Numer Algor (2012) 60:263–277
DOI 10.1007/s11075-012-9561-2

ORIGINAL PAPER

Handling infeasibility in a large-scale nonlinear optimization algorithm

Jose Mario Martínez · Leandro da Fonseca Prudente

Received: 3 November 2011 / Accepted: 8 March 2012 /
Published online: 24 March 2012

Abstract Practical Nonlinear Programming algorithms may converge to infeasible points. It is sensible to detect this situation as quickly as possible, in order to have time to change initial approximations and parameters, with the aim of obtaining convergence to acceptable solutions in further runs. In this paper, a recently introduced Augmented Lagrangian algorithm is modified in such a way that the probability of quick detection of asymptotic infeasibility is enhanced. The modified algorithm preserves the property of convergence to stationary points of the sum of squares of infeasibilities without harming the convergence to KKT points in feasible cases.

Keywords Augmented lagrangians · Nonlinear programming · Algorithms · Numerical experiments

1 Introduction

In Constrained Optimization, one aims to find the lowest possible value of an objective function within a given domain. Global Optimization is very hard, especially in large-scale problems: full guarantee that a given point is a global minimizer of a continuous function can be obtained, if additional properties of

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

J. M. Martínez (✉) · L. F. Prudente
Department of Applied Mathematics, IMECC-UNICAMP,
University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil
e-mail: martinez@ime.unicamp.br

L. F. Prudente
e-mail: lprudente@ime.unicamp.br

the function are not available, only after visiting a dense set in the domain. Of course, such a search is impossible except in low-dimensional problems.

Affordable optimization algorithms (usually called "local algorithms" in the global optimization literature) generate sequences of iterates that, in the limit, use to find, in the best case, Karush-Kuhn-Tucker (KKT) points. In the worst case only (infeasible) stationary points for some infeasibility measure are found. In the second case, one suspects that the problem is infeasible. However, every affordable optimization algorithm can converge to an infeasible point, even when feasible points exist. Therefore, optimization users that wish to find feasible and optimal solutions of practical problems usually change the initial approximation and/or the algorithmic parameters of the algorithm when an almost infeasible point is found. The effectiveness of this trial-and-error process is, in part, related with the ability of the algorithm of stopping quickly when the generated sequence is faded to converge to an infeasible point.

As a consequence, practical optimization algorithms should be effective, not only for finding solutions of the problems (when they exist), but also for finding infeasible points, when there is no alternative. This feature is never considered in comparative numerical studies. Usually, convergence to an infeasible point is computed as a failure, without taking into account that a "quick failure" gives rise to the possibility of making better in reasonable computer time, whereas a "slow failure" does not. Clearly, practical users that wish to solve effectively their problems, should prefer algorithms in which failure detection is as fast as possible.

In this paper we focus our analysis in algorithms of Augmented Lagrangian type [1, 8, 14, 21, 22, 25]. In particular, in the algorithm introduced in [1], whose computer implementation (called Algencan) is available in www.ime.usp.br/∼egbirgin/tango, the iterates x^k are computed as approximate minimizers (with increasing precision) of augmented Lagrangians in which multipliers and penalty parameters are updated. The "increasing precision" requirement makes it very difficult to solve subproblems when the penalty parameter goes to infinity, which is necessarily the case when a feasible point is not found. Consequently, Algencan may employ a lot of computer time to declare that the problem is, perhaps, infeasible. Here, we will observe that, in that case, the same convergence results are obtained using bounded away from zero tolerances for solving the subproblems. This fact motivates the employment of dynamic adaptive tolerances that depend on the degree of infeasibility and complementarity at each iterate x^k. Adaptive precision control for optimality depending on infeasibility measures has been considered, with different purposes or in different contexts, in [1, 14–16, 20]. In these works the main preoccupation is to guarantee that the subproblem solution is accurate enough if the point is almost feasible, whereas in our case we want to take advantage of the fact that one does not need great accuracy in the absence of near-feasibility. The problem of quick detection of infeasibility has been considered in [11] in the context of a Sequential Quadratic Programming (SQP) method. In many other algorithms (which employ SQP, filters, or Interior Point techniques) specific restoration procedures are responsible for the detection of minimizers of the infeasibility. See, among others, [7, 12, 13, 17–19, 27, 28]. An interesting result where the detection of infeasibility does not depend on restoration procedures at all may be found in [26].

In this paper we will show that the convergence properties of [1] are preserved by the new algorithm and that efficiency is improved in the sense discussed in this introduction.

Notation If $v \in I\!R^n$, $v = (v_1, \ldots, v_n)$, we denote $v_+ = (\max\{0, v_1\}, \ldots, \max\{0, v_n\})$. If $K = (k_1, k_2, \ldots) \subseteq I\!N$ (with $k_j < k_{j+1}$ for all j), we denote $K \underset{\infty}{\subset} I\!N$. The symbol $\|\cdot\|$ will denote the Euclidian norm. For all $z \in I\!R^n$, $P(z)$ will denote the Euclidean projection of z on the box Ω. If $y \in I\!R^n$, its i-th component will be denoted by $[y]_i$ or by y_i, if this does not lead to confusion.

2 Algorithm

The problem considered in this paper is:

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

where $h : I\!R^n \to I\!R^m$, $g : I\!R^n \to I\!R^p$, $f : I\!R^n \to I\!R$ are smooth and $\Omega \subset I\!R^n$ is a bounded n-dimensional box given by

$$\Omega = \{x \in I\!R^n \mid a_i \le x_i \le b_i \; \forall\, i = 1, \ldots, n\}.$$

The Augmented Lagrangian function [21, 22, 25] will be defined by:

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

for all $x \in \Omega$, $\rho > 0$, $\lambda \in I\!R^m$, $\mu \in I\!R^p_+$.

Below we describe the basic Algencan algorithm, which differs from the one stated in [1] only in the stopping criterion for the subproblem. In the original Algencan one imposes that the convergence tolerances for the subproblems $\{\varepsilon_k\}$ should tend to zero, whereas in Algorithm 2.1 this requirement is relaxed.

Algorithm 2.1

Let $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $0 < \tau < 1$. Let $\bar{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}]$, $i = 1, \ldots, m$, $\bar{\mu}_i^1 \in [0, \mu_{\max}]$, $i = 1, \ldots, p$, and $\rho_1 > 0$. Initialize $k \leftarrow 1$. We assume that $\{\varepsilon_k\}$ is a bounded sequence of positive numbers.

Step 1. Find $x^k \in \Omega$ as an approximate minimizer of $L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k)$ on Ω. By this we mean that, for all $i = 1, \ldots, n$, we have:

$$\begin{aligned} &\left|\left[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)\right]_i\right| \min\left\{1, \left|x_i^k - a_i\right|\right\} \\ &\quad \le \varepsilon_k \ \text{ if } \ \left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i > 0 \end{aligned} \tag{2}$$

and

$$\begin{aligned} &\left|\left[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)\right]_i\right| \min\left\{1, \left|x_i^k - b_i\right|\right\} \\ &\quad \le \varepsilon_k \ \text{ if } \ \left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i < 0. \end{aligned} \tag{3}$$

Step 2. Define

$$V_i^k = \min\left\{-g_i\left(x^k\right), \frac{\bar{\mu}_i^k}{\rho_k}\right\}, i = 1, \ldots, p.$$

If $k = 1$ or

$$\max\left\{\left\|h\left(x^k\right)\right\|_\infty, \left\|V^k\right\|_\infty\right\} \le \tau \ \max\left\{\left\|h\left(x^{k-1}\right)\right\|_\infty, \left\|V^{k-1}\right\|_\infty\right\}, \tag{4}$$

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

Step 3. Compute $\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}]$, $i = 1, \ldots, m$ and $\bar{\mu}_i^{k+1} \in [0, \mu_{\max}]$, $i = 1, \ldots, p$. Set $k \leftarrow k + 1$ and go to Step 1.

Let us give a natural interpretation of the criteria (2) and (3). Consider the linear approximation of $L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k)$ at x^k. Clearly, a minimizer z of this linear approximation on Ω satisfies

$$z_i = a_i \ \text{ if } \ \left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i > 0$$

and

$$z_i = b_i \ \text{ if } \ \left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i < 0.$$

Therefore, the modulus of the difference between $L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)$ and the value of the linear approximation at its minimizer is $|\sum_{i=1}^n [\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)]_i \ (x_i - c_i)|$, where $c_i = a_i$ if $[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)]_i > 0$, and $c_i = b_i$ if $[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)]_i < 0$. Therefore, the condition $|[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)]_i \ (x_i^k - c_i)| \le \varepsilon_k$ $(i = 1, \ldots, n)$ guarantees that the modulus of the difference between $L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)$ and the linear approximation minimum is smaller than $n\varepsilon_k$. This interpretation shows that the condition $|[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)]_i \ (x_i^k - c_i)| \le \varepsilon_k$ could be unnecessarily rigid, since the bounds that define Ω could be very far from x^k (sometimes artificially big bounds are employed in practical optimization). This is the reason why, in (2) and (3), $|x_i^k - c_i|$ is replaced by $\min\{1, |x_i^k - c_i|\}$.

The solvability of the subproblem (2) and (3) is guaranteed employing standard bound-optimization solvers. In Algencan, it is used the active-set projected-gradient method introduced in [10]. In [1] it was assumed that $\lim_{k\to\infty} \varepsilon_k = 0$. Here we wish to exploit the case in which ε_k does not tend to zero. Several results follow almost exactly as in [1], only observing that the assumption $\varepsilon_k \to 0$ is not used in those cases.

If $s, t \geq 0$ and $st \leq \varepsilon$ we have that $\min\{s, t\} \leq \sqrt{\varepsilon}$. Using this obvious fact, we note that (2) and (3) imply that

$$\left|\min\left\{\left[\nabla L_{\rho_k}\left(x, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i, 1, x_i^k - a_i\right\} \leq \sqrt{\varepsilon_k} \text{ if } \left[\nabla L_{\rho_k}\left(x, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i > 0 \quad (5)$$

and

$$\left|\min\left\{-\left[\nabla L_{\rho_k}\left(x, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i, 1, b_i - x_i^k\right\}\right| \leq \sqrt{\varepsilon_k} \text{ if } \left[\nabla L_{\rho_k}\left(x, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i < 0 \quad (6)$$

By direct calculation, (5) and (6) imply that, if $\varepsilon_k < 1$,

$$\left\| P\left(x^k - \nabla L_{\rho_k}\left(x, \bar{\lambda}^k, \bar{\mu}^k\right)\right] - x^k \right\|_\infty \leq \sqrt{\varepsilon_k}. \quad (7)$$

From now on, we define

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

and

$$\mu^{k+1} = \left(\bar{\mu}^k + \rho_k g\left(x^k\right)\right)_+ \quad (9)$$

for all $k \in I\!N$. At each subproblem solution the projected gradient of the Lagrangian (with multipliers λ^{k+1} and μ^{k+1}) is smaller than the prescribed tolerance. In general, one chooses $\bar{\lambda}^{k+1}$ and $\bar{\mu}^{k+1}$ as the projections of λ^{k+1} and μ^{k+1} on $[\lambda_{\min}, \lambda_{\max}]^m$ and $[0, \mu_{\max}]^p$, respectively. The Lagrange multipliers approximations λ^{k+1} and μ^{k+1} may be unbounded, unlike their safeguarded counterparts, which are kept bounded to preserve stability.

Lemma 2.1 *Assume that $\{x^k\}_{k\in K}$ is a subsequence of the sequence generated by Algorithm* 2.1 *and $x^* = \lim_{k\in K} x^k$. Then, for all $i = 1, \ldots, p$, if $g_i(x^*) < 0$ and $k \in K$ is large enough, one has that $\mu_i^{k+1} = 0$.*

Proof If $\{\rho_k\}$ is bounded we have that $\lim_{k\to\infty} V_i^k = 0$ for all $i = 1, \ldots, p$. If $g_i(x^*) < 0$ this implies that $\lim_{k\in K} \bar{\mu}_i^k/\rho_k = 0$. By the boundesness of $\{\rho_k\}$ it turns out that $\lim_{k\in K} \bar{\mu}_i^k = 0$. Therefore, for $k \in K$ large enough, $(\bar{\mu}_i^k + \rho_k g_i(x^k))_+ = 0$. So, $\mu_i^{k+1} = 0$ for $k \in K$ large enough. If ρ_k tends to infinity and $g_i(x^*) < 0$, by the boundedness of $\{\bar{\mu}^k\}$, one has that $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$ for $k \in K$ large enough. Therefore, $\mu_i^{k+1} = 0$ as in the bounded case. □

Lemma 2.2 *Under the assumptions of Lemma* 2.1*, for all $k \in K$ large enough, if $\varepsilon_k < 1$, one has:*

$$\left\| P\left(x^k - \left[\nabla f\left(x^k\right) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla h_i\left(x^k\right) + \sum_{g_i(x^*) \ge 0} \mu_i^{k+1} \nabla g_i\left(x^k\right)\right]\right) - x^k \right\|_\infty \le \sqrt{\varepsilon_k}. \quad (10)$$

Proof Use (7)–(9) and Lemma 2.1. □

The following result shows that, under the assumptions of previous lemmas, in the case that $\{\rho_k\}$ is bounded and $\varepsilon_k \to 0$, limit points satisfy classical optimality conditions for local minimization. As previous ones, this result is proved in [1] in the context of a more complete convergence theorem.

Lemma 2.3 *Let us assume again that the hypotheses of Lemma* 2.1 *hold. Suppose, further, that the sequence $\{\rho_k\}$ is bounded. Then, x^* is feasible. Moreover, if $\lim_{k \in K} \varepsilon_k = 0$, x^* satisfies the Karush-Kuhn-Tucker (KKT) conditions for constrained local minimization*

Proof By (4) and the boundedness of $\{\rho_k\}$ we obtain that x^* is feasible. By the boundedness of $\{\rho_k\}$, $\{\bar{\lambda}^k\}$, $\{\bar{\mu}^k\}$, $\{h(x^k)\}$, and $\{g(x^k)\}$, and the definitions (8) and (9) we have that the sequences $\{\lambda^{k+1}\}$ and $\{\mu^{k+1}\}$ are bounded. Therefore, taking limits in (10) for an appropriate subsequence, we obtain the desired result. □

It remains to analyze the behavior of Algorithm 2.1 when $\rho_k \to \infty$. Lemma 2.4 below shows that, in that case, limit points are stationary points of an infeasibility measure, even without the requirement that ε_k tends to zero.

Lemma 2.4 *Under the assumptions of Lemma* 2.1*, if $\lim_{k \to \infty} \rho_k = \infty$, the limit point x^* is a stationary point of*

$$\textit{Minimize } \|h(x)\|^2 + \|g(x)_+\|^2 \ \textit{ subject to } \ x \in \Omega. \quad (11)$$

Proof Suppose that

$$\left[\nabla\left[\left\|h\left(x^*\right)\right\|^2 + \left\|\left(g\left(x^*\right)\right)_+\right\|^2\right]\right]_i > 0. \quad (12)$$

Then, by the boundedness of $\{\nabla f(x^k)\}$, $\{\bar{\lambda}^k\}$, and $\{\bar{\mu}^k\}$, and the continuity of h and g, since $\rho_k \to \infty$, we have that, for $k \in K$ large enough,

$$\left[(1/\rho_k)\,\nabla f\left(x^k\right) + (1/2)\nabla\left[\left\|h\left(x^k\right) + \bar{\lambda}^k/\rho_k\right\|^2 + \left\|\left(g\left(x^k\right) + \bar{\mu}^k/\rho_k\right)_+\right\|^2\right]\right]_i > 0.$$

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

$$\left[\nabla f\left(x^k\right) + (\rho_k/2)\,\nabla\left[\left\|h\left(x^k\right) + \bar{\lambda}^k/\rho_k\right\|^2 + \left\|\left(g\left(x^k\right) + \bar{\mu}^k/\rho_k\right)_+\right\|^2\right]\right]_i > 0.$$

Therefore, by (2),

$$\left|\left[\nabla f\left(x^k\right) + (\rho_k/2)\,\nabla\left[\left\|h\left(x^k\right) + \bar{\lambda}^k/\rho_k\right\|^2 + \left\|\left(g\left(x^k\right) + \bar{\mu}^k/\rho_k\right)_+\right\|^2\right]\right]_i\right. \times \min\left\{1, x_i^k - a_i\right\} \le \varepsilon_k \tag{13}$$

for $k \in K$ large enough. Dividing both members of (13) by ρ_k, using continuity of f, h, and g, and boundedness of $\{\bar{\lambda}^k\}$, $\{\bar{\mu}^k\}$, and $\{\varepsilon_k\}$, and taking limits, we obtain:

$$\left|\left[\nabla\left[\left\|h\left(x^*\right)\right\|^2 + \left\|\left(g\left(x^*\right)\right)_+\right\|^2\right]\right]_i\right. \min\left\{1, x_i^* - a_i\right\} = 0.$$

Therefore, the inequality (12) implies that the partial derivative $[\nabla[\|h(x^*)\|^2 + \|(g(x^*))_+\|^2]_i$ is non-positive whenever $x_i^* > a_i$.

In an analogous way, we prove that, $[\nabla[\|h(x^*)\|^2 + \|(g(x^*))_+\|^2]_i$ is non-negative whenever $x_i^* < b_i$. These two facts imply that x^* is an stationary point of (11). □

Finally, in Lemma 2.5, we recall the result of [1] that shows that KKT conditions hold, in the limit, if a weak constraint qualification takes place. In [1] the Constant Positive Linear Dependence (CPLD) condition [5, 24] was invoked in this context. However, it has been recently shown that this result holds under even weaker constraint qualifications [3, 4].

Lemma 2.5 *Assume the hypotheses of Lemma* 2.4, *with* $\rho_k \to \infty$. *Suppose, further, that a weak constraint qualification (CPLD or the weaker ones presented in* [3, 4]*) holds at the feasible limit point* x^* *and that* $\lim_{k\in K} \varepsilon_k = 0$. *Then, the KKT conditions hold at* x^*.

Proof See [1] using (7). □

The lemmas above show that, discarding rounding errors, and assuming that infinite time is available, Algorithm 2.1 only fails in the case of convergence to infeasible points, and, in this case, one necessarily has that ρ_k tends to infinity and all the limit points are probably local minimizers of the infeasibility measure. The case in which KKT conditions do not hold at a feasible limit point due to lack of fulfillment of a constraint qualification cannot be considered a failure, since, even in this case, as shown by Lemma 2.2, approximate KKT conditions remain to hold (see [2, 6]), at least when $\varepsilon_k \to 0$.

In practical terms, the case in which Algencan converges to an infeasible point is usually detected by the growth of the penalty parameter ρ_k. When ρ_k becomes very large, Algencan stops with a message of possible infeasibility.

3 An adaptive stopping criterion for the subproblems

In Section 2 we showed that, when the limit is infeasible, the requirement $\varepsilon_k \to 0$ is not necessary at all, and, in practice, excessively small values of ε_k may contribute to increase computer time for solving subproblems. Therefore, variations of Algorithm 2.1 in which ε_k does not tend to zero in the case of infeasibility are necessary. An appropriate definition of ε_k by means of which this tolerance does not tend to zero unless strictly necessary, is given below. Let us define first a continuous increasing "forcing" function $\varphi : \mathbb{R}_+ \to \mathbb{R}$, such that $\varphi(0) = 0$. For all $k \in \mathbb{N}$ we define:

$$\varepsilon_k = \varphi \left[\left\| h\left(x^k\right) \right\| + \left\| \min\left\{ -g\left(x^k\right), \bar{\mu}^k / \rho_k \right\} \right\| \right]. \tag{14}$$

Clearly, with the definition (14), we have that $\lim_{k \in K} \varepsilon_k = 0$ only if $\lim_{k \in K} \|h(x^k)\| + \|g(x^k)_+\| = 0$ and, in this case, the limit of x^k for $k \in K$ must be feasible. If the limit of x^k for $k \in K$ is infeasible, the tolerances ε_k do not tend to zero and, so, we expect that the corresponding subproblems will be solved much faster than when we set $\varepsilon_k \to 0$. Note that, due to the convergence theory of the subproblem solver (see [10]), the requirements (2) and (3) can always be fulfilled if $\varepsilon_k > 0$ is fixed.

On the other hand, since ε_k is not defined before the execution of the subproblem solver, it is possible that, for all the internal iterations of this subalgorithm, the requirements (2) and (3) remain unfulfilled when one uses the definition (14). This means that the convergence analysis of Algorithm 2.1 with the definition (14) must consider two possibilities. In the first one, the subproblem solver always returns satisfying (2), (3) and (14). The second possibility is that, for some value of k, the subproblem solver is not able to stop because the projected gradient of the Augmented Lagrangian tends to zero slower than the feasibility-complementarity measure (14). We consider these possibilities in the following two theorems.

Theorem 3.1 *Assume that $\{x^k\}$ is an infinite sequence generated by Algorithm* 2.1 *with ε_k defined by* (14). *(This means that the subproblem solver is always able to satisfy* (2) *and* (3) *with* (14).*) Let $K \underset{\infty}{\subset} \mathbb{N}$ and $x^* \in \Omega$ be such that $\lim_{k \in K} x^k = x^*$. Then:*

- *x^* is a (feasible or infeasible) stationary point of* (11).
- *If $\{\rho_k\}$ is bounded, then x^* is feasible and satisfies the KKT conditions.*
- *If $\{\rho_k\}$ is unbounded, x^* is feasible, and satisfies a weak constraint qualification (as CPLD or the ones introduced in* [3, 4]*), then x^* satisfies the KKT conditions.*
- *If x^* is feasible, then, independently of constraint qualifications, we have that*

$$\lim_{k \in K} \left\| P\left(x^k - \left[\nabla f\left(x^k\right) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla h_i\left(x^k\right) + \sum_{g_i(x^*)=0} \mu_i^{k+1} \nabla g_i\left(x^k\right) \right] \right) - x^k \right\| = 0.$$

Proof By the boundedness of Ω, $\{\bar{\lambda}^k\}$, and $\{\bar{\mu}^k\}$, the continuity of h and g, and the definition (14), we have that the sequence $\{\varepsilon_k\}$ is bounded. Then, if $\rho_k \to \infty$, the first part of the thesis follows from Lemma 2.4.

In the case that $\{\rho_k\}$ is bounded, by (4) and (14), we have that $\lim_{k\to\infty} \varepsilon_k = 0$. Then, the second part of the thesis follows from Lemma 2.3.

Consider now the third part of the thesis. Since $h(x^*) = 0$ and $g(x^*) \le 0$, we have that

$$\lim_{k \in K} \left\| h\left(x^k\right) \right\| + \left\| g\left(x^k\right)_+ \right\| = 0. \tag{15}$$

By the boundedness of $\{\bar{\mu}^k\}$ we have that $\bar{\mu}^k/\rho_k \to 0$. Therefore, by (14) and (15), it turns out that $\lim_{k\in K} \varepsilon_k = 0$. Then, the desired result follows from Lemma 2.5.

The proof of the last part of the thesis follows from Lemma 2.2, after proving that $\lim_{k\in K} \varepsilon_k = 0$. This fact follows from (4) and (14) in the case that $\{\rho_k\}$ is bounded and is deduced from (14), (15), and the boundedness of $\bar{\lambda}^k$ and $\{\bar{\mu}^k\}$ when $\{\rho_k\}$ tends to infinity. This completes the proof of the theorem. □

It remains to consider the case in which Algorithm 2.1, with the definition (14), is not able to generate an infinite sequence $\{x^k\}$. This means that, for some finite value of k, the subproblem solver cannot fulfill the requirements (2), (3), and (14). A few words about "reasonable" subproblem solvers are necessary. A typical subproblem solver that aims to obtain (2) and (3) addresses the box-constrained minimization problem:

$$\text{Minimize } L_{\rho_k}\left(x, \bar{\lambda}^k, \bar{\mu}^k\right) \text{ subject to } x \in \Omega. \tag{16}$$

Well established affordable algorithms for handling (16) exist. These algorithms (as the one presented in [10], which is used in Algencan) typically generate a sequence $\{x^{k,\ell}\}$ that converges to a stationary point $x \in \Omega$. This implies that the sequence $\{x^{k,\ell}\}$ satisfies

$$\left|\lim_{\ell\to\infty} \left|\left[\nabla L_{\rho_k}\left(x^{k,\ell}, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i \left(x_i^{k,\ell} - a_i\right)\right| = 0 \text{ if } \left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i > 0 \quad (17)\right.$$

and

$$\left|\lim_{\ell\to\infty} \left|\left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i \left(x_i^k - b_i\right)\right| = 0 \text{ if } \left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i < 0 \quad (18)\right.$$

Therefore, if $\varepsilon_k > 0$ is given in advance, the fulfillment of (2) and (3) is guaranteed for ℓ large enough.

The problem is that, with the definition (14), the stopping tolerance for the subproblem solver is dependent of the current internal iterate $x^{k,\ell}$. As a consequence, in spite of going to zero, the quantities $|[\nabla L_{\rho_k}(x^{k,\ell}, \bar{\lambda}^k, \bar{\mu}^k)]_i\ (x_i^{k,\ell} - a_i)|$ (for $[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)]_i > 0$) or $|[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)]_i\ (x_i^k - b_i)|$ (for $[\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)]_i < 0$) could be always greater than $\varphi[\|h(x^{k,\ell})\| + \|\min\{-g(x^{k,\ell}), \bar{\mu}^k/\rho_k\|]$. This situation is considered in the following theorem.

Theorem 3.2 *Assume that, for some $k \in I\!N$, the sequence generated by Algorithm* 2.1*, with the definition* (14)*, and employing a subproblem solver with the properties* (17) *and* (18)*, cannot satisfy* (2)–(3)*. Then, every limit point of the sequence $\{x^{k,\ell}\}_{\ell \in I\!N}$ is feasible and satisfy the KKT conditions of* (1)*.*

Proof By the hypothesis of the theorem and (14), for all $\ell \in I\!N$ we have:

$$\max\left\{\left|\left[\nabla L_{\rho_k}\left(x^{k,\ell}, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i \left(x_i^{k,\ell} - a_i\right)\right| \left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i > 0\right\}$$
$$> \varphi\left[\left\|h\left(x^{k,\ell}\right)\right\| + \left\|\min\left\{-g\left(x^{k,\ell}\right), \bar{\mu}^k/\rho_k\right\}\right\|\right]$$

or

$$\max\left\{\left|\left[\nabla L_{\rho_k}\left(x^{k,\ell}, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i \left(x_i^{k,\ell} - b_i\right)\right| \left[\nabla L_{\rho_k}\left(x^k, \bar{\lambda}^k, \bar{\mu}^k\right)\right]_i < 0\right\}$$
$$> \varphi\left[\left\|h\left(x^{k,\ell}\right)\right\| + \left\|\min\left\{-g\left(x^{k,\ell}\right), \bar{\mu}^k/\rho_k\right\}\right\|\right]$$

Therefore, by (17) and (18),

$$\lim_{\ell\to\infty} \varphi\left[\left\|h\left(x^{k,\ell}\right)\right\| + \left\|\min\left\{-g\left(x^{k,\ell}\right), \bar{\mu}^k/\rho_k\right\}\right\|\right] = 0. \quad (19)$$

Let $x^* \in \Omega$ be a limit point of $\{x^{k,\ell}\}_{\ell \in I\!N}$. (Limit points necessarily exist as Ω is compact.) Taking limits in (19) we see that x^* is feasible. Moreover, if

$g_i(x^*) < 0$, (19) implies that $\bar{\mu}_i^k = 0$. Therefore, by (17), (18), and the deduction of (7),

$$\lim_{\ell \to \infty} \left\| P\left(x^{k,\ell} - \left[\nabla f\left(x^{k,\ell}\right) + (\rho_k/2) \sum_{i=1}^{m} \nabla \left(h_i\left(x^{k,\ell}\right) + \bar{\lambda}_i^k/\rho_k \right)^2 \right.\right.\right.$$
$$\left.\left.\left. + \ (\rho_k/2) \sum_{g_i(x^*)=0} \nabla \left(g_i\left(x^{k,\ell}\right) + \bar{\mu}_i^k/\rho_k \right)_+^2 \right] \right) - x^{k,\ell} \right\| = 0. \quad (20)$$

Taking limits in (20) for $\ell \to \infty$, we get:

$$\left\| P\left(x^* - \left[\nabla f\left(x^*\right) + \sum_{i=1}^{m} \bar{\lambda}_i^k \nabla h_i\left(x^*\right) + \sum_{g_i(x^*)=0} \bar{\mu}_i^k \nabla g_i\left(x^*\right) \right] \right) - x^* \right\| = 0.$$

Therefore, the limit point x^* is feasible and satisfies the KKT conditions of (1). □

Therefore, Theorem 3.2 says that the case in which the subproblem solver does not stop corresponds to the case in which the current approximation to the multipliers $\bar{\lambda}^k$ and $\bar{\mu}^k$ are the true multipliers for the fulfillment of KKT conditions at a limit point of the inner sequence $\{x^{k,\ell}\}$.

4 Numerical examples

In order to illustrate the reliability of our approach, we ran the original Algencan and the modified version introduced here both with feasible and infeasible problems. In the feasible problems, there exists a local minimizer of the sum of squares of the infeasibilities that attracts the iterations for most initial points. We will show that, in these cases, the modified algorithm is more efficient than the original one and that the performance of both algorithms is approximately the same when convergence to the solution of the original problem occurs. In the infeasible cases the performance of the modified algorithm is remarkably better than that of the original Algencan.

4.1 Feasible examples

We considered the family of nonlinear programming problems given by

$$\text{Minimize} \sum_{i=1}^{\frac{n}{2}} 4x_{2i-1}^2 + 2x_{2i-1}x_{2i} + 2x_{2i}^2 - 22x_{2i-1} - 2x_{2i} \quad (21)$$

subject to

$$\left[\left(x_{2i} - x_{2i-1}^2 \right)^2 + 1 \right] (x_{2i-1} - x_{2i} - 18) = 0, \quad (22)$$

Table 1 Feasible problems: Results for $n = 1000$

Initial Point	Original	Modified
$(-10, 10, -10, 10, \ldots)$	22, 64, 0, 0.22	23, 32, 0, 0.15
$(10, 10, 10, 10, \ldots)$	21, 62, 3, 0.26	22, 29, 0, 0.14
$(-10, -10, -10, -10, \ldots)$	22, 62, 0, 0.20	24, 37, 0, 0.16
$(10, -10, 10, -10, \ldots)$	2, 15, 2, 0.10	8, 9, 2, 0.10
$(0, 0, 0, 0, \ldots)$	25, 100, 0, 0.37	26, 29, 0, 0.12

for $i = 1, \ldots, n/2$, and

$$x \in [-10, 10]^n. \tag{23}$$

The objective function is a variable-dimensioned convex quadratic and each constraint is defined by the product of two functions, the first of which is not smaller than 1 for all $x \in I\!R^n$, and the second vanishes in the set $x_{2i-1} - x_{2i} - 18 = 0$. Therefore, the feasible set is a “hidden” polytope, but, due to the presence of the factors $[(x_{2i} - x_{2i-1}^2)^2 + 1]$, iterative solvers only find feasible points when the initial approximation is close to the corner $(10, -10, 10, -10, \ldots)$. For most initial approximations, nonlinear programming solvers tend to find stationary infeasible points. Therefore, with the aim of getting solutions of the hidden quadratic programming problem, it is important that, in the case of infeasible limit, the solver should run as fast as possible. In this way, it could be possible to change initial points with the hope of getting one leading to the true solution.

With this in mind, we ran the ordinary version of Algencan and the modification given in Algorithm 2.1 with (2), (3), and (14). We employed different initial points and the dimensions $n = 1000$ and $n = 10,000$. In (14) we used $\varphi(t) \equiv t$ as forcing function. For each experiment we report a 4-uple (Outer, Inner, Newton, Time). Outer is the number of Algencan (outer) iterations. Inner is the number of iterations executed by the subproblem solver [10]. In some cases, when Algencan judges that the current approximation is close to a solution, some Newton-like iterations are executed in order to speed up the convergence [9]. Finally, Time is the computer time (in seconds) employed by each algorithm. Note that Time should be roughly proportional to Inner.

The results are given in Tables 1 and 2. Only in the case in which the initial point is $(10, -10, 10, -10, \ldots)$ convergence took place to the true minimizer of

Table 2 Feasible problems: Results for $n = 10000$

Initial Point	Original	Modified
$(-10, 10, -10, 10, \ldots)$	20, 60, 0, 1.89	23, 32, 0, 1.16
$(10, 10, 10, 10, \ldots)$	20, 77, 3, 2.82	22, 29, 0, 1.12
$(-10, -10, -10, -10, \ldots)$	20, 58, 0, 1.72	24, 37, 0, 1.34
$(10, -10, 10, -10, \ldots)$	2, 15, 1, 0.74	8, 9, 2, 0.84
$(0, 0, 0, 0, \ldots)$	26, 110, 0, 3.48	27, 29, 0, 1.11

Table 3 Infeasible problems: Results for $n = 1000$

Initial Point	Original	Modified
$(-5, 5, -5, 5, \ldots)$	20 , 72 , 3, 0.06	22 , 29, 0, 0.02
$(5, 5, 5, 5, \ldots)$	20, 70, 2, 0.06	22, 25, 0, 0.03
$(-5, -5, -5, -5, \ldots)$	20, 69, 2, 0.07	22, 30, 0, 0.03
$(5, -5, 5, -5, \ldots)$	22, 88, 3, 0.11	25, 27, 0, 0.03
$(0, 0, 0, 0, \ldots)$	25, 103, 0, 0.09	26, 27, 0, 0.03

the problem. In the other cases, both algorithms converged to stationary points of the infeasibility, as expected. In these cases, the algorithms stopped when the penalty parameter became greater than 10^{20}, generally involving around 20–25 outer iterations. The performance of the modified Algorithm 2.1 was consistently better than the one of Algencan in these cases.

4.2 Infeasible Examples

In what follows we present results for problems in which the feasible region is empty.

The objective function is (21) and the equality constraints are given by (22). However, the box constraints (23) are replaced by

$$x \in [-8, 8]^n. \tag{24}$$

The problem defined by (21), (22) and (24) has empty feasible region. At the global minimizer of the sum of squares of equality constraints subject to the bounds (24) is $(0.5, 0.22176, \ldots)$ and the maximal modulus of infeasibilities is 17.356. On the other hand, the norm of the gradient of the squared norm of infeasibility is $\approx 10^{-20}$. Using all the tested initial points, both the original Algencan and the modified version introduced here converged to the minimizer of infeasibilities. Some examples are given in Tables 3 and 4. Observe that the modified algorithm with adaptive criterion employed much less inner iterations per major iteration (always less than 2) than the original Algencan. Consequently, the computer time employed for converging to the minimizer of infeasibility was considerably reduced.

Table 4 Infeasible problems: Results for $n = 10000$

Initial Point	Original	Modified
$(-5, 5, -5, 5, \ldots)$	19, 416, 1, 3.70	22, 29, 0, 0.24
$(5, 5, 5, 5, \ldots)$	20, 77, 3, 1.07	22, 25, 0, 0.20
$(-5, -5, -5, -5, \ldots)$	20, 83, 2, 0.81	22, 30, 0, 0.24
$(5, -5, 5, -5, \ldots)$	22, 83, 3, 0.88	26, 27, 0, 0.23
$(0, 0, 0, 0, \ldots)$	26, 87, 0, 0.63	27, 27, 0, 0.21

5 Conclusions

Constrained optimization algorithms must be prepared to run with different initial points in order to enhance the chances of convergence to global minimizers, even without using global minimization strategies. One of the conditions for the reliability of this approach is that the nonlinear programming algorithm should detect as fast as possible the case in which convergence to an infeasible point occurs. In the case of the Algencan version of the Augmented Lagrangian method, we showed that, in order to converge to stationary points of the infeasibility, the tolerance for optimality does not need to tend to zero. Exploiting this property, we suggested a specific way to measure infeasibility and to test optimality at each outer iteration of the Augmented Lagrangian algorithm. We proved that, in this way, the convergence properties of Algencan [1] are preserved. In particular, the algorithm converges to stationary points of the sum of squares of infeasibilities, which are generally local minimizers of this measure. Other alternatives for maintaining reasonable levels of optimality tolerances in the presence of possible infeasibility have been proposed in [23]. Preliminary experience shows that computer time is considerably reduced when one employs the new strategy in the case of convergence to infeasible points. In the case of convergence to nonlinear programming solutions the computer time does not change. We conjecture that strategies like this should contribute to increase the reliability of repeated applications of affordable algorithms for finding better local (and perhaps global) minimizers of optimization problems.

Acknowledgements This paper was presented at the International Conference on Scientific Computing (SC2011) held in Sardinia (October 2011) to celebrate the 70th birthday of Claude Brezinski and Sebastiano Seatzu and the 20th anniversary of the journal Numerical Algorithms. Thanks are given to the organizer of the Optimization Cluster, Marcos Raydan, for his invitation to deliver a talk at this meeting.

The authors are indebted to two anonymous referees for insightful comments addressing the organization of the paper and the meaning of the presented results.

References

1. 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)
2. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**, 627–641 (2011)
3. Andreani, R., Haeser, G., Schuverdt, M. L., Silva, P. J. S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. (2011). doi:10.1007/s10107-011-0456-0
4. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new constraint qualifications and applications. Technical report, available in Optimization online (2011)
5. 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)
6. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**, 3533–3554 (2010)

7. Benson, H.Y., Shanno, D.F., Vanderbei, R.J.: Interior-point methods for nonconvex nonlinear programming: filter methods and merit functions. Comput. Optim. Appl. **23**, 257–272 (2002)
8. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont (1999)
9. Birgin, E.G., Martínez, J.M.: Improving ultimate convergence of an Augmented Lagrangian method. Optim Methods Softw **23**, 177–195 (2008)
10. Birgin, E.G., Martínez J.M.: Large-scale active-set box-constrained optimization method with spectral projected gradients. Comput. Optim. Appl. **23**, 101–125 (2002)
11. Byrd, R.H., Curtis, F.E. Nocedal, J.: Infeasibility detection and SQP methods for nonlinear optimization. SIAM J. Optim. **20**, 2281–2299 (2010)
12. Byrd, R.H., Gilbert, J.Ch., Nocedal, J.: A trust region method based on interior point techniques for nonlinear programming. Math. Program. **89**, 149185 (2000)
13. Byrd, R.H., Nocedal, J., Waltz, R.A.: KNITRO: an integrated package for nonlinear optimization. In: Di Pillo, G., Roma, M. (eds.) Large-Scale Nonlinear Optimization, pp. 3559. Springer, New York (2006)
14. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust region methods. In: MPS/SIAM Series on Optimization. SIAM, Philadelphia (2000)
15. Dostál, Z.: Optimal Quadratic Programming Algorithms. Springer, New York (2009)
16. Dostál, Z., Friedlander, A., Santos, S.A.: Augmented Lagrangians with adaptive precision control for quadratic programming with simple bounds and equality constraints. SIAM J. Optim. **13**, 1120–1140 (2003)
17. Fletcher, R., Gould, N.I.M., Leyffer, S., Toint, Ph.L., Wächter, A.: Global convergence of trust-region SQP-filter algorithms for general nonlinear programming. SIAM J. Optim. **13**, 635–659 (2002)
18. Fletcher, R., Leyffer, S.: Nonlinear programming without a penalty function. Math. Program. **91**, 239–269 (2002)
19. Fletcher, R., Leyffer, S., Toint, Ph.L.: On the global convergence of a filter-SQP algorithm. SIAM J. Optim. **13**, 4459 (2002)
20. Hager, W.W.: Analysis and implementation of a dual algorithm for constrained optimization. J. Optim. Theory Appl. **79**, 427–462 (1993)
21. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
22. 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)
23. Prudente, L.F.: Augmented Lagrangian approaches in nonlinear programming. Doctoral dissertation, State University of Campinas (in preparation)
24. Qi, L., Wei, L. Qi, Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
25. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM J. Control Optim. **12**, 268–285 (1974)
26. Shen, C., Xue, W., Pu, D.: A filter SQP algorithm without a feasibility restoration phase. Comput. Appl. Math. **28**, 167–194 (2009)
27. Shen, C., Leyffer, S., Fletcher, R.: A nonmonotone filter method for nonlinear optimization. Comput. Optim. Appl. (2011). doi:10.1007/s10589-011-9430-2
28. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. **106**, 25–57 (2006)