Global Nonlinear Programming with possible infeasibility and finite termination

E. G. Birgin
Department of Computer Science IME-USP, University of São Paulo
Rua do Matão 1010, Cidade Universitária, 05508-090, São Paulo SP, Brazil
e-mail: egbirgin@ime.usp.br

J. M. Martínez
Department of Applied Mathematics, IMECC, University of Campinas
Rua Sérgio Buarque de Holanda 651, Cidade Universitria Zeferino Vaz, 13083-859, Campinas, Brazil
e-mail: martinez@ime.unicamp.br

L. F. Prudente
Department of Applied Mathematics, IMECC, University of Campinas
Rua Sérgio Buarque de Holanda 651, Cidade Universitria Zeferino Vaz, 13083-859, Campinas, Brazil
e-mail: lprudente@ime.unicamp.br

Abstract

In a recent paper, Birgin, Floudas and Martínez introduced a novel Augmented Lagrangian method for global optimization. In their approach, Augmented Lagrangian subproblems are solved using the αBB method and convergence to global minimizers was obtained assuming feasibility of the original problem. In the present research, the algorithm mentioned above will be improved in several crucial aspects. On one hand, feasibility of the problem will not be required. Possible infeasibility will be detected in finite time by the new algorithms and optimal infeasibility results will be proved. On the other hand, finite termination results that guarantee optimality and/or feasibility up to any required precision will be provided.

Key words: deterministic global optimization, Augmented Lagrangians, nonlinear programming.

1 Introduction

Many practical models require to solve global optimization problems involving continuous functions and constraints. Algorithms for solving non-trivial optimization problems are always iterative. Sometimes, for practical purposes, one only needs optimality properties at the limit points. In many other cases, one wishes to find an iterate x^k for which it can be proved that feasibility and optimality hold up to some previously established precision. Moreover, in the case that no feasible point exists, a certificate of infeasibility could also be required. In simple-constrained cases, several well-known algorithms accomplish that purpose. This is the case of the αBB algorithm Adjiman et al. (1996, 1998a,b), Androulakis et al. (1995), that has been used in Birgin et al. (2010) as subproblem solver in the context of an Augmented Lagrangian method.

The numerical algorithm introduced in Birgin et al. (2010) for constrained global optimization was based on the Powell-Hestenes-Rockafellar (PHR) Augmented Lagrangian approach. An implementation in which subproblems were solved by means of the αBB method was described and tested in Birgin et al. (2010). The convergence theory assumed that the nonlinear programming problem is feasible and it was proved that limit points of sequences generated by the algorithm are ε-global minimizers, where ε is a given positive tolerance. However, a test for verifying ε-optimality at an iterate x^k was not provided. As a consequence, the stopping criterion employed in the numerical implementation was not directly related with ε-optimality and relied on heuristic considerations. This gap will be closed in the present paper. On one hand, we will not restrict the range of applications to feasible problems. Infeasible cases may also be handled by the methods analyzed in our present contribution, where we will prove that possible infeasibility can be detected in finite time by means of a computable test. On the other hand, we will introduce a practical stopping criterion that guarantees that, at the approximate solution provided by the algorithm, feasibility holds up to some prescribed tolerance and the objective function value is the optimal one up to tolerance ε.

Global optimization theory also clarifies practical algorithmic properties of "local" optimization algorithms, which use to converge quickly to stationary points. We recall that the Augmented Lagrangian methodology based on the PHR approach has been successfully used for defining practical nonlinear programming algorithms Andreani et al. (2007, 2008), Birgin et al. (2005), Conn et al. (2000). In the local optimization field, which requires near-stationarity (instead of near global optimality) at subproblems, convergence to KKT points was proved using the Constant Positive Linear Dependence constraint qualification Andreani et al. (2005). Convergence results involving sequential optimality conditions that do not need constraint qualifications at all were presented in Andreani et al. (2010, 2011).

The Algencan code, available in `www.ime.usp.br/~egbirgin/` and based on the theory presented in Andreani et al. (2007), has been improved several times in the last few years and, in practice, has been shown to converge to global minimizers more frequently than other Nonlinear Programming solvers. There exist many global optimization techniques for nonlinear programming. The main appeal of the Augmented Lagrangian approach in this context is that the structure of this method makes it possible to take advantage of global optimization algorithms for simpler problems. In Birgin et al. (2010) and the present paper we exploit the ability of αBB for solving linearly constrained global optimization problems, which has been corroborated in many applied papers. In order to take advantage of the αBB potentialities, Augmented Lagrangian subproblems are "over-restricted" by means of linear constraints that simplify subproblem resolutions and do not affect successful search of global minimizers. Because of the necessity of dealing with infeasible problems, the definition of the additional constraints has been modified in the present contribution with respect to the one given in Birgin et al. (2010).

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.

2 Algorithm

The problem considered in this paper is:

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

where $h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p, f : \mathbb{R}^n \to \mathbb{R}$ are continuous and $\Omega \subset \mathbb{R}^n$ is compact. In general, Ω is defined by "easy" constraints such as linear constraints and box constraints. Since all the iterates x^k generated by our methods will belong to Ω, the constraints related with this set may be called "non-relaxable" in the sense of Audet and Dennis (2009).

The Augmented Lagrangian function will be defined by:

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

for all $x \in \Omega, \rho > 0, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p_+$.

At each (outer) iteration, the algorithm considered in this section minimizes the Augmented Lagrangian, with precision ε_k, on the set $\Omega \cap P_k$, where $P_k \subseteq \mathbb{R}^n$ is built in order to facilitate the work of a subproblem solver like αBB. The assumptions required for the tolerances $\{\varepsilon_k\}$ and the auxiliary sets $\{P_k\}$ are given below.

Assumption A1. The sequence of positive tolerances $\{\varepsilon_k\}$ is bounded.

Assumption A2. The sets P_k are closed and the set of global minimizers of (1) is contained in P_k for all $k \in \mathbb{N}$.

The sequence $\{\varepsilon_k\}$ may be defined in an external or an internal way, in different implementations. In the external case, the sequence is given as a parameter of the algorithm. If one decides for an internal definition, each tolerance ε_{k+1} is defined only after the computation of x^k as a result of the process evolution. Except in the case that one of the sets $\Omega \cap P_k$ is found to be empty, we will consider that the algorithm defined here generates an infinite sequence $\{x^k\}$ and we will prove theoretical properties of this sequence. Later, we will see that the generated sequence may be stopped satisfying stopping criteria that guarantee feasibility and optimality, or, perhaps, infeasibility. Observe that the existence of global minimizers is not guaranteed at all, since the feasible set could be empty. In this case Assumption A2 is trivially satisfied. In Birgin et al. (2010) the existence of a global minimizer was an assumption on the problem and the sets P_k were assumed to contain at least one global minimizer.

Algorithm 2.1

Let $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $0 < \tau < 1$. Let $\lambda_i^1 \in [\lambda_{\min}, \lambda_{\max}], i = 1, \ldots, m$, $\mu_i^1 \in [0, \mu_{\max}], i = 1, \ldots, p$, and $\rho_1 > 0$. Initialize $k \leftarrow 1$.

Step 1.1 If $\Omega \cap P_k$ is found to be empty, stop the execution of the algorithm.

Step 1.2 Find $x^k \in \Omega \cap P_k$ such that:

$$L_{\rho_k}(x^k, \lambda^k, \mu^k) \le L_{\rho_k}(x, \lambda^k, \mu^k) + \varepsilon_k \tag{3}$$

for all $x \in \Omega \cap P_k$.

Step 2. Define

$$V_i^k = \min\left\{ -g_i(x^k), \frac{\mu_i^k}{\rho_k} \right\}, i = 1, \dots, p.$$

If $k = 1$ or

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

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

Step 3. Compute $\lambda_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}], i = 1, \dots, m$ and $\mu_i^{k+1} \in [0, \mu_{\max}], i = 1, \dots, p$. Set $k \leftarrow k + 1$ and go to Step 1.

Algorithm 2.1 has been presented above without a stopping criterion, except in the case in which emptiness of $\Omega \cap P_k$ is detected. Therefore, in this ideal form, the algorithm generally generates an infinite sequence. The solvability of the subproblems (3) is guaranteed, if $\Omega \cap P_k$ is a bounded polytope, employing global optimization algorithms as αBB.

Although infinite-sequence properties do not satisfy our requirements of getting feasibility and optimality certificates in finite time, results concerning the behavior of the infinite sequence potentially generated by the algorithm help to understand its practical properties.

Theorem 2.1. *Assume that $\{x^k\}$ is an infinite sequence generated by Algorithm 2.1. Let $K \underset{\infty}{\subset} \mathbb{N}$ and $x^* \in \Omega$ be such that $\lim_{k \in K} x^k = x^*$. (Such subsequence exists since Ω is compact.) Then, for all $z \in \Omega$ such that z is a limit point of $\{z^k\}_{k \in K}$, with $z^k \in \Omega \cap P_k$ for all $k \in K$, we have:*

$$\|h(x^*)\|^2 + \|g(x^*)_+\| \le \|h(z)\|^2 + \|g(z)_+\|^2. \tag{5}$$

In particular, if the problem (1) is feasible, every limit point of an infinite sequence generated by Algorithm 2.1 is feasible.

Proof. In the case that $\{\rho_k\}$ is bounded, we have, by (4), that $\lim_{k\to\infty} \|h(x^k)\| + \|g(x^k)_+\| = 0$. Taking limits for $k \in K$ this implies that $\|h(x^*)\| + \|g(x^*)_+\| = 0$, which trivially implies (5).

Consider now the case in which $\rho_k \to \infty$. Let $z \in \Omega$, $K_1 \underset{\infty}{\subset} K$ be such that

$$\lim_{k \in K_1} z^k = z,$$

with $z^k \in \Omega \cap P_k$ for all $k \in K_1$. By (3), we have:

$$L_{\rho_k}(x^k, \lambda^k, \mu^k) \le L_{\rho_k}(z^k, \lambda^k, \mu^k) + \varepsilon_k$$

for all $k \in K_1$. This implies that, for all $k \in K_1$,

$$\frac{\rho_k}{2}\left[\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] \le \frac{\rho_k}{2}\left[\left\|h(z^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(z^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] + \varepsilon_k + f(z^k) - f(x^k).$$

Therefore,

$$\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2 \le \left[\left\|h(z^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(z^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] + \frac{2(\varepsilon_k + f(z^k) - f(x^k))}{\rho_k}.$$

Since $\{\varepsilon_k\}$ is bounded, ρ_k tends to infinity, and Ω is compact, the inequality (5) follows, taking limits for $k \in K_1$, by the continuity of f, h, and g. $\square$

In the case that $\Omega \subseteq P_k$ for all k, Theorem 2.1 says that any limit point is a global minimizer of the infeasibility measure $\|h(x)\|^2 + \|g(x)_+\|^2$ onto Ω. In particular, if the problem is feasible, every limit point is feasible. It is interesting to observe that the tolerances ε_k do not necessarily tend to zero, in order to obtain the thesis of Theorem 2.1. Moreover, although in the algorithm we assume that λ^k and μ^k are bounded, in the proof we only need that the quotients λ^k/ρ_k and μ^k/ρ_k tend to zero as ρ_k tends to infinity.

In the following theorem we prove that infeasibility can be detected in finite time. Let us define, for all $k \in I\!N$, $c_k > 0$ by:

$$|f(z) - f(x^k)| \le c_k \quad \text{for all} \quad z \in \Omega \cap P_k. \tag{6}$$

Note that c_k may be computed using interval calculations as in the αBB algorithm. Clearly, since f is continuous and Ω is bounded, the sequence $\{c_k\}$ may be chosen to be bounded.

Theorem 2.2. *Assume that $\{x^k\}$ is a sequence generated by Algorithm 2.1 and, for all $k \in I\!N$, the set $\Omega \cap P_k$ is non-empty. Then, the problem (1) is infeasible if and only if there exists $k \in I\!N$ such that*

$$\frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\frac{\mu^k}{\rho_k}\right\|^2\right] - \frac{\rho_k}{2}\left[\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] + \varepsilon_k < -c_k. \tag{7}$$

Proof. Suppose that the feasible region of (1) is non-empty. Then there exists a global minimizer z such that $z \in \Omega \cap P_k$ for all $k \in I\!N$. Therefore,

$$f(x^k) + \frac{\rho_k}{2}\left[\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] \le f(z) + \frac{\rho_k}{2}\left[\left\|h(z) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(z) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] + \varepsilon_k.$$

Thus,

$$\frac{\rho_k}{2}\left[\left\|h(z) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(z) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] - \frac{\rho_k}{2}\left[\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] \ge f(x^k) - f(z) - \varepsilon_k. \tag{8}$$

Since $h(z) = 0$ and $g(z) \le 0$, we have:

$$\left\|h(z) + \frac{\lambda^k}{\rho_k}\right\|^2 = \left\|\frac{\lambda^k}{\rho_k}\right\|^2 \text{ and } \left\|\left(g(z) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2 \le \left\|\frac{\mu^k}{\rho_k}\right\|^2.$$

Then, by (8),

$$\frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\frac{\mu^k}{\rho_k}\right\|^2\right] - \frac{\rho_k}{2}\left[\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] \ge f(x^k) - f(z) - \varepsilon_k.$$

Therefore, by (6),

$$\frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\frac{\mu^k}{\rho_k}\right\|^2\right] - \frac{\rho_k}{2}\left[\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] + \varepsilon_k \ge -c_k$$

for all $k \in I\!N$. This means that the infeasibility test (7) fails to be fulfilled for all $k \in I\!N$.

Reciprocally, suppose that problem (1) is infeasible. In this case ρ_k tends to infinity. This implies that the sequence $\{x^k\}$ admits an infeasible limit point $x^* \in \Omega$. So, for some subsequence, the quantity $\|h(x^k) + \lambda^k/\rho_k\|^2 + \|(g(x^k) + \mu^k/\rho_k)_+\|^2$ is bounded away from zero. Since

$$-\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\frac{\mu^k}{\rho_k}\right\|^2\right] - \frac{2(\varepsilon_k + c_k)}{\rho_k}$$

tends to zero, it turns out that, for k large enough, the test (7) is fulfilled. □

In the following theorem we prove another asymptotic convergence result, this time connected with optimality, instead of feasibility.

Theorem 2.3. *Assume that* $\{x^k\}$ *is an infinite sequence generated by Algorithm 2.1,* $\lim_{k\to\infty}\varepsilon_k = 0$, *and problem (1) is feasible. Then, every limit point of* $\{x^k\}$ *is a global solution of (1).*

Proof. Let $K \underset{\infty}{\subset} \mathbb{N}$ and $x^* \in \Omega$ be such that $\lim_{k\in K} x^k = x^*$. Since the feasible set is non-empty and compact, problem (1) admits a global minimizer $z \in \Omega$. By Assumption A2, $z \in P_k$ for all $k \in \mathbb{N}$. We consider two cases: $\rho_k \to \infty$ and $\{\rho_k\}$ bounded.
Case 1 ($\rho_k \to \infty$): By the definition of the algorithm:

$$f(x^k)+\frac{\rho_k}{2}\left[\left\|h(x^k)+\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\left(g(x^k)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\right]\le f(z)+\frac{\rho_k}{2}\left[\left\|h(z)+\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\left(g(z)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\right]+\varepsilon_k \quad (9)$$

for all $k \in \mathbb{N}$. Since $h(z) = 0$ and $g(z) \le 0$, we have:

$$\left\|h(z)+\frac{\lambda^k}{\rho_k}\right\|^2=\left\|\frac{\lambda^k}{\rho_k}\right\|^2 \text{ and } \left\|\left(g(z)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\le\left\|\frac{\mu^k}{\rho_k}\right\|^2.$$

Therefore, by (9),

$$f(x^k)\le f(x^k)+\frac{\rho_k}{2}\left[\left\|h(x^k)+\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\left(g(x^k)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\right]\le f(z)+\frac{\|\lambda^k\|^2}{2\rho_k}+\frac{\|\mu^k\|^2}{2\rho_k}+\varepsilon_k.$$

Taking limits for $k \in K$, using that $\lim_{k\in K}\|\lambda^k\|^2/\rho_k = \lim_{k\in K}\|\mu^k\|^2/\rho_k = 0$, and $\lim_{k\in K}\varepsilon_k = 0$, by the continuity of f and the convergence of x^k, we get:

$$f(x^*)\le f(z).$$

Since z is a global minimizer, it turns out that x^* is a global minimizer, as we wanted to prove.

Case 2 ($\{\rho_k\}$ bounded): In this case, we have that $\rho_k = \rho_{k_0}$ for all $k \ge k_0$. Therefore, by the definition of Algorithm 2.1, we have:

$$f(x^k)+\frac{\rho_{k_0}}{2}\left[\left\|h(x^k)+\frac{\lambda^k}{\rho_{k_0}}\right\|^2+\left\|\left(g(x^k)+\frac{\mu^k}{\rho_{k_0}}\right)_+\right\|^2\right]\le f(z)+\frac{\rho_{k_0}}{2}\left[\left\|h(z)+\frac{\lambda^k}{\rho_{k_0}}\right\|^2+\left\|\left(g(z)+\frac{\mu^k}{\rho_{k_0}}\right)_+\right\|^2\right]+\varepsilon_k$$

for all $k \ge k_0$. Since $g(z) \le 0$ and $\mu^k/\rho_{k_0} \ge 0$,

$$\left\|\left(g(z)+\frac{\mu^k}{\rho_{k_0}}\right)_+\right\|^2\le\left\|\frac{\mu^k}{\rho_{k_0}}\right\|^2.$$

Thus, since $h(z) = 0$,

$$f(x^k)+\frac{\rho_{k_0}}{2}\left[\left\|h(x^k)+\frac{\lambda^k}{\rho_{k_0}}\right\|^2+\left\|\left(g(x^k)+\frac{\mu^k}{\rho_{k_0}}\right)_+\right\|^2\right]\le f(z)+\frac{\rho_{k_0}}{2}\left[\left\|\frac{\lambda^k}{\rho_{k_0}}\right\|^2+\left\|\frac{\mu^k}{\rho_{k_0}}\right\|^2\right]+\varepsilon_k \quad (10)$$

for all $k \ge k_0$. Let us take now $\varepsilon > 0$ arbitrarily small. Suppose, for a moment, that $g_i(x^*) < 0$. Since $\lim_{k\to\infty}\min\{-g_i(x^k), \mu_i^k/\rho_{k_0}\} = 0$, we have that

$$\lim_{k\in K}\mu_i^k/\rho_{k_0} = 0. \quad (11)$$

This implies that $(g_i(x^k)+\mu_i^k/\rho_{k_0})_+ = 0$ for $k \in K$ large enough. Therefore, for $k \in K$ large enough, $\sum_{i=1}^p (g_i(x^k)+\mu_i^k/\rho_{k_0})_+^2 = \sum_{g_i(x^*)=0}(g_i(x^k)+\mu_i^k/\rho_{k_0})_+^2$. Thus, by (10), for $k \in K$ large enough we have:

$$f(x^k)+\frac{\rho_{k_0}}{2}\left[\sum_{i=1}^m\left(h_i(x^k)+\frac{\lambda_i^k}{\rho_{k_0}}\right)^2+\sum_{g_i(x^*)=0}\left(g_i(x^k)+\frac{\mu_i^k}{\rho_{k_0}}\right)_+^2\right]$$

$$\leq f(z)+\frac{\rho_{k_0}}{2}\left[\sum_{i=1}^m\left(\frac{\lambda_i^k}{\rho_{k_0}}\right)^2+\sum_{g_i(x^*)=0}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2+\sum_{g_i(x^*)<0}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2\right]+\varepsilon_k.$$

By (11), we deduce that, for $k \in K$ large enough,

$$f(x^k)+\frac{\rho_{k_0}}{2}\left[\sum_{i=1}^m\left(h_i(x^k)+\frac{\lambda_i^k}{\rho_{k_0}}\right)^2+\sum_{g_i(x^*)=0}\left(g_i(x^k)+\frac{\mu_i^k}{\rho_{k_0}}\right)_+^2\right]$$

$$\leq f(z)+\frac{\rho_{k_0}}{2}\left[\sum_{i=1}^m\left(\frac{\lambda_i^k}{\rho_{k_0}}\right)^2+\sum_{g_i(x^*)=0}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2\right]+\varepsilon_k+\varepsilon. \tag{12}$$

For $k \in K$ large enough, by the boundedness of λ_i^k/ρ_{k_0} and the fact that $h(x^k)\to 0$, we have that

$$\frac{\rho_{k_0}}{2}\sum_{i=1}^m\left[h_i(x^k)^2+2h_i(x^k)\frac{\lambda_i^k}{\rho_{k_0}}\right]\geq -\varepsilon.$$

Therefore, by (12),

$$f(x^k)+\frac{\rho_{k_0}}{2}\left[\sum_{i=1}^m\left(\frac{\lambda_i^k}{\rho_{k_0}}\right)^2+\sum_{g_i(x^*)=0}\left(g_i(x^k)+\frac{\mu_i^k}{\rho_{k_0}}\right)_+^2\right]\leq f(z)+\frac{\rho_{k_0}}{2}\left[\sum_{i=1}^m\left(\frac{\lambda_i^k}{\rho_{k_0}}\right)^2+\sum_{g_i(x^*)=0}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2\right]+\varepsilon_k+2\varepsilon.$$

Thus, there exists $k_1 \geq k_0$ such that for all $k \in K$ such that $k \geq k_1$, we have that

$$f(x^k)+\frac{\rho_{k_0}}{2}\left[\sum_{g_i(x^*)=0}\left(g_i(x^k)+\frac{\mu_i^k}{\rho_{k_0}}\right)_+^2\right]\leq f(z)+\frac{\rho_{k_0}}{2}\left[\sum_{g_i(x^*)=0}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2\right]+\varepsilon_k+2\varepsilon. \tag{13}$$

Define

$$I=\{i\in\{1,\dots,p\}\mid g_i(x^*)=0\}$$

and

$$K_1=\{k\in K\mid k\geq k_1\}.$$

For each $i \in I$, we define

$$K_+(i)=\{k\in K_1\mid g_i(x^k)+\mu_i^k/\rho_{k_0}\geq 0\}$$

and

$$K_-(i)=\{k\in K_1\mid g_i(x^k)+\mu_i^k/\rho_{k_0}< 0\}.$$

Obviously, for all $i \in I$, $K_1 = K_+(i)\cup K_-(i)$. Let us fix $i \in I$. For k large enough, since $g_i(x^*)=0$, by the continuity of g_i and the boundedness of μ_i^k/ρ_{k_0}, we have that:

$$\frac{\rho_{k_0}}{2}\left(g_i(x^k)^2+\frac{2g_i(x^k)\mu_i^k}{\rho_{k_0}}\right)\geq -\varepsilon.$$

Therefore,

$$\frac{\rho_{k_0}}{2}\left[g_i(x^k)^2+\frac{2g_i(x^k)\mu_i^k}{\rho_{k_0}}+\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2\right]\geq\frac{\rho_{k_0}}{2}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2-\varepsilon.$$

Thus, for $k \in K_+(i)$ large enough,

$$\frac{\rho_{k_0}}{2}\left(g_i(x^k)+\frac{\mu_i^k}{\rho_{k_0}}\right)_+^2 \geq \frac{\rho_{k_0}}{2}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2 - \varepsilon. \tag{14}$$

Now, if $k \in K_-(i)$, we have that $-g_i(x^k) > \mu_i^k/\rho_{k_0}$. So, since $g_i(x^k)$ tends to zero, for $k \in K_-(i)$ large enough we have that $(\rho_{k_0}/2)(\mu_i^k/\rho_{k_0})^2 \leq \varepsilon$. Therefore,

$$\frac{\rho_{k_0}}{2}\left(g_i(x^k)+\frac{\mu_i^k}{\rho_{k_0}}\right)_+^2 = 0 \geq \frac{\rho_{k_0}}{2}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2 - \varepsilon. \tag{15}$$

Combining (14) and (15) and taking k large enough, we obtain:

$$f(x^k)+\frac{\rho_{k_0}}{2}\left[\sum_{g_i(x^*)=0}\left(g_i(x^k)+\frac{\mu_i^k}{\rho_{k_0}}\right)_+^2\right] \geq f(x^k)+\frac{\rho_{k_0}}{2}\left[\sum_{g_i(x^*)=0}\left(\frac{\mu_i^k}{\rho_{k_0}}\right)^2\right] - p\varepsilon. \tag{16}$$

Then, by (13) and (16), for $k \in K$ large enough we have that

$$f(x^k) \leq f(z)+\varepsilon_k+(2+p)\varepsilon.$$

Since $\lim_{k\in K}\varepsilon_k = 0$ and ε is arbitrarily small, it turns out that $\lim_{k\in K} f(x^k) = f(z)$ and, so, x^* is a global minimizer as we wanted to prove. □

Theorem 2.4. *Assume that* $\{x^k\}$ *is an infinite sequence generated by Algorithm 2.1. Let* $\varepsilon \in \mathbb{R}$ *(note that* ε *may be negative) and* $k \in \mathbb{N}$ *such that*

$$\frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\frac{\mu^k}{\rho_k}\right\|^2\right] - \frac{\rho_k}{2}\left[\left\|h(x^k)+\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\left(g(x^k)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] \leq \varepsilon. \tag{17}$$

Then

$$f(x^k) \leq f(z)+\varepsilon+\varepsilon_k, \tag{18}$$

for all feasible point z.

Proof. Let $z \in \Omega$ be a feasible point of (1). By the definition of Algorithm 2.1, we have that

$$f(x^k)+\frac{\rho_k}{2}\left[\left\|h(x^k)+\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\left(g(x^k)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] \leq f(z)+\frac{\rho_k}{2}\left[\left\|h(z)+\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\left(g(z)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\right]+\varepsilon_k \tag{19}$$

for all $k \in \mathbb{N}$. Moreover, since

$$\left\|h(z)+\frac{\lambda^k}{\rho_k}\right\|^2 = \left\|\frac{\lambda^k}{\rho_k}\right\|^2 \text{ and } \left\|\left(g(z)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2 \leq \left\|\frac{\mu^k}{\rho_k}\right\|^2, \tag{20}$$

we obtain:

$$f(x^k)+\frac{\rho_k}{2}\left[\left\|h(x^k)+\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\left(g(x^k)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] \leq f(z)+\frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\frac{\mu^k}{\rho_k}\right\|^2\right]+\varepsilon_k. \tag{21}$$

Assuming that (17) takes place, we have

$$f(x^k)+\frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\frac{\mu^k}{\rho_k}\right\|^2\right]-\varepsilon \leq f(x^k)+\frac{\rho_k}{2}\left[\left\|h(x^k)+\frac{\lambda^k}{\rho_k}\right\|^2+\left\|\left(g(x^k)+\frac{\mu^k}{\rho_k}\right)_+\right\|^2\right]. \tag{22}$$

Hence, by (22) and (21), we have

$$f(x^k) + \frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\frac{\mu^k}{\rho_k}\right\|^2\right] - \varepsilon \le f(z) + \frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\frac{\mu^k}{\rho_k}\right\|^2\right] + \varepsilon_k. \tag{23}$$

Simplifying the expression (23), we obtain:

$$f(x^k) \le f(z) + \varepsilon + \varepsilon_k,$$

as we wanted to prove. □

Theorem 2.5. *Assume that $\{x^k\}$ is an infinite sequence generated by Algorithm 2.1. Suppose that (1) is feasible and $\lim_{k\to\infty}\varepsilon_k = 0$. Let ε be an arbitrary positive number. Then, there exists $k \in I\!N$ such that*

$$\frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\frac{\mu^k}{\rho_k}\right\|^2\right] - \frac{\rho_k}{2}\left[\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right] \le \varepsilon. \tag{24}$$

Proof. By the compactness of Ω, there exists $K \underset{\infty}{\subset} I\!N$ and $x^* \in \Omega$ such that $\lim_{k\in K} x^k = x^*$ and, by Theorem 2.1, x^* is feasible. Suppose that ρ_k tends to infinity. Note that the left-hand side of (24) is bounded by $(\|\lambda^k\|^2 + \|\mu^k\|^2)/(2\rho_k)$ that tends to zero, by the boundedness of λ^k and μ^k. Thus, we obtain (24) for k large enough.

Consider now the case in which $\{\rho_k\}$ is bounded. For all $i = 1,\dots,m$ we have that $(\rho_k/2)[h_i(x^k) + \lambda_i^k/\rho_k]^2 = (\rho_k/2)[h_i(x^k)^2 + 2h_i(x^k)\lambda_i^k/\rho_k + (\lambda_i^k/\rho_k)^2]$. Since ρ_k is bounded, λ^k is bounded, and $h_i(x^k) \to 0$ there exists $k_0(i) \in K$ such that $(\rho_k/2)[h_i(x^k) + \lambda_i^k/\rho_k]^2 \ge (\rho_k/2)(\lambda_i^k/\rho_k)^2 - \varepsilon/m$ for all $k \in K$, $k \ge k_0(i)$. Taking $k_0 = \max\{k_0(i)\}$ we obtain that, for all $k \in K, k \ge k_0$, $i = 1,\dots,m$,

$$\frac{\rho_k}{2}\left(\frac{\lambda_i^k}{\rho_k}\right)^2 - \frac{\rho_k}{2}\left(h_i(x^k) + \frac{\lambda_i^k}{\rho_k}\right)^2 \le \frac{\varepsilon}{m}. \tag{25}$$

Assume that $g_i(x^*) < 0$. Then, as in Case 2 of the proof of Theorem 2.3, since $\lim_{k\to\infty}\min\{-g_i(x^k), \mu_i^k/\rho_k\} = 0$, we have that $\lim_{k\in K}\mu_i^k/\rho_k = 0$. Thus, there exists $k_1(i) \ge k_0$ such that $(g_i(x^k) + \mu_i^k/\rho_k)_+ = 0$ for all $k \in K, k \ge k_1(i)$. Therefore, since $\mu_i^k/\rho_k \to 0$, there exists $k_2(i) \ge k_1(i)$ such that

$$\frac{\rho_k}{2}\left(\frac{\mu_i^k}{\rho_k}\right)^2 - \frac{\rho_k}{2}\left(g_i(x^k) + \frac{\mu_i^k}{\rho_k}\right)_+^2 \le \frac{\varepsilon}{p} \tag{26}$$

for all $k \in K, k \ge k_2(i)$. Taking $k_2 = \max\{k_2(i)\}$, we obtain that (26) holds for all $k \in K, k \ge k_2$ whenever $g_i(x^*) < 0$.

Now, as in the proof of Theorem 2.3, define

$$I = \{i \in \{1,\dots,p\} \mid g_i(x^*) = 0\}$$

and

$$K_1 = \{k \in K \mid k \ge k_2\}.$$

For each $i \in I$, we define

$$K_+(i) = \{k \in K_1 \mid g_i(x^k) + \mu_i^k/\rho_k \ge 0\}$$

and

$$K_-(i) = \{k \in K_1 \mid g_i(x^k) + \mu_i^k/\rho_k < 0\}.$$

Let us fix $i \in I$. For k large enough, since $g_i(x^*) = 0$, by the continuity of g_i and the boundedness of μ_i^k/ρ_k, we have that:

$$\frac{\rho_k}{2}\left(g_i(x^k)^2 + \frac{2g_i(x^k)\mu_i^k}{\rho_k}\right) \geq -\frac{\varepsilon}{p}.$$

Therefore,

$$\frac{\rho_k}{2}\left[g_i(x^k)^2 + \frac{2g_i(x^k)\mu_i^k}{\rho_k} + \left(\frac{\mu_i^k}{\rho_k}\right)^2\right] \geq \frac{\rho_k}{2}\left(\frac{\mu_i^k}{\rho_k}\right)^2 - \frac{\varepsilon}{p}.$$

Thus, for $k \in K_+(i)$ large enough,

$$\frac{\rho_k}{2}\left(\frac{\mu_i^k}{\rho_k}\right)^2 - \frac{\rho_k}{2}\left(g_i(x^k) + \frac{\mu_i^k}{\rho_k}\right)_+^2 \leq \frac{\varepsilon}{p}. \tag{27}$$

Now, if $k \in K_-(i)$, we have that $-g_i(x^k) > \mu_i^k/\rho_k$. So, since $g_i(x^k)$ tends to zero, for $k \in K_-(i)$ large enough we have that $(\rho_k/2)(\mu_i^k/\rho_k)^2 \leq \varepsilon/p$. Therefore,

$$\frac{\rho_k}{2}\left(\frac{\mu_i^k}{\rho_k}\right)^2 - \frac{\rho_k}{2}\left(g_i(x^k) + \frac{\mu_i^k}{\rho_k}\right)_+^2 \leq \frac{\varepsilon}{p}. \tag{28}$$

.

By (26), (27), and (28),

$$\frac{\rho_k}{2}\left(\frac{\mu_i^k}{\rho_k}\right)^2 - \frac{\rho_k}{2}\left(g_i(x^k) + \frac{\mu_i^k}{\rho_k}\right)_+^2 \leq \frac{\varepsilon}{p} \tag{29}$$

for all $i = 1, \ldots, p$.

Taking the summation for $i = 1, \ldots, m$ in (25) and for $i = 1, \ldots, p$ in (29) we obtain the desired result. □

Due to the results proved above, we are able to define a variation of Algorithm 2.1, for which we can guarantee finite termination with certificates of infeasibility or optimality up to given precisions. For defining Algorithm 2.2, we assume that $\varepsilon_{feas} > 0$ and $\varepsilon_{\text{opt}} > 0$ are user-given tolerances for feasibility and optimality respectively. On the other hand, we will maintain Assumptions A1 and A2, which concern boundedness of $\{\varepsilon_k\}$ and the inclusion property for the sets P_k.

Algorithm 2.2

Let $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $0 < \tau < 1$. Let $\lambda_i^1 \in [\lambda_{\min}, \lambda_{\max}], i = 1, \ldots, m$, $\mu_i^1 \in [0, \mu_{\max}], i = 1, \ldots, p$, and $\rho_1 > 0$. Assume that $\{\bar{\varepsilon}_k\}$ is a bounded positive sequence and initialize $k \leftarrow 1$.

Step 1 Solve the subproblem

Solve, using global optimization on the set $\Omega \cap P_k$, the subproblem

$$\text{Minimize } L_{\rho_k}(x, \lambda^k, \mu^k) \quad \text{subject to} \quad x \in \Omega \cap P_k. \tag{30}$$

If, in the process of solving (30), the set $\Omega \cap P_k$ is detected to be empty, stop the execution of Algorithm 2.2 declaring **Infeasibility**. Otherwise, define $x^k \in \Omega \cap P_k$ as an approximate solution of (30) that satisfies (3) for some $\varepsilon_k \leq \bar{\varepsilon}_k$.

Step 2 Test Infeasibility

Compute $c_k > 0$ such that $|f(x^k) - f(z)| \leq c_k$ for all $z \in \Omega \cap P_k$ and define

$$\gamma_k = \frac{\rho_k}{2}\left[\left\|\frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\frac{\mu^k}{\rho_k}\right\|^2\right] - \frac{\rho_k}{2}\left[\left\|h(x^k) + \frac{\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \frac{\mu^k}{\rho_k}\right)_+\right\|^2\right].$$

If

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

stop the execution of the algorithm declaring **Infeasibility**.

Step 3 Test Feasibility and optimality

If

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

stop the execution of the algorithm declaring **Solution found**.

Step 4 Update penalty parameter

Define

$$V_i^k = \min\left\{-g_i(x^k), \frac{\mu_i^k}{\rho_k}\right\}, i = 1, \dots, p.$$

If $k = 1$ or

$$\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\},$$

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

Step 5. Update multipliers

Compute $\lambda_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}], i = 1, \dots, m$ and $\mu_i^{k+1} \in [0, \mu_{\max}], i = 1, \dots, p$. Set $k \leftarrow k+1$ and go to Step 1.

Theorem 2.6. *Assume that Algorithm 2.1 is executed with the condition that,* $\lim_{k\to\infty} \varepsilon_k = 0$. *Then, the execution finishes in a finite number of iterations with one of the following diagnostics:*

1. **Infeasibility**, *which means that, guaranteedly, no feasible point of (1) exists;*

2. **Solution found**, *in the case that the final point* x^k *is guaranteed to satisfy*

$$\|h(x^k)\| + \|g(x^k)_+\| \leq \varepsilon_{feas}$$

and

$$f(x^k) \leq f(z) + \varepsilon_{opt}$$

for all $z \in \Omega$ *such that* $h(z) = 0$ *and* $g(z) \leq 0$.

Proof. The proof follows straightforwardly from Theorems 2.2, 2.4, and 2.5. □

Numerical experiments showing how the new algorithm and results are related to practical computations can be found in Birgin et al. (2012), as well as a variation of the method presented here that allows the Augmented Lagrangian subproblems to be solved without requiring unnecessary potentially high precisions in the intermediate steps of the method.

References

C. S. Adjiman, I. P. Androulakis, C. D. Maranas, and C. A. Floudas. A global optimization method α-bb for process design. *Computers and Chemical Engineering*, 20:419–424, 1996.

C. S. Adjiman, I. P. Androulakis, and C. A. Floudas. A global optimization method, α-bb, for general twice-differentiable constrained nlps – ii. implementation and computational results. *Computers and Chemical Engineering*, 22:1159–1179, 1998a.

C. S. Adjiman, S. Dallwig, C. A. Floudas, and A. Neumaier. A global optimization method, α-bb, for general twice-differentiable constrained nlps – i. theoretical advances. *Computers and Chemical Engineering*, 22:1137–1158, 1998b.

R. Andreani, J. M. Martínez, and M. L. Schuverdt. On the relation between the constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125:473–485, 2005.

R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18:1286–1309, 2007.

R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 111:5–32, 2008.

R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20:139–162, 2010.

R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization (Print)*, 60:627–641, 2011.

I. P. Androulakis, C. D. Maranas, and C. A. Floudas. α-bb: A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization*, 7:337–363, 1995.

C. Audet and J. E. Dennis. A progressive barrier for derivative-free nonlinear programming. *SIAM Journal of Optimization*, 20:445–472, 2009.

E. G. Birgin, R. Castillo, and J. M. Martínez. Numerical comparison of augmented lagrangian algorithms for nonconvex problems. *Computational Optimization and Applications*, 31:31–56, 2005.

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

E. G. Birgin, , J. M. Martínez, and L. F. Prudente. Global nonlinear programming with possible infeasibility and finite termination. Technical report, Department of Applied Mathematics, University of Campinas, 2012.

A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust Region Methods*. MOS-SIAM Series on Optimization, SIAM, Philadephia, PA, 2000.