Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification

R. Andreani [*] E. G. Birgin [†] J. M. Martínez [‡] M. L. Schuverdt [§]

August 6, 2004, 11.44 hs.

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 $F : I\!R^n \to I\!R^n$, $h : I\!R^n \to I\!R^m$, $\ell, u \in I\!R^n, \ell < u$, $\Omega = \{x \in I\!R^n \mid \ell \leq x \leq u\}$.

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 on a closed and convex set A. A point $x \in \Omega$ 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

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

[*]Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization 76.79.1008-00, FAPESP (Grant 01-04597-4) and CNPq. e-mail: andreani@ime.unicamp.br

[†]Department of Computer Science IME-USP, University of São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-090, São Paulo SP, Brazil. This author was supported by PRONEX-Optimization 76.79.1008-00, FAPESP (Grants 01-04597-4 and 02-00094-0) and CNPq (Grant 300151/00-4). e-mail: egbirgin@ime.usp.br

[‡]Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization 76.79.1008-00, FAPESP (Grant 01-04597-4) and CNPq. e-mail: martinez@ime.unicamp.br

[§]Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization 76.79.1008-00 and FAPESP (Grants 01-04597-4 and 02-00832-1). e-mail: schuverd@ime.unicamp.br

If $F = \nabla f$ for some $f : I\!\!R^n \to I\!\!R$, the equations (1) represent the KKT optimality conditions of the minimization problem

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

The KKT form (1) allows one to consider more general situations, as equilibrium problems, variational inequalities and some variations. See [21, 23, 26, 27] and references therein. The introduction of the Fischer-Burmeister function [23] made it possible to reduce KKT systems to suitable semismooth nonlinear systems of equations. See [29, 30, 31]. Many authors used the semismooth approach to obtain interesting algorithms for solving KKT systems. See [18, 19, 20, 36].

The most influential work on practical Augmented Lagrangian algorithms for minimization with equality constraints and bounds was the paper by Conn, Gould and Toint [11], on which the LANCELOT package [9] is based. Convergence of the algorithm presented in [11] was proved under the assumption that the gradients of the general constraints and the active bounds at any limit point are linearly independent. See, also, [10]. In the present paper we do not use this assumption at all. Firstly, we characterize the situations in which infeasible limit points might exist using weaker assumptions than the linear independence condition. Moreover, the fact that feasible limit points are KKT points will follow using the Constant Positive Linear Dependence (CPLD) condition [32], which has been recently proved to be a constraint qualification [1] and is far more general than the regularity condition and other popular constraint qualifications. We use regularity, following closely the development of [11], 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.

We denote:

$$I\!\!R_+ = \{t \in I\!\!R \mid t \geq 0\},$$

$$I\!\!R_{++} = \{t \in I\!\!R \mid t > 0\},$$

$$I\!\!N = \{0, 1, 2, \ldots\},$$

$$\{e_1, \ldots, e_n\} \quad \text{the canonical basis of } I\!\!R^n.$$

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 $x \in \Omega$, $\lambda \in I\!R^m$, $\rho \in I\!R_{++}$ we define

$$G_1(x, \lambda, \rho) = F(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \rho \sum_{i=1}^{m} h_i(x) \nabla h_i(x).$$

For $\rho \in I\!R^m_{++}$ we define

$$G_2(x, \lambda, \rho) = F(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{i=1}^{m} \rho_i h_i(x) \nabla h_i(x).$$

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

$$L_1(x, \lambda, \rho) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \frac{\rho}{2} \sum_{i=1}^{m} h_i(x)^2$$

and, for $\rho \in I\!R^m_{++}$,

$$L_2(x, \lambda, \rho) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \frac{1}{2} \sum_{i=1}^{m} \rho_i h_i(x)^2.$$

In these cases we have that $\nabla L_1 = G_1$ and $\nabla L_2 = G_2$. The functions L_1 and L_2 are one-parameter and many-parameters Augmented Lagrangians associated to the problem (2).

The mappings G_1 and G_2 will be used to define one-parameter and many-parameters Augmented Lagrangian algorithms for solving the general KKT problem (1). These algorithms are described below.

Algorithm 1.

Let $x_0 \in \Omega$, $\tau \in [0, 1)$, $\gamma > 1$, $-\infty < \bar{\lambda}_{min} < \bar{\lambda}_{max} < \infty$, $\rho_1 > 0$, $\bar{\lambda}_1 \in [\bar{\lambda}_{min}, \bar{\lambda}_{max}]^m$. Let $\{\varepsilon_k\}_{k \in I\!N} \subset I\!R_{++}$ 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

$$\|\mathcal{P}_\Omega[x_k - G_1(x_k, \bar{\lambda}_k, \rho_k)] - x_k\|_\infty \le \varepsilon_k. \tag{3}$$

Step 3. *Estimate multipliers*

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

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

If $h(x_k) = 0$ and $\mathcal{P}_\Omega[x_k - G_1(x_k, \bar{\lambda}_k, \rho_k)] - x_k = 0$ 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

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

Step 4. *Update the penalty parameter*

If

$$\|h(x_k)\|_\infty \leq \tau \|h(x_{k-1})\|_\infty \tag{6}$$

define

$$\rho_{k+1} = \rho_k,$$

else, define

$$\rho_{k+1} = \gamma \rho_k.$$

Step 5. *Begin a new iteration*

Set $k \leftarrow k + 1$. Go to Step 2.

Algorithm 2 only differs from Algorithm 1 in the way in which penalty parameters are updated. In Algorithm 2 we use one penalty parameter for each constraint.

Algorithm 2.

Let $x_0 \in \Omega$, $\tau \in [0, 1)$, $\gamma > 1$, $-\infty < \bar{\lambda}_{min} < \bar{\lambda}_{max} < \infty$, $\rho_1 \in I\!R^m_{++}$, $\bar{\lambda}_1 \in [\bar{\lambda}_{min}, \bar{\lambda}_{max}]^m$. Let $\{\varepsilon_k\}_{k \in I\!N} \subset I\!R_{++}$ 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

$$\|\mathcal{P}_\Omega[x_k - G_2(x_k, \bar{\lambda}_k, \rho_k)] - x_k\|_\infty \leq \varepsilon_k. \tag{7}$$

Step 3. *Estimate multipliers*

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

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k) \tag{8}$$

If $h(x_k) = 0$ and $\mathcal{P}_\Omega[x_k - G_2(x_k, \bar{\lambda}_k, \rho_k)] - x_k = 0$ terminate the execution of the algorithm. (As in Algorithm 1, x_k is a KKT point and λ_{k+1} is the vector of Lagrange multipliers.)

Compute

$$\bar{\lambda}_{k+1} \in [\bar{\lambda}_{min}, \bar{\lambda}_{max}]^m. \tag{9}$$

Step 4. *Update the penalty parameters*

For all $i = 1, \ldots, m$, if

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

define

$$[\rho_{k+1}]_i = [\rho_k]_i.$$

Else, define

$$[\rho_{k+1}]_i = \gamma[\rho_k]_i.$$

Step 5. *Begin a new iteration*

Set $k \leftarrow k+1$. Go to Step 2.

Remarks.

1. The difference between Algorithms 1 and 2 relies on the updating formula for the penalty parameters. In the case in which Algorithm 2 updates at least one penalty parameter, Algorithm 1 updates its unique penalty parameter. In such a situation, other penalty parameters may remain unchanged in Algorithm 2. Therefore, the penalty parameters at Algorithm 2 tend to be smaller than the penalty parameter at Algorithm 1.

2. The global convergence results to be presented in the following sections are independent of the choice of $\bar{\lambda}_{k+1}$ in (5) and (9).

3. The Augmented Lagrangian algorithms are based on the resolution of the inner problems (3) and (7). In the minimization case ($F = \nabla f$) the most reasonable way for obtaining these conditions is to solve (approximately) the box-constrained minimization problem

$$\text{Minimize } L_1(x, \bar{\lambda}_k, \rho_k) \text{ subject to } x \in \Omega \tag{10}$$

in the case of Algorithm 1, and

$$\text{Minimize } L_2(x, \bar{\lambda}_k, \rho_k) \text{ subject to } x \in \Omega \tag{11}$$

in the case of Algorithm 2. Both (10) and (11) are box-constrained minimization problems. 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) and (7) exist and can be obtained by available algorithms in more general problems have been analyzed in many recent papers. See [18, 19, 20, 21, 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.

Briefly speaking, in this section we show that Algorithm 1 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 2 we will show that the set of 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 obvious property:

$$\|\mathcal{P}_\Omega(u + tv) - u\|_2 \le \|\mathcal{P}_\Omega(u + v) - u\|_2 \;\; \forall\, u \in \Omega, v \in I\!R^n, t \in [0, 1]. \tag{12}$$

Theorem 3.1. *Assume that the sequence $\{x_k\}$ is generated by Algorithm 1 and that x_* is a limit point. Then, x_* is a stationary point of the problem*

$$\begin{array}{c} \textit{Minimize } \|h(x)\|_2^2 \\ \textit{subject to } x \in \Omega. \end{array} \tag{13}$$

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

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

$$\lim_{k \to \infty} \|\mathcal{P}_\Omega[x_k - F(x_k) - \sum_{i=1}^m ([\bar{\lambda}_k]_i + \rho_k h_i(x_k)) \nabla h_i(x_k)] - x_k\|_2 = 0. \tag{14}$$

By (6), if $\{\rho_k\}_{k \in K}$ is bounded we have that $h(x_*) = 0$, so x_* is a stationary point of (13). Assume that $\{\rho_k\}_{k \in K}$ is unbounded. Since $\{\rho_k\}$ is nondecreasing, we have that

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

Then, $\rho_k > 1$ for $k \in K$ large enough. So, using (12) with

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

we have, by (14), that

$$\lim_{k \to \infty} \left\| \mathcal{P}_\Omega \left[x_k - \frac{F(x_k)}{\rho_k} - \sum_{i=1}^m \left(\frac{[\bar{\lambda}_k]_i}{\rho_k} + h_i(x_k) \right) \nabla h_i(x_k) \right] - x_k \right\|_2 = 0. \tag{16}$$

By (15) and (16), since $\{\bar{\lambda}_k\}_{k \in K}$ is bounded, 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 (13), 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

$$\begin{array}{c} \text{Minimize } \sum_{i=1}^m w_i h_i(x)^2 \\ \text{subject to } x \in \Omega, \end{array} \tag{17}$$

and

$$\sum_{i=1}^m w_i h_i(x_*)^2 > 0. \tag{18}$$

Theorem 3.2. *Let $\{x_k\}$ be a sequence generated by Algorithm 2. 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

$$\|h(x_k)\|_\infty \geq \varepsilon \tag{19}$$

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

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

Let K be an infinite subset of $I\!N$ such that

$$\|\rho_k\|_\infty > \|\rho_{k-1}\|_\infty \;\; \forall\, k \in K. \tag{20}$$

Let K_1 be an infinite subset of K and $j \in \{1, \ldots, m\}$ be such that

$$\|\rho_k\|_\infty = [\rho_k]_j \;\; \forall\, k \in K_1. \tag{21}$$

Then, by (20),

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

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 (19),

$$|h_j(x_{k-1})| > \tau\varepsilon \;\; \forall \;\, k \in K_1. \tag{22}$$

Moreover, by the definition of the algorithm, (20) and (21), we have:

$$[\rho_{k-1}]_j \geq \frac{\|\rho_{k-1}\|_\infty}{\gamma} \;\; \forall \;\, k \in K_1. \tag{23}$$

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 (22) we have that

$$h_j(x_*) \neq 0. \tag{24}$$

By (7) and the equivalence of norms in $I\!R^n$, we have:

$$\lim_{k\to\infty} \|\mathcal{P}_\Omega[x_k - F(x_k) - \sum_{i=1}^{m}([\bar\lambda_k]_i + [\rho_k]_i h_i(x_k))\nabla h_i(x_k)] - x_k\|_2 = 0. \tag{25}$$

Clearly $\|\rho_k\|_\infty > 1$ for $k \in K_2$ large enough. So, using (12) 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 (25), that

$$\lim_{k\in K_2}\left\|\mathcal{P}_\Omega\left[x_k-\frac{F(x_k)}{\|\rho_k\|_\infty}-\sum_{i=1}^m\left(\frac{[\bar\lambda_k]_i}{\|\rho_k\|_\infty}+\frac{[\rho_k]_i}{\|\rho_k\|_\infty}h_i(x_k)\right)\nabla h_i(x_k)\right]-x_k\right\|_2=0. \tag{26}$$

But

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

Therefore, there exist $K_3\subset K_2$ and $w\in I\!R^m_+$ such that

$$\lim_{k\in K_1}\frac{[\rho_k]_i}{\|\rho_k\|_\infty}=w_i \quad \forall i=1,\dots,m.$$

Moreover, by (23),

$$w_j>0. \tag{27}$$

Since $\{\bar\lambda_k\}_{k\in K_1}$ is bounded, taking limits for $k\in K_1$ in (26), 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 (17). By (24) and (27), the condition (18) also takes place. Therefore, x_* is a degenerate infeasible point. □

Remark. Clearly, any infeasible stationary point of (13) must be degenerate. Moreover, if x is infeasible and degenerate, by (18) and the KKT conditions of (17), 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

$$h(x)=x=0\in I\!R^1,\ \ -1\le x\le 1. \tag{28}$$

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 [11] 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 [11]). See, also, [10]. Under this assumption it is proved that the limit points are feasible. By the considerations above, we see that Assumption AS3 is stronger than assuming the nonexistence of infeasible degenerate points. In other words, the assumption on the problem that guarantees that AS3 holds is that the gradients of equality constraints and active bound constraints are linearly independent at all the points of the box (not merely at the feasible points). This assumption does not hold in (28) and is much stronger than assuming that there are no degenerate points in Ω.

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 [32]. More recently [1], it was proved that this condition is a constraint qualification. A feasible point satisfies CPLD if one of the following situations take place:

1. The point satisfies the Mangasarian-Fromovitz constraint qualification [28, 35].
2. The point does not satisfy the Mangasarian-Fromovitz constraint qualification but any set of positive linear dependent gradients of active constraints (including equality constraints) remains linearly dependent in a neighborhood of the point.

Obviously, the CPLD condition is weaker than the Mangasarian-Fromovitz constraint qualification. The AS3 condition of [11], when applied only to feasible points, is the classical regularity assumption (linear independence of the gradients of active constraints). In [1] examples where the CPLD condition holds but the Mangasarian-Fromovitz condition does not were given. 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 regularity condition.

Theorem 4.1. *Assume that $\{x_k\}$ is a sequence generated by Algorithm 1 or by Algorithm 2 and that x_* is a feasible limit point that satisfies the CPLD constraint qualification. Then, x_* is a KKT point.*

Proof. Let us write, for Algorithm 1,

$$G^k = G_1(x_k, \bar{\lambda}_k, \rho_k)$$

and, for Algorithm 2,

$$G^k = G_2(x_k, \bar{\lambda}_k, \rho_k).$$

Define, for all $k \in I\!N$,

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

Therefore, $v_k \in I\!R^n$ solves

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

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

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 $k \in I\!N$,

$$v_k - x_k + G^k + \sum_{i=1}^n [\mu_k^u]_i e_i - \sum_{i=1}^n [\mu_k^\ell]_i e_i = 0 \tag{29}$$

and

$$[\mu_k^u]_i(u_i - [x_k]_i) = [\mu_k^\ell]_i(\ell_i - [x_k]_i) = 0 \;\; \forall \;\; i = 1, \ldots, n. \tag{30}$$

By (3) and (7),

$$\lim_{k\to\infty} v_k - x_k = 0,$$

then, by (29),

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

So, defining λ_{k+1} as in (4) and (8),

$$\lim_{k\to\infty} F(x_k) + \nabla h(x_k)\lambda_{k+1} + \sum_{i=1}^{n} [\mu_k^u]_i e_i - \sum_{i=1}^{n} [\mu_k^\ell]_i e_i = 0. \tag{31}$$

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

$$\lim_{k\in K} x_k = x_*. \tag{32}$$

Define:

$$I_\ell = \{i \in \{1, \ldots, n\} \mid [x_*]_i = \ell_i\},$$
$$I_u = \{i \in \{1, \ldots, n\} \mid [x_*]_i = u_i\},$$
$$I_0 = \{i \in \{1, \ldots, n\} \mid \ell_i < [x_*]_i < u_i\}.$$

By (32), there exists $k_0 \in I\!\!N$ such that for all $k \in K, k \geq k_0$,

$$i \in I_0 \Rightarrow \ell_i < [x_k]_i < u_i.$$

So, by (30), for all $k \in K, k \geq k_0$, we have:

$$[\mu_k^u]_i = 0 \; \forall \; i \notin I_u$$

and

$$[\mu_k^\ell]_i = 0 \; \forall \; i \notin I_\ell.$$

So, by (31),

$$\lim_{k\in K} F(x_k) + \nabla h(x_k)\lambda_{k+1} + \sum_{i\in I_u} [\mu_k^u]_i e_i - \sum_{i\in I_\ell} [\mu_k^\ell]_i e_i = 0.$$

Define, for $k \in K, k \geq k_0$,

$$E_k = F(x_k) + \nabla h(x_k)\lambda_{k+1} + \sum_{i\in I_u} [\mu_k^u]_i e_i - \sum_{i\in I_\ell} [\mu_k^\ell]_i e_i.$$

Clearly,

$$\lim_{k\in K} E_k = 0$$

and

$$E_k - F(x_k) = \nabla h(x_k)\lambda_{k+1} + \sum_{i\in I_u}[\mu_k^u]_i e_i - \sum_{i\in I_\ell}[\mu_k^\ell]_i e_i \tag{33}$$

for all $k \in K, k \geq k_0$.

By Caratheodory's Theorem of Cones (see [2], page 689), for all $k \in K, k \geq k_0$ there exist

$$\tilde{I}_k \subset \{1,\ldots,m\},\ \tilde{I}_{uk} \subset I_u,\ \tilde{I}_{\ell k} \subset I_\ell,$$

$$[\tilde{\lambda}_k]_i \ \forall\, i \in \tilde{I}_k,\ \ [\tilde{\mu}_k^u]_i \geq 0 \ \forall\, i \in \tilde{I}_{uk},\ \ [\tilde{\mu}_k^\ell]_i \geq 0 \ \forall\, i \in \tilde{I}_{\ell k}$$

such that the vectors

$$\{\nabla h_i(x_k)\}_{i\in\tilde{I}_k},\ \{e_i\}_{i\in\tilde{I}_{uk}},\ \{-e_i\}_{i\in\tilde{I}_{\ell k}}$$

are linearly independent and

$$E_k - F(x_k) = \sum_{i\in\tilde{I}_k}[\tilde{\lambda}_k]_i \nabla h_i(x_k) + \sum_{i\in\tilde{I}_{uk}}[\tilde{\mu}_k^u]_i e_i - \sum_{i\in\tilde{I}_{\ell k}}[\tilde{\mu}_k^\ell]_i e_i. \tag{34}$$

Since there is only a finite number of possible sets $\tilde{I}_k, \tilde{I}_{uk}, \tilde{I}_{\ell k}$, there exists an infinite set of indices

$$K_1 \subset \{k \in K \mid k \geq k_0\}$$

such that

$$\tilde{I}_k = \tilde{I}, \tilde{I}_{uk} = \tilde{I}_u, \tilde{I}_{\ell k} = \tilde{I}_\ell$$

for all $k \in K_1$.

Therefore, by (34),

$$E_k - F(x_k) = \sum_{i\in\tilde{I}}[\tilde{\lambda}_k]_i \nabla h_i(x_k) + \sum_{i\in\tilde{I}_u}[\tilde{\mu}_k^u]_i e_i - \sum_{i\in\tilde{I}_\ell}[\tilde{\mu}_k^\ell]_i e_i \tag{35}$$

and the vectors

$$\{\nabla h_i(x_k)\}_{i\in\tilde{I}}, \{e_i\}_{i\in\tilde{I}_u}, \{-e_i\}_{i\in\tilde{I}_\ell} \quad \text{are linearly independent} \tag{36}$$

for all $k \in K_1$.

Define

$$S_k = \max\{\max\{|[\tilde{\lambda}_k]_i|, i \in \tilde{I}\},\ \max\{[\tilde{\mu}_k^u]_i, i \in \tilde{I}_u\},\ \max\{[\tilde{\mu}_k^\ell]_i, i \in \tilde{I}_\ell\}\}.$$

We consider two possibilities:

- $\{S_k\}_{k\in K_1}$ bounded;
- $\{S_k\}_{k\in K_1}$ unbounded.

In the first case there exists K_2, an infinite subset of K_1, such that

$$\lim_{k\in K_2}[\tilde{\lambda}_k]_i = \tilde{\lambda}_i,$$

$$\lim_{k\in K_2}[\tilde{\mu}_k^u]_i = \tilde{\mu}_i^u \geq 0 \;\forall\, i \in \tilde{I}_u$$

and

$$\lim_{k\in K_2}[\tilde{\mu}_k^\ell]_i = \tilde{\mu}_i^\ell \geq 0 \;\forall\, i \in \tilde{I}_\ell.$$

So, taking limits in (35) for $k \in K_2$ we obtain that x_* is a KKT point.

Suppose now that $\{S_k\}_{k\in K_1}$ is unbounded. Let K_3 be an infinite subset of K_1 such that $\lim_{k\in K_3} S_k = \infty$ and $S_k > 1$ for all $k \in K_3$. Dividing both sides of (35) by S_k for all $k \in K_3$, we get:

$$\frac{E_k - F(x_k)}{S_k} = \sum_{i\in\tilde{I}} \frac{[\tilde{\lambda}_k]_i}{S_k}\nabla h_i(x_k) + \sum_{i\in\tilde{I}_u} \frac{[\tilde{\mu}_k^u]_i}{S_k} e_i - \sum_{i\in\tilde{I}_\ell} \frac{[\tilde{\mu}_k^\ell]_i}{S_k} e_i. \tag{37}$$

By the definition of S_k, this quantity is the modulus of one of the coefficients $[\tilde{\lambda}_k]_i, [\tilde{\mu}_k^u]_i, [\tilde{\mu}_k^\ell]_i$ that occur in (37). Therefore, we can extract an infinite set $K_4 \subset K_3$ such that, for all $k \in K_4$, S_k is the modulus of the same coefficient. But, since

$$\left|\frac{[\tilde{\lambda}_k]_i}{S_k}\right| \leq 1, \left|\frac{[\tilde{\mu}_k^u]_i}{S_k}\right| \leq 1, \left|\frac{[\tilde{\mu}_k^\ell]_i}{S_k}\right| \leq 1,$$

there exists K_5, an infinite subset of K_4, such that

$$\lim_{k\in K_5}\frac{[\tilde{\lambda}_k]_i}{S_k} = \tilde{\lambda}_i,\; \lim_{k\in K_5}\frac{[\tilde{\mu}_k^u]_i}{S_k} = \tilde{\mu}_i^u \geq 0,\; \lim_{k\in K_5}\frac{[\tilde{\mu}_k^\ell]_i}{S_k} = \tilde{\mu}_i^\ell \geq 0.$$

Then, taking limits on both sides of (37) for $k \in K_5$ we obtain that

$$\sum_{i\in\tilde{I}} \tilde{\lambda}_i \nabla h_i(x_*) + \sum_{i\in\tilde{I}_u} \tilde{\mu}_i^u e_i - \sum_{i\in\tilde{I}_\ell} \tilde{\mu}_i^\ell e_i = 0.$$

Moreover, by the choice of K_4, the modulus of at least one of the coefficients $\tilde{\lambda}_i, \tilde{\mu}_i^u, \tilde{\mu}_i^\ell$ is equal to 1. Therefore, the gradients

$$\{\nabla h_i(x_*)\}_{i\in\tilde{I}},\;\; \{e_i\}_{i\in\tilde{I}_u},\;\; \{-e_i\}_{i\in\tilde{I}_\ell}$$

are positively linearly dependent. By the CPLD condition the gradients

$$\{\nabla h_i(x)\}_{i\in\tilde{I}},\;\; \{e_i\}_{i\in\tilde{I}_u},\;\; \{-e_i\}_{i\in\tilde{I}_\ell}$$

must be linearly dependent in a neighborhood of x_*. This contradicts (36). So, the theorem is proved. □

5 Boundedness of the penalty parameters

In this section we assume that the sequence $\{x_k\}$, generated by Algorithm 1 or by Algorithm 2, converges to a KKT point $x_* \in \Omega$. To simplify the arguments, as in [11], we assume without loss of generality that $[x_*]_i < u_i$ for all $i = 1, \ldots, n$ and that $\ell_i = 0$ for all $i = 1, \ldots, n$. The Lagrange multipliers associated to x_* will be denoted $\lambda_* \in I\!R^m$.

We will assume that $F'(x)$ and $\nabla^2 h_i(x)$ exist and are Lipschitz-continuous for all $x \in \Omega$.

Finally, we will also make use of the following *Nonsingularity Assumption*:

Assumption NS *Define*

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

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

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 $K \subset J_2$.

Assumption NS, which corresponds to Assumption AS5 of [11], will be supposed to be true all along the section. Moreover, we will also assume that the computation of $\bar{\lambda}_k$ at Step 2 of both algorithms is:

$$[\bar{\lambda}_k]_i = \max\{\bar{\lambda}_{min}, \min\{\bar{\lambda}_{max}, \{[\lambda_k]_i\}\}\} \tag{38}$$

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

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

$$[\bar{\lambda}_{min}]_i < [\lambda_*]_i < [\bar{\lambda}_{max}]_i \ \ \forall \ \ i = 1, \ldots, m.$$

Lemma 5.1. *Assume that the sequence $\{x_k\}$ is generated by Algorithm 1 and that, for all $k \in I\!N$,*

$$\lambda_{k+1} = \bar{\lambda}_k + \rho_k h(x_k).$$

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 $k \geq k_0$,

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq a_1 \varepsilon_k + a_2 \|x_k - x_*\|_\infty. \tag{39}$$

Moreover, if $\rho_{k_0} \geq \bar{\rho}$, we have:

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

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq a_5\varepsilon_k + a_6 \frac{\|\bar{\lambda}_k - \lambda_*\|_\infty}{\rho_k} \tag{40}$$

and

$$\|h(x_k)\|_\infty \leq a_5\varepsilon_k \frac{1}{\rho_k} + (1 + \frac{a_6}{\rho_k}) \frac{\|\bar{\lambda}_k - \lambda_*\|_\infty}{\rho_k}. \tag{41}$$

Proof. The proof is identical to the ones of Lemmas 4.3 and 5.1 of [11], replacing μ_k by $1/\rho_k$ and using the equivalence of norms in $I\!R^n$. □

Lemma 5.2. *Assume that the sequence $\{x_k\}$ is generated by Algorithm 1. Then, there exists $k_0 \in I\!N$ such that for all $k \geq k_0$,*

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

Proof. By (39) there exists $k_1 \in I\!N$ tal que

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq a_1\varepsilon_k + a_2\|x_k - x_*\|_\infty \text{ for all } k \geq k_1. \tag{42}$$

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 (42) we obtain that there exists $k_0 \geq k_1$ such that

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq \epsilon \text{ for all } k \geq k_0.$$

By the definition of ϵ we obtain the desired result. □

Theorem 5.1 *Assume that the sequence $\{x_k\}$ is generated by Algorithm 1 and that ε_k is such that*

$$\varepsilon_k = min\{\varepsilon'_k, \|h(x_k)\|_\infty\} \tag{43}$$

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 5.2. Then, for all $k \geq k_0$, we have that $\bar{\lambda}_k = \lambda_k$.

Assume that $\rho_k \to \infty$. By (41) and (43) there exists $k_1 \geq k_0$ such that

$$\|h(x_k)\|_\infty \leq (1 + \frac{a_6}{\rho_k})(\frac{1}{1 - \frac{a_5}{\rho_k}}) \frac{\|\lambda_k - \lambda_*\|_\infty}{\rho_k} \text{ for all } k \geq k_1 \tag{44}$$

whenever $\frac{a_5}{\rho_k} < 1$.

Since $\lambda_k = \lambda_{k-1} + \rho_{k-1}h(x_{k-1})$ we get

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

Then, by (40) and (43),

$$\|\lambda_k - \lambda_*\|_\infty \leq \frac{1}{{a_6}^{-1} - {\rho_{k-1}}^{-1}}(1 + \frac{a_5}{a_6})\|h(x_{k-1})\|_\infty. \tag{45}$$

Replacing (45) in (44) we obtain:

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

whenever

$$m_k = \frac{\tilde{m}}{\rho_k},$$

where $\tilde{m}$ is a positive constant.

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

From now on, if the sequence $\{x_k\}$ is generated by Algorithm 2, we define

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

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

The following result corresponds to Lemmas 4.3 and 5.1 of [11], adapted for several penalty parameters.

Lemma 5.3. *Assume that the sequence $\{x_k\}$ is computed by Algorithm 2. 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 $k \ge k_0$,*

$$\|\lambda_{k+1} - \lambda_*\|_\infty \le b_1\varepsilon_k + b_2\|x_k - x_*\|_\infty. \tag{46}$$

If $[\rho_{k_0}]_i \ge \overline{\rho}$ for all $i \in I_\infty$, then

$$\|x_k - x_*\|_\infty \le \alpha_1\varepsilon_k + \alpha_2\eta_k + \alpha_3 \sum_{i\in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} \tag{47}$$

and

$$\|\lambda_{k+1} - \lambda_*\|_\infty \le \alpha_4\varepsilon_k + \alpha_5\|h(x_k)\|_\infty. \tag{48}$$

Proof. The proof of (46) is identical to the one of (39).

Using the definition of $G_2(x_k, \bar{\lambda}_k, \rho_k)$, defining I_F as in the formula (5.6) of [11] and taking $\Delta x_k = \|(x_k - x_*)_{[I_F]}\|_2$, we obtain the following version of the inequality (5.27) of Lemma 5.1 of [11]:

$$\left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\|_2 \le M(b_{14}\varepsilon_k + \|h(x_k)\|_2 + b_{11}(\Delta x_k)^2 + b_{12}\Delta x_k \varepsilon_k + b_{13}{\varepsilon_k}^2) \tag{49}$$

for all $k \ge k_0$, with $b_{11} = a_7 + a_9 + a_8 b_2$, $b_{12} = 2(a_7 + a_9) + a_8(b_{10} + b_2)$, $b_{13} = a_7 + a_9 + a_8 b_{10}$, $b_{10} = b_1 + b_2$ and for the constants a_7, a_8 and a_9 considered in the proof of our Lemma 5.1.

By (8) and (46),

$$|h_i(x_k)| = \frac{|[\lambda_{k+1}]_i - [\bar{\lambda}_k]_i|}{[\rho_k]_i} \le \frac{|[\lambda_{k+1}]_i - [\lambda_*]_i| + |[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}$$

$$\leq \frac{b_1\varepsilon_k + b_2\|x_k - x_*\|_2 + |[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}. \tag{50}$$

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 $\sum_{i=1}^n a_i^2 \leq (\sum_{i=1}^n a_i)^2$ for $a_i \geq 0, i = 1,\ldots,n$, and the inequality (50) for all $i \in I_\infty$, we obtain:

$$\|h(x_k)\|_2 \leq \eta_k + \sum_{i\in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + |I_\infty| \frac{(b_1\varepsilon_k + b_2\|x_k - x_*\|_2)}{\rho_k}. \tag{51}$$

Recall that

$$\|x_k - x_*\|_2 \leq \Delta x_k + \varepsilon_k. \tag{52}$$

By (52), replacing (51) in (49), we get

$$\left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\|_2 \leq M(b_{14}\varepsilon_k + \eta_k + \sum_{i\in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}$$

$$+\tilde{b}_{10}\frac{\varepsilon_k}{\rho_k} + b_2|I_\infty|\frac{\Delta x_k}{\rho_k} + b_{11}{\Delta x_k}^2 + b_{12}\Delta x_k\varepsilon_k + b_{13}{\varepsilon_k}^2), \tag{53}$$

where $\tilde{b}_{10} = |I_\infty|(b_1 + b_2)$.

Now, if k is large enough,

$$\varepsilon_k \leq \min\{1, \frac{1}{4Mb_{12}}\} \text{ and } \Delta x_k \leq \frac{1}{4Mb_{11}}. \tag{54}$$

Define $\bar{\rho} = \max\{1, 4|I_\infty|Mb_2\}$. If k is large enough, $[\rho_k]_i \geq \bar{\rho}$ for all $i \in I_\infty$. By (53) and (54) we get:

$$\Delta x_k = \|(x_k - x_*)_{[I_F]}\|_2 \leq 4M((b_{14} + \tilde{b}_{10} + b_{13})\varepsilon_k + \eta_k + \sum_{i\in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}).$$

So, by (52), using the equivalence of norms in $I\!R^n$, we obtain:

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

for suitable constants α_1, α_2 and α_3. This proves (47).

Let us now prove (48). Using (54) in the inequality (49) we obtain:

$$\Delta x_k = \|(x_k - x_*)_{[I_F]}\|_2 \leq \frac{\Delta x_k}{2} + M(b_{15}\varepsilon_k + \|h(x_k)\|_2),$$

where $b_{15} = b_{13} + b_{14}$. Therefore,

$$\Delta x_k \leq 2M(b_{15}\varepsilon_k + \|h(x_k)\|_2). \tag{55}$$

By (52), (55) and the equivalence of norms in $I\!R^n$, we obtain:

$$\|x_k - x_*\|_\infty \le \alpha_6 \varepsilon_k + \alpha_7 \|h(x_k)\|_\infty \tag{56}$$

for suitable constants α_6 and α_7.

Replacing (56) in (46) and using again the equivalence of norms in $I\!R^n$, we obtain the inequality

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

for suitable constants α_4 and α_5. Then, (48) is proved. □

Lemma 5.4. *Assume that the sequence $\{x_k\}$ is computed by Algorithm 2. Then, there exists $k_0 \in I\!N$ such that, for all $k \ge k_0$,*

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

Proof. By (46), the proof is the same of Lemma 5.2. □

Theorem 5.2. *Assume that the sequence $\{x_k\}$ is computed by Algorithm 2 and that ε_k is such that*

$$\varepsilon_k = min\{\varepsilon_{k-1}, \|h(x_k)\|_\infty, \varepsilon'_k\} \tag{57}$$

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,

$$\eta_k = \sum_{i \in I_a} |h_i(x_k)| \le |I_a| \tau \|h(x_{k-1})\|_\infty. \tag{58}$$

Let $k \ge \tilde{k} = \max_{i \in I_a}\{k_0, k_1(i)\}$, where k_0 is obtained as in Lema 5.4.

By (8),

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

So, by (46),

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

Thus, by (47),

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

Now, by (48) with λ_k replacing λ_{k+1}, (57) implies that

$$|[\lambda_k]_i - [\lambda_*]_i| \le \|\lambda_k - \lambda_*\|_\infty \le (\alpha_4 + \alpha_5)\|h(x_{k-1})\|_\infty \quad i = 1, \ldots, m. \tag{59}$$

Since $\varepsilon_k \le \varepsilon_{k-1} \le \|h(x_{k-1})\|_\infty$, combining (58)-(59) we obtain:

$$|h_{i_0}(x_k)| \le m_k(i_0)\|h(x_{k-1})\|_\infty,$$

where

$$m_k(i_0) = \frac{\tilde{m}}{[\rho_k]_{i_0}}$$

and $\tilde{m} > 0$.

Since $m_k(i_0) \to 0$, there exists $\tilde{k}(i_0) \ge \tilde{k}$ such that

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

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

6 Numerical experiments

Our main objective regarding this set of experiments is to decide between Algorithm 1 and Algorithm 2. From the theoretical point of view, Algorithm 1 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 2. Thus, in principle, Algorithm 2 might converge to infeasible points more often than Algorithm 1. On the other hand, Algorithm 2 tends to increase the penalty parameters less frequently than Algorithm 1, a fact that has a positive influence on the conditioning of the subproblems.

However, we are also interested on 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 1.

   Option TWO: Algorithm 2.

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 1 and

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

   for Algorithm 2.

Therefore, 24 different methods are defined. Observe that, when $\varepsilon_{min} = 0$, the option ADPT corresponds to the theoretical hypotheses used in Section 4 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 (10) and (11) at Step 2 of both algorithms we used GENCAN [4] with its default parameters. The resulting code (Augmented Lagrangian with GENCAN) will be called ALGENCAN.

2. We computed the Lagrange multipliers estimates using (4), (8) and (38).

3. We set $\rho_1 = 10$ for Algorithm 1, $[\rho_1]_i = 10$ for all i for Algorithm 2 and $\gamma = 10$ for both algorithms.

4. The algorithms were stopped declaring *Convergence* when

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

   and

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

   We used $\varepsilon_{min} = 10^{-4}$.

5. 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 option -O4 to optimize the code.

We considered all the nonlinear programming problems with equality constraints and bounds of the CUTE collection [8]. As a whole, we tried to solve 128 problems.

Consider a fixed problem and let $x_{\text{final}}^{(M)}, M = 1, \ldots, 24$, 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 $t^{(M)}, M = 1, \ldots, 24$, be the computer CPU time that method M used to arrive to $x_{\text{final}}^{(M)}$. 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 fastests* method for the problem when

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

These definitions are the same used in [3] for comparing different Augmented Lagrangian formulae.

We are interested in comparing the 24 variants of Augmented Lagrangian algorithms with respect to Feasibility, Robustness and Efficiency. We say that a 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 fastests* method 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 fastests 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 2 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 2 is better than Algorithm 1. 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 1 does.

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)

Table 1: Performance of ALGENCAN

- In general, using a fixed small convergence criterion in the subproblems ($\varepsilon_k = \varepsilon_{min}$) is better than using inexact 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. It is not be very relevant the choice of $\bar{\lambda}_{max}$ between 10^6 and 10^{20}. 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 [11] 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 use 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 [14] 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 [18, 20, 21, 23, 26, 27, 36]) 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 1 seems to be strictly contained in the set of possible infeasible limit points of Algorithm 2. This could indicate that Algorithm 2 converges to infeasible points more frequently than Algorithm 1. However, this property was not confirmed by numerical experiments, which clearly indicate that Algorithm 2 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 3.2 can be obtained for Algorithm 2.

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 regularity assumptions (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 constraint qualifications under which convergence to KKT points can be guaranteed remains open.

The superiority of Algorithm 2 over Algorithm 1 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 [5, 6, 7, 12, 33, 34] 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 in mechanical applications [15, 16, 17]. 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 [15], 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. 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 [10].) However, many times the intersection of Ω with the general constraints $h(x) = 0$ is very complicate. 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 have been done in practical applications for many years. The convergence properties of these combinations using weak constraint qualifications deserve to be studied.

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 [13]). We compared the resulting algorithm with the PATH method for solving directly the original problem. On one hand, we confirmed the following warning of [22]: "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. We also acknowledge Sergio Ventura for computational insight.

References

[1] R. Andreani, J. M. Martínez and M. L. Schuverdt. *On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification.* To appear in Journal of Optimization Theory and Applications.

[2] D. P. Bertsekas. *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts, 1999.

[3] E. G. Birgin, R. Castillo and J. M. Martínez. *Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems.* To appear in Computational Optimization and Applications.

[4] 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).

[5] 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).

[6] 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).

[7] 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).

[8] 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).

[9] A. R. Conn, N. I. M. Gould and Ph. L. Toint. *LANCELOT: A Fortran package for large scale nonlinear optimization.* Springer-Verlag, Berlin, 1992.

[10] 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).

[11] 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).

[12] 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.* To appear in Journal of Optimization Theory and Applications.

[13] 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).

[14] E. D. Dolan and J. J. Moré. *Benchmarking optimization software with performance profiles*, Mathematical Programming 91, pp. 201–213 (2002).

[15] Z. Dostál. *Inexact semi-monotonic Augmented Lagrangians with optimal feasibility convergence for convex bound and equality constrained quadratic programming.* To appear in SIAM Journal on Numerical Analysis.

[16] 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).

[17] 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).

[18] 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).

[19] 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.

[20] 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).

[21] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementary Problems*, Volume II, Springer Series in Operation Research, Springer (2000).

[22] M. C. Ferris and T. S. Munson. *PATH 4.6 User Manual*, http://www.gams.com/solvers/solvers.htm#PATH.

[23] A. Fischer. *A special Newton-type optimization method*, Optimization 24, pp. 269–284 (1992).

[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] J-S. Pang. *Newton methods for B-differentiable equations*, Mathematics of Operations Research 13, pp. 311–341 (1990).

[30] J-S. Pang and L. Qi. *Nonsmooth equations: motivation and algorithms*, SIAM Journal on Optimization 3, pp. 443–465 (1993).

[31] L. Qi. *Convergence Analysis of some algorithms for solving nonsmooth equation*, Mathematics of Operations Research 18, pp. 227–244 (1993).

[32] 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).

[33] M. Raydan. *On the Barzilai and Borwein choice of steplength for the gradient method*, IMA Journal of Numerical Analysis 13, pp. 321–326 (1993).

[34] M. Raydan. *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, SIAM Journal on Optimization 7, pp. 26–33 (1997).

[35] R. T. Rockafellar. *Lagrange multipliers and optimality*, SIAM Review 35, pp. 183–238 (1993).

[36] N. Yamashita and M. Fukushima. *Nonsmooth equation based BFGS methods for solving KKT systems in mathematical programming*, Journal of Optimization Theory and Applications 100, pp 123–167 (2001).