Comput Optim Appl (2010) 46: 229–245
DOI 10.1007/s10589-009-9271-4

Addressing the greediness phenomenon in Nonlinear Programming by means of Proximal Augmented Lagrangians

Emerson V. Castelani · André L.M. Martinez · J.M. Martínez · B.F. Svaiter

Received: 3 November 2008 / Published online: 24 July 2009

Abstract When one solves Nonlinear Programming problems by means of algorithms that use merit criteria combining the objective function and penalty feasibility terms, a phenomenon called *greediness* may occur. Unconstrained minimizers attract the iterates at early stages of the calculations and, so, the penalty parameter needs to grow excessively, in such a way that ill-conditioning harms the overall convergence. In this paper a regularization approach is suggested to overcome this difficulty. An Augmented Lagrangian method is defined with the addition of a regularization term that inhibits the possibility that the iterates go far from a reference point. Convergence proofs and numerical examples are given.

Keywords Nonlinear programming · Greediness · Augmented Lagrangian method · Regularization

E.V. Castelani was supported by PRONEX-Optimization (PRONEX—CNPq/FAPERJ E-26/171.164/2003—APQ1), FAPESP (Grants 06/53768-0 and 05-56773-1) and CNPq.

A.L.M. Martinez was supported by PRONEX-Optimization (PRONEX—CNPq/FAPERJ E-26/171.164/2003—APQ1), FAPESP (Grants 06/53768-0 and 05-56773-1) and CAPES.

J.M. Martínez was supported by PRONEX-Optimization (PRONEX—CNPq/FAPERJ E-26/171.164/2003—APQ1), FAPESP (Grant 06/53768-0) and CNPq.

E.V. Castelani (✉) · A.L.M. Martinez · J.M. Martínez
Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas, SP, Brazil
e-mail: emersonvitor@gmail.com

A.L.M. Martinez
e-mail: andrelmmartinez@yahoo.com.br

J.M. Martínez
e-mail: martinez@ime.unicamp.br

B.F. Svaiter
Institute of Pure and Applied Mathematics (IMPA), Rio de Janeiro, Brazil

1 Introduction

When Penalty-Lagrangian methods are used for solving constrained minimization problems, outer iterations consist of minimizing a merit function that combines the objective function of the problem and a penalty term that inhibits infeasibility.

The penalty parameter is set to be small at the first outer iterations, because large penalty parameters tend to produce ill-conditioned subproblems [4–6, 9]. As a consequence, the first iterations tend to privilege optimality over feasibility and, so, very infeasible points may be computed at the beginning, even when the initial approximation is feasible or nearly feasible [3]. This phenomenon is called *greediness* in the present paper. Unfortunately, greediness may take place even when a feasible initial point is available.

Here we consider the greediness phenomenon associated with the application of the Powell-Hestenes-Rockafellar (PHR) Augmented Lagrangian method to nonconvex problems [4, 7, 10, 13]. We employ the approach of [2], by means of which feasible limit points satisfy the KKT conditions under the CPLD (Constant Positive Linear Dependence) constraint qualification [1, 11]. The theoretical results of [2] show that the method always converges to stationary points of the sum-of-squares measure of infeasibility, which, of course, includes feasible points, but infeasible limits are possible, being, in general, local minimizers of the infeasibility. This limitation is unavoidable because, ultimately, the problem may have no feasible points. However, it is reasonable to provide the algorithms with practical procedures that avoid convergence to local-nonglobal infeasibility minimizers, at least when feasible points exist.

In 1973, dealing with the convex programming problem with inequality constraints, Rockafellar [12] introduced the Proximal Augmented Lagrangian method. At each iteration of this method an unconstrained minimizer of the augmented Lagrangian with a proximal point penalization is computed. In the present paper we employ a similar regularization procedure to inhibit greediness in nonconvex problems. Augmented Lagrangian regularizations were used with different purposes in the convex programming literature. See, for example, [8]. The objective function used in the subproblems of Algencan [2] is modified here in such a way that the distance with respect a reference point, updated at every iteration, is penalized. We will show that the convergence properties of [2] are preserved and that, in practice, the modification does not harm the behavior of the algorithm when greediness is not present. Finally, we show that the performance of the algorithm is improved in situations where greediness is observed when applying Algencan.[1]

This paper is organized as follows. Two versions of the algorithm are presented in Sect. 2. Convergence results are proved in Sect. 3. Numerical experiments are given in Sect. 4. Section 5 contains final remarks and lines for future research.

Notation The symbol $\|\cdot\|$ denotes the Euclidean norm throughout the paper, although many times it can be replaced by an arbitrary norm.

We denote $K_1 \underset{\infty}{\subset} K_2$ to indicate that K_1 is an infinite subsequence of indices contained in K_2.

[1] Algencan is available in www.ime.usp.br/~egbirgin/tango.

2 Algorithms

The problem considered from now on is:

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & h(x) = 0, \quad g(x) \le 0, \quad x \in \Omega. \end{array} \tag{1}$$

The set Ω will be given by *lower-level constraints* of the form

$$\underline{h}(x) = 0, \quad \underline{g}(x) \le 0.$$

In the most simple cases, Ω will take the form of an n-dimensional box:

$$\Omega = \{x \in \mathbb{R}^n \mid \ell \le x \le u\}.$$

We will assume that the functions $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p, \underline{h} : \mathbb{R}^n \to \mathbb{R}^{\underline{m}}, \underline{g} : \mathbb{R}^n \to \mathbb{R}^{\underline{p}}$ have continuous first derivatives on $\mathbb{R}^n$.

Given $\rho > 0$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p_+$, $x \in \mathbb{R}^n$ the usual PHR- Augmented Lagrangian $L_\rho(x, \lambda, \mu)$ is given by:

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

At each (outer) iteration, Algencan [2] minimizes (approximately) $L_\rho(x, \lambda, \mu)$ subject to $x \in \Omega$.

The Regularized Augmented Lagrangian employed in this paper uses two additional parameters that are updated at every outer iteration: The reference point $\bar{x}$ and the regularization parameter γ. Therefore, we define:

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

At each iteration of the algorithms an approximate solution of the subproblem

$$\begin{array}{ll} \text{Minimize} & L_{\rho_k,\gamma_k,\bar{x}^k}(x, \bar{\lambda}^k, \bar{\mu}^k) \\ \text{s. t.} & \underline{h}(x) = 0, \quad \underline{g}(x) \le 0 \end{array} \tag{2}$$

will be computed. The objective function of (2) has continuous first derivatives although its second derivatives are generally discontinuous.

Let us define now the main algorithms presented in this paper.

Algorithm 2.1 The parameters that define the algorithm are: $\tau \in [0, 1)$, $\eta > 1, \lambda_{min} < \lambda_{max}, \mu_{max} > 0, \beta > 0$. At the first outer iteration we use a penalty parameter $\rho_1 > 0$ and safeguarded Lagrange multipliers estimates $\bar{\lambda}^1 \in \mathbb{R}^m$ and $\bar{\mu}^1 \in \mathbb{R}^p$ such that

$$\bar{\lambda}_i^1 \in [\lambda_{min}, \lambda_{max}] \quad \forall i = 1, \ldots, m \quad \text{and} \quad \mu_i^1 \in [0, \mu_{max}] \quad \forall i = 1, \ldots, p.$$

The initial regularization parameter is $\gamma_1 \geq 0$. We also assume that $x^0 \in \mathbb{R}^n$ is an arbitrary initial point that coincides with the initial reference point $\bar{x}^0$.

Finally, $\{\varepsilon_k\}$ is a sequence of positive numbers that satisfies

$$\lim_{k \to \infty} \varepsilon_k = 0.$$

Step 1. *Initialization.*
Set $k \leftarrow 1$.

Step 2. *Solve the subproblem.*
Compute $x^k \in \mathbb{R}^n$ such that there exist $v^k \in \mathbb{R}^{\underline{m}}$, $w^k \in \mathbb{R}^{\underline{p}}$ satisfying

$$\left\| \nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) + \gamma_k(x^k - \bar{x}^{k-1}) + \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^{\underline{p}} w_i^k \nabla \underline{g}_i(x^k) \right\| \leq \varepsilon_k, \tag{3}$$

$$w^k \geq 0, \qquad \underline{g}(x^k) \leq \varepsilon_k, \tag{4}$$

$$\underline{g}_i(x^k) < -\varepsilon_k \quad \Rightarrow \quad w_i^k = 0 \quad \text{for all } i = 1, \ldots, \underline{p}, \tag{5}$$

$$\|\underline{h}(x^k)\| \leq \varepsilon_k. \tag{6}$$

Step 3. *Estimate multipliers.*
For all $i = 1, \ldots, m$, compute

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k). \tag{7}$$

For all $i = 1, \ldots, p$, compute

$$\mu_i^{k+1} = \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \tag{8}$$

and

$$V_i^k = \max\left\{ g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k} \right\}.$$

Step 4. *Update penalty parameter.*
Define

$$R_k = \max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\}. \tag{9}$$

If $k > 1$ and

$$R_k > \tau R_{k-1},$$

define

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

Else, define

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

Step 5. *Update reference point and regularization parameter.*
If

$$R_k = \min\{R_0, \dots, R_k\} \tag{10}$$

define $\bar{x}^k = x^k$. Else, define $\bar{x}^k = \bar{x}^{k-1}$.
For all $i = 1, m$, compute:

$$\bar{\lambda}_i^{k+1} \in [\lambda_{min}, \lambda_{max}]. \tag{11}$$

For all $i = 1, \dots, p$, compute:

$$\bar{\mu}_i^{k+1} \in [0, \mu_{max}]. \tag{12}$$

Choose $\gamma_{k+1} \geq 0$ in such a way that

$$\gamma_{k+1} \leq \min\{\gamma_k, \beta R_k\}. \tag{13}$$

Remarks

1. The conditions (3–6) say that x^k is an approximate KKT point of the subproblem (2). Therefore, at each outer iteration we aim to minimize (approximately) the regularized augmented Lagrangian subject to the lower level constraints defined by the set Ω. Since the objective function of (2) is differentiable, first-order algorithms may be used for this purpose.
2. Steps 3 and 4 of Algorithm 2.1 are as in the main algorithm of [2]. At Step 5 the reference point and the regularization parameter are updated. The idea is that the reference point should be the best previous iterate in terms of the feasibility-complementarity measure R_k. In other words, the regularization does not allow the next iteration to be very far from the most feasible point already computed. On the other hand, the condition (13) imposes that the regularization parameter must tend to zero when R_k goes to zero. Thus, the effect of regularization tends to disappear when the algorithm is going well in terms of getting feasible points.
3. If (10) takes place the safeguarded multipliers $\bar{\lambda}$ and $\bar{\mu}$ in (11) and (12) will be computed as follows:

$$\bar{\lambda}_i^{k+1} = \max\{\lambda_{min}, \min\{\lambda_{max}, \lambda_i^{k+1}\}\},$$

for $i = 1, \dots, m$ and

$$\bar{\mu}_i^{k+1} = \min\{\mu_i^{k+1}, \mu_{max}\}$$

for $i = 1, \dots, p$.
When (10) does not hold we define $\bar{\lambda}^{k+1} = \bar{\lambda}^k$ and $\bar{\mu}^{k+1} = \bar{\mu}^k$.

4. If the initial point is feasible and the initial inequality multipliers are chosen in such a way that complementarity takes place, we have $R_0 = 0$ and the condition (10) may be too restrictive for updating the reference point. In this case it is convenient to replace (10) by

$$R_k = \min\{R_{tol}, R_1, \dots, R_k\}, \tag{14}$$

where $R_{tol} > 0$ is a given initial tolerance. The use of (14) instead of (10) does not affect the convergence theory.

In the proof of the global convergence of Algorithm 2.1 we will use the fact that, if $K \underset{\infty}{\subset} \mathbb{N}$,

$$\lim_{k\in K} R_k = 0 \quad \Rightarrow \quad \lim_{k\to\infty} \gamma_k = 0.$$

This property is true because $\lim_{k\in K} R_k = 0$ implies, by (13), that $\lim_{k\in K} \gamma_{k+1} = 0$. So, since $\gamma_{k+1} \le \gamma_k$ for all k, it follows that $\lim_{k\to\infty} \gamma_k = 0$. This means that the requirement

$$\gamma_{k+1} \le \gamma_k \quad \forall k, \tag{15}$$

imposed by (13) seems to be necessary for proving the convergence theorem. However, the condition (15) contradicts common sense because, perhaps, one would like to increase the regularization parameter if feasibility deteriorates. In other words, reasonable regularization strategies could require only that $\gamma_{k+1} \le \beta R_k$, but not that $\gamma_{k+1} \le \gamma_k$.

We will see that, with a suitable modification of Algorithm 2.1, monotonicity conditions on γ_k may be eliminated.

Algorithm 2.2 This algorithm coincides with Algorithm 2.1 except that:

- For the solution of the subproblems we employ an iterative minimization algorithm that uses $\bar{x}^{k-1}$ as initial point and guarantees that

$$L_{\rho_k,\gamma_k,\bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \le L_{\rho_k,\gamma_k,\bar{x}^{k-1}}(\bar{x}^{k-1}, \bar{\lambda}^k, \bar{\mu}^k). \tag{16}$$

- The condition (13) is replaced by

$$\gamma_{k+1} \le \beta R_k. \tag{17}$$

The condition (16) is quite natural if one uses a monotone minimization algorithm that preserves feasibility of the lower level set for solving the subproblems. In most cases, the lower level set is simple enough to make it possible the employment of such an algorithm.

The crucial result that allows one to relax (17) is the following.

Lemma 2.3 *Let $\{x^k\}$ be a bounded sequence generated by Algorithm 2.2 and suppose that there exists an infinite set of indices K such that*

$$\lim_{k\in K} R_k = 0.$$

Then,

$$\lim_{k\to\infty} R_k = 0. \tag{18}$$

Proof If $\{\rho_k\}$ is bounded we have that $R_k \le \tau R_{k-1}$ for all k large enough, so the thesis is proved.

Let us assume, from now on, that $\lim_{k\to\infty} \rho_k = \infty$.

By (16), we have that, for all $k \in \mathbb{N}$,

$$f(x^k) + \frac{\rho_k}{2}\left\{\sum_{i=1}^{m}\left[h_i(x_k) + \frac{\bar{\lambda}_i^k}{\rho_k}\right]^2 + \sum_{i=1}^{p}\left[\max\left(0, g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k}\right)\right]^2\right\}$$
$$+ \frac{\gamma_k}{2}\|x^k - \bar{x}^{k-1}\|^2$$
$$\le f(\bar{x}^{k-1}) + \frac{\rho_k}{2}\left\{\sum_{i=1}^{m}\left[h_i(\bar{x}_{k-1}) + \frac{\bar{\lambda}_i^k}{\rho_k}\right]^2 + \sum_{i=1}^{p}\left[\max\left(0, g_i(\bar{x}^{k-1}) + \frac{\bar{\mu}_i^k}{\rho_k}\right)\right]^2\right\}.$$

Dividing by ρ_k we get:

$$\frac{1}{\rho_k} f(x^k) + \frac{1}{2}\left\{\sum_{i=1}^{m}\left[h_i(x_k) + \frac{\bar{\lambda}_i^k}{\rho_k}\right]^2 + \sum_{i=1}^{p}\left[\max\left(0, g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k}\right)\right]^2\right\}$$
$$+ \frac{\gamma_k}{2\rho_k}\|x^k - \bar{x}^{k-1}\|^2$$
$$\le \frac{1}{\rho_k} f(\bar{x}^{k-1}) + \frac{1}{2}\left\{\sum_{i=1}^{m}\left[h_i(\bar{x}^{k-1}) + \frac{\bar{\lambda}_i^k}{\rho_k}\right]^2 + \sum_{i=1}^{p}\left[\max\left(0, g_i(\bar{x}^{k-1}) + \frac{\bar{\mu}_i^k}{\rho_k}\right)\right]^2\right\}. \tag{19}$$

By the definition of $\bar{x}^{k-1}$ we can write:

$$\{\bar{x}^0, \bar{x}^1, \bar{x}^2, \ldots\} = \{x^{k_0}, x^{k_1}, x^{k_2}, \ldots\}$$

where $k_0 \le k_1 \le k_2 \le \cdots$. Moreover, since $\lim_{k\in K} R_k = 0$, we have that

$$\lim_{j\to\infty} R_{k_j} = 0. \tag{20}$$

Clearly, (20) implies that

$$\lim_{j\to\infty} \|h(x^{k_j})\|^2 + \sum_{i=1}^{p} \max\{0, g(x^{k_j})\}^2 = 0.$$

Therefore,

$$\lim_{k\to\infty} \|h(\bar{x}^{k-1})\|^2 + \sum_{i=1}^{p} \max\{0, g(\bar{x}^{k-1})\}^2 = 0.$$

Therefore, the right-hand side of (19) tends to zero when k tends to infinity.

Thus,

$$\lim_{k\to\infty} \frac{1}{2}\left\{\sum_{i=1}^{m}\left[h_i(x_k)+\frac{\bar{\lambda}_i^k}{\rho_k}\right]^2+\sum_{i=1}^{p}\left[\max\left(0, g_i(x^k)+\frac{\bar{\mu}_i^k}{\rho_k}\right)\right]^2\right\}=0.$$

Since $\rho_k \to \infty$ and $\bar{\mu}^k$, $\bar{\lambda}^k$ are bounded, this implies that

$$\lim_{k\to\infty} \|h(x^k)\| = 0 \tag{21}$$

and

$$\lim_{k\to\infty} \max\{0, g_i(x^k)\} = 0 \quad \forall i = 1, \ldots, p.$$

Then,

$$\lim_{k\to\infty} V_i^k = 0 \quad \forall i = 1, \ldots, p. \tag{22}$$

By (9), (21) and (22) we obtain (18). □

Lemma 2.4 *Let* $\{x^k\}$ *be a bounded sequence generated by Algorithm 2.2 and suppose that there exists an infinite set of indices* K *such that*

$$\lim_{k\in K} R_k = 0.$$

Then,

$$\lim_{k\to\infty} \gamma_k = 0.$$

Proof The desired result follows from (17) and Lemma 2.3. □

3 Convergence

In this section we address the global convergence of Algorithms 2.1 and 2.2. Essentially, we will show that the addition of the regularization term does not affect the properties proved in [2] for the standard version of Algencan.

The main global convergence results are given below. Lemma 3.1 shows that, as a result of solving (3–6) one obtains an approximate KKT point of the problem of minimizing a regularized Lagrangian subject to $x \in \Omega$.

Lemma 3.1 *Assume that* $\{x^k\}$ *is a sequence generated by Algorithm 2.1 or Algorithm 2.2. Then, for all* $k = 1, 2, \ldots$ *we have*:

$$\begin{aligned}&\|\nabla f(x^k)+\nabla h(x^k)\lambda^{k+1}+\nabla g(x^k)\mu^{k+1}+\gamma_k(x^k-\bar{x}^{k-1})\\&\quad+\nabla \underline{h}(x^k)v^k+\nabla \underline{g}(x^k)w^k\| \le \varepsilon_k,\end{aligned}$$

where

$$w^k \geq 0, \qquad w_i^k = 0 \quad \textit{whenever } \underline{g}_i(x^k) < -\varepsilon_k,$$

$$\underline{g}_i(x^k) \leq \varepsilon_k \quad \forall i = 1, \dots, \underline{p}, \qquad \|\underline{h}(x^k)\| \leq \varepsilon_k.$$

Proof The proof follows from (3–6) using the definitions (7) and (8). □

Lemma 3.2 below says that, if $\{x^k\}_{k\in K}$ is a convergent subsequence and an inequality constraint is strictly satisfied at the limit point, then the corresponding estimated Lagrange multiplier is necessarily null if k is large enough. This property holds even if the limit point is infeasible. This result corresponds, in Algencan, to formula (4.10) of [2, Theorem 4.2]. It is interesting to show that it also holds for the regularized method. The result has an independent interest in terms of stopping criteria. It shows that, essentially, the algorithm takes care of complementarity and that the algorithm designer needs only to look at feasibility and optimality for deciding if an iterate is good enough. Moreover, since optimality is guaranteed by the solution of the subproblems, we may think that the iterates follow a path that satisfies optimality and complementarity, aiming a feasible point along this path. The arguments for proving this property are all contained in the proof of Theorem 4.2 of [2]. However, we found it interesting to isolate them from that specific context in order to stress the independent interest of the result.

Lemma 3.2 *Assume that the sequence $\{x^k\}$ is generated by Algorithm 2.1 or Algorithm 2.2 and K is an infinite sequence of indices such that*

$$\lim_{k\in K} x^k = x^*.$$

Then, for all $k \in K$ large enough,

$$g_i(x^*) < 0 \quad \Rightarrow \quad \mu_i^{k+1} = 0$$

and

$$\underline{g}_i(x^*) < 0 \quad \Rightarrow \quad w_i^k = 0.$$

Proof By Lemma 3.1, we have that for all $k \in K$, there exist $w^k \in \mathbb{R}_+^{\underline{p}}$, $\delta^k \in \mathbb{R}^n$ such that $\|\delta^k\| \leq \varepsilon_k$ and

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i^{k+1} \nabla g_i(x^k) + \gamma_k(x^k - \bar{x}^{k-1})$$
$$+ \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^{\underline{p}} w_i^k \nabla \underline{g}_i(x^k) = \delta^k.$$

By (8), $\mu^{k+1} \in \mathbb{R}_+^p$ for all $k \in \mathbb{N}$.

Suppose that $\underline{g}_i(x^*) < 0$. Then, there exists $k_1 \in \mathbb{N}$ such that $\forall k \in K$, $k \geq k_1$, $\underline{g}_i(x^k) < -\varepsilon_k$. Then, by (5),

$$w_i^k = 0 \quad \forall k \in K, k \geq k_1.$$

Now, assume that $g_i(x^*) < 0$. In this case, there exists $k_2 \geq k_1$ such that

$$g_i(x^k) < c < 0 \quad \forall k \in K, k \geq k_2.$$

We consider two cases:

1. $\{\rho_k\}$ is unbounded.
2. $\{\rho_k\}$ is bounded.

In the first case, we have that $\lim_{k\in K} \rho_k = \infty$. Since $\{\bar{\mu}_i^k\}$ is bounded, there exists $k_3 \geq k_2$ such that, for all $k \in K, k \geq k_3$,

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

By the definition of μ^{k+1} this implies that

$$\mu_i^{k+1} = 0 \quad \forall k \in K, k \geq k_3.$$

Now, consider the case in which $\{\rho_k\}$ is bounded. In this case, by Step 4,

$$\lim_{k\to\infty} V_i^k = 0.$$

Therefore, since $g_i(x^k) < c < 0$ for $k \in K$ large enough,

$$\lim_{k\in K} \bar{\mu}_i^k = 0.$$

Then, for $k \in K$ large enough,

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

By the definition of μ^{k+1}, there exists $k_4 \in \mathbb{N}$ such that $\mu_i^{k+1} = 0$ for all $k \in K$, $k \geq k_4$.

Therefore, there exists $k_5 \geq \max\{k_3, k_4\}$ such that for all $k \in K, k \geq k_5$,

$$g_i(x^*) < 0 \quad \Rightarrow \quad \mu_i^{k+1} = 0 \quad \text{and} \quad \underline{g}_i(x^*) < 0 \quad \Rightarrow \quad w_i^k = 0,$$

as we wanted to prove. □

The asymptotic joint behavior of the regularization parameter, together with the penalty parameter and the reference point is described in Lemma 3.3. Thanks to this result we are able to reproduce the convergence results of Algencan [2].

Lemma 3.3 *Assume that $\{x^k\}$ is a bounded sequence generated by Algorithm 2.1 or Algorithm 2.2. Then,*

$$\lim_{k\to\infty} \frac{\gamma_k}{\rho_k} \|x^k - \bar{x}^{k-1}\| = 0. \tag{23}$$

Proof The sequence $\{\bar{x}^k\}$ is bounded since, by definition, $\{\bar{x}^k\} \subset \{x^k\}$.

Since $\{\bar{\mu}^k\}$ is also bounded, we have that $\{R_k\}$ is bounded. Therefore, since $\gamma_{k+1} \le \beta R_k$ for all k, $\{\gamma_k\}$ is bounded.

If $\{\rho_k\}$ is bounded, then the penalty parameter ceases to increase for k large enough. Therefore, $\lim_{k\to\infty} R_k = 0$. Then, since $\gamma_{k+1} \le \beta R_k$, we have that $\lim_{k\to\infty} \gamma_k = 0$. Since $\|x^k - \bar{x}^{k-1}\|$ is bounded, it turns out that (23) takes place.

If $\{\rho_k\}$ tends to infinity, the boundedness of $\{\gamma_k\}$ and $\|x^k - \bar{x}^{k-1}\|$ also implies that (23) holds. □

Theorem 3.1 *Assume that x^* is a limit point of a bounded sequence generated by Algorithm 2.1 or Algorithm 2.2. Then*:

1. *At least one of the following two possibilities holds*:
   - *The point x^* fulfills the KKT conditions of the problem*

$$\begin{array}{lll} \textit{Minimize} & \|h(x)\|^2 + \|g(x)_+\|^2 & \\ \textit{s.t.} & \underline{h}(x) = 0, & \underline{g}(x) \le 0. \end{array}$$

   - *The CPLD constraint qualification is not fulfilled at x^* for the lower level constraints $\underline{h}(x) = 0$, $\underline{g}(x) \le 0$.*
2. *If x^* is a feasible point of (1) then at least one of the following two possibilities holds*:
   - *The point x^* fulfills the KKT conditions of (1).*
   - *The CPLD constraint qualification is not satisfied at x^* for the constraints $h(x) = 0$, $g(x) \le 0$, $\underline{h}(x) = 0$, $\underline{g}(x) \le 0$.*

Proof Let $K \underset{\infty}{\subset} \mathbb{N}$ be such that

$$\lim_{k\in K} x^k = x^*.$$

The first part of the thesis is obtained following the same sequence of arguments used in Theorem 4.1 of [2]. The only difference is that we need to use (23) at the proper places. (For this reason we included the assumption that the whole sequence $\{x^k\}$ is bounded.)

Let us prove the second part of the statement. By (3–6) and Lemma 3.2, for $k \in K$ large enough we have that:

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{g_i(x^*)=0} \mu_i^{k+1} \nabla g_i(x^k) + \gamma_k (x^k - \bar{x}^{k-1})$$

$$+ \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{\underline{g}_j(x^*)=0} w_j^k \nabla \underline{g}_j(x^k) = \underline{\delta}^k,$$

where $\mu^{k+1} \in \mathbb{R}_+^p$, $w^k \in \mathbb{R}_+^{\underline{p}}$ and $\|\underline{\delta}^k\| \to 0$.

Now, since x^* is feasible, we have that $\lim_{k \in K} R_k = 0$. In the case of Algorithm 2.1 we have, by (13), that

$$\lim_{k \to \infty} \gamma_k = 0. \tag{24}$$

In the case of Algorithm 2.2, (24) follows from Lemma 2.4. Therefore, defining

$$\delta^k = \underline{\delta}^k - \gamma_k (x^k - \bar{x}^{k-1}),$$

we have that, for $k \in K$ large enough,

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{g_i(x^*)=0} \mu_i^{k+1} \nabla g_i(x^k)$$

$$+ \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{\underline{g}_j(x^*)=0} w_j^k \nabla \underline{g}_j(x^k) = \delta^k,$$

where $\mu^{k+1} \in \mathbb{R}_+^p$, $w^k \in \mathbb{R}_+^{\underline{p}}$ and $\|\delta^k\| \to 0$.

From this point on, we may follow the arguments of Theorem 4.2 of [2] for obtaining the desired result. □

4 Implementation and numerical examples

We implemented Algorithm 2.2 (RAlgencan) under the general framework of the Algencan package, maintaining the default parameters and internal choices of this implementation as well as the routines used for solving the subproblems. We used the version of Algencan available in March 2007. The tolerances for feasibility and optimality were $\varepsilon_{feas} = 10^{-8}$ and $\varepsilon_{opt} = 10^{-4}$. As initial point for the algorithm that solves subproblem k, we took the reference point $\bar{x}^{k-1}$. In this way we guarantee the fulfillment of (16).

We employed the following strategy for updating the regularization parameter:

- We use (14) with $R_{tol} = \max\{R_0, 1\}$ and $\gamma_1 = 0$.

- If $R_k = \min\{R_{tol}, R_1, \ldots, R_k\}$ we define $\gamma_{k+1} = 0$. Else, we define

$$\gamma_{k+1} = \min\{\beta R_k, \gamma_k + 1\}. \tag{25}$$

In well behaved cases Algencan satisfies $R_{tol} \geq R_k \geq R_{k+1}$ at most iterations. This is the reason why the performance of RAlgencan is almost identical to the performance of Algencan in those situations. As an example, we considered the family of problems defined by:

$$\begin{aligned} &\text{Minimize} \quad \sum_{i>j} \frac{1}{\|P_i - P_j\|} \\ &\text{subject to} \quad P_k \in \mathbb{R}^{n_{dim}}, \quad \|P_k\| \leq 1, \quad k = 1, \ldots, n_{pun} \end{aligned}$$

for different values of n_{dim}, n_{pun}. We wrote $x = (P_1, \ldots, P_{n_{pun}})$ and we defined the initial point by

$$x_i^0 = \sin(i), \quad i = 1, \ldots, n.$$

We ran Algencan and RAlgencan for the problems defined by $n_{dim} = 3$, $n_{pun} = 10, 20, \ldots, 100$. Both algorithms obtained the same solutions in the 10 problems. In 6 problems ($n_{pun} = 10, 20, 40, 60, 90, 100$) both used the same number of iterations. Algencan converged in 27 iterations for $n_{pun} = 30$ and in 25 iterations for $n_{pun} = 50$. In these problems RAlgencan employed 28 iterations. On the other hand, RAlgencan converged in 19 iterations for $n_{pun} = 70$ and in 18 iterations for $n_{pun} = 80$. Algencan needed 23 and 21 iterations respectively for those problems.

Then, we considered the following 6 problems, in which Algencan exhibited the greedy behavior. In all these problems the lower-level constraint set Ω was the (artificial) box $[-10^{20}, 10^{20}]^n$.

Problem 1

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{n} x_i^3 \\ &\text{subject to} \quad g_i(x) \equiv -x_i \leq 0, \quad i = 1, \ldots, n. \end{aligned}$$

Initial point: $(-7, \ldots, -7)$, $n = 100$.

Results

- Algencan stopped at iteration 9, due to impossibility of improving feasibility during 9 consecutive iterations. The final norm of the constraint was $\approx 10^{20}$. The final objective function value was -10^{60}.
- RAlgencan detected greediness at the first iteration with $f(x^1) = -10^{62}$ and $R_1 \approx 10^{20}$. So, x^0 was used as reference point and $\gamma_2 = 1$. The same situation was repeated the next 5 iterations, in which the regularization parameter was updated. At iteration 7 the minimization subproblem used $\gamma_7 = 6$. The solution of the seventh subproblem was finally acceptable, with $R_7 \approx 0.004$ and $f(x^7) \approx -0.98$. As a consequence, the algorithm chose $\gamma_8 = 0$ and convergence ocurred at iteration 9, with the fulfillment of the feasibility-optimality convergence criterion.

Problem 2

$$\begin{array}{ll} \text{Minimize} & -x_1x_2x_3 \\ \text{subject to} & h_1(x) \equiv x_1 - 4.2\sin(x_4)^2 = 0, \\ & h_2(x) \equiv x_2 - 4.2\sin(x_5)^2 = 0, \\ & h_3(x) \equiv x_3 - 4.2\sin(x_6)^2 = 0, \\ & h_4(x) \equiv x_1 + 2x_2 + 2x_3 - 7.2\sin(x_7)^2 = 0. \end{array}$$

Initial point: $(1, 2, 3, 4, 5, 6, 7)$.

Results

- At the initial point we have $f(x^0) \approx -6$ and $R_0 \approx 7.9$. The first iteration of Algencan was greedy, obtaining $f(x^1) \approx -10^6$, $R_1 \approx 10^{20}$. Algencan could not improve feasibility in the next iterations, in spite of the growth of the penalty parameters. The algorithm stopped at iteration 9, due to an unacceptably big penalty parameter.
- At the first iteration, with $\gamma_1 = 0$, RAlgencan obtained, of course, the same iterate x^1 as Algencan did. Therefore, the algorithm took x^0 as reference point and increased the regularization parameter. With $\gamma_2 = 1$, x^2 was still very close to x^1 but, with $\gamma_3 = 2$, a quite reasonable iterate x^3 was computed, with $f(x^3) \approx -2.76$ and $R_3 \approx 0.12$. As a consequence, RAlgencan chose $\gamma_4 = 0$. However, due to moderate deterioration of feasibility, after 4 iterations, it was necessary to take $\gamma_8 = 1$ and $\gamma_9 = 2$. This was enough to find a good improvement of R_k, so that $\gamma_{10} = \gamma_{11} = 0$ and convergence criteria were met at x^{11}.

Problem 3

$$\begin{array}{ll} \text{Minimize} & -x_1x_2^3 \\ \text{subject to} & h_1(x) \equiv x_1x_2 - 4\sin(x_1)^2 = 0. \end{array}$$

Initial point: $(1, 1)$.

Results

- Algencan computes a very greedy infeasible first outer iteration. Therefore, the penalty parameter is increased many times. After each penalty parameter increase, feasibility is slightly improved. However, after 40 outer iterations the method stops with a very large penalty parameter and a very infeasible point. The final norm of the constraint is greater than 10^{20}.
- After the first greedy iteration, RAlgencan uses x^0 as reference point associated with $\gamma_2 = 1$ as regularization parameter. The second iteration was also greedy and γ_k needed to be increased once more. With $\gamma_3 = 2$ greediness was overcome and the algorithm obtained $f(x^3) \approx -43$ and $R_3 \approx 0.46$. From this iteration on, accuracy improved monotonically, and the regularization parameter was zero. Convergence criteria were met at iteration 18, with $f(x^{18}) \approx -30$.

Problem 4

$$\begin{array}{ll} \text{Minimize} & -x_1\exp(-x_1x_2) \\ \text{subject to} & h_1(x) \equiv -(x_1+1)^3 + 3(x_1+1)^2 - 1.5 + x_2 = 0. \end{array}$$

Initial point: $(1, -1.5)$.

Results

- Algencan exhibited greediness at the first iteration, with $f(x^1) \approx -\infty$ and $R_1 \approx 10^{12}$. The penalty parameter was increased during 9 consecutive iterations without modification of this state of facts. Divergence was declared at iteration 21.
- The first iteration was greedy, therefore, RAlgencan increased the regularization parameter and the reference point for the second iteration was the initial point x^0. As a consequence, the second iteration was less infeasible than the first. However, since its infeasibility is substantially greater than the one of the initial point, the regularization parameter increased. Increase of γ_k continued until iteration 7. With $\gamma_7 = 6$ we obtained $f(x^7) \approx -2.3$ and $R_7 = 10^{-4}$. From then on, the algorithm used $\gamma_k = 0$. Convergence ocurred at iteration 10, with $f(x^{10}) \approx -2.28$.

Problem 5

$$\begin{aligned} \text{Minimize} \quad & -\sum_{i=1}^{n}(x_i^8 + x_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i^2 \le 1. \end{aligned}$$

Initial point: $(0.1, \ldots, 0.1)$, $n = 50$.

Results

- At the initial point we have $f(x^0) \approx -5$, $R_0 = 0$. The first iteration of Algencan leads to $f(x^1) = -10^{99}$, $R_1 \approx 10^{26}$. The situation was not modified in the next 20 iterations and divergence was declared at iteration 21.
- After the first greedy iteration, Ralgencan took x^0 as reference point and ran the second subproblem with $\gamma_2 = 1$. As a result, greediness was corrected, obtaining $f(x^2) \approx -8$ and $R_2 \approx 0.29$. At the following iterations the algorithm used $\gamma_k = 0$. Convergence ocurred at iteration 15 with $f(x^{15}) = -0.71$.

Problem 6

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^{n}\varphi(x_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i^2 \le 1, \end{aligned}$$

where $\varphi(t) = \log(\cos(t))$ if $\cos(t) > 0$; $\varphi(t) = -10^{30}$ otherwise.

We used $n = 100$.

Initial point: $(1/n, \ldots, 1/n)$.

Results The peculiarity of this problem is that the objective function is defined arbitrarily as $f(x) = -10^{30}$ at the points x where, in principle, $f(x)$ would not be defined. All these points are infeasible.

- At the initial (feasible) point we have $f(x^0) \approx -0.005$ and $R_0 = 0$. The initial iterate computed by Algencan gave $f(x^1) = -10^{30}$ and $R_1 \approx 250$. At all the subsequent Algencan iterates the objective function value was -10^{30}. The feasibility-complementarity measure R_k oscilated around 250 and divergence was declared at iteration 22.
- The second iteration of RAlgencan was computed with $\gamma_2 = 1$ and was not able to correct the initial greediness. However, with $\gamma_3 = 2$ and preserving x^0 as reference point, we obtained $f(x^3) \approx -0.002$ and $R_3 = 0$. With $\gamma_4 = 0$, iteration 4 was greedy again, giving $f(x^4) = -10^{30}$ and $R_4 \approx 246$. Greediness was corrected at iteration 5, with $\gamma_5 = 1$, $f(x^5) = -0.5$ and $R_5 \approx 10^{-4}$. Since R_5 was not the minimum of $\{R_{tol}, R_1, \ldots, R_5\}$ the algorithm chose, according to (25), $\gamma_6 = \beta R_5 \approx 0.102$. This mild regularization produced again a greedy iteration $f(x^6) = -10^{30}$, $R_6 \approx 246$. So, the algorithm defined $\gamma_7 \approx 1.102$. Greediness was corrected once more, allowing the algorithm to choose $\gamma_8 = 0.006$ due to the limitation βR_k. Alternate greedy and non-greedy iterations followed until iteration 12. With $\gamma_{12} = 1$ we got $f(x^{12}) \approx -0.5$ and $R_{12} \approx 10^{-8}$. Finally, convergence was detected at iteration 14.

5 Final remarks

The greediness phenomenon is defined in this paper as the tendency of some nonlinear programming methods to find very infeasible points with very small functional values, in general, at the first iterations. This phenomenon may occur even when one has good (perhaps feasible) initial approximations to the solution of the problem. Frequently, the points found in this way are local minimizers of the infeasibility measure and the method converges to them even from the theoretical point of view.

The remedy proposed in this paper to alleviate this inconvenient is to replace the usual Augmented Lagrangian iteration by a regularized iteration with respect to a suitably defined reference point. We showed that the resulting methods preserve the global convergence properties of the original algorithm. From the practical point of view, we showed that the modification proposed does not seem to harm the behavior of Algencan (the method introduced in [2]) when this algorithm behaves well. We also provided some evidence that, in situations where Algencan fails because of the greediness effect, the regularization procedure may help to put the algorithm in the correct trajectory.

An additional consequence of the regularization strategy is that conditioning of the subproblems is improved. Usually, this type of improvement is accompanied by a decrease of the speed of convergence. However, this phenomenon was not observed in our numerical experiments using problems without greediness. This is due to the particular strategy employed for reducing the regularization parameter, that tends to zero when feasibility-complementarity progresses. Therefore, it may be recommendable to incorporate the regularization strategy as a standard procedure of Algencan.

Considering that the greediness phenomenon is a serious drawback for the application of generally useful nonlinear programming algorithms that use globalizing devices combining feasibility and optimality, we think that further research may be expected, along the following lines:

- Alternative heuristic choices of regularization parameters, preserving the theoretical requirements of Algorithms 2.1 or 2.2.
- Anti-greedy strategies not based on regularization.
- Warm-start procedures based on a priori knowledge of the quality of the initial point.

Acknowledgements We are indebted to two anonymous referees for their useful and insightful comments.

References

1. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On the relation between the constant positive linear dependence condition and quasinormality constraint qualification. J. Optim. Theory Appl. **125**, 473–485 (2005)
2. 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)
3. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Nashua (2000).
4. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust Region Methods, MPS/SIAM Series on Optimization. Philadelphia, SIAM (2000)
5. Fiacco, A.V., McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York (1968)
6. Fletcher, R.: Practical Methods of Optimization. Academic Press, London (1987)
7. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
8. Iusem, A.N.: Augmented Lagrangian methods and proximal point methods for convex optimization. Investig. Oper. **8**, 11–49 (1999)
9. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1999)
10. 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)
11. Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
12. Rockafellar, R.T.: A dual approach to solving nonlinear programming problems by unconstrained optimization. Math. Program. **5**, 354–373 (1973)
13. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM J. Control Optim. **12**, 268–285 (1974)