Optimality properties of an Augmented Lagrangian method on infeasible problems*

E. G. Birgin† J. M. Martínez ‡ L. F. Prudente §

March 7, 2014

Abstract

Sometimes, the feasible set of an optimization problem that one aims to solve using a Nonlinear Programming algorithm is empty. In this case, two characteristics of the algorithm are desirable. On the one hand, the algorithm should converge to a minimizer of some infeasibility measure. On the other hand, one may wish to find a point with minimal infeasibility for which some optimality condition, with respect to the objective function, holds. Ideally, the algorithm should converge to a minimizer of the objective function subject to minimal infeasibility. In this paper the behavior of an Augmented Lagrangian algorithm with respect to those properties will be studied.

Key words: Nonlinear Programming, infeasible domains, Augmented Lagrangians, algorithms, numerical experiments.

1 Introduction

We wish to consider optimization problems in which the feasible region may be empty. Rigorously speaking, such problems have no solutions at all and, so, a desirable property of algorithms is to detect infeasibility as soon as possible [11, 19, 21, 22, 23, 30, 31, 46, 32, 39, 43, 47, 48]. However, in many practical situations one is interested in optimizing the function, admitting some level of infeasibility. For example, when the constraints involve measurements or modeling errors, their complete accurate satisfaction may be irrelevant and minimizers subject to moderate infeasibility may be useful.

*This work was supported by PRONEX-CNPq/FAPERJ E-26/111.449/2010-APQ1, FAPESP 2010/10133-0, 2013/05475-7, and 2013/07375-0, Capes/MES-Cuba 226/2012, Capes/Procad NF 21/2009, and CNPq 474160/2013-0.

†Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil. e-mail: egbirgin@ime.usp.br

‡Department of Applied Mathematics, Institute of Mathematics, Statistics, and Scientific Computing, University of Campinas, Campinas, SP, Brazil. e-mail: martinez@ime.unicamp.br

§Institute of Mathematics and Statistics, Federal University of Goiás, Goiânia, GO, Brazil. e-mail: lfprudente@ufg.br

This motivated us to study the behavior of optimization algorithms in the case of infeasibility. In recent papers [19, 39] we considered the problem of early detection of this fact, without concerns about the quality of the limit point obtained in terms of optimality. In the present paper we will adopt the point of view that it is relevant to distinguish between different infeasible points according to their objective function values. In this sense, it is important to distinguish between relaxable and non-relaxable constraints [9, 10]. Non-relaxable constraints are those which are necessarily satisfied at all iterations of the algorithms. On the other hand, relaxable constraints need only to be satisfied approximately or asymptotically. It is interesting to recognize as soon as possible the infeasibility of non-relaxable constraints (due, for example, to incompatibility of definitions) because the model has no sense at all if these constraints are not fulfilled. On the other hand, one should be tolerant with respect to non-fulfillment of relaxable constraints and one should take into account the objective function value in that case.

In Section 2 of this paper we define a conceptual algorithm of Augmented Lagrangian type for solving Nonlinear Programming problems. Essentially, the algorithm corresponds to the ones introduced in [1] and [16], for "local" and global optimization, respectively. One of the differences is that the complementarity measure employed for deciding whether to increase the penalty parameter will be defined here as the product between multipliers and constraints. With a suitable modification involving the updating of the multipliers we will prove that the global version of the algorithm, in which it is assumed that the subproblems are solved globally, converges to optimizers of the objective function on the set of minimizers of infeasibility. The version of the main algorithm presented in Section 3 corresponds, with some modifications that include measuring complementarity by products, to the algorithm introduced in [1]. In this case we show, as in [1], that limit points are stationary points for the infeasibility measure and, in addition, we prove that these points satisfy an optimality condition on a subset of minimizers of infeasibility. In Section 4 we establish the consequences of the results of Section 3 in the presence of weak constraint qualifications. Experiments will be presented in Section 5 and conclusions will be given in Section 6.

Notation. If $v \in I\!R^n$, $v = (v_1, \dots, v_n)^T$, we denote $v_+ = (\max(0, v_1), \dots, \max(0, v_n))^T$. If $K = (k_1, k_2, \dots) \subseteq I\!N$ (with $k_j < k_{j+1}$ for all j), we denote $K \underset{\infty}{\subset} I\!N$. The symbol $\|\cdot\|$ will denote the Euclidean norm. If $h : I\!R^n \to I\!R^m$, we denote $\nabla h(x) = (\nabla h_1(x), \dots, \nabla h_m(x)) \in I\!R^{n \times m}$.

2 Conceptual Algorithm and global interpretations

We will consider the optimization problem defined by

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

where $h : I\!\!R^n \to I\!\!R^m$, $g : I\!\!R^n \to I\!\!R^p$, and $f : I\!\!R^n \to I\!\!R$ are smooth and $\Omega \subseteq I\!\!R^n$ is closed. The Lagrangian function will be defined by

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{p} \mu_i g_i(x)$$

for all $x \in \Omega$, $x \in I\!\!R^n$, $\lambda \in I\!\!R^m$, and $\mu \in I\!\!R^p_+$, whereas the Augmented Lagrangian [25, 35, 42, 45] will be given by

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

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

Below we define the main algorithm employed in this paper. The algorithm will be "conceptual" in the sense that each outer iteration will be given by the "approximate minimization" of the Augmented Lagrangian subject to the non-relaxable set Ω. "Approximate minimization" is a deliberately ambiguous denomination that will be subject to different interpretations along the paper.

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 $k \leftarrow 1$.

Step 1. Find $x^k \in \Omega$ as an approximate solution of

$$\text{Minimize } L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k) \text{ subject to } x \in \Omega. \tag{2}$$

Step 2. Compute multipliers

$$\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k) \tag{3}$$

and

$$\mu^{k+1} = (\bar{\mu}^k + \rho_k g(x^k))_+. \tag{4}$$

Step 3. Define

$$V_i^k = \mu_i^{k+1} g_i(x^k) \text{ for } i = 1, \ldots, p.$$

If $k = 1$ or

$$\max\{\|h(x^k)\|, \|g(x^k)_+\|, \|V^k\|\} \le \tau \max\{\|h(x^{k-1})\|, \|g(x^{k-1})_+\|, \|V^{k-1}\|\}, \tag{5}$$

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

Step 4. Compute $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\bar{\mu}^{k+1} \in [0, \mu_{\max}]^p$ in such a way that if $\lambda^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\mu^{k+1} \in [0, \mu_{\max}]^p$ then $\bar{\lambda}^{k+1} = \lambda^{k+1}$ and $\bar{\mu}^{k+1} = \mu^{k+1}$.

As in [1] and other Augmented Lagrangian algorithms [25, 26, 27], the penalty parameter is increased only if enough progress, in terms of feasibility and complementarity, has not been obtained at iteration k. Due to scaling reasons, we adopted here the product form to measure complementarity. The comparison between the products $\mu_i^{k+1} g_i(x^k)$ and $\mu_i^k g_i(x^{k-1})$ is more invariant under scaling than the comparisons that involve $\min\{-g_i(x^k), \mu_i^{k+1}\}$ or $\min\{-g_i(x^k), \bar{\mu}_i^k/\rho_k\}$ used in [1] and [15], respectively. Moreover, the product form is connected with the Complementary Approximate KKT condition (CAKKT) recently introduced in [7], where it was proved that, in the Augmented Lagrangian context of [1], convergence to feasible points implies convergence to null products of the constraints values with their corresponding Lagrange multipliers approximations, independently of the fulfillment of constraint qualifications.

The assumption below gives a (global) interpretation for the approximate minimization of the subproblem that defines the outer iteration, which corresponds to the global Augmented Lagrangian method of [16].

Assumption 2.1 *Assume that $\{\varepsilon_k\}$ is a sequence such that $\lim_{k\to\infty} \varepsilon_k = 0$ and $\varepsilon_k \geq 0$ for all $k \in I\!N$. At Step 1 of Algorithm 2.1, the point $x^k \in \Omega$ is such that*

$$L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k) + \varepsilon_k \tag{6}$$

for all $x \in \Omega$.

Assumption 2.1 requires the approximate global minimization of the subproblems. This strong assumption will allow us to prove strong results. The global minimization requirement is impractical for many real-world problems of interest. However, it is well known that global optimization tools have been developed and successfully applied to specific classes of real-world problems. See [16] and references therein.

The proofs of Theorems 2.1 and 2.2 are small variations of the ones of Theorems 1 and 2 of [16, pp. 142–145]. Theorem 2.1 will show that, under Assumption 2.1, Algorithm 2.1 finds global minimizers of the infeasibility measure $\|h(x)\|^2 + \|g(x)_+\|^2$. In particular, if the problem is feasible, the algorithm finds feasible points.

Theorem 2.1 *Let $\{x^k\}$ be a sequence generated by Algorithm 2.1 under Assumption 2.1. If x^* is a limit point of $\{x^k\}$, we have that*

$$\|h(x^*)\|^2 + \|g(x^*)_+\|^2 \leq \|h(x)\|^2 + \|g(x)_+\|^2$$

for all $x \in \Omega$.

Theorem 2.2 states that, when the problem is feasible, the algorithm with Assumption 2.1 finds global minimizers of (1).

Theorem 2.2 *Suppose that, at Step 1 of Algorithm 2.1, x^k is chosen according to Assumption 2.1. Assume, moreover, that the feasible region of problem (2) is non-empty and x^* is a limit point of the sequence $\{x^k\}$. Then, x^* is feasible and*

$$f(x^*) \leq f(x)$$

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

Theorems 2.1 and 2.2 say that Algorithm 2.1 always converges to minimizers of infeasibility and that, in the feasible case, convergence occurs to optimal points. However, these theorems do not provide any information about the quality of the minimal infeasible points in terms of optimality. The following definition of the auxiliary problem $PA(x^*)$ will help us to understand the behavior of Algorithm 2.1 in infeasible cases.

For all $x^* \in \Omega$, we denote by $PA(x^*)$ the following auxiliary problem:

$$\begin{array}{lrcl} \text{Minimize} & f(x) & & \\ \text{subject to} & h(x) - h(x^*) & = & 0 \\ & g_i(x) - g_i(x^*) & \leq & 0 \text{ for all } i \text{ such that } g_i(x^*) > 0, \\ & g_i(x) & \leq & 0 \text{ for all } i \text{ such that } g_i(x^*) \leq 0, \\ & x \in \Omega. & & \end{array} \tag{7}$$

Let x^* be a global minimizer of the infeasibility. Although the feasible set of (7) is contained in the set of global minimizers of the infeasibility, the latter is generally bigger than the former. In other words, solutions of $PA(x^*)$ may not be minimizers of the objective function subject to minimal infeasibility.

Let us prove that any limit point x^* of the sequence generated by Algorithm 2.1, with Assumption 2.1, is necessarily a global minimizer of $PA(x^*)$.

Theorem 2.3 *Assume that $x^* \in \Omega$ is a limit point of a sequence generated by Algorithm 2.1, where the subproblems at Step 1 are computed according to Assumption 2.1. Then, for all feasible point x of problem $PA(x^*)$, we have that $f(x^*) \leq f(x)$.*

Proof. If $h(x^*) = 0$ and $g(x^*) \leq 0$, the thesis follows from Theorems 2.1 and 2.2. Let us assume, from now on, that x^* is infeasible for problem (1). This implies, by Step 3, that $\lim_{k\to\infty} \rho_k = \infty$. Let x be an arbitrary feasible point of $PA(x^*)$. Then, $h(x) = h(x^*)$, $g_i(x) \leq 0$ for all i such that $g_i(x^*) \leq 0$, and $g_i(x) \leq g_i(x^*)$ for all i such that $g_i(x^*) > 0$. By Theorem 2.1, x^* is a global minimizer of infeasibility, therefore, by the definition of $PA(x^*)$,

$$\sum_{i=1}^{m} h_i(x^*)^2 + \sum_{g_i(x^*)>0} g_i(x^*)^2 \leq \sum_{i=1}^{m} h_i(x)^2 + \sum_{g_i(x^*)>0} g_i(x)^2.$$

Since $h(x^*) = h(x)$ we obtain

$$\sum_{g_i(x^*)>0} g_i(x^*)^2 \leq \sum_{g_i(x^*)>0} g_i(x)^2. \tag{8}$$

But, by the definition of $PA(x^*)$, we have that $g_i(x) \leq g_i(x^*)$ whenever $g_i(x^*) > 0$. Then, by (8), $g_i(x) = g_i(x^*)$ for all i such that $g_i(x^*) > 0$.

Therefore, $h(x) = h(x^*)$, $g_i(x) \leq 0$ if $g_i(x^*) \leq 0$, and $g_i(x) = g_i(x^*)$ if $g_i(x^*) > 0$, for all $i = 1, \ldots, p$. Let $K \underset{\infty}{\subset} I\!N$ such that $\lim_{k\in K} x^k = x^*$. Since $x^k \in \Omega$ for all k and Ω is closed, we have that $x^* \in \Omega$. By (6),

$$\begin{aligned} &f(x^k) + \tfrac{\rho_k}{2}\left[\left\|h(x^k) + \tfrac{\bar\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x^k) + \tfrac{\bar\mu^k}{\rho_k}\right)_+\right\|^2\right] \\ &\le f(x) + \tfrac{\rho_k}{2}\left[\left\|h(x) + \tfrac{\bar\lambda^k}{\rho_k}\right\|^2 + \left\|\left(g(x) + \tfrac{\bar\mu^k}{\rho_k}\right)_+\right\|^2\right] + \varepsilon_k, \end{aligned} \tag{9}$$

for all $k \in K$. Note that

$$\begin{aligned} &\left\|\left(g(x) + \tfrac{\bar\mu^k}{\rho_k}\right)_+\right\|^2 = \textstyle\sum_{i=1}^p \max\left(0, g_i(x) + \tfrac{\bar\mu_i{}^k}{\rho_k}\right)^2 \\ &\le \textstyle\sum_{i=1}^p \left(g_i(x)_+ + \tfrac{\bar\mu_i^k}{\rho_k}\right)^2 = \|g(x)_+\|^2 + \tfrac{2}{\rho_k}(\bar\mu^k)^T g(x)_+ + \left(\tfrac{\|\bar\mu^k\|}{\rho_k}\right)^2 \end{aligned} \tag{10}$$

and

$$\begin{aligned} &\left\|\left(g(x^k) + \tfrac{\bar\mu^k}{\rho_k}\right)_+\right\|^2 = \textstyle\sum_{i=1}^p \max\left(0, g_i(x^k) + \tfrac{\bar\mu_i^k}{\rho_k}\right)^2 \\ &\ge \textstyle\sum_{g_i(x^k)>0} \left(g_i(x^k) + \tfrac{\bar\mu_i^k}{\rho_k}\right)^2 \ge \left\|g(x^k)_+\right\|^2 + \tfrac{2}{\rho_k}(\bar\mu^k)^T g(x^k)_+. \end{aligned} \tag{11}$$

So, by (9), (10), and (11),

$$\begin{aligned} &f(x^k) + (\bar\lambda^k)^T h(x^k) + (\bar\mu^k)^T g(x^k)_+ + \tfrac{\rho_k}{2}\left[\left\|h(x^k)\right\|^2 + \left\|g(x^k)_+\right\|^2\right] \\ &\le f(x) + (\bar\lambda^k)^T h(x) + (\bar\mu^k)^T g(x)_+ + \tfrac{\rho_k}{2}\left[\|h(x)\|^2 + \|g(x)_+\|^2\right] + \frac{\left\|\bar\mu^k\right\|^2}{2\rho_k} + \varepsilon_k. \end{aligned}$$

Therefore,

$$\begin{aligned} &f(x^k) + (\bar\lambda^k)^T[h(x^k) - h(x)] + (\bar\mu^k)^T[g(x^k)_+ - g(x)_+] \\ &+ \tfrac{\rho_k}{2}\left[\left(\left\|h(x^k)\right\|^2 + \left\|g(x^k)_+\right\|^2\right) - \left(\|h(x)\|^2 + \|g(x)_+\|^2\right)\right] - \frac{\left\|\bar\mu^k\right\|^2}{2\rho_k} \le f(x) + \varepsilon_k. \end{aligned}$$

By the definition of x, $h(x) = h(x^*)$ and $\|h(x)\|^2 + \|g(x)_+\|^2 = \|h(x^*)\|^2 + \|g(x^*)_+\|^2$, so

$$\begin{aligned} &f(x^k) + (\bar\lambda^k)^T[h(x^k) - h(x^*)] + (\bar\mu^k)^T[g(x^k)_+ - g(x)_+] \\ &+ \tfrac{\rho_k}{2}\left[\left(\left\|h(x^k)\right\|^2 + \left\|g(x^k)_+\right\|^2\right) - \left(\|h(x^*)\|^2 + \|g(x^*)_+\|^2\right)\right] - \frac{\left\|\bar\mu^k\right\|^2}{2\rho_k} \le f(x) + \varepsilon_k. \end{aligned}$$

Define $I = \{i \in \{1, \ldots, p\} \mid g_i(x^*) > 0\}$. For each $i \notin I$, since $g_i(x) \le 0$, we have $\bar\mu_i^k[g_i(x^k)_+ - g_i(x)_+] \ge 0$. Therefore, by Theorem 2.1,

$$f(x^k) + (\bar\lambda^k)^T[h(x^k) - h(x^*)] + \sum_{i\in I} \bar\mu_i^k[g_i(x^k) - g_i(x)] - \frac{\left\|\bar\mu^k\right\|^2}{2\rho_k} \le f(x) + \varepsilon_k,$$

for $k \in K$ large enough. By continuity of f, h, g, the boundedness of $\{\bar\lambda^k\}$ and $\{\bar\mu^k\}$, and the fact that $\lim_{k\in K} g_i(x^k) = g_i(x^*) = g_i(x)$ for each $i \in I$, taking limits in K on both sides of this inequality, we obtain the desired result. $\square$

Counter-Example

We will show that the result of Theorem 2.3 cannot be improved, in the sense that limit points generated by Algorithm 2.1 are not necessarily minimizers of the objective function subject to minimal infeasibility.

Consider the problem

$$\begin{array}{lrcl} \text{Minimize} & -x & & \\ \text{subject to} & x-1 & = & 0 \\ & x+1 & = & 0 \\ & 2(x^2-1) & = & 0 \\ & x\in\Omega & \equiv & \mathbb{R}. \end{array} \tag{12}$$

Clearly, this problem is infeasible and the points that minimize infeasibility are $-\sqrt{3}/2$ and $\sqrt{3}/2$. The latter is the minimizer of the objective function on the set of minimizers of infeasibility. The Augmented Lagrangian is

$$\begin{aligned} L_{\rho_k}(x,\bar\lambda^k,\bar\mu^k) &= -x+\bar\lambda_1^k(x-1)+\bar\lambda_2^k(x+1)+2\bar\lambda_3^k(x^2-1)+\tfrac{\rho_k}{2}\|h(x)\|^2+c_1 \\ &= (\bar\lambda_1^k+\bar\lambda_2^k-1)x+2\bar\lambda_3^k x^2+\tfrac{\rho_k}{2}\|h(x)\|^2+c, \end{aligned}$$

where c_1 and c are constants. Let us define $\lambda_{\min}=-1$ and $\lambda_{\max}=3$.

Since the problem is infeasible, we certainly have that $\rho_k\to\infty$. We have that $h_1(x)=x-1<0$ for x close to $\sqrt{3}/2$ or $-\sqrt{3}/2$, therefore, $\lambda_1^k=\bar\lambda_1^{k-1}+\rho_{k-1}h_1(x^{k-1})\to-\infty$ for any subsequence that converges to points that minimize infeasibility. By Step 4 of Algorithm 2.1, using the choice of projecting onto the safeguarding interval, we have that $\bar\lambda_1^k=\lambda_{\min}$ for k large enough. Analogously, in the case of $h_2(x)=x+1$, we have that $\bar\lambda_2^k=\lambda_{\max}$ and for $h_3(x)=2(x^2-1)$ we have that $\bar\lambda_3^k=\lambda_{\min}$ for k large enough. Then, for k large enough, we have

$$L_{\rho_k}(x,\bar\lambda^k,\bar\mu^k)=(\lambda_{\min}+\lambda_{\max}-1)x+2\lambda_{\min}x^2+\frac{\rho_k}{2}\|h(x)\|^2+c.$$

Then, by the definition of $\lambda_{\min}$ and $\lambda_{\max}$,

$$L_{\rho_k}(x,\bar\lambda^k,\bar\mu^k)=x-2x^2+\frac{\rho_k}{2}\|h(x)\|^2+c.$$

Clearly, if x is close to $-\sqrt{3}/2$ and y is close to $\sqrt{3}/2$ one has that

$$L_{\rho_k}(x,\bar\lambda^k,\bar\mu^k)<L_{\rho_k}(y,\bar\lambda^k,\bar\mu^k).$$

As a consequence, $-\sqrt{3}/2$ will be the limit point of the sequence $\{x^k\}$ generated by Algorithm 2.1 under Assumption 2.1 when applied to problem (12), for the particular choice $\lambda_{\min}=-1$ and $\lambda_{\max}=3$ and computing $\bar\lambda^{k+1}$ as the projection of λ^{k+1} onto $[\lambda_{\min},\lambda_{\max}]^m$. However, this limit point, in spite of being a minimizer of the infeasibility, does not minimize f on the set $\{-\sqrt{3}/2,\sqrt{3}/2\}$.

The counter-example above shows that the property of minimizing the objective function subject to minimal infeasibility does not hold for arbitrary choices of the approximate Lagrange multipliers. In the following assumption we state a simple updating rule that essentially says that multipliers should be updated with the standard first-order formulae (3) and (4) except when they exceed the safeguarding bounds. In this last case, the multipliers are reset to zero and the next iteration is essentially a quadratic penalty iteration. In the practical implementation of Algorithm 2.1, one uses artificial safeguarding bounds for the Lagrange multipliers [1]. If the safeguards of the multipliers are large enough and, in addition, the problem is feasible and some constraint qualification is satisfied, the safeguarding rule will not be activated. In these cases the alternative choice will not affect the behavior of the algorithm.

In the Augmented Lagrangian approach the quantities λ_i/ρ and μ_i/ρ represent shifts of the equality and inequality constraints, respectively. Roughly speaking one expects that modest penalization with respect to a suitably shifted constraint should produce similar effect as extreme penalization without shifts. Obviously, when one is forced to use very large (extreme) penalty parameters shifting the constraints makes no sense at all. This is the reason why we preserve Lagrange multipliers between finite bounds, forcing the shifts to go to zero as ρ tends to infinity. Due to the updating rule of multipliers the case in which λ^{k+1} or μ^{k+1} tend to infinity is generally associated with $\rho \to \infty$ too. So, when λ^{k+1} or μ^{k+1} exceed the safeguarding bounds it is sensible to annihilate the shifts, thus setting $\bar{\lambda}^{k+1} = 0$ and $\bar{\mu}^{k+1} = 0$. This is stated in the assumption below.

Assumption 2.2 *At Step 4 of Algorithm 2.1, if $\lambda^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\mu^{k+1} \in [0, \mu_{\max}]^p$, we define*

$$\bar{\lambda}^{k+1} = \lambda^{k+1} \;\; \textit{and} \;\; \bar{\mu}^{k+1} = \mu^{k+1}.$$

Otherwise, we define $\bar{\lambda}^{k+1} = 0$ and $\bar{\mu}^{k+1} = 0$.

Theorem 2.4 *Assume that $x^* \in \Omega$ is a limit point of a sequence generated by Algorithm 2.1, with Assumptions 2.1 and 2.2. Then, $\|h(x^*)\|^2 + \|g(x^*)_+\|^2 \le \|h(x)\|^2 + \|g(x)_+\|^2$ for all $x \in \Omega$, and*

$$f(x^*) \le f(x) \;\; \textit{for all} \;\; x \in \Omega \;\; \textit{such that} \;\; \|h(x)\|^2 + \|g(x)_+\|^2 = \|h(x^*)\|^2 + \|g(x^*)_+\|^2.$$

Proof. If $h(x^*) = 0$ and $g(x^*) \le 0$, the thesis follows from Theorems 2.1 and 2.2. Let us assume, from now on, that $\|h(x^*)\|^2 + \|g(x^*)_+\|^2 = c > 0$. This implies, by Step 3, that $\lim_{k\to\infty} \rho_k = \infty$. Since, by Theorem 2.1, x^* is a global minimizer of $\|h(x)\|^2 + \|g(x)_+\|^2$, it turns out that, for all $k \in I\!N$, $\|h(x^k)\|^2 + \|g(x^k)_+\|^2 \ge c$. Since by Theorem 2.1 any other limit point of $\{x^k\}$ is infeasible, by (3), (4), the boundedness of $\{\bar{\lambda}^k\}$ and $\{\bar{\mu}^k\}$, and the fact that ρ_k tends to infinity, we have that, for all k large enough, either $\lambda^{k+1} \notin [\lambda_{\min}, \lambda_{\max}]^m$ or $\mu^{k+1} \notin [0, \mu_{\max}]^p$. Therefore, by Assumption 2.2, there exists $k_0 \in I\!N$ such that for all $k \ge k_0$ we have that $\bar{\lambda}^k = 0$ and $\bar{\mu}^k = 0$.

Let $K \underset{\infty}{\subset} \{k_0, k_0+1, k_0+2, \dots\}$ be such that $\lim_{k\in K} x^k = x^*$. By Assumption 2.1 and the fact that $\|\bar{\lambda}^k\| = \|\bar{\mu}^k\| = 0$ we have that, for all $x \in \Omega$,

$$f(x^k) + \frac{\rho_k}{2}\left[\|h(x^k)\|^2 + \|g(x^k)_+\|^2\right] \le f(x) + \frac{\rho_k}{2}\left[\|h(x)\|^2 + \|g(x)_+\|^2\right] + \varepsilon_k, \tag{13}$$

for all $k \in K$ such that $k \geq k_0$. In particular, if $x \in \Omega$ is such that $\|h(x)\|^2 + \|g(x)_+\|^2 = \|h(x^*)\|^2 + \|g(x^*)_+\|^2$ we have that x is a global minimizer of the infeasibility on Ω, thus

$$\frac{\rho_k}{2}\left[\|h(x^k)\|^2 + \|g(x^k)_+\|^2\right] \geq \frac{\rho_k}{2}\left[\|h(x)\|^2 + \|g(x)_+\|^2\right].$$

Therefore, by (13) and Assumption 2.1,

$$f(x^k) \leq f(x) + \varepsilon_k \text{ for all } k \in K.$$

By the continuity of f, taking limits on both sides of this inequality, we obtain the desired result. $\square$

3 Local interpretations

In Section 2 we presented an algorithm that exhibits the desirable property of converging to global minimizers of the objective function, subject to minimal infeasibility. This algorithm depends on the interpretation of Step 1 in terms of global optimization of the subproblem (2). Although implementations of Algorithm 2.1 with those characteristics exist (see, for example, [16]), in the case of large-scale problems solving (2) by means of cheaper "local" algorithms in generally preferred. In order to describe the local version of Algorithm 2.1, following the lines of [1], we define the non-relaxable set Ω by

$$\Omega = \{x \in I\!R^n \mid \underline{h}(x) = 0, \underline{g}(x) \leq 0\}, \tag{14}$$

where $\underline{h} : I\!R^n \to I\!R^{\underline{m}}$ and $\underline{g} : I\!R^n \to I\!R^{\underline{p}}$ are continuously differentiable.

Assumption 3.1 defines the approximate solution of (2) in terms of first-order optimality conditions for the minimization of the Augmented Lagrangian subject to $\underline{h}(x) = 0$ and $\underline{g}(x) \leq 0$.

Assumption 3.1 *At Step 1 of Algorithm 2.1, the point $x^k \in I\!R^n$ is such that there exist $v^k \in I\!R^{\underline{m}}$ and $w^k \in I\!R^{\underline{p}}_+$ satisfying*

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) + \nabla \underline{h}(x^k) v^k + \nabla \underline{g}(x^k) w^k\| \leq \varepsilon_k, \tag{15}$$

$$\underline{h}(x^k) = 0, \ \underline{g}(x^k) \leq 0, \ |\underline{g}_i(x^k) w_i^k| \leq \varepsilon_k \ \text{ for all } \ i = 1, \ldots, \underline{p}, \tag{16}$$

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

Asking feasibility with respect to the constraints $\underline{h}$ and $\underline{g}$ in Assumption 3.1 is sensible when these constraints have favorable structure. Under Assumption 3.1, Algorithm 2.1 is similar to the Augmented Lagrangian algorithm defined in [1], the main difference being that complementarity is measured by the product form instead of the minimum. Practical implementations of the Augmented Lagrangian algorithm defined in [1] considering box-constraints and affine-constraints for the non-relaxable set Ω have been developed in [1, 2, 14, 18] and [8], respectively. In this paper we wish to emphasize the properties related to convergence towards infeasible points.

Let us define now Approximate KKT points in the sense of [3].

Definition 3.1. *We say that a feasible point x^* for problem (1) satisfies the Approximate KKT conditions (AKKT) with respect to a problem of the form (1), with Ω given by (14), if there exist sequences $\{x^k\} \subseteq I\!R^n$, $\{\lambda^k\} \subseteq I\!R^m$, $\{\mu^k\} \subseteq I\!R^p_+$, $\{v^k\} \in I\!R^{\underline{m}}$, $\{w^k\} \in I\!R^{\underline{p}}_+$, such that*

$$\lim_{k\to\infty} x^k = x^*,$$

$$\lim_{k\to\infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k + \nabla \underline{h}(x^k)v^k + \nabla \underline{g}(x^k)w^k\| = 0, \tag{17}$$

$$\lim_{k\to\infty} \min\{-g_i(x^k), \mu_i^k\} = 0 \textit{ for all } i = 1, \ldots, p, \tag{18}$$

and

$$\lim_{k\to\infty} \min\{-\underline{g}_i(x^k), w_i^k\} = 0 \textit{ for all } i = 1, \ldots, \underline{p}. \tag{19}$$

When x^* satisfies AKKT we will say that x^* is an AKKT point.

Proposition 3.1. *A feasible point x^* satisfies AKKT with respect to the problem (1), with the definition (14) if and only if there exist sequences $\{x^k\} \subseteq I\!R^n$, $\{\lambda^k\} \subseteq I\!R^m$, $\{\mu^k\} \subseteq I\!R^p_+$, $\{v^k\} \in I\!R^{\underline{m}}$, $\{w^k\} \in I\!R^{\underline{p}}_+$, such that, for all $k \in I\!N$*

$$\lim_{k\to\infty} x^k = x^*,$$

$$\lim_{k\to\infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k + \nabla \underline{h}(x^k)v^k + \nabla \underline{g}(x^k)w^k\| = 0, \tag{20}$$

$$\mu_i^k = 0 \textit{ for all } i = 1, \ldots, p \textit{ such that } g_i(x^*) < 0, \tag{21}$$

and

$$w_i^k = 0 \textit{ for all } i = 1, \ldots, \underline{p} \textit{ such that } \underline{g}_i(x^*) < 0. \tag{22}$$

Proof. This proof was essentially given in Lemma 2.1 of [3, p. 629]. □

Theorem 3.1. *Assume that the sequence $\{x^k\}$ is generated by Algorithm 2.1 under Assumption 3.1 and let x^* be a limit point of $\{x^k\}$. Then, x^* satisfies the AKKT conditions associated with the problem*

$$\textit{Minimize } \frac{1}{2}\left(\|h(x)\|^2 + \|g(x)_+\|^2\right) \quad \textit{subject to } \underline{h}(x) = 0, \ \underline{g}(x) \le 0. \tag{23}$$

Proof. By Assumption 3.1 we have that $\underline{h}(x^*) = 0$ and $\underline{g}(x^*) \le 0$. Consider first the case in which the sequence of penalty parameters $\{\rho_k\}$ is bounded. Then, by (5), we have that

$\lim_{k\to\infty} \|h(x^k)\| = \lim_{k\to\infty} \|g(x^k)_+\| = 0$. Therefore, $h(x^*) = 0$ and $g(x^*)_+ = 0$. Therefore, x^* is an unconstrained global minimizer of $\frac{1}{2}(\|h(x)\|^2 + \|g(x)_+\|^2)$. Thus, the gradient of this function vanishes at x^*. Then, the AKKT conditions corresponding to (23) hold trivially defining all multipliers equal to zero in (17–19).

Now, consider the case in which ρ_k tends to infinity. Let $K \underset{\infty}{\subset} I\!N$ be such that $\lim_{k\in K} x^k = x^*$. By (15), for all $k \in K$ we have that

$$\begin{aligned}\Big\| \nabla f(x^k) + \rho_k \Big\{ \textstyle\sum_{i=1}^m \nabla h_i(x^k) \Big[h_i(x^k) + \frac{\bar\lambda_i^k}{\rho_k} \Big] + \textstyle\sum_{i=1}^p \nabla g_i(x^k) \Big[\max(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}) \Big] \Big\} \\ + \nabla \underline{h}(x^k) v^k + \nabla \underline{g}(x^k) w^k \Big\| \leq \varepsilon_k.\end{aligned} \tag{24}$$

Since $\lim_{k\in K} x^k = x^*$, the sequence $\{\nabla f(x^k)\}$ is bounded for $k \in K$. Dividing both sides of (24) by ρ_k we obtain:

$$\lim_{k\in K} \sum_{i=1}^m \nabla h_i(x^k) \left[h_i(x^k) + \frac{\bar\lambda_i^k}{\rho_k} \right] + \sum_{i=1}^p \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right] + \nabla \underline{h}(x^k) \frac{v^k}{\rho_k} + \nabla \underline{g}(x^k) \frac{w^k}{\rho_k} = 0. \tag{25}$$

Now, by (16) we have that $\lim_{k\to\infty} w_i^k \underline{g}_i(x^k) = 0$. Therefore, if $\underline{g}_i(x^*) < 0$, we have that $w_i^k \to 0$. Then, by (25), and the boundedness of $\bar\lambda_i^k$,

$$\lim_{k\in K} \sum_{i=1}^m \nabla h_i(x^k) h_i(x^k) + \sum_{i=1}^p \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right] + \nabla \underline{h}(x^k) \frac{v^k}{\rho_k} + \sum_{\underline{g}_i(x^*)=0} \nabla \underline{g}_i(x^k) \frac{w_i^k}{\rho_k} = 0. \tag{26}$$

Now,

$$\begin{aligned}\textstyle\sum_{i=1}^p \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right] = \textstyle\sum_{g_i(x^*)<0} \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right] \\ + \textstyle\sum_{g_i(x^*)=0} \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right] + \textstyle\sum_{g_i(x^*)>0} \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right].\end{aligned}$$

If $g_i(x^*) < 0$, then $g_i(x^k) < 0$ for k large enough, so, since $\rho_k \to \infty$ and $\bar\mu^k$ is bounded, the first term of the right-hand side of the equality above vanishes for k large enough. Thus,

$$\lim_{k\in K} \sum_{g_i(x^*)<0} \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right] = \lim_{k\in K} \sum_{g_i(x^*)<0} \nabla g_i(x^k) g_i(x^k)_+ = 0. \tag{27}$$

Analogously,

$$\lim_{k\in K} \sum_{g_i(x^*)=0} \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right] = \lim_{k\in K} \sum_{g_i(x^*)=0} \nabla g_i(x^k) g_i(x^k)_+ = 0. \tag{28}$$

Finally, by the boundedness of $\bar\mu^k$,

$$\lim_{k\in K} \sum_{g_i(x^*)>0} \nabla g_i(x^k) \left[\max\left(0, g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right) \right] = \lim_{k\in K} \sum_{g_i(x^*)>0} \nabla g_i(x^k) g_i(x^k). \tag{29}$$

By (26), (27), (28), and (29), we have:

$$\lim_{k\in K}\sum_{i=1}^m \nabla h_i(x^k)h_i(x^k)+\sum_{i=1}^p \nabla g_i(x^k)\left[\max\left(0, g_i(x^k)+\frac{\bar\mu_i^k}{\rho_k}\right)\right]+\nabla \underline{h}(x^k)\frac{v^k}{\rho_k}+\sum_{\underline{g}_i(x^*)=0}\nabla \underline{g}_i(x^k)\frac{w_i^k}{\rho_k}$$
$$=\lim_{k\in K}\nabla h(x^k)h(x^k)+\nabla g(x^k)g(x^k)_+ +\nabla \underline{h}(x^k)\frac{v^k}{\rho_k}+\sum_{\underline{g}_i(x^*)=0}\nabla \underline{g}_i(x^k)\frac{w_i^k}{\rho_k}=0.$$

Thus,

$$\lim_{k\in K}\tfrac{1}{2}\nabla(\|h(x^k)\|^2+\|g(x^k)_+\|^2))+\nabla \underline{h}(x^k)\frac{v^k}{\rho_k}+\sum_{\underline{g}_i(x^*)=0}\nabla \underline{g}_i(x^k)\frac{w_i^k}{\rho_k}$$
$$=\lim_{k\in K}\nabla h(x^k)h(x^k)+\nabla g(x^k)g(x^k)_+ +\nabla \underline{h}(x^k)\frac{v^k}{\rho_k}+\sum_{\underline{g}_i(x^*)=0}\nabla \underline{g}_i(x^k)\frac{w_i^k}{\rho_k}=0.$$

By Proposition 3.1 this implies the desired result. □

Remark. In Assumption 3.1 we imposed that $\lim\limits_{k\to\infty} w_i^k \underline{g}_i(x^k)=0$. However, it is easy to see that the thesis of Theorem 3.1 holds under the weaker condition that $|w_i^k \underline{g}_i(x^k)|$ is bounded. This seems to indicate that the product form of complementarity is more adequate than the "min" form of complementarity for declaring that a point is an approximate solution of the subproblem in the presence of possible infeasibility.

Theorem 3.2. *Assume that the sequence $\{x^k\}$ is generated by Algorithm 2.1 under Assumption 3.1. Let x^* be a limit point of $\{x^k\}$. Then, x^* satisfies the AKKT conditions related with problem PA(x^*) (defined by (7) and (14)).*

Proof. Let $K \underset{\infty}{\subset} \mathbb{N}$ be such that $\lim_{k\in K} x^k = x^*$. By (3), (4), and (15) there exist $v^k\in\mathbb{R}^{\underline{m}}$ and $w^k\in\mathbb{R}_+^{\underline{p}}$ such that

$$\lim_{k\to\infty}\left\|\nabla\mathcal{L}(x^k,\lambda^{k+1},\mu^{k+1})+\sum_{i=1}^{\underline{m}} v_i^k\nabla \underline{h}_i(x^k)+\sum_{i=1}^{\underline{p}} w_i^k\nabla \underline{g}_i(x^k)\right\|=0. \tag{30}$$

By (16) we have that

$$\lim_{k\in K}\|\underline{h}(x^k)\|=0 \ \text{ and } \ \lim_{k\in K}\min\{-\underline{g}_i(x^k), w_i^k\}=0, \tag{31}$$

for all $i=1,\dots,\underline{p}$.

If ρ_k is bounded, by (5), we have that

$$\lim_{k\to\infty}\|h(x^k)\|=0, \lim_{k\to\infty}\|g(x^k)_+\|=0,$$

and, since $\lim_{k\to\infty}\mu_i^{k+1}g_i(x^k)=0$,

$$\lim_{k\to\infty}\min\{-g_i(x^k),\mu_i^{k+1}\}=0$$

for all $i=1,\dots,p$. Then, x^* is feasible, $h(x^*)=0$, $g(x^*)\le 0$, and the thesis follows trivially.

Consider the case in which $\lim_{k\to\infty} \rho_k = \infty$. Suppose first that $g_i(x^*) < 0$. By the boundedness of $\{\bar\mu^k\}$ and the updating formula (4) we have that, for $k \in K$ large enough, $\mu_i^{k+1} = 0$. So,

$$\lim_{k\in K} \min\{-g_i(x^k), \mu_i^{k+1}\} = 0, \tag{32}$$

for all i such that $g_i(x^*) \le 0$. Finally, suppose that $g_i(x^*) > 0$. By the definition of $PA(x^*)$, it turns out that

$$\lim_{k\in K} \min\{-(g_i(x^k) - g_i(x^*)), \mu_i^{k+1}\} = 0. \tag{33}$$

Therefore, by (30), (31), (32), and (33), x^* is an AKKT point of $PA(x^*)$. □

We finish this section proving that, under Assumptions 2.2 and 3.1, if Algorithm 2.1 fails to find approximate feasible points, then every limit point satisfies the AKKT optimality conditions related with the minimization of the objective function with only one relaxable constraint that says that the sum of squares of infeasibilities is the presumably minimal one. As in the case of Theorem 2.4 this is due to the fact that, under Assumption 2.2, the algorithm switches to the quadratic penalty method in this case.

Theorem 3.3. *Assume that the sequence $\{x^k\}$ is generated by Algorithm 2.1 under Assumptions 2.2 and 3.1 and that the sequence $\{\|h(x^k)\|^2 + \|g(x^k)_+\|^2\}$ is bounded away from zero for k large enough. Suppose that $\{x^k\}$ admits at least one limit point x^*. Define, for all $x \in I\!R^n$,*

$$\Gamma(x) = \|h(x)\|^2 + \|g(x)_+\|^2.$$

Then, x^* satisfies the AKKT conditions related with the problem

$$\text{Minimize } f(x) \text{ subject to } \Gamma(x) - \Gamma(x^*) = 0,\ \underline{h}(x) = 0, \text{ and } \underline{g}(x) \le 0. \tag{34}$$

Proof. Since the sequence $\{\|h(x^k)\|^2 + \|g(x^k)_+\|^2\}$ is bounded away from zero, by Step 3 of Algorithm 2.1, we have that $\lim_{k\to\infty} \rho_k = \infty$. Thus, by the boundedness of $\{\bar\lambda^k\}$ and $\{\bar\mu^k\}$, we deduce that $\lim_{k\to\infty} \|\lambda^{k+1}\| + \|\mu^{k+1}\| = \infty$. Then, by Assumption 2.2 we have that $\bar\lambda^k = 0$ and $\bar\mu^k = 0$ for k large enough.

Therefore, by Assumption 3.1 there exist sequences $\{v^k\} \subseteq I\!R^{\underline{m}}$ and $\{w^k\} \subseteq I\!R^{\underline{p}}_+$ such that

$$\lim_{k\to\infty} \nabla f(x^k) + \frac{\rho_k}{2}\nabla\Gamma(x^k) + \nabla \underline{h}(x^k) v^k + \nabla \underline{g}(x^k) w^k = 0,$$

$$\underline{h}(x^k) = 0,\ \underline{g}(x^k) \le 0, \text{ and } \lim_{k\to\infty} \underline{g}_i(x^k) w_i^k = 0 \text{ for all } i = 1, \dots, \underline{p}.$$

Thus,

$$\lim_{k\to\infty} \min\{-\underline{g}_i(x^k), w_i^k\} = 0 \text{ for all } i = 1, \dots, \underline{p}.$$

Since x^* is a feasible point of problem (34), we deduce that this point satisfies the AKKT condition of (34). □

The "ideal" property of finding minimizers of the objective function under minimal global infeasibility is not achievable by affordable algorithms. On the one hand the problem of finding global minimizers of infeasibility is a global optimization problem the solution of which, in the absence of suitable assumptions on the problem, can be obtained only under evaluation on a dense set. By the same reason, without dense-set evaluations, it is not possible to guarantee global minimization of the objective function subject to minimal infeasibility. Therefore, the best that we can prove using standard affordable algorithms is that limit points satisfy some optimality condition related to the minimization of f subject to the fact that some optimality condition related to the minimization of infeasibility holds. This is the type of result proved in Theorems 3.2 and 3.3.

4 Results under constraint qualifications

Theorem 3.2 establishes that the main local algorithm satisfies an optimality property in the infeasible case: every limit point, infeasible or not, fulfills the Approximate KKT conditions relative to problem (7). It is interesting to study the conditions under which this property implies that the limit point fulfills the KKT conditions of (7). Clearly, this is not always true. Consider, for example, the problem of minimizing x subject to $x^2+1=0$. Although the AKKT condition holds at the solution $x^*=0$ and, in fact, our algorithm converges to x^* independently of the initial point, the auxiliary problem $PA(x^*)$ is

$$\text{Minimize } x \text{ subject to } x^2 = 0$$

and the KKT conditions do not hold for this problem.

In this section we will prove that, when a limit point x^* satisfies some constraint qualifications, the KKT conditions of $PA(x^*)$ hold at x^*. In the last few years many weak constraint qualifications were introduced in [4, 5, 36, 37, 40, 41]. In particular, the Constant Positive Generators (CPG) condition introduced in [5] is one of the weakest constraint qualification that allows one that prove that AKKT implies KKT. The definition of the CPG condition for the minimization of a function $f(x)$ with m' constraints $H_i(x)=0$ and p' constraints $G_i(x)\leq 0$ is given below.

Definition 4.1. Assume that $H(x^*)=0$ and $G(x^*)\leq 0$. Define $I=\{1,\ldots,m'\}$. Let $J\subseteq\{1,\ldots,p'\}$ be the indices of the active inequality constraints at x^*. Let J_- be set of indices $\ell\in J$ such that, for all $\ell\in J_-$, there exist $\lambda_1,\ldots,\lambda_{m'}\in I\!R$ and $\mu_j\in I\!R_+$ for all $j\in J$, such that

$$-\nabla G_\ell(x^*) = \sum_{i=1}^{m'} \lambda_i \nabla H_i(x^*) + \sum_{j\in J} \mu_j \nabla G_j(x^*). \tag{35}$$

Define $J_+ = J - J_-$. We say that the Constant Positive Generator (CPG) condition holds at x^* if there exists $I'\subseteq I$ and $J'\subseteq J_-$ such that

1. The gradients $\nabla H_i(x^*)$ and $\nabla G_j(x^*)$ indexed by $i\in I'$ and $j\in J'$ are linearly independent.

2. For all x in a neighborhood of x^*, if

$$z = \sum_{i=1}^{m'} \lambda_i' \nabla H_i(x) + \sum_{j \in J} \mu_j' \nabla G_j(x),$$

with $\mu_j' \geq 0$ for all $j \in J$, then for all $i \in I'$, $\ell \in J'$, and $j \in J_+$, there exist $\lambda_i'' \in I\!R$, $\lambda_\ell''' \in I\!R$, and $\mu_j'' \in I\!R_+$ such that

$$z = \sum_{i \in I'} \lambda_i'' \nabla H_i(x) + \sum_{\ell \in J'} \lambda_\ell''' \nabla G_\ell(x) + \sum_{j \in J_+} \mu_j'' \nabla G_j(x).$$

Theorem 4.1. *Assume that the sequence $\{x^k\}$ is generated by Algorithm 2.1 under Assumption 3.1. Let x^* be a limit point of $\{x^k\}$ and assume that the CPG constraint qualification with respect to the problem $PA(x^*)$, given by (7) and (14), is fulfilled at x^*, with the obvious definition of H, G, m', and p'. Then, x^* satisfies the KKT conditions related with $PA(x^*)$.*

Proof. By Theorem 3.2 x^* satisfies the AKKT conditions with respect to the problem $PA(x^*)$ with the suitable definition of H, G, m', and p'. Then, by Theorem 3.1 of [5], x^* satisfies the KKT conditions of the problem of minimizing $f(x)$ subject to $H(x) = 0$ and $G(x) \leq 0$. This completes the proof. □

Constraint qualifications are properties of the constraints of Nonlinear Programming problems that, when satisfied at a local minimizer x^*, independently of the objective function, imply that x^* fulfills the KKT conditions. In other words, if CQ is a constraint qualification, the proposition "KKT or not-CQ" is a necessary optimality condition. As a consequence, weak constraint qualifications produce strong optimality conditions. CPG is a constraint qualification weaker than LICQ (Linear Independence of the gradients of active constraints), MFCQ (Mangasarian-Fromovitz), and CPLD (Constant Positive Linear Dependence) [5]. The linearity of active constraints also implies CPG. As a consequence of the results of this section and Section 3 we conclude that Algorithm 2.1 converges to limit points that satisfy strong optimality conditions with respect to the auxiliary problem (7).

5 Experiments

We implemented Algorithm 2.1 with Assumptions 2.2 and 3.1 as a straightforward modification of Algencan, the Augmented Lagrangian method introduced in [1] (available at the TANGO Project web page [49]), that considers box constraints for the non-relaxable set Ω. Basically, our modification only affects the criterion for increasing the penalty parameter, so the modified method will be called "modified Algencan".

5.1 Preliminary experiments

In a first set of experiments, that can be found in [50, 51], using all the problems from the CUTEr collection [33], we performed an exhaustive numerical comparison between the modified

Algencan and the original version of Algencan. We did not find meaningful differences between the behavior of both algorithms, neither in feasible, nor in infeasible problems. Although we believe that the product form of measuring complementarity is less prone to scaling inconveniences than the ones based on the minimum operator and that the product form has the theoretical advantage of being connected to the strong sequential optimality condition CAKKT [7], these features does not seem to be enough to produce practical effects. In a second set of experiments, that can be found in [52], we compared the modified Algencan with Ipopt [48], a well established interior point method for constrained optimization that is not prepared for optimal convergence to points that minimize infeasibility. The practical interpretation of the theory presented in this paper was corroborated by the performance of the modified Algencan on infeasible problems, that outperformed the interior point algorithm. Of course, this does not mean that Algencan should be more efficient than Ipopt in different (perhaps feasible) situations.

5.2 Large-scale infeasible problems

In a third set of experiments, we analyzed the performance of the modified Algencan method on a set of large-scale infeasible problems related to image reconstruction (see, for example, [12, 28]). Consider a signal $t \in \mathbb{R}^n$ for which it is known that it has a sparse representation in the space generated by the columns of a matrix $\Psi^T \in \mathbb{R}^{n\times n}$, i.e. there exists $s \in \mathbb{R}^n$ with a few non-null elements such that $t = \Psi^T s$. Let $\Phi \in \mathbb{R}^{m\times n}$ be a matrix that represents the task of performing $m < n$ observations $b \in \mathbb{R}^m$ that are linear combinations of the elements of the n-dimensional signal t, i.e. $b = \Phi t$. The problem of recovering the signal t from the observations b consists of finding a vector $s \in \mathbb{R}^n$ with the smallest possible number of non-null elements such that $\Phi\Psi^T s = b$. Matrices Φ and Ψ are implicitly defined sparse matrices such that computing a single matrix-vector product is very cheap, but $A = \Phi\Psi^T$ may be dense and computing it may be very expensive. In Compressive Sensing, the basis pursuit problem [24] consists on

$$\text{Minimize } \|s\|_1 \text{ subject to } \Phi\Psi^T s = b. \tag{36}$$

In Algencan, the Augmented Lagrangian subproblems are solved (at Step 1 of Algorithm 2.1) with an active-set method [17] that uses spectral projected gradients [20] for leaving the faces. Within the faces, there are several alternatives for tackling the "unconstrained" subproblems. When applying Algencan to problem (36) (or its reformulation below), since computing the full matrix $\Phi\Psi^T$ is computationally unaffordable, a truncated Newton approach will be considered. Therefore, Newtonian systems will be solved by conjugate gradients and only matrix-vector products will be required.

By the change of variables $s = u - v$ with $u \geq 0$ and $v \geq 0$, problem (36) can be reformulated as the linear programming problem given by

$$\text{Minimize } \sum_{i=1}^{n} u_i + v_i \text{ subject to } \Phi\Psi^T(u - v) = b,\ u \geq 0,\ v \geq 0. \tag{37}$$

Consider the "original" signal $t \in \mathbb{R}^n$ with $n = 256 \times 256 = 65{,}536$ pixels (each one representing a grey-scale between 0 and 1), depicted on Figure 1. Let Ψ be the $n \times n$ (orthogonal) Haar transform matrix (see, for example, [29, Chapter 6]) and let $\Phi \in \mathbb{R}^{m\times n}$ be the (full rank) highly sparse binary permuted block diagonal measurement matrix (with $L = 2$ block diagonal matrices; see [34] for details). The sparse representation of t in the space generated by the columns of Ψ^T is given by $s = \Psi t$ and it has only $k = 6{,}391$ non-null elements, which corresponds to a density of $k/n \times 100\% \approx 9.75\%$. Consider eight different linear programming problems of the form (37) with $m = \alpha k$ and $\alpha \in \{1.5, 2.0, 2.5, \dots, 5.0\}$. The problems have $2n = 131{,}072$ variables and m equality constraints plus the bound constraints. Starting from the least-squares solution $(u^0, v^0) = ([s^0]_+, [-s^0]_+)$, with $s^0 = \Theta^T(\Theta\Theta^T)^{-1}b$ and $\Theta = \Phi\Psi^T$, Algencan (with all its default parameters, that include a tolerance $\varepsilon_{\rm feas} = 10^{-8}$ for feasibility) solves the eight (feasible) problems using much less time than the time needed to perform the strongly unrecommended task of computing the matrix Θ only once. Figure 2 illustrates the solutions found considering different numbers of measurements m.

Figure 1: Phantom "original" image with $n = 256 \times 256 = 65{,}536$ pixels.

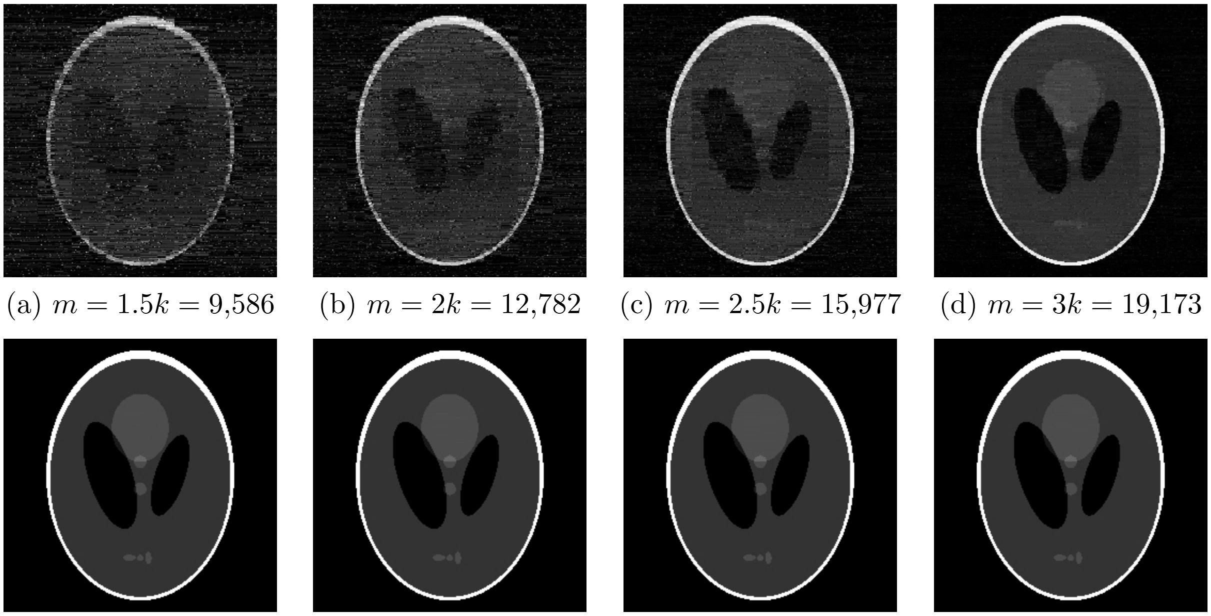
The key point in the numerical experiments described in the paragraph above is that noise *was not* considered at all in the process of obtaining the observations b. Assume now that there is noise in the acquisition of the observations b. In this case, it may be expected the linear system $\Phi\Psi^T s = b$ to be incompatible. If the level of noise is known or can be estimated, the problem to be solved may be given by

$$\text{Minimize } \|s\|_1 \text{ subject to } \|\Phi\Psi^T s - b\|_2 \le \sigma, \tag{38}$$

where σ is an estimation of the Euclidean norm of the residual of the presumable incompatible linear system $\Phi\Psi^T s = b$. Problem (38) is known as basis pursuit denoise problem [12]. If the level of noise of the observations is unknown, a sequence of problems of the form (38), considering different values for σ, may be solved. This alternative is strongly related to a second alternative that consists of solving a sequence of problems of the form

$$\text{Minimize } \|s\|_1 + \gamma\|\Phi\Psi^T s - b\|_2^2, \tag{39}$$

where γ is a parameter. Solving sequences of problems of the form (38) or (39) can be seen as applying classical scalarization methods such as the ϵ-constraint method or the weights method,



(a) $m = 1.5k = 9{,}586$ (b) $m = 2k = 12{,}782$ (c) $m = 2.5k = 15{,}977$ (d) $m = 3k = 19{,}173$

(e) $m = 3.5k = 22{,}368$ (f) $m = 4k = 25{,}564$ (g) $m = 4.5k = 28{,}759$ (h) $m = 5k = 31{,}955$

Figure 2: Recovered images obtained as solutions of the linear programming problem (37) with different number of constraints (observations).

respectively, to an underlying biobjective optimization problem. See, for example, [12, 38] and the references therein. In the present numerical experiments, we explore a different alternative: trying "to solve" with the modified Algencan method an infeasible problem of the form (36).

Assume that observations b are of the form $b = \Phi t + \eta$, where $\eta_i \in [-\varepsilon_i, \varepsilon_i]$ are random variables with uniform distribution and that the levels of noise ε_i are unknown. Moreover, to be sure that we are in fact dealing with an empty feasible set, we consider that observations are given by $\tilde{b} = \Phi t + \tilde{\eta}$ and $\hat{b} = \Phi t + \hat{\eta}$, where $\tilde{\eta}$ and $\hat{\eta}$ are random variables as already described. In this case, the set of $2m$ constraints that express the desire of finding a signal in accordance with the observations is given by the incompatible set of linear equations

$$\begin{pmatrix} \Phi\Psi^T \\ \Phi\Psi^T \end{pmatrix} s = \begin{pmatrix} \tilde{b} \\ \hat{b} \end{pmatrix}. \tag{40}$$

Thus, the (infeasible) linear programming problem that should be solved is given by

$$\begin{array}{ll} \text{Minimize} & \sum_{i=1}^n u_i + v_i \\ \text{subject to} & \Phi\Psi^T(u - v) = \tilde{b}, \\ & \Phi\Psi^T(u - v) = \hat{b}, \\ & u \geq 0, \ v \geq 0. \end{array} \tag{41}$$

We considered instances of problem (41) with $m = 3.5k = 22{,}368$ and three different levels of noise $\varepsilon_i = \xi |b_i|$ for $i = 1, \dots, m$, with $\xi \in \{0.01, 0.1, 0.5\}$. The considered initial point was the least-squares solution $(u^0, v^0) = ([s^0]_+, [-s^0]_+)$, with $s^0 = \Theta^T(\Theta\Theta^T)^{-1}\tilde{b}$. Figure 3 shows the "solutions" found. In all cases, as expected, the objective function seems to have been minimized subject to minimal infeasibility. "Solutions" found should be compared with the one depicted in Figure 2e, that corresponds to the case $\xi = 0$. In that case, the solution found is feasible with tolerance $\varepsilon_{\text{feas}} = 10^{-8}$ for the sup-norm of the infeasibility and the objective function value (ℓ_1-norm of s) at the solution is $f(s^*) \approx 3303.27$. For the infeasible problems with $\xi \in \{0.01, 0.1, 0.5\}$, the sup-norm of the infeasibility at the "solutions" was of the order of 10^{-2}, 0.1, and 1, respectively, while the optimal objective function values were approx. 3313.10, 3431.25, and 4238.36, respectively. The values of the infeasibility at the "solutions" were verified by solving, independently, the problem of minimizing the squared Euclidean norm of the linear constraints of problem (41) subject to the bound constraints. Note that, without estimating the level of noise of the observations' acquisition, "solutions" obtained by the modified Algencan applied to problem (41) appear to be similar to the ones obtained by Algencan when applied to problem (37). While this experiments suggest that a method that minimizes the objective function subject to minimum infeasibility may be a useful tool in the field of images reconstruction, the reader should be warned that such a claim is out of the scope of the present work, where image reconstruction problems were used to illustrate the practical behaviour of the modified Algencan method.

(a) 1% ($\xi = 0.01$) (b) 10% ($\xi = 0.1$) (c) 50% ($\xi = 0.5$)

Figure 3: Recovered images obtained as "solutions" of the infeasible problem (41) with different levels of noise $\varepsilon_i = \xi|b_i|$ for $i = 1, \ldots, m$ in the generation of $\tilde{b}$ and $\hat{b}$.

6 Final Remarks

In many practical situations we do not know whether the constrained optimization problem that we try to solve is feasible or not. In these cases it is advantageous to use algorithms that quickly detect possible infeasibility. However, there are cases in which it is relevant to distinguish between different infeasible points. In those cases, under similar levels of infeasibility, one may prefer the points at which the objective function value is as small as possible. This is a motivation for studying "optimality properties" of algorithms that could converge to infeasible points. In the context of global optimization, in which we assume that it is possible to solve simple subproblems up to global minimization with arbitrary precision, it is simple to develop Augmented Lagrangian algorithms that converge to minimizers of the objective function subject to minimal infeasibility. However, these algorithms are in general not affordable if the number of variables or constraints is large. It is interesting, therefore, to study the properties of affordable algorithms (which generally converge to feasible points and satisfy first-order optimality conditions) in the case that convergence to a feasible point does not occur. In this paper we addressed this task in the case of an Augmented Lagrangian method. Although, in this case, it is impossible to prove convergence to global minimizers subject to minimal infeasibility, we were able to detect a simpler feasible auxiliary problem such that the proposed algorithm converges to the auxiliary problem's feasible stationary points. Modifications made on an standard Augmented Lagrangian method do not affect its classical convergence properties and the properties of the so far modified algorithm were studied from the point of view of global convergence. In order to corroborate the theoretical properties, we performed some numerical experiments with geometric appeal, in which we illustrated that behaviour of the analyzed Augmented Lagrangian method in infeasible problems when the criterion of minimal objective function is considered to be relevant.

References

[1] 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, pp. 1286–1309, 2007.

[2] 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, pp. 5–32, 2008.

[3] R. Andreani, G. Haeser, and J. M. Martínez, On sequential optimality conditions for smooth constrained optimization, *Optimization* 60, pp. 627–641, 2011.

[4] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, A relaxed constant positive linear dependence constraint qualification and applications, to appear in *Mathematical Programming* (DOI: 10.1007/s10107-011-0456-0).

[5] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, Two new weak constraint qualifications and applications, *SIAM Journal on Optimization* 22, pp. 1109–1135, 2012.

[6] 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, pp. 473–485, 2005.

[7] 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, pp. 3533–3554, 2010.

[8] M. Andretta, E. G. Birgin, and J. M. Martínez, Partial spectral projected gradient method with active-set strategy for linearly constrained optimization, *Numerical Algorithms* 53, pp. 23–52, 2010.

[9] C. Audet and J. E. Dennis Jr., Mesh adaptive direct search algorithms for constrained optimization, *SIAM Journal on Optimization* 17, pp. 188–217, 2006.

[10] C. Audet and J. E. Dennis Jr., A progressive barrier for derivative-free nonlinear programming, *SIAM Journal on Optimization* 20, pp. 445–472, 2009.

[11] H. Y. Benson, D. F. Shanno, and R. J. Vanderbei, Interior-point methods for nonconvex nonlinear programming: Filter methods and merit functions, *Computational Optimization and Applications* 23, pp. 257–272, 2002.

[12] E. van den Berg and M. P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, *SIAM Journal on Scientific Computing* 31, pp. 890–912, 2008.

[13] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts, 1999.

[14] E. G. Birgin, R. Castillo and J. M. Martínez, Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems, *Computational Optimization and Applications* 31, pp. 31–55, 2005.

[15] E. G. Birgin, D. Fernández, and J. M. Martínez, The boundedness of penalty parameters in an Augmented Lagrangian method with constrained subproblems, *Optimization Methods and Software* 27, pp. 1001–1024, 2012.

[16] 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, pp. 139–162, 2010.

[17] E. G. Birgin and J. M. Martínez, Large-scale active-set box-constrained optimization method with spectral projected gradients, *Computational Optimization and Applications* 23, pp. 101–125, 2002.

[18] E. G. Birgin and J. M. Martínez, Improving ultimate convergence of an Augmented Lagrangian method, *Optimization Methods and Software* 23, pp. 177–195, 2008.

[19] E. G. Birgin, J. M. Martínez, and L. F. Prudente, Augmented Lagrangians with possible infeasibility and finite termination for global nonlinear programming, *Journal of Global Optimization* 58, pp. 207–242, 2014.

[20] E. G. Birgin, J. M. Martínez, and M. Raydan, Nonmonotone spectral projected gradient methods on convex sets, *SIAM Journal on Optimization* 10, pp. 1196–1211, 2000.

[21] R. H. Byrd, F. E. Curtis, and J. Nocedal, Infeasibility detection and SQP methods for nonlinear optimization, *SIAM Journal on Optimization* 20, pp. 2281–2299, 2010.

[22] R. H. Byrd, J. Ch. Gilbert, and J. Nocedal, A trust region method based on interior point techniques for nonlinear programming, *Mathematical Programming* 89, pp. 149–185, 2000.

[23] R. H. Byrd, J. Nocedal, and R.A. Waltz, KNITRO: An integrated package for nonlinear optimization, in *Large-Scale Nonlinear Optimization*, G. Di Pillo and M. Roma eds., pp. 35–59. Springer US, 2006.

[24] S. S. Chen, D. L. Donoho, and M. A. Saunders, Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing* 20, pp. 33–61, 1998.

[25] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[26] Z. Dostál, *Optimal Quadratic Programming Algorithms*, Springer, New York, 2009.

[27] Z. Dostál, A. Friedlander, and S. A. Santos, Augmented Lagrangians with adaptive precision control for quadratic programming with simple bounds and equality constraints, *SIAM Journal on Optimization* 13, pp. 1120–1140, 2003.

[28] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE Journal of Selected Topics in Signal Processing* 1, pp. 586–597, 2007.

[29] P. Van Fleet, *Discrete Wavelet Transformations: An Elementary Approach with Applications*, John Wiley and Sons Inc., New York, 2008.

[30] R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint, and A. Wächter, Global convergence of trust-region SQP-filter algorithms for general nonlinear programming, *SIAM Journal on Optimization* 13, pp. 635–659, 2002.

[31] R. Fletcher and S. Leyffer, Nonlinear programming without a penalty function, *Mathematical Programming* 91, pp. 239–269, 2002.

[32] R. Fletcher, S. Leyffer, and Ph. L. Toint, On the global convergence of a filter-SQP algorithm, *SIAM Journal on Optimization* 13, pp. 44–59, 2002.

[33] N. I. M. Gould, D. Orban, and Ph. L. Toint, CUTEr and SifDec: A Constrained and Unconstrained Testing Environment, revisited, *ACM Transactions on Mathematical Software* 29, pp. 373–394, 2003.

[34] Z. He, T. Ogawa, and M. Haseyama, The simplest measurement matrix for compressed sensing of natural images, in *17th IEEE International Conference on Image Processing (ICIP)*, pp. 4301–4304, 2010.

[35] M. R. Hestenes, Multiplier and gradient methods, *Journal of Optimization Theory and Applications* 4, pp. 303–320, 1969.

[36] S. Lu, Implications of the constant rank constraint qualification, *Mathematical Programming* 126, pp. 365–392, 2009.

[37] S. Lu, Relation between the constant rank and the relaxed constant rank constraint qualifications, *Optimization* 61, pp. 555–566, 2012.

[38] K. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, 1999.

[39] J. M. Martínez and L. F. Prudente, Handling infeasibility in a large-scale nonlinear optimization algorithm, *Numerical Algorithms* 60, pp. 263–277, 2012.

[40] L. Minchenko and S. Stakhovski, On relaxed constant rank regularity conditions in mathematical programming, *Optimization* 60, pp. 429–440, 2011.

[41] L. Minchenko and S. Stakhovski, Parametric nonlinear programming problems under the relaxed constant rank condition, *SIAM Journal on Optimization* 1, pp. 314–332, 2011.

[42] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in *Optimization*, R. Fletcher ed., Academic Press, New York, NY, pp. 283–298, 1969.

[43] L. F. Prudente, Augmented Lagrangian approaches in Nonlinear Programming, Doctoral Dissertation, State University of Campinas, Campinas, SP, Brazil, 2012.

[44] L. Qi and Z. Wei, On the constant positive linear dependence condition and its application to SQP methods, *SIAM Journal on Optimization* 10, pp. 963–981, 2000.

[45] R. T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, *SIAM Journal on Control and Optimization* 12, pp. 268–285, 1974.

[46] C. Shen, S. Leyffer, and R. Fletcher, A nonmonotone filter method for nonlinear optimization, *Computational Optimization and Applications* 52, pp. 583–607, 2012.

[47] C. Shen, W. Xue, and D. Pu, A filter SQP algorithm without a feasibility restoration phase, *Computational and Applied Mathematics* 28, pp. 167–194, 2009.

[48] A. Wächter and L. T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Mathematical Programming* 106, pp. 25–57, 2006.

[49] `http://www.ime.usp.br/~egbirgin/tango/`

[50] `http://www.ime.unicamp.br/~martinez/algencancuteprudente.pdf`

[51] `http://www.ime.unicamp.br/~martinez/algencanprodutocute.pdf`

[52] `http://www.ime.unicamp.br/~martinez/experimentsmpru3.pdf`