

# A new sequential optimality condition for constrained optimization and algorithmic consequences <sup>\*</sup>

Roberto Andreani <sup>†</sup>      J. M. Martínez <sup>‡</sup>      B. F. Svaiter <sup>§</sup>

March 27, 2010.

## Abstract

Necessary first-order sequential optimality conditions provide adequate theoretical tools to justify stopping criteria for nonlinear programming solvers. Sequential optimality conditions are satisfied by local minimizers of optimization problems independently of the fulfillment of constraint qualifications. A new condition of this type is introduced in the present paper. It will be proved that a well established Augmented Lagrangian algorithm produces sequences whose limits satisfy the new condition. Practical consequences will be discussed.

**Key words:** Nonlinear Programming, Optimality Conditions, Approximate KKT conditions, Stopping criteria.

**AMS Subject Classification:** 90C30, 49K99, 65K05.

## 1 Introduction

Practical algorithms for solving large nonlinear programming problems are iterative. Consequently, implementations include stopping criteria that generally indicate when the current iterate is close to a solution. Computer codes usually test the approximate fulfillment of the KKT conditions [23]. This means that, with some small tolerance, one tests whether the point is feasible, the gradient of the Lagrangian is null and complementarity conditions are satisfied. This procedure is theoretically justified because it is possible to prove that every local minimizer is the limit of a sequence of points that satisfy the approximate KKT test with tolerances going to zero [2, 25]. In terms of [2], this means that every local minimizer satisfies an Approximate KKT condition (AKKT) [2, 25]. This property holds independently of the fulfillment of constraint qualifications and even in the case that the local minimizer does not satisfy the exact KKT conditions. For example, in the problem of minimizing  $x$  subject to  $x^2 = 0$  the solution  $x^* = 0$  does not satisfy KKT but satisfies AKKT.

In critical situations, the mere fulfillment of a KKT approximate criterion may lead to wrong conclusions. Consider the problem

$$\text{Minimize } \frac{(x_2 - 2)^2}{2} \text{ subject to } x_1 = 0, x_1 x_2 = 0. \quad (1) \quad \{\mathbf{p1}\}$$

---

<sup>\*</sup>This work was supported by PRONEX-CNPq/FAPERJ Grant E-26/171.164/2003 - APQ1, FAPESP Grants 05/02163/8 and 06/53768-0, CNPq (480101/2008-6, 303583/2008-8) and FAPERJ: E-26/102.821/2008.

<sup>†</sup>Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas SP, Brazil. E-mail: andreani@ime.unicamp.br

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

<sup>§</sup>Institute of Pure and Applied Mathematics (IMPA), Rio de Janeiro, Brazil.

The solution of this problem is  $(0, 2)^T$ . Consider the point  $(\varepsilon, 1)^T$  for  $\varepsilon > 0$ , small. The gradient of the objective function at this point is  $(0, -1)^T$  and the gradients of the constraints are  $(1, 0)^T$  and  $(1, \varepsilon)^T$ . Therefore, the gradient of the objective function is a linear combination of the gradients of the constraints with coefficients  $1/\varepsilon$  and  $-1/\varepsilon$ . Moreover, the point  $(\varepsilon, 1)^T$  is almost feasible in the sense that the sup-norm of the constraints vector is  $\varepsilon$ . This means that for arbitrarily small  $\varepsilon > 0$ , the point  $(\varepsilon, 1)^T$  fulfills any sensible practical KKT test. This simple example suggests that stronger requirements are necessary to declare practical convergence of numerical optimization methods and that, consequently, stronger sequential optimality conditions should be encountered. Unfortunately, the Approximate Gradient Projection (AGP) and Linear Approximate Gradient Projection (L-AGP) conditions introduced in [22] and [2] are also satisfied at the wrong point  $(0, 1)^T$  and stopping tests based on them would be misleading too.

In this paper we introduce the ‘‘Complementary Approximate KKT’’ (CAKKT) sequential optimality condition as a remedy for situations as the one described above. In the CAKKT stopping test we require, in addition to the usual approximate KKT test, that the product of each multiplier with the corresponding constraint value must be small. In Section 3 we will see that this requirement is not satisfied in the example above, therefore the ‘‘wrong’’ point  $(0, 1)^T$  does not satisfy CAKKT.

The role of sequential optimality conditions in practical optimization may be better understood by means of the comparison with classical ‘‘punctual’’ optimality conditions. When one uses an iterative algorithm to solve a constrained optimization problem we need to decide if a computed iterate is an acceptable solution or not. A punctual optimality condition is necessarily satisfied at a local minimizer but not at ‘‘approximately local minimizers’’. Since, in the iterative framework, one never gets exact solutions, numerical practice always leads to test the ‘‘approximate fulfillment’’ of punctual optimality conditions. More specifically, practical codes usually test relaxed versions of the KKT conditions. This could be a paradoxical decision, because KKT conditions do not need to be fulfilled at local minimizers, if constraint qualifications do not hold [7]. Nevertheless, computational practice can be justified because it is possible to show that any local minimizer has the property of having arbitrarily close neighbors that ‘‘approximately fulfill’’ KKT conditions [2]. Now, the approximate fulfillment of KKT conditions may have different definitions. The implementation of the strong definition given in this paper may help optimization solvers avoid stopping at false approximate minimizers.

This paper is organized as follows:

In Section 2 we survey some results on sequential optimality conditions that will be useful to understand the main results of the paper. In Section 3 we define rigorously the new condition and we prove that local minimizers necessarily satisfy it. In Section 4 we prove that CAKKT is stronger than the AGP condition given in [2, 22] and we show that CAKKT is a sufficient optimality condition in convex problems. In Section 5 we prove that the Augmented Lagrangian method with lower-level constraints defined in [1] produces CAKKT sequences if one assumes that a sum-of-squares infeasibility measure satisfies the Kurdyka-Lojasiewicz inequality [17, 18]. Conclusions will be stated in the final section of this paper.

**Notation.**

- $\mathcal{N} = \{0, 1, 2, \dots\}$ .
- The symbol  $\|\cdot\|$  denotes an arbitrary norm.
- If  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  we denote  $\nabla h = (\nabla h_1, \dots, \nabla h_m)$ .
- $\mathbb{R}_+ = \{t \in \mathbb{R} \mid t \geq 0\}$ .
- If  $v \in \mathbb{R}^n$ , we denote  $v_+ = (\max\{v_1, 0\}, \dots, \max\{v_n, 0\})^T$ .

- If  $v \in \mathbb{R}^n$ , we denote  $v_- = (\min\{v_1, 0\}, \dots, \min\{v_n, 0\})^T$ .
- If  $a \in \mathbb{R}$ , we denote  $a_+^2 = (a_+)^2 = a_+ a_+$ .
- $B(x, \delta) = \{z \in \mathbb{R}^n \mid \|z - x\| \leq \delta\}$ .
- $P_\Omega(x)$  is the Euclidean projection of  $x$  on  $\Omega$ .

## 2 Preliminaries

{background}

We consider the Nonlinear Programming problem in the form

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0, \quad g(x) \leq 0, \quad (2) \quad \{\text{nlp}\}$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$  have continuous first derivatives.

The KKT conditions are fulfilled at a point  $x \in \mathbb{R}^n$  if

$$h(x) = 0, \quad g(x) \leq 0, \quad (3) \quad \{\text{feasibles}\}$$

and there exist  $\lambda \in \mathbb{R}^m$ ,  $\mu \in \mathbb{R}_+^p$  such that

$$\nabla f(x) + \nabla h(x)\lambda + \nabla g(x)\mu = 0, \quad (4) \quad \{\text{lagranja}\}$$

where  $\mu_i = 0$  for all  $i$  such that  $g_i(x) < 0$ .

Many reformulations of the KKT conditions are possible. For example, the KKT conditions hold at  $x \in \mathbb{R}^n$  if and only if there exist  $\lambda \in \mathbb{R}^m$ ,  $\mu \in \mathbb{R}^p$  such that  $h(x) = 0$ , (4) takes place and

$$\min\{-g_i(x), \mu_i\} = 0 \quad \text{for all } i = 1, \dots, p. \quad (5) \quad \{\text{minimocom}\}$$

Clearly, (5) is equivalent to

$$g_i(x) \leq 0, \quad \mu_i \geq 0, \quad \mu_i g_i(x) = 0 \quad \text{for all } i = 1, \dots, p. \quad (6) \quad \{\text{prodcom}\}$$

Usually, nonlinear programming algorithms stop declaring “success” when KKT conditions are approximately satisfied. However, although different formulations of the KKT conditions are equivalent, the *approximate* KKT fulfillment depends on the formulation. For example, given a small positive tolerance  $\varepsilon > 0$ , (5) induces a stopping criterion given by:

$$\|\nabla f(x) + \nabla h(x)\lambda + \nabla g(x)\mu\| \leq \varepsilon, \quad (7) \quad \{\text{stoplagraco}\}$$

$$\|h(x)\| \leq \varepsilon \quad (8) \quad \{\text{stophache}\}$$

and

$$|\min\{-g_i(x), \mu_i\}| \leq \varepsilon, \quad i = 1, \dots, p. \quad (9) \quad \{\text{stopminimo}\}$$

On the other hand, the stopping criterion induced by (6) involves (7), (8) and

$$g_i(x) \leq \varepsilon, \quad \mu_i \geq -\varepsilon, \quad |g_i(x)\mu_i| \leq \varepsilon, \quad i = 1, \dots, p. \quad (10) \quad \{\text{stoprodu}\}$$

In practice, different tolerances are used for testing the approximate fulfillment of different relations.

Stopping criteria based on approximate fulfillment of the KKT conditions make sense even in the case that exact KKT conditions in the limit do not hold. The reason is that one is generally able to prove that, given a local minimizer  $x^*$  and an arbitrary tolerance  $\varepsilon > 0$ , there exists a point

$x$  such that  $\|x - x^*\| \leq \varepsilon$  and the approximate KKT criterion is fulfilled with tolerance  $\varepsilon$ . The fulfillment of KKT conditions at a local minimizer depends on constraint qualifications, but their approximate fulfillment at an arbitrarily close point does not [2, 22].

We will formulate these concepts in terms of “sequential optimality conditions”. The Approximate KKT condition (AKKT) introduced in [2] corresponds to the approximate KKT criterion based on (7), (8) and (9). We say that a feasible point  $x^*$  satisfies AKKT if there exist sequences  $\{x^k\} \subset \mathbb{R}^n$ ,  $\{\lambda^k\} \subset \mathbb{R}^m$ ,  $\{\mu^k\} \subset \mathbb{R}^p$  such that

$$\lim_{k \rightarrow \infty} x^k = x^*, \quad (11) \quad \{\text{limitexk}\}$$

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| = 0 \quad (12) \quad \{\text{lagracok}\}$$

and

$$\lim_{k \rightarrow \infty} \min\{-g_i(x^k), \mu_i^k\} = 0, \quad \text{for all } i = 1, \dots, p. \quad (13) \quad \{\text{complemenk}\}$$

In [2] it has been proved that every local minimizer of (2) necessarily satisfies AKKT.

We say that  $x^*$  satisfies the optimality condition AKKT(strong) if there exist sequences  $\{x^k\} \subset \mathbb{R}^n$ ,  $\{\lambda^k\} \subset \mathbb{R}^m$ ,  $\{\mu^k\} \subset \mathbb{R}_+^p$  such that (11), (12) are verified and, in addition,

$$\lim_{k \rightarrow \infty} \mu_i^k g_i(x^k) = 0 \quad \text{for all } i = 1, \dots, p. \quad (14) \quad \{\text{prodacero}\}$$

The condition (14) corresponds to the requirement (10). It is easy to see that AKKT(strong) implies AKKT but the reciprocal is not true.

In the present paper we present an even stronger optimality condition called CAKKT (Complementary Approximate KKT condition). In addition to (11), (12) and (14), CAKKT requires that

$$\lim_{k \rightarrow \infty} \lambda_i^k h_i(x^k) = 0 \quad \text{for all } i = 1, \dots, m. \quad (15) \quad \{\text{prodaceroh}\}$$

Clearly, CAKKT is even stronger than AKKT(strong) since the latter does not make any requirement with respect to complementarity related to equality constraints.

In order to illustrate the strength of CAKKT, we will recall in this paper the Approximate Gradient Projection (AGP) optimality condition introduced by Martínez and Svaiter in [22]. We say that the feasible point  $x^*$  satisfies AGP if there exists a sequence  $\{x^k\}$  such that

$$\lim_{k \rightarrow \infty} \|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\| = 0, \quad (16) \quad \{\text{defiagp}\}$$

where:

$$\Omega_k = \{x \in \mathbb{R}^n \mid \nabla h(x^k)^T(x - x^k) = 0, \min\{g_i(x^k), 0\} + \nabla g_i(x^k)^T(x - x^k) \leq 0, i = 1, \dots, p\}. \quad (17) \quad \{\text{omegak}\}$$

The attractiveness of AGP is that it does not involve Lagrange multipliers estimates. If  $x^k$  also satisfies all the linear (equality or inequality) constraints involved in (2) we say that the Linear AGP condition (L-AGP) is satisfied. Properties of L-AGP are discussed in [2].

All the sequential optimality conditions discussed above are obviously satisfied if the KKT conditions hold. The interesting property is that they are also satisfied at local minimizers when - due to the lack of fulfillment of constraint qualifications - the KKT conditions do not hold. This state of facts suggests that sequential optimality conditions may be contemplated from two different points of view.

1. Proximity to local minimizers: The CAKKT condition is stronger than other approximate versions of the KKT conditions. This is essentially due to the presence of a complementarity requirement involving equality constraints. In the presence of a nonlinear programming (2) users are interested in finding minimizers, not merely stationary points. We will show that the set of local minimizers is contained in the set of CAKKT points and the set of CAKKT points is contained in the set of points that satisfy other approximate KKT conditions. This suggests that, when we approximate a CAKKT point, we have better chances of being close to a local minimizer than when we approximate KKT in some alternative way. This argument is essentially qualitative and we do not have any quantitative evidence, up to now, of its accuracy. In particular, we do not know how to estimate the distance of  $x$  to a local minimizer, given the degree of fulfillment of CAKKT at  $x$ . However, the set-theoretical argument displayed above suggests that this is an issue that deserves both practical and theoretical investigation. Observe that, in the case of problem (1), the CAKKT condition detects that  $(\varepsilon, 1)^T$  is not close to the solution  $(0, 2)^T$  but the AKKT condition does not.

In this paper we are not concerned with the fact that a nonlinear programming problem may be formulated in different ways, giving rise to different objective functions and constraints. Some reformulations may have better conditioning properties than others. Our arguments here assume a given formulation. Interaction between reformulations and effectiveness of sequential optimality conditions remain to be investigated.

2. Efficiency of algorithms: When a practical algorithm for solving (2) converges to a feasible point that satisfies some constraint qualification, this feasible point is generally guaranteed to fulfill the KKT conditions. More generally, limit points of practical algorithms usually satisfy some sequential optimality condition, although only a few methods have been analyzed from this point of view up to now. In [2] it was shown that not all the sequential optimality conditions are equivalent. Here we will show that a particular sequential optimality condition (CAKKT) is strictly stronger than the others we know, its strength being mainly represented by approximate complementarity with respect to equality constraints. The question arises whether particular algorithms are convergent to such CAKKT points. The obvious conjecture is that algorithms having this convergence property have more chances to converge to minimizers. Of course, many other factors are involved in algorithmic efficiency. Nevertheless, it is interesting to detect theoretical properties having practical interpretations that may be corroborated (or not) by experiments.

### 3 CAKKT is a necessary optimality condition

{dos}

In this section we prove that CAKKT is a necessary optimality condition independently of the fulfillment of constraint qualifications.

As in (11), (12) and (14), we say that  $x^* \in \mathbb{R}^n$  fulfills the Complementary Approximate-KKT condition (CAKKT) for problem (2) if

$$h(x^*) = 0, g(x^*) \leq 0$$

and there exists a sequence  $\{x^k\}$  that converges to  $x^*$  and satisfies:

- For all  $k \in \mathbb{N}$  there exist  $\lambda^k \in \mathbb{R}^m, \mu^k \in \mathbb{R}_+^p$  such that

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| = 0 \tag{18} \quad \{\text{sakkt1}\}$$

and

$$\lim_{k \rightarrow \infty} \sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{i=1}^p |\mu_i^k g_i(x^k)| = 0. \quad (19) \quad \{\text{sakkt3}\}$$

Points that satisfy the CAKKT condition will be called ‘‘CAKKT points’’. If  $x^*$  satisfies the KKT conditions necessarily fulfills CAKKT, taking  $x^k = x^*$ ,  $\lambda^k = \lambda^*$ ,  $\mu^k = \mu^*$  for all  $k \in \mathbb{N}$ , where  $\lambda^* \in \mathbb{R}^m$ ,  $\mu^* \in \mathbb{R}_+^p$  are the Lagrange multipliers associated with  $x^*$ . The interesting cases are the ones in which the KKT conditions do not hold.

In the following lemma we show that the non-negativity of  $\mu^k$  in the definition of CAKKT can be relaxed.

**Lemma 3.1.** *A feasible point  $x^*$  satisfies the CAKKT condition if and only if there exist sequences  $\{x^k\} \subset \mathbb{R}^n$ ,  $\{\lambda^k\} \subset \mathbb{R}^m$ ,  $\{\mu^k\} \subset \mathbb{R}^p$  such that  $\lim_{k \rightarrow \infty} x^k = x^*$ , (18) and (19) hold and, in addition, there exists a nonnegative sequence  $\varepsilon_k$  that tends to zero such that*

$$\mu_i^k \geq -\varepsilon_k \quad \text{for all } i = 1, \dots, p, k \in \mathbb{N}. \quad (20) \quad \{\text{relaxed}\}$$

*Proof.* The fact that CAKKT implies (20) is trivial. On the other hand, if (20) holds, by the continuity of  $\nabla g$ , it is easy to see that (18) and (19) remain true replacing  $\mu_i^k$  by  $\max\{\mu_i^k, 0\}$ .  $\square$

**Lemma 3.2.** *Assume that the feasible point  $x^*$  satisfies the CAKKT condition. Then, there exist sequences  $\{x^k\} \subset \mathbb{R}^n$ ,  $\{\lambda^k\} \subset \mathbb{R}^m$ ,  $\{\mu^k\} \subset \mathbb{R}_+^p$  such that  $\lim_{k \rightarrow \infty} x^k = x^*$ , (18) and (19) hold and, in addition,*

$$\mu_i^k = 0 \quad \text{for all } i \text{ such that } g_i(x^*) < 0. \quad (21) \quad \{\text{sakkt2}\}$$

*Proof.* Assume that  $x^*$  satisfies CAKKT and let  $x^k, \lambda^k, \mu^k$  be such that (18) and (19) hold.

If  $g_i(x^*) < 0$ , then, by the continuity of  $g_i$  and (19) one has that

$$\lim_{k \rightarrow \infty} \mu_i^k = 0. \quad (22) \quad \{\text{limu}\}$$

Define, for all  $i$  such that  $g_i(x^*) < 0$ ,  $k \in \mathbb{N}$ ,  $\tilde{\mu}_i^k = 0$ . Clearly, (19) and (21) hold if one replaces  $\mu_i^k$  by  $\tilde{\mu}_i^k$ . Moreover, by (22) and the boundedness of  $\nabla g_i(x^k)$ , (18) also holds replacing  $\mu_i^k$  by  $\tilde{\mu}_i^k$ . This completes the proof.  $\square$

Properties (18) and (19) provide the natural stopping criterion associated with CAKKT. Given small positive tolerances  $\varepsilon_{feas}, \varepsilon_{opt}, \varepsilon_{mult}$  corresponding to feasibility, optimality (18), and the new condition (19), an algorithm that aims to solve (2) should be stopped declaring ‘‘convergence’’ when, for suitable multipliers  $\lambda^k \in \mathbb{R}^m$ ,  $\mu^k \in \mathbb{R}_+^p$ ,

$$\|h(x^k)\| \leq \varepsilon_{feas}, \|g(x^k)_+\| \leq \varepsilon_{feas}, \quad (23) \quad \{\text{stop1}\}$$

$$\|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| \leq \varepsilon_{opt} \quad (24) \quad \{\text{stop2}\}$$

and

$$|\lambda_i^k h_i(x^k)| \leq \varepsilon_{mult}, |\mu_j^k g_j(x^k)| \leq \varepsilon_{mult} \quad \text{for all } i = 1, \dots, m, j = 1, \dots, p. \quad (25) \quad \{\text{stop4}\}$$

In the nonlinear programming software Algenca [1]<sup>1</sup> and other Augmented Lagrangian algorithms [11] the convergence stopping criterion is given by (23), (24) and

$$\mu_i^k = 0 \quad \text{whenever } g_i(x^k) < -\varepsilon_{comp}. \quad (26) \quad \{\text{stop3}\}$$

<sup>1</sup>Algenca is available at [www.ime.usp.br/~egbirgin/tango](http://www.ime.usp.br/~egbirgin/tango).

In order to show that this criterion might not be sufficient to detect good approximations to the solution, let us come back to the example given in the introduction of this paper, where  $n = 2, m = 2, p = 0$ ,

$$f(x_1, x_2) = \frac{(x_2 - 2)^2}{2}. \quad (27) \quad \{\text{ex1a}\}$$

$$h_1(x_1, x_2) = x_1, h_2(x_1, x_2) = x_1 x_2. \quad (28) \quad \{\text{ex1b}\}$$

Taking  $\varepsilon_{feas} = \varepsilon_{opt} = \varepsilon_{comp} = \varepsilon_{mult} = \varepsilon$  and  $\|\cdot\| = \|\cdot\|_\infty$  one has that the point  $x^k = (\varepsilon, 1)^T$  satisfies (23), (24), (26) with  $\lambda_1^k = -1/\varepsilon$  and  $\lambda_2^k = 1/\varepsilon$ . However,

$$\lambda_1^k h_1(x^k) = -1, \quad \lambda_2^k h_2(x^k) = 1.$$

Therefore  $x^k$  does not fulfill (25) if  $\varepsilon_{mult} < 1$ .

The proof that CAKKT is a genuine necessary optimality condition is given below. This proof uses a penalty reduction technique employed in [2, 7, 16, 22] for analyzing different optimality conditions.

**Theorem 3.1.** *Let  $x^*$  be a local minimizer of (2). Then,  $x^*$  satisfies CAKKT.*

*Proof.* Let  $\delta > 0$  be such that  $f(x^*) \leq f(x)$  for all feasible  $x$  such that  $\|x - x^*\| \leq \delta$ . Consider the problem

$$\text{Minimize } f(x) + \|x - x^*\|_2^2 \quad \text{subject to } h(x) = 0, g(x) \leq 0, x \in B(x^*, \delta). \quad (29) \quad \{\text{nlpaux}\}$$

Clearly,  $x^*$  is the unique solution of (29). Let  $x^k$  be a solution of

$$\text{Minimize } f(x) + \|x - x^*\|_2^2 + \frac{\rho_k}{2} \left[ \|h(x)\|_2^2 + \sum_{i=1}^p g_i(x)_+^2 \right] \quad \text{subject to } x \in B(x^*, \delta). \quad (30) \quad \{\text{labelar}\}$$

By the compactness of  $B(x^*, \delta)$ ,  $x^k$  is well defined for all  $k$ . Moreover, since  $x^* \in B(x^*, \delta)$ ,  $h(x^*) = 0$  and  $g(x^*)_+ = 0$ ,

$$f(x^k) + \|x^k - x^*\|_2^2 + \frac{\rho_k}{2} \left[ \|h(x^k)\|_2^2 + \sum_{i=1}^p g_i(x^k)_+^2 \right] \leq f(x^*). \quad (31) \quad \{\text{lonuevo}\}$$

By the convergence theory of external penalty methods [12] one has that  $\lim_{k \rightarrow \infty} x^k = x^*$ . Therefore, by (31) and the continuity of  $f$ ,

$$\lim_{k \rightarrow \infty} \left[ \|x^k - x^*\|_2^2 + \frac{\rho_k}{2} \left[ \|h(x^k)\|_2^2 + \sum_{i=1}^p g_i(x^k)_+^2 \right] \right] = 0.$$

Thus,

$$\lim_{k \rightarrow \infty} \left[ \sum_{i=1}^m \rho_k h_i(x^k)^2 + \sum_{i=1}^p \rho_k g_i(x^k)_+^2 \right] = 0.$$

Defining

$$\lambda^k = \rho_k h(x^k), \quad \mu^k = \rho_k g(x^k)_+, \quad (32) \quad \{\text{losmul}\}$$

we obtain:

$$\lim_{k \rightarrow \infty} \left[ \sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{i=1}^p \mu_i^k g_i(x^k)_+ \right] = 0. \quad (33) \quad \{\text{yaya}\}$$

By (32),  $\mu_i^k = 0$  when  $g_i(x^k) < 0$ . Therefore, by (33),

$$\lim_{k \rightarrow \infty} \left[ \sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{i=1}^p |\mu_i^k g_i(x^k)| \right] = 0. \quad (34) \quad \{\text{dieciocho}\}$$

Thus, (19) follows from (32) and (34).

The proof of (18) is standard. For  $k$  large enough, one has that  $\|x^k - x^*\| < \delta$ , therefore, the gradient of the objective function must vanish. Thus, by (30),

$$\nabla f(x^k) + 2(x^k - x^*) + \sum_{i=1}^m \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{i=1}^p \rho_k g_i(x^k)_+ \nabla g_i(x^k) = 0$$

By (32), since  $\|x^k - x^*\| \rightarrow 0$  we have that

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k) + \nabla h(x^k) \lambda^k + \nabla g(x^k) \mu^k\| = 0.$$

Thus, (18) is proved. □

## 4 Strength of the CAKKT condition

In this section we deal with the problem of proximity between points that satisfy CAKKT and local minimizers, from a qualitative point of view. We will show that if a point  $x^*$  satisfies CAKKT, then it satisfies KKT or fails to satisfy a weak constraint qualification. We will prove this result using an intermediate result: we prove first that the fulfillment of CAKKT implies the fulfillment of AGP and that the reciprocal property is not true. AGP was proved to be a strong optimality condition in the sense that it implies KKT or the absence of a weak constraint qualification in [2, ?, 25]. In [2] it was proved that AGP implies AKKT. These results suggest that points that satisfy CAKKT are more likely to be local minimizers than points that merely satisfy AGP, or points that do not satisfy the weak constraint qualification. Therefore, points that approximate a CAKKT point have more chances to be close to local minimizers than points that approximate AGP.

In this section we are concerned with the first issue addressed in Section 2 of this paper: The results presented here support the conjecture that the fulfillment of CAKKT is a better indicator of proximity to local minimizers than the fulfillment of other approximate KKT conditions.

A necessary optimality condition should be as strong as possible. Moreover, as we will see in the next section, algorithmically oriented optimality conditions should be associated with some implementable nonlinear programming algorithm. A plausible conjecture is that the property of converging to points that satisfy strong necessary optimality conditions is linked to the practical efficiency of the algorithm.

In this section we will see that CAKKT is strong. We will show that any CAKKT point satisfies the KKT conditions or fails to fulfill the Constant Positive Linear Dependence constraint qualification (CPLD)<sup>2</sup>. The CPLD condition was introduced in [24] and, in [4], it was proved that it implies the quasinormality constraint qualification [7]. Since CPLD is a weak constraint qualification (strictly weaker than the Mangasarian-Fromovitz condition [19]) the property

$$\text{KKT or not-CPLD} \quad (35) \quad \{\text{not}\}$$

---

<sup>2</sup>A feasible point  $x^*$  is said to satisfy the CPLD condition if the existence of linear dependent gradients of active constraints with positive coefficients implies that the same gradients are linearly dependent in a neighborhood of  $x^*$  [24, 4].

is a strong necessary optimality condition. The property that CAKKT implies (35) will follow as a corollary of a stronger result. For stating this result we must recall a different sequential optimality condition, introduced in [22], analyzed in [2] and employed in several algorithmically oriented papers ([13, 14, 15, 20, 21] and others).

Recall that a feasible point  $x^*$  satisfies the Approximate Gradient Projection (AGP) property if there exist a sequence  $\{x^k\}$  that converges to  $x^*$  and satisfies

$$\lim_{k \rightarrow \infty} \|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\| = 0, \quad (36) \quad \{\text{projecero}\}$$

where  $\Omega_k$  is the set of points  $x \in \mathbb{R}^n$  that satisfy:

$$\nabla h(x^k)^T(x - x^k) = 0 \quad (37) \quad \{\text{agp1}\}$$

and

$$\nabla g(x^k)^T(x - x^k) + g(x^k)_- \leq 0. \quad (38) \quad \{\text{agp2}\}$$

Note that  $x^k$  always belong to the polytope defined by (37) and (38).

If, in addition,  $x^k$  fulfills the linear (equality or inequality) constraints of (2) defined by a set of indices  $I_{lin}$  we say that  $x^*$  satisfies the Linear AGP (LAGP) condition associated with  $I_{lin}$ . It can be shown that LAGP is strictly stronger than AGP [2]. Moreover, both AGP and LAGP are strictly stronger than (35) [2, 14].

Let us show now that CAKKT implies AGP.

**Theorem 4.1.** *Assume that  $x^*$  is a feasible CAKKT point of (2). Then,  $x^*$  satisfies the AGP condition. Moreover, if all the elements of a sequence  $\{x^k\}$  associated with the CAKKT definition fulfill all the linear constraints corresponding to a set of indices  $I_{lin}$ , then  $x^*$  satisfies the LAGP condition associated with  $I_{lin}$ . Finally, if  $x^*$  satisfies the CPLD constraint qualification, this point fulfills the KKT optimality conditions.*

*Proof.* Assume that  $\{x^k\} \subset \mathbb{R}^n$  converges to  $x^*$  and satisfies (18) and (19). Let  $y^k$  be the solution of

$$\text{Minimize } \|[x^k - \nabla f(x^k)] - y\|_2^2 \quad (39) \quad \{\text{pnino}\}$$

subject to  $y \in \Omega_k$ , where  $\Omega_k$  is the set of points defined by

$$\begin{aligned} \nabla h_i(x^k)^T(y - x^k) &= 0, i = 1, \dots, m, \\ \nabla g_i(x^k)^T(y - x^k) &\geq 0, \text{ if } g_i(x^k) \geq 0, \\ g_i(x^k) + \nabla g_i(x^k)^T(y - x^k) &\geq 0 \text{ if } g_i(x^k) < 0. \end{aligned}$$

The objective function of (39) is a strictly convex quadratic and  $\Omega_k$  is defined by linear constraints. Since  $x^k \in \Omega_k$  one has that  $\Omega_k$  is nonempty and, so,  $y^k$  exists and is the unique solution of this problem. We wish to show that  $\lim_{k \rightarrow \infty} \|y^k - x^k\| = 0$ . Since the constraints of (39) are linear, the KKT conditions are fulfilled at  $y^k$ . Therefore, there exist  $\{\hat{\lambda}^k\} \subset \mathbb{R}^m$ ,  $\{\hat{\mu}^k\} \subset \mathbb{R}_+^p$  such that:

$$[y^k - x^k] + \nabla f(x^k) + \nabla h(x^k)\hat{\lambda}^k + \nabla g(x^k)\hat{\mu}^k = 0, \quad (40) \quad \{\text{nino1}\}$$

$$\nabla h_i(x^k)^T(y^k - x^k) = 0, i = 1, \dots, m, \quad (41) \quad \{\text{nino2}\}$$

$$\nabla g_i(x^k)^T(y^k - x^k) \leq 0, \text{ if } g_i(x^k) \geq 0, \quad (42) \quad \{\text{nino3}\}$$

$$g_i(x^k) + \nabla g_i(x^k)^T(y^k - x^k) \leq 0 \text{ if } g_i(x^k) < 0. \quad (43) \quad \{\text{nino4}\}$$

$$\hat{\mu}_i^k \nabla g_i(x^k)^T(y^k - x^k) = 0 \text{ if } g_i(x^k) \geq 0, \quad (44) \quad \{\text{nino5}\}$$

and

$$\widehat{\mu}_i^k g_i(x^k) + \widehat{\mu}_i^k \nabla g_i(x^k)^T (y^k - x^k) = 0 \text{ if } g_i(x^k) < 0. \quad (45) \quad \{\text{nino6}\}$$

By (41–44), pre-multiplying (40) by  $(y^k - x^k)^T$  we obtain:

$$\|y^k - x^k\|_2^2 + \nabla f(x^k)^T (y^k - x^k) + \sum_{g_i(x^k) < 0} \widehat{\mu}_i^k \nabla g_i(x^k)^T (y^k - x^k) = 0. \quad (46) \quad \{\text{nino7}\}$$

By (45), when  $g_i(x^k) < 0$ , we have that

$$\widehat{\mu}_i^k \nabla g_i(x^k)^T (y^k - x^k) = -\widehat{\mu}_i^k g_i(x^k).$$

Therefore, by (46),

$$\|y^k - x^k\|_2^2 + \nabla f(x^k)^T (y^k - x^k) = \sum_{g_i(x^k) < 0} \widehat{\mu}_i^k g_i(x^k)^T.$$

Then, since  $\widehat{\mu}^k \geq 0$ , we have:

$$\|y^k - x^k\|_2^2 \leq -\nabla f(x^k)^T (y^k - x^k). \quad (47) \quad \{\text{menoraro}\}$$

Now, by (18), there exist sequences  $\{\lambda^k\} \subset \mathbb{R}^p$ ,  $\{\mu^k\} \subset \mathbb{R}_+^p$ ,  $\{v^k\} \subset \mathbb{R}^n$  such that

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

Therefore,

$$-\nabla f(x^k)^T (y^k - x^k) = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k)^T (y^k - x^k) + \sum_{i=1}^p \mu_i^k \nabla g_i(x^k)^T (y^k - x^k) - (y^k - x^k)^T v^k.$$

Thus, by (41),

$$\begin{aligned} -\nabla f(x^k)^T (y^k - x^k) &= \sum_{i=1}^p \mu_i^k \nabla g_i(x^k)^T (y^k - x^k) - (y^k - x^k)^T v^k \\ &= \sum_{g_i(x^k) < 0} \mu_i^k \nabla g_i(x^k)^T (y^k - x^k) + \sum_{g_i(x^k) \geq 0} \mu_i^k \nabla g_i(x^k)^T (y^k - x^k) - (y^k - x^k)^T v^k. \end{aligned}$$

By (42), since  $\mu^k \geq 0$ , one has that  $\mu_i^k \nabla g_i(x^k)^T (y^k - x^k) \leq 0$  whenever  $g_i(x^k) \geq 0$ , therefore,

$$\begin{aligned} -\nabla f(x^k)^T (y^k - x^k) &\leq \sum_{g_i(x^k) < 0} \mu_i^k \nabla g_i(x^k)^T (y^k - x^k) - (y^k - x^k)^T v^k \\ &= \sum_{g_i(x^k) < 0} \mu_i^k [g_i(x^k) + \nabla g_i(x^k)^T (y^k - x^k)] - \sum_{g_i(x^k) < 0} \mu_i^k g_i(x^k) - (y^k - x^k)^T v^k. \end{aligned}$$

Thus, by (43), since  $\mu^k \geq 0$ , we have:

$$-\nabla f(x^k)^T (y^k - x^k) \leq - \sum_{g_i(x^k) < 0} \mu_i^k g_i(x^k) - (y^k - x^k)^T v^k \leq - \sum_{g_i(x^k) < 0} \mu_i^k g_i(x^k) + \|v^k\|_2 \|y^k - x^k\|_2.$$

Therefore, by (47),

$$\|y^k - x^k\|_2^2 \leq \sum_{g_i(x^k) < 0} |\mu_i^k g_i(x^k)| + \|v^k\|_2 \|y^k - x^k\|_2. \quad (48) \quad \{\text{apafin}\}$$

By (19),  $\lim_{k \rightarrow \infty} |\mu_i^k g_i(x^k)| = 0$  for all  $i$ , so the sequence  $\{\|y^k - x^k\|\}$  is bounded and, taking limits in both sides of (48), we obtain that  $\lim_{k \rightarrow \infty} \|y^k - x^k\| = 0$  as we wanted to prove. Therefore,  $x^*$  satisfies the AGP condition.

The second part of the proof is immediate. If  $\{x^k\}$  satisfies all the linear constraints corresponding to the indices in  $I_{lin}$ , it satisfies the LAGP condition associated to this set.

Finally, if  $x^*$  satisfies the CPLD constraint qualification, the fulfillment of KKT follows from the fulfillment of AGP [2, 14, 25]. □

We will show now that CAKKT is strictly stronger than AGP. (Recall that AGP is strictly stronger than AKKT [2].) Consider, once more, the problem defined by (27), (28). Define  $x^k = (1/k, 1)^T$ . Clearly,  $\Omega_k = \{x^k\}$  for all  $k \in \mathbb{N}$ . Therefore,  $P_{\Omega_k}(x^k - \nabla f(x^k)) = x^k$  and  $\|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\| = 0$  for all  $k \in \mathbb{N}$ . Therefore,  $x^* = (0, 1)^T$  satisfies AGP.

Let us show now that a sequence  $x^k$  fulfilling (18) and (19) cannot exist. If such a sequence exists, we have  $x^k = (x_1^{(k)}, x_2^{(k)})^T$  and  $\lambda^k = (\lambda_1^{(k)}, \lambda_2^{(k)})^T$  satisfying:

$$\lim_{k \rightarrow \infty} x_1^{(k)} = 0, \quad \lim_{k \rightarrow \infty} x_2^{(k)} = 1, \quad (49) \quad \{\text{cont0}\}$$

such that, by (18),

$$\lim_{k \rightarrow \infty} \begin{pmatrix} 0 \\ x_2^{(k)} - 2 \end{pmatrix} + \lambda_1^{(k)} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2^{(k)} \begin{pmatrix} x_2^{(k)} \\ x_1^{(k)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (50) \quad \{\text{cont1}\}$$

and, by (19),

$$\lim_{k \rightarrow \infty} \lambda_1^{(k)} x_1^{(k)} = 0 \quad (51) \quad \{\text{cont2}\}$$

and

$$\lim_{k \rightarrow \infty} \lambda_2^{(k)} x_1^{(k)} x_2^{(k)} = 0. \quad (52) \quad \{\text{cont3}\}$$

By (50) we have:

$$\lim_{k \rightarrow \infty} \lambda_1^{(k)} + \lambda_2^{(k)} x_2^{(k)} = 0$$

and

$$\lim_{k \rightarrow \infty} x_2^{(k)} + \lambda_2^{(k)} x_1^{(k)} = 2. \quad (53) \quad \{\text{cont5}\}$$

By (49) and (53), we have:

$$\lim_{k \rightarrow \infty} \lambda_2^{(k)} x_1^{(k)} = 1.$$

Therefore, since  $\lim_{k \rightarrow \infty} x_2^{(k)} = 1$ ,

$$\lim_{k \rightarrow \infty} \lambda_2^{(k)} x_1^{(k)} x_2^{(k)} = \lim_{k \rightarrow \infty} x_2^{(k)} = 1.$$

This contradicts (52). Therefore, a sequence satisfying (49–52) cannot exist. Thus  $x^*$  is not a CAKKT point. By Theorem 4.1, CAKKT is strictly stronger than AGP.

We finish this section with an additional strongness result. In fact, we will show that, in the convex case, CAKKT is a sufficient optimality condition for global minimizers. As a consequence, in convex problems, CAKKT is equivalent to global minimization.

**Theorem 4.2.** *Assume that, in problem (2), the functions  $f$  and  $g_i, i = 1, \dots, p$  are convex and  $h_1, \dots, h_m$  are affine. Let  $x^*$  be a feasible point that satisfies the CAKKT condition. Then,  $x^*$  is a global minimizer of (2).*

*Proof.* Assume that  $\{x^k\}, \{\lambda^k\}, \{\mu^k\}$  are given by (18), (19). Let  $z$  be a feasible point of (2). By the convexity of  $f$  and the constraints, we have, for all  $k \in \mathbb{N}$ ,

$$\begin{aligned} f(z) &\geq f(x^k) + \nabla f(x^k)^T(z - x^k), \\ h_i(z) &= h_i(x^k) + \nabla h_i(x^k)^T(z - x^k) = 0, i = 1, \dots, m, \\ g_i(z) &\geq g_i(x^k) + \nabla g_i(x^k)^T(z - x^k), i = 1, \dots, p. \end{aligned}$$

Therefore, since  $h(z) = 0$  and  $g(z) \leq 0$ ,

$$\begin{aligned} f(z) &\geq f(x^k) + \nabla f(x^k)^T(z - x^k) + \sum_{i=1}^m \lambda_i^k h_i(z) + \sum_{i=1}^p \mu_i^k g_i(z) \\ &\geq f(x^k) + \nabla f(x^k)^T(z - x^k) + \sum_{i=1}^m \lambda_i^k [h_i(x^k) + \nabla h_i(x^k)^T(z - x^k)] + \sum_{i=1}^p \mu_i^k [g_i(x^k) + \nabla g_i(x^k)^T(z - x^k)] \\ &= f(x^k) + [\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla g_i(x^k)]^T(z - x^k) + \sum_{i=1}^m \lambda_i^k h_i(x^k) + \sum_{i=1}^p \mu_i^k g_i(x^k). \end{aligned}$$

Thus,

$$f(z) \geq \lim_{k \rightarrow \infty} f(x^k) + \lim_{k \rightarrow \infty} [(z - x^k)^T [\nabla f(x^k) + \nabla h(x^k) \lambda^k + \nabla g(x^k) \mu^k]] + \lim_{k \rightarrow \infty} [\sum_{i=1}^m \lambda_i^k h_i(x^k) + \sum_{i=1}^p \mu_i^k g_i(x^k)].$$

Then, by the continuity of  $f$  and the properties (18), (19), we have that  $f(z) \geq f(x^*)$ .  $\square$

## 5 A practical algorithm that generates CAKKT points

This section concerns the second issue mentioned in Section 2 of the present paper. Since we know that the CAKKT optimality condition is strong, the obvious conjecture is that algorithms that guaranteedly find CAKKT points may be, in some sense, more efficient than algorithms that may converge to feasible points that do not fulfill the CAKKT condition.

A first step along this direction will be to discuss an implementable algorithm that generates sequences converging to CAKKT points. We exclude from our analysis “global optimization algorithms” like the one introduced in [8] that guaranteedly converge to global minimizers using more expensive procedures than the ones generally affordable in everyday practical optimization. Algorithms that converge to global minimizers obviously satisfy CAKKT, since even local minimizers satisfy this condition, as shown in Theorem 3.1.

{cuatro}

Our results in this section make use a generalization of the Kurdyka-Lojasiewicz inequality [5, 9, 17, 18].

One says that the continuously differentiable function  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$  satisfies the Kurdyka-Lojasiewicz inequality [5, 17, 18] at  $\bar{x}$  if there exists  $\delta > 0$ ,  $\theta \in (0, 1)$ ,  $c > 0$  such that, for all  $x \in B(\bar{x}, \delta)$ ,

$$|\Phi(x) - \Phi(\bar{x})|^\theta \leq c \|\nabla \Phi(x)\|. \quad (54) \quad \{\mathbf{k1}\}$$

The properties of functions that satisfy this inequality have been studied in several recent papers in connection with minimization methods, complexity theory and asymptotic analysis of partial differential equations and tame geometry [5, 6, 9, 10]. Smooth functions satisfy this inequality under fairly weak conditions.

We will use an even weaker form of the Kurdyka-Lojasiewicz inequality. Let us say that the smooth function  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$  satisfies the Generalized Kurdyka-Lojasiewicz (GKL) inequality at  $\bar{x}$  if there exist  $\delta > 0$ ,  $\varphi : B(\bar{x}, \varepsilon)$  such that  $\lim_{x \rightarrow \bar{x}} \varphi(x) = 0$  and for all  $x \in B(\bar{x}, \delta)$ , one has:

$$|\Phi(x) - \Phi(\bar{x})| \leq \varphi(x) \|\nabla \Phi(x)\|. \quad (55) \quad \{\mathbf{gk1}\}$$

Clearly, the fulfillment of (54) implies that (55) holds, but the reciprocal is not true. To see this, define

$$\Phi(x) = e^{-\frac{1}{|x|}} \text{ if } x \neq 0,$$

$$\Phi(0) = 0.$$

This function satisfies GKL at  $\bar{x} = 0$ , since

$$\frac{|\Phi(x) - \Phi(0)|}{|\Phi'(x)|} = x^2 \rightarrow 0.$$

However, for all  $\theta \in (0, 1)$ ,

$$\frac{|\Phi(x) - \Phi(0)|^\theta}{|\Phi'(x)|} = x^2 e^{\frac{1-\theta}{|x|}} \rightarrow \infty.$$

Therefore,  $\Phi$  does not satisfy KL.

We will analyze the Augmented Lagrangian algorithm with arbitrary lower level constraints described in [1]<sup>3</sup>. For the description of this algorithm, let us formulate the nonlinear programming problem in the form:

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0, \quad g(x) \leq 0, \quad \underline{h}(x) = 0, \quad \underline{g}(x) \leq 0, \quad (56) \quad \{\mathbf{n1p2}\}$$

where  $f, h, g$  are as in (2) and  $\underline{h} : \mathbb{R}^n \rightarrow \mathbb{R}^m, \underline{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$  are continuously differentiable.

We say that  $\underline{h}(x) = 0, \underline{g}(x) \leq 0$  are “lower-level constraints”. These constraints are usually simpler than the “upper-level constraints”  $h(x) = 0, g(x) \leq 0$ , which means that minimizing (only with) lower-level constraints is easier than minimizing with general constraints. Frequently, lower-level constraints are given by upper and lower bounds of the form  $\ell \leq x \leq u$ .

For all  $x \in \mathbb{R}^n, \bar{\lambda} \in \mathbb{R}^m, \bar{\mu} \in \mathbb{R}_+^p, \rho > 0$ , we define the “displaced infeasibility” by:

$$\Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x) = \frac{1}{2} \left[ \left\| h(x) + \frac{\bar{\lambda}}{\rho} \right\|_2^2 + \left\| \left( g(x) + \frac{\bar{\mu}}{\rho} \right)_+ \right\|_2^2 \right]. \quad (57) \quad \{\mathbf{fibar}\}$$

The augmented Lagrangian  $L_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x)$  is given by:

$$L_{\bar{\lambda}, \bar{\mu}, \rho}(x) = f(x) + \rho \Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x). \quad (58) \quad \{\mathbf{Lbar}\}$$

---

<sup>3</sup>A freely available implementation of this method may be found in [www.ime.usp.br/~egbirgin/tango](http://www.ime.usp.br/~egbirgin/tango).

The general description of the Augmented Lagrangian algorithm with lower-level constraints [1] is given below. In [1] it was proved that feasible limit points that satisfy the CPLD constraint qualification [4, 24] are KKT points.

**Algorithm 5.1.** Let  $\varepsilon_k \downarrow 0$ ,  $\bar{\lambda}^k \in [\lambda_{min}, \lambda_{max}]^m$ ,  $\bar{\mu}^k \in [0, \mu_{max}]^p$  for all  $k \in \mathbb{N}$ ,  $\rho_1 > 0$ ,  $r \in (0, 1)$ ,  $\gamma > 1$ .

For all  $k = 1, 2, \dots$  we compute  $x^k \in \mathbb{R}^n$ ,  $\underline{\lambda}^k \in \mathbb{R}^m$ ,  $\underline{\mu}^k \in \mathbb{R}_+^p$  such that:

$$\|\nabla L_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) + \nabla \underline{h}(x^k) \underline{\lambda}^k + \nabla \underline{g}(x^k) \underline{\mu}^k\| \leq \varepsilon_k, \quad (59) \quad \{\text{due}\}$$

$$\|\underline{h}(x^k)\| \leq \varepsilon_k, \|\underline{g}(x^k)_+\| \leq \varepsilon_k,$$

and

$$\underline{\mu}_i^k = 0 \text{ whenever } \underline{g}_i(x^k) < -\varepsilon_k.$$

We define, for all  $i = 1, \dots, p$ ,

$$V_i^k = \max \left\{ g_i(x^k), \frac{-\bar{\mu}_i^k}{\rho_k} \right\}. \quad (60) \quad \{\text{vik}\}$$

If  $k = 1$  or

$$\max\{\|h(x^k)\|, \|V^k\|\} \leq r \max\{\|h(x^{k-1})\|, \|V^{k-1}\|\} \quad (61) \quad \{\text{max}\}$$

we define  $\rho_{k+1} \geq \rho_k$ . Else, we define  $\rho_{k+1} \geq \gamma \rho_k$ .

#### Remarks.

- In general, we define  $\rho_{k+1} = \rho_k$  if (61) holds and  $\rho_{k+1} = \gamma \rho_k$  otherwise [1]. Here we prefer to use the more general form, in which  $\rho_{k+1} \geq \rho_k$  and  $\rho_{k+1} \geq \gamma \rho_k$ , for theoretical reasons. The global convergence results of [1] hold for this formulation without modifications and the more general formulation is useful to analyze more general optimization problems [3].
- In Theorem 5.1 we will assume that the approximate Lagrange multipliers  $\underline{\lambda}^k, \underline{\mu}^k$  associated with lower-level constraints are bounded. This is a reasonable assumption because, in practical terms, lower-level constraints should be simple (complicated constraints should be included in the upper level) and complications due to the lack of fulfillment of constraint qualifications should not be expected in the lower level. The theorem obviously holds if there are no lower-level constraints at all, so that all the constraints are submitted to the penalty-Lagrangian treatment. Theorem 5.1 also includes, as a particular case, the classical External Penalty method [12] with an approximate stopping criterion for the subproblem.
- The convergence results of [1] apply to Algorithm 5.1. In [1] it was proved that feasible limit points that satisfy the CPLD constraint qualification are KKT points. Under the GKL assumption, Theorem 5.1 presents a stronger result, showing that feasible limit points satisfy the CAKKT condition.

**Theorem 5.1.** Assume that  $x^*$  is a feasible limit point of a sequence generated by Algorithm 5.1 and that the sequences  $\{\underline{\lambda}^k\}, \{\underline{\mu}^k\}$  are bounded. In addition, assume that there exist  $\delta > 0, \bar{\rho} > 0$ ,  $\varphi : B(x^*, \delta) \rightarrow \mathbb{R}^n$ ,  $\lim_{x \rightarrow x^*} \varphi(x) = 0$ , such that, for all  $x \in B(x^*, \delta)$ ,  $\bar{\lambda} \in [\lambda_{min}, \lambda_{max}]$ ,  $\bar{\mu} \in [0, \mu_{max}]$ ,  $\rho \geq \bar{\rho}$ ,

$$|\Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x) - \Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x^*)| \leq \varphi(x) \|\nabla \Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x)\|. \quad (62) \quad \{\text{k12}\}$$

Then,  $x^*$  satisfies CAKKT.

*Proof.* Define, for all  $k = 1, 2, \dots$ ,

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k), \quad \mu^k = (\bar{\mu}^k + \rho_k g(x^k))_+. \quad (63) \quad \{\text{lamu}\}$$

By (58), (59) and (63) one has:

$$\|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k + \nabla \underline{h}(x^k)\underline{\lambda}^k + \nabla \underline{g}(x^k)\underline{\mu}^k\| \leq \varepsilon_k$$

for all  $k = 1, 2, \dots$

Therefore, we only need to prove (19), both for the upper-level and the lower-level constraints. Without loss of generality we assume that the whole sequence  $\{x^k\}$  converges to  $x^*$ .

The proof of (19) for the lower-level constraints (replacing  $\lambda^k$  by  $\underline{\lambda}^k$ ,  $\mu^k$  by  $\underline{\mu}^k$ ,  $h$  by  $\underline{h}$  and  $g$  by  $\underline{g}$ ) follows trivially from the feasibility of  $x^*$ , the continuity of  $\underline{h}, \underline{g}$  and the boundedness of  $\{\underline{\lambda}^k, \underline{\mu}^k\}$ .

For proving (19) in the case of upper-level constraints, let us consider first the case in which  $i \in \{1, \dots, p\}$  is such that  $g_i(x^*) = 0$  and  $g_i(x^k) \leq 0$  for an infinite set of indices  $K_1$ . By (63), for all  $k \in K_1$ , we have:

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

Then, for all  $k \in K_1$ , since  $g_i(x^k) \leq 0$ , we have that  $0 \leq \mu_i^k \leq \bar{\mu}_i^k \leq \mu_{max}$ . Then,  $\{\mu_i^k, k \in K_1\}$  is bounded. So, by the continuity of  $g_i$  and the fact that  $g_i(x^*) = 0$ , we get:

$$\lim_{k \in K_1} \mu_i^k g_i(x^k) = 0. \quad (64) \quad \{\text{paracom}\}$$

This result will be used later in the proof.

Consider now the case in which  $\{\rho_k\}$  is bounded above. By (63), we have that  $\{\lambda^k\}$  and  $\{\mu^k\}$  are bounded. Moreover, by the choice of  $\rho_{k+1}$ , defining  $V^k$  as in (60), we have that  $\lim_{k \rightarrow \infty} \|V^k\| = 0$ . Clearly, since  $x^*$  is feasible,  $\lim_{k \rightarrow \infty} \lambda_i^k h_i(x^k) = 0$  and, if  $g_i(x^*) = 0$ ,  $\lim_{k \rightarrow \infty} \mu_i^k g_i(x^k) = 0$ . In the case that  $g_i(x^*) < 0$ , since  $V^k \rightarrow 0$ , we have that

$$\lim_{k \rightarrow \infty} \frac{\bar{\mu}_i^k}{\rho_k} = 0.$$

Then, by the boundedness of  $\{\rho_k\}$ ,

$$\lim_{k \rightarrow \infty} \bar{\mu}_i^k = 0.$$

Thus,  $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$  for  $k$  large enough. Therefore, by (63),  $\mu_i^k = 0$  for  $k$  large enough. This implies that  $\lim_{k \rightarrow \infty} \mu_i^k g_i(x^k) = 0$  also in the case that  $g_i(x^*) < 0$ .

This completes the proof of (19) in the case that  $\{\rho_k\}$  is bounded.

Let us consider now the case in which  $\lim_{k \rightarrow \infty} \rho_k = \infty$ . By (58), (59), the continuity of  $f, \nabla \underline{h}, \nabla \underline{g}$  and the boundedness of  $\{\underline{\lambda}^k\}, \{\underline{\mu}^k\}$ , we have that  $\{\rho_k \|\nabla \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k)\|\}$  is bounded.

By (62), for  $k$  large enough we have:

$$|\Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) - \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^*)| \leq \varphi(x^k) \|\nabla \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k)\|.$$

Since  $x^k \rightarrow x^*$  and  $\varphi(x^k) \rightarrow 0$ , we obtain:

$$|\Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) - \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^*)| \leq c_1(k) \|\nabla \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k)\|,$$

with  $\lim_{k \rightarrow \infty} c_1(k) = 0$ . Multiplying by  $\rho_k$  and using the boundedness of  $\{\rho_k \|\nabla \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k)\|\}$ , we obtain:

$$\lim_{k \rightarrow \infty} \rho_k |\Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) - \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^*)| = 0. \quad (65) \quad \{\text{esto}\}$$

Now,

$$\begin{aligned}
\Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^*) &= \frac{1}{2} \left[ \sum_{i=1}^m \left[ h_i(x^*) + \frac{\bar{\lambda}_i^k}{\rho_k} \right]^2 + \sum_{i=1}^p \max \left\{ g_i(x^*) + \frac{\bar{\mu}_i^k}{\rho_k}, 0 \right\}^2 \right] \\
&= \frac{1}{2} \left[ \sum_{i=1}^m \left[ \frac{\bar{\lambda}_i^k}{\rho_k} \right]^2 + \sum_{i=1}^p \max \left\{ \frac{\bar{\mu}_i^k + \rho_k g_i(x^*)}{\rho_k}, 0 \right\}^2 \right] \\
&= \frac{1}{2\rho_k^2} \left[ \sum_{i=1}^m (\bar{\lambda}_i^k)^2 + \sum_{i=1}^p \max \{ \bar{\mu}_i^k + \rho_k g_i(x^*), 0 \}^2 \right].
\end{aligned}$$

If  $g_i(x^*) < 0$  then, for  $k$  large enough,  $\max\{\bar{\mu}_i^k + \rho_k g_i(x^*), 0\} = 0$ . Therefore,

$$\rho_k \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^*) \leq \frac{1}{2\rho_k} \left[ \sum_{i=1}^m (\bar{\lambda}_i^k)^2 + \sum_{i=1}^p (\bar{\mu}_i^k)^2 \right].$$

Therefore,  $\lim_{k \rightarrow \infty} \rho_k \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^*) = 0$ . Thus, by (65),

$$\lim_{k \rightarrow \infty} \rho_k \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) = 0. \quad (66) \quad \{\text{estotro}\}$$

But

$$\begin{aligned}
\rho_k \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) &= \frac{\rho_k}{2} \left[ \sum_{i=1}^m \left[ h_i(x^k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right]^2 + \sum_{i=1}^p \max \left\{ g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k}, 0 \right\}^2 \right] \\
&= \frac{1}{2} \sum_{i=1}^m \left| \bar{\lambda}_i^k + \rho_k h_i(x^k) \right| \left| h_i(x^k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right| + \sum_{i=1}^p \left( \bar{\mu}_i^k + \rho_k g_i(x^k) \right)_+ \left( g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right)_+ \\
&= \frac{1}{2} \sum_{i=1}^m |\lambda_i^k| \left| h_i(x^k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right| + \sum_{i=1}^p \mu_i^k \left( g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right)_+.
\end{aligned}$$

Thus, by (66),

$$\lim_{k \rightarrow \infty} \lambda_i^k \left[ h_i(x^k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right] = 0 \quad (67) \quad \{\text{casilam}\}$$

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

$$\lim_{k \rightarrow \infty} \mu_i^k \left( g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right)_+ = 0 \quad (68) \quad \{\text{casimu}\}$$

for  $i = 1, \dots, p$ .

Now,

$$\lambda_i^k \left[ h_i(x^k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right] = \lambda_i^k h_i(x^k) + \lambda_i^k \frac{\bar{\lambda}_i^k}{\rho_k} \quad (69) \quad \{\text{quala}\}$$

and

$$\lambda_i^k \frac{\bar{\lambda}_i^k}{\rho_k} = (\bar{\lambda}_i^k + \rho_k h_i(x^k)) \frac{\bar{\lambda}_i^k}{\rho_k} = \frac{(\bar{\lambda}_i^k)^2}{\rho_k} + \bar{\lambda}_i^k h_i(x^k) \rightarrow 0.$$

Therefore, by (67) and (69), since  $\rho_k \rightarrow \infty$ , we have that

$$\lim_{k \rightarrow \infty} \lambda_i^k h_i(x^k) = 0 \quad (70) \quad \{\text{lambache}\}$$

for all  $i = 1, \dots, m$ .

By (63), if  $g_i(x^*) < 0$  we have that  $\mu_i^k = 0$  for  $k$  large enough. This implies that  $\lim_{k \rightarrow \infty} \mu_i^k g_i(x^k) = 0$  in this case.

By (64), if  $g_i(x^*) = 0$  and  $g_i(x^k) < 0$  for infinitely many indices  $k \in K_1$ , one has that  $\lim_{k \in K_1} \mu_i^k g_i(x^k) = 0$ .

Therefore, it remains to analyze the case in which  $g_i(x^k) \geq 0$  for infinitely many indices  $k \in K_2$ . In this case, we have:

$$\mu_i^k \left( g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right)_+ = \left( \mu_i^k g_i(x^k) + \frac{\bar{\mu}_i^k \mu_i^k}{\rho_k} \right)_+ \quad (71) \quad \{\text{quamu}\}$$

and, by (63),

$$\frac{\mu_i^k \bar{\mu}_i^k}{\rho_k} = \frac{[\bar{\mu}_i^k + \rho_k g_i(x^k)]_+ \bar{\mu}_i^k}{\rho_k} = \left( \frac{(\bar{\mu}_i^k)^2}{\rho_k} + \bar{\mu}_i^k g_i(x^k) \right)_+.$$

Then, since  $g_i(x^k) \rightarrow 0$  and  $\rho_k \rightarrow \infty$ , we deduce that  $\lim_{k \in K_2} \mu_i^k \bar{\mu}_i^k / \rho_k = 0$ . By (68) and (71), this implies that  $\lim_{k \in K_2} (\mu_i^k g_i(x^k))_+ = 0$ . Since, in this case,  $g_i(x^k) \geq 0$  for all  $k$ , we have that

$$\lim_{k \in K_2} \mu_i^k g_i(x^k) = 0. \quad (72) \quad \{\text{yaesta}\}$$

Since  $K_1 \cup K_2 = \{1, 2, 3, \dots\}$ , (64) and (72) imply that, for all  $i = 1, \dots, p$ ,

$$\lim_{k \rightarrow \infty} \mu_i^k g_i(x^k) = 0. \quad (73) \quad \{\text{ahora}\}$$

By (70) and (73) the proof is complete.  $\square$

### Counter-example

We will show that, if the GKL assumption does not hold, Algorithm 5.1 may generate a sequence that does not satisfy (19).

Consider the problem

$$\text{Minimize } x \text{ subject to } h(x) = 0, \quad (74) \quad \{\text{probcontra}\}$$

where

$$h(x) = x^4 \sin\left(\frac{1}{x}\right) \text{ if } x \neq 0,$$

and  $h(0) = 0$ .

We will use Algorithm 5.1 with  $\bar{\lambda}^k = 0$  for all  $k$ . Therefore, the algorithm reduces to an external penalty method. We will show that, for a choice of  $\rho_k \rightarrow \infty$ , the sequence  $\{x_k\}$  generated by the algorithm tends to  $x_* = 0$  and does not satisfy

$$\lim_{k \rightarrow \infty} x_k h(x_k) = 0.$$

We have:

$$h'(x) = x^2 \left[ 4x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) \right] \quad (75) \quad \{\text{hprima}\}$$

if  $x \neq 0$ ,  $h'(0) = 0$ .

Let us define, for all  $x \neq 0$ ,

$$R(x) = 4x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) + x^3 \quad (76) \quad \{\text{errex}\}$$

and, for all  $k = 1, 2, \dots$ ,

$$z_k = \frac{1}{2k\pi}, \quad y_k = \frac{1}{(2k + 1/2)\pi}. \quad (77) \quad \{\text{zetak}\}$$

Then, for all  $k = 1, 2, \dots$ ,

$$R(z_k) = -1 + z_k^3 < 0$$

and

$$R(y_k) = 4y_k + y_k^3 > 0.$$

Therefore, for all  $k = 1, 2, \dots$  there exists  $x_k \in (y_k, z_k)$  such that

$$R(x_k) = 0. \tag{78} \quad \{\text{errekero}\}$$

By (77) we have that  $y_k \rightarrow 0, z_k \rightarrow 0$ . Therefore, since  $x_k \in (y_k, z_k)$ ,  $\lim_{k \rightarrow \infty} x_k = 0$ . Moreover, since  $x_k \in (y_k, z_k)$  we have that

$$1/z_k < 1/x_k < 1/y_k$$

and, by (77),

$$2k\pi < 1/x_k < (2k + 1/2)\pi.$$

Therefore, for all  $k = 1, 2, \dots$ ,

$$\sin\left(\frac{1}{x_k}\right) > 0. \tag{79} \quad \{\text{two}\}$$

Since  $x_k \rightarrow 0$ , we have that  $\lim_{k \rightarrow \infty} x_k \sin(1/x_k) = 0$ . Then, by (76) and (78),

$$\lim_{k \rightarrow \infty} \cos\left(\frac{1}{x_k}\right) = 0. \tag{80} \quad \{\text{three}\}$$

By (79) and (80), we have that

$$\lim_{k \rightarrow \infty} \sin\left(\frac{1}{x_k}\right) = 1.$$

Therefore, for  $k$  large enough,

$$\sin\left(\frac{1}{x_k}\right) > \frac{1}{2}. \tag{81} \quad \{\text{medio}\}$$

By (75) and (78),

$$h'(x_k) = -x_k^5 \tag{82} \quad \{\text{hprimaquinta}\}$$

for all  $k = 1, 2, \dots$

Now, for  $k$  large enough, define:

$$\rho_k = \frac{-1}{h(x_k)h'(x_k)}. \tag{83} \quad \{\text{definerhok}\}$$

By (82), we have:

$$\rho_k = \frac{-1}{x^4 \sin(1/x_k)(-x_k^5)} = \frac{1}{x_k^9 \sin(1/x_k)}.$$

By (81),  $\rho_k$  is well defined for  $k$  large enough and  $\lim_{k \rightarrow \infty} \rho_k = \infty$ . Taking an appropriate subsequence we may assume, without loss of generality, that

$$\rho_{k+1} \geq \gamma \rho_k \tag{84} \quad \{\text{masgama}\}$$

for all  $k = 1, 2, \dots$

Now, by (58), in this case we have:

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

Thus, by (83),

$$\nabla L_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x_k) = 1 + \rho_k h(x_k) h'(x_k) = 0.$$

This means that the sequence  $\{x_k\}$  is defined by the application of Algorithm 5.1 to the problem (74), with  $\bar{\lambda}^k = 0$  for all  $k$  and the penalty parameters given by (83).

Now, assume that, for all  $k = 1, 2, \dots$ ,  $\lambda_k$  is an approximate Lagrange multiplier such that (18) holds. Then:

$$\lim_{k \rightarrow \infty} 1 + \lambda_k h'(x_k) = 0.$$

Therefore, by (82),

$$\lim_{k \rightarrow \infty} 1 - \lambda_k x_k^5 = 0.$$

and, so,

$$\lim_{k \rightarrow \infty} \lambda_k x_k^5 = 1. \tag{85} \quad \{\text{vaprauno}\}$$

Now,

$$\lambda_k h(x_k) = \lambda_k x_k^4 \sin(1/x_k) = \lambda_k x_k^5 \frac{\sin(1/x_k)}{x_k}.$$

By (81) and (85), since  $x_k \rightarrow 0$ , we deduce that

$$\lim_{k \rightarrow \infty} \lambda_k h(x_k) = \infty.$$

Therefore, (19) does not hold.

## 6 Conclusions

In the first section of this paper we motivated the introduction of a new strong approximate KKT condition using a simple example that shows that points that approximate fulfill KKT conditions may be far from true minimizers. It can be argued, however, that efficient nonlinear programming solvers will not compute these kinds of points and, thus, they will ultimately succeed in the optimization purpose. In fact, we showed in Section 4 that this is the case of the Augmented Lagrangian method defined in [1]. Nevertheless, “wrong points” that approximately satisfy classical KKT conditions may occur as initial approximations to the solution of a problem, when the user has no control of starting points, as is usually the case when the optimization problem is a part of a more complex model. In practical terms, this indicates that the stopping criterion associated with the CAKKT condition (including the fact that the product between multipliers and constraints must be small, even in the case of equality constraints) should be used in practical nonlinear optimization codes.

Few research has been dedicated to the study of optimality conditions in the case that KKT does not hold. Since, as shown in this paper, local minimizers satisfy CAKKT, this corresponds to the case in which some multipliers “are infinity”. However, many practical problems may have this characteristic and, thus, algorithms should be equipped with adequate procedures to deal with this anomaly. Moreover, numerical behavior of practical algorithms in the case of very large multipliers probably emulates the non-KKT case. Very likely, well-established implemented optimization algorithms include heuristics that make it possible to cope degenerate situations, but it is also plausible that many numerical phenomena may be explained in terms of the theoretical behavior in the presence of divergent sequences of multipliers.

A popular point of view in numerical optimization is that one always tries to solve a KKT system, with an obvious preference to solutions that represent minimizers. However, there is not a unique way to define a KKT system, although all the KKT formulations have the same

exact solutions. For example, consider the constrained optimization problem with only equality constraints. In this case there is a general agreement that the KKT (Lagrange) system of equations is:

$$\nabla f(x) + \nabla h(x)\lambda = 0, \quad h(x) = 0. \quad (86) \quad \{\text{lgrange}\}$$

In fact (86) is a square nonlinear system and variations of Newton's method are often successful for solving it. Now, (86) is obviously equivalent to:

$$\nabla f(x) + \nabla h(x)\lambda = 0, \quad h(x) = 0, \quad \lambda_i h_i(x) = 0, \quad i = 1, \dots, m \quad (87) \quad \{\text{lgrange2}\}$$

but equivalence disappears when, roughly speaking, we admit that multipliers may be infinity. In this case, the approximate fulfillment of (86) corresponds to the approximate KKT condition, but the approximate fulfillment of (87) gives rise to the CAKKT condition. In this sense, the systems are not equivalent and, as we saw before, we have good reasons to prefer the rectangular form (87).

As it is well known, Sequential Quadratic Programming (SQP) algorithms correspond to Newton's method applied to the nonlinear system given by the KKT conditions [23]. A natural question, addressed by one of the referees of the first version of this paper, is whether SQP sequences that converge to feasible points always provide multipliers satisfying the CAKKT relation (19). Similar questions may be formulated with respect to all modern practical algorithms. The rectangular form (87) seems to suggest that the answer is negative in the case of SQP. When one applies Newton's method to the KKT system associated with the problem of minimize  $f(x) \equiv x$  subject to  $h(x) \equiv x^2 = 0$  one gets a sequence  $(x_k, \lambda_k)$  such that  $x_k \rightarrow 0, \lambda_k \rightarrow \infty, \lambda_k h(x_k) \rightarrow 0$  but  $\|\nabla f(x_k) + \nabla h(x_k)\lambda_k\|$  converges to 1/3. Therefore, although  $x^* = 0$  satisfies CAKKT for an appropriate sequence of Lagrange multipliers, the multipliers provided by Newton's method are not the ones that guarantee the fulfillment of the CAKKT conditions. Curiously, the complementarity condition is satisfied by this sequence but the Lagrange condition (18) does not.

## Acknowledgements

We are indebted to two anonymous referees whose careful comments and insights helped us a lot to improve the present paper.

## 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, G. Haeser and J. M. Martínez, On sequential optimality conditions for smooth constrained optimization, *Optimization*, to appear.
- [3] R. Andreani, J. M. Martínez, L. Martínez and F. S. Yano, Low Order-Value Optimization and applications, *Journal of Global Optimization* 43, pp. 1-10, 2009.
- [4] R. Andreani, J. M. Martínez and M. L. Schuverdt, On the relation between the Constant Positive Linear Dependence condition and quasnormality constraint qualification, *Journal of Optimization Theory and Applications* 125, pp. 473–485, 2005.

- [5] H. Attouch and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Mathematical Programming* 116, pp. 5-16, 2009.
- [6] H. Attouch, J. Bolte, P. Redont and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Lojasiewicz inequality, to appear in *Mathematics of Operations Research*.
- [7] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Mass., 1999.
- [8] E. G. Birgin, C. A. Floudas and J. M. Martínez, Global minimization using an augmented Lagrangian method with variable lower-level constraints, to appear in *Mathematical Programming*.
- [9] J. Bolte, A. Daniilidis and A. S. Lewis, The Lojasiewicz inequality for nonsmooth sub-analytic functions with applications to subgradient dynamical systems, *SIAM Journal on Optimization* 17, pp. 1205–1223, 2006.
- [10] J. Bolte, A. Daniilidis, O. Ley and L. Mazet, Characterizations of Lojasiewicz inequalities and applications: subgradient flows, talweg, convexity, to appear in *Transactions of American Mathematical Society*.
- [11] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *Trust Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, 2000.
- [12] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.
- [13] A. Fischer and A. Friedlander, A new line search inexact restoration approach for nonlinear programming, to appear in *Computational Optimization and Applications*.
- [14] M. A. Gomes-Ruggiero, J. M. Martínez and S. A. Santos, Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints, *SIAM Journal on Scientific Computing* 31, pp. 1628-1652, 2009.
- [15] C. C. Gonzaga, E. W. Karas and M. Vanti, A globally convergent filter method for Nonlinear Programming, *SIAM Journal on Optimization* 14, pp. 646–669, 2003.
- [16] M. R. Hestenes, *Optimization Theory: The finite dimensional case*, Wiley, NY, 1975.
- [17] K. Kurdyka, On gradients of functions definable in o-minimal structures, *Annales de l'Institut Fourier* 48, pp. 769-783, 1998.
- [18] S. Lojasiewicz, Sur les trajectoires du gradient d'une fonction analytique, Seminari di Geometria 1982-1983, Università di Bologna, Istituto di Geometria, Dipartimento di Matematica, 1984, pp. 115-117.
- [19] O. L. Mangasarian and S. Fromovitz, The Fritz-John necessary optimality conditions in presence of equality and inequality constraints, *Journal of Mathematical Analysis and Applications* 17 pp. 37–47, 1967.
- [20] J. M. Martínez, Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. *Journal of Optimization Theory and Applications* 111, pp. 39-58 ,2001.

- [21] J. M. Martínez and E. A. Pilotta. Inexact restoration algorithms for constrained optimization, *Journal of Optimization Theory and Applications* 104, pp. 135-163, 2000.
- [22] J. M. Martínez and B. F. Svaiter, A practical optimality condition without constraint qualifications for nonlinear programming, *Journal of Optimization Theory and Applications* 118, pp. 117–133, 2003.
- [23] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.
- [24] 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.
- [25] M. L. Schuverdt, *Métodos de Lagrangiano Aumentado com convergência usando a condição de dependência linear positiva constante*, Tese de Doutorado, Departamento de Matemática Aplicada, Universidade Estadual de Campinas, 2006.