CrossMark

Augmented Lagrangian methods for nonlinear programming with possible infeasibility

M. L. N. Gonçalves · J. G. Melo · L. F. Prudente

Abstract In this paper, we consider a nonlinear programming problem for which the constraint set may be infeasible. We propose an algorithm based on a large family of augmented Lagrangian functions and analyze its global convergence properties taking into account the possible infeasibility of the problem. We show that, in a finite number of iterations, the algorithm stops detecting the infeasibility of the problem or finds an approximate feasible/optimal solution with any required precision. We illustrate, by means of numerical experiments, that our algorithm is reliable for different Lagrangian/penalty functions proposed in the literature.

1 Introduction

It is well known that augmented Lagrangian methods are efficient to solve constrained optimization problems in different areas of applications. The augmented Lagrangian for optimization problems with equality constraints was proposed by Hestenes [12] and Powell [23] and extended by Rockafellar for inequality constraints [24]. The Powell–Hestenes–Rockafellar (PHR) augmented Lagrangian function has been widely used in practice, see for example [1,8] related to the packages Algencan and Lancelot, respectively.

In each iteration of an augmented Lagrangian method, a subproblem with simple constraints is approximately solved. One drawback of the PHR approach is that the objective functions of the subproblems are not twice continuously differentiable, which can result in

M. L. N. Gonçalves · J. G. Melo · L. F. Prudente (✉)
Institute of Mathematics and Statistics, Federal University of Goias, Campus II,
Caixa Postal 131, Goiânia, GO CEP 74001-970, Brazil
e-mail: lfprudente@ufg.br

M. L. N. Gonçalves
e-mail: maxlng@ufg.br

J. G. Melo
e-mail: jefferson@ufg.br

difficulties to apply a Newton-like method. This motivates the study of different augmented Lagrangian functions. Many nonlinear Lagrangian/penalty functions have been proposed and/or analyzed by many authors, see for example [2,3,6,11,14,16,17,21,22,25,29].

It is desirable that optimization algorithms have effective stopping criteria. Ideally, an algorithm should stop at an iterate for which it is possible to ensure feasibility and optimality up to arbitrarily small given precisions. On the other hand, if the problem is infeasible, the algorithm should stop with a certificate of infeasibility. In the context of global optimization and PHR augmented Lagrangian methods, this issue was considered in [5]. In the local optimization field, strategies handling possible infeasibility were considered, for example, in [7,19,30] in the context of SQP, augmented Lagrangian and Interior Point methods, respectively. In some practical applications, it may be useful to accept some level of infeasibility (for example when the constraints involve modeling errors) and distinguish infeasible points according to their objective function values. These issues were considered in [4].

In this paper, we consider a large family of augmented Lagrangian methods introduced by Kort and Bertsekas [2,14]. We propose an algorithm based on this class and, accepting inexact global solutions of the subproblems, study its asymptotic convergence properties. We prove that the limit points of the sequence generated by the algorithm are global minimizers of an infeasibility measure, which is specified for some penalty functions. Based on these results, a second algorithm is proposed, which stops after a finite number of iterations at an approximate feasible/optimal solution or detects that the problem is infeasible. In particular, we extend some results of [5] to our more general setting. We also present some numerical experiments illustrating the applicability of the algorithm.

This paper is organized as follows. In Sect. 2, we consider the optimization problem and the class of Lagrangian/penalty functions to be analyzed. In Sect. 3, we propose an algorithm and examine its asymptotic convergence properties. We also propose a second algorithm that takes into account the possible infeasibility of the problem. In Sect. 4, we present some numerical experiments analyzing our algorithm for some Lagrangian/penalty functions. Last section is devoted to conclusions and final remarks.

Notation The α-norm and maximum norm of $x \in \mathbb{R}^m$ are denoted, respectively, by $\|x\|_\alpha := \left(\sum_{i=1}^m |x_i|^\alpha\right)^{1/\alpha}$ and $\|x\|_\infty := \max(|x_1|, \ldots, |x_m|)$, when $\alpha = 2$, we use the notation $\|\cdot\|$. We denote $\mathbb{R}_+$ and $\mathbb{R}_{++}$ the sets of non-negative and strictly positive real numbers, respectively. Let $v \in \mathbb{R}^n$, then $v_+ := (\max\{0, v_1\}, \ldots, \max\{0, v_n\})$. For $K = (k_1, k_2, \ldots) \subseteq \mathbb{N}$ (with $k_j < k_{j+1}$ for all j), we denote $K \underset{\infty}{\subset} \mathbb{N}$.

2 General Lagrangian function

In this paper, we consider the inequality constrained optimization problem

$$
\begin{aligned}
\text{Minimize} \quad & f(x) \\
\text{subject to} \quad & g(x) \leq 0 \\
& x \in \Omega,
\end{aligned}
\tag{1}
$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are continuous and $\Omega \subset \mathbb{R}^n$ is a nonempty compact set. In general, Ω is defined by simple constraints such as affine constraints or box constraints.

In order to solve problem (1) or identify if the feasible set is empty, we propose an algorithm based on the general class of Lagrangians introduced by Kort and Bertsekas[2,14]

$$L(x, \lambda, \rho) = f(x) + \frac{1}{\rho} \sum_{i=1}^{m} W(\rho g_i(x), \lambda_i), \tag{2}$$

where $x \in \Omega$, $\lambda \in \mathbb{R}_+^m$, $\rho > 0$, and the penalty function $W : \mathbb{R}^2 \to \mathbb{R}$ satisfies the following conditions:

(H1) W is continuously differentiable on $\mathbb{R} \times]0, \infty[$ and possesses, for all $s \in \mathbb{R}$, the right derivative

$$\lim_{t \to 0^+} \frac{W(s, t) - W(s, 0)}{t};$$

(H2) $W(s, .)$ is concave on $\mathbb{R}_+$ for each fixed $s \in \mathbb{R}$;

(H3) for each fixed $t \in \mathbb{R}_+$, $W(., t)$ is convex on $\mathbb{R}$ and satisfies the following strict convexity condition: If $s_0 > 0$ or $W_s'(s_0, t) > 0$, then $W(s, t) - W(s_0, t) > (s - s_0) W_s'(s_0, t)$ for $s \neq s_0$, where $W_s'(s, t)$ denotes the partial derivative of $W(s, t)$ with respect to s;

(H4) $W(0, t) = 0$, $W_s'(0, t) = t$, for all $t \in \mathbb{R}_+$;

(H5) $\lim_{s \to -\infty} W_s'(s, t) = 0$, for all $t \in \mathbb{R}_+$;

(H6) $\inf_{s \in \mathbb{R}} W(s, t) > -\infty$, for all $t \in \mathbb{R}_+$;

(H7) $\lim_{s \to \infty} \frac{W(s,t)}{s} = \infty$, for all $t \in \mathbb{R}_+$.

Remark 1 Since by (H3) $W(., t)$ is convex, it follows that $W'(., t)$ is monotone. Hence, (H5) implies

$$W_s'(s, t) \geq 0 \qquad \forall (s, t) \in \mathbb{R} \times \mathbb{R}_+,$$

which shows that $W(\cdot, t)$ is non-decreasing for every $t \in \mathbb{R}_+$. It is also possible to show that (H7) is equivalent to $\lim_{s \to \infty} W_s'(s, t) = \infty$, for all $t \in \mathbb{R}_+$.

The family of penalty functions satisfying (H1)–(H7) is very large. Next, we present two subclasses of functions satisfying these conditions. The first one generalizes the classical quadratic augmented Lagrangian, and the second one contains some exponential type functions. In particular, every function in the second family is twice continuously differentiable. For more details on these functions and other examples, see for instance [2].

Example 1 Let $W : \mathbb{R}^2 \to \mathbb{R}$ such that

$$W(s, t) = \begin{cases} ts + \phi(s) & \text{if } t + \phi'(s) \geq 0, \\ \min_{\tau \in \mathbb{R}} [t\tau + \phi(\tau)] & \text{otherwise,} \end{cases} \tag{3}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ satisfies the following conditions:

(E1) ϕ is continuously differentiable and strictly convex on $\mathbb{R}$;

(E2) $\phi(0) = 0$, $\phi'(0) = 0$, $\lim_{s \to -\infty} \phi'(s) = -\infty$;

(E3) $\lim_{s \to \infty} \frac{\phi(s)}{|s|} = \infty$.

It is not difficult to prove that the penalty function $W(s, t)$ defined in (3) satisfies (H1)–(H7), see [2]. In the following, we consider some examples of ϕ satisfying conditions (E1)–(E3):

(i) $\phi(s) = \alpha^{-1} |s|^\alpha$, $\alpha > 1$;

(ii) $\phi(s) = \alpha^{-1} |s|^\alpha + (1/2)s^2$, $\alpha > 1$;

(iii) $\phi(s) = \cosh(s) - 1$.

Taking $\phi(s) = (1/2)s^2$ in (3), then

$$W(s, t) = \begin{cases} ts + (1/2)s^2 & \text{if } s \geq -t, \\ -(1/2)t^2 & \text{if } s < -t. \end{cases} \tag{4}$$

Hence, after simple calculations, the corresponding augmented Lagrangian function (2) can be rewritten as

$$L(x, \lambda, \rho) = f(x) + \frac{1}{2\rho} \sum_{i=1}^{m} \{[\max\{0, \lambda_i + \rho g_i(x)\}]^2 - (\lambda_i)^2\}, \tag{5}$$

which is the PHR augmented Lagrangian function [24]. Now, if $\phi(s) = \alpha^{-1}|s|^\alpha$, where $\alpha > 1$, or $\phi(s) = \cosh(s) - 1$, we obtain the following penalty functions, respectively,

$$W(s, t) = \begin{cases} ts + \alpha^{-1}|s|^\alpha & \text{if } t + \alpha^{-1}\frac{d}{ds}|s|^\alpha \geq 0, \\ \min_{\tau \in \mathbb{R}}[t\tau + \alpha^{-1}|\tau|^\alpha] & \text{otherwise,} \end{cases} \tag{6}$$

and

$$W(s, t) = \begin{cases} ts + \cosh(s) - 1 & \text{if } t + \sinh(s) \geq 0, \\ t\sinh^{-1}(-t) + \cosh(\sinh^{-1}(-t)) - 1 & \text{otherwise.} \end{cases} \tag{7}$$

(Recall that $\sinh^{-1}(x) = \ln(x + \sqrt{x^2 + 1})$.)

We mention that, similarly to the quadratic augmented Lagrangian, the penalty functions of Example 1 may not be twice continuously differentiable. Next, we present some examples of penalty functions with the advantage that the corresponding augmented Lagrangian functions are twice continuously differentiable, see [2] for details.

Example 2 Let $W : \mathbb{R}^2 \to \mathbb{R}$ such that

$$W(s, t) = t\psi(s) + \xi(s), \tag{8}$$

where $\psi, \xi : \mathbb{R} \to \mathbb{R}$ satisfy the following conditions:

(A1) ψ and ξ are twice continuously differentiable functions and convex;
(A2) $\psi(0) = 0$, $\psi'(0) = 1$ and $\psi''(s) > 0$ for all $s \in \mathbb{R}$;
(A3) $\lim_{s \to -\infty} \psi(s) > -\infty$, $\lim_{s \to -\infty} \psi'(s) = 0$ and $\lim_{s \to \infty} \psi'(s) = \infty$;
(A4) $\xi(s) = 0$ for all $s \leq 0$ and $\xi(s) > 0$ for all $s > 0$;
(A5) $\xi''(0) = 0$ and $\lim_{s \to \infty} \xi'(s) = \infty$.

An interesting example in this family of penalty functions is obtained by taking $\psi(s) = e^s - 1$, $\xi(s) = \frac{1}{3}[\max\{0, s\}]^3$. In this case,

$$W(s, t) = t(e^s - 1) + \frac{1}{3}[\max\{0, s\}]^3 \tag{9}$$

and the associated augmented Lagrangian is

$$L(x, \lambda, \rho) = f(x) + \frac{1}{\rho} \sum_{i=1}^{m} \left(\lambda_i(e^{\rho g_i(x)} - 1) + \frac{1}{3}[\max\{0, \rho g_i(x)\}]^3\right).$$

Another example of a twice continuously differentiable penalty function is obtained by taking ψ and ξ as

$$\psi(s) = \begin{cases} s + s^2 & \text{if } s \geq 0, \\ s/(1 - s) & \text{if } s < 0, \end{cases} \quad \text{and} \quad \xi(s) = \begin{cases} s^3 & \text{if } s \geq 0, \\ 0 & \text{if } s < 0. \end{cases}$$

Thus, the corresponding penalty function (8) becomes

$$W(s,t) = \begin{cases} ts + ts^2 + s^3 & \text{if } s \geq 0, \\ ts/(1-s) & \text{if } s < 0. \end{cases} \tag{10}$$

In the following, we present a lemma which is essential for the analysis of our algorithm.

Lemma 1 Let $s \in \mathbb{R}$, $t \geq 0$ and a bounded sequence $\{t_k\} \subset \mathbb{R}_+$. Then

(i) $W'_t(s,t) \geq s$;

(ii) $sW'_s(s,t) \geq W(s,t) \geq tW'_t(s,t) \geq st$;

(iii) $W'_s(s,t) = t$ implies $s \leq 0$, $st = 0$;

(iv) $\lim_{k \to \infty} \frac{1}{r_k} \inf_{s \in \mathbb{R}} W(s,t_k) = 0$, where $r_k \to \infty$.

Proof For the proofs of items (i)–(iii), see [2, Proposition 5.1]. Let us prove item (iv). It follows by assumptions (H4) and (H6) that

$$0 \geq \inf_{s \in \mathbb{R}} W(s,t) > -\infty.$$

From (H2) we have $q(t) := \inf_{s \in \mathbb{R}} W(s,t)$ is concave $[0, +\infty[$. Let $0 < t_k \leq b$, $v_k \in \partial q(t_k)$ and $v \in \partial q(b)$. Therefore, using the supergradient inequality $q(0) \leq q(t_k) - v_k t_k$ and the non-increasing property of superdifferential ($t_k \leq b \Rightarrow v \leq v_k$), we obtain

$$q(t_k) \geq q(0) - b|v|, \qquad 0 < t_k \leq b.$$

Since the last estimate also holds for $t_k = 0$ and $q(t) \leq 0$ for all $t \geq 0$, it follows that $q(t_k)$ is bounded for $0 \leq t_k \leq b$. Therefore, as $r_k \to \infty$, we have

$$\lim_{k \to \infty} \frac{1}{r_k} \inf_{s \in \mathbb{R}} W(s,t^k) = 0.$$

$\square$

3 Algorithms

In this section, we study global convergence properties of an augmented Lagrangian method associated with the Lagrangian function defined in (2). We state the algorithm and show that its generated sequence satisfies an asymptotic infeasibility measure. Based on this result, we show some infeasibility measures satisfied by the limit points of the sequence generated by the algorithm when specific examples of penalty functions are considered. We also analyze some properties of the algorithm, which are essential for the feasibility/infeasibility test presented in the second algorithm. In order to state the algorithm, let $\{\varepsilon_k\} \subset \mathbb{R}_+$ be a bounded sequence, which is used to estimate the quality of approximate solutions of the augmented Lagrangian subproblems. Also, let us define $\sigma(x, \lambda, \rho) := (\sigma_1(x, \lambda, \rho), \ldots, \sigma_m(x, \lambda, \rho))^T$ with

$$\sigma_i(x, \lambda, \rho) := W'_s(\rho g_i(x), \lambda_i) - \lambda_i, \quad \text{for } i = 1, \ldots, m. \tag{11}$$

It follows by item (iii) of Lemma 1 that $\|\sigma(x, \lambda, \rho)\|_\infty$ may be used to measure the feasibility and complementarity condition of an approximate solution x and a multiplier λ. We shall consider this measure in order to update the penalty parameter.

Algorithm 1 Let $\gamma > 1$, $0 < \tau < 1$, $\rho_1 > 0$, $\lambda_{\max} > 0$, and $\lambda^1 \in \mathbb{R}^m$ such that $\lambda_i^1 \in [0, \lambda_{\max}]$, for all $i = 1, \ldots, m$. Assume that $\{\varepsilon_k\}$ is a bounded positive sequence and initialize $k \leftarrow 1$.

Step 1. **(Subproblem)**
Find $x_k \in \Omega$ such that

$$L(x^k, \lambda^k, \rho_k) \leq L(x, \lambda^k, \rho_k) + \varepsilon_k \tag{12}$$

for all $x \in \Omega$.

Step 2. **(Update penalty parameter)**
If $k = 1$ or

$$\|\sigma(x^k, \lambda^k, \rho_k)\|_\infty \leq \tau \, \|\sigma(x^{k-1}, \lambda^{k-1}, \rho_{k-1})\|_\infty, \tag{13}$$

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

Step 3. **(Update multipliers)**
Compute $\lambda_i^{k+1} \in [0, \lambda_{\max}]$, for all $i = 1, \ldots, m$.
Set $k \leftarrow k + 1$ and go to Step 1.

We mention that, as in many augmented Lagrangian methods, solving the subproblems is usually a difficult task. It is assumed that to solve a sequence of optimization problems with simple constraints is much less difficult than to solve the constrained optimization problem (1). In order to alleviate this task, we will accept approximate solutions of the subproblems. In the first iterations the tolerance ε_k may be considered not so stringent and one shall start to consider it smaller during the execution of the algorithm. Moreover, it follows from Remark 1 that, if problem (1) is convex, then the subproblems are also convex, which implies that every stationary point is a global solution.

In order to prove the theoretical results, we require the boundedness of the Lagrange multipliers, see Step 3 of Algorithm 1. In practice, the rule to update the Lagrangian multipliers is determinant for the performance of the algorithm. For the family of augmented Lagrangian functions studied in this paper, we consider the safeguard strategy to update the Lagrangian multipliers given by

$$\lambda_i^{k+1} = \min\{W_s'(\rho_k g_i(x^k), \lambda_i^k), \lambda_{\max}\}, \text{ for all } k, \text{ and } i = 1, \ldots, m. \tag{14}$$

A similar strategy is also considered in [1,3,5,16].

In the following, we present some properties of Algorithm 1. These properties will be summarized and they will be the motivations for our last algorithm, for which we will introduce an infeasibility test. Let us consider the function $M : \mathbb{R}^m \times \mathbb{R}_+^m \times \mathbb{R}_{++} \to \mathbb{R}$ defined by

$$M(y, \lambda, \rho) = \frac{1}{\rho} \sum_{i=1}^m W(\rho y_i, \lambda_i). \tag{15}$$

Proposition 2 *Assume that $\{x^k\}$ is a sequence generated by Algorithm 1. Then, for all $z \in \Omega$,*

$$\limsup_{k \to \infty} \frac{M(g(x^k), \lambda^k, \rho_k)}{\beta(\rho_k)} \leq \limsup_{k \to \infty} \frac{M(g(z)_+, \lambda^k, \rho_k)}{\beta(\rho_k)}, \tag{16}$$

for any continuous function $\beta : \,]0, \infty[\, \to \mathbb{R}_{++}$ such that $\lim_{s \to \infty} \beta(s) = \infty$. In particular, if problem (1) is feasible, then every limit point of $\{x^k\}$ is feasible.

Proof Let $x^* \in \Omega$ be an arbitrary limit point of $\{x^k\}$ (such a point exists because Ω is compact). Let $K \subset_\infty \mathbb{N}$ such that $\lim_{k \in K} x^k = x^*$. As $\{\rho_k\}$ is non-decreasing, we may consider two cases: $\{\rho_k\}$ bounded and $\rho_k \to \infty$.

Case 1 ($\{\rho_k\}$ bounded): By Step 2 of Algorithm 1, we have $\lim_{k \to \infty} \|\sigma(x^k, \lambda^k, \rho_k)\|_\infty = 0$, or equivalently,

$$\lim_{k \to \infty} [W_s'(\rho_k g_i(x^k), \lambda_i^k) - \lambda_i^k] = 0, \quad i = 1, \ldots, m. \tag{17}$$

Since $\{\lambda^k\}$ and $\{\rho_k\}$ are bounded, we may assume, without loss of generality, $\lambda^k \to \bar{\lambda} \geq 0$, and $\rho_k \to \bar{\rho} > 0$. So, (17) implies

$$W_s'(\bar{\rho} g_i(x^*), \bar{\lambda}_i) = \bar{\lambda}_i \quad i = 1, \ldots, m.$$

This result, combined with item (iii) of Lemma 1, gives $g(x^*) \leq 0$, which trivially implies the second part of the proposition when $\{\rho_k\}$ is bounded. Now, since $g_i(x^*) \leq 0$, for all $i = 1, \ldots, m$, $W(\cdot, t)$ is non-decreasing for every $t \geq 0$ and $W(0, \bar{\lambda}_i) = 0$, we obtain

$$\limsup_{k \to \infty} \frac{1}{\rho_k \beta(\rho_k)} \sum_{i=1}^{m} W(\rho_k g_i(x^k), \lambda_i^k) \leq 0. \tag{18}$$

Similarly, observing that $0 \leq \max\{0, g_i(z)\} = g_i(z)_+$, we have

$$\limsup_{k \to \infty} \frac{1}{\rho_k \beta(\rho_k)} \sum_{i=1}^{m} W(\rho_k g_i(z)_+, \lambda_i^k) = \frac{1}{\bar{\rho}\beta(\bar{\rho})} \sum_{i=1}^{m} W(\bar{\rho} g_i(z)_+, \bar{\lambda}_i)$$

$$\geq \frac{1}{\bar{\rho}\beta(\bar{\rho})} \sum_{i=1}^{m} W(0, \bar{\lambda}_i) = 0. \tag{19}$$

Therefore, (16) is implied by (18) and (19).

Case 2 ($\rho_k \to \infty$): For any $z \in \Omega$, we have, by Step 1 of Algorithm 1

$$L(x^k, \lambda^k, \rho_k) \leq L(z, \lambda^k, \rho_k) + \varepsilon_k, \quad \forall \, k.$$

Since $g_i(z) \leq g_i(z)_+$, for all $i = 1, \ldots, m$, $W(\cdot, t)$ is non-decreasing for every $t \geq 0$ and taking into account (2), the last inequality becomes

$$f(x^k) + \frac{1}{\rho_k} \sum_{i=1}^{m} W(\rho_k g_i(x^k), \lambda_i^k) \leq f(z) + \frac{1}{\rho_k} \sum_{i=1}^{m} W(\rho_k g_i(z)_+, \lambda_i^k) + \varepsilon_k, \quad \forall \, k.$$

Hence, using the functions β and M [see (15)], it follows from the last inequality that

$$\frac{M(g(x^k), \lambda^k, \rho_k)}{\beta(\rho_k)} \leq \frac{M(g(z)_+, \lambda^k, \rho_k)}{\beta(\rho_k)} + \frac{\varepsilon_k + f(z) - f(x^k)}{\beta(\rho_k)}.$$

As $\{\varepsilon_k\}$ is bounded, f is continuous, $\lim_{\rho_k \to \infty} \beta(\rho_k) = \infty$, and Ω is compact, then (16) follows from the last inequality. Thus, the first part of the proposition is concluded.

Observing that when $\{\rho_k\}$ is bounded, the last assertion of the proposition has already been proved, we may assume that $\rho_k \to \infty$. Let $z \in \Omega$ be a global minimizer of (1). Hence, using that $\max\{0, g_i(z)\} = 0$, $W(0, \lambda_i^k) = 0$ and (16), we obtain

$$\limsup_{k \to \infty} \frac{M(g(x^k), \lambda^k, \rho_k)}{\beta(\rho_k)} \leq 0, \tag{20}$$

for any continuous function $\beta : \,]0, \infty[\, \to \mathbb{R}_{++}$ such that $\lim_{s \to \infty} \beta(s) = \infty$. Suppose, by contradiction, that x^* is not feasible for problem (1). Thus, there exists some i_0 such that $g_{i_0}(x^*) > 0$. Hence, for some $\delta > 0$, it follows from the continuity of g_{i_0} that there exist

$k_0 > 0$ such that $g_{i_0}(x^k) \geq \delta$ for all $k \geq k_0, k \in K$. Consequently, using that $W(\cdot, t)$ is non-decreasing for every $t \geq 0$, we obtain

$$M(g(x^k), \lambda^k, \rho_k) = \frac{1}{\rho_k} \sum_{i \neq i_0} W(\rho_k g_i(x^k), \lambda_i^k) + \frac{W(\rho_k g_{i_0}(x^k), \lambda_{i_0}^k)}{\rho_k}$$

$$\geq \frac{1}{\rho_k} \sum_{i \neq i_0} W(\rho_k g_i(x^k), \lambda_i^k) + \frac{W(\rho_k \delta, \lambda_{i_0}^k)}{\rho_k}$$

$$\geq \frac{1}{\rho_k} \sum_{i \neq i_0} \inf_{s \in \mathbb{R}} W(s, \lambda_i^k) + \frac{W(\rho_k \delta, 0)}{\rho_k}, \tag{21}$$

where we have used properties of infimum and item (i) of Lemma 1 in the last inequality. It follows by item (iv) of Lemma 1 and properties of function β that

$$\lim_{k \to \infty} \frac{1}{\rho_k \beta(\rho_k)} \sum_{i \neq i_0} \inf_{s \in \mathbb{R}} W(s, \lambda_i^k) = 0. \tag{22}$$

Now, taking $\beta(s) = W(\delta s, 0)/s$, (H7) implies $\lim_{s \to \infty} \beta(s) = \infty$. Hence, it follows by (21) and (22) that

$$\limsup_{k \to \infty} \frac{M(g(x^k), \lambda^k, \rho_k)}{\beta(\rho_k)} \geq 1,$$

which is a contradiction with (20), concluding the proof. $\square$

Proposition 2 presents an asymptotic infeasibility measure associated with the general augmented Lagrangian function (2). In the following, we specify this infeasibility measure for some specific penalty functions.

Corollary 3 *Let $\{x^k\}$ be an infinite sequence generated by* Algorithm 1 *with W defined in* (6). *For any x^* limit point of $\{x^k\}$, and for all $z \in \Omega$ we have*

$$\|g(x^*)_+\|_\alpha \leq \|g(z)_+\|_\alpha. \tag{23}$$

Proof Let $K \underset{\infty}{\subseteq} \mathbb{N}$ be such that $\lim_{k \in K} x^k = x^*$. Using the definition in (6) and $\rho_k > 0$ for all $k \in \mathbb{N}$, we obtain

$$W(\rho_k g_i(x^k), \lambda_i^k)$$
$$= \begin{cases} \rho_k \lambda_i^k g_i(x^k) + \alpha^{-1} \rho_k^\alpha |g_i(x^k)|^\alpha & \text{if } g_i(x^k) \geq 0, \\ \rho_k \lambda_i^k g_i(x^k) + \alpha^{-1} \rho_k^\alpha |g_i(x^k)|^\alpha & \text{if } g_i(x^k) < 0, \ \lambda_i^k \geq \rho_k^{\alpha-1} |g_i(x^k)|^{\alpha-1}, \\ (1-\alpha)\alpha^{-1}(\lambda_i^k)^{\alpha/(\alpha-1)} & \text{if } g_i(x^k) < 0, \ \lambda_i^k < \rho_k^{\alpha-1} |g_i(x^k)|^{\alpha-1}. \end{cases}$$

First, consider the case in which $\rho_k \to \infty$. It follows from this expression of W that

$$\limsup_{k \in K} \sum_{i=1}^m \frac{W(\rho_k g_i(x^k), \lambda_i^k)}{\rho_k^\alpha} = \sum_{i=1}^m \alpha^{-1} |\max\{0, g_i(x^*)\}|^\alpha.$$

It is easy to see that a similar argument shows

$$\limsup_{k \in K} \sum_{i=1}^m \frac{W(\rho_k \max\{0, g_i(z)\}, \lambda_i^k)}{\rho_k^\alpha} = \sum_{i=1}^m \alpha^{-1} |\max\{0, g_i(z)\}|^\alpha.$$

Therefore, by Proposition 2 with $\beta(s) = s^{\alpha-1}$, we conclude that (23) holds in the case $\rho_k \to \infty$.

In the case $\{\rho_k\}$ bounded, similarly to the first part of the proof of Proposition 2, we obtain that each limit point x^* of $\{x^k\}$ is feasible for problem (1), which trivially implies (23). ☐

Remark 2 For the PHR augmented Lagrangian function (5), Corollary 3 implies the infeasibility measure $\|g(x^*)_+\| \leq \|g(z)_+\|$, for every $z \in \Omega$ and x^* limit point of $\{x^k\}$ (see [5, Theorem 2.1] for a similar result). If we consider the penalty function (10), similarly to the proof of Corollary 3 with $\beta(s) = s^2$, we obtain $\|g(x^*)_+\|_3 \leq \|g(z)_+\|_3$, for every $z \in \Omega$ and x^* limit point of $\{x^k\}$.

Next, we present an infeasibility measure for the limit points of the sequence $\{x^k\}$ generated by Algorithm 1 when the exponential type function is considered.

Corollary 4 Assume that $\{x^k\}$ is a sequence generated by Algorithm 1 with W defined in (9). Let $x^* \in \Omega$ be a limit point of $\{x^k\}$ such that $\lim_{k \in K} x^k = x^*$ with $K \subseteq_\infty \mathbb{N}$. Suppose that if $g_i(x^*) > 0$ then $\limsup_{j \in K} \lambda_i^j > 0$. Then, for all $z \in \Omega$,

$$\|g(x^*)_+\|_\infty \leq \|g(z)_+\|_\infty. \tag{24}$$

Proof Let us assume that problem (1) is infeasible, otherwise, it follows from the last statement of Proposition 2 that (24) trivially holds. Suppose by contradiction that there exists $z \in \Omega$ such that $\|g(x^*)_+\|_\infty > \|g(z)_+\|_\infty$. Since problem (1) is infeasible, $\|g(z)_+\|_\infty > 0$ and $\rho_k \to \infty$ (see Step 2 of Algorithm 1). Therefore, there exists an index $i_0 \in \{1, \ldots, m\}$ and $\delta > 0$ such that

$$g_{i_0}(x^*) = g_{i_0}(x^*)_+ = \|g(x^*)_+\|_\infty > \|g(z)_+\|_\infty + \delta.$$

As g is continuous, there exists $k_0 > 0$ such that $g_{i_0}(x^k) > \|g(z)_+\|_\infty + \delta$, for any $k \geq k_0, k \in K$. In particular, it follows from our assumptions that $\limsup_{k \in K} \lambda_{i_0}^k > 0$. Observing that $0 \leq \lambda_i^k \leq \lambda_{\max}$ and considering the function M of (15) and W defined in (9), it follows that

$$M(g(x^k), \lambda^k, \rho_k) = \frac{1}{\rho_k} \sum_{i=1}^m \lambda_i^k (e^{\rho_k g_i(x^k)} - 1)$$

$$+ \frac{1}{3}[\max\{0, \rho_k g_i(x^k)\}]^3 \geq \frac{\lambda_{i_0}^k e^{\rho_k g_{i_0}(x^k)}}{\rho_k} - \frac{m\lambda_{\max}}{\rho_k}.$$

Therefore, since z is fixed, defining $\beta(s) = e^{s\|g(z)_+\|_\infty}$ we have

$$\limsup_{k \in K} \frac{M(g(x^k), \lambda^k, \rho_k)}{\beta(\rho_k)} \geq \limsup_{k \in K} \frac{1}{\rho_k} \left[\lambda_{i_0}^k e^{\rho_k(g_{i_0}(x^k) - \|g(z)_+\|_\infty)} - \frac{m\lambda_{\max}}{e^{\rho_k\|g(z)_+\|_\infty}} \right]$$

$$\geq \limsup_{k \in K} \lambda_{i_0}^k \frac{e^{\rho_k \delta}}{\rho_k}.$$

As $\limsup_{k \in K} \lambda_{i_0}^k > 0$, we obtain $\limsup_{k \in K} (\lambda_{i_0}^k e^{\rho_k \delta})/\rho_k = \infty$. Combining this result with previous inequalities, it yields

$$\limsup_{k \in K} \frac{M(g(x^k), \lambda^k, \rho_k)}{\beta(\rho_k)} = \infty. \tag{25}$$

On the other hand,

$$
\begin{aligned}
\frac{M(g(z)_+, \lambda^k, \rho_k)}{\beta(\rho_k)} &= \frac{1}{e^{\rho_k \|g(z)_+\|_\infty}} \sum_{i=1}^m \left[\frac{\lambda_i^k}{\rho_k} (e^{\rho_k g_i(z)_+} - 1) + \frac{1}{3\rho_k} (\max\{0, \rho_k g_i(z)_+\})^3 \right] \\
&\leq \sum_{i=1}^m \left[\frac{\lambda_i^k}{\rho_k} \left(e^{\rho_k (g_i(z)_+ - \|g(z)_+\|_\infty)} \right) + \frac{\rho_k^2 (\max\{0, g_i(z)_+\})^3}{3 e^{\rho_k \|g(z)_+\|_\infty}} \right] \\
&\leq \sum_{i=1}^m \frac{\lambda_i^k}{\rho_k} + \sum_{i=1}^m \frac{\rho_k^2}{3 e^{\rho_k \|g(z)_+\|_\infty}} (\max\{0, g_i(z)_+\})^3 \\
&\leq \frac{m\lambda_{\max}}{\rho_k} + \sum_{i=1}^m \frac{\rho_k^2}{3 e^{\rho_k \|g(z)_+\|_\infty}} (\max\{0, g_i(z)_+\})^3,
\end{aligned}
\tag{26}
$$

where the first and third inequalities are due to fact that the multipliers are non-negative and bounded by $\lambda_{\max}$, respectively. In the second one, we used $g_i(z)_+ \leq \|g(z)_+\|_\infty$, for $i = 1, \ldots, m$. Since $\{\rho_k\}$ tends to infinity, we obtain from (26)

$$
\limsup_{k \to \infty} \frac{M(g(z)_+, \lambda^k, \rho_k)}{\beta(\rho_k)} \leq 0,
$$

which combined with (25) gives a contradiction, by Proposition 2. The proof of the corollary follows.   □

Remark 3 We mention that the update rule for the Lagrange multipliers is general, see Step 3 of Algorithm 1. Therefore, the additional asymptotic condition of Corollary 4 may be forced during the progress of the algorithm. If the update rule (14) is considered, then the asymptotic condition is satisfied when $\|x_{k+1} - x_k\| \to 0$. We also mention that Corollary 4 holds for the penalty function defined in (7).

Next, we present an estimate which is used in Algorithm 2 to define a criterion to detect an approximate solution of problem (1). We shall consider for any k,

$$
\gamma_k := -M(g(x^k), \lambda^k, \rho_k) = -\frac{1}{\rho_k} \sum_{i=1}^m W(\rho_k g_i(x^k), \lambda_i^k).
\tag{27}
$$

Lemma 5 *Assume that $\{x^k\}$ is an infinite sequence generated by* Algorithm 1. *Suppose that problem (1) is feasible. Then, for any k*

$$
f(x^k) \leq f(z) + \gamma_k + \varepsilon_k,
$$

for all feasible point z.

Proof By Step 1 of Algorithm (1) and (2), we have

$$
f(x^k) + \frac{1}{\rho_k} \sum_{i=1}^m W(\rho_k g_i(x^k), \lambda_i^k) \leq f(z) + \frac{1}{\rho_k} \sum_{i=1}^m W(\rho_k g_i(z), \lambda_i^k) + \varepsilon_k, \quad \forall k, \text{ and } z \in \Omega.
\tag{28}
$$

Let z be a feasible point. Therefore, since $g(z) \leq 0$, $W(0, t) = 0$ for $t \geq 0$ and $W(\cdot, t)$ is non-decreasing for every $t \geq 0$, we have

$$
W(\rho_k g_i(z), \lambda_i^k) \leq W(0, \lambda_i^k) = 0, \quad \forall k \text{ and } i = 1, \ldots, m.
$$

Combining this result with (27) and (28), we obtain

$$f(x^k) \leq f(z) + \gamma_k + \varepsilon_k,$$

which proves the lemma.   □

In the following theorem, we present a criterion to detect infeasibility after a finite number of iterations of Algorithm 1. We assume that it is possible to obtain a bounded sequence $\{c_k\}$ satisfying

$$|f(x^k) - f(z)| \leq c_k \quad \text{for all} \quad z \in \Omega, \, \forall k. \tag{29}$$

In the numerical experiments section, we describe a way to obtain c_k. Clearly, since f is continuous and Ω is bounded, the sequence $\{c_k\}$ may be assumed to be bounded.

Theorem 6 *Assume that $\{x^k\}$ is an infinite sequence generated by* Algorithm 1. *Then, problem* (1) *is infeasible if, and only if, there exists $k \in \mathbb{N}$ such that*

$$\gamma_k + \varepsilon_k < -c_k, \tag{30}$$

where γ_k is defined in (27).

Proof Suppose the feasible region of (1) is non-empty. Therefore, it follows by (29) and Lemma 5 that, for all feasible point z,

$$\gamma_k + \varepsilon_k \geq f(x^k) - f(z) \geq -c_k, \qquad \forall k.$$

This means that the infeasibility test (30) fails to be fulfilled.

Reciprocally, suppose that problem (1) is infeasible. In this case, it follows from Step 2 of Algorithm 1 that $\rho_k \to \infty$. Also, for any subsequence $\{x^k\}_{k \in K}$, there exist $k_0 > 0$ and $i_0 \in \{1, \ldots, m\}$ such that $g_{i_0}(x^k) > \delta$ for some $\delta > 0$, and $k \in K, k \geq k_0$. Consequently, using $W(\cdot, t)$ is non-decreasing for every $t \geq 0$, we obtain for all $k \in K, k \geq k_0$

$$
\begin{aligned}
\gamma_k &= -\frac{1}{\rho_k} \sum_{i=1}^{m} W(\rho_k g_i(x^k), \lambda_i^k) \\
&= -\frac{1}{\rho_k} \sum_{i \neq i_0} W(\rho_k g_i(x^k), \lambda_i^k) - \frac{W(\rho_k g_{i_0}(x^k), \lambda_{i_0}^k)}{\rho_k} \\
&\leq -\frac{1}{\rho_k} \sum_{i \neq i_0} W(\rho_k g_i(x^k), \lambda_i^k) - \frac{W(\rho_k \delta, \lambda_{i_0}^k)}{\rho_k} \\
&\leq -\frac{1}{\rho_k} \sum_{i \neq i_0}^{m} \inf_{s \in \mathbb{R}} W(s, \lambda_i^k) - \frac{W(\rho_k \delta, 0)}{\rho_k},
\end{aligned} \tag{31}
$$

where we have used properties of infimum and item (i) of Lemma 1 to get the last inequality. It follows by item (iv) of Lemma 1 that

$$\lim_{k \to \infty} \frac{1}{\rho_k} \sum_{i \neq i_0}^{m} \inf_{s \in \mathbb{R}} W(s, \lambda_i^k) = 0.$$

On the other hand, (H7) implies

$$\lim_{k \to \infty} \frac{W(\rho_k \delta, 0)}{\rho_k} = \infty.$$

Therefore, it follows from (31) that

$$\lim_{k \to \infty} \gamma_k = -\infty.$$

Hence, the boundedness of $\{\varepsilon_k\}$ and $\{c_k\}$ implies the fulfillment of test (30), for k large enough. $\qquad \square$

In the following theorem, we show that an optimality condition holds after a finite number of iterations of Algorithm 1, under the assumption that problem (1) is feasible and $\{\varepsilon_k\}$ tends to zero.

Theorem 7 *Assume that $\{x^k\}$ is an infinite sequence generated by* Algorithm 1. *Let $K \subseteq_\infty \mathbb{N}$ and $x^* \in \Omega$ be such that $\lim_{k \in K} x^k = x^*$. Suppose that problem (1) is feasible and $\lim_{k \to \infty} \varepsilon_k = 0$. Then,*

$$\lim_{k \in K} \gamma_k = 0, \tag{32}$$

where γ_k is defined in (27). As a consequence, given $\varepsilon_{opt} > 0$, for large enough k,

$$\gamma_k + \varepsilon_k \le \varepsilon_{opt},$$

which in turn implies

$$f(x^k) \le f(z) + \varepsilon_{opt}, \qquad \forall \, z \text{ feasible.}$$

Proof It follows from Proposition 2 that x^* is feasible for problem (1). Since the feasible region of (1) is non-empty and compact, there exists a global minimizer $z \in \Omega$. By Lemma 5,

$$f(x^k) \le f(z) + \gamma_k + \varepsilon_k, \qquad \forall \, k. \tag{33}$$

First, we shall prove (32) in the case that $\rho_k \to \infty$. For this, note that item (iv) of Lemma 1 implies

$$\lim_{k \to \infty} \frac{1}{\rho_k} \sum_{i=0}^{m} \inf_{s \in \mathbb{R}} W(s, \lambda_i^k) = 0.$$

Using the definition of γ_k in (27), properties of infimum and the last equality, we obtain

$$\limsup_{k \in K} \gamma_k = \limsup_{k \in K} -\frac{1}{\rho_k} \sum_{i=1}^{m} W(\rho_k g_i(x^k), \lambda_i^k) \le \limsup_{k \in K} -\frac{1}{\rho_k} \sum_{i=1}^{m} \inf_{s \in \mathbb{R}} W(s, \lambda_i^k) = 0. \tag{34}$$

On the other hand, taking the infimum limit for $k \in K$ in (33) and using $\varepsilon_k \to 0$ and $f(z) \le f(x^*)$, we obtain

$$\liminf_{k \in K} \gamma_k \ge 0,$$

which, combined with (34), proves (32) in the case that $\rho_k \to \infty$.

Now, if $\{\rho_k\}$ is bounded, by step 2 of Algorithm 1, we have $\lim_{k \to \infty} \|\sigma(x^k, \lambda^k, \rho_k)\|_\infty = 0$, or equivalently,

$$\lim_{k \to \infty} [W_s'(\rho_k g_i(x^k), \lambda_i^k) - \lambda_i^k] = 0, \quad i = 1, \ldots, m. \tag{35}$$

As $\{\lambda^k\}$ and $\{\rho_k\}$ are bounded, we can assume that $\lambda^k \to \bar{\lambda} \geq 0$ and $\rho_k \to \bar{\rho} > 0$. So, (35) implies

$$W_s'(\bar{\rho} g_i(x^*), \bar{\lambda}_i) = \bar{\lambda}_i \quad i = 1, \ldots, m, \tag{36}$$

which, combined with item (iii) of Lemma 1, gives

$$g(x^*) \leq 0, \quad \bar{\lambda}_i g_i(x^*) = 0, \quad i = 1, \ldots, m. \tag{37}$$

Now, using item (ii) of Lemma 1, we have for all k and i

$$\lambda_i^k g_i(x^k) \leq \frac{1}{\rho_k} W(\rho_k g_i(x^k), \lambda_i^k) \leq g_i(x^k) W_s'(\rho_k g_i(x^k), \lambda_i^k).$$

Taking the limit for $k \in K$ in the last inequality, together with (36) and (37), it follows that

$$\lim_{k \to \infty} \frac{1}{\rho_k} [W(\rho_k g_i(x^k), \lambda_i^k)] = \bar{\lambda}_i g_i(x^*) = 0, \quad i = 1, \ldots, m,$$

which, combined with (27), yields (32) in the case $\{\rho_k\}$ bounded. Thus, the proof of (32) is concluded.

The second statement of the theorem follows from the first result and the assumption $\varepsilon_k \to 0$. In order to establish the last part of the theorem, combine the second one and Lemma 5. □

Due to the previous theorems, we are able to define a variation of Algorithm 1, for which we can guarantee finite termination with certificates of infeasibility or optimality up to given precisions. To state Algorithm 2, we assume that $\varepsilon_{\text{feas}} > 0$ and $\varepsilon_{\text{opt}} > 0$ are user-given tolerances for feasibility and optimality respectively.

Algorithm 2 *Let* $\gamma > 1, 0 < \tau < 1, \rho_1 > 0, \lambda_{\max} > 0$, *and* $\lambda^1 \in \mathbb{R}^m$ *such that* $\lambda_i^1 \in [0, \lambda_{\max}]$, *for all* $i = 1, \ldots, m$. *Assume that* $\{\varepsilon_k\}$ *is a bounded positive sequence and initialize* $k \leftarrow 1$.

Step 1 **(Subproblem)**
  Obtain $x^k \in \Omega$ *satisfying*

$$L(x^k, \lambda^k, \rho_k) \leq L(z, \lambda^k, \rho_k) + \varepsilon_k, \quad \forall z \in \Omega.$$

Step 2 **(Infeasibility test)**
  Let

$$\gamma_k := -\frac{1}{\rho_k} \sum_{i=1}^m [W(\rho_k g_i(x^k), \lambda_i^k)].$$

  Compute $c_k > 0$ *such that* $|f(x^k) - f(z)| \leq c_k$ *for all* $z \in \Omega$. *If*

$$\gamma_k + \varepsilon_k < -c_k,$$

  stop declaring **Infeasible problem**.
Step 3 **(Feasibility and optimality test)**
  If

$$\|g(x^k)_+\|_\infty \leq \varepsilon_{feas} \quad and \quad \gamma_k + \varepsilon_k \leq \varepsilon_{opt},$$

  stop declaring **Solution found**.

Step 4 **(Updating the penalty parameter)**
    If $k = 1$ or

$$\|\sigma(x^k, \lambda^k, \rho_k)\|_\infty \leq \tau \|\sigma(x^{k-1}, \lambda^{k-1}, \rho_{k-1})\|_\infty,$$

then define $\rho_{k+1} = \rho_k$. Otherwise, $\rho_{k+1} := \gamma \rho_k$.
Step 5 **(Updating the multipliers)**
    Compute $\lambda_i^{k+1} = \min\{W_s'(\rho_k g_i(x^k), \lambda_i^k), \lambda_{\max}\}$, for all $i = 1, \ldots, m$.
    Set $k \leftarrow k + 1$ and go to Step 1.

The next result is an immediate consequence of Theorems 6 and 7.

Theorem 8 *Let $\{x_k\}$ be generated by* Algorithm 2 *with $\lim_{k\to\infty} \varepsilon_k = 0$. Then, after a finite number of iterations, the algorithm finds an approximate solution x^k satisfying*

$$\|g(x^k)_+\| \leq \varepsilon_{feas}, \qquad f(x^k) \leq f(z) + \varepsilon_{opt}$$

for all $z \in \Omega$ such that $g(z) \leq 0$, or stops, certifying the problem is infeasible.

4 Numerical experiments

This section summarizes the results of the numerical experiments we carried out in order to verify the effectiveness of Algorithm 2, for solving problem (1) or identifying if it is infeasible. We implemented Algorithm 2 considering five different augmented Lagrangian functions (2) with the penalty functions (4), (6) with $\alpha = 4/3$, (7), (9) and (10), which will be denoted by augmented Lagrangian of type (ALtype) 1, 2, 3, 4 and 5, respectively.

Our algorithm was implemented as modification of the software Algencan, which is a Fortran 77 code, free available in http://www.ime.usp.br/~egbirgin/tango/ and based on the theory presented in [1]. Similarly to [3], our main modifications are the implementation of different augmented Lagrangian functions. The main difference of our codes and the ones implemented in [3] is a *multistart strategy* to solve the subproblems and the introduction of stopping criteria, described in Steps 2 and 3 of Algorithm 2. The experiments were run on a 3.4 GHz Intel(R) i7 with 4 processors, 8Gb of RAM and Linux operating system.

4.1 Implementation details

The performance of algorithms is, in general, sensitive to the choice of the initial parameters. For comparison purposes, we decided to keep the same configuration of the initial algorithmic parameters for all augmented Lagrangians versions. We set $\varepsilon_{feas} = \varepsilon_{opt} = 10^{-4}, \lambda^1 = 0, \lambda_{\max} = 10^{10}, \rho_1 = 10, \gamma = 10$ and $\tau = 0.5$. We put $\varepsilon_1 = 10^{-4}$ and $\varepsilon_k = \max\{10^{-(k+2)}, 10^{-8}\}$ for $k \geq 2$.

In all considered problems the set Ω is a box $\{x \in \mathbb{R}^n \mid l \leq x \leq u\}$. We use the local optimization package Gencan to solve the box-constrained subproblems. Gencan is an active-set method which uses a truncated Newton approach with line searches within the faces whereas, for leaving the faces, uses spectral projected gradient iterations, see [1] and references therein. In all implemented versions, we replaced each Hessian of the augmented Lagrangian-vector product by an incremental quotient, using the approximation

$$\nabla^2 L(x, \lambda, \rho)d \approx \frac{1}{t}(\nabla L(x + td, \lambda, \rho) - \nabla L(x, \lambda, \rho))$$

with a small value of t.

Step 1 of Algorithm 2 requires the global solution of the subproblems. However, we use Gencan, which is a local optimization algorithm, to solve the subproblems. In each outer iteration, we ran Gencan 30 times using different randomly generated initial points (feasible with respect to the box Ω) and we considered as solution the point with the smallest value of the objective function of the subproblem. Adopting this strategy, we assume that we are able to find a global minimizer of each subproblem. Let us call this strategy *multistart*.

For simplicity, let us denote the objective function of the k-th subproblem $L(x, \lambda^k, \rho_k)$ by $L(x)$. Given a tolerance $\bar{\varepsilon}_k > 0$, Gencan stops at point y if

$$\|y - P_\Omega(y - \nabla L(y))\| \leq \bar{\varepsilon}_k, \tag{38}$$

where $P_\Omega(x)$ is the Euclidean projection of x on the box Ω. Therefore, the inequality (38) holds for y equal to x^k. The tolerance $\bar{\varepsilon}_k$ used in the stopping criterion (38) of the k-th subproblem is related to the gap ε_k at Step 1 of Algorithm 2 by

$$\bar{\varepsilon}_k = \frac{\varepsilon_k}{\max\{1, d(x^k) + \|\nabla L(x^k)\|\}}, \tag{39}$$

where

$$d(x) = \|\max\{P_\Omega(x - \nabla L(x)) - l, u - P_\Omega(x - \nabla L(x))\}\|.$$

Let us justify the relation (39). Denote by x_*^k a global minimizer of the k-th subproblem. Since the box Ω is compact and convex, using properties of the Euclidean projection, we obtain

$$\left\langle x^k - \nabla L(x^k) - P_\Omega(x^k - \nabla L(x^k)), x_*^k - P_\Omega(x^k - \nabla L(x^k)) \right\rangle \leq 0.$$

Then,

$$\left\langle \nabla L(x^k), P_\Omega(x^k - \nabla L(x^k)) - x_*^k \right\rangle + \left\langle x^k - P_\Omega(x^k - \nabla L(x^k)), x_*^k - P_\Omega(x^k - \nabla L(x^k)) \right\rangle \leq 0,$$

and, by the Cauchy–Schwarz inequality,

$$\left\langle \nabla L(x^k), P_\Omega(x^k - \nabla L(x^k)) - x_*^k \right\rangle \leq \left\| x^k - P_\Omega(x^k - \nabla L(x^k)) \right\| \left\| x_*^k - P_\Omega(x^k - \nabla L(x^k)) \right\|.$$

Therefore, by the stopping criterion (38) and the definition of $d(x)$, we have

$$\left\langle \nabla L(x^k), P_\Omega(x^k - \nabla L(x^k)) - x_*^k \right\rangle \leq \bar{\varepsilon}_k d(x^k). \tag{40}$$

If $L(x)$ is a convex function, it follows by the gradient inequality that

$$L(x_*^k) \geq L(x^k) + \left\langle \nabla L(x^k), x_*^k - x^k \right\rangle,$$

or, equivalently,

$$L(x_*^k) \geq L(x^k) + \left\langle \nabla L(x^k), P_\Omega(x^k - \nabla L(x^k)) - x^k \right\rangle + \left\langle \nabla L(x^k), x_*^k - P_\Omega(x^k - \nabla L(x^k)) \right\rangle.$$

Hence, using the Cauchy–Schwarz inequality, stopping criterion (38) and (40), we obtain

$$L(x_*^k) \geq L(x^k) - \bar{\varepsilon}_k \left(d(x^k) + \|\nabla L(x^k)\| \right),$$

which, combined with (39), gives the desired inequality in Step 1 of Algorithm 2. Now, consider the case where $L(x)$ is not a convex function. If we assume that x^k is close enough

to x_*^k then the linear approximation $L(x^k) + \langle \nabla L(x^k), x_*^k - x^k \rangle$ of $L(x_*^k)$ is "acceptable". Therefore, we can proceed as in the convex case to conclude that the inequality in Step 1 of Algorithm 2 is "approximately" satisfied.

The constant c_k at Step 2 of Algorithm 2 is computed as suggested in [5]. By interval arithmetic, an interval $[f^{\min}, f^{\max}]$ is computed, such that $f^{\min} \leq f(x) \leq f^{\max}$ for all $x \in \Omega$. So, c_k is calculate by

$$c_k = \max\{f(x^k) - f^{\min}, f^{\max} - f(x^k)\}.$$

For interval analysis calculations we use the Intlib library [13].

4.2 Some feasible examples

The main purpose of this section is to illustrate the performance of Algorithm 2 with the five types of augmented Lagragian functions for some feasible problems. In the following experiments, we considered the set of problems with inequality constraints analyzed in [5]. These problems come from the global optimization literature. Each problem is feasible and the optimal value of the objective function is known. Therefore, it is possible to check if a particular solution given by Algorithm 2 is indeed a global minimizer.

In this section, we are also interested to check whether the *multistart* strategy to solve the subproblems and the relation (39) are appropriate or not. The reliability of these approaches will be corroborated if the problems are successfully solved. We understand that a problem has been successfully solved if Algorithm 2 stops at Step 3 and the final iterate x^* satisfies $f(x^*) \leq f(z) + \varepsilon_{opt}$ for all feasible z. In the following, we describe the problems considered in these section.

Problem 1. [9]

Minimize $-\left(\dfrac{k_1 x_1}{(1 + k_1 x_1)(1 + k_3 x_1)(1 + k_4 x_2)} + \dfrac{k_2 x_2}{(1 + k_1 x_1)(1 + k_2 x_2)(1 + k_4 x_2)} \right)$

subject to $\sqrt{x_1} + \sqrt{x_2} \leq 4$

$\qquad (10^{-5}, 10^{-5}) \leq x \leq (16, 16)$

$\qquad k_1 = 9.755988 \times 10^{-2}; k_2 = 0.99; k_3 = 3.919080; k_4 = 0.90k_3$

Problem 2. [26]

Minimize $-x_1 - x_2$

subject to $x_1 x_2 \leq 4$

$\qquad (0, 0) \leq x \leq (6, 4)$

Problem 3. [15]

Minimize $29.4x_1 + 18x_2$

subject to $-x_1 + 0.2458x_1^2/x^2 \leq -6$

$\qquad (0, 10^{-5}) \leq x \leq (115.8, 30)$

Problem 4. [10]

Minimize $x_1 + x_2$

subject to $x_1^2 + x_2^2 \leq 4$

$\qquad -x_1^2 - x_2^2 \leq -1$

$\qquad x_1 - x_2 \leq 1$

$\qquad -x_1 + x_2 \leq 1$

$\qquad (-2, -2) \leq x \leq (2, 2)$

Problem 5. [18]

Minimize $x_1^4 - 14x_1^2 + 24x_1 - x_2^2$

subject to $x_2 - x_1^2 - 2x_1 \leq -2$

$\qquad -x_1 + x_2 \leq 8$

$\qquad (-8, 0) \leq x \leq (10, 10)$

Problem 6. [28]

Minimize $2x_1 + x_2$

subject to $-16x_1 x_2 \leq -1$

$\qquad -4x_1^2 - 4x_2^2 \leq -1$

$\qquad (0, 0) \leq x \leq (1, 1)$

Problem 7. [28]

Minimize $-2x_1 x_2$

subject to $4x_1 x_2 + 2x_1 + 2x_2 \leq 3$

$\qquad (0, 0) \leq x \leq (1, 1)$

Table 1 Performance of Algorithm 2 with the five augmented Lagrangian functions in the problems 1–7

Problem	n	m	ALtype	Iter	$f(x^*)$	$\|g(x^*)_+\|_\infty$	$\gamma_k + \varepsilon_k$	Global
Problem 1	2	1	1	2	−3.8881364D−01	2.91D−05	9.78D−05	Y
			2	2	−3.8881470D−01	4.32D−05	9.76D−05	Y
			3	2	−3.8881363D−01	2.90D−05	9.78D−05	Y
			4	5	−3.8881148D−01	5.52D−07	5.83D−08	Y
			5	4	−3.8881400D−01	3.40D−05	−1.56D−06	Y
Problem 2	2	1	1	3	−6.6666667D+00	4.74D−11	1.00D−05	Y
			2	3	−6.6666667D+00	0.00D+00	1.00D−05	Y
			3	3	−6.6666667D+00	0.00D+00	1.00D−05	Y
			4	3	−6.6666667D+00	1.61D−08	1.00D−05	Y
			5	3	−6.6666667D+00	3.75D−08	9.99D−06	Y
Problem 3	2	1	1	3	3.7629193D+02	0.00D+00	1.02D−05	Y
			2	6	3.7629193D+02	8.17D−14	9.99D−09	Y
			3	3	3.7629193D+02	0.00D+00	1.00D−05	Y
			4	2	3.7629193D+02	2.12D−09	9.99D−05	Y
			5	3	3.7629193D+02	0.00D+00	1.00D−05	Y
Problem 4	2	4	1	3	−2.8284274D+00	6.77D−07	9.76D−06	Y
			2	3	−2.8284271D+00	5.06D−11	1.00D−05	Y
			3	3	−2.8284274D+00	6.64D−07	9.77D−06	Y
			4	3	−2.8284302D+00	8.73D−06	6.91D−06	Y
			5	3	−2.8284276D+00	1.31D−06	9.54D−06	Y
Problem 5	2	2	1	6	−1.1870483D+02	0.00D+00	2.96D−05	Y
			2	7	−1.1870491D+02	1.54D−05	−5.20D−05	Y
			3	7	−1.1870493D+02	2.05D−05	−7.08D−05	Y
			4	5	−1.1870497D+02	3.32D−05	−1.15D−04	Y
			5	4	−1.1870480D+02	0.00D+00	6.35D−05	Y
Problem 6	2	2	1	3	7.4178118D−01	3.12D−06	9.22D−06	Y
			2	3	7.4178196D−01	8.76D−12	1.00D−05	Y
			3	3	7.4178118D−01	3.08D−06	9.22D−06	Y
			4	4	7.4178442D−01	0.00D+00	3.47D−06	Y
			5	3	7.4177556D−01	2.52D−05	3.60D−06	Y
Problem 7	2	1	1	3	−5.0000006D−01	2.47D−07	9.94D−06	Y
			2	3	−5.0000000D−01	0.00D+00	1.00D−05	Y
			3	3	−5.0000006D−01	2.45D−07	9.94D−06	Y
			4	3	−5.0000198D−01	7.94D−06	8.02D−06	Y
			5	3	−5.0000028D−01	1.13D−06	9.72D−06	Y

Table 1 shows the performance of Algorithm 2 considering the five types of augmented Lagrangian functions. In the table, the first three columns identify the problem and the number of variables and constraints. "ALtype" identifies the considered augmented Lagrangian function, "Iter" is the number of augmented Lagrangian iterations, $f(x^*)$ is the value of the objective function at the final iterate x^*, $\|g(x^*)_+\|_\infty$ is the infeasibility measurement at x^*, $\gamma_k + \varepsilon_k$ is the computed value at Steps 2 and 3 in the last iteration and "Global" iden-

tifies whether the final iterate is a global solution, in the sense that $f(x^*) \leq f(z) + \varepsilon_{opt}$ for all feasible z, or not.

In all instances, Algorithm 2 stopped at Step 3 in a finite number of iterations reporting *solution found*. In fact, note that the values of $\gamma_k + \varepsilon_k$ in Table 1 are equal to or less than $\varepsilon_{opt} = 10^{-4}$. The last column reports that all problems have been successfully solved by Algorithm 2 for each augmented Lagrangian type. Analyzing the column "Iter" of Table 1, we may note that, for these problems, the algorithm has a similar performance for all augmented Lagrangian functions considered. Moreover, these results allow us to conclude that the *multistart* strategy to solve the subproblems and the relation (39) are appropriate.

4.3 Kissing problems

In this section, our aim is to illustrate the behavior of our algorithm to detect if a problem is infeasible. For this, we analyze our infeasible test for some infeasible problems. In order to be self-contained, we also consider some feasible problems.

Given a set of $nSph$ N-dimensional unit spheres, we consider the problem of finding the coordinates of the sphere centers in such a way that each one of the spheres touches another given one, with no overlap and maximizing the sum of the squared distances between the sphere centers.

Let us denote by $C_i \in \mathbb{R}^N$ the coordinate vector of the N-dimensional sphere center for all $i = 1, \ldots, nSph$. This problem can be modeled as follows:

$$
\begin{aligned}
\text{Maximize } & \sum_{i<j}^{N} \left\| C_i - C_j \right\|^2 \\
\text{subject to } & \|C_i\|^2 = 1, \qquad i = 1, \ldots, nSph, \\
& \left\| C_i - C_j \right\|^2 \geq 1, \ \forall i < j, \\
& -1 \leq C_i \leq 1, \qquad i = 1, \ldots, nSph.
\end{aligned}
\tag{41}
$$

The first set of constraints requires that each considered sphere touches another one centered at the origin, while the second set determines that the spheres must be placed without overlapping. Simple devices can be used to transform problem (41) to the form (1). For example, we may rewrite each equality constraint as two inequality constraints. After making the necessary modifications, problem (41) has $n = N \times nSph$ variables, $2 \times nSph$ constraints arising from the first set and another C_2^{nSph} arising from the second one, totaling $m = 2 \times nSph + C_2^{nSph}$ constraints beyond the box-constraints.

The *kissing number* is defined as the maximum number of non-overlapping N-dimensional unit spheres that can be arranged such that each one touches another given N-dimensional unit sphere. When $N = 2$ the kissing number is equal 6, while for $N = 3$ and $N = 4$ is equal 12 and 24, respectively, [20,27]. Therefore, for $N = 2$, problem (41) is feasible if $1 \leq nSph \leq 6$ and infeasible if $nSph > 6$. Analogously, feasible instances of problem (41) are obtained when $N = 3$ and $1 \leq nSph \leq 12$ or when $N = 4$ and $1 \leq nSph \leq 24$. Infeasible instances may be generated with $N = 3$ and $nSph > 12$ or with $N = 4$ and $nSph > 24$.

For illustrative purposes, we considered six instances of problem (41) with $N = 2, nSph = 3, 6, 7, 10, 12$ and 15. The first two instances are feasible while the remaining ones are infeasible. The problems were solved with different types of augmented Lagrangians, specified in the Fig. 1. For all instances Algorithm 2 behaves according to theory. For the feasible instances Algorithm 2 stopped at Step 3 reporting *solution found* while for infeasible ones stopped at Step 2 declaring *infeasibility*. Figure 1 illustrates the "solutions".
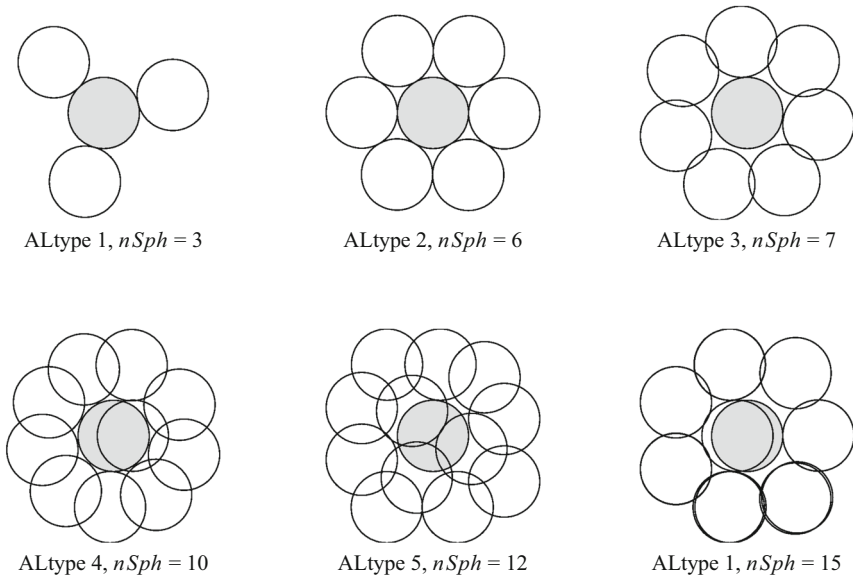
Fig. 1 Graphical representation of six instances of the kissing problem with $N = 2$ and $nSph = 3, 6, 7, 10, 12$ and 15

For the infeasible instances illustrated in Fig. 1, Algorithm 2 stopped at outer iteration 4, 3, 3 and 3, respectively. Despite that Algorithm 2 stopped with few iterations, the nice arrangement of the circles corroborates the result of Proposition 2. In some sense, Proposition 2 establishes that the limit points of the sequence generated by Algorithm 2 are global minimizers of an infeasibility measure. See Corollary 3 for ALtype 1 and 2, Remark 2 for ALtype 5, Corollary 4 for ALtype 4 and Remark 3 for ALtype 3.

In order to verify the applicability of our approach, we run Algorithm 2 for feasible and infeasible instances of problem (41) considering the five types of augmented Lagrangians. Table 2 shows the performance of Algorithm 2 with the five augmented Lagrangian functions for three feasible instances of the problem (41). These problems correspond to dimension $N = 2$, 3 and 4 with $nSph$ equal to their respective *kissing number*. Therefore, it is reasonable to believe that these are the most difficult feasible problems for each considered dimension. In the table, the first four columns identify the dimension N, the number of N-dimensional spheres $nSph$ and the number of variables and constraints. "ALtype" identifies the considered augmented Lagrangian function, "Iter" is the number of augmented Lagrangian iterations, $f(x^*)$ is the value of the objective function at the final iterate x^*, $\|g(x^*)_+\|_\infty$ is the infeasibility measurement at x^*, $\gamma_k + \varepsilon_k$ and $-c_k$ are the computed values at Steps 2 and 3 in the last iteration. In all the three feasible problems for all augmented Lagrangian types, Algorithm 2 stopped at Step 3 reporting *solution found*. In fact, note that the values of $\gamma_k + \varepsilon_k$ in Table 2 is less than $\varepsilon_{opt} = 10^{-4}$.

Table 3 shows the performance of Algorithm 2 with the five augmented Lagrangian functions for six infeasible instances of the problem (41). Three out of the six considered instances may be the most difficult infeasible problems for each considered dimension. These problems correspond to dimension $N = 2$, 3 and 4 with $nSph$ equal to their respective *kissing number* plus one. Table 3 contains the same information as Table 2. In all infeasible instances, Algorithm 2 stopped at Step 2 reporting *infeasibility*. In fact, comparing the last two columns of

Table 2 Performance of Algorithm 2 with the five augmented Lagrangian functions for three feasible instances of the kissing problem

N	nSph	n	m	ALtype	Iter	$f(x^*)$	$\|g(x^*)_+\|_\infty$	$\gamma_k + \varepsilon_k$	$-c_k$
2	6	12	27	1	3	$-3.6000000\text{D}+01$	$1.95\text{D}-08$	$9.84\text{D}-06$	$-8.40\text{D}+01$
				2	4	$-3.6000000\text{D}+01$	$2.32\text{D}-12$	$1.00\text{D}-06$	$-8.40\text{D}+01$
				3	3	$-3.6000000\text{D}+01$	$1.65\text{D}-08$	$9.83\text{D}-06$	$-8.40\text{D}+01$
				4	3	$-3.6000000\text{D}+01$	$5.95\text{D}-06$	$1.01\text{D}-05$	$-8.40\text{D}+01$
				5	4	$-3.5999999\text{D}+01$	$1.70\text{D}-05$	$1.94\text{D}-06$	$-8.40\text{D}+01$
3	12	36	90	1	3	$-1.4400000\text{D}+02$	$8.43\text{D}-08$	$9.56\text{D}-06$	$-6.48\text{D}+02$
				2	5	$-1.4400000\text{D}+02$	$5.01\text{D}-12$	$1.00\text{D}-07$	$-6.48\text{D}+02$
				3	3	$-1.4400000\text{D}+02$	$6.20\text{D}-08$	$9.25\text{D}-06$	$-6.48\text{D}+02$
				4	3	$-1.4400000\text{D}+02$	$7.30\text{D}-05$	$1.26\text{D}-05$	$-6.48\text{D}+02$
				5	4	$-1.4400000\text{D}+02$	$7.14\text{D}-05$	$9.97\text{D}-06$	$-6.48\text{D}+02$
4	24	96	324	1	3	$-5.7600005\text{D}+02$	$1.76\text{D}-07$	$-3.93\text{D}-05$	$-3.84\text{D}+03$
				2	10	$-5.7600000\text{D}+02$	$4.68\text{D}-13$	$1.00\text{D}-08$	$-3.84\text{D}+03$
				3	7	$-5.7600000\text{D}+02$	$9.16\text{D}-06$	$1.21\text{D}-06$	$-3.84\text{D}+03$
				4	6	$-5.7600000\text{D}+02$	$6.94\text{D}-06$	$-1.00\text{D}-06$	$-3.84\text{D}+03$
				5	5	$-5.7600000\text{D}+02$	$2.80\text{D}-05$	$1.35\text{D}-06$	$-3.84\text{D}+03$

Table 3, $\gamma_k + \varepsilon_k < -c_k$ in all cases. The results in Tables 2 and 3 show that the performance of Algorithm 2 is in agreement with the theoretical part.

4.4 Comments on the numerical experiments

In order to compare the performance of the five type of Lagrangian functions, we may consider the amount of augmented Lagrangian iterations (columns "Iter") of the problems presented in Sects. 4.2 and 4.3. For the ten feasible problems of our numerical experiments, the numbers of augmented Lagrangian iterations are 32, 46, 37, 37 and 36 for ALtype 1, 2, 3, 4 and 5, respectively. For the six infeasible problems, the numbers of augmented Lagrangian iterations are 38, 62, 25, 24 and 23 for ALtype 1, 2, 3, 4 and 5, respectively.

Based on our numerical experiments, we may conclude that Algorithm 2 with ALtype 2 presents the worst performance for both feasible and infeasible instances. It is interesting to point out that augmented Lagrangians for ALtype 3, 4 and 5 demonstrated to be as efficient as the PHR augmented Lagrangian (ALtype 1) for feasible instances. Moreover, we observe that Algorithm 2 with these Lagrangians detected infeasibility in less iterations than PHR, which may reflect in computational savings. These motivate to continue the research on general augmented Lagrangian methods for nonlinear programming under possible infeasibility.

5 Conclusions

In this article, we considered a nonlinear programming problem for which the constraint set may be infeasible. We studied the augmented Lagrangian method based on the penalty functions of Kort and Bertsekas. A general algorithm was stated and its convergence properties were analyzed. For some specific penalty functions, we showed that every limit point of the sequence generated by the algorithm is a global minimizer of an infeasibility measure. Some

Table 3 Performance of Algorithm 2 with the five augmented Lagrangian functions for six infeasible instances of the kissing problem

Dim	nSph	n	m	ALtype	Iter	$f(x^*)$	$\|g(x^*)_+\|_\infty$	$\gamma_k + \varepsilon_k$	$-c_k$
2	7	14	35	1	4	−5.3627698D+01	1.83D−01	−1.57D+02	−1.14D+02
				2	8	−4.8000008D+01	1.00D+00	−1.59D+02	−1.20D+02
				3	3	−5.4366890D+01	1.65D−01	−5.76D+05	−1.14D+02
				4	3	−5.0569965D+01	2.25D−01	−2.80D+02	−1.17D+02
				5	3	−5.3257638D+01	1.86D−01	−5.95D+02	−1.15D+02
3	13	39	104	1	8	−1.7798952D+02	7.22D−02	−4.09D+03	−7.58D+02
				2	11	−1.7364253D+02	1.80D−01	−9.28D+02	−7.62D+02
				3	5	−1.7655173D+02	4.50D−02	−3.90D+17	−7.59D+02
				4	5	−1.5466360D+02	7.07D−01	−1.00D+20	−7.81D+02
				5	5	−1.7537088D+02	6.35D−02	−8.13D+03	−7.61D+02
3	20	60	230	1	5	−4.9839222D+02	3.10D−01	−1.55D+04	−1.78D+03
				2	9	−4.3008527D+02	1.00D+00	−2.38D+03	−1.85D+03
				3	3	−4.8605914D+02	2.18D−01	−6.19D+08	−1.79D+03
				4	3	−4.1554159D+02	4.19D−01	−5.81D+03	−1.86D+03
				5	3	−4.6204436D+02	2.93D−01	−1.13D+04	−1.82D+03
4	25	100	350	1	8	−6.5410561D+02	9.60D−02	−5.64D+03	−4.15D+03
				2	13	−6.5071172D+02	3.84D−01	−7.83D+03	−4.15D+03
				3	5	−6.4911245D+02	3.97D−02	−2.67D+15	−4.15D+03
				4	5	−6.2793040D+02	1.30D−01	−1.22D+04	−4.17D+03
				5	5	−6.4328594D+02	6.38D−02	−1.10D+04	−4.16D+03
4	30	120	495	1	8	−1.0099839D+03	1.86D−01	−4.26D+04	−5.95D+03
				2	11	−9.8167065D+02	1.00D+00	−7.59D+03	−5.98D+03
				3	6	−8.8718074D+02	1.00D+00	−1.70D+21	−6.07D+03
				4	5	−8.9740635D+02	7.10D−01	−4.00D+20	−6.06D+03
				5	4	−9.5306975D+02	1.55D−01	−2.49D+05	−6.01D+03
4	40	160	860	1	5	−1.9918823D+03	3.33D−01	−2.40D+04	−1.05D+04
				2	10	−1.8693087D+03	1.00D+00	−1.09D+04	−1.06D+04
				3	3	−1.8866072D+03	1.86D−01	−3.22D+07	−1.06D+04
				4	3	−1.6579181D+03	3.94D−01	−1.13D+04	−1.08D+04
				5	3	−1.8143639D+03	2.59D−01	−1.95D+04	−1.07D+04

results regarding optimality are also presented. Based on these results, a second algorithm is proposed, which stops after a finite number of iterations at an approximate solution or detects that the problem is infeasible. We also considered some numerical experiments illustrating the applicability of the algorithm. It is interesting to point out that some other augmented Lagrangians presented similar computational performance to the classical quadratic augmented Lagrangian (PHR). Besides, for infeasible problems, our algorithm based on these other Lagrangians detected infeasibility in less iterations than PHR. These numerical results motivate to continue the studies and analysis of general Lagrangian methods.

References

1. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**(4), 1286–1309 (2008)
2. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York (1982)
3. 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)
4. Birgin, E.G., Martínez, J. M., Prudente, L.F.: Optimality properties of an augmented Lagrangian method on infeasible problems. Comput. Optim. Appl. doi:10.1007/s10589-014-9685-5
5. 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**(2), 207–242 (2014)
6. Burachik, R.S., Iusem, A.N., Melo, J.G.: A primal dual modified subgradient algorithm with sharp Lagrangian. J. Glob. Optim. **46**(3), 347–361 (2010)
7. Byrd, R., Curtis, F., Nocedal, J.: Infeasibility detection and SQP methods for nonlinear optimization. SIAM J. Optim. **20**(5), 2281–2299 (2010)
8. Conn, A.R., Gould, N.I.M., Toint, PhL: Lancelot: A FORTRAN Package for Large-Scale Nonlinear Optimization (Release A). Springer, Secaucus (1992)
9. Floudas, C.A.: Deterministic Global Optimization: Theory, Methods and Application. Kluwer Academic Publishers, Dordrecht (1999)
10. Floudas, C.A., Visweswaran, V.: A global optimization algorithm (GOP) for certain classes of nonconvex NLPS II. Application of theory and test problems. Comput. Chem. Eng. **14**, 1419–1434 (1990)
11. Griva, I., Polyak, R.: Primal-dual nonlinear rescaling method with dynamic scaling parameter update. Math. Program. **106**(2), 237–259 (2006)
12. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**(5), 303–320 (1969)
13. Kearfott, R.B., Dawande, M., Du, K., Hu, C.: Algorithm 737: Intlib & mdash: a portable Fortran 77 interval standard-function library. ACM Trans. Math. Softw. **20**(4), 447–459 (1994)
14. Kort, B., Bertsekas, D.P.: Combined primal-dual and penalty methods for convex programming. SIAM J. Control Optim. **14**(2), 268–294 (1976)
15. Liebman, J., Lasdon, L., Schrage, L., Waren, A.: Modeling and Optimization with GINO. The Scientic Press, Palo Alto (1986)
16. Luo, H.Z., Sun, X.L., Li, D.: On the convergence of augmented Lagrangian methods for constrained global optimization. SIAM J. Optim. **18**(4), 1209–1230 (2007)
17. Mangasarian, O.: Unconstrained Lagrangians in nonlinear programming. SIAM J. Control Optim. **13**(4), 772–791 (1975)
18. Manousiouthakis, M., Sourlas, D.: A global optimization approach to rationally constrained rational programming. Chem. Eng. Comm. **115**(1), 127–147 (1992)
19. Martínez, J.M., Prudente, L.F.: Handling infeasibility in a large-scale nonlinear optimization algorithm. Numer. Algorithms **60**(2), 263–277 (2012)
20. Musin, O.R.: The kissing number in four dimensions. Ann. Math. **168**, 1–32 (2008)
21. Polyak, R.: Modified barrier functions (theory and methods). Math. Program. **54**(1–3), 177–222 (1992)
22. Polyak, R., Griva, I.: Primal-dual nonlinear rescaling method for convex optimization. J. Optim. Theory Appl. **122**(1), 111–156 (2004)
23. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)
24. Rockafellar, R.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM J. Control Optim. **12**(2), 268–285 (1974)
25. Rubinov, A.M., Yang, X.Q.: Lagrange-Type Functions in Constrained Non-convex Optimization. Kluwer Academic Publishers, Dordrecht (2003)
26. Sahinidis, N.V., Grossmann, I.E.: Reformulation of the multiperiod milp model for capacity expansion of chemical processes. Oper. Res. **40**(1), S127–S144 (1992)
27. Schtte, K., Waerden, B.L.: Das problem der dreizehn Kugeln. Math. Ann. **125**(1), 325–334 (1952)
28. Swaney, R.E.: Global Solution of Algebraic Nonlinear Programs. AIChE Annual Meeting, Chicago (1990)
29. Wang, C.Y., Li, D.: Unified theory of augmented Lagrangian methods for constrained global optimization. J. Glob. Optim. **44**(3), 433–458 (2009)
30. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. **106**(1), 25–57 (2006)