", *Investigacion Operativa*, 8 (1999) 11–49.
- [14] Jongen, H.Th., Jonker, P., and Twilt, F., "Critical sets in parametric optimization", *Mathematical Programming*, 34(3) (1986) 333–353.
- [15] Jongen, H.Th., Jonker, P. and Twilt, F., "On one-parametere families of optimization problems: Equality constraints", *Journal of Optimization Theory and Applications*, 48 (1) (1986) 141–161.
- [16] Li, D., and Sun, X.L., "Local convexification of the Lagrangian function in non-convex optimization", *Journal of Optimization Theory and Applications*, 104 (1) (2000) 109–120.
- [17] Lin, Z., and Song, D., "A continuation method for solving convex programming via Fisher reformulation", *Optimization*, 44 (3) (1998) 291–302.
- [18] Luenberger, D.G., *Linear and Nonlinear Programming*, Addison-Wesley Inc., Reading, Massachusetts, Second edition, 1984.
- [19] Schmidt, R., *Eine Modifizierte Standard Einbettung zur Behandlung von Gleichungs und Ungleichungs Restriktionen*, Master's thesis, Humboldt Universitaet zu Berlin, 2000.
- [20] Song, X., "A non-interior pathfollowing method for convex quadratic programming problems with bound constraints", *Computational Optimization and Applications*, 27 (3) (2004) 285–303.
- [21] Watson, L., "Theory of globally convergent probability-one homotopies for nonlinear programming", *SIAM Journal on Optimization*, 11 (3) (2000) 761–780.
- [22] Wright, S.J., "A path-following interior point algorithm for linear and quadratic problems", *Annals of Operations Research*, 62 (1994) 103–130.