J Optim Theory Appl
DOI 10.1007/s10957-015-0735-7

Convergence Results of an Augmented Lagrangian Method Using the Exponential Penalty Function

Nélida Echebest[1] · **María Daniela Sánchez**[2] · **María Laura Schuverdt**[2]

Received: 22 September 2014 / Accepted: 8 April 2015

Abstract In the present research, an Augmented Lagrangian method with the use of the exponential penalty function for solving inequality constraints problems is considered. Global convergence is proved using the constant positive generator constraint qualification when the subproblem is solved in an approximate form. Since this constraint qualification was defined recently, the present convergence result is new for the Augmented Lagrangian method based on the exponential penalty function. Boundedness of the penalty parameters is proved considering classical conditions. Three illustrative examples are presented.

Keywords Nonlinear programming · Augmented Lagrangian methods · The exponential penalty function · Global convergence · Constraint qualifications

Mathematics Subject Classification 65K05 · 90C30 · 49K99

Communicated by Vaithilingam Jeyakumar.

✉ María Laura Schuverdt
schuverd@mate.unlp.edu.ar

Nélida Echebest
opti@mate.unlp.edu.ar

María Daniela Sánchez
dsanchez@mate.unlp.edu.ar

1 Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata, Bs. As., Argentina

2 CONICET, Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata, Bs. As., Argentina

Published online: 25 April 2015

1 Introduction

Augmented Lagrangian methods are powerful tools for solving nonlinear optimization problems. The main idea is to approximate the original problem by an iterative sequence of subproblems considerably easier to solve. In each subproblem, for a fixed penalty parameter and Lagrange multiplier estimate, an Augmented Lagrangian function is approximately minimized. Once the approximate solution is found, new parameters are defined and a new iteration starts. In practical implementations, the most classical and widely used Augmented Lagrangian function, due to Powell–Hestenes–Rockafellar [1–3], is based on the quadratic penalty function (called PHR function). In the literature, many effective algorithms based on the PHR function can be found; see, for example, the algorithms defined in [4–12] for solving general nonlinear smooth problems, [13] for solving derivative free problems, [14–17] for mathematical problems with degenerate or complementarity constraints and [18,19] for global optimization problems.

In the present research, we consider the Augmented Lagrangian methodology with the use of the exponential penalty function. This function was introduced in the context of convex programming [20,21]. Non-quadratic penalty functions have been extensively studied during the last decades; see, for example, the seminal book [6,8,22–26] and references therein. To understand the connection between penalization and image space analysis, at a more general level in the field of optimization, see the fundamental book [27] and references therein.

The Augmented Lagrangian method defined here follows the idea introduced in [4, 5], but the difference relies on the penalty function used. In this context, we were able to obtain global convergence to first-order stationary points considering the exponential penalty function under a weak constraint qualification without convexity assumptions. We have proved that a limit point which satisfies the constant positive generator (CPG) constraint qualification [28] fulfils the Karush–Kuhn–Tucker (KKT) conditions. This global convergence under the CPG constraint qualification is a novel result in the optimization field.

It is well established in the literature that, when the penalty parameter is very large, the subproblems can be very difficult to solve in an effective form [9]. For this reason, it is important to establish appropriate sufficient conditions to guarantee that the sequence of penalty parameters does not need to increase indefinitely in order to achieve a solution. This local convergence analysis strongly depends on the measure defined to control the penalty parameter.

In this work, the measure used to increase the penalty parameter comes from the measure used in the quadratic case [4,5] adapted for the exponential one. Thus, based on the ideas in [5,23] we will be able to obtain the boundedness of the penalty parameter sequence, in the case of the exponential Augmented Lagrangian algorithm, with the use of: the linear independence of the active constraints gradients, the positive definite of the Hessian of the Lagrangian function in the orthogonal subspace to the gradients of the active constraints and the strict complementarity condition.

This paper has been organized as follows. In Sect. 2, we have described the main algorithm and we have presented the global convergence results. Section 3 devoted to prove boundedness of the penalty parameters. In Sect. 4, we are going to show three illustrative examples. Conclusions are given in Sect. 5 and lines for future research in Sect. 6.

2 The Exponential Augmented Lagrangian Method and Global Convergence

For the reader's convenience, we present first some notations. Throughout this paper, $\mathbb{N} = \{0, 1, 2, \ldots\}$ is the set of all natural numbers, $\mathbb{R}$ denotes the set of real numbers and $\mathbb{R}^n$ denotes the Euclidean space with dimension n. The set of non-negative numbers, $\mathbb{R}_+$, is defined by $\mathbb{R}_+ := \{t \in \mathbb{R}: t \geq 0\}$, and the set of positive real numbers, $\mathbb{R}_{++}$, is defined by $\mathbb{R}_{++} := \{t \in \mathbb{R}: t > 0\}$. For a vector $v \in \mathbb{R}^n$, v_i is the $i-$th component of the vector v. For a vector-valued function $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $F := (f_1, \ldots, f_m)$, the Jacobian matrix is denoted by $\nabla F(x)$ and defined by $\nabla F(x) := (\nabla f_1(x), \ldots, \nabla f_m(x)) \in \mathbb{R}^{n\times m}$. For a subset $K = \{k_0, k_1, k_2, \ldots\} \subset \mathbb{N}\, (k_{j+1} > k_j \forall j)$, we denote $\lim_{k\in K} x^k = \lim_{j\to\infty} x^{k_j}$.

For a vector $y \in \mathbb{R}^n$ and a set $J \subset \{1, \ldots, n\}$, y_J denotes the subvector composed from the components y_i, $i \in J$. Analogously, for a matrix B a set $J \subset \{1, \ldots, n\}$, B_J denotes the matrix composed by taking the columns of B indexed by J. Finally, $\|\cdot\|$ is an arbitrary vector norm.

We will consider the inequality constrained nonlinear optimization problem:

$$\min f(x) \text{ subject to } g(x) \leq 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ admit continuous first derivatives on an open domain containing the feasible set. We do not assume convexity. The Augmented Lagrangian method has a very rich theory, when applied to convex problems; see, e.g. [6,20,21,24,25].

In this work, we consider the Augmented Lagrangian function based on the exponential penalty function:

$$L(x, \mu, \rho) = f(x) + \sum_{i=1}^{p} \frac{\mu_i}{\rho}\left(e^{\rho g_i(x)} - 1\right).$$

As we have already mentioned, the most standard Augmented Lagrangian function is based on the quadratic penalty function and called PHR function:

$$L_Q(x, \mu, \rho) = f(x) + \frac{1}{2\rho}\sum_{i=1}^{p}(\max\{0, \mu_i + \rho g_i(x)\})^2,$$

for all $x \in \mathbb{R}^n$, $\rho \in \mathbb{R}_{++}$, $\mu \in \mathbb{R}^p_+$.

In [8], the authors implemented 65 methods of the Augmented Lagrangian class for nonlinear optimization problems with inequality constraints, using the same framework with respect to stopping criteria, precision, subproblem solver and other algorithmic parameters. Even though the authors concluded the superiority of the use of the quadratic penalty function over the other non-quadratic ones, this conclusion strongly depends on the internal algorithm used for solving the subproblem. The discontinuity of the second derivatives of the Augmented Lagrangian function can slow down the rate of convergence and cause failures. In our opinion, this is enough to study the Augmented Lagrangian approach considering a different option rather than the quadratic penalty function.

Other important motivation to consider non-quadratic penalty functions is the possibility to develop an Augmented Lagrangian method with convergence to points verifying second-order optimality conditions. In this case, it is desirable to have a twice differentiable Augmented Lagrangian function. Thus, as it is mentioned in [25], Newton-type methods can be used for the corresponding subproblem minimization more effectively.

As we have mentioned in the Introduction, in the first step of the Augmented Lagrangian algorithm, an approximate solution of the subproblem 'minimize $L(x, \mu^k, \rho_k)$' is found. In the second step, new estimates of the multipliers are computed, and finally, in the third step, the penalty parameter is updated according to some infeasibility and complementarity measure.

Algorithm 2.1 Let $x^0 \in \mathbb{R}^n$ be an arbitrary initial point. The given parameters for the execution of the algorithm are: $0 \le \tau < 1$, $\gamma > 1$, $0 \le \bar{\mu}_i^{max} < \infty$, $\forall\, i = 1, \ldots, p$, $\rho_0 \in \mathbb{R}_{++}$, $\bar{\mu}_i^0, \bar{\mu}_i^1 \in [0, \bar{\mu}_i^{max}]$, $\sigma_i^0 = \frac{\bar{\mu}_i^1 - \bar{\mu}_i^0}{\rho_0}$, $i = 1, \ldots, p$. Finally, $\{\varepsilon_k\} \subset \mathbb{R}_+$ is a sequence of tolerance parameters such that $\lim_{k\to\infty} \varepsilon_k = 0$.
Initialize $k \leftarrow 1$.

Step 1. Solving the subproblem. Compute (if possible) $x^k \in \mathbb{R}^n$, satisfying

$$\|\nabla L(x^k, \bar{\mu}^k, \rho_k)\|_\infty \le \varepsilon_k. \tag{2}$$

If it is not possible, stop the execution of the algorithm.

Step 2. Estimate new multipliers and define a new infeasibility and complementarity measure. For $j = 1, \ldots, p$, compute

$$\mu_j^{k+1} = \bar{\mu}_j^k e^{\rho_k g_j(x^k)} \tag{3}$$

$$\bar{\mu}_j^{k+1} = P_{[0, \bar{\mu}_j^{max}]}(\mu_j^{k+1}) \tag{4}$$

$$\sigma_j^k = \frac{\mu_j^{k+1} - \bar{\mu}_j^k}{\rho_k} \tag{5}$$

Step 3. Update the penalty parameter.
If $\|\sigma^k\|_\infty \le \tau \|\sigma^{k-1}\|_\infty$, define $\rho_{k+1} = \rho_k$.
Else, define $\rho_{k+1} = \gamma \rho_k$.

Step 4. Set $k \leftarrow k + 1$ and go to Step 1.

Formula (3) is the usual first-order multiplier estimate, it is a multiplicative form, in contrast to the quadratic penalty form, where the dual update is additive:

$$\mu_j^{k+1} = \max\{0, \bar{\mu}_j^k + \rho_k g_j(x^k)\}; \tag{6}$$

see [4,5,7,9,12]. Formula (4) is the safeguarded counterpart of the multiplier estimate. This idea was introduced in [4,5] and also used in [9].

Formula (5) is the parameter that measures the progress in terms of infeasibility and complementarity. It comes from the measure used in the quadratic Augmented Lagrangian method defined in [4,5] updated for the exponential case. In those papers, the formula used is $\sigma_j^k = \max\left\{g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k}\right\}$ that corresponds to formula (5) using the definition (6). Thus, measure (5) can be seen as the appropriate measure for the exponential case. In [8], the paper where the authors compare the 65 different Augmented Lagrangian methods, the penalty parameter not increases when: $\max\{0, g_j(x^k)\} \leq \tau \max\{0, g_j(x^{k-1})\}$ and $|g_j(x^k)\mu_j^k| \leq \tau|g_j(x^{k-1})\mu_j^{k-1}|$ for all $j = 1, \ldots, p$. However, with this measure, paper [8] does not include any local convergence analysis.

During the last years, it has been proved that quadratic Augmented Lagrangian methods, which solve the subproblem in an inexact form, converge to stationary point under different constraint qualifications. The weakest constraint qualification used to prove global convergence of an inexact quadratic Augmented Lagrangian method is the constant positive generator constraint qualification defined in [28]. We will establish the definition here for completeness.

Given a tuple $V = (v_1, v_2, \ldots, v_K)$ of vectors in $\mathbb{R}^n$ and $\mathcal{I}, \mathcal{J} \subset \{1, 2, \ldots, K\}$, a pair of index sets. A *positive combination* of elements of V associated with the (ordered) pair $(\mathcal{I}, \mathcal{J})$ is a vector in the form $\sum_{i\in\mathcal{I}} \lambda_i v_i + \sum_{j\in\mathcal{J}} \mu_j v_j$, with $\mu_j \geq 0, \forall j \in \mathcal{J}$. The set of all such positive combinations is a cone called the *positive span of V associated with* $(\mathcal{I}, \mathcal{J})$, and it is denoted by $span_+(\mathcal{I}, \mathcal{J}; V)$.

The pair $(\mathcal{I}, \mathcal{J})$, when V is clear from the context, is said to be positively linearly independent, iff the only way to write the zero vector using positive combinations is to use trivial coefficients. Otherwise, the pair is positively linearly dependent.

Associated with the set V, it is possible to define the following sets $\mathcal{J}_- := \{j \in \mathcal{J} \ : \ -v_j \in span_+(\mathcal{I}, \mathcal{J}; V)\}$ and $\mathcal{J}_+ := \mathcal{J} \setminus \mathcal{J}_-$.

Definition 2.1 [28] Consider the general nonlinear optimization problem

$$\begin{aligned} &\min\ f_0(x) \\ &\text{s.t. } f_i(x) = 0,\ i = 1, \ldots, m;\ f_j(x) \leq 0,\ j = m+1, \ldots, m+p. \end{aligned}$$

For $y \in \mathbb{R}^n$, define $Gf(y) := (\nabla f_1(y), \nabla f_2(y), \ldots, \nabla f_{m+p}(y))$.

Let x be a feasible point and define the index sets $\mathcal{I} := \{1, 2, \ldots, m\}$ and $\mathcal{J} = A(x) := \{j = m+1, \ldots m+p : f_j(x) = 0\}$, the set of active inequality constraints. We say that the *constant positive generator* (CPG) condition holds at x iff there is a positively linearly independent spanning pair $(\mathcal{I}', \mathcal{J}_+)$ of $span_+(\mathcal{I}, \mathcal{J}; Gf(x))$ such that, for all y in a neighbourhood of x:

$$span_+(\mathcal{I}', \mathcal{J}_+; Gf(y)) \supset span_+(\mathcal{I}, \mathcal{J}; Gf(y)).$$

In [28], it is proved that the constant positive generator condition is a constraint qualification weaker than the linear independence constraint qualification, Mangasarian–Fromovitz constraint qualification (MFCQ) [29,30], constant rank constraint qualification [31], constant positive linear dependence (CPLD) constraint qualification [30], relaxed constant rank constraint qualification [32] and relaxed constant positive linear dependence (RCPLD) [33]. Paper [28] also discusses when an error bound holds using the new constraint qualification and shows that many algorithms, not only the quadratic Augmented Lagrangian method, converge under the weak constant positive generator condition.

Before proving the global convergence result related to Algorithm 2.1, we will establish the following proposition, which will be also used later on.

Proposition 2.1 *Let* $\{x^k\}$ *be a sequence generated by Algorithm 2.1. Assume that* x^* *is a feasible limit point of* $\{x^k\}$ *and that* $K \subset \mathbb{N}$ *is such that* $\lim_{k\in K} x^k = x^*$. *Then,* $\lim_{k\in K} \mu_j^{k+1} = 0, \quad \forall j \notin A(x^*)$.

Proof For each $j \notin A(x^*)$, let us consider the following two cases:

- If $\{\rho_k\}$ is bounded, by Step 3 of the Algorithm 2.1 we have that $\lim_{k\in K} \sigma_j^k = 0$. Since $g_j(x^*) < 0$, we obtain that $\lim_{k\in K} \dfrac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} \neq 0$ and this implies that $\lim_{k\in K} \bar{\mu}_j^k = 0$. Thus, by the definition of μ_j^{k+1} it is true that $\lim_{k\in K} \mu_j^{k+1} = 0$.
- If $\{\rho_k\}$ is unbounded, since $g_j(x^*) < 0$, we have that $\lim_{k\in K} e^{\rho_k g_j(x^k)} = 0$. Since $\{\bar{\mu}_j^k\}_{k\in K}$ is bounded, we obtain that $\lim_{k\in K} \mu_j^{k+1} = 0$.

Thus, the thesis is proved. □

Let us state now the global convergence theorem.

Theorem 2.1 *Let* $\{x^k\}$ *be a sequence generated by Algorithm 2.1. Assume that* x^* *is a limit point of* $\{x^k\}$ *that satisfies the constant positive generator constraint qualification related to the feasible set of the problem (1). Then,* x^* *is a Karush–Kuhn–Tucker point of the original problem (1).*

Moreover, if x^* *satisfies the Mangasarian–Fromovitz constraint qualification and* $\{x^k\}_{k\in K}$ *is a subsequence that converges to* x^*, *the set* $\{\|\mu^{k+1}\|\}_{k\in K}$ *is bounded.*

Proof For all $k \in \mathbb{N}$, by (2) and (3) there exist $\delta_k \in \mathbb{R}^n$ such that $\|\delta_k\| \leq \varepsilon_k$ and $\nabla f(x^k) + \sum_{j=1}^{p} \mu_j^{k+1} \nabla g_j(x^k) = \delta_k$.

Let be $K \subset \mathbb{N}$ such that $\lim_{k\in K} x^k = x^*$. Then,

$$\nabla f(x^k) + \sum_{j\in A(x^*)} \mu_j^{k+1} \nabla g_j(x^k) = \delta_k - \sum_{j\notin A(x^*)} \mu_j^{k+1} \nabla g_j(x^k). \tag{7}$$

Let $(\mathcal{I}', \mathcal{J}_+)$ be the positively linearly independent spanning pair given in the definition of CPG and consider $V := (\nabla g_1(x^*), \ldots, \nabla g_p(x^*))$. Then, for k sufficiently large there must be scalars $\tilde{\mu}_j^k$, $j \in \mathcal{I}' \cup \mathcal{J}_+$ such that $\tilde{\mu}_j^k \geq 0$ if $j \in \mathcal{J}_+$ and

$$\nabla f(x^k) + \sum_{i\in \mathcal{I}'} \tilde{\mu}_i^k \nabla g_i(x^k) + \sum_{j\in \mathcal{J}_+} \tilde{\mu}_j^k \nabla g_j(x^k) = \delta_k - \sum_{j\notin A(x^*)} \mu_j^{k+1} \nabla g_j(x^k). \tag{8}$$

Define, for all $k \in K$, $M_k := \max\{|\tilde{\mu}_i^k|, i \in \mathcal{I}', \tilde{\mu}_j^k, j \in \mathcal{J}_+\}$.

1. If M_k has a bounded subsequence then there are subsequences $\hat{\mu}_j^k$, for $j \in \mathcal{I}' \cup \mathcal{J}_+$, $\hat{\mu}_j^k \geq 0$ if $j \in \mathcal{J}_+$ and limits $\hat{\mu}_j^*$, $j \in \mathcal{I}' \cup \mathcal{J}_+$, $\hat{\mu}_j^* \geq 0$ if $j \in \mathcal{J}_+$ such that, taking limits on both sides of (8) and using Proposition 2.1, we arrive at $\nabla f(x^*) + \sum_{i \in \mathcal{I}'} \hat{\mu}_i^* \nabla g_i(x^*) + \sum_{j \in \mathcal{J}_+} \hat{\mu}_j^* \nabla g_j(x^*) = 0$.

Then, since $\mathcal{I} = \emptyset$ for the original problem (1) and using that the linear combination $\sum_{i \in \mathcal{I}'} \hat{\mu}_i^* \nabla g_i(x^*) + \sum_{j \in \mathcal{J}_+} \hat{\mu}_j^* \nabla g_j(x^*) \in span_+(\mathcal{J}; V)$, we obtain that x^* is a KKT point.

2. If $M_k \to \infty$, we can divide (8) by M_k and get

$$\frac{\nabla f(x^k)}{M_k} + \sum_{i \in \mathcal{I}'} \frac{\tilde{\mu}_i^k}{M_k} \nabla g_i(x^k) + \sum_{j \in \mathcal{J}_+} \frac{\tilde{\mu}_j^k}{M_k} \nabla g_j(x^k) = \frac{\delta_k}{M_k} - \sum_{j \notin A(x^*)} \frac{\mu_j^{k+1}}{M_k} \nabla g_j(x^k).$$

By taking limits in the last equation and using Proposition 2.1, we came to a conclusion that there is a contradiction to the fact that $(\mathcal{I}', \mathcal{J}_+)$ is positively linearly independent.

The second part can be obtained using similar arguments. □

Remark 2.1 It is worth noting how the previous proofs differ from the corresponding results in the quadratic case, for example in [28]. The proof of Theorem 2.1 is strongly based on the definition of the constant positive generator constraint qualification, in the same way as in [28]. The main difference relies on the fact that, in the quadratic case, the last term in the equality (7) is absent for k large enough. This is due to the use of formula (6).

Note that, since Theorem 2.1 uses the weaker CPG constraint qualification, this global convergence result is new for the Augmented Lagrangian algorithm using the exponential penalty function (Algorithm 2.1). This result shows that, theoretically speaking, the quadratic and the exponential penalty functions enjoy similar convergence properties.

3 Boundedness of the Penalty Parameters

We note that Augmented Lagrangian methods, as opposed to penalty methods, attempt to locate the optimum value of the problem keeping the sequence $\{\rho_k\}$ bounded in order to avoid the ill-conditioning in the limit. Because of that, in this section we will study conditions under which the sequence of penalty parameters in Algorithm 2.1 is bounded.

When inequality constraints are presented, the usual hypotheses needed for proving the local convergence result, among other algorithmic conditions, are: the sufficient second-order optimality condition (the Hessian of the Lagrangian function is positive definite in the orthogonal subspace to the gradients of the active constraints) and the linear independence of the active constraints gradients together with the strict complementarity condition. These are, for example, the conditions used in [5] to prove the boundedness of the penalty parameter for the quadratic Augmented

Lagrangian algorithm ALGENCAN and also in [23] to analyse the asymptotic behaviour when non-quadratic functions are considered. Similar but slightly more restrictive are the conditions considered in [12] to obtain boundedness of the penalty parameter for LANCELOT. Recently, in [9] the authors proved that the penalty parameter remains bounded for a quadratic Augmented Lagrangian algorithm under the following assumptions: instead of the linear independence of the active constraints, they assume that Mangasarian–Fromovitz holds and the vector of Lagrange multipliers is unique, and they employ a second-order sufficient optimality condition that does not involve the strict complementarity condition. It is important to mention that in [9] the measure used to update the penalty parameter is different from the one used in [4,5,8,12]. In [9], the authors increase the penalty parameter considering the measure

$$\sigma(x,\mu) = \left\| \begin{Bmatrix} \nabla L_Q(x,\mu,\rho) \\ \min\{-g(x),\mu\} \end{Bmatrix} \right\| \tag{9}$$

and the boundedness property is obtained with the use of the local error bound theory [34,35].

In [8], the paper shows a systematic comparison of several Augmented Lagrangian algorithms including the exponential penalty function; however, the authors do not include any analysis related to the boundedness of the penalty parameter sequence.

In this section, we work on the following assumptions.

Assumption 1 The sequence $\{x^k\}$ is generated by the application of the Algorithm 2.1 and $\lim_{k\to\infty} x^k = x^*$. The functions f and g admit continuous second derivatives in a neighbourhood of the feasible point x^*.

Let us suppose that the set of indexes of the active constraints at x^* is $A(x^*) := \{1,\dots,q\}$ and $J := \{q+1,\dots,p\}$. Define the Lagrangian function $l(x,\mu) := f(x) + \sum_{i=1}^{p} \mu_i g_i(x)$.

Assumption 2 The gradients $\{\nabla g_1(x^*),\dots,\nabla g_q(x^*)\}$ are linearly independent.

Assumption 3 The following second-order sufficient condition for local minimizers is satisfied at (x^*,μ^*). For all $z \in T(x^*)$, $z \neq 0$, $z^T \nabla^2 l(x^*,\mu^*) z > 0$, where $T(x^*) := \{z \in \mathbb{R}^n : \nabla g_i(x^*)^T z = 0, \quad \text{for all } i \in A(x^*)\}$ is the tangent subspace.

Also, we assume that the strict complementarity condition holds: $\mu_i^* > 0$ for all $i = 1,\dots,q$.

Assumption 4 For all $i = 1,\dots,p$, $0 \le \mu_i^* < \bar{\mu}_i^{max}$.

The following Lemma states that if x^* is a KKT point that verifies Assumption 3, then x^* is a local minimizer of the exponential Augmented Lagrangian function $L(x,\mu^*,\rho)$ for all $\rho \ge \bar{\rho}$. This result was proved effective for the quadratic Augmented Lagrangian function in the literature [5,7].

Lemma 3.1 *Suppose that x^* is a Karush–Kuhn–Tucker point of* (1) *and that Assumption* 3 *holds. Then, there exists $\bar{\rho} > 0$ such that, for all $\rho \ge \bar{\rho}$, $\nabla^2_{xx} L(x^*,\mu^*,\rho)$ is positive definite.*

Proof By the definition of $L(x, \mu, \rho)$ and the assumption, we have that

$$\nabla^2_{xx} L(x^*, \mu^*, \rho) = \nabla^2_{xx} l(x^*, \mu^*) + \rho \sum_{i \in A(x^*)} \mu_i^* \nabla g_i(x^*) \nabla g_i(x^*)^T.$$

Let us define $P := \nabla^2_{xx} l(x^*, \mu^*)$ and $Q := \sum_{i \in A(x^*)} \mu_i^* \nabla g_i(x^*) \nabla g_i(x^*)^T$. Then, for all $z \neq 0$, $z \in T(x^*)$ we have that $z^T Q z = \sum_{i \in A(x^*)} \mu_i^* \|\nabla g_i(x^*)^T z\|^2 = 0$ and, using Assumption 3, we obtain that $z^T P z > 0$. Thus, using Lemma 1.25 from [6] we obtain the desired result. □

Below we give some results needed to provide estimates of an approximate solution of $L(x, \mu, \rho)$ to x^* and the corresponding Lagrange multiplier estimate to μ^*. These ideas are further developed in [5,23].

Lemma 3.2 *Suppose that Assumptions 2 and 3 hold. Consider $\bar{\rho} > 0$ as in the previous Lemma. Then, for all $r \in [0, \frac{1}{\bar{\rho}}]$, the following matrix*

$$H = \begin{pmatrix} \nabla^2_{xx} l(x^*, \mu^*) & [\nabla g(x^*)]_{A(x^*)} & [\nabla g(x^*)]_J \\ [\nabla g(x^*)]^T_{A(x^*)} & \left[-\frac{r}{\mu^*}\right]_{A(x^*)} & 0 \\ 0 & 0 & I \end{pmatrix}$$

is non-singular.

Proof If H is singular, there exists $\bar{x} = (y, z) \in \mathbb{R}^{n+p}$, $(y, z) \neq 0$ such that $H\bar{x} = 0$.

Let us consider first the case $0 < r \leq \frac{1}{\bar{\rho}}$. Then, $z_j = 0$ for $j = q+1, \ldots, p$ and

$$\nabla^2_{xx} l(x^*, \mu^*) y + \sum_{i=1}^{p} \nabla g_i(x^*) z_i = 0 \tag{10}$$

$$\nabla g_i(x^*)^T y - \frac{r}{\mu_i^*} z_i = 0 \qquad i = 1, \ldots, q. \tag{11}$$

By replacing z_i from (11) in (10), we obtain that

$$y^T \nabla^2_{xx} l(x^*, \mu^*) y + \sum_{i=1}^{q} \frac{\mu_i^*}{r} ||\nabla g_i(x^*)^T y||^2 = 0.$$

Therefore, $y^T \nabla^2_{xx} L(x^*, \mu^*, \frac{1}{r}) y = 0$ for all $r = \frac{1}{\rho}$ such that $\rho \geq \bar{\rho}$, and using the previous Lemma, it must be true that $y = 0$. Therefore, by the definition of z, it must be $z = 0$ obtaining a contradiction.

The case $r = 0$ can be proved using similar arguments. □

Lemma 3.3 *Suppose that Assumptions 1–3 hold. Then, there exist differentiable functions $x(\mu, r, \alpha)$ and $y(\mu, r, \alpha)$ such that for $r > 0$:*

1. *The functions are solutions of the following system*

$$\nabla f(x)+\sum_{i=1}^{p} y_i \nabla g_i(x)=\alpha$$
$$g_i(x)-rln\left(\frac{y_i}{\mu_i}\right)=0, i=1,\ldots,q$$
$$y_i-\mu_i e^{\frac{g_i(x)}{r}}=0, i=q+1,\ldots,p.$$

2. *If* r, $\|\alpha\|$ *and* $\|\mu-\mu^*\|$ *are small enough, then:*
   (a) $\|x(\mu,r,\alpha)-x^*\|\le M\max\{r\|\mu-\mu^*\|,\|\alpha\|\}$
   (b) $\|y(\mu,r,\alpha)-\mu^*\|\le M\max\{r\|\mu-\mu^*\|,\|\alpha\|\}$.

Proof For $\rho>0$, let us consider the following system of equations in the variables $(x,y,\mu,\rho,\alpha)\in\mathbb{R}^{n+p+p+1+n}$:

$$\begin{array}{ll}\nabla f(x)+\sum_{i=1}^{p} y_i\nabla g_i(x)-\alpha=0 & \\ g_i(x)-\frac{1}{\rho}ln(\frac{y_i}{\mu_i})=0 & i=1,\ldots,q\\ y_j-\mu_j e^{\rho g_j(x)}=0 & j=q+1,\ldots,p.\end{array}$$

By defining the variable $r=\frac{1}{\rho}$, we obtain the system

$$\Theta(x,y,\mu,r,\alpha):=\begin{pmatrix}\nabla f(x)+\sum_{i=1}^{p} y_i\nabla g_i(x)-\alpha\\ g_1(x)-rln(\frac{y_1}{\mu_1})\\ \vdots\\ g_q(x)-rln(\frac{y_q}{\mu_q})\\ y_{q+1}-\mu_{q+1}e^{\frac{g_{q+1}(x)}{r}}\\ \vdots\\ y_p-\mu_p e^{\frac{g_p(x)}{r}}\end{pmatrix}=0.$$

By Assumption 2, for $0<r\le\frac{1}{\rho}$ we have that $\Theta(x^*,\mu^*,\mu^*,r,0)=0$. Moreover, the Jacobian matrix of Θ with respect to (x,y) computed at $(x^*,\mu^*,\mu^*,r,0)$ is the matrix: $H=\begin{pmatrix}\nabla^2_{xx}l(x^*,\mu^*) & [\nabla g(x^*)]_{A(x^*)} & [\nabla g(x^*)]_J\\ [\nabla g(x^*)]^T_{A(x^*)} & [-\frac{r}{\mu^*}]_{A(x^*)} & 0\\ 0 & 0 & I\end{pmatrix}$. By Lemma 3.2, this matrix is non-singular for all $r\in[0,\frac{1}{\rho}]$, and by continuity, the norm of its inverse is bounded in a neighbourhood of $(x^*,\mu^*,\mu^*,r,0)$ uniformly. By using the implicit function theorem, there is a neighbourhood V_r of $(\mu^*,r,0)$ and differentiable functions $x(\mu,r,\alpha)$, $y(\mu,r,\alpha)$ such that, for all $(\mu,r,\alpha)\in V_r$, $0<r\le\frac{1}{\rho}$:

$$\nabla f(x(\mu,r,\alpha)) + \sum_{i=1}^{p} \nabla g(x(\mu,r,\alpha)) y(\mu,r,\alpha) - \alpha = 0$$
$$g_i(x(\mu,r,\alpha)) - r ln\left(\frac{y_i(\mu,r,\alpha)}{\mu_i}\right) = 0 \qquad i \in A(x^*)$$
$$y_j(\mu,r,\alpha) - \mu_j e^{\frac{g_j(x(\mu,r,\alpha))}{r}} = 0 \qquad j \in J.$$

By using the mean value theorem, there is $c \in \mathbb{R}^{m+2}$ such that

$$\|y(\mu,r,\alpha) - \mu^*\| \le \|\nabla_\mu y(c)\| \|\mu - \mu^*\| + \|\nabla_r y(c)\| r + \|\nabla_\alpha y(c)\| \|\alpha\|. \tag{12}$$

By computing the first derivatives of $\Theta(x(\mu,r,\alpha), y(\mu,r,\alpha), \mu, r, \alpha) = 0$ with respect to (μ,r,α), we have that there is a constant C such that $\|\nabla_\alpha y(c)\| \le C\|\nabla_\alpha \Theta\|$, $\|\nabla_r y(c)\| \le C\|\nabla_r \Theta\|$ and $\|\nabla_\mu y(c)\| \le C\|\nabla_\mu \Theta\|$. Let us find bounds for the quantities $\|\nabla_\mu \Theta\|$, $\|\nabla_r \Theta\|$ and $\|\nabla_\alpha \Theta\|$.

(i) $\nabla_\alpha \Theta(x,y,\mu,r,\alpha) = (-I \;\; 0 \;\; 0)$, thus, $\|\nabla_\alpha \Theta\| = 1$

(ii) $\nabla_\mu \Theta(x,y,\mu,r,\alpha) = (0 \quad [\frac{r}{\mu}]_{A(x^*)} \quad [-e^{\frac{g(x)}{r}}]_J)$. Then, if $j \in J$, for $r > 0$ small enough, $|e^{\frac{g_j(x)}{r}}| \le m_1 r$, and if $i \in A(x^*)$, then $\mu_i^* > 0$ and for all μ_i near μ_i^* and we have that $|\frac{r}{\mu_i}| \le m_2 r$. Therefore, $\|\nabla_\mu \Theta\| \le mr$.

(iii) $\nabla_r \Theta(x,y,\mu,r,\alpha) = \left(0 \;\; [-ln(\frac{y}{\mu})]_{A(x^*)} \;\; [\frac{\mu e^{\frac{g(x)}{r}}}{r^2}]_J\right)$. Then, if $j \in J$, for $r > 0$ small enough, $|\frac{\mu_j e^{\frac{g_j(x)}{r}} g_j(x)}{r^2}| \le m_3 \mu_j$, and if $i \in A(x^*)$, we have that $|-ln(\frac{y_i}{\mu_i})| \le m_4 |y_i - \mu_i|$. Therefore, we obtain the bound:

$\|\nabla_r \Theta\| \le \tilde{m} \max\left\{\max_{i \in A(x^*)} \{|y_i - \mu_i|\}, \max_{j \in J}\{\mu_j\}\right\} \le \tilde{m}(\|y-\mu\| + \|\mu - \mu^*\|)$. Then, replacing the bounds obtained in (i)–(iii) in (12) there is $\tilde{M} > 0$ such that $\|y(\mu,r,\alpha) - \mu^*\| \le \tilde{M}(2r\|\mu-\mu^*\| + r\|y(\mu,r,\alpha)-\mu^*\| + \|\alpha\|)$, and by choosing r small enough such that $1 - r\tilde{M} > 0$, we have that

$$\|y(\mu,r,\alpha) - \mu^*\| \le \frac{\tilde{M}}{1 - r\tilde{M}}(\|\alpha\| + 2r\|\mu - \mu^*\|).$$

Then, there is a neighbourhood N of $(\mu^*, r, 0)$ such that, for all $(\mu,r,\alpha) \in N$,

$$\|y(\mu,r,\alpha) - \mu^*\| \le M \max\{r\|\mu - \mu^*\|, \|\alpha\|\}$$

as we wanted to prove.

By using the same ideas, it is possible to prove $2(a)$. □

Proposition 3.1 *Suppose that Assumptions 1 and 2 hold.*

1. *Then,* $\lim_{k\to\infty} \mu^{k+1} = \mu^*$.
2. *Suppose that also Assumption 4 holds, then* $\mu^k = \bar{\mu}^k$ *for* k *large enough.*

Proof 1. By Assumption 2, the matrix $\nabla g_A(x^*) := [\nabla g(x^*)]_{A(x^*)}$ has full rank, and since x^* is a KKT point, $\mu_A^* = (\nabla g_A(x^*)^T \nabla g_A(x^*))^{-1}(-\nabla g_A(x^*)^T \nabla f(x^*))$. By using the stopping criterion, we have that

$$\nabla f(x^k) + \nabla g_A(x^k)\mu_A^{k+1} = \delta_k - \sum_{i \notin A(x^*)} \mu_i^{k+1} \nabla g_i(x^k).$$

By Assumption 2, let us consider k large enough such that $\{\nabla g_i(x^k)\}_{i \in A(x^*)}$ is linearly independent. Then, as we did before we have that

$$\mu_A^{k+1} = (\nabla g_A(x^k)^T \nabla g_A(x^k))^{-1}\left[\nabla g_A(x^k)^T \left(\delta_k - \sum_{i \notin A(x^*)} \mu_i^{k+1} \nabla g_i(x^k) - \nabla f(x^k)\right)\right].$$

By taking limits in the last equality and using that $\delta_k \to 0$ and Proposition 2.1, we have that $\lim\limits_{k\to\infty} \mu_A^{k+1} = (\nabla g_A(x^*)^T \nabla g_A(x^*))^{-1}(-\nabla g_A(x^*)^T \nabla f(x^*)) = \mu_A^*$.

If $i \notin A(x^*)$, the result follows from Proposition 2.1.

2. It is a consequence of item 1, Assumption 4 and the definition of $\bar{\mu}^{k+1}$. □

Observe that, as opposed to the quadratic case, the term $\sum_{i \notin A(x^*)} \mu_i^{k+1} \nabla g_i(x^k)$—mentioned in Remark 2.1—appears in the proof of the previous proposition.

Corollary 3.1 *Suppose that Assumptions 1–3 hold and assume that* $\lim\limits_{k\to\infty} \rho_k = \infty$. *Define* $\alpha^k \in \mathbb{R}^n$ *such that* $\|\alpha^k\| \le \varepsilon_k$. *Then, there exists* $M > 0$ *such that, for all* $k \in \mathbb{N}$, $\|x^k - x^*\| \le M \max\left\{\frac{\|\bar{\mu}^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\}$.

And, if $\mu^{k+1} = \bar{\mu}^k e^{\rho_k g(x^k)}$,

$$\|\mu^{k+1} - \mu^*\| \le M \max\left\{\frac{\|\bar{\mu}^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\}. \tag{13}$$

Proof Let us consider $k_0 \in \mathbb{N}$ such that $\rho_k \ge \bar{\rho}$ for all $k \ge k_0$ for $\bar{\rho}$ defined in Lemma 3.1.

Since the sequence $\{\bar{\mu}^k\}$ is bounded and $\|\alpha^k\| \to 0$, there exists $k_1 \ge k_0$ such that $\|\alpha_k\|$ and $\frac{1}{\rho_k}\|\bar{\mu}^k - \mu^*\|$ are small enough for all $k \ge k_1$.

By Lemma 3.3, define $(x^k, y^k) := (x(\bar{\mu}^k, \frac{1}{\rho_k}, \alpha_k), y(\bar{\mu}^k, \frac{1}{\rho_k}, \alpha_k))$. Then, $\nabla f(x^k) + \sum_{i=1}^p y_i^k \nabla g_i(x^k) = \alpha^k$, and by using that $\|\alpha^k\| \le \varepsilon_k$, we obtain that $\|x^k - x^*\| \le M \max\left\{\frac{\|\bar{\mu}^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\}$.

If $y^k = y(\bar{\mu}^k, \frac{1}{\rho_k}, \alpha_k)$, by item 1 of Lemma 3.3 we have that $y^k = \bar{\mu}^k e^{\rho_k g(x^k)}$ and then $y^k = \mu^{k+1}$. Thus, we obtain that $\|\mu^{k+1} - \mu^*\| \le M \max\left\{\frac{\|\bar{\mu}^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\}$ as we wanted to prove. □

Theorem 3.1 *Suppose that Assumptions 1–4 are satisfied by the sequence generated by Algorithm 2.1. In addition, assume that there exists a sequence $\eta_k \to 0$ such that, for all $k \in \mathbb{N}$,*

$$\varepsilon_k \le \eta_k \|\sigma^k\|. \tag{14}$$

Then, the sequence of penalty parameters $\{\rho_k\}$ is bounded.

Proof Assume, by contradiction, that $\lim_{k\to\infty} \rho_k = \infty$. Let us consider $k_1 \ge k_0$ large enough such that, for item 2 of Proposition 3.1, $\bar{\mu}^k = \mu^k$ for all $k \ge k_1$.

By the definition of σ^k and using Corollary 3.1, we have that, for all $k \ge k_1$,

$$\|\sigma^k\| = \left\|\frac{\mu^{k+1} - \mu^k}{\rho_k}\right\| \le \left\|\frac{\mu^{k+1} - \mu^*}{\rho_k}\right\| + \left\|\frac{\mu^* - \mu^k}{\rho_k}\right\| \le$$
$$\le \frac{M}{\rho_k} \max\left\{\frac{\|\mu^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\} + \frac{\|\mu^k - \mu^*\|}{\rho_k}.$$

We consider two possibilities:

1. If $\max\left\{\frac{\|\mu^k-\mu^*\|}{\rho_k}, \varepsilon_k\right\} = \frac{\|\mu^k-\mu^*\|}{\rho_k}$, then $\|\sigma^k\| \le \left(\frac{M}{\rho_k} + 1\right)\frac{\|\mu^k-\mu^*\|}{\rho_k}$.
2. If $\max\left\{\frac{\|\mu^k-\mu^*\|}{\rho_k}, \varepsilon_k\right\} = \varepsilon_k$, then $\|\sigma^k\| \le \frac{M}{\rho_k}\eta_k\|\sigma^k\| + \frac{\|\mu^k-\mu^*\|}{\rho_k}$, and we have that, for k large enough such that $\rho_k - M\eta_k > 0$, the inequality: $\|\sigma^k\| \le \frac{\rho_k}{\rho_k - M\eta_k}\frac{\|\mu^k-\mu^*\|}{\rho_k}$, holds.

Thus, for k large enough such that $\rho_k - M\eta_k > 0$ we obtain that

$$\|\sigma^k\| \le \frac{\|\mu^k - \mu^*\|}{\rho_k} \max\left\{\frac{M}{\rho_k} + 1, \frac{\rho_k}{\rho_k - M\eta_k}\right\}. \tag{15}$$

For another side, from (13)–(14): $\|\mu^k - \mu^*\| \le M\left(\frac{\|\mu^{k-1}-\mu^*\|}{\rho_{k-1}} + \eta_{k-1}\|\sigma^{k-1}\|\right)$. Thus,

$$\frac{\|\mu^{k-1} - \mu^*\|}{\rho_{k-1}} \ge \frac{\|\mu^k - \mu^*\|}{M} - \eta_{k-1}\|\sigma^{k-1}\|. \tag{16}$$

By definition of σ^{k-1} and considering (16),

$$\|\sigma^{k-1}\| = \frac{\|\mu^k - \mu^{k-1}\|}{\rho_{k-1}} \ge \frac{\|\mu^k - \mu^*\|}{M} - \eta_{k-1}\|\sigma^{k-1}\| - \frac{\|\mu^k - \mu^*\|}{\rho_{k-1}}.$$

Therefore, $(1 + \eta_{k-1})\|\sigma^{k-1}\| \ge \|\mu^k - \mu^*\|\left(\frac{1}{M} - \frac{1}{\rho_{k-1}}\right) \ge \frac{1}{2M}\|\mu^k - \mu^*\|$, and we have that $\|\mu^k - \mu^*\| \le 2M(1 + \eta_{k-1})\|\sigma^{k-1}\|$.

Thus, by (15), $\|\sigma^k\| \le \frac{2M(1+\eta_{k-1})}{\rho_k} \max\left\{\frac{M}{\rho_k} + 1, \frac{\rho_k}{\rho_k - M\eta_k}\right\}\|\sigma^{k-1}\|$.

If we assume that $\rho_k \to \infty$ and $\eta_k \to 0$, there exists k_2 such that

$\frac{2M(1+\eta_{k-1})}{\rho_k}\max\{\frac{M}{\rho_k}+1, \frac{\rho_k}{\rho_k-M\eta_k}\} < \tau$ for all $k \geq k_2$ and this shows that $\|\sigma^k\| \leq \tau\|\sigma^{k-1}\|, \forall k \geq k_2$. So, $\rho_{k+1} = \rho_k$ for all $k \geq k_2$ contradicting the hypothesis. Thus, $\{\rho_k\}$ is bounded. □

Remark 3.1 The proof of the previous theorem can be deduced from the ideas presented in [5], adapted to the exponential penalty function. It is worth noting that the authors' method of proof in [5]—to obtain the same thesis—is based on the reduction to the original problem to a problem with only equality constraints by using slack variables (see Section 5.2 in [5]). This technique cannot be adopted when the exponential penalty function is considered for inequality constraints.

4 Illustrative Examples

The objective of this section is to illustrate the behaviour of the exponential Augmented Lagrangian algorithm introduced in this paper by providing three small constrained problems. The main idea is to show that there are specific problems which the exponential penalty function can be more useful than the quadratic one to find local/global minimizers.

We have considered the ALGENCAN solver from TANGO software in www.ime.usp.br/~egbirgin/tango/. ALGENCAN is a Fortran code for general nonlinear programming which is able to solve extremely large size problems with moderate computer time. The general algorithm is of Augmented Lagrangian type; it is based on the multiplier method described in [5] which uses the quadratic penalty function.

We have introduced modifications of some subroutines of ALGENCAN with the objective of considering the exponential penalty function and its derivatives instead of the quadratic ones.

Other modifications of ALGENCAN that we have introduced are: the multiplier estimate is computed using formula (3) and the parameter measuring the progress in terms of infeasibility and complementarity is (5).

In [4,5], the infeasibility–complementarity measure for the quadratic case was $\sigma_j^k = \max\left\{g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k}\right\}$, and in the experiments in [9], the measure considered was $\sigma_j^k = \max\{g_j(x^k), -\bar{\mu}_j^k\}$, differently from the one used to obtain the convergence results, cited in (9).

As we have defined in Step 3 of Algorithm 2.1, the penalty parameter for inequality constraints is updated according to the following criteria: if $\|\sigma^k\| \leq \tau\|\sigma^{k-1}\|$ define $\rho_{k+1} = \rho_k$, else define $\rho_{k+1} = \gamma\rho_k$, where $\tau < 1, \gamma > 1$.

We have adopted the default parameters of ALGENCAN with respect to maximum number of inner and outer iterations, stopping criteria, precision of subproblems and other algorithmic parameters of subproblem solver. Also, we have used the same values for the parameters μ_1, ρ_1, τ,γ and $\bar{\mu}^{max}$. The modified version of ALGENCAN with the use of the exponential penalty function was called ALExp.

The experiments were run on a personal computer with INTEL (R) Core (TM) 2 Duo CPU E8400 at 3.00 GHz and 3.23 GB of RAM.

Example 4.1 For $x \in \mathbb{R}^2$, we consider the following indefinite quadratic problem:

$$\text{minimize } \left(x_1^2 - x_2^2\right) \qquad \text{s. t.} \qquad \|x\|^2 \leq 1.$$

By using the initial point (0.5, 0), ALGENCAN converges to the interior saddle point (0, 0) in one iteration, whereas ALExp converges to the global minimizer $(0, -1)$ in two iterations.

Example 4.2 For $x \in \mathbb{R}^2$, we consider the following problem (Problem 19 in [36]):

$$\begin{array}{rc} \text{minimize} & \left(x_1^4 - 14x_1^2 + 24x_1 - x_2^2\right) \\ \text{s. t.} & -x_1 + x_2 - 8 \leq 0, \ x_2 - x_1^2 - 2x_1 + 2 \leq 0 \\ & -8 \leq x_1 \leq 10, 0 \leq x_2 \leq 11. \end{array}$$

By using the initial point (0, 1), ALGENCAN converges to the saddle point (2, 0), with the functional value $f = 8.00$, in two iterations, whereas ALExp converges to the local minimizer (2.702, 10.702), with the functional value $f = -98.597$, in 10 iterations.

We have observed a similar behaviour considering other different initial points, for example: $(0, a)$ with $a = 2, 3, \ldots, 7$.

We also consider the following example to show the necessity to define methods based on second-order information in order to obtain global minimizers instead of local or saddle points.

Example 4.3 For $x \in \mathbb{R}^2$, we consider the following problem (Problem 1 in [36], Example 2 in [37]):

$$\begin{array}{rc} \text{minimize} & (-x_1 - x_2) \\ \text{s. t.} & x_1 x_2 \leq 4, \ 0 \leq x_1 \leq 6, \ 0 \leq x_2 \leq 4. \end{array}$$

For this problem, we have observed that by using the initial feasible point (2, 2), both methods stay in this saddle point.

5 Perspectives

The analysis we presented in this paper can be extended in several fields of interest. For instance, it is possible to verify whether we can replace the exponential penalty function by more general penalty functions, provided that they agree with some conditions the exponential function verifies. This extension, although less direct, is also worthy to investigate.

Example 4.3 shows the situation in which a saddle point is found considering methods based on first-order information. In our opinion, this example is an important motivation to study the second-order approach of the Augmented Lagrangian algorithm. Since the Lagrangian function of the problem in Example 4.3 has a negative curvature direction in (2, 2), we could use this information—as it was suggested in

[37]—to move to a stationary point verifying second-order necessary optimality conditions. This will be the subject of future research in the context of non-quadratic twice differentiable penalty functions. We believe that this study can be the key to define a method to solve optimization problems with degenerate or complementarity constraints. Our expectations for a good behaviour of these methods on degenerate or complementarity constraints problems are based on the attractive global and local convergence properties that the Augmented Lagrangian methods have.

Finally, since Augmented Lagrangian methods proceed by sequential resolution of simple (generally unconstrained or box-constrained) problems and they are useful in different contexts in the optimization field, we are interested in using the technique for solving the smooth multiobjective problem. As far as we know, the investigation of this approach for constrained multicriteria minimization problems is still very limited.

6 Conclusions

We have proposed an Augmented Lagrangian algorithm based on the exponential penalty function for inequality constrained optimization. One potential feature of our algorithm is that it possesses global convergence to first-order stationary points under the weak constant positive generator constraint qualification (CPG). We have proved that the proposed algorithm guarantees that the penalty parameter remains bounded away from zero; this happens when it is applied to solve problems where the linear independence constraint qualification, the strict complementarity and the second-order sufficient optimality condition are satisfied.

Acknowledgments The authors are grateful to the anonymous referees for their contribution to the final preparation of the paper.

References

1. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
2. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization. Academic Press, New York, pp. 283–298 (1969)
3. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM J. Control Optim. **12**, 0363–0129 (1974). Collection of articles dedicated to the memory of Lucien W. Neustadt
4. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math. Program. **112**, 5–32 (2008)
5. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**, 1286–1309 (2007)
6. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Athena Scientific, Belmont (1996)
7. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Belmont (1999)
8. Birgin, E.G., Castillo, R.A., Martínez, J.M.: Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. Comput. Optim. Appl. **31**(1), 31–55 (2005)
9. Birgin, E.G., Fernández, D., Martínez, J.M.: The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems. Optim. Methods Softw. **27**(6), 1001–1024 (2012)
10. Conn, A.R., Gould, N.I.M., Sartenaer, A., Toint, PhL: Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. SIAM J. Optim. **6**, 674–703 (1996)

11. Conn, A.R., Gould, N.I.M., Toint, PhL: A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. SIAM J. Numer. Anal. **28**, 545–572 (1991)
12. Conn, A.R., Gould, N.I.M., Toint, PhL: LANCELOT: A Fortran Package for Large Scale Nonlinear Optimization. Springer, Berlin (1992)
13. Diniz-Ehrhardt, M.A., Martínez, J.M., Pedroso, L.G.: Derivative-free methods for nonlinear programming with general lower-level constraints. Comput. Appl. Math. **30**, 19–52 (2011)
14. Huang, X.X., Yang, X.Q., Teo, K.L.: Partial augmented Lagrangian method and mathematical programs with complementarity constraints. J. Glob. Optim. **35**, 235–254 (2006)
15. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. SIAM J. Optim. **22**(4), 1579–1606 (2012)
16. Luo, H.Z., Sun, X.L., Xu, Y.F.: Convergence properties of modified and partially-augmented Lagrangian methods for mathematical programs with complementarity constraints. J. Optim. Theory Appl. **145**(3), 489–506 (2010)
17. Luo, H.Z., Sun, X.L., Xu, Y.F., Wu, H.X.: On the convergence properties of modified augmented Lagrangian methods for mathematical programming with complementarity constraints. J. Glob. Optim. **46**(2), 217–232 (2010)
18. Birgin, E.G., Floudas, C.A., Martínez, M.: Global minimization using an augmented Lagrangian method with variable lower-level constraints. Math. Program. **125**(1, Ser. A), 139–162 (2010)
19. Birgin, E.G., Martínez, J.M., Prudente, L.F.: Augmented Lagrangians with possible infeasibility and finite termination for global nonlinear programming. J. Glob. Optim. **58**, 207–242 (2014)
20. Kort, B.W., Bertsekas, D.P.: Multiplier methods for convex programming. In: Proceedings of the IEEE Decision and Control Conference, pp. 260–264. San Diego, CA (1973)
21. Kort, B.W., Bertsekas, D.P.: Combined primal-dual and penalty methods for convex programming. SIAM J. Control Optim. **35**, 1142–1168 (1976)
22. Cominetti, R., Dussault, J.P.: Stable exponential-penalty algorithm with superlinear convergence. J. Optim. Theory Appl. **83**(2), 285–309 (1994)
23. Dussault, J.P.: Augmented non-quadratic penalty algorithms. Math. Program. **99**, 467–486 (2004)
24. Iusem, A.N.: Augmented Lagrangian methods and proximal point methods for convex optimization. Invest. Op. **8**, 11–50 (1999)
25. Tseng, P., Bertsekas, D.P.: On the convergence of the exponential multiplier method for convex programming. Math. Program. **60**, 1–19 (1993)
26. Wang, C., Li, D.: Unified theory of augmented Lagrangian methods for constrained global optimization. J. Glob. Optim. **44**, 433–458 (2009)
27. Giannessi, F.: Constrained Optimization and Image Space Analysis, vol. 1. Separation of Sets and Optimality Conditions, vol. I. Springer, New York (2005)
28. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**(3), 1109–1135 (2012)
29. Mangasarian, O.L., Fromovitz, S.: Fritz John optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37–47 (1967)
30. Rockafellar, R.T.: Lagrange multipliers and optimality. SIAM Rev. **35**, 183–238 (1993)
31. Janin, R.: Directional derivative of the marginal function in nonlinear programming. Math. Program. Study **21**, 110–126 (1984)
32. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011)
33. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**(1–2, Ser. A), 255–273 (2012)
34. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**(1, Ser. A), 91–124 (2002)
35. Hager, W.W., Gowda, M.S.: Stability in the presence of degeneracy and error estimation. Math. Program. **85**(1, Ser. A), 181–192 (1999)
36. Ryoo, N.V., Sahinidis, N.D.: Global optimization of nonconvex NLPs and MINLPs with applications in process design. Comput. Chem. Eng. **19**, 551–566 (1995)
37. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Second-order negative-curvature methods for box-constrained and general constrained optimization. Comput. Optim. Appl. **45**, 209–236 (2010)