# A new sequential optimality condition for constrained optimization and algorithmic consequences *

Roberto Andreani † J. M. Martínez ‡ Benar F. Svaiter §

November 13, 2009.

**Abstract**

Necessary first-order sequential optimality conditions provide adequate theoretical tools to justify stopping criteria for nonlinear programming solvers. These conditions are satisfied by local minimizers of optimization problems independently of the fulfillment of constraint qualifications. A new strong sequential optimality condition is introduced in the present paper. A proof that a well established Augmented Lagrangian algorithm produces sequences whose limits satisfy the new condition is given. 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 nonlinear programming problems are iterative. Consequently, implementations include stopping criteria that generally indicate that the current iterate is close to a solution. Computer codes usually test the approximate fulfillment of the KKT conditions [20]. 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 $x^*$ is the limit of a sequence of points that satisfy the approximate KKT test with tolerances going to zero [2, 22]. In other words, every local minimizer satisfies an Approximate KKT condition (AKKT) [2, 22]. 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 the KKT approximate criterion may lead to wrong conclusions. Consider the problem

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

*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 .

†Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas SP, Brazil. E-mail: andreani@ime.unicamp.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 Pure and Applied Mathematics (IMPA), Rio de Janeiro, Brazil.

The solution of this problem is given by $x_1^* = 0, x_2^* = 2$. Consider the point $(\varepsilon, 1)$ for $\varepsilon > 0$, small. The gradient of the objective function at this point is $(0, -1)$ and the gradients of the constraints are $(1, 0)$ and $(1, \varepsilon)$. 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 $(\varepsilon, 1)$ is almost feasible in the sense that the sup-norm of the constraints vector is $\varepsilon$. This means that for arbitrary small $\varepsilon > 0$, the point $(\varepsilon, 1)$ 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 AGP and L-AGP conditions introduced in [19] and [2] are satisfied by the wrong point $(0, 1)$ and stopping tests based on them would be misleading too.

In this paper we introduce the Complementary AKKT (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. Observe that this requirement is not satisfied in the example above, therefore the "wrong" point $(0, 1)$ 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. At a first sight this is a paradoxal decision, because KKT conditions do not need to be fulfilled at local minimizers, if constraint qualifications [6] do not hold. Nevertheless, computational practice can be justified because it is possible to show that any local minimizer has the property of having arbitrary 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 to avoid stopping at false approximate minimizers, as shown in the example above.

This paper is organized as follows:

In Section 2 we define rigorously the CAKKT condition and we prove that local minimizers necessarily satisfy it. In Section 3 we prove that CAKKT is stronger than the L-AGP condition given in [2] and we show that CAKKT is a sufficient optimality condition in convex problems. In Section 4 we prove that the Augmented Lagrangian method defined in [1] produces CAKKT sequences if one assumes that a sum-of-squares infeasiblity measure satisfies the Kurdyka-Lojasiewicz inequality [14, 15]. Conclusions will be stated in the final section of this paper.

**Notation.**

- $I\!N = \{0, 1, 2, \ldots\}$.
- $\|\cdot\|$ denotes an arbitrary norm.
- If $h : I\!R^n \to I\!R^m$ we denote $\nabla h = (\nabla h_1, \ldots, \nabla h_m)$.
- $I\!R_+ = \{t \in I\!R \mid t \geq 0\}$.
- If $v \in I\!R^n$, we denote $v_+ = (\max\{v_1, 0\}, \ldots, \max\{v_n, 0\})^T$.
- If $a \in I\!R$, we denote $a_+^2 = (a_+)^2 = a_+ a_+$.
- If $v \in I\!R^n$, we denote $v_- = (\min\{v_1, 0\}, \ldots, \min\{v_n, 0\})^T$.

- $B(x,\delta) = \{z \in I\!R^n \mid \|z - x\| \leq \delta\}$.
- $P_\Omega(x)$ is the Euclidean projection of $x$ on $\Omega$.

## 2 CAKKT is a necessary optimality condition

We consider the Nonlinear Programming problem in the form

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

where $f : I\!R^n \to I\!R$, $h : I\!R^n \to I\!R^m$, $g : I\!R^n \to I\!R^p$ are smooth.

We say that $x^* \in I\!R^n$ fulfills the Complementary Approximate-KKT condition (CAKKT) if

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

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

- For all $k \in I\!N$ there exist $\lambda^k \in I\!R^m$, $\mu^k \in I\!R^p_+$ such that

  $$\lim_{k\to\infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| = 0 \tag{3}$$

  and

  $$\lim_{k\to\infty} \lambda_i^k h_i(x^k) = \lim_{k\to\infty} \mu_j^k g_j(x^k) = 0 \text{ for all } i = 1,\ldots,m, j = 1,\ldots,p. \tag{4}$$

Points that satisfy the CAKKT condition will be called "CAKKT points". Note that points $x^*$ that satisfy the KKT conditions necessarily fulfill CAKKT, taking $x^k = x^*, \lambda^k = \lambda^*, \mu^k = \mu^*$ for all $k \in I\!N$, where $\lambda^* \in I\!R^m, \mu^* \in I\!R^p_+$ are the Lagrange multipliers associated with $x^*$. The interesting cases are the ones in which KKT does not hold.

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

**Lemma 2.1**. *The feasible point $x^*$ satisfies the CAKKT condition if, and only if, there exist sequences $\{x^k\} \subset I\!R^n$, $\{\lambda^k\} \subset I\!R^m$, $\{\mu^k\} \subset I\!R^p$ such that $\lim_{k\to\infty} x^k = x^*$, (3) and (4) hold and, in addition, there exists a nonnegative sequence $\varepsilon_k$ that tends to zero and*

$$\mu_i^k \geq -\varepsilon_k \quad \textit{for all} \quad i = 1,\ldots,p,\ k \in I\!N. \tag{5}$$

*Proof.* The fact that CAKKT implies (5) is trivial. On the other hand, if (5) takes place, by the continuity of $\nabla g$, it is easy to see that (3) and (4) remain true replacing $\mu_i^k$ by $\max\{\mu_i^k, 0\}$. □

**Lemma 2.2**. *Assume that the feasible point $x^*$ satisfies the CAKKT condition. Then, there exist sequences $\{x^k\} \subset I\!R^n, \{\lambda^k\} \subset I\!R^m, \{\mu^k\} \subset I\!R^p_+$ such that $\lim_{k\to\infty} x^k = x^*$, (3) and (4) hold and, in addition,*

$$\mu_i^k = 0 \textit{ for all } i \textit{ such that } g_i(x^*) < 0. \tag{6}$$

*Proof.* Assume that $x^*$ satisfies CAKKT and let $x^k, \lambda^k, \mu^k$ be such that (3) and (4) hold.
If $g_i(x^*) < 0$, then, by the continuity of $g_i$ and (4) one has that

$$\lim_{k\to\infty} \mu_i^k = 0. \tag{7}$$

Define, for all $i$ such that $g_i(x^*) < 0$, $k \in I\!N$, $\tilde{\mu}_i^k = 0$. Clearly, (4) and (6) hold if one replaces $\mu_i^k$ by $\tilde{\mu}_i^k$. Moreover, by (7) and the boundedness of $\nabla g_i(x^k)$, (3) also holds replacing $\mu_i^k$ by $\tilde{\mu}_i^k$. This completes the proof. □

We say [2] that a feasible point $x^*$ satisfies the weaker condition AKKT if there exists sequences $\{x^k\} \subset I\!R^n, \{\lambda^k\} \subset I\!R^m, \{\mu^k\} \subset I\!R^p_+$ such that $\lim_{k\to\infty} x^k = x^*$ and (3, 6) hold.

Properties (3) and (4) provide the natural stopping criterion associated with CAKKT. Given small positive tolerances $\varepsilon_{feas}, \varepsilon_{opt}, \varepsilon_{mult}$ corresponding to feasibility, optimality (3), and the new condition (4), an algorithm that aims to solve (2) should be stopped declaring "convergence" when, for suitable multipliers $\lambda^k \in I\!R^m$, $\mu^k \in I\!R^p_+$,

$$\|h(x^k)\| \leq \varepsilon_{feas}, \|g(x^k)_+\| \leq \varepsilon_{feas}, \tag{8}$$

$$\|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| \leq \varepsilon_{opt} \tag{9}$$

and

$$|\lambda_i^k h_i(x^k)| \leq \varepsilon_{mult}, \quad |\mu_j^k g_j(x^k)| \leq \varepsilon_{mult} \text{ for all } i = 1, \ldots, m, \ j = 1, \ldots, p. \tag{10}$$

In the nonlinear programming software Algencan [1] [1] and other Augmented Lagrangian algorithms [8] the convergence stopping criterion is given by (8), (9) and

$$\mu_i^k = 0 \text{ whenever } g_i(x^k) < -\varepsilon_{comp}. \tag{11}$$

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 $n = 2, m = 2, p = 0$,

$$f(x_1, x_2) = \frac{(x_2 - 2)^2}{2}. \tag{12}$$

$$h_1(x_1, x_2) = x_1, h_2(x_1, x_2) = x_1 x_2. \tag{13}$$

Taking $\varepsilon_{feas} = \varepsilon_{opt} = \varepsilon_{comp} = \varepsilon_{mult} = \varepsilon$, one has (with $\|\cdot\| = \|\cdot\|_\infty$) that the point $x^k = (\varepsilon, 1)^T$ satisfies (8,9,11) 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 (10).

The proof that CAKKT is a genuine necessary optimality condition is given below. This proof uses a penalty reduction technique employed in [2, 6, 13, 19] for analyzing different optimality conditions.

**Theorem 2.1.** *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

$$\text{Minimize } f(x) + \|x - x^*\|_2^2 \text{ subject to } h(x) = 0, g(x) \leq 0, x \in B(x^*, \delta). \tag{14}$$

Clearly, $x^*$ is the unique solution of (14). Let $x^k$ be a solution of

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

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

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 $g(x^*)_+ = 0$,

$$f(x^k) + \|x^k - x^*\|_2^2 + \rho_k \left[ \|h(x^k)\|_2^2 + \sum_{i=1}^{p} g_i(x^k)_+^2 \right] \leq f(x^*). \tag{15}$$

By the convergence theory of external penalty methods [9] one has that $\lim_{k \to \infty} x^k = x^*$. Therefore, by (15) and the continuity of $f$,

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

Thus,

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

Defining

$$\lambda^k = 2\rho_k h(x^k), \quad \mu^k = 2\rho_k g(x^k)_+, \tag{16}$$

we obtain:

$$\lim_{k \to \infty} \left[ \sum_{i=1}^{m} |\lambda_i^k h_i(x^k)| + \sum_{i=1}^{p} \mu_i^k g_i(x^k)_+ \right] = 0. \tag{17}$$

By (16), $\mu_i^k = 0$ when $g_i(x^k) < 0$. Therefore, by (17),

$$\lim_{k \to \infty} \left[ \sum_{i=1}^{m} |\lambda_i^k h_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k g_i(x^k)| \right] = 0. \tag{18}$$

Thus, (4) follows from (16) and (18).

The proof of (3) is standard. For $k$ large enough, one has that $\|x^k - x^*\| < \delta$, therefore, the gradient of the objective function must vanish. Thus,

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

By (16), since $\|x^k - x^*\| \to 0$ we have that

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

Thus, (3) is proved. □

# 3 Strongness of the CAKKT condition

A necessary optimality condition should be as strong as possible. Moreover, as we will see in the next section, practical (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) [2]. The CPLD condition was introduced in [21] and, in [4], it was proved that it implies the quasinormality constraint qualification [6]. Since CPLD is a weak constraint qualification (strictly weaker than the Mangasarian-Fromovitz condition [16]) the property

$$\text{KKT or not-CPLD} \tag{19}$$

is a strong necessary optimality condition. The property that CAKKT implies (19) will follow as a corollary of a stronger result. For stating this result we must recall a different sequential optimality condition, introduced in [19], analyzed in [2] and employed in several algorithmically oriented papers ([10, 11, 12, 17, 18] and others).

We say 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

$$\lim_{k\to\infty} \|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\| = 0, \tag{20}$$

where $\Omega_k$ is the set of points $x \in I\!R^n$ that satisfy:

$$\nabla h(x^k)^T (x - x^k) = 0 \tag{21}$$

and

$$\nabla g(x^k)^T (x - x^k) + g(x^k)_- \le 0. \tag{22}$$

Note that $x^k$ always belong to the polytope defined by (21, 22).

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 [2] that LAGP is strictly stronger than AGP. Moreover, both AGP and LAGP are strictly stronger than (19) [2, 11].

Let us show now that CAKKT implies AGP.

**Theorem 3.1.** *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 $I_{lin}$.*

*Proof.* Assume that $\{x^k\} \subset I\!R^n$ converges to $x^*$ and satisfies (3, 4). Let $y^k$ be the solution of

$$\text{Minimize } \|[x^k - \nabla f(x^k)] - y\|_2^2 \tag{23}$$

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

$$\nabla h_i(x^k)^T (y - x^k) = 0, i = 1, \dots, m,$$

$$\nabla g_i(x^k)^T (y - x^k) \ge 0, \text{ if } g_i(x^k) \ge 0,$$

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

[2]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 $x^*$.

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 $\lim_{k\to\infty} \|y^k - x^k\| = 0$. Since the constraints of (23) are linear, the KKT conditions are verified at $y^k$. Therefore, there exist $\{\widehat{\lambda}^k\} \subset I\!R^m$, $\{\widehat{\mu}^k\} \subset I\!R^p_+$ such that:

$$[y^k - x^k] + \nabla f(x^k) + \nabla h(x^k)\widehat{\lambda}^k + \nabla g(x^k)\widehat{\mu}^k = 0, \tag{24}$$

$$\nabla h_i(x^k)^T(y^k - x^k) = 0, i = 1, \dots, m, \tag{25}$$

$$\nabla g_i(x^k)^T(y^k - x^k) \le 0, \text{ if } g_i(x^k) \ge 0, \tag{26}$$

$$g_i(x^k) + \nabla g_i(x^k)^T(y^k - x^k) \le 0 \text{ if } g_i(x^k) < 0. \tag{27}$$

$$\widehat{\mu}_i^k \nabla g_i(x^k)^T(y^k - x^k) = 0 \text{ if } g_i(x^k) \ge 0, \tag{28}$$

and

$$\widehat{\mu}_i^k g_i(x^k) + \widehat{\mu}_i^k \nabla g_i(x^k)^T(y^k - x^k) = 0 \text{ if } g_i(x^k) < 0. \tag{29}$$

By (25–28), pre-multiplying (24) by $(y^k - x^k)^T$ we obtain:

$$\|y^k - x^k\|_2^2 + \nabla f(x^k)^T(y^k - x^k) + \sum_{g_i(x^k)<0} \widehat{\mu}_i^k \nabla g_i(x^k)^T(y^k - x^k) = 0. \tag{30}$$

By (29), 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 (30),

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

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

$$\|y^k - x^k\|_2^2 \le -\nabla f(x^k)^T(y^k - x^k). \tag{31}$$

Now, by (3), there exist sequences $\{\lambda^k\} \subset I\!R^p$, $\{\mu^k\} \subset I\!R^p_+$, $\{v^k\} \subset I\!R^n$ such that

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

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 (25),

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

$$= \sum_{g_i(x^k)<0} \mu_i^k \nabla g_i(x^k)^T(y^k - x^k) + \sum_{g_i(x^k)\ge 0} \mu_i^k \nabla g_i(x^k)^T(y^k - x^k) - (y^k - x^k)^T v^k.$$

By (26), 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,

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

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

Thus, by (27), 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 (31),

$$\|y^k - x^k\|_2^2 \leq \sum_{g_i(x^k)<0} |\mu_i^k g_i(x^k)| + \|v^k\|_2 \|y^k - x^k\|_2. \tag{32}$$

By (4), $\lim_{k\to\infty} |\mu_i^k g_i(x^k)| = 0$ for all $i$, so the sequence $\{\|y^k - x^k\|\}$ is bounded and, taking limits in both sides of (32), we obtain that $\lim_{k\to\infty} \|y^k - x^k\| = 0$ 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. □

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 (12, 13). Define $x^k = (1/k, 1)^T$. Clearly, $\Omega_k = \{x^k\}$ for all $k \in I\!N$. 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 $k \in I\!N$. Therefore, $x^* = (0, 1)^T$ satisfies AGP.

Let us show now that a sequence $x^k$ fulfilling (3,4) 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:

$$\lim_{k\to\infty} x_1^{(k)} = 0, \ \lim_{k\to\infty} x_2^{(k)} = 1, \tag{33}$$

$$\lim_{k\to\infty} \begin{pmatrix} 0 \\ x_2^{(k)} - 2 \end{pmatrix} + \lambda_1^{(k)} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2^{(k)} \begin{pmatrix} x_2^{(k)} \\ x_1^{(k)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{34}$$

$$\lim_{k\to\infty} \lambda_1^{(k)} x_1^{(k)} = 0 \tag{35}$$

and

$$\lim_{k\to\infty} \lambda_2^{(k)} x_1^{(k)} x_2^{(k)} = 0. \tag{36}$$

By (34) we have:

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

and

$$\lim_{k\to\infty} x_2^{(k)} + \lambda_2^{(k)} x_1^{(k)} = 2. \tag{37}$$

By (33) and (37), we have:

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

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

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

This contradicts (36). Therefore, a sequence satisfying (33–36) cannot exist. Therefore $x^*$ is not a CAKKT point. By Theorem 3.1, 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 3.2.** *Assume that, in problem (2), the functions $f$ and $g_i, i = 1, \ldots, p$ are convex and $h_1, \ldots, h_m$ 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 (3,4). Let $z$ be a feasible point of (2). By the convexity of $f$ and the constraints, we have, for all $k \in I\!N$,

$$f(z) \geq f(x^k) + \nabla f(x^k)^T(z - x^k),$$

$$h_i(z) = h_i(x^k) + \nabla h_i(x^k)^T(z - x^k) = 0, i = 1, \ldots, m,$$

$$g_i(z) \geq g_i(x^k) + \nabla g_i(x^k)^T(z - x^k), i = 1, \ldots, p.$$

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

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

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

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

Thus,

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

Then, by the continuity of $f$ and the properties (3, 4), we have that $f(z) \geq f(x^*)$. □

# 4 A practical algorithm that generates CAKKT points

In this section we discuss an implementable algorithm that generates sequences converging to CAKKT points. We exclude from our analysis "global optimization algorithms" like the one introduced in [7] that guaranteedly converge to global minimizers using more expensive procedures that 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 2.1.

Our results in this section make use a generalization of the Kurdyka-Lojasiewicz (KL) inequality [5, 14, 15].

We say that the smooth function $\Phi : I\!R^n \to I\!R$ satisfies the GKL inequality at $\bar{x}$ if there exist $\delta > 0, \varphi : B(\bar{x}, \varepsilon)$ such that $\lim_{x \to \bar{x}} \varphi(x) = 0$ and for all $x \in B(\bar{x}, \delta)$, one has:

$$|\Phi(x) - \Phi(\bar{x})| \leq \varphi(x) \|\nabla \Phi(x)\|. \tag{38}$$

One says that $\Phi : I\!R^n \to I\!R$ satisfies the Kurdyka-Lojasiewicz (KL) inequality [5, 14, 15] at $\bar{x}$ if there exists $\delta > 0$, $\theta \in (0, 1)$, $c > 0$ such that, for all $x \in B(\bar{x}, \delta)$,

$$|\Phi(x) - \Phi(\bar{x})|^{\theta} \leq c \|\nabla \Phi(x)\|. \tag{39}$$

Clearly, the fulfillment of (39) implies that (38) holds, but the reciprocal is not true. To see this, define

$$\Phi(x) = e^{-\frac{1}{|x|}}$$

if $x \neq 0$, $\Phi(0) = 0$.

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

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

However,

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

Therefore, $\Phi$ does not satisfy KL.

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

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0, g(x) \leq 0, \underline{h}(x) = 0, \underline{g}(x) \leq 0, \tag{40}$$

where $f, h, g$ are as in (2) and $\underline{h} : I\!R^n \to I\!R^{\underline{m}}, \underline{g} : I\!R^n \to I\!R^{\underline{p}}$ 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 $x \in I\!R^n, \bar{\lambda} \in I\!R^m$, $\bar{\mu} \in I\!R^p_+$, $\rho > 0$, we define the "displaced infeasibility" by:

$$\Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x) = \frac{1}{2} \left[ \left\| h(x) + \frac{\bar{\lambda}}{\rho} \right\|_2^2 + \left\| \left( g(x) + \frac{\bar{\mu}}{\rho} \right)_+ \right\|_2^2 \right]. \tag{41}$$

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

$$L_{\bar{\lambda}, \bar{\mu}, \rho}(x) = f(x) + \rho \Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x). \tag{42}$$

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, 21] are KKT points.

[3] A freely available implementation of this method may be found in www.ime.usp.br/∼egbirgin/tango.

**Algorithm 4.1.** Let $\varepsilon_k \downarrow 0$, $\bar{\lambda}^k \in [\lambda_{min}, \lambda_{max}]^m$, $\bar{\mu}^k \in [0, \mu_{max}]^p$ for all $k \in I\!N$, $\rho_1 > 0$, $r \in (0,1)$, $\gamma > 1$.

For all $k = 1, 2, \ldots$ we compute $x^k \in I\!R^n$, $\underline{\lambda}^k \in I\!R^{\underline{m}}$, $\underline{\mu}^k \in I\!R_+^{\underline{p}}$ such that:

$$\|\nabla L_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) + \nabla \underline{h}(x^k)\underline{\lambda}^k + \nabla \underline{g}(x^k)\underline{\mu}^k\| \leq \varepsilon_k, \tag{43}$$

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

and

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

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

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

If $k = 1$ or

$$\max\{\|h(x^k)\|, \|V^k\|\} \leq r \max\{\|h(x^{k-1})\|, \|V^{k-1}\|\} \tag{44}$$

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

**Remarks.**

- In general [1], we define $\rho_{k+1} = \rho_k$ if (44) holds and $\rho_{k+1} = \gamma \rho_k$ otherwise. 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 the convergence theorem below we 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 (complicate 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. In any case, 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 4.2 also includes, as a particular case, the classical External Penalty method [9] with an approximate stopping criterion for the subproblem.

- The convergence results of [1] apply to Algorithm 4.1. In [1] it was proved that feasible limit points that satisfy the CPLD constraint qualification are KKT points. Under the GKL assumption, Theorem 4.1 presents a stronger result, showing that feasible limit points satisfy the CAKKT condition.

**Theorem 4.1.** *Assume that $x^*$ is a feasible limit point of a sequence generated by Algorithm 4.1 and that the sequences $\{\underline{\lambda}^k\}, \{\underline{\mu}^k\}$ are bounded. In addition, assume that there exist $\delta > 0, \bar{\rho} > 0$, $\varphi : B(x^*, \delta) \to I\!R^n$, $\lim_{x \to x^*} \varphi(x) = 0$, such that, for all $x \in B(x^*, \delta), \bar{\lambda} \in [\lambda_{min}, \lambda_{max}], \bar{\mu} \in [0, \mu_{max}]$, $\rho \geq \bar{\rho}$,*

$$|\Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x) - \Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x^*)| \leq \varphi(x) \|\nabla \Phi_{\bar{\lambda}, \bar{\mu}, \rho}(x)\|. \tag{45}$$

*Then, $x^*$ satisfies CAKKT.*

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

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k), \;\; \mu^k = (\bar{\mu}^k + \rho_k g(x^k))_+. \tag{46}$$

By (42), (43) and (46) one has:

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

for all $k = 1, 2, \ldots$.

Therefore, we only need to prove (4), 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 (4) 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 (4) in the case of upper-level constraints, let us consider first the case in which $i \in \{1, \ldots, p\}$ is such that $g_i(x^*) = 0$ and $g_i(x^k) \le 0$ for an infinite set of indices $K_1$. By (46), 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 $g_i(x^k) \le 0$, we have that $0 \le \mu_i^k \le \bar{\mu}_i^k \le \mu_{max}$. 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:

$$\lim_{k \in K_1} \mu_i^k g_i(x^k) = 0. \tag{47}$$

This result will be used later in the proof.

Consider now the case in which $\{\rho_k\}$ is bounded. By (46), we have that $\{\lambda^k\}$ and $\{\mu^k\}$ are bounded. Moreover, by the choice of $\rho_{k+1}$, we have that $\lim_{k\to\infty} \|V^k\| = 0$. Clearly, since $x^*$ is feasible, $\lim_{k\to\infty} \lambda_i^k h_i(x^k) = 0$ and, if $g_i(x^*) = 0$, $\lim_{k\to\infty} \mu_i^k g_i(x^k) = 0$. In the case that $g_i(x^*) < 0$, since $V^k \to 0$, we have that

$$\lim_{k\to\infty} \frac{\bar{\mu}_i^k}{\rho_k} = 0,$$

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

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

Thus, $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$ for $k$ large enough. Therefore, by (46), $\mu_i^k = 0$ for $k$ large enough. This implies that $\lim_{k\to\infty} \mu_i^k g_i(x^k) = 0$ also in the case that $g_i(x^*) < 0$.

This completes the proof of (4) in the case that $\{\rho_k\}$ is bounded.

Let us consider now the case in which $\lim_{k\to\infty} \rho_k = \infty$. By (42), (43) and 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 (45), for $k$ large enough we have:

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

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

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

with $\lim_{k\to\infty} c_1(k) = 0$. 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:

$$\lim_{k\to\infty} \rho_k |\Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) - \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^*)| = 0. \tag{48}$$

Now,

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

$$= \frac{1}{2}\left[\sum_{i=1}^{m}\left[\frac{\bar\lambda_i^k}{\rho_k}\right]^2 + \sum_{i=1}^{p}\max\left\{\frac{\bar\mu_i^k + \rho_k g_i(x^*)}{\rho_k}, 0\right\}^2\right]$$

$$= \frac{1}{2\rho_k^2}\left[\sum_{i=1}^{m}(\bar\lambda_i^k)^2 + \sum_{i=1}^{p}\max\{\bar\mu_i^k + \rho_k g_i(x^*), 0\}^2\right].$$

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

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

Therefore, $\lim_{k\to\infty} \rho_k \Phi_{\bar\lambda^k,\bar\mu^k,\rho_k}(x^*) = 0$. Thus, by (48),

$$\lim_{k\to\infty} \rho_k \Phi_{\bar\lambda^k,\bar\mu^k,\rho_k}(x^k) = 0. \tag{49}$$

But

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

$$= \frac{1}{2}\sum_{i=1}^{m}\left|\bar\lambda_i^k + \rho_k h_i(x^k)\right|\left|h_i(x^k) + \frac{\bar\lambda_i^k}{\rho_k}\right| + \sum_{i=1}^{p}\left(\bar\mu_i^k + \rho_k g_i(x^k)\right)_+\left(g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right)_+$$

$$= \frac{1}{2}\sum_{i=1}^{m}|\lambda_i^k|\left|h_i(x^k) + \frac{\bar\lambda_i^k}{\rho_k}\right| + \sum_{i=1}^{p}\mu_i^k\left(g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right)_+.$$

Thus, by (49),

$$\lim_{k\to\infty} \lambda_i^k\left[h_i(x^k) + \frac{\bar\lambda_i^k}{\rho_k}\right] = 0 \tag{50}$$

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

$$\lim_{k\to\infty} \mu_i^k\left(g_i(x^k) + \frac{\bar\mu_i^k}{\rho_k}\right)_+ = 0 \tag{51}$$

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

Now,

$$\lambda_i^k\left[h_i(x^k) + \frac{\bar\lambda_i^k}{\rho_k}\right] = \lambda_i^k h_i(x^k) + \lambda_i^k\frac{\bar\lambda_i^k}{\rho_k} \tag{52}$$

and

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

Therefore, by (50) and (52),

$$\lim_{k\to\infty} \lambda_i^k h_i(x^k) = 0 \tag{53}$$

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

By (46), if $g_i(x^*) < 0$ we have that $\mu_i^k = 0$ for $k$ large enough. This implies that $\lim_{k\to\infty} \mu_i^k g_i(x^k) = 0$ in this case.

By (47), 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:

$$\mu_i^k \left( g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right)_+ = \left( \mu_i^k g_i(x^k) + \frac{\bar{\mu}_i^k \mu_i^k}{\rho_k} \right)_+ \tag{54}$$

and

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

Then, since $g_i(x^k) \to 0$, we deduce that $\lim_{k \in K_2} \mu_i^k \bar{\mu}_i^k / \rho_k = 0$. By (51) and (54), 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

$$\lim_{k \in K_2} \mu_i^k g_i(x^k) = 0. \tag{55}$$

Since $K_1 \cup K_2 = \{1, 2, 3, \ldots\}$, (47) and (55) imply that, for all $i = 1, \ldots, p$,

$$\lim_{k \to \infty} \mu_i^k g_i(x^k) = 0. \tag{56}$$

By (53) and (56) the proof is complete also for the case in which $\rho_k$ tends to infinity. □

**Counter-example**

We will show that, if the GKL assumption does not hold, Algorithm 4.1 may generate a sequence that does not satisfy (4).

Consider the problem

$$\text{Minimize } x \ \text{ subject to } \ h(x) = 0, \tag{57}$$

where

$$h(x) = x^4 \sin\left(\frac{1}{x}\right)$$

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

We will use Algorithm 4.1 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 $\rho_k \to \infty$, the sequence $\{x_k\}$ generated by the algorithm tends to $x_* = 0$ and does not satisfy

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

We have:

$$h'(x) = x^2 \left[ 4x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) \right] \tag{58}$$

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

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

$$R(x) = 4x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) + x^3 \tag{59}$$

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

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

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

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

and

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

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

$$R(x_k) = 0. \tag{60}$$

Clearly, $\lim_{k\to\infty} x_k = 0$. Moreover, since $x_k \in (y_k, z_k)$ we have that

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

and, consequently,

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

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

$$\sin\left(\frac{1}{x_k}\right) > 0. \tag{61}$$

By (59) and (60), since $x_k \to 0$, we have:

$$\lim_{k\to\infty} \cos\left(\frac{1}{x_k}\right) = 0. \tag{62}$$

By (61) and (62), we have that

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

Therefore, for $k$ large enough,

$$\sin\left(\frac{1}{x_k}\right) > \frac{1}{2}. \tag{63}$$

By (60) and (58),

$$h'(x_k) = -x_k^5 \tag{64}$$

for all $k = 1, 2, \ldots$.

Now, define, for $k$ large enough:

$$\rho_k = \frac{-1}{h(x_k)h'(x_k)}. \tag{65}$$

By (64), 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 (63), $\rho_k$ is well defined for $k$ large enough and $\lim_{k\to\infty} \rho_k = \infty$. Taking an appropriate subsequence we may assume, without loss of generality, that

$$\rho_{k+1} \geq \gamma \rho_k \tag{66}$$

for all $k = 1, 2, \ldots$.

Now, by (42), 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 (65),

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

This means that the sequence $\{x_k\}$ is defined by the application of Algorithm 4.1 to the problem (57), with $\bar{\lambda}^k = 0$ for all $k$ and the penalty parameters given by (65).

Now, assume that, for all $k = 1, 2, \ldots$, $\lambda_k$ is an approximate Lagrange multiplier such that (3) holds. Then:

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

Therefore, by (64),

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

and, so,

$$\lim_{k\to\infty} \lambda_k x_k^5 = 1. \tag{67}$$

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 (67) and (63), since $x_k \to 0$, we deduce that

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

Therefore, (4) does not hold.

# 5 Conclusions

In the first section of this paper we motivated the introduction of a new strong approximate KKT condition using a simple example which 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 this kind 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 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, as a matter of fact, 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, of course, all the KKT formulations have the same exact solutions. To fix ideas, 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:

$$\nabla f(x) + \nabla h(x)\lambda = 0, \quad h(x) = 0. \tag{68}$$

In fact (68) is a square nonlinear system and variations of Newton's method are many times successful for solving it. Now, (68) is obviously equivalent to:

$$\nabla f(x) + \nabla h(x)\lambda = 0, \; h(x) = 0, \; \lambda_i h_i(x) = 0, i = 1, \ldots, m \tag{69}$$

but equivalence blows up when, roughly speaking, we admit that multipliers may be infinity. In this case, the approximate fulfillment of (68) corresponds to the approximate KKT condition, but the approximate fulfillment of (69) 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 (69).

# 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 quasinormality 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] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Mass., 1999.

[7] 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.*

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

[9] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.

[10] A. Fischer and A. Friedlander, A new line search inexact restoration approach for nonlinear programming, to appear in *Computational Optimization and Applications.*

[11] 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.

[12] 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.

[13] M. R. Hestenes, *Optimization Theory: The finite dimensional case*, Wiley, NY, 1975.

[14] K. Kurdyka, On gradients of functions definable in o-minimal structures, *Annales de l'Institut Fourier* 48, pp. 769-783, 1998.

[15] S. Lojasiewicz, Sur les trajectories du gradient d'une fonction analytique, Seminari di Geometria 1982-1983, Universitá di Bologna, Istituto di Geometria, Dipartamento di Matematica, 1984, pp. 115-117.

[16] 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.

[17] 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.

[18] J. M. Martínez and E. A. Pilotta. Inexact restoration algorithms for constrained optimization, *Journal of Optimization Theory and Applications* 104, pp. 135-163, 2000.

[19] 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.

[20] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[21] 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.

[22] 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.