The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems — Source link ↗

ErnestoG. Birgin, Damián Fernández, José Mario Martínez

Institutions: University of São Paulo, State University of Campinas

Topics: Penalty method, Augmented Lagrangian method, Lagrangian relaxation, Lagrange multiplier and Nonlinear programming

Related papers:

- On Augmented Lagrangian Methods with General Lower-Level Constraints

- Multiplier and gradient methods

- Augmented Lagrangian methods under the constant positive linear dependence constraint qualification

- A method for nonlinear constraints in minimization problems

- Global minimization using an Augmented Lagrangian method with variable lower-level constraints

Share this paper: 🅕 🅣 🅛 ✉

On the Boundedness of Penalty Parameters in an Augmented Lagrangian Method with Constrained Subproblems[*]

Ernesto G. Birgin [†]   Damián Fernández [‡]   J. M. Martínez[‡]

January 10, 2011

Abstract

Augmented Lagrangian methods are effective tools for solving large-scale nonlinear programming problems. At each outer iteration a minimization subproblem with simple constraints, whose objective function depends on updated Lagrange multipliers and penalty parameters, is approximately solved. When the penalty parameter becomes very large the subproblem is difficult, therefore the effectiveness of this approach is associated with boundedness of penalty parameters. In this paper it is proved that, under more natural assumptions than the ones up to now employed, penalty parameters are bounded. For proving the new boundedness result, the original algorithm has been slightly modified. Numerical consequences of the modifications are discussed and computational experiments are presented.

Key words: Nonlinear Programming, Augmented Lagrangian methods, Penalty parameters, Numerical experiments.

AMS Subject Classification: 90C30, 49K99, 65K05.

1 Introduction

We consider the nonlinear programming problem in the form:

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0, \ g(x) \leq 0, \ x \in \Omega, \tag{1}$$

where

$$\Omega = \{x \in I\!\!R^n \mid \underline{h}(x) = 0, \ \underline{g}(x) \leq 0\} \tag{2}$$

and $f : I\!\!R^n \to I\!\!R$, $h : I\!\!R^n \to I\!\!R^m$, $g : I\!\!R^n \to I\!\!R^p$, $\underline{h} : I\!\!R^n \to I\!\!R^{\underline{m}}$, $\underline{g} : I\!\!R^n \to I\!\!R^{\underline{p}}$ have continuous first derivatives.

The constraints defined by $h(x) = 0$ and $g(x) \leq 0$ will be included in the augmented Lagrangian definition. On the other hand, the constraints that define the set Ω are generally simple, in the sense that efficient methods for optimization within Ω are available. In many practical cases Ω is a box or a bounded polytope.

[†]Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, São Paulo SP, Brazil. e-mail: egbirgin@ime.usp.br

[‡]Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas, SP, Brazil. e-mail: {dfernan | martinez}@ime.unicamp.br

Given $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$, we define the Powell-Hestenes-Rockafellar (PHR) [11, 12, 20, 29, 32] augmented Lagrangian by:

$$L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left[\sum_{i=1}^{m} \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{i=1}^{p} \max \left\{ 0, g_i(x) + \frac{\mu_i}{\rho} \right\}^2 \right]. \qquad (3)$$

PHR-like augmented Lagrangian methods are based on the iterative minimization of $L_\rho(x, \lambda, \mu)$ with respect to $x \in \Omega$, followed by convenient updates of λ, μ and ρ. In this paper we consider variations of the implementation introduced in [1], by means of which it is possible to prove that feasible limit points that satisfy the Constant Positive Linear Dependence (CPLD) constraint qualification [3, 30] fulfill the Karush-Kuhn-Tucker (KKT) conditions[1]. The CPLD condition was introduced in [30] and its status as a constraint qualification was elucidated in [3]. This condition is weaker than the Mangasarian-Fromovitz (MFCQ) constraint qualification [25] and, thus, it is also weaker than the Linear Independence constraint qualification (LICQ) [17]. Therefore, first-order convergence results based on the CPLD constraint qualification are stronger than results based on LICQ or MFCQ.

The method introduced in [1] has been implemented as Algencan, a freely available software that can be found in *http://www.ime.usp.br/∼egbirgin/tango/*. Algencan is especially effective in problems with many inequality constraints and in problems in which the structure of the Lagrangian Hessian matrix is not suitable for sparse matrix factorizations. A version of Algencan for global optimization has been described in [8]. Modifications with convergence to points that satisfy second order optimality conditions were given in [2].

When the penalty parameter is very large, the subproblem associated with the minimization of the augmented Lagrangian may be very difficult, in the sense that unitary stabilized Newtonian steps corresponding to this problem may not provide decrease of the augmented Lagrangian, which makes it hard to evaluate progress during the subproblem resolution process. For this reason, it is important to develop strategies that preserve convergence maintaining moderate values of the penalty parameters. Under stronger conditions than the ones required for proving that limit points satisfy the KKT conditions, it was proved in [1] that penalty parameters remain bounded.

In the present paper we prove that, for a modified form of Algencan, penalty parameters remain bounded under more natural assumptions than the ones used in [1]. We will present numerical experiments showing that the modifications introduced here do not impair the performance of previous versions of Algencan. We will show that first-order convergence results proved in [1] are preserved for the new version of Algencan. From the point of view of penalty parameter boundedness, the differences between the new results and the ones proved in [1] are:

1. In [1] the LICQ condition is assumed to hold at a limit point x^*. Here we only assume that MFCQ holds and the vector of Lagrange multipliers associated with x^* is unique.

2. In [1] it is assumed that strict complementarity holds at x^*, with respect to the active constraints of type $g_i(x) \leq 0$. Moreover, for obtaining the boundedness results of [1], it is assumed that the Hessian of the Lagrangian is positive definite on the orthogonal subspace to the gradients of all active constraints. In the present paper we employ a second order sufficient optimality condition that does not involve strict complementarity at all.

The paper is organized as follows. In Section 2 we describe the PHR augmented Lagrangian method considered in this research and we prove that this method preserves the global convergence

---

[1]The CPLD condition says that whenever some gradients of active constraints are linearly dependent at a feasible point, and the coefficients corresponding to inequality constraints are non-negative, then the same gradients remain linearly dependent in a neighborhood of the point.

properties of [1]. In Section 3 we prove boundedness of the penalty parameter and some results on the speed of convergence. Section 4 is devoted to numerical questions and implementation. Conclusions and lines for future research are stated in Section 5.

Notation. Vectors will be denoted by columns. For any vector v with components v_i we denote v_+ the vector whose components are $\max\{0, v_i\}$. Given $h : I\!\!R^n \to I\!\!R^m$, we denote $\nabla h(x) = (\nabla h_1(x), \dots, \nabla h_m(x))$. The symbol $\|\cdot\|$ denotes an arbitrary norm. The closed ball $\{x \in I\!\!R^n \mid \|x - x^*\| \le \delta\}$ will be denoted $B(x^*, \delta)$. We denote $I\!\!R_+ = \{x \in I\!\!R^n \mid x \ge 0\}$. If $v, w \in I\!\!R^n$ we denote $\min\{v, w\} \in I\!\!R^n$ the vector whose components are $\min\{v_i, w_i\}$. We write $\varepsilon_k \downarrow 0$ to denote that $\{\varepsilon_k\}$ is a sequence of nonnegative numbers that tends to zero.

2 Algorithm

Given $x \in I\!\!R^n$, $\lambda \in I\!\!R^n$, $\mu \in I\!\!R_+^p$, $v \in I\!\!R^{\underline{m}}$, $w \in I\!\!R_+^{\underline{p}}$ we define the Lagrangian $\mathcal{L}(x, \lambda, \mu, v, w)$ by

$$\mathcal{L}(x, \lambda, \mu, v, w) = \mathcal{L}_{\text{upper}}(x, \lambda, \mu) + \sum_{i=1}^{\underline{m}} v_i \underline{h}_i(x) + \sum_{i=1}^{\underline{p}} w_i \underline{g}_i(x),$$

where

$$\mathcal{L}_{\text{upper}}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{p} \mu_i g_i(x).$$

We also define:

$$S(x, \lambda, \mu, v, w) = \begin{pmatrix} \nabla \mathcal{L}(x, \lambda, \mu, v, w) \\ h(x) \\ \min\{-g(x), \mu\} \\ \underline{h}(x) \\ \min\{-\underline{g}(x), w\} \end{pmatrix} \tag{4}$$

and $\sigma(x, \lambda, \mu, v, w) = \|S(x, \lambda, \mu, v, w)\|$. The Karush-Kuhn-Tucker (KKT) conditions may be written:

$$\sigma(x, \lambda, \mu, v, w) = 0.$$

Algorithm 2.1. Let $\varepsilon_k \downarrow 0$, $\lambda_{\min} \le \lambda_{\max}$, $\mu_{\max} \ge 0$, $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^1 \in [0, \mu_{\max}]^p$, $\rho_1 > 0$, $r \in (0, 1)$, $\gamma > 1$. Initialize $k \leftarrow 1$.

Step 1. Compute x^k as an approximate stationary point of

$$\text{Minimize } L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k) \quad \text{subject to} \quad \underline{h}(x) = 0, \;\; \underline{g}(x) \le 0. \tag{5}$$

We use, as approximate minimization criterion, the existence of $v^k \in I\!\!R^{\underline{m}}, w^k \in I\!\!R_+^{\underline{p}}$, such that:

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) + \nabla \underline{h}(x^k) v^k + \nabla \underline{g}(x^k) w^k\| \le \varepsilon_k, \tag{6}$$

$$\|\underline{h}(x^k)\| \le \varepsilon_k, \;\; \|\min\{-\underline{g}(x^k), w^k\}\| \le \varepsilon_k. \tag{7}$$

Step 2. Compute

$$\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k) \tag{8}$$

and

$$\mu^{k+1} = (\bar{\mu}^k + \rho_k g(x^k))_+. \tag{9}$$

3

Step 4. If $\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) = 0$, stop. (In this case x^k satisfies the KKT conditions of the problem.)

Step 5. If $k = 1$ or

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \le r\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}) \tag{10}$$

we define

$$\rho_{k+1} \ge \rho_k. \tag{11}$$

Else, we define

$$\rho_{k+1} \ge \gamma \rho_k. \tag{12}$$

Step 6. For all $i = 1, \ldots, m$, $j = 1, \ldots, p$, compute $\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{max}]$ and $\bar{\mu}_j^{k+1} \in [0, \mu_{max}]$ in such a way that

$$\bar{\lambda}_i^{k+1} = \lambda_i^{k+1} \quad \text{if} \quad \lambda_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}] \tag{13}$$

and

$$\bar{\mu}_j^{k+1} = \mu_j^{k+1} \quad \text{if} \quad \mu_j^{k+1} \in [0, \mu_{\max}]. \tag{14}$$

Set $k \leftarrow k + 1$ and go to Step 1.

In the following lemma we give a sufficient condition for well-definiteness of the sequence generated by Algorithm 2.1.

Lemma 2.1 *Assume that Ω is bounded and that, for all $x \in \Omega$ a constraint qualification with respect to the constraints $\underline{h}(x) = 0, \underline{g}(x) \le 0$ is satisfied. Then, for all $k = 1, 2, \ldots$ there exists $x^k \in \mathbb{R}^n$ satisfying (6) and (7).*

Proof. Let x be a solution of (5). The existence of x is guaranteed by the compactness of Ω and the continuity of $L_{\rho_k}, \underline{h}, \underline{g}$. Since x satisfies a constraint qualification, the KKT conditions are fulfilled at this point. This implies that (6) and (7) hold with $x^k = x$. $\qquad\square$

Note that x^k does not need to be a global minimizer of (5), since we only need the fulfillment of (6) and (7). If Ω is simple enough there exist practical algorithms that find KKT points of (5) with arbitrarily high precision. For example, in Algencan, when Ω is a box, one employs the globally convergent algorithm Gencan [9] that satisfies the desired convergence requirements.

The main difference between Algorithm 2.1 and Algorithm 3.1 of [1] is that in [1] one uses the test

$$\left\| \begin{pmatrix} h(x^k) \\ \min\{-g(x^k), \bar{\mu}^k/\rho_k\} \end{pmatrix} \right\| \le r \left\| \begin{pmatrix} h(x^{k-1}) \\ \min\{-g(x^{k-1}), \bar{\mu}^{k-1}/\rho_{k-1}\} \end{pmatrix} \right\| \tag{15}$$

instead of (10) to decide how to update the penalty parameter ρ_k. Note that, by (6), (7), (8) and (9), one has that $\|\nabla \mathcal{L}(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k)\| \le \varepsilon_k$, $\|\underline{h}(x^k)\| \le \varepsilon_k$, and $\|\min\{-g(x^k), w^k\}\| \le \varepsilon_k$ for all k. Therefore, by (4), since ε_k tends to zero, the test (10) essentially verifies feasibility and complementarity with respect to the penalized constraints. Dual-feasibility progress and feasibility-complementarity improvement with respect to the simple subproblem constraints is guaranteed by (6) and (7), respectively.

Another minor difference is that, in [1] one defines $\rho_{k+1} = \rho_k$ instead of (11) and $\rho_{k+1} = \gamma \rho_k$ instead of (12). Obviously, (11) and (12) are more general than the corresponding choices in [1]. We decided to use (11-12) here because we want to consider the possibility of increasing the penalty parameter even in situations in which (10) holds. The use of (11) and (12) does not alter at all

the global convergence results of [1].

Lemma 2.2. *Assume that $\{x^k\}$ is generated by Algorithm 2.1. Then, for all $k = 1, 2, \ldots,$ $i = 1, \ldots, p$, we have:*

$$|\min\{-g_i(x^k), \mu_i^{k+1}\}| \leq \frac{|\mu_i^{k+1} - \bar{\mu}_i^k|}{\rho_k}. \tag{16}$$

Proof. If $g_i(x^k) \geq 0$, since $\mu_i^{k+1} \geq 0$ we have that

$$\min\{-g_i(x^k), \mu_i^{k+1}\} = -g_i(x^k). \tag{17}$$

Since $\bar{\mu}_i^k + \rho_k g_i(x^k) \geq 0$, we have that $\mu^{k+1} = \bar{\mu}_i^k + \rho_k g_i(x^k)$ in this case. Therefore,

$$\frac{\mu_i^{k+1} - \bar{\mu}_i^k}{\rho_k} = g_i(x^k). \tag{18}$$

By (17) and (18), the thesis holds in the case $g_i(x^k) \geq 0$.

If $g_i(x^k) \leq -\bar{\mu}_i^k/\rho_k$, we have that $\mu^{k+1} = 0$. Therefore, $\min\{-g_i(x^k), \mu_i^{k+1}\} = 0$ and (16) is also fulfilled. Consider, finally, the case in which

$$-\bar{\mu}_i^k/\rho_k < g_i(x^k) < 0.$$

In this case $\mu^{k+1} = \bar{\mu}_i^k + \rho_k g_i(x^k)$ and (18) takes place. Therefore, if (17) holds, the thesis follows trivially. It remains to consider the case in which

$$\min\{-g_i(x^k), \mu_i^{k+1}\} = \mu_i^{k+1}. \tag{19}$$

Then, $\mu_i^{k+1} \leq |g_i(x^k)|$. So, by (18) and (19), the thesis also follows in this case. $\qquad \square$

Lemma 2.3. *Assume that $\{x^k\}$ is generated by Algorithm 2.1 and $\{x^k\}_{k \in K}$ is a subsequence that converges to $x^* \in \mathbb{R}^n$. Assume that $g_i(x^*) < 0$ for all $i \in I$ and $\underline{g}_j(x^*) < 0$ for all $j \in J$. Then, there exists a sequence $\{\varepsilon_k'\}_{k \in K}$ such that $\varepsilon_k' \downarrow 0$ and, for all $k \in K$ large enough:*

$$\|\nabla f(x^k) + \nabla h(x^k)\lambda^{k+1} + \sum_{i \notin I} \mu_i^{k+1} \nabla g_i(x^k) + \nabla \underline{h}(x^k)v^k + \sum_{j \notin J} w_j^k \nabla \underline{g}_j(x^k)\| \leq \varepsilon_k', \tag{20}$$

$$\|\underline{h}(x^k)\| \leq \varepsilon_k' \quad and \quad \|\underline{g}(x^k)_+\| \leq \varepsilon_k'. \tag{21}$$

Moreover, if, in addition, the point x^* is feasible and satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ), the sequences $\{\lambda^{k+1}, \mu^{k+1}, v^k, w^k\}_{k \in K}$ are bounded. In this case, x^* satisfies the KKT conditions and, if there is only one vector of multipliers $(\lambda^*, \mu^*, v^*, w^*)$ associated with x^*, we have:

$$\lim_{k \in K} (\lambda^{k+1}, \mu^{k+1}, v^k, w^k) = (\lambda^*, \mu^*, v^*, w^*). \tag{22}$$

Proof. By (6), (8) and (9) we have that

$$\|\nabla \mathcal{L}_{\text{upper}}(x^k, \lambda^{k+1}, \mu^{k+1}) + \nabla \underline{h}(x^k)v^k + \nabla \underline{g}(x^k)w^k\| \leq \varepsilon_k \tag{23}$$

for all $k \in K$. By (7) and the continuity of $\underline{g}$ we have that $\lim_{k \in K} w_j^k = 0$ for all $j \in J$. Therefore, by the continuity of $\nabla \underline{g}$,

$$\lim_{k \in K} \nabla \underline{g}_j(x^k)w_j^k = 0$$

5

for all $j \in J$. Therefore, by (23),

$$\lim_{k \in K} \|\nabla \mathcal{L}_{\text{upper}}(x^k, \lambda^{k+1}, \mu^{k+1}) + \nabla \underline{h}(x^k) v^k + \sum_{j \notin J} \nabla \underline{g}_j(x^k) w_j^k\| = 0. \tag{24}$$

Now suppose that $\{\rho_k\}$ is bounded. In this case, by (10), we have that

$$\lim_{k \in K} \min\{-g_i(x^k), \mu_i^{k+1}\} = 0$$

for all $i = 1, \ldots, p$. Since $g_i(x^*) < 0$ for all $i \in I$, by the continuity of g we have that

$$\lim_{k \in K} \mu_i^{k+1} = 0 \quad \text{for all} \quad i \in I.$$

By the continuity of ∇g this implies that

$$\lim_{k \in K} \nabla g_i(x^k) \mu_i^{k+1} = 0 \quad \text{for all} \quad i \in I.$$

Then, by (24),

$$\lim_{k \in K} \|\nabla f(x^k) + \nabla h(x^k) \lambda^{k+1} + \sum_{i \notin I} \nabla g_i(x^k) \mu_i^{k+1} + \nabla \underline{h}(x^k) v^k + \sum_{j \notin J} \nabla \underline{g}_j(x^k) w_j^k\| = 0. \tag{25}$$

Finally, suppose that $\{\rho_k\}$ tends to infinity. Since $\{\bar{\mu}^k\}$ is bounded and $g_i(x^*) < 0$ for $i \in I$, we have that $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$ for $i \in I$ and $k \in K$ large enough. This implies that, $\mu_i^{k+1} = 0$ for $i \in I$ and $k \in K$ large enough. This completes the proof of the first part of the Lemma.

For proving the second part, assume that the sequence $\{\lambda^{k+1}, \mu^{k+1}, v^k, w^k\}_{k \in K}$ is not bounded. Therefore, $\lim_{k \in K} M_k = \infty$, where

$$M_k = \max\{\|\lambda^{k+1}\|_\infty, \|\mu^{k+1}\|_\infty, \|v^k\|_\infty, \|w^k\|_\infty\}.$$

Then, the sequence $\{\lambda^{k+1}/M_k, \mu^{k+1}/M_k, v^k/M_k, w^k/M_k\}_{k \in K}$ is bounded. It turns out that some subsequence is convergent and its limit cannot be null, since, by the choice of M_k infinitely many elements have modulus equal to 1. Taking limits for this subsequence and using the first part of the theorem, we obtain that MFCQ cannot hold.

Then, if MFCQ holds, every limit of a convergent subsequence of $(\lambda^{k+1}, \mu^{k+1}, v^k, w^k)$ defines a set of KKT multipliers. Therefore, (22) holds in the case that the multipliers are unique. $\qquad \square$

Lemma 2.3 makes it easy to prove the following global convergence theorem employing several arguments used in Theorems 4.1 and 4.2 of [1].

Theorem 2.1. *Assume that $\{x^k\}$ is a sequence generated by Algorithm 2.1 and x^* is a limit point. Then:*

1. *The subproblem constraints $\underline{h}(x) = 0, \underline{g}(x) \leq 0$ hold at x^*. Moreover, if $\{\rho_k\}$ is bounded, x^* is feasible.*

2. *If x^* satisfies the CPLD constraint qualification with respect to the subproblem constraints, then the KKT conditions of the problem*

$$\textit{Minimize } \|h(x)\|_2^2 + \|g(x)_+\|_2^2 \quad \textit{subject to} \quad \underline{h}(x) = 0, \underline{g}(x) \leq 0 \tag{26}$$

hold at x^*.

6

3. *If x^* is feasible and satisfies the CPLD constraint qualification with respect to all the constraints of problem (1), then x^* fulfills the KKT conditions of this problem.*

Proof. By (7), since ε_k tends to zero, we have that $\underline{h}(x^*) = 0, \underline{g}(x^*) = 0$. If the sequence $\{\rho_k\}$ is bounded, then, for all k large enough, (10) holds. Therefore,

$$\lim_{k \to \infty} \sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) = 0.$$

Thus,

$$\lim_{k \to \infty} \|h(x^k)\| = \lim_{k \to \infty} \|\underline{h}(x^k)\| = \lim_{k \to \infty} \|\min\{-g(x^k), \mu^{k+1}\}\| = \lim_{k \to \infty} \|\min\{-\underline{g}(x^k), w^k\}\| = 0.$$

This implies that

$$\|h(x^*)\| = \|\underline{h}(x^*)\| = \|g(x^*)_+\| = \|\underline{g}(x^*)_+\| = 0.$$

Thus, x^* is feasible.

Let us prove now the third part of the thesis. Assume that K is an infinite sequence of indices such that $\lim_{k \in K} x^k = x^*$ and x^* is feasible. By Lemma 2.3, (20) holds for $k \in K$ large enough (say, $k \in K_1$). Define, for all $k \in K_1$,

$$G_k = \nabla h(x^k)\lambda^{k+1} + \sum_{i \notin I} \mu_i^{k+1} \nabla g_i(x^k) + \nabla \underline{h}(x^k) v^k + \sum_{j \notin J} w_j^k \nabla \underline{g}_j(x^k). \tag{27}$$

By (20) we have that

$$\|\nabla f(x^k) + G_k\| \le \varepsilon_k' \tag{28}$$

for all $k \in K_1$. By the convergence of $\{x^k\}$ for $k \in K_1$ and the continuity of ∇f, we have that $\{G_k\}$ is bounded.

By Caratheodory's Theorem ([6], page 689), for all $k \in K_1$ there exist $\widehat{\lambda}^{k+1} \in I\!\!R^m$, $\widehat{\mu}^{k+1} \in I\!\!R^p_+$, $\widehat{v}^k \in I\!\!R^{\underline{m}}$, $\widehat{w}^k \in I\!\!R^{\underline{p}}_+$, $I_1(k) \subseteq \{1, \ldots, m\}$, $I_2(k) \subseteq \{1, \ldots, p\} - I$, $I_3(k) \subseteq \{1, \ldots, \underline{m}\}$, $I_4(k) \subseteq \{1, \ldots, \underline{p}\} - J$ such that

$$G_k = \sum_{i \in I_1(k)} \widehat{\lambda}_i^{k+1} \nabla h_i(x^k) + \sum_{i \in I_2(k)} \widehat{\mu}_i^{k+1} \nabla g_i(x^k) + \sum_{i \in I_3(k)} \widehat{v}^k \nabla \underline{h}_i(x^k) + \sum_{i \in I_4(k)} \widehat{w}_j^k \nabla \underline{g}_j(x^k) \tag{29}$$

and the gradients $\nabla h_i(x^k), i \in I_1(k)$, $\nabla g_i(x^k), i \in I_2(k)$, $\nabla \underline{h}_i(x^k), i \in I_3(k)$ and $\nabla \underline{g}_i(x^k), i \in I_4(k)$ are linearly independent.

Taking an appropriate subsequence $K_2 \subseteq K_1$ we may assume that $I_j = I_j(k)$ for all $k \in K_2$, $j = 1, 2, 3, 4$.

Define, for all $k \in K_1$,

$$b_k = \max\{\max\{|\widehat{\lambda}_i^{k+1}|, i \in I_1\}, \max\{|u_i^k|, i \in I_3\}, \max\{\widehat{\mu}_i^{k+1}, i \in I_2\}, \max\{\widehat{v}_i^k, i \in I_4\}\}.$$

If $\{b_k\}$ is bounded, taking limits on both sides of (28) and using (29) for an appropriate subsequence, we obtain the KKT condition.

Let us assume now that the CPLD constraint qualification relative to all the constraints of (1) is satisfied at x^*. Assume that $\{b_k\}$ is unbounded. By (29), dividing both sides of (28) by $\{b_k\}$, we obtain:

$$\frac{G_k}{b_k} = \sum_{i \in I_1} \frac{\widehat{\lambda}_i^{k+1}}{b_k} \nabla h_i(x^k) + \sum_{i \in I_2} \frac{\widehat{\mu}_i^{k+1}}{b_k} \nabla g_i(x^k) + \sum_{i \in I_3} \frac{\widehat{v}^k}{b_k} \nabla \underline{g}_i(x^k) + \sum_{i \in I_4} \frac{\widehat{w}_j^k}{b_k} \nabla \underline{g}_i(x^k). \tag{30}$$

Since $\{G_k\}$ is bounded, the left-hand side of (30) tends to zero. Therefore, taking limits for an appropriate subsequence we obtain that a nontrivial linear combination of the gradients $\nabla h_i(x^*), i \in I_1$, $\nabla g_i(x^*), i \in I_2$, $\nabla \underline{h}_i(x^*), i \in I_3$ and $\nabla \underline{g}_i(x^*), i \in I_4$, is null, with non-negative coefficients corresponding to the gradients $\nabla g_i(x^*)$ and $\nabla \underline{g}_i(x^*)$. Thus, by the CPLD condition, the gradients $\nabla h_i(x), i \in I_1$, $\nabla g_i(x), i \in I_2$, $\nabla \underline{h}_i(x), i \in I_3$ and $\nabla \underline{g}_i(x), i \in I_4$ are linearly dependent in a neighborhood of x^*. This contradicts the linear independence of the gradients that appear in the second term of (29).

Finally, let us prove the second part of the thesis. By (8), (9) and (20), if $\lim_{k \in K} x^k = x^*$, for $k \in K$ large enough we have:

$$\|\nabla f(x^k) + \sum_{i=1}^m (\bar{\lambda}_i^k + \rho_k h_i(x^k))\nabla h_i(x^k) + \sum_{g_i(x^*) \geq 0} (\bar{\mu}_i^k + \rho_k g_i(x^k))_+ \nabla g_i(x^k)$$

$$+ \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{\underline{g}_i(x^*) \geq 0} w_i^k \nabla \underline{g}_i(x^k)\| \leq \varepsilon_k'. \tag{31}$$

If $\{\rho_k\}$ is bounded the thesis follows from the first part of the theorem. Assuming that ρ_k tends to infinity, dividing both sides of (31) by ρ_k, using that $\bar{\lambda}^k$ and $\bar{\mu}^k$ are bounded and that $f, \nabla h, \nabla g$ are continuous we deduce that there exists a sequence $\varepsilon_k'' \downarrow 0$ such that

$$\|\sum_{i=1}^m h_i(x^k)\nabla h_i(x^k) + \sum_{g_i(x^*) \geq 0} g_i(x^k)_+ \nabla g_i(x^k) + \sum_{i=1}^{\underline{m}} (v_i^k)' \nabla \underline{h}_i(x^k) + \sum_{\underline{g}_i(x^*) \geq 0} (w_i^k)' \nabla \underline{g}_i(x^k)\| \leq \varepsilon_k'',$$
$$\tag{32}$$

where $(v_i^k)' = v_i^k/\rho_k$ for all $i = 1, \ldots, \underline{m}$ and $(w_i^k)' = w_i^k/\rho_k$ for all i such that $\underline{g}_i(x^*) \geq 0$.

Let us define $\Phi(x) = \frac{1}{2}(\|h(x)\|_2^2 + \|g(x)_+\|_2^2)$. Then, by (32),

$$\|\nabla \Phi(x^k) + \sum_{i=1}^{\underline{m}} (v_i^k)' \nabla \underline{h}_i(x^k) + \sum_{\underline{g}_i(x^*) \geq 0} (w_i^k)' \nabla \underline{g}_i(x^k)\| \leq \varepsilon_k''. \tag{33}$$

Assume that the CPLD condition relative to the constraints that define Ω is satisfied at x^*. Then, we deduce that x^* is a KKT point of (26) as in the proof of the third part of the thesis, using Caratheodory's Theorem in (33). $\qquad \square$

3 Penalty Boundedness and Order of Convergence

In this section we use the following assumptions.

Assumption A1. The set Ω is compact and a constraint qualification with respect to $\underline{h}(x) = 0, \underline{g}(x) \leq 0$ is satisfied for all $x \in \Omega$. (Then, by Lemma 2.1, the sequence $\{x^k\}$, generated by Algorithm 2.1, is well defined.)

Assumption A2. $\{x^k\}$ converges to a feasible point x^*.

Assumption A3. The Mangasarian-Fromovitz constraint qualification is satisfied at x^* and there is only one vector $(\lambda^*, \mu^*, v^*, w^*)$ of associated multipliers.

Assumption A4. The functions $f, h, g, \underline{h}, \underline{g}$ are twice continuously differentiable at x^* and the second order sufficient optimality condition is fulfilled at x^*. This means that KKT conditions hold at x^* and that there exist associated Lagrange multipliers $\{\lambda^*, \mu^*, v^*, w^*\}$ such that

$$d^T \nabla^2 \mathcal{L}(x^*, \lambda^*, \mu^*, v^*, w^*) d > 0$$

for all non-null vector $d \in \mathbb{R}^n$ satisfying

$$\nabla f(x^*)^T d \leq 0,$$

$$\nabla h(x^*)^T d = 0, \nabla g_i(x^*)^T d \leq 0 \text{ for } i \in \{1, \ldots, p\} \text{ with } g_i(x^*) = 0,$$

$$\nabla \underline{h}(x^*)^T d = 0, \nabla \underline{g}_i(x^*)^T d \leq 0 \text{ for } i \in \{1, \ldots, \underline{p}\} \text{ with } \underline{g}_i(x^*) = 0.$$

Assumption A5. For all $k = 2, 3, \ldots$, we define $\varepsilon_k \to 0$ in such a way that

$$\varepsilon_k \leq \chi(\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1})),$$

where $\chi : \mathbb{R}^+ \to \mathbb{R}^+$ is such that $\lim_{t \to 0} \chi(t)/t = 0$.

Assumption A6. For all $k = 1, 2, \ldots$, when (10) holds, we choose $\rho_{k+1} = \rho_k$ in (11).

Remark. Assume that the feasible point x^* is such that $g_i(x^*) = 0$ if and only if $i \in A \subseteq \{1, \ldots, p\}$, $\underline{g}(x^*) < 0$ and x^* satisfies KKT with multipliers $\lambda^*, \mu^*, v^*, w^*$. Suppose that x^* satisfies the conditions employed in [1] to prove boundedness of the penalty parameters. Let us prove that, in this case, Assumption A4 necesarily holds. Let $d \in \mathbb{R}^n$, $d \neq 0$, be such that

$$\nabla f(x^*)^T d \leq 0 \tag{34}$$

and

$$\nabla h(x^*)^T d = 0, \nabla g_i(x^*)^T d \leq 0 \text{ for } i \in A, \nabla \underline{h}(x^*)^T d = 0. \tag{35}$$

By the KKT conditions,

$$\nabla f(x^*) + \nabla h(x^*) \lambda^* + \nabla \underline{h}(x^*) v^* + \sum_{i \in A} \nabla g(x^*) \mu_i^* = 0. \tag{36}$$

By the strict complementarity assumption used in [1], we have that $\mu^* > 0$. Thus, by (35) and (36), if there exists $i \in A$ such that $\nabla g_i(x^*)^T d < 0$, we obtain that $\nabla f(x^*)^T d > 0$, contradicting (34). Therefore, $\nabla g_i(x^*)^T d = 0$ for all $i \in A$. Thus, the second order assumption of [1] implies that

$$d^T \nabla^2 \mathcal{L}(x^*, \lambda^*, \mu^*, v^*, w^*) d > 0.$$

This means that Assumption A4 holds. On the other hand, Assumption A4 is satisfied in trivial situations in which the assumptions of [1] are not fulfilled. For example, consider the problem

$$\text{Minimize } x^2 \text{ subject } g_1(x) \leq 0,$$

with $g_1(x) \equiv x$. In this case, the strict complementarity assumption of [1] does not hold, but Assumption A4 is trivially verified.

Lemma 3.1. *Let Assumptions A1, A2 and A3 hold. Then,*

$$\lim_{k \to \infty} \lambda^k = \lambda^*, \ \lim_{k \to \infty} \mu^k = \mu^*, \ \lim_{k \to \infty} v^k = v^*, \ \lim_{k \to \infty} w^k = w^*.$$

Proof. The result follows from (22) using the convergence of the whole sequence $\{x^k\}$. □

Lemma 3.2. *Let Assumptions A1, A2, A3 and A4 hold. Then, there exist $k_0 \in \{1, 2, \ldots\}$, $\beta_1, \beta_2 > 0$ such that, for all $k \geq k_0$,*

$$\beta_1 \|(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) - (x^*, \lambda^*, \mu^*, v^*, w^*)\|$$

$$\leq \sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k)$$

$$\leq \beta_2 \|(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) - (x^*, \lambda^*, \mu^*, v^*, w^*)\|$$

Proof. The result follows from Lemma 3.1 and Assumption A4, using the local error bound theory [16, 19] □

By Lemma 3.1, under Assumptions A1, A2 and A3, the sequences $\{\lambda^k\}$ and $\{\mu^k\}$ are bounded. Therefore, by (13) and (14), if $\mu_{\max}, \lambda_{\max}$ and $-\lambda_{min}$ are large enough, we will have that, asymptotically,

$$\lambda^k = \bar{\lambda}^k, \text{ and } \mu^k = \bar{\mu}^k \tag{37}$$

This fact justifies the following assumption which, in turn, will be used in several of the forthcoming results.

Assumption B. There exists $k_0 \in I\!N$ such that $\lambda^k = \bar{\lambda}^k$ and $\mu^k = \bar{\mu}^k$ for all $k \geq k_0$.

The fulfillment of Assumption B depends on the choice of the safeguarding parameters $\lambda_{\min}$, $\lambda_{\max}$ and $\mu_{\max}$. Algorithm 2.1 admits arbitrary choices for these parameters, and the first-order convergence results are independent of these choices. However, if we fail to choose sufficiently large safeguards and, consequently, Assumption B fails to hold, boundedness of the penalty parameters will not be guaranteed.

The technique of safeguarding the Lagrange multiplier estimates allows one to prove the first-order convergence results of Section 2. This technique is also used in [1]. If one decides not to use safeguards at all, convergence to KKT points can only be guaranteed assuming boundedness of the multipliers estimates, or providing a sufficient assumption on the problem that implies such boundedness. Unfortunately, problem assumptions that imply boundedness of $\{\lambda^k, \mu^k\}$ tend to be very strong. Many times it is assumed that the gradients of active and violated constraints at all the limit points of the algorithmic sequence are linearly independent [12]. Since, in principle, every point in Ω might be a limit point, in order to ensure linearly independence at the cluster points we would need to assume that the gradients of active and violated constraints are linearly independent at all the points of Ω. This property does not hold even in very simple cases. Therefore, we consider, as in [1], that safeguarding multipliers is a reasonable price to pay for obtaining first-order convergence under weak problem assumptions. On the other hand, we are able to prove boundedness of the sequence $\{\lambda^k, \mu^k\}$ under MFCQ, but we need sufficiently large safeguarding parameters to contain that sequence. Again, we think that this is a reasonable price to pay. Last, but not least, sensible computer implementations do not accept extremely large multiplier estimates, which would impair the performance of subproblem solvers.

Lemma 3.3. *Suppose that Assumptions A1–A5 and B hold. Then, there exists $k_1 \in \{1, 2, \ldots\}, c_1, c_4 > 0$ and a sequence $\eta_k \to 0$ such that, for all $k \geq k_1$,*

$$\left(1 - \frac{c_4}{\rho_k}\right) \sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \leq \left(c_1 \eta_k + \frac{c_4}{\rho_k}\right) \sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}), \tag{38}$$

where

$$\eta_k = \frac{\chi(\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}))}{\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1})}. \tag{39}$$

Proof. By (6), (7), (8), (9), Assumption A1 and norm equivalences, there exists $c > 0$ such that

$$\left\| \begin{pmatrix} \nabla\mathcal{L}(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \\ \underline{h}(x^k) \\ \min\{-\underline{g}(x^k), w^k\} \end{pmatrix} \right\| \le c\varepsilon_k$$

for all $k = 1, 2, \ldots$. Therefore, there exist $c_1, c_2 > 0$ such that, for all $k = 1, 2, \ldots$,

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \le c_1\varepsilon_k + c_2(\|h(x^k)\| + \|\min\{-g(x^k), \mu^{k+1}\}\|).$$

Thus, by (8), (16) and Assumption B, there exists $c_3 > 0$ such that, for all $k \ge k_1$,

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \le c_1\varepsilon_k + \frac{c_3}{\rho_k}(\|\lambda^{k+1} - \lambda^k)\| + \|\mu^{k+1} - \mu^k\|).$$

Therefore, for all $k \ge k_1$,

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \le c_1\varepsilon_k + \frac{c_3}{\rho_k}(\|\lambda^{k+1} - \lambda^*\| + \|\lambda^k - \lambda^*\| + \|\mu^{k+1} - \mu^*\| + \|\mu^k - \mu^*\|).$$

Thus, by Lemma 3.2, there exists $c_4 > 0$ such that, for all $k \ge k_1$,

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \le c_1\varepsilon_k + \frac{c_4}{\rho_k}[\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) + \sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1})].$$

Therefore, by Assumption A5, for all $k \in k_1$ we have:

$$\left(1 - \frac{c_4}{\rho_k}\right)\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \le c_1\chi(\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1})) + \frac{c_4}{\rho_k}\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}).$$

By Lemma 3.1, Lemma 3.2 and Assumption A5 we have that

$$\lim_{k \to \infty} \eta_k = 0,$$

where η_k is defined by (39). Therefore, for $k \ge k_1$,

$$\left(1 - \frac{c_4}{\rho_k}\right)\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \le \left(c_1\eta_k + \frac{c_4}{\rho_k}\right)\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}),$$

as we wanted to prove. $\square$

Theorem 3.1 *Suppose that Assumptions A1–A6 and B hold. Then, the sequence of penalty parameters $\{\rho_k\}$ is bounded.*

Proof. Assume, by contradiction, that $\lim_{k \to \infty} \rho_k = \infty$. By Lemma 3.3, there exists k_1 such that (38) holds for $k \ge k_1$. Let $k_2 \ge k_1$ be such that

$$\left(1 - \frac{c_4}{\rho_k}\right) > \frac{1}{2}$$

for all $k \ge k_2$. Then, by (38), for $k \ge k_2$ we have:

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \le 2\left(c_1\eta_k + \frac{c_4}{\rho_k}\right)\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}). \tag{40}$$

Let $k_3 \geq k_2$ be such that

$$2\left(c_1 \eta_k + \frac{c_4}{\rho_k}\right) \leq r$$

for all $k \geq k_3$. Then, for all $k \geq k_3$,

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \leq r\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}).$$

Therefore, by (10) and Assumption A6, one has that $\rho_{k+1} = \rho_k$ for all $k \geq k_3$. This contradicts the assumption that ρ_k tends to infinity. So, the theorem is proved. $\qquad\square$

Corollary 3.1. *Under the assumptions of Theorem 3.1, the sequence $(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k)$ converges to $(x^*, \lambda^*, \mu^*, v^*, w^*)$ with R-linear convergence rate equal to r.*

Proof. By Theorem 3.1 and Assumption A6, for all k large enough one has:

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \leq r\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}).$$

Therefore, the result follows from Lemma 3.2. $\qquad\square$

The necessity of maintaining bounded penalty parameters is controversial. The linear systems associated with the solution of the subproblem (6-7) may be decomposed in such a way that ill-conditioning due to large ρ_k disappears. As a consequence, the inconveniences associated with big penalty parameters are limited to the following fact: when ρ_k is large, unitary Newtonian steps associated with (6-7) are less likely to give penalty function decrease than in the cases in which ρ_k is small.

However, if arbitrary large penalty parameters are employed, higher convergence orders can be obtained. It is not clear which of these effects is preponderant. With the aim of establishing the benefits that can be obtained from large penalty parameters we introduce the following Assumption C, which may replace Assumption A6. Note that Assumption C is compatible with the updates (11) and (12).

Assumption C. The sequence of penalty parameters $\{\rho_k\}$ tends to infinity.

Theorem 3.2 *Suppose that Assumptions A1–A5 and C hold. Then, $(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k)$ converges to $(x^*, \lambda^*, \mu^*, v^*, w^*)$ Q-superlinearly. Moreover, if $\chi(t) = O(t^2)$ and*

$$\frac{1}{\rho_k} = O(\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}))$$

the convergence is Q-quadratic.

Proof. Since $\rho_k \to \infty$ we obtain (40) as in the proof of Theorem 3.1. This shows that the sequence $\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k)$ tends to zero superlinearly. By Lemma 3.2 the Q-superlinear convergence of $(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k)$ follows.

Moreover, by (39), for $k \geq k_2$ we have:

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \leq 2\left(c_1 \frac{\chi(\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}))}{\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1})} + \frac{c_4}{\rho_k}\right)\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1})$$

for all $k \geq k_2$.

If $\chi(t) = O(t^2)$ and $\frac{1}{\rho_k} = O(\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1}))$ we have:

$$\sigma(x^k, \lambda^{k+1}, \mu^{k+1}, v^k, w^k) \leq O(\sigma(x^{k-1}, \lambda^k, \mu^k, v^{k-1}, w^{k-1})^2).$$

for all $k \geq k_2$. Therefore, the desired result follows from Lemma 3.2. □

Now we want to show that well-definiteness of the algorithm is plausible even if we do not assume compactness of Ω, provided that the initial point is close enough to an isolated solution x^*. We are going to make use of the following Assumption D.

Assumption D. The point x^* is feasible with respect to (1) and satisfies the CPLD constraint qualification.

Lemma 3.4. *Suppose that Assumption D holds. Then, there exists $\delta > 0$ such that x satisfies the CPLD constraint qualification corresponding to the constraints of (1) for all feasible point x belonging to $B(x^*, \delta)$.*

Proof. Assume, without loss of generality, that the constraints of (1) are $h(x) = 0, g(x) \leq 0$. Suppose that the thesis is not true. Therefore, there exists a sequence of feasible points $\{x^k\}$ that converges to x^* and such that its members do not satisfy the CPLD constraint qualification. Thus, for all $k = 1, 2, \ldots$ there exists $I_k \subseteq \{1, \ldots, m\}$, $J_k \subseteq \{1, \ldots, p\}$, $\lambda_i^k, i \in I_k$, $\mu_j^k \geq 0, j \in J_k$ such that $\sum_{i \in I_k} |\lambda_i^k| + \sum_{j \in J_k} \mu_j^k > 0$,

$$\sum_{i \in I_k} \lambda_i^k \nabla h_i(x^k) + \sum_{j \in J_k} \mu_j^k \nabla g_j(x^k) = 0 \tag{41}$$

and the gradients $\nabla h_i(y^k), i \in I_k$, $\nabla g_j(y^k), j \in J_k$ are linearly independent for some y^k such that $\|y^k - x^k\| \leq 1/k$.

Since the number of constraints is finite, we may assume, without loss of generality, that $I_k = I, J_k = J$ for all k.

Dividing (41) by $\max\{|\lambda_i^k|, \mu_j^k, i \in I, j \in J\}$ we may consider that all the coefficients of (41) are in a compact set and that the maximum modulus is equal to 1 for all k. Then, taking limits in (41) for an appropriate subsequence, we obtain that there exist $\lambda_i, i \in I$, $\mu_j, j \in J$ with $\sum_{i \in I} |\lambda_i| + \sum_{j \in J} \mu_j > 0$ such that

$$\sum_{i \in I} \lambda_i \nabla h_i(x^*) + \sum_{j \in J} \mu_j \nabla g_j(x^*) = 0. \tag{42}$$

However, the sequence y^k tends to x^* and the gradients $\nabla h_i(y^k), i \in I$, $\nabla g_j(y^k), j \in J$ are linearly independent. Thus, by (42), the CPLD condition does not hold at x^*. □

Lemma 3.5. *Suppose that Assumption D holds. Then, there exists $\delta > 0$ such that x satisfies the CPLD constraint qualification corresponding to the constraints Ω for all $x \in \Omega$ belonging to $B(x^*, \delta)$.*

Proof. By Assumption D and the definition of the constant linear dependence condition, x^* satisfies the CPLD constraint qualification with respect to the subset of constraints Ω. Therefore, the thesis follows from Lemma 3.4 □

Without loss of generality, we assume that the same δ appears in Lemmas 3.4 and 3.5.

13

Theorem 3.3. *Suppose that Assumptions A4 and D hold. Then, there exists $\underline{\rho} > 0$ such that, if $\rho_1 \geq \underline{\rho}$, there exists a sequence $\{x^k\}$ generated by Algorithm 2.1 that converges to x^*.*

Proof. By Assumption A4 and Theorem 2.2 of [31], there exist $\gamma, \delta > 0$ such that

$$f(x) \geq f(x^*) + \gamma \|x - x^*\|^2 \tag{43}$$

for all $x \in \Omega$ such that $h(x) = 0, g(x) \leq 0$ and $\|x - x^*\| \leq \delta$. Hence, x^* is the unique global minimizer of the auxiliary problem

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0, g(x) \leq 0, x \in \Omega, \|x - x^*\| \leq \delta/2. \tag{44}$$

Define
$$I = \{i \in \{1, \ldots, p\} \mid g_i(x^*) < 0\}.$$

Shrinking δ, by the continuity of g, we may assume that $g_i(x) < 0$ for all $i \in I$, $x \in B(x^*, \delta)$. For all $\rho > 0$, $\bar{\lambda} \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu} \in [0, \mu_{\max}]^p$ we define $x(\rho, \bar{\lambda}, \bar{\mu})$ as a global minimizer of $L_\rho(x, \bar{\lambda}, \bar{\mu})$ subject to $x \in \Omega$, $\|x - x^*\| \leq \delta/2$. By the compactness of the domain and the continuity of the objective function, $x(\rho, \bar{\lambda}, \bar{\mu})$ is well defined.

We want to show that, if ρ is large enough, one has

$$\|x(\rho, \bar{\lambda}, \bar{\mu}) - x^*\| < \delta/2. \tag{45}$$

If this is not the case, there exists $\rho_k \to \infty$, $\bar{\lambda}^k \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^k \in [0, \mu_{\max}]^p$ such that, defining

$$z^k = x(\rho_k, \bar{\lambda}^k, \bar{\mu}^k),$$

we have
$$\|z^k - x^*\| = \delta/2$$

for all k. However, by standard adapted arguments of external penalty methods (see, for example, Theorem 2 of [8]) the sequence $\{z^k\}$ converges to x^*, the unique global minimizer of (44). This is a contradiction. Therefore, there exists $\underline{\rho} > 0$ such that for $\rho \geq \underline{\rho}$, we have that (45) holds, independently of $\bar{\lambda}, \bar{\mu}$. Now, taking $\rho_1 > \underline{\rho}$ and defining $x^k = x(\rho_k, \bar{\lambda}^k, \bar{\mu}^k)$ for all k, by Lemma 3.5 and (45) one has that (6) and (7) are trivially satisfied. Therefore, the sequence x^k may be thought as being generated by Algorithm 2.1.

If remains to prove that $\lim x^k = x^*$. If ρ_k tends to infinity, this is a consequence of the penalty argument (Theorem 2 of [8]). Assume now that $\{\rho_k\}$ is bounded. Then, by (10) we have that

$$\lim_{k \to \infty} \|h(x^k)\| = \lim_{k \to \infty} \|\min\{-g(x^k), \mu^{k+1}\}\| = 0.$$

Let $\bar{x}$ be a limit point of $\{x^k\}$. Let K be an infinite sequence of indices such that $\lim_{k \in K} x^k = \bar{x}$. Clearly, $\bar{x}$ is feasible and belongs to $B(x^*, \delta/2)$. Let $K_1 \subseteq K$ be an infinite sequence of indices such that $\lim_{k \in K_1} \bar{\lambda}^k = \bar{\lambda}$, $\lim_{k \in K_1} \bar{\mu}^k = \bar{\mu}$. By the boundedness of $\{\rho_k\}$ there exists $\bar{\rho} > 0$ such that $\rho_k = \bar{\rho}$ for k large enough. By the definition of x^k we have that:

$$L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq L_{\rho_k}(x^*, \bar{\lambda}^k, \bar{\mu}^k)$$

for all k. Therefore,

$$f(x^k) + \frac{\rho_k}{2} \left(\sum_{i=1}^m \left(h_i(x^k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right)^2 + \sum_{i=1}^p \left(g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right)_+^2 \right)$$

14

$$\leq f(x^*) + \frac{\rho_k}{2}\left(\sum_{i=1}^{m}\left(h_i(x^*) + \frac{\bar{\lambda}_i^k}{\rho_k}\right)^2 + \sum_{i=1}^{p}\left(g_i(x^*) + \frac{\bar{\mu}_i^k}{\rho_k}\right)_+^2\right)$$

for all k. Taking limits for $k \in K_1$ we get:

$$f(\bar{x}) + \frac{\bar{\rho}}{2}\left(\sum_{i=1}^{m}\left(h_i(\bar{x}) + \frac{\bar{\lambda}_i}{\bar{\rho}}\right)^2 + \sum_{i=1}^{p}\left(g_i(\bar{x}) + \frac{\bar{\mu}_i}{\bar{\rho}}\right)_+^2\right)$$

$$\leq f(x^*) + \frac{\bar{\rho}}{2}\left(\sum_{i=1}^{m}\left(h_i(x^*) + \frac{\bar{\lambda}_i}{\bar{\rho}}\right)^2 + \sum_{i=1}^{p}\left(g_i(x^*) + \frac{\bar{\mu}_i}{\bar{\rho}}\right)_+^2\right). \tag{46}$$

By the boundedness of $\{\rho_k\}$ we have that (10) holds for all k large enough. Therefore,

$$\lim_{k\to\infty}\|h(x^k)\| = \lim_{k\to\infty}\|\min\{-g(x^k), \mu^{k+1}\}\| = 0. \tag{47}$$

By (47), $\mu_i^{k+1} = 0$ for $i \in I$ and k large enough. By (13) and (14), this implies that $\bar{\mu}_i^k = 0$ for $i \in I$ and k large enough. So, $\bar{\mu}_i = 0$ for all $i \in I$. Clearly, we also have that $\|h(x^*)\| = \|h(\bar{x})\| = 0$, $g(x^*) \leq 0$ and $g(\bar{x}) \leq 0$. Therefore, by (46),

$$f(\bar{x}) + \frac{\bar{\rho}}{2}\left(\sum_{i=1}^{m}\left(\frac{\bar{\lambda}_i}{\bar{\rho}}\right)^2 + \sum_{i\notin I}\left(\frac{\bar{\mu}_i}{\bar{\rho}}\right)_+^2\right) \leq f(x^*) + \frac{\bar{\rho}}{2}\left(\sum_{i=1}^{m}\left(\frac{\bar{\lambda}_i}{\bar{\rho}}\right)^2 + \sum_{i\notin I}\left(\frac{\bar{\mu}_i}{\bar{\rho}}\right)_+^2\right).$$

Thus, $f(\bar{x}) \leq f(x^*)$. Since x^* is the unique global minimizer of (44) this implies that $\bar{x} = x^*$. Since $\bar{x}$ was defined as an arbitrary limit point of the bounded sequence $\{x^k\}$, it turns out that $\lim_{k\to\infty} x^k = x^*$, as we wanted to prove. $\qquad\square$

4   Numerical Experiments

Since the introduction of Algencan in [1], many modifications were made, so that the 2010 version of this software is considerably better than the one that was described and tested in [1]. In the present paper, theoretical arguments motivate some additional modifications. Section 3 of the present paper indicate that Algorithm 2.1 possesses more natural convergence properties than the previous version of Algencan. It is natural, therefore, to test the new version against the former one. Essentially, we want to report three types of experiments. Firstly, we want to test whether experiments corroborate the hypothesis that the new algorithm is, at least, as efficient as the traditional Algencan. Secondly, we employ a large set of classical test problems to study the practical behavior of penalty parameters when Algencan converges to a KKT point. In a final set of of experiments we compare the new version of Algencan to a well established Interior Point method, in order to place our solver within the context of NLP algorithms. Subsections 4.1 and 4.2 correspond to the first objective of the experimentation. In Subsection 4.3 the practical behavior of the penalty parameter is analysed. Subsection 4.4 corresponds to the comparison.

4.1   Criterion for Increasing the Penalty Parameter

The point of view of Algorithm 2.1 is that one aims to solve the nonsmooth system of equations (with unknowns x, λ, μ, v, w) given by

$$\nabla f(x) + \nabla h(x)\lambda + \nabla g(x)\mu + \nabla \underline{h}(x)v + \nabla \underline{g}(x)w = 0, \tag{48}$$

15

$$h(x) = 0, \ \min\{-g(x), \mu\} = 0, \tag{49}$$

$$\underline{h}(x) = 0, \ \min\{-\underline{g}(x), w\} = 0. \tag{50}$$

If, using a penalty parameter ρ_k, one obtains enough decrease of the residual of this system, the penalty parameter is maintained (or increased slowly). Otherwise, we quickly increase the penalty parameter.

In Algorithm 3.1 of [1] one maintains the penalty parameter if the quantity

$$\left\| \begin{pmatrix} h(x^k) \\ \min\{-g(x^k), \frac{\bar{\mu}^k}{\rho_k}\} \end{pmatrix} \right\| \tag{51}$$

decreased enough with respect to its value at the previous iteration.

Therefore, there are two main differences between the criteria used in [1] and the one used in this paper. On one hand, in Algorithm 3.1 of [1] the increasing-penalty criterion does not involve (48) and (50). The reason is that, in [1], it is considered that the progress with respect to dual feasibility (48), feasibility with respect to Ω, and complementarity (50) is already guaranteed by the subproblem requirements (6) and (7). On the other hand, the main difference between the increasing criteria of [1] and Algorithm 2.1 of the present paper is that, in [1], the quantity that measures feasibility-complementarity with respect to a penalized inequality constraint is

$$\min\left\{-g_i(x^k), \frac{\bar{\mu}_i^k}{\rho_k}\right\} \tag{52}$$

whereas in Algorithm 2.1 we use

$$\min\{-g_i(x^k), \mu_i^{k+1}\} \tag{53}$$

for the same purpose.

The employment of (53) does not need further explanation since it is obviously connected with the residual of the KKT system. On the other hand, the employment of (52) is more subtle and may be connected with the essence of the Augmented Lagrangian idea. By (3), at each subproblem resolution one minimizes the objective function penalized by the non-fulfillment of the constraints, where the constraints are shifted by the scaled Lagrange multipliers $\bar{\lambda}/\rho$ and $\bar{\mu}/\rho$. This means that x^k is the result of minimizing a penalized function with shifted constraints. Note that the comparison between $-g_i(x^k)$ and $\bar{\mu}_i^k/\rho_k$ involved in (52) is, in this sense, well scaled, whereas the comparison in (53) could be badly scaled.

If x^k was obtained using a tiny shift $\bar{\mu}_i^k/\rho_k$ then the quantity (52) is small (in modulus) if and only if x^k is almost feasible with respect to g_i. On the other hand, if x^k was obtained using a big shift we can be happy with $g_i(x^k)$ only if this value is close to zero. (A very negative value of $g_i(x^k)$ could have been caused merely by the size of the shift.)

In this experiment we aim to observe the practical differences between Algorithm 3.1 of [1] and Algorithm 2.1. With this objective in mind, we implemented Algorithm 2.1 as a minor modification of Algencan 2.3.2 (the current available implementation of Algorithm 3.1 of [1]). This means that in the numerical experiments of the present subsection, as well as in the remaining of the paper, we will restrict ourselves to the case in which the simple constraints constraints are given only by bounds. In other words, we assume that Ω in (2) is defined by

$$\Omega = [\ell, u] \equiv \{x \in I\!\!R^n \mid \ell \le x \le u\},$$

defining $\underline{m} = 0$, $\underline{p} = 2n$, $\underline{g}_i(x) = l_i - x_i$ and $\underline{g}_{n+i}(x) = x_i - u_i$ for $i = 1, \dots, n$. All the other constraints of the problem will be considered in the definition of the augmented Lagrangian.

Algencan solves the bound-constrained Augmented Lagrangian subproblem using Gencan [9] with some modifications suggested in [4, 10]. We used as a stopping criterion of Gencan

$$\|P_{[\ell,u]}(x - \nabla L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k)) - x\|_\infty \le \varepsilon_k. \tag{54}$$

Given x^k that satisfies (54), there exists $w^k \ge 0$ such that (6) and (7) hold. In fact, we may compute $w^k \ge 0$ such that

$$\left\| \begin{pmatrix} \nabla \mathcal{L}(x^k, \lambda^{k+1}, \mu^{k+1}, w^k) \\ \min\{-\underline{g}(x^k), w^k\} \end{pmatrix} \right\|_\infty = \|P_{[\ell,u]}(x^k - \nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)) - x^k\|_\infty.$$

Moreover, if we set $\varepsilon_k \le r \|P_{[\ell,u]}(x^{k-1} - \nabla L_{\rho_{k-1}}(x^{k-1}, \bar{\lambda}^{k-1}, \bar{\mu}^{k-1})) - x^{k-1}\|_\infty$, we have that (6) and (7) imply

$$\left\| \begin{pmatrix} \nabla \mathcal{L}(x^k, \lambda^{k+1}, \mu^{k+1}, w^k) \\ \min\{-\underline{g}(x^k), w^k\} \end{pmatrix} \right\|_\infty \le r \left\| \begin{pmatrix} \nabla \mathcal{L}(x^{k-1}, \lambda^k, \mu^k, w^{k-1}) \\ \min\{-\underline{g}(x^{k-1}), w^{k-1}\} \end{pmatrix} \right\|_\infty.$$

The convergence theory of Gencan guarantees that an approximate solution of the subproblem x^k that satisfies (54) can always be found [9]. Therefore, the test (10) reduces to

$$\left\| \begin{pmatrix} h(x^k) \\ \min\{-g(x^k), \mu^{k+1}\} \end{pmatrix} \right\|_\infty \le r \left\| \begin{pmatrix} h(x^{k-1}) \\ \min\{-g(x^{k-1}), \mu^k\} \end{pmatrix} \right\|_\infty. \tag{55}$$

In the implementation of Algorithm 2.1, we set

$$\varepsilon_k = \min\{\bar{\varepsilon}_k, r \|P_{[\ell,u]}(x^{k-1} - \nabla L_{\rho_{k-1}}(x^{k-1}, \bar{\lambda}^{k-1}, \bar{\mu}^{k-1})) - x^{k-1}\|_\infty\},$$

where $\{\bar{\varepsilon}_k\}$ is a sequence of positive numbers than tends to zero employed in the current Algencan version. The test (10) was implemented in the form (55). When (55) is fulfilled, we choose $\rho_{k+1} = \rho_k$ in (11). This version of the algorithm will be called Algorithm 2.1.v1 from now on.

We compared Algorithm 2.1.v1 against Algorithm 3.1 of [1] using all the (648) nonlinear programming problems from the Cuter collection. We excluded only feasibility problems, bound-constrained and unconstrained problems. Both methods used all the Algencan default parameters. The CPU time of each pair method/problem was limited by 60 secs. (This limitation should be taken into account in the global performance analysis.)

The numerical performances of the methods were very similar. Using performance profiles [14] and number of iterations of the inner solver as performance measurement, we verified that the efficiency of both methods is 78% and 77%, respectively, while their robustness is 81% in both cases (a performance profile graphic would show two superimposed curves). The two methods found equivalent minima in 521 problems. (We say that f_1 and f_2 are equivalent if

$$[|f_1 - f_2| \le \max\{10^{-10}, 10^{-6} \min\{|f_1|, |f_2|\}\}] \text{ or } [f_1 \le -10^{20} \text{ and } f_2 \le -10^{20}].)$$

Both methods found infeasible final iterates in 122 problems and found different minima in 4 cases. In 2 cases one version found better solutions while the other method found better solutions in the other 2 cases. In one problem Algorithm 2.1.v1 found a solution while Algorithm 3.1 of [1] failed to find a feasible point. We warn that these results may be quantitatively different if one increases the 60-seconds time tolerance. However, if we increase the time tolerance both methods seem to improve their performance in the same proportion. From these experiments we conclude that the two algorithms compared in this section are essentially equivalent from the practical point of view.

4.2 Boundedness of Penalty Parameter and Rate of Convergence

In Section 3 we showed that, in the implementation of Algorithm 2.1 we must choose between a strategy that tries to maintain boundedness of penalty parameters getting a linear rate of convergence or an strategy that allows penalty parameter growth achieving superlinear convergence. Here we wish to design an experiment that aims to indicate which strategy is the best.

Let us define Algorithm 2.1.v2 and Algorithm 2.1.v3. Algorithm 2.1.v2 uses

$$\varepsilon_k = \min\left\{\bar{\varepsilon}_k, \left\| \begin{pmatrix} P_{[\ell,u]}(x^{k-1} - \nabla L_{\rho_{k-1}}(x^{k-1}, \bar{\lambda}^{k-1}, \bar{\mu}^{k-1})) - x^{k-1} \\ h(x^{k-1}) \\ \min\{-g(x^{k-1}), \mu^k\} \end{pmatrix} \right\|_\infty \right\},$$

while Algorithm 2.1.v3 uses

$$\varepsilon_k = \min\left\{\bar{\varepsilon}_k, \left\| \begin{pmatrix} P_{[\ell,u]}(x^{k-1} - \nabla L_{\rho_{k-1}}(x^{k-1}, \bar{\lambda}^{k-1}, \bar{\mu}^{k-1})) - x^{k-1} \\ h(x^{k-1}) \\ \min\{-g(x^{k-1}), \mu^k\} \end{pmatrix} \right\|_\infty^2 \right\}.$$

In Algorithm 2.1.v2 we set

$$\rho_{k+1} = 2\rho_k$$

in (11), while in Algorithm 2.1.v3 we set

$$\rho_{k+1} = \max\left\{2\rho_k, \left\| \begin{pmatrix} P_{[\ell,u]}(x^k - \nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)) - x^k \\ h(x^k) \\ \min\{-g(x^k), \mu^{k+1}\} \end{pmatrix} \right\|_\infty^{-1} \right\}$$

in (11). Roughly speaking, by the theory of Section 3, Algorithm 2.1.v2 converges superlinearly and the convergence of Algorithm 2.1.v3 is quadratic. We compared Algorithm 2.1.v2 and Algorithm 2.1.v3 against Algorithm 2.1.v1 under the same conditions and using the same set of problems of the previous subsection.

Figure 1 shows the performance profiles comparing the superlinear and quadratic versions of Algorithm 2.1 against the original linearly convergent version. As expected, there is gain of approximately 5% in efficiency at the cost of a small robustness deterioration. This evokes the common knowledge about the behavior of shortcut penalty methods, where one uses a very large external penalty parameter from the beginning. The shortcut strategy may find very quickly suitable minimizers in many easy problems but fails to solve many difficult problems in which more conservative penalty strategies succeed [17].

Algorithm 2.1.v1 and Algorithm 2.1.v2 found the same minimum in 486 problems and infeasible final iterates in 119 problems (recall the rather severe time limitation). They found different minima in 36 cases (23 better solutions were found by Algorithm 2.1.v1 while 13 better solutions were found by Algorithm 2.1.v2). Algorithm 2.1.v1 was the only method to find a solution in 4 problems (Algorithm 2.1.v2 found no feasible point in these cases), while the opposite situation occurred in 3 problems. Comparing Algorithm 2.1.v1 and Algorithm 2.1.v3, both found the same minimum in 474 problems and infeasible final iterates in 115 problems. They found different minima in 41 cases (23 better solutions were found by Algorithm 2.1.v1 while 18 better solutions were found by Algorithm 2.1.v3). Algorithm 2.1.v1 was the only method to find a solution in 11 problems (Algorithm 2.1.v3 found no feasible point in these cases), while the opposite situation occurred in 7 problems.

The efficiency gain of Algorithm 2.1.v2 and Algorithm 2.1.v3 with respect to Algorithm 2.1.v1 was approximately 5%, in terms of performance profiles. Roughly speaking, this means that if our time tolerance is very small, the superlinear and quadratic algorithms solve 5% more problems than the linearly convergent algorithm that maintains bounded penalty parameters.

Algorithm 2.1.v1 employed, on average, 12% more outer iterations than its superlinear counterpart and 18% more outer iterations than the quadratic algorithm, considering only the problems where the algorithms obtained the same solution. However, the increased difficulty of the subproblems used in the superlinear methods was reflected by the fact that the number of inner (Gencan) iterations employed by Algorithms 2.1.v2 and 2.1.v3 was 21% and 44% greater than the total number of Gencan iterations used by Algorithm 2.1.v1.

Considering that, in general, the capacity of solving more problems in reasonable time (robustness) is more important than the capacity of solving many problems in minimal time (efficiency), our conclusion from these experiments is that the version of Algorithm 2.1 that has the chance of maintaining bounded penalty parameters (v1) should be preferred over the ones in which penalty parameters go to infinity.

4.3   Behavior of the Penalty Parameter in Experiments

In this experiment we aim to evaluate the asymptotic behavior of the penalty parameter on Algorithm 2.1.v1. With this purpose, we consider the 465 problems (out of 648) for which Algorithm 2.1.v1 satisfied its stopping criterion related to success, i.e., it found an approximate KKT point. Recall the sixty-seconds CPU time limit.

Figure 2 shows the distribution of the 465 problems for which Algorithm 2.1.v1 found a KKT point, according to the number of times the penalty parameter was increased during the execution of the method. The penalty parameter can potentially be increased at every iteration of the method. Therefore, it is important to stress that while the average number of outer iterations was approximately 9.69, the average number of times the penalty parameter was increased is 1.76. Moreover, the penalty parameter remains constant, in average, in the last 4.36 iterations of the method.

4.4   Location and Packing Problems

One of the most noticeable practical features of Algencan is its ability to deal efficiently with problems with many inequality constraints and problems in which the Hessian of the Lagrangian does not exhibit a friendly structure. Many Packing and Location problems involve those characteristics. Our interest in those problems comes from its application to the first stage of Molecular Dynamics processes, in which we are involved [26, 28].

Given a function $\Phi : I\!\!R^{nd} \to I\!\!R$, and a set of np points $p_i \in I\!\!R^{nd}, i = 1, \ldots, np$, the type of problems considered in this section is

$$\text{Maximize } d \quad \text{subject to} \quad \Phi(p_i) \leq 0, \ \forall i \text{ and } \|p_i - p_j\| \geq d, \ \forall j > i. \tag{56}$$

In problem (56) we have $n = nd \times np + 1$, $p = np \times (np-1)/2 + np$, $m = 0$. Therefore, the number of variables and, especially, inequality constraints, may be very large.

We compared Algencan with the interior-point solver Ipopt [33] for problems of this type. In the numerical experiments below we considered Algencan 2.3.5 and Ipopt 3.8.3. Ipopt was installed with all the required ASL, Blas, Lapack and Harwell subroutines (including MA57) and without Methis and Mumps. The experiments were performed in a 2.4GHz Intel Core2 Quad Q6600 with 8.0GB of RAM memory and Linux Operating System.

Consider the problem

$$\begin{array}{rlll}
\text{Maximize} & d & & \\
\text{subject to} & d(p_i, p_j) & \geq & d, \quad \forall \, j > i, \\
& \Phi(p_i) & \leq & 2, \quad \forall \, i,
\end{array}$$
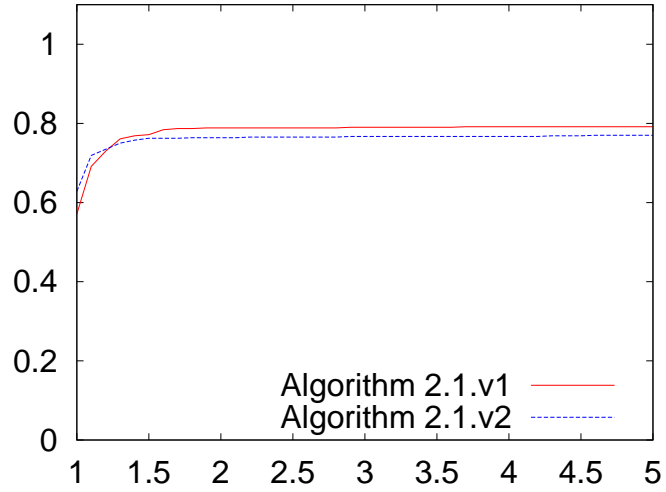
(a)
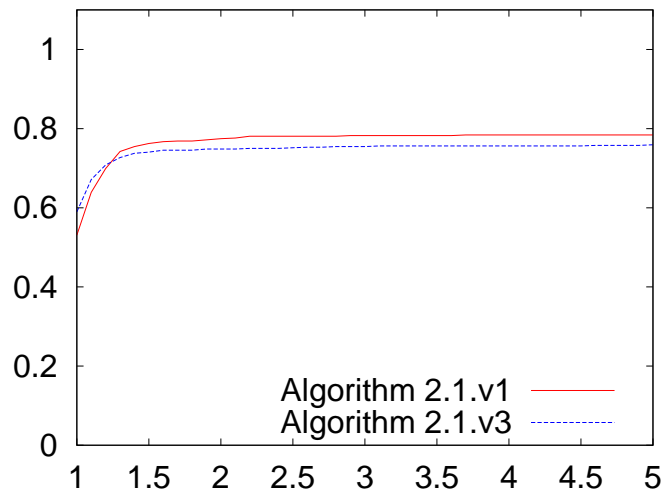
(b)

Figure 1: Performance profile comparing the (a) superlinear and (b) quadratic versions of Algorithm 2.1 against the linearly convergent version. As expected, there is a 5% efficiency gain at the cost of a small robustness deterioration; resembling the empirical knowledge about short-cut strategies.
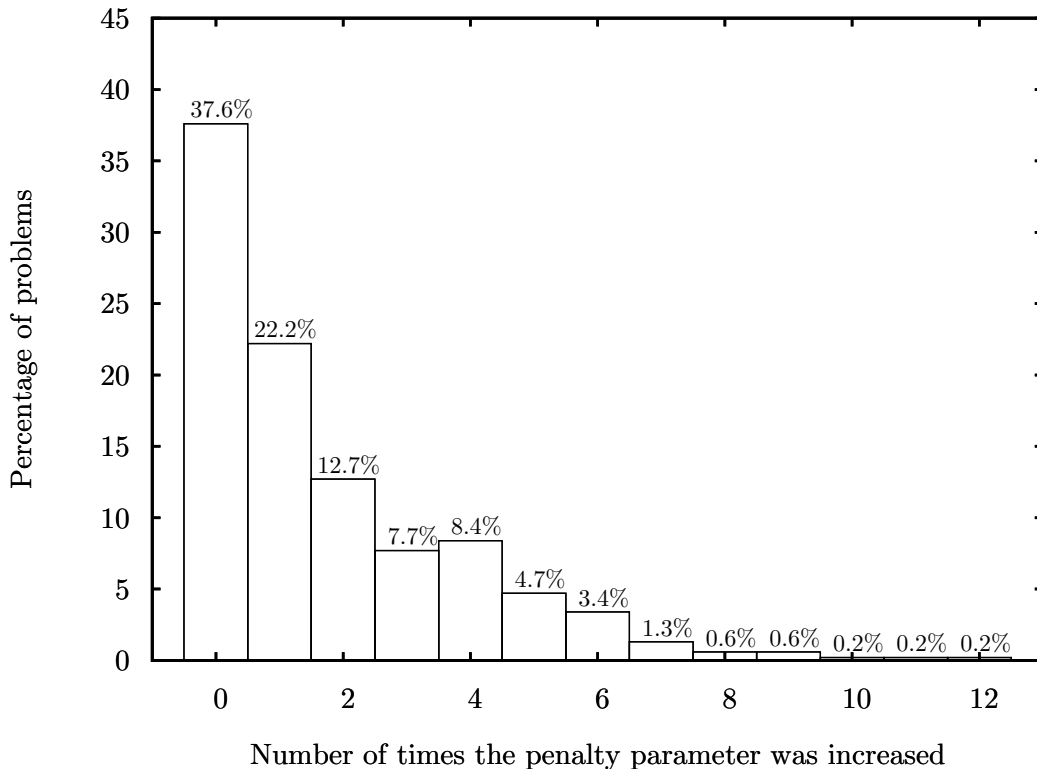
Figure 2: Analysis of the asymptotic behavior of the penalty parameter on Algorithm 2.1.v1.

where $d(\cdot, \cdot)$ represents the squared Euclidean distance and $\Phi(x, y) = 10(y - x^2)^2 + (1 - x)^2$ is the Rosenbrock function with $\rho = 10$.

We ran Algencan and Ipopt with the AMPL formulation of the problem. All the default Algencan and Ipopt parameters were used, including a feasibility tolerance of 10^{-8}. We generated 15 different instances varying the number of points $np \in \{10, 20, \ldots, 150\}$. For each instance, we ran Algencan and Ipopt starting from a random initial point with $p_i \in [-1, 3] \times [-1, 7]$ and $d = 0$. Table 1 shows, for each problem, the number of points np, the number of variables n and the number of constraints m, and, for each method, the CPU Time in seconds (Time) and the reported objective function value at the final point, that corresponds with the smallest squared Euclidean distance $d \equiv \text{dist}_{\min}^2$ between every pair of points.

Table 1 shows that in the nine instances with up to 90 points, Ipopt was faster than Algencan but Algencan found slightly better solutions in eight out of the nine instances. In those nine instances both methods stopped satisfying their own stopping criteria related to success. In the remaining six larger instances Algencan also stopped declaring success, while the Ipopt exit deserves further explanations. In instances with $np = 100, 110, 120, 130$ Ipopt stopped declaring *EXIT: Solved To Acceptable Level* which means that feasibility is not strictly satisfied. In the instance with $np = 140$ a solution similar to the one obtained by Algencan was found. In the instance with $np = 150$ Ipopt was abruptly terminated with the exit message *EXIT: Not enough memory. Ipopt 3.8.3: Unknown Error.* In an additional instance with $np = 200$ ($n = 401$ and $m = 20100$), Ipopt stopped after

633.52 seconds of CPU time declaring *EXIT: Converged to a point of local infeasibility. Problem may be infeasible.*. On the other hand, Figure 3 illustrates the solution found by Algencan for this instance, with $\text{dist}^2_{\min} = 0.0138687$ using 2897.68 seconds of CPU time.

Problem data			Algencan		Ipopt	
np	n	m	Time	$d \equiv \text{dist}^2_{\min}$	Time	$d \equiv \text{dist}^2_{\min}$
10	21	55	0.74	0.7346560	0.05	0.7186290
20	41	210	1.82	0.2523750	0.27	0.2334550
30	61	465	2.31	0.1366320	0.62	0.1327630
40	81	820	4.01	0.0957827	1.40	0.0939190
50	101	1275	6.34	0.0701519	2.10	0.0700105
60	121	1830	11.89	0.0569928	3.10	0.0557956
70	141	2485	15.46	0.0478792	5.50	0.0481858
80	161	3240	31.69	0.0410347	28.44	0.0293272
90	181	4095	43.86	0.0356818	33.80	0.0296515
100	201	5050	98.44	0.0312704	11.90	7.06e−09
110	221	6105	73.58	0.0277161	8.14	3.95e−09
120	241	7260	83.03	0.0254531	13.40	-1.25e−06
130	261	8515	210.20	0.0228939	16.93	-3.36e−08
140	281	9870	796.93	0.0210594	146.21	0.0208684
150	301	11325	274.53	0.0195162		

Table 1: Performance of Algencan and Ipopt in a set of packing problems.
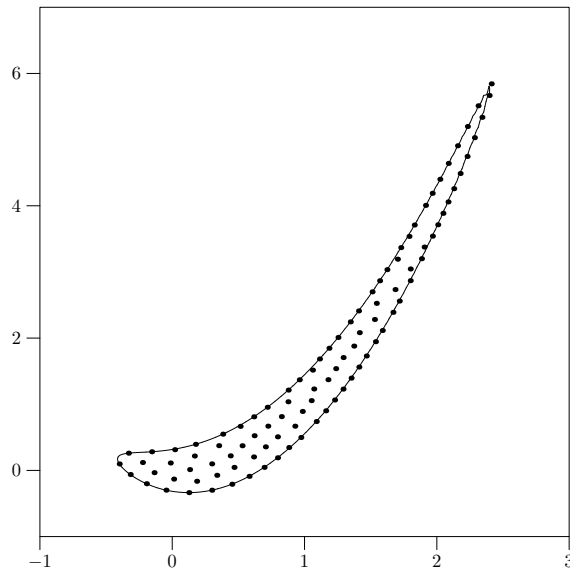

Figure 3: Solution found by Algencan for the problem of maximizing the smallest distance between every pair of $np = 200$ points within a given level curve of the Rosenbrock function.

5   Final Remarks

The main objectives of numerical optimization theory are to explain practical behavior and to suggest improvements of computational methods. When, in the Augmented Lagrangian framework, penalty parameters become very large, the theoretical order of convergence improves but the subproblems become more difficult. As a consequence, when penalty parameters are allowed to grow indefinitely, the number of outer iterations tends to decrease but the total number of inner iterations increases. It is possible to solve subproblems with very large penalty parameters by means of Newtonian stabilized decomposition of optimality conditions [5, 15, 18, 27, 34]. However, these procedures do not eliminate completely the inconveniences of using extreme penalty parameters because the objective function of the subproblems decreases at the trial point computed by the stabilized iteration only when the inner iterate is very close to the solution. As a consequence, short steps are taken far from the solution of the subproblem and subproblems are hard to solve. Our experiments seem to indicate that strategies that maintain bounded penalty parameters tend to be more robust than strategies in which the penalty parameters go to infinity. Therefore, it is important to study sufficient conditions under which one can guarantee that the sequence of penalty parameters computed by Augmented Lagrangian algorithms like Algencan is bounded.

In this paper we proved that a variation of the available (March 2010) version of Algencan produces bounded penalty parameters under more natural conditions than the ones given in the paper in which Algencan was introduced [1]. We investigated numerically whether the conditions given here can be relaxed. The answer seems to be affirmative: In numerical experiments the penalty parameters remained bounded even in conditions in which weak constraint qualifications do not hold. Therefore, finding weaker conditions for penalty parameter boundedness seems to be a promising research area with theoretical and practical significance.

Many alternative Augmented Lagrangian methods, not based on the PHR function, exist. See [7] for an exhaustive comparison among many methods of this type and [22, 23, 24] for very recent variations of the basic scheme. Some of these methods may be more efficient than the PHR scheme for specific families of nonlinear programming problems. It would be interesting to discover weak sufficient conditions under which these methods preserve bounded penalty parameters.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, On Augmented Lagrangian Methods with general lower-level constraints, *SIAM Journal on Optimization* 18, pp. 1286–1309, 2007.

[2] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, Second-order negative-curvature methods for box-constrained and general constrained optimization, *Computational Optimization and Applications* 45, pp. 209–236, 2010.

[3] R. Andreani, J. M. Martínez and M. L. Schuverdt, On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification, *Journal of Optimization Theory and Applications* 125, pp. 473–485, 2005.

[4] M. Andretta, E. G. Birgin and J. M. Martínez, Practical active-set Euclidian trust-region method with spectral projected gradients for bound-constrained minimization, *Optimization* 54, pp. 305–325, 2005.

[5] M. C. Bartholomew–Biggs, Recursive quadratic programming methods based on the augmented Lagrangian, *Mathematical Programming Study* 31, pp. 21–41, 1987.

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

[7] 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.

[8] E. G. Birgin, C. A. Floudas and J. M. Martínez, Global minimization using an augmented Lagrangian method with variable lower-level constraints, *Mathematical Programming* 125, pp. 139–162, 2010.

[9] 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.

[10] E. G. Birgin and J. M. Martínez, Structured minimal-memory inexact quasi-Newton method and secant preconditioners for Augmented Lagrangian Optimization, *Computational Optimization and Applications* 39, pp. 1–16, 2008.

[11] J. D. Buys, Dual Algorithms for Constrained Optimization Problems, Doctoral Dissertation, University of Leiden, Leiden, the Netherlands, 1972.

[12] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *Trust Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[13] H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed., New York, Springer-Verlag, 1999.

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

[15] J. P. Dussault, Improved convergence order for augmented penalty algorithms, *Computational Optimization and Applications* 44, pp. 373–383, 2009.

[16] A. Fischer, Local behavior of an iterative framework for generalized equations with nonisolated solutions, *Mathematical Programming* 94, pp. 91–124, 2002.

[17] R. Fletcher, *Practical Methods of Optimization*, Academic Press, London, 1987.

[18] N. I. M. Gould, On the accurate determination of search directions for simple differentiable penalty functions, *IMA Journal of Numerical Analysis* 6, pp. 357–372, 1986.

[19] W. W. Hager and M. S. Gowda, Stability in the presence of degeneracy and error estimation, *Mathematical Programming* 85, pp. 181–192, 1999.

[20] M. R. Hestenes, Multiplier and gradient methods, *Journal of Optimization Theory and Applications* 45, pp. 303–320, 1969.

[21] N. Krejić, J. M. Martínez, M. P. Mello and E. A. Pilotta, Validation of an Augmented Lagrangian algorithm with a Gauss-Newton Hessian approximation using a set of hard-spheres problems, *Computational Optimization and Applications* 16, pp. 247–263, 2000.

[22] H-Z Luo, X-L Sun and D. Li, On the convergence of augmented Lagrangian methods for constrained global optimization, *SIAM Journal on Optimization* 18, pp. 1209-1230, 2007.

[23] H-Z Luo, X-L Sun and H-X Wu, Convergence properties of augmented Lagrangian methods for constrained global optimization, *Optimization Methods and Software* 23, pp. 763–778, 2008.

[24] H-Z Luo, X-L Sun, Y-F Xu and H-X Wu, On the convergence properties of modified augmented Lagrangian methods for mathematical programming with complementarity constraints, *Journal of Global Optimization* 46, pp. 217–232, 2010.

[25] 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.

[26] J. M. Martínez and L. Martínez, Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking, *Journal of Computational Chemistry* 24, pp. 819–825, 2003.

[27] J. M. Martínez and L. T. Santos, Some new theoretical results on recursive quadratic programming algorithms, *Journal of Optimization Theory and Applications* 97, pp. 435–454, 1998.

[28] L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, Packmol: A package for building initial configurations for molecular dynamics simulations, *Journal of Computational Chemistry* 30, pp. 2157–2164, 2009.

[29] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in *Optimization*, R. Fletcher (ed.), Academic Press, New York, NY, pp. 283–298, 1969.

[30] 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.

[31] S. M. Robinson, Generalized equations and their solutions, Part II: Applications to Nonlinear Programming, *Mathematical Programming Study* 19, pp. 200–221, 1982.

[32] R. T. Rockafellar, A dual approach for solving nonlinear programming problems by unconstrained optimization, *Mathematical Programming* 5, pp. 354–373, 1973.

[33] A. Wächter and L. T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Mathematical Programming* 106, pp. 25–57, 2006.

[34] S. J. Wright, Superlinear convergence of a stabilized SQP method to a degenerate solution, *Computational Optimization and Applications* 11, pp. 253–275, 1998.