Mathematical Programming manuscript No.
(will be inserted by the editor)

R. Andreani · E. G. Birgin · J. M. Martínez · M. L. Schuverdt

Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification*

Received: date / Revised version: date

Abstract. Two Augmented Lagrangian algorithms for solving KKT systems are introduced. The algorithms differ in the way in which penalty parameters are updated. Possibly infeasible accumulation points are characterized. It is proved that feasible limit points that satisfy the Constant Positive Linear Dependence constraint qualification are KKT solutions. Boundedness of the penalty parameters is proved under suitable assumptions. Numerical experiments are presented.

Key words. Nonlinear programming – Augmented Lagrangian methods – KKT systems – numerical experiments.

1. Introduction

Let $F : I\!R^n \to I\!R^n$, $h : I\!R^n \to I\!R^m$ and $\Omega = \{x \in I\!R^n \mid \ell \le x \le u\}$, where $\ell, u \in I\!R^n$, $\ell < u$. Assume that h admits continuous first derivatives on an open set that contains Ω and denote

$$\nabla h(x) = (\nabla h_1(x), \ldots, \nabla h_m(x)) = h'(x)^T \in I\!R^{n \times m}.$$

Let $\mathcal{P}_A$ denote the Euclidian projection operator onto a closed and convex set A. A point $x \in I\!R^n$ is said to be a KKT point of the problem defined by F, h and Ω if there exists $\lambda \in I\!R^m$ such that

$$\mathcal{P}_\Omega[x - F(x) - \nabla h(x)\lambda] - x = 0, \quad h(x) = 0, \quad x \in \Omega. \tag{1}$$

KKT points are connected with the solution of Variational Inequality Problems. The Variational Inequality Problem (VIP) defined by $(F, \mathcal{D})$ (see, for example, [34]) consists of finding $x \in \mathcal{D}$ such that $F(x)^T d \ge 0$ for all d in the tangent cone $T_\mathcal{D}(x)$ ([3], page 343). Defining $\mathcal{D} = \{x \in \Omega \mid h(x) = 0\}$, and under

R. Andreani, J. M. Martínez and M. L. Schuverdt: Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970, Campinas SP, Brazil. e-mail: `{andreani|martinez|schuverd}@ime.unicamp.br`

E. G. Birgin: Department of Computer Science IME-USP, University of São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-090, São Paulo SP, Brazil. e-mail: `egbirgin@ime.usp.br`

Mathematics Subject Classification (1991): 65K05, 65K10

* The authors were supported by PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1, FAPESP (Grants 2001/04597-4, 2002/00094-0, 2003/09169-6, 2002/00832-1 and 2005/56773-1) and CNPq.

appropriate constraint qualifications, it turns out that the solutions of the VIP are KKT points. In particular, if $f : I\!R^n \to I\!R$ is differentiable and $F = \nabla f$, the equations (1) represent the KKT optimality conditions of the minimization problem

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0, \; x \in \Omega. \tag{2}$$

The most influential work on practical Augmented Lagrangian algorithms for minimization with equality constraints and bounds (problem (2)) was the paper by Conn, Gould and Toint [12], on which the LANCELOT package [10] is based. Convergence of the algorithm presented in [12] was proved under the assumption that the gradients of the general constraints and the active bounds at any limit point are linearly independent. In [11] the authors extended the method of [12] to the case where linear inequality constraints are treated separately and also to the case where different penalty parameters are associated with different general constraints.

In the present paper we introduce two Augmented Lagrangian algorithms for solving (1). For proving global convergence, we do not use the Linear Independence Constraint Qualification (LICQ) at all. On one hand, we characterize the situations in which infeasible limit points might exist using weaker assumptions than the LICQ. On the other hand, the fact that feasible limit points are KKT points will follow using the Constant Positive Linear Dependence (CPLD) condition [29], which has been recently proved to be a constraint qualification [2] and is far more general than the LICQ and other popular constraint qualifications. We use the LICQ only for proving boundedness of the penalty parameters.

This paper is organized as follows. The two main algorithms are introduced in Section 2. In Section 3 we characterize the infeasible points that could be limit points of the algorithms. In Section 4 it is proved that, if the CPLD constraint qualification holds at a feasible limit point, then this point must be KKT. In Section 5 we prove boundedness of the penalty parameters. In Section 6 we present numerical experiments. Conclusions and lines for future research are given in Section 7.

Notation.

Throughout this work, $[v]_i$ is the $i-$th component of the vector v. We also denote $v_i = [v]_i$ if this does not lead to confusion.

$I\!R_+$ denotes the set of nonnegative real numbers and $I\!R_{++}$ denotes the set of positive real numbers.

If J_1 and J_2 are subsets of $\{1, \ldots, n\}$, $B_{[J_1,J_2]}$ is the matrix formed by taking the rows and columns of B indexed by J_1 and J_2 respectively and $B_{[J_1]}$ is the matrix formed by taking the columns of B indexed by J_1. If $y \in I\!R^n$, $y_{[J_1]}$ is the vector formed by taking the components y_i such that $i \in J_1$.

2. Model algorithms

From here on we assume that F is continuous. Given $x \in \Omega$, $\lambda \in I\!R^m$, $\rho \in I\!R^m_{++}$ we define

$$G(x, \lambda, \rho) = F(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{i=1}^{m} \rho_i h_i(x) \nabla h_i(x).$$

If the KKT system is originated in a smooth minimization problem, the mapping F is the gradient of some $f : I\!R^n \to I\!R$. In this case we define, for $\rho \in I\!R^m_{++}$,

$$L(x, \lambda, \rho) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \frac{1}{2} \sum_{i=1}^{m} \rho_i h_i(x)^2.$$

This is the definition of the Augmented Lagrangian used in [11]. In this case we have that $\nabla L = G$. The function L is the Augmented Lagrangian associated with the problem (2).

The mapping G will be used to define one-parameter and many-parameters Augmented Lagrangian algorithms for solving the general KKT problem (1). These two algorithms (A1 and A2) are described below. They are presented as two instances of the general Algorithm A.

Algorithm A.

Let $x_0 \in \Omega$, $\tau \in [0, 1)$, $\gamma > 1$, $-\infty < \bar{\lambda}_{min} \leq 0 \leq \bar{\lambda}_{max} < \infty$, $\rho_1 \in I\!R^m_{++}$ (in Algorithm A1 $[\rho_1]_i = \|\rho_1\|_\infty$ for all $i = 1, \ldots, m$), $\bar{\lambda}_1 \in [\bar{\lambda}_{min}, \bar{\lambda}_{max}]^m$. Let $\{\varepsilon_k\}_{k \in I\!N} \subseteq I\!R_{++}$ be a sequence that converges to zero.

Step 1. *Initialization*

Set $k \leftarrow 1$.

Step 2. *Solving the subproblem*

Compute $x_k \in \Omega$ such that

$$\|\mathcal{P}_\Omega[x_k - G(x_k, \bar{\lambda}_k, \rho_k)] - x_k\|_\infty \leq \varepsilon_k. \quad (3)$$

Step 3. *Estimate multipliers*

Define, for all $i = 1, \ldots, m$,

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k). \quad (4)$$

If $h(x_k) = 0$ and $\mathcal{P}_\Omega[x_k - G(x_k, \bar{\lambda}_k, \rho_k)] - x_k = 0$ terminate the execution of the algorithm. (In this case, x_k is a KKT point and λ_{k+1} is the associated vector of Lagrange multipliers.)

Compute

$$\bar{\lambda}_{k+1} \in [\bar{\lambda}_{min}, \bar{\lambda}_{max}]^m. \quad (5)$$

Step 4. *Update the penalty parameters*

Define $\Gamma_k = \{i \in \{1, \ldots, m\} \mid |h_i(x_k)| > \tau \|h(x_{k-1})\|_\infty\}$. If $\Gamma_k = \emptyset$, define $\rho_{k+1} = \rho_k$. Else,

– In Algorithm A1, define $\rho_{k+1} = \gamma \rho_k$.

– In Algorithm A2, define $[\rho_{k+1}]_i = \gamma[\rho_k]_i$ if $i \in \Gamma_k$ and $[\rho_{k+1}]_i = [\rho_k]_i$ if $i \notin \Gamma_k$.

Step 5. *Begin a new iteration*
Set $k \leftarrow k+1$. Go to Step 2.

Remark 1. (i) Algorithm A2 only differs from Algorithm A1 in the way in which penalty parameters are updated. In Algorithm A2, as in [11], more than one penalty parameter is used per iteration. In the case in which Algorithm A2 updates at least one penalty parameter, Algorithm A1 updates its unique penalty parameter. In such a situation, other penalty parameters may remain unchanged in Algorithm A2. Therefore, the penalty parameters in Algorithm A2 tend to be smaller than the penalty parameter in Algorithm A1.

(ii) The global convergence results to be presented in the following sections are independent of the choice of $\bar{\lambda}_{k+1}$ in (5). Whenever possible, we will choose $\bar{\lambda}_{k+1} = \lambda_{k+1}$ but, as a matter of fact, the definition (4) is not used at all in the forthcoming Sections 3 and 4. If one chooses $\bar{\lambda}_{k+1} = 0$ for all k, Algorithms A1 and A2 turn out to be External Penalty methods.

(iii) The Augmented Lagrangian algorithms are based on the resolution of the inner problems (3). In the minimization case ($F = \nabla f$) the most reasonable way for obtaining these conditions is to solve (approximately) the minimization problem

$$\text{Minimize } L(x, \bar{\lambda}_k, \rho_k) \text{ subject to } x \in \Omega. \tag{6}$$

This is a box-constrained minimization problem. Since Ω is compact, minimizers exist and stationary points can be obtained up to any arbitrary precision using reasonable algorithms. Sufficient conditions under which points that satisfy (3) exist and can be obtained by available algorithms in more general problems have been analyzed in many recent papers. See [19–22,26].

3. Convergence to feasible points

At a KKT point we have that $h(x) = 0$ and $x \in \Omega$. Points that satisfy these two conditions are called *feasible*. It would be nice to have algorithms that find feasible points in every situation, but this is impossible. (In an extreme case, feasible points might not exist at all.) Therefore, it is important to study the behavior of algorithms with respect to infeasibility.

In this section we show that Algorithm A1 always converges to stationary points of the problem of minimizing $\|h(x)\|_2^2$ subject to $\ell \le x \le u$. In the case of Algorithm A2 we will show that the possible limit points must be solutions of a weighted least-squares problem involving the constraints.

In the proof of both theorems we will use the following well known property:

$$\|\mathcal{P}_\Omega(u + tv) - u\|_2 \le \|\mathcal{P}_\Omega(u + v) - u\|_2 \;\; \forall\, u \in \Omega, v \in I\!R^n, t \in [0, 1]. \tag{7}$$

Theorem 1. *Assume that the sequence $\{x_k\}$ is generated by Algorithm A1 and that x_* is a limit point. Then, x_* is a stationary point of the problem*

$$\begin{array}{c} \textit{Minimize } \|h(x)\|_2^2 \\ \textit{subject to } x \in \Omega. \end{array} \tag{8}$$

Proof. Let $K \subseteq I\!N$ be such that $\lim_{k\in K} x_k = x_*$. Let us denote

$$\hat{\rho}_k = \|\rho_k\|_\infty \text{ for all } k \in I\!N.$$

Clearly, $\hat{\rho}_k = [\rho_k]_i$ for all $i = 1, \ldots, m$.

By (3) and the equivalence of norms in $I\!R^n$, we have that

$$\lim_{k\to\infty} \|\mathcal{P}_\Omega[x_k - F(x_k) - \sum_{i=1}^m ([\bar{\lambda}_k]_i + \hat{\rho}_k h_i(x_k))\nabla h_i(x_k)] - x_k\|_2 = 0. \tag{9}$$

By Step 4 of Algorithm A, if $\{\rho_k\}_{k\in K}$ is bounded we have that $h(x_*) = 0$, so x_* is a stationary point of (8).

Assume that $\{\rho_k\}_{k\in K}$ is unbounded. Since $\{\hat{\rho}_k\}$ is nondecreasing, we have that

$$\lim_{k\to\infty} \hat{\rho}_k = \infty. \tag{10}$$

Then, $\hat{\rho}_k > 1$ for $k \in K$ large enough. So, using (7) with

$$u = x_k,\ v = -F(x_k) - \sum_{i=1}^m ([\bar{\lambda}_k]_i + \hat{\rho}_k h_i(x_k))\nabla h_i(x_k),\ t = 1/\hat{\rho}_k,$$

we have, by (9), that

$$\lim_{k\to\infty} \left\| \mathcal{P}_\Omega \left[x_k - \frac{F(x_k)}{\hat{\rho}_k} - \sum_{i=1}^m \left(\frac{[\bar{\lambda}_k]_i}{\hat{\rho}_k} + h_i(x_k) \right) \nabla h_i(x_k) \right] - x_k \right\|_2 = 0. \tag{11}$$

By (5), (10), (11) and the continuity of F we obtain:

$$\|\mathcal{P}_\Omega[x_* - \sum_{i=1}^m h_i(x_*)\nabla h_i(x_*)] - x_*\|_2 = 0.$$

This means that x_* is a stationary point of (8), as we wanted to prove. □

We say that an infeasible point $x_* \in \Omega$ is *degenerate* if there exists $w \in I\!R^m_+$ such that x_* is a stationary point of the weighted least-squares problem

$$\begin{array}{c} \text{Minimize } \sum_{i=1}^m w_i h_i(x)^2 \\ \text{subject to } x \in \Omega, \end{array} \tag{12}$$

and

$$w_i > 0 \text{ for some } i \text{ such that } h_i(x_*) \neq 0. \tag{13}$$

Theorem 2. *Let $\{x_k\}$ be a sequence generated by Algorithm A2. Then, at least one of the following possibilities hold:*

1. *The sequence admits a feasible limit point.*
2. *The sequence admits an infeasible degenerate limit point.*

Proof. Assume that all the limit points of the sequence $\{x_k\}$ are infeasible. Therefore, there exists $\varepsilon > 0$ such that

$$\|h(x_k)\|_\infty \geq \varepsilon \tag{14}$$

for all $k \in I\!N$. This implies that

$$\lim_{k\to\infty} \|\rho_k\|_\infty = \infty. \tag{15}$$

Let K be an infinite subset of indices such that

$$\|\rho_k\|_\infty > \|\rho_{k-1}\|_\infty \;\; \forall\, k \in K. \tag{16}$$

Since $\{1,\ldots,m\}$ is a finite set, there exists K_1, an infinite subset of K, and $j \in \{1,\ldots,m\}$ such that

$$\|\rho_k\|_\infty = [\rho_k]_j \;\; \forall\, k \in K_1. \tag{17}$$

Then, by (16) and Step 4 of Algorithm A2,

$$[\rho_k]_j = \gamma[\rho_{k-1}]_j \;\; \forall\, k \in K_1. \tag{18}$$

By the definition of the algorithm, we have that, for all $k \in K_1$,

$$|h_j(x_{k-1})| > \tau\|h(x_{k-2})\|_\infty.$$

So, by (14),

$$|h_j(x_{k-1})| > \tau\varepsilon \;\; \forall \;\; k \in K_1. \tag{19}$$

Moreover, by (16), (17) and (18), we have:

$$[\rho_{k-1}]_j \geq \frac{\|\rho_{k-1}\|_\infty}{\gamma} \;\; \forall \;\; k \in K_1. \tag{20}$$

Let K_2 be an infinite subset of indices of $\{k-1\}_{k\in K_1}$ such that

$$\lim_{k\in K_2} x_k = x_*.$$

By (19) we have that

$$h_j(x_*) \neq 0. \tag{21}$$

By (3) and the equivalence of norms in $I\!R^n$, we have:

$$\lim_{k\to\infty} \|\mathcal{P}_\Omega[x_k - F(x_k) - \sum_{i=1}^m ([\bar\lambda_k]_i + [\rho_k]_i h_i(x_k))\nabla h_i(x_k)] - x_k\|_2 = 0. \tag{22}$$

By (15), $\|\rho_k\|_\infty > 1$ for $k \in K_2$ large enough. So, using (7) with

$$u = x_k,\; v = -F(x_k) - \sum_{i=1}^m ([\bar\lambda_k]_i + [\rho_k]_i h_i(x_k))\nabla h_i(x_k),\; t = 1/\|\rho_k\|_\infty,$$

we have, by (22), that

$$\lim_{k\in K_2}\left\|\mathcal{P}_\Omega\left[x_k-\frac{F(x_k)}{\|\rho_k\|_\infty}-\sum_{i=1}^m\left(\frac{[\bar\lambda_k]_i}{\|\rho_k\|_\infty}+\frac{[\rho_k]_i}{\|\rho_k\|_\infty}h_i(x_k)\right)\nabla h_i(x_k)\right]-x_k\right\|_2=0. \tag{23}$$

But

$$\frac{[\rho_k]_i}{\|\rho_k\|_\infty}\le 1 \quad \forall i=1,\ldots,m.$$

Therefore, there exist $K_3\subseteq K_2$ and $w\in \mathbb{R}^m_+$ such that

$$\lim_{k\in K_3}\frac{[\rho_k]_i}{\|\rho_k\|_\infty}=w_i \quad \forall i=1,\ldots,m.$$

Moreover, by (20),

$$w_j>0. \tag{24}$$

Since $\{\bar\lambda_k\}_{k\in\mathbb{N}}$ is bounded, taking limits for $k\in K_3$ in (23), by (15) and the continuity of F, we get:

$$\|\mathcal{P}_\Omega[x_*-\sum_{i=1}^m w_i h_i(x_*)\nabla h_i(x_*)]-x_*\|_2=0.$$

So, x_* is a stationary point of (12). By (21) and (24), the condition (13) also takes place. Therefore, x_* is a degenerate infeasible point. □

Remark 2. Clearly, any infeasible stationary point of (8) must be degenerate. Moreover, if x is infeasible and degenerate, by (13) and the KKT conditions of (12), the gradients of the equality constraints and the active bound constraints are linearly dependent. The reciprocal is not true. In fact, consider the set of constraints

$$h(x)\equiv x=0\in\mathbb{R}^1, \ \ -1\le x\le 1.$$

At the points $z=-1$ and $z=1$ the gradients of equality constraints and active bound constraints are linearly dependent but these points are not degenerate. In [12] it is assumed that, at all the limit points of the sequence generated by the Augmented Lagrangian algorithm, the gradients of equality constraints and active bound constraints are linearly independent (Assumption AS3 of [12]). Under this assumption it is proved that all the limit points are feasible. The feasibility of all the limit points generated by Algorithm A1 also holds from our Theorem 1, if we assume that limit points are nondegenerate. For Algorithm A2, the corresponding result comes from Theorem 2: under the weak assumption of nondegeneracy we only obtain the weaker result that there exists at least one feasible limit point.

4. Convergence to optimal points

In this section we investigate under which conditions a feasible limit point of a sequence generated by the Augmented Lagrangian algorithms is a KKT point. The main result is that a feasible limit point is KKT if it satisfies the Constant Positive Linear Dependence condition (CPLD). The CPLD condition was introduced by Qi and Wei in [29]. More recently [2], it was proved that this condition is a constraint qualification. Assume that the constraints of a problem are $h(x) = 0, g(x) \leq 0$, where $h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$ and that $\bar{x}$ is a feasible point such that $g_i(\bar{x}) = 0$ for all $i \in I$, $g_i(\bar{x}) < 0$ for all $i \notin I$. We say that $\bar{x}$ satisfies the CPLD condition if the existence of $J_1 \subseteq \{1, \ldots, m\}$, $J_2 \subseteq I$, $\{\lambda_i\}_{i \in J_1} \subseteq \mathbb{R}$, $\{\mu_i\}_{i \in J_2} \subseteq \mathbb{R}_+$ such that

$$\sum_{i \in J_1} \lambda_i \nabla h_i(\bar{x}) + \sum_{i \in J_2} \mu_i \nabla g_i(\bar{x}) = 0$$

and

$$\sum_{i \in J_1} |\lambda_i| + \sum_{i \in J_2} \mu_i > 0$$

implies that the gradients $\{\nabla h_i(x)\}_{i \in J_1} \cup \{\nabla g_i(x)\}_{i \in J_2}$ are linearly dependent for all x in a neighborhood of $\bar{x}$.

Clearly, if the Mangasarian-Fromovitz constraint qualification [28,33] holds, the CPLD condition holds as well, but the reciprocal is not true.

The AS3 condition of [12], when applied only to feasible points, is the classical LICQ assumption (linear independence of the gradients of active constraints). Of course, at points that do not satisfy Mangasarian-Fromovitz the gradients of active constraints are linearly dependent. Therefore, convergence results based on the CPLD condition are stronger than convergence results that assume the classical LICQ.

In Theorem 3 we prove that, if a feasible limit point of an algorithmic sequence satisfies the CPLD constraint qualification, then this point must be KKT.

Theorem 3. *Assume that $\{x_k\}$ is a sequence generated by Algorithm A and that x_* is a feasible limit point that satisfies the CPLD constraint qualification. Then, x_* is a KKT point.*

Proof. Let us write

$$G^k = G(x_k, \bar{\lambda}_k, \rho_k).$$

Define, for all $k \in \mathbb{N}$,

$$v_k = \mathcal{P}_{\Omega}(x_k - G^k).$$

Therefore, $v_k \in \mathbb{R}^n$ solves

$$\text{Minimize } \|v - (x_k - G^k)\|_2^2$$

$$\text{subject to } \ell \leq v \leq u.$$

By the KKT conditions of this problem, there exist $\mu_k^u \in I\!R_+^n$, $\mu_k^\ell \in I\!R_+^n$ such that, for all $k \in I\!N$,

$$v_k - x_k + G^k + \sum_{i=1}^n [\mu_k^u]_i e_i - \sum_{i=1}^n [\mu_k^\ell]_i e_i = 0 \tag{25}$$

and

$$[\mu_k^u]_i (u_i - [x_k]_i) = [\mu_k^\ell]_i (\ell_i - [x_k]_i) = 0 \;\; \forall \;\; i = 1, \dots, n. \tag{26}$$

By (3),

$$\lim_{k\to\infty} (v_k - x_k) = 0.$$

Then, by (25),

$$\lim_{k\to\infty} \left(G^k + \sum_{i=1}^n [\mu_k^u]_i e_i - \sum_{i=1}^n [\mu_k^\ell]_i e_i \right) = 0.$$

So, defining λ_{k+1} as in (4),

$$\lim_{k\to\infty} \left(F(x_k) + \nabla h(x_k) \lambda_{k+1} + \sum_{i=1}^n [\mu_k^u]_i e_i - \sum_{i=1}^n [\mu_k^\ell]_i e_i \right) = 0. \tag{27}$$

Assume now that K is an infinite subset of $I\!N$ such that

$$\lim_{k\in K} x_k = x_*.$$

Since x_* is feasible, by the continuity of h we have that

$$\lim_{k\in K} \|h(x_k)\| = 0. \tag{28}$$

By (26), (27) and (28), since $\ell \le x_k \le u$ for all k, the subsequence $\{x_k\}_{k\in K}$ is an approximate KKT sequence in the sense of [29] (Definition 2.5). (In [29] one has $F = \nabla f$ but the extension of the definition to a general F is straightfoward.) Therefore, as in the proof of Theorem 2.7 of [29], we obtain that x_* is a KKT point. $\square$

5. Boundedness of the penalty parameters

In this section we assume that the sequence $\{x_k\}$, generated by Algorithm A1 or by Algorithm A2, converges to a KKT point $x_* \in \Omega$. To simplify the arguments, as in [12], we assume without loss of generality that $[x_*]_i < u_i$ for all $i = 1, \dots, n$ and that $\ell_i = 0$ for all $i = 1, \dots, n$. The Lagrange multipliers associated with x_* will be denoted $\lambda_* \in I\!R^m$. This vector will be unique by future assumptions. We will assume that $F'(x)$ and $\nabla^2 h_i(x)$ exist and are Lipschitz-continuous for all $x \in \Omega$.

Many definitions and proofs of this section invoke arguments used in [12]. We will mention all the cases in which this occurs.

Assumption NS *Define*

$$J_1 = \{\ i \in \{1,\dots,n\} \mid [F(x_*) + \sum_{j=1}^{m} [\lambda_*]_j \nabla h_j(x_*)]_i = 0 \text{ and } [x_*]_i > 0\}$$

$$J_2 = \{\ i \in \{1,\dots,n\} \mid [F(x_*) + \sum_{j=1}^{m} [\lambda_*]_j \nabla h_j(x_*)]_i = 0 \text{ and } [x_*]_i = 0\}.$$

Then, the matrix

$$\begin{pmatrix} [F'(x_*) + \sum_{j=1}^{m} [\lambda_*]_j \nabla^2 h_j(x_*)]_{[J,J]} & (h'(x_*)_{[J]})^T \\ h'(x_*)_{[J]} & 0 \end{pmatrix}$$

is nonsingular for all $J = J_1 \cup K$ such that $K \subseteq J_2$.

Assumption NS, which corresponds to Assumption AS5 of [12], will be supposed to be true all along this section. Clearly, the fulfillment of NS implies that the gradients of active constraints at x_* are linearly independent.

We will also assume that the computation of $\bar\lambda_k$ at Step 3 of both algorithms is:

$$[\bar\lambda_k]_i = \max\{\bar\lambda_{min}, \min\{\bar\lambda_{max}, [\lambda_k]_i\}\} \tag{29}$$

for all $i = 1,\dots,m$.

Finally, we will assume that the true Lagrange multipliers $[\lambda_*]_i$ satisfy

$$\bar\lambda_{min} < [\lambda_*]_i < \bar\lambda_{max} \ \ \forall\ \ i = 1,\dots,m. \tag{30}$$

In the case of Algorithm A1 it will be useful to denote, as before, $\hat\rho_k = \|\rho_k\|_\infty$.

Lemma 1. *Assume that the sequence $\{x_k\}$ is generated by Algorithm A1. Then, there exist $k_0 \in I\!N$, $\bar\rho, a_1, a_2, a_3, a_4, a_5, a_6 > 0$ such that, for all $k \geq k_0$,*

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq a_1\varepsilon_k + a_2\|x_k - x_*\|_\infty, \tag{31}$$

and, if $\hat\rho_{k_0} \geq \bar\rho$:

$$\|x_k - x_*\|_\infty \leq a_3\varepsilon_k + a_4 \frac{\|\bar\lambda_k - \lambda_*\|_\infty}{\hat\rho_k},$$

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq a_5\varepsilon_k + a_6 \frac{\|\bar\lambda_k - \lambda_*\|_\infty}{\hat\rho_k} \tag{32}$$

and

$$\|h(x_k)\|_\infty \leq a_5\varepsilon_k \frac{1}{\hat\rho_k} + \left(1 + \frac{a_6}{\hat\rho_k}\right) \frac{\|\bar\lambda_k - \lambda_*\|_\infty}{\hat\rho_k}. \tag{33}$$

Proof. The proof is identical to the ones of Lemmas 4.3 and 5.1 of [12], using NS, replacing μ_k by $1/\hat\rho_k$ and using the equivalence of norms in $I\!R^n$. □

Lemma 2 and Theorem 4 below complete the penalty boundedness proof for Algorithm A1. These proofs are specific for the updating rule of this algorithm, since the proofs of [12] do not apply to our case.

Lemma 2. *Assume that the sequence $\{x_k\}$ is generated by Algorithm A1. Then, there exists $k_0 \in I\!N$ such that for all $k \geq k_0$,*

$$\bar{\lambda}_k = \lambda_k.$$

Proof. By (31) there exists $k_1 \in I\!N$ such that

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq a_1 \varepsilon_k + a_2 \|x_k - x_*\|_\infty \text{ for all } k \geq k_1. \tag{34}$$

Define $\epsilon = \frac{1}{2}\min_i\{[\lambda_*]_i - \bar{\lambda}_{min}, \bar{\lambda}_{max} - [\lambda_*]_i\} > 0$. Since $\|x_k - x_*\|_\infty \to 0$ and $\varepsilon_k \to 0$, by (34) we obtain that

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq \epsilon \text{ for } k \text{ large enough.}$$

By (29), (30) and the definition of ϵ we obtain the desired result. □

In the following theorem we prove that, if a suitable adaptive choice of the convergence criterion of the subproblems is used, Lagrange multipliers are bounded in Algorithm A1.

Theorem 4. *Assume that the sequence $\{x_k\}$ is generated by Algorithm A1 and that ε_k is such that*

$$\varepsilon_k = min\{\varepsilon'_k, \|h(x_k)\|_\infty\} \tag{35}$$

where $\{\varepsilon'_k\}$ is a decreasing sequence that tends to zero. Then, the sequence of penalty parameters $\{\rho_k\}$ is bounded.

Proof. Let k_0 be as in Lemma 2. Then, for all $k \geq k_0$, we have that $\bar{\lambda}_k = \lambda_k$.

Assume that $\hat{\rho}_k \to \infty$. By (33) and (35) there exists $k_1 \geq k_0$ such that $a_5/\hat{\rho}_k < 1$ and

$$\|h(x_k)\|_\infty \leq \left(1 + \frac{a_6}{\hat{\rho}_k}\right)\left(\frac{1}{1 - \frac{a_5}{\hat{\rho}_k}}\right)\frac{\|\lambda_k - \lambda_*\|_\infty}{\hat{\rho}_k} \tag{36}$$

for all $k \geq k_1$. Since $\lambda_k = \lambda_{k-1} + \hat{\rho}_{k-1} h(x_{k-1})$ we get

$$\|h(x_{k-1})\|_\infty = \frac{\|\lambda_k - \lambda_{k-1}\|_\infty}{\hat{\rho}_{k-1}} \geq \frac{\|\lambda_{k-1} - \lambda_*\|_\infty}{\hat{\rho}_{k-1}} - \frac{\|\lambda_k - \lambda_*\|_\infty}{\hat{\rho}_{k-1}}.$$

Then, by (32) and (35), if k is large enough we have that:

$$\|\lambda_k - \lambda_*\|_\infty \leq a_5 \varepsilon_{k-1} + a_6 \frac{\|\lambda_{k-1} - \lambda_*\|_\infty}{\hat{\rho}_{k-1}}$$

$$\leq a_5 \|h(x_{k-1})\|_\infty + a_6 \left(\|h(x_{k-1})\|_\infty + \frac{\|\lambda_k - \lambda_*\|_\infty}{\hat{\rho}_{k-1}}\right).$$

Therefore, if k is large enough (so that $a_6/\hat{\rho}_{k-1} < 1$) we get:

$$\|\lambda_k - \lambda_*\|_\infty \leq \frac{1}{\frac{1}{a_6} - \frac{1}{\hat{\rho}_{k-1}}} \left(1 + \frac{a_5}{a_6}\right) \|h(x_{k-1})\|_\infty. \tag{37}$$

Combining (37) and (36) we obtain:

$$\|h(x_k)\|_\infty \leq \frac{\pi_k}{\hat{\rho}_k} \|h(x_{k-1})\|_\infty$$

with

$$\pi_k = \left(1 + \frac{a_6}{\hat{\rho}_k}\right) \left(\frac{1}{1 - \frac{a_5}{\hat{\rho}_k}}\right) \frac{1}{\frac{1}{a_6} - \frac{1}{\hat{\rho}_{k-1}}} \left(1 + \frac{a_5}{a_6}\right).$$

Since $\lim_{k\to\infty} \frac{\pi_k}{\hat{\rho}_k} = 0$, there exists$k_2 \geq k_1$ such that $\frac{\pi_k}{\hat{\rho}_k} < \tau$ and $\rho_{k+1} = \rho_k$ for all $k \geq k_2$. This is a contradiction. □

Remark 3. The choice of the tolerance ε_k in (35) deserves some explanation. In this case ε'_k is given and tends to zero. Therefore, by (35), the sequence ε_k tends to zero as required by Algorithm A1. However, when the rule (35) is adopted, ε_k is not given before the resolution of each subproblem. In other words, the inner algorithm used to solve each subproblem stops (returning the approximate solution x_k) only when the condition

$$\|\mathcal{P}_\Omega[x_k - G(x_k, \bar{\lambda}_k, \rho_k)] - x_k\|_\infty \leq \min\{\varepsilon'_k, \|h(x_k)\|_\infty\}$$

is fulfilled.

In the rest of this section we consider the Augmented Lagrangian method with several penalty parameters defined by Algorithm A2. Several definitions and proofs will be adapted from the ones given in [11,12]. This is the case of the definitions that precede Lemma 3.

From now on, the sequence $\{x_k\}$ is generated by Algorithm A2. Define

$$I_\infty = \{i \in \{1, \ldots, m\} \mid [\rho_k]_i \to \infty\}, \quad I_a = \{i \in \{1, \ldots, m\} \mid [\rho_k]_i \text{ is bounded }\},$$

$$\tilde{\rho}_k = \min_{i \in I_\infty} \{[\rho_k]_i\}, \quad \eta_k = \sum_{i \in I_a} |h_i(x_k)|.$$

Given the iterate $x_k \in \Omega$ and $i \in \{1, \ldots, n\}$, we have two possibilities for each component $[x_k]_i$:

(i) $0 \leq [x_k]_i \leq [G(x_k, \bar{\lambda}_k, \rho_k)]_i$, or
(ii) $[G(x_k, \bar{\lambda}_k, \rho_k)]_i < [x_k]_i$.

A variable $[x_k]_i$ is said to be *dominated* at the point x_k if $[x_k]_i$ satisfies (i). If $[x_k]_i$ satisfies (ii) the variable $[x_k]_i$ is said to be *floating*.

If the variable $[x_k]_i$ is dominated, we have that

$$[\mathcal{P}_\Omega[x_k - G(x_k, \bar{\lambda}_k, \rho_k)] - x_k]_i = -[x_k]_i. \tag{38}$$

On the other hand, if the variable $[x_k]_i$ is floating, we have:

$$[\mathcal{P}_\Omega[x_k - G(x_k, \bar{\lambda}_k, \rho_k)] - x_k]_i = -[G(x_k, \bar{\lambda}_k, \rho_k)]_i. \tag{39}$$

Let us define:

$$I_1 = \{i \in \{1, \dots, n\} \mid [x_k]_i \text{ is floating for all } k \text{ large enough and } [x_*]_i > 0\}, \tag{40}$$

$$I_2 = \{i \in \{1, \dots, n\} \mid [x_k]_i \text{ is dominated for all } k \text{ large enough}\}. \tag{41}$$

The following result corresponds to Lemmas 4.3 and 5.1 of [12], adapted for several penalty parameters.

Lemma 3. *Assume that the sequence $\{x_k\}$ is computed by Algorithm A2. There exists $k_0 \in I\!N$ and positive constants b_1, b_2, $\overline{\rho}$, α_1, α_2, α_3, α_4, α_5 such that, for all $k \geq k_0$,*

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq b_1 \varepsilon_k + b_2 \|x_k - x_*\|_\infty, \tag{42}$$

and, if $[\rho_{k_0}]_i \geq \overline{\rho}$ for all $i \in I_\infty$,

$$\|x_k - x_*\|_\infty \leq \alpha_1 \varepsilon_k + \alpha_2 \eta_k + \alpha_3 \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} \tag{43}$$

and

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq \alpha_4 \varepsilon_k + \alpha_5 \|h(x_k)\|_\infty. \tag{44}$$

Proof. The proof of (42) is identical to the one of (31). The arguments used below to prove (46)–(51) may be found in [12] (proof of Lemma 5.1, pages 558–561).

Let k be large enough, so that the sets I_1, I_2 defined in (40) and (41) are well determined. Let I_3 be the set of the remaining indices (dominated or floating). For all k large enough (say, $k \geq k_0$), define $I_4(k), I_5(k)$ such that

(i) $I_4(k) \cap I_5(k) = \emptyset$ and $I_4(k) \cup I_5(k) = I_3$;
(ii) The indices $i \in I_4(k)$ correspond to floating variables;
(iii) The indices $i \in I_5(k)$ correspond to dominated variables.

Let $\mathcal{K} = \bigcup_{k \geq k_0} \{(I_4(k), I_5(k)\}$. Clearly, $\mathcal{K}$ is a finite set. For all $(I_4, I_5) \in \mathcal{K}$, consider the set of indices $\mathcal{I}(I_4, I_5)$ such that $I_4(k) = I_4, I_5(k) = I_5$ for all $k \in \mathcal{I}(I_4, I_5)$. Obviously,

$$\{k \in I\!N \mid k \geq k_0\} = \bigcup_{(I_4, I_5) \in \mathcal{K}} \mathcal{I}(I_4, I_5), \tag{45}$$

where the union on the right-hand side of (45) involves a finite number of sets. Therefore, it is sufficient to prove the lemma for each set $K = \mathcal{I}(I_4, I_5)$. In this way, the constants $\alpha_1, \dots, \alpha_5$ will depend on K. (Say, $\alpha_i = \alpha_i(K), i = 1, \dots, 5$.) At the end we can take $\alpha_i = \max\{\alpha_i(K)\}$ and (43)–(44) will be true.

So, fixing $k \in K = \mathcal{I}(I_4, I_5)$ we define:

$$I_F = I_1 \cup I_4 \quad \text{and} \quad I_D = I_2 \cup I_5.$$

Thus, the variables in I_F are floating whereas the variables in I_D are dominated.

Define $T(x,\lambda) = F(x) + \nabla h(x)\lambda$. By the definition of G and (4) we have that $T(x_k, \lambda_{k+1}) = G(x_k, \bar{\lambda}_k, \rho_k)$. Define $H_l(x,\lambda) = T_x^{'}(x,\lambda) = F^{'}(x) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x)$, where the derivatives are taken with respect to x. Using Taylor's formula on each term of the expression $T(x_k, \lambda_{k+1})$ in a neighborhood of (x_*, λ_*) and on $h(x_k)$ in a neighborhood of x_* (see details in [12], page 559), we obtain:

$$\begin{pmatrix} H_l(x_*,\lambda_*) & h^{'}(x_*)^T \\ h^{'}(x_*) & 0 \end{pmatrix} \begin{pmatrix} x_k - x_* \\ \lambda_{k+1} - \lambda_* \end{pmatrix} = \begin{pmatrix} G(x_k, \bar{\lambda}_k, \rho_k) - T(x_*, \lambda_*) \\ h(x_k) \end{pmatrix} - \begin{pmatrix} r_1 + r_2 \\ r_3 \end{pmatrix}, \tag{46}$$

where

$$r_1(x_k, x_*, \lambda_{k+1}) = \int_0^1 [H_l(x_k + s(x_* - x_k), \lambda_{k+1}) - H_l(x_*, \lambda_{k+1})](x_k - x_*)ds,$$

$$r_2(x_k, x_*, \lambda_{k+1}, \lambda_*) = \sum_{j=1}^m ([\lambda_{k+1}]_j - [\lambda_*]_j)\nabla^2 h_j(x_*)(x_k - x_*)$$

and

$$[r_3(x_k, x_*)]_i = \int_0^1 s \int_0^1 (x_k - x_*)^T \nabla^2 h_i(x_* + ts(x_k - x_*))(x_k - x_*)dtds.$$

By (42), $\lim_{k\to\infty} \lambda_{k+1} = \lambda_*$. So, by the Lipschitz-continuity of $F^{'}(x)$ and $\nabla^2 h_i(x)$ in a neighborhood of x_*, we get:

$$\begin{gathered} \|r_1(x_k, x_*, \lambda_{k+1})\|_2 \le a_7 \|x_k - x_*\|_2^2, \\ \|r_2(x_k, x_*, \lambda_{k+1}, \lambda_*)\|_2 \le a_8 \|x_k - x_*\|_2 \|\lambda_{k+1} - \lambda_*\|_2 \text{ and} \\ \|r_3(x_k, x_*)\|_2 \le a_9 \|x_k - x_*\|_2^2 \end{gathered} \tag{47}$$

for positive constants a_7, a_8 and a_9.

By (42) we have that λ_{k+1} tends to λ_* and, so, $G(x_k, \bar{\lambda}_k, \rho_k)$ tends to $T(x_*, \lambda_*)$. Therefore, as in Lemma 2.1 of [12], we obtain that $[x_*]_i = 0$ when $i \in I_D$ and $[T(x_*, \lambda_*)]_i = 0$ when $i \in I_F$. Thus, equation (46) may be written as

$$\begin{pmatrix} H_l(x_*,\lambda_*)_{[I_F,I_F]} & H_l(x_*,\lambda_*)_{[I_F,I_D]} & h^{'}(x_*)^T_{[I_F]} \\ H_l(x_*,\lambda_*)_{[I_D,I_F]} & H_l(x_*,\lambda_*)_{[I_D,I_D]} & h^{'}(x_*)^T_{[I_D]} \\ h^{'}(x_*)_{[I_F]} & h^{'}(x_*)_{[I_D]} & 0 \end{pmatrix} \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ (x_k)_{[I_D]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix}$$

$$= \begin{pmatrix} G(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]} \\ (G(x_k, \bar{\lambda}_k, \rho_k) - T(x_*, \lambda_*)_{[I_D]} \\ h(x_k) \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_{[I_F]} \\ (r_1 + r_2)_{[I_D]} \\ r_3 \end{pmatrix}.$$

Therefore, after some manipulation, we obtain:

$$\begin{pmatrix} H_l(x_*, \lambda_*)_{[I_F,I_F]} & h'(x_*)^T_{[I_F]} \\ h'(x_*)_{[I_F]} & 0 \end{pmatrix} \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix}$$
$$= \begin{pmatrix} G(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]} - H_l(x_*, \lambda_*)_{[I_F,I_D]}(x_k)_{[I_D]} \\ h(x_k) - h'(x_*)_{[I_D]}(x_k)_{[I_D]} \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_{[I_F]} \\ r_3 \end{pmatrix}. \tag{48}$$

Since $[x_*]_i = 0$ for all $i \in I_D$ and, by (3) and (38), $\|(x_k)_{[I_D]}\|_2 \le \sqrt{|I_D|}\varepsilon_k$, we get:

$$\|x_k - x_*\|_2 \le \|(x_k - x_*)_{[I_F]}\|_2 + \sqrt{|I_D|}\varepsilon_k. \tag{49}$$

Define now $\Delta x_k = \|(x_k - x_*)_{[I_F]}\|_2$. Combining (42) and (49) we obtain:

$$\|\lambda_{k+1} - \lambda_*\|_2 \le a_{10}\varepsilon_k + a_{11}\Delta x_k, \tag{50}$$

with $a_{10} = \sqrt{m}(b_1 + \sqrt{|I_D|}b_2), a_{11} = \sqrt{m}b_2$. Moreover, by (47), (49) and (50),

$$\left\| \begin{pmatrix} (r_1 + r_2)_{[I_F]} \\ r_3 \end{pmatrix} \right\|_2 \le a_{12}(\Delta x_k)^2 + a_{13}\Delta x_k \varepsilon_k + a_{14}{\varepsilon_k}^2 \tag{51}$$

with $a_{12} = a_7 + a_9 + a_8 a_{11}$, $a_{13} = 2\sqrt{|I_D|}(a_7 + a_9) + a_8(\sqrt{|I_D|}a_{11} + a_{10})$ and $a_{14} = |I_D|(a_7 + a_9) + \sqrt{|I_D|}a_8 a_{10}$.

From here to the end of the proof, the arguments used are not present in [12].

Since $\|(x_k)_{[I_D]}\|_2 \le \sqrt{|I_D|}\varepsilon_k$ and, by (3) and (39), $\|G(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]}\|_2 \le \sqrt{|I_F|}\varepsilon_k$ we obtain:

$$\left\| \begin{pmatrix} G(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]} - H_l(x_*, \lambda_*)_{[I_F,I_D]}(x_k)_{[I_D]} \\ h(x_k) - h'(x_*)_{[I_D]}(x_k)_{[I_D]} \end{pmatrix} \right\|_2 \le a_{15}\varepsilon_k + \|h(x_k)\|_2,$$

with $a_{15} = \sqrt{n}\left(1 + \left\| \begin{pmatrix} H_l(x_*, \lambda_*)_{[I_F,I_D]} \\ h'(x_*)_{[I_D]} \end{pmatrix} \right\|_2\right)$.

By Assumption NS, the left-hand side matrix of (48) is nonsingular. Let M be the norm of its inverse. Multiplying both sides of the equation by this inverse and taking norms, we obtain:

$$\left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\|_2 \le M(a_{15}\varepsilon_k + \|h(x_k)\|_2 + a_{12}(\Delta x_k)^2 + a_{13}\Delta x_k\varepsilon_k + a_{14}{\varepsilon_k}^2). \tag{52}$$

By (4) and (42),

$$|h_i(x_k)| = \frac{|[\lambda_{k+1}]_i - [\bar{\lambda}_k]_i|}{[\rho_k]_i} \le \frac{|[\lambda_{k+1}]_i - [\lambda_*]_i| + |[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}$$
$$\le \frac{b_1\varepsilon_k + b_2\|x_k - x_*\|_2 + |[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}. \tag{53}$$

Now,

$$\|h(x_k)\|_2^2 = \sum_{i\in I_\infty} |h_i(x_k)|^2 + \sum_{i\in I_a} |h_i(x_k)|^2.$$

So, using that $\sum_{i=1}^n a_i^2 \le (\sum_{i=1}^n a_i)^2$ for $a_i \ge 0, i = 1,\ldots,n$, and the inequality (53) for all $i \in I_\infty$, we obtain:

$$\|h(x_k)\|_2 \le \eta_k + \sum_{i\in I_\infty} \frac{|[\bar\lambda_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + |I_\infty| \frac{(b_1\varepsilon_k + b_2\|x_k - x_*\|_2)}{\tilde\rho_k}. \tag{54}$$

By (49), combining (54) and (52), we get

$$\left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\|_2$$

$$\le M\left(a_{15}\varepsilon_k + \eta_k + \sum_{i\in I_\infty} \frac{|[\bar\lambda_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + a_{16}\frac{\varepsilon_k}{\tilde\rho_k} + b_2|I_\infty|\frac{\Delta x_k}{\tilde\rho_k} + a_{12}(\Delta x_k)^2 + a_{13}\Delta x_k\varepsilon_k + a_{14}{\varepsilon_k}^2\right), \tag{55}$$

where $a_{16} = |I_\infty|(b_1 + b_2\sqrt{|I_D|})$.

Now, if k is large enough,

$$\varepsilon_k \le \min\left\{1, \frac{1}{4Ma_{13}}\right\} \tag{56}$$

and

$$\Delta x_k \le \frac{1}{4Ma_{12}}. \tag{57}$$

By (55) and (56), we have that

$$\Delta x_k = \|(x_k - x_*)_{[I_F]}\|_2$$

$$\le M\left(a_{15}\varepsilon_k + \eta_k + \sum_{i\in I_\infty} \frac{|[\bar\lambda_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + a_{16}\frac{\varepsilon_k}{\tilde\rho_k} + b_2|I_\infty|\frac{\Delta x_k}{\tilde\rho_k} + a_{12}(\Delta x_k)^2 + \frac{\Delta x_k}{4M} + a_{14}\varepsilon_k\right). \tag{58}$$

Then, by (57) and (58),

$$\Delta x_k \le M\left(a_{15}\varepsilon_k + \eta_k + \sum_{i\in I_\infty} \frac{|[\bar\lambda_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + a_{16}\frac{\varepsilon_k}{\tilde\rho_k} + b_2|I_\infty|\frac{\Delta x_k}{\tilde\rho_k} + \frac{\Delta x_k}{4M} + \frac{\Delta x_k}{4M} + a_{14}\varepsilon_k\right). \tag{59}$$

Define $\bar\rho = \max\{1, 4|I_\infty|Mb_2\}$. If k is large enough, $[\rho_k]_i \ge \bar\rho$ for all $i \in I_\infty$. By (59) we get:

$$\Delta x_k \le 4M\left((a_{15} + a_{16} + a_{14})\varepsilon_k + \eta_k + \sum_{i\in I_\infty} \frac{|[\bar\lambda_k]_i - [\lambda_*]_i|}{[\rho_k]_i}\right). \tag{60}$$

So, by (49) and (60), we obtain:

$$\|x_k - x_*\|_\infty \le \|x_k - x_*\|_2 \le \alpha_1 \varepsilon_k + \alpha_2 \eta_k + \alpha_3 \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i},$$

where $\alpha_1 = 4M(a_{15} + a_{16} + a_{14}) + \sqrt{|I_D|}$ and $\alpha_2 = \alpha_3 = 4M$. This proves (43).

Let us now prove (44). Using (56) and (57) in the inequality (52) we obtain:

$$\Delta x_k = \|(x_k - x_*)_{[I_F]}\|_2 \le \frac{\Delta x_k}{2} + M(a_{17}\varepsilon_k + \|h(x_k)\|_2),$$

where $a_{17} = a_{14} + a_{15}$. Therefore,

$$\Delta x_k \le 2M(a_{17}\varepsilon_k + \|h(x_k)\|_2). \tag{61}$$

By (49) and (61) we obtain:

$$\|x_k - x_*\|_\infty \le \|x_k - x_*\|_2 \le \alpha_6 \varepsilon_k + \alpha_7 \|h(x_k)\|_\infty, \tag{62}$$

where $\alpha_6 = 2Ma_{17} + \sqrt{|I_D|}$ and $\alpha_7 = 2M\sqrt{m}$. Combining (62) and (42) we obtain the inequality

$$\|\lambda_{k+1} - \lambda_*\|_\infty \le \alpha_4 \varepsilon_k + \alpha_5 \|h(x_k)\|_\infty,$$

where $\alpha_4 = b_1 + b_2\alpha_6$ and $\alpha_5 = b_2\alpha_7$. Then, (44) is proved. □

Lemma 4. *Assume that the sequence $\{x_k\}$ is computed by Algorithm A2. Then, there exists $k_0 \in I\!N$ such that, for all $k \ge k_0$,*

$$\bar{\lambda}_k = \lambda_k.$$

Proof. By (42), the proof is the same than that of Lemma 2. □

Theorem 5 is the final result of this section. We will prove that, under a different adaptive choice of the stopping criterion used in the subproblems, the penalty parameters are bounded for Algorithm A2.

Theorem 5. *Assume that the sequence $\{x_k\}$ is computed by Algorithm A2 and that ε_k is such that*

$$\varepsilon_k = min\{\varepsilon_{k-1}, \|h(x_k)\|_\infty, \varepsilon'_k\} \tag{63}$$

where $\{\varepsilon'_k\}$ is a decreasing sequence that converges to zero. Then the sequence $\{\rho_k\}$ is bounded.

Proof. Suppose that $I_\infty \ne \emptyset$. Let $i_0 \in I_\infty$.

For all $i \in I_a$ there exists $k_1(i)$ such that for all $k \ge k_1(i)$, $[\rho_{k+1}]_i = [\rho_k]_i$. If k is large enough we have that, for all $i \in I_a$,

$$|h_i(x_k)| \le \tau \|h(x_{k-1})\|_\infty.$$

Then,

$$\eta_k = \sum_{i \in I_a} |h_i(x_k)| \leq |I_a| \tau \|h(x_{k-1})\|_\infty. \tag{64}$$

Let $k \geq \tilde{k} = \max_{i \in I_a} \{k_0, k_1(i)\}$, where k_0 is obtained as in Lemma 4. By (4),

$$|h_{i_0}(x_k)| = \frac{|[\lambda_{k+1}]_{i_0} - [\lambda_k]_{i_0}|}{[\rho_k]_{i_0}} \leq \frac{|[\lambda_{k+1}]_{i_0} - [\lambda_*]_{i_0}| + |[\lambda_k]_{i_0} - [\lambda_*]_{i_0}|}{[\rho_k]_{i_0}}.$$

So, by (42),

$$|h_{i_0}(x_k)| \leq \frac{b_1 \varepsilon_k + b_2 \|x_k - x_*\|_\infty + |[\lambda_k]_{i_0} - [\lambda_*]_{i_0}|}{[\rho_k]_{i_0}}.$$

Thus, by (43),

$$|h_{i_0}(x_k)| \leq \frac{1}{[\rho_k]_{i_0}} \left[(b_1 + b_2 \alpha_1) \varepsilon_k + b_2 \alpha_2 \eta_k + b_2 \alpha_3 \sum_{i \in I_\infty} \frac{|[\lambda_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + |[\lambda_k]_{i_0} - [\lambda_*]_{i_0}| \right]. \tag{65}$$

Now, by (44) with λ_k replacing λ_{k+1}, (63) implies that

$$|[\lambda_k]_i - [\lambda_*]_i| \leq \|\lambda_k - \lambda_*\|_\infty \leq (\alpha_4 + \alpha_5) \|h(x_{k-1})\|_\infty \quad i = 1, \ldots, m. \tag{66}$$

Since $\varepsilon_k \leq \varepsilon_{k-1} \leq \|h(x_{k-1})\|_\infty$, combining (64), (65) and (66), we obtain:

$$|h_{i_0}(x_k)| \leq \frac{\pi_k(i_0)}{[\rho_k]_{i_0}} \|h(x_{k-1})\|_\infty,$$

where

$$\pi_k(i_0) = (b_1 + b_2 \alpha_2) + b_2 \alpha_2 |I_a| \tau + \left(b_2 \alpha_3 \sum_{i \in I_\infty} \frac{1}{[\rho_k]_i} + 1 \right) (\alpha_4 + \alpha_5).$$

Since $\frac{\pi_k(i_0)}{[\rho_k]_{i_0}} \to 0$, there exists $\tilde{k}(i_0) \geq \tilde{k}$ such that

$$|h_{i_0}(x_k)| \leq \tau \|h(x_{k-1})\|_\infty$$

for all $k \geq \tilde{k}(i_0)$. Therefore, $[\rho_{k+1}]_{i_0} = [\rho_k]_{i_0}$. This is a contradiction. □

Remark 4. As in the case of Theorem 4, the choice of ε_k that satisfies (63) is adaptive. In other words, the precision needed for solving each subproblem depends on the level of infeasibility of the approximate solution. The sequence $\{\varepsilon'_k\}$ is given and the stopping criterion at each subproblem is

$$\|\mathcal{P}_\Omega[x_k - G(x_k, \bar{\lambda}_k, \rho_k)] - x_k\|_\infty \leq \min\{\varepsilon_{k-1}, \|h(x_k)\|_\infty, \varepsilon'_k\},$$

where ε_{k-1} is defined by (63). So, as in Theorem 4, the inner algorithm that solves the subproblem returns x_k only when this stopping condition is fulfilled.

6. Numerical experiments

Our main objective regarding this set of experiments is to decide between Algorithm A1 and Algorithm A2. From the theoretical point of view, Algorithm A1 has the advantage that the set of possible infeasible limit points seems to be smaller than the set of possible infeasible limit points of Algorithm A2. Thus, in principle, Algorithm A2 might converge to infeasible points more often than Algorithm A1. On the other hand, Algorithm A2 tends to increase the penalty parameters less frequently than Algorithm A1, a fact that has a positive influence on the conditioning of the subproblems.

However, we are also interested in testing several different options for the implementation of the algorithms. Namely: the best values for $\bar{\lambda}_{min}$ and $\bar{\lambda}_{max}$ (large or small?), the best value for the tolerance τ that determines the increase of penalty parameters and the strategy for choosing ε_k.

Summing up, the practical algorithms to be tested are defined by:

1. Strategy for updating penalty parameters
   Option ONE: Algorithm A1.
   Option TWO: Algorithm A2.
2. Choice of the safeguarded Lagrange multiplier approximations
   Option BIG: $\bar{\lambda}_{max} = -\bar{\lambda}_{min} = 10^{20}$.
   Option SMALL: $\bar{\lambda}_{max} = -\bar{\lambda}_{min} = 10^{6}$.
3. Tolerance for improvement of feasibility
   Option TIGHT: $\tau = 0.1$.
   Option LOOSE: $\tau = 0.5$.
4. Strategy for convergence criterion of subproblems
   Option FIX : $\varepsilon_k = \varepsilon_{min} \geq 0$ for all k.
   Option INEX: $\varepsilon_k = \max\{0.1^k, \varepsilon_{min}\}$ for all k.
   Option ADPT: $\varepsilon'_k = \max\{0.1^k, \varepsilon_{min}\}$ for all k,
   $$\varepsilon_k = \max\{\varepsilon_{min}, \min\{\varepsilon'_k, \|h(x_k)\|_\infty\}\}$$
   for Algorithm A1 and
   $$\varepsilon_k = \max\{\varepsilon_{min}, \min\{\varepsilon_{k-1}, \varepsilon'_k, \|h(x_k)\|_\infty\}\}$$
   for Algorithm A2.

Therefore, 24 different methods are defined. Observe that, when $\varepsilon_{min} = 0$, the option ADPT corresponds to the theoretical hypotheses used in Section 5 to prove boundedness of the penalty parameters. Obviously, in practical (floating point) computations we must choose some small $\varepsilon_{min} > 0$.

The implementation decisions that are common to all the options were the following:

1. For solving the box-constrained minimization subproblems (6) at Step 2 of both algorithms we used GENCAN [5] with its default parameters. The resulting code (Augmented Lagrangian with GENCAN) will be called ALGENCAN.

2. We computed the Lagrange multipliers estimates using (4) and (29).
3. We set $[\rho_1]_i = 10$ for all $i = 1, \dots, m$ and $\gamma = 10$ for both algorithms.
4. The algorithms were stopped declaring *Convergence* when

$$\|\mathcal{P}_\Omega(x_k - F(x_k) - \nabla h(x_k)\lambda_{k+1}) - x_k\|_\infty \le \varepsilon_{min}$$

and

$$\|h(x_k)\|_\infty \le \varepsilon_{min}.$$

We used $\varepsilon_{min} = 10^{-4}$. An execution is stopped declaring *Time exceeded* if the algorithm runs during 10 minutes without achieving *Convergence*. Other stopping criteria were inhibited in order to ensure an homogeneous comparison.

All experiments were done in a Sun Fire 880 with 8 900 Mhz UltraSPARC III Processors, 32 Gb of RAM memory, running SunOS 5.8. The codes were written in FORTRAN 77 and compiled with Forte Developer 7 Fortran 95 7.0 2002/03/09. We used the optimizing option -O. The codes used in this study are available for download in the TANGO webpage (`www.ime.usp.br/~egbirgin/tango`).

We considered all the nonlinear programming problems with equality constraints and bounds of the CUTE collection [9]. As a whole, we tried to solve 128 problems.

Consider a fixed problem and let $x_{\text{final}}^{(M)}, M = 1, \dots, 24$, be the final point of method M applied to that problem. In this numerical study we say that $x_{\text{final}}^{(M)}$ is feasible if

$$\left\| h\left(x_{\text{final}}^{(M)}\right)\right\|_\infty \le \varepsilon_{min}.$$

We define

$$f_{\text{best}} = \min_M \left\{ f\left(x_{\text{final}}^{(M)}\right) \mid x_{\text{final}}^{(M)} \text{ is feasible} \right\}.$$

We say that method M *found a solution* of the problem if $x_{\text{final}}^{(M)}$ is feasible and

$$f\left(x_{\text{final}}^{(M)}\right) \le f_{\text{best}} + 10^{-3}|f_{\text{best}}| + 10^{-6}.$$

Let $t^{(M)}, M = 1, \dots, 24$, be the computer CPU time that method M used to find a solution. If the method did not find a solution we define $t^{(M)} = \infty$. We define

$$t_{\text{best}} = \min_M \{t^{(M)} \mid \text{ method } M \text{ found a solution}\},$$

and we say that method M is *one of the fastest* methods for the problem when

$$t^{(M)} \le t_{\text{best}} + 0.01\, t_{\text{best}} < \infty.$$

These definitions are the same used in [4] for comparing different Augmented Lagrangian formulae.

We are interested in comparing the 24 variants of Augmented Lagrangian algorithms with respect to Robustness, Feasibility and Efficiency. We say that a

Table 1. Performance of ALGENCAN

Method				Performance		
Strategy for updating penalty parameters	Choice of the safeguarded Lagrange multiplier approximations	Tolerance for improvement of feasibility	Strategy for convergence criterion of subproblems	R	F	E
TWO	LOOSE	BIG	FIX	1(96)	1(102)	1(56)
TWO	LOOSE	BIG	INEX	1(96)	2(101)	13(28)
ONE	LOOSE	BIG	FIX	3(95)	11(100)	5(52)
TWO	TIGHT	BIG	FIX	3(95)	11(100)	4(53)
TWO	TIGHT	SMALL	FIX	3(95)	11(100)	3(54)
TWO	LOOSE	SMALL	FIX	3(95)	2(101)	2(55)
TWO	LOOSE	SMALL	INEX	3(95)	2(101)	14(27)
ONE	TIGHT	BIG	FIX	8(94)	11(100)	5(52)
ONE	LOOSE	BIG	ADPT	8(94)	16(99)	23(12)
ONE	LOOSE	SMALL	ADPT	8(94)	19(98)	23(12)
TWO	TIGHT	BIG	ADPT	8(94)	2(101)	15(26)
TWO	TIGHT	SMALL	ADPT	8(94)	2(101)	16(25)
TWO	LOOSE	BIG	ADPT	8(94)	2(101)	22(15)
TWO	LOOSE	SMALL	ADPT	8(94)	2(101)	22(15)
ONE	TIGHT	BIG	INEX	15(93)	11(100)	12(36)
ONE	TIGHT	SMALL	INEX	15(93)	16(99)	11(37)
ONE	LOOSE	BIG	INEX	15(93)	16(99)	17(24)
ONE	LOOSE	SMALL	FIX	15(93)	21(97)	5(52)
ONE	LOOSE	SMALL	INEX	15(93)	19(98)	17(24)
TWO	TIGHT	BIG	INEX	15(93)	2(101)	9(38)
TWO	TIGHT	SMALL	INEX	15(93)	2(101)	9(38)
ONE	TIGHT	BIG	ADPT	22(92)	23(95)	19(21)
ONE	TIGHT	SMALL	FIX	22(92)	21(97)	5(52)
ONE	TIGHT	SMALL	ADPT	22(92)	23(95)	19(21)

particular algorithm is *robust* for solving some problem if it *finds the solution* of the problem according to the criterion defined above. We say that it is *feasible* if it finds a feasible point and we say that it is *efficient* if it is *one of the fastest* methods for solving the problem. In Table 1 we report, for each combination of parameters, the number of problems in which the corresponding algorithm was robust, feasible and efficient, respectively. More precisely, the symbol $p(q)$ under column R indicates that the algorithm found the solution of q problems, according the criterion above and that its rank with respect to robustness was p. The symbol $p(q)$ under column F means that the algorithm found a feasible point in q cases and ranked p with respect to feasibility. The same symbol under column E means that the algorithm was one of the fastest in q cases and ranked p with respect to this criterion.

Some preliminary conclusions may be drawn by inspection of Table 1.

- One of the methods (Algorithm A2 with $\tau = 0.5, \bar{\lambda}_{max} = 10^{20}, \varepsilon_k \equiv \varepsilon_{min}$) appears to be the best one, considering feasibility, robustness and efficiency.
- Algorithm A2 is better than Algorithm A1. This means that using different penalty parameter and increasing separately each of them is better than

increasing "all" the penalty parameters when the improvement of just one constraint is not enough, as Algorithm A1 does.

- In general, using a fixed small convergence criterion in the subproblems ($\varepsilon_k = \varepsilon_{min}$) is better than using different choices of ε_k at least in terms of efficiency. With respect to feasibility and robustness the different choices of ε_k are equivalent.
- The option LOOSE for increasing the penalty parameter is slightly better than the option TIGHT. The choice of $\bar{\lambda}_{max}$ between 10^6 and 10^{20} is not very relevant. Preliminary experiments showed that smaller values of $\bar{\lambda}_{max}$ are not convenient.

In order to test the consistency of our algorithms we compared our winner Augmented Lagrangian algorithm with the default version of LANCELOT [12] and with the same version with true Hessians and without preconditioners. The last one is more adequate since the version of GENCAN that we use does not employ preconditioners at all. It must be observed that GENCAN does not use true Hessians either. Matrix-vector products involving Hessians are replaced by incremental gradient quotients in GENCAN. ALGENCAN was more efficient and robust than the version of LANCELOT without preconditioners. It was also more efficient than the preconditioned LANCELOT but not as robust as this method. The corresponding performance profile [15] is shown in Figure 1.

7. Conclusions

Augmented Lagrangian methods are useful tools for solving many practical nonconvex minimization problems with equality constraints and bounds. Its extension to KKT systems and, in consequence, to a wide variety of equilibrium problems (see [26,27]) is straightforward. We presented two Augmented Lagrangian algorithms for this purpose. They differ only in the way in which penalty parameters are updated. There seems to be an important difference between these two algorithms with respect to convergence properties. According to our feasibility results the set of possible infeasible limit points of Algorithm A1 seems to be strictly contained in the set of possible infeasible limit points of Algorithm A2. This could indicate that Algorithm A2 converges to infeasible points more frequently than Algorithm A1. However, this property was not confirmed by numerical experiments, which indicate that Algorithm A2 is better. So, it seems that maintaining moderate values of the penalty parameters is the more important feature for explaining the practical performance. However, it is still an open problem if stronger results than Theorem 2 can be obtained for Algorithm A2.

The question about convergence to optimal (KKT) points is also relevant. Up to our knowledge, convergence to KKT points of algorithms of this type had been obtained only using assumptions on the linear independence of active constraints. Here we proved that a much better constraint qualification (CPLD) can be used with the same purpose. Again, the problem of finding even weaker con-

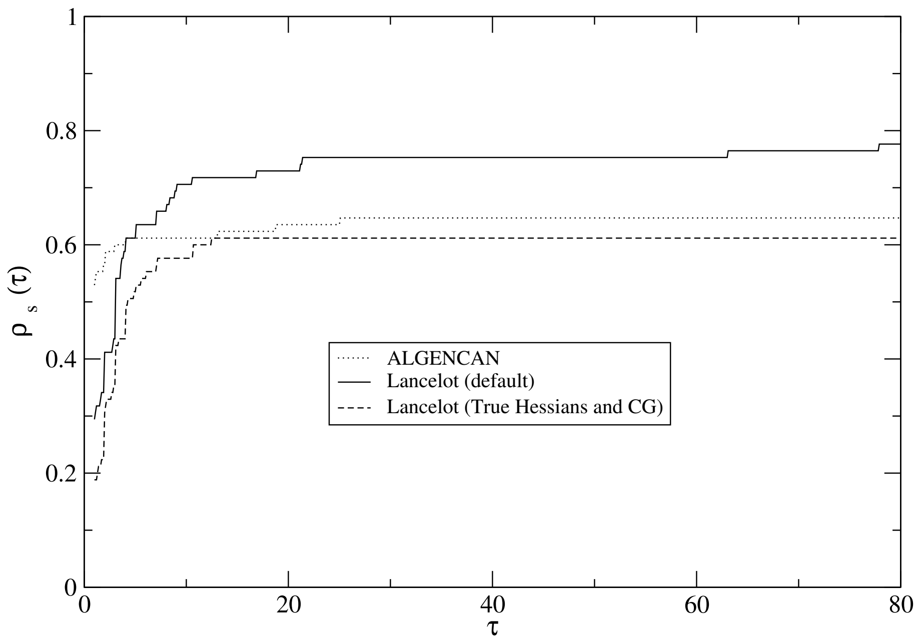

Fig. 1. ALGENCAN versus LANCELOT.

straint qualifications under which convergence to KKT points can be guaranteed remains open.

The superiority of Algorithm A2 over Algorithm A1 in numerical experiments was not a surprise since every optimization practitioner is conscious of the effect of large penalty parameters on the conditioning of the subproblems and, hence, on the overall performance of Augmented Lagrangian and penalty methods. A little bit more surprising was the (slight) superiority of the algorithms based on accurate resolution of the subproblems over the ones based on inexact resolution. Careful inspection of some specific cases lead us to the following explanation for that behavior. On one hand, GENCAN, the algorithm used to solve the subproblems is an inexact-Newton method whose behavior is many times similar to Newton's method especially when the iterate is close to the solution. This implies that, after satisfying a loose convergence criterion, the amount of effort needed for satisfying a strict convergence criterion is usually small. In these cases it is not worthwhile to interrupt the execution for defining a new subproblem. (One would be "abandoning Newton" precisely in the region where it is more efficient!) On the other hand, the formula used for updating the Lagrange multipliers is a first-order formula motivated by the assumption of exact solution of the subproblems. When the resolution is inexact, other updating formulae ([24], p. 291) might be more efficient (although, of course, more costly).

The conclusion about the relative efficiency of solving accurately or inaccurately the subproblem may change if one uses different box-constrained solvers. The excellent behavior of the spectral gradient method for very large convex constrained minimization [6–8,13,30,31] is a strong motivation for pursuing the research on inexact stopping criteria for the subproblems, since in this case quadratic or superlinear convergence is not expected.

Valuable research has been done in the last 10 years in Augmented Lagrangian methods for solving quadratic problems originated from mechanical applications [16–18]. Adaptive criteria that depend on feasibility of the current point (as in the assumptions of our penalty boundedness theorems) have been successfully used and justified from several different points of view. (Antecedents of these practical strategies can be found in [25].) More recently [16], Dostál showed that, for some convex quadratic programming problems, an updating strategy based on the increase of the Augmented Lagrangian function have interesting theoretical and practical properties. The extension of his philosophy to the general nonquadratic and nonconvex case must be investigated.

The recent development of efficient sequential quadratic programming, interior-point and restoration methods for nonlinear programming motivates a different line of Augmented Lagrangian research. The "easy" set Ω does not need to be a box and, in fact, it does not need to be "easy" at all if a suitable algorithm for minimizing on it is available. (The case in which Ω is a general polytope was considered in [11].) However, many times the intersection of Ω with the general constraints $h(x) = 0$ is very complicated. In these cases, using the Augmented Lagrangian approach to deal with the general constraints and a different nonlinear programming algorithm to deal with the subproblems is attractive. Certainly, this has been done in practical applications for many years. The convergence properties of these combinations using weak constraint qualifications is considered in a separate report [1].

Inequality constraints in the original problem can be reduced to equality and box constraints by means of the addition of slack variables and bounds. However, it is interesting to consider directly Augmented Lagrangian methods that deal with inequality constraints without that reformulation. The most popular Augmented Lagrangian function for inequality constraints [32] can be obtained by reducing the inequality constrained problem to an equality constrained one, with the help of squared slack variables, and applying the equality Augmented Lagrangian to the new problem. After some manipulation, squared slack variables are eliminated and an Augmented Lagrangian without auxiliary variables arises [3]. Many alternative Augmented Lagrangians with better smoothness properties than the classical one have been introduced. It is possible to obtain feasibility and global convergence results for methods based on many inequality Augmented Lagrangians after removing a considerable number of technical difficulties [1,4]. However, results on the boundedness of the penalty parameters are harder to obtain. In particular, strict complementarity at the limit point seems to be necessary for obtaining such results. This assumption is not used at all in the present paper.

We presented our methods and theory considering KKT systems and not merely minimization problems to stress the applicability of the Augmented Lagrangian strategy to the general KKT case. We performed several experiments for general KKT systems, where the algorithm used for solving the subproblems was the well known PATH solver (see [14]). We compared the resulting algorithm with the PATH method for solving directly the original problem. On one hand, we confirmed the following warning of [23]: "Typically, singularity [of the Jacobian] does not cause a lot of problems and the algorithm [PATH] can handle the situation appropriately. However, an excessive number of singularities are cause of concern. A further indication of possible singularities at the solution is the lack of quadratic convergence to the solution". In fact, for some tested problems, the effect of singularity of the Jacobian was more serious in the direct application of PATH to the original problem than in the "Augmented Lagrangian with PATH" algorithm. In many other situations the direct application of PATH to the KKT system was more efficient. Clearly, the Augmented Lagrangian framework intensely exploits the minimization structure of the problem when the source of the KKT system is nonlinear programming and loses this advantage when the KKT system is general. However, much research is necessary in order to evaluate the potentiality of the Augmented Lagrangian for equilibrium problems, variational inequalities and related problems.

Acknowledgements. We are indebted to the *Advanced Database Laboratory (LABD)* and *Bioinfo* at the Institute of Mathematics and Statistics of the University of São Paulo for computer facilities and Sergio Ventura for computational insight. Moreover, we are deeply indebted to two anonymous referees whose insightful comments helped us a lot to improve the quality of the paper.

References

1. R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, *On Augmented Lagrangian methods with general lower-level constraints.* Technical Report MCDO-050303, Department of Applied Mathematics, UNICAMP, Brazil, 2005.
2. R. Andreani, J. M. Martínez and M. L. Schuverdt. *On the relation between Constant Positive Linear Dependence condition and quasinormality constraint qualification*, Journal of Optimization Theory and Applications 125, pp. 473–485 (2005).
3. D. P. Bertsekas. *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts, 2003.
4. E. G. Birgin, R. Castillo and J. M. Martínez. *Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems*, Computational Optimization and Applications 31, pp. 31–56 (2005).
5. E. G. Birgin and J. M. Martínez. *Large-scale active-set box-constrained optimization method with spectral projected gradients*, Computational Optimization and Applications 23, pp. 101–125 (2002).
6. E. G. Birgin, J. M. Martínez and M. Raydan. *Nonmonotone spectral projected gradient methods on convex sets*, SIAM Journal on Optimization 10, pp. 1196–1211 (2000).
7. E. G. Birgin, J. M. Martínez and M. Raydan. *Algorithm 813: SPG - Software for convex-constrained optimization*, ACM Transactions on Mathematical Software 27, pp. 340–349 (2001).

8. E. G. Birgin, J. M. Martínez and M. Raydan. *Inexact Spectral Projected Gradient methods on convex sets*, IMA Journal on Numerical Analysis 23, pp. 539–559 (2003).
9. I. Bongartz, A. R. Conn, N. I. M. Gould and Ph. L. Toint. *CUTE: constrained and unconstrained testing environment*, ACM Transactions on Mathematical Software 21, pp. 123–160 (1995).
10. A. R. Conn, N. I. M. Gould and Ph. L. Toint. *LANCELOT: A Fortran package for large scale nonlinear optimization.* Springer-Verlag, Berlin, 1992.
11. A. R. Conn, N. I. M. Gould, A. Sartenaer and Ph. L. Toint. *Convergence properties of an Augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints*, SIAM Journal on Optimization 6, pp. 674–703 (1996).
12. A. R. Conn, N. I. M. Gould and Ph. L. Toint. *A globally convergent Augmented Lagrangian algorithm for optimization with general constraints and simple bounds*, SIAM Journal on Numerical Analysis 28, pp. 545–572 (1991).
13. M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, J. M. Martínez and S. A. Santos. *Augmented Lagrangian algorithms based on the spectral projected gradient for solving nonlinear programming problems*, Journal of Optimization Theory and Applications 123, pp. 497–517 (2004).
14. S. P. Dirkse and M. C. Ferris. *The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems*, Optimization Methods and Software 5, pp. 123–156 (1995).
15. E. D. Dolan and J. J. Moré. *Benchmarking optimization software with performance profiles*, Mathematical Programming 91, pp. 201–213 (2002).
16. Z. Dostál. *Inexact semimonotonic augmented Lagrangians with optimal feasibility convergence for convex bound and equality constrained quadratic programming*, SIAM Journal on Numerical Analysis 43, pp. 96–115 (2005).
17. Z. Dostál, A. Friedlander and S. A. Santos. *Augmented Lagrangian with adaptive precision control for quadratic programming with simple bounds and equality constraints*, SIAM Journal on Optimization 13, pp. 1120–1140 (2003).
18. Z. Dostál, F. A. M. Gomes and S. A. Santos. *Duality based domain decomposition with natural coarse space for variational inequalities*, Journal of Computational and Applied Mathematics 126, pp. 397–415 (2000).
19. F. Facchinei, A. Fischer and C. Kanzow. *Regularity properties of a semismooth reformulation of variational inequalities*, SIAM Journal on Optimization 8, pp. 850–869 (1998).
20. F. Facchinei, A. Fischer and C. Kanzow. *A semismooth Newton method for variational inequalities: The case of box constraints.* In: M. C. Ferris and J.-S. Pang (eds.): Complementarity and Variational Problems: State of the Art. SIAM, Philadelphia, 1997, pp. 76–90.
21. F. Facchinei, A. Fischer, C. Kanzow and J.-M. Peng. *A simply constrained reformulation of KKT systems arising from variational inequalities*, Applied Mathematics and Optimization 40, pp. 19–37 (1999).
22. F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementary Problems*, Springer Series in Operation Research, Springer (2000).
23. M. C. Ferris and T. S. Munson. *PATH 4.6 User Manual*, http://www.gams.com/solvers/solvers.htm#PATH.
24. R. Fletcher. *Practical Methods of Optimization*, Academic Press, London, 1987.
25. W. W. Hager. *Analysis and implementation of a dual algorithm for constrained optimization*, Journal of Optimization Theory and Applications 79, pp. 37–71 (1993).
26. P. T. Harker and J.-S. Pang. *Finite-dimensional variational inequality and nonlinear complementary problems: a survey of theory, algorithms and applications*, Mathematical Programming 48, pp. 161–220 (1990).
27. Z.-Q. Luo, J.-S. Pang and D. Ralph. *Mathematical programs with equilibrium constraints*, Cambridge, University Press, 1996.
28. O. L. Mangasarian and S. Fromovitz. *The Fritz-John necessary optimality conditions in presence of equality and inequality constraints*, Journal of Mathematical Analysis and Applications 17, pp. 37–47 (1967)
29. L. Qi and Z. Wei. *On the constant positive linear dependence condition and its application to SQP methods*, SIAM Journal on Optimization 10, pp. 963–981 (2000).
30. M. Raydan. *On the Barzilai and Borwein choice of steplength for the gradient method*, IMA Journal of Numerical Analysis 13, pp. 321–326 (1993).
31. M. Raydan. *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, SIAM Journal on Optimization 7, pp. 26–33 (1997).
32. R. T. Rockafellar. *The multiplier method of Hestenes and Powell applied to convex programming*, Journal of Optimization Theory and Applications 12, pp. 555–562 (1973).

33. R. T. Rockafellar. *Lagrange multipliers and optimality*, SIAM Review 35, pp. 183–238 (1993).
34. A. Shapiro. *Sensitivity analysis of parameterized variational inequalities*, Mathematics of Operations Research 30, pp. 109-126 (2005).