SIAM J. OPTIM.
Vol. 20, No. 6, pp. 3533–3554

A NEW SEQUENTIAL OPTIMALITY CONDITION FOR CONSTRAINED OPTIMIZATION AND ALGORITHMIC CONSEQUENCES*

ROBERTO ANDREANI†, J. M. MARTÍNEZ†, AND B. F. SVAITER‡

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 is proved that a well-established augmented Lagrangian algorithm produces sequences whose limits satisfy the new condition. Practical consequences are discussed.

Key words. nonlinear programming, optimality conditions, approximate KKT conditions, stopping criteria

AMS subject classifications. 90C30, 49K99, 65K05

DOI. 10.1137/090777189

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 [25]. 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, 27]. In terms of [2], this means that every local minimizer satisfies an approximate KKT (AKKT) condition [2, 27]. 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

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

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

*Received by the editors November 16, 2009; accepted for publication (in revised form) October 6, 2010; published electronically December 16, 2010. 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 (300900/2009-0 480101/2008-6, 303583/2008-8), and FAPERJ: E-26/102.821/2008.
http://www.siam.org/journals/siopt/20-6/77718.html
†Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas SP, Brazil (andreani@ime.unicamp.br, martinez@ime.unicamp.br).
‡Institute of Pure and Applied Mathematics, Rio de Janeiro, Brazil (benar@impa.br).

constraints vector is ε. 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 (LAGP) conditions introduced in [24] 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 such as the one described above. In the CAKKT stopping test we require, in addition to the usual AKKT 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 "pointwise" 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 pointwise optimality condition is necessarily satisfied at a local minimizer but not at "approximate local minimizers." Since, in the iterative framework, one never gets exact solutions, numerical practice always leads to test the "approximate fulfillment" of pointwise optimality conditions. More specifically, practical codes usually test relaxed versions of the KKT conditions. This could be a paradoxal 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, 24] 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 infeasiblity measure satisfies the Lojasiewicz inequality [19, 20]. Conclusions will be stated in the final section of this paper.

Notation.

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

2. Preliminaries.

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

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

$$h(x) = 0, \;\; g(x) \leq 0, \tag{3}$$

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

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

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

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

Clearly, (5) is equivalent to

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

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

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

$$\|h(x)\| \leq \varepsilon, \tag{8}$$

and

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

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

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

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 for this 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 AKKT criterion is fulfilled with tolerance ε. 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, 24].

We will formulate these concepts in terms of "sequential optimality conditions." The AKKT condition introduced in [2] corresponds to the AKKT 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

$$\lim_{k\to\infty} x^k = x^*, \tag{11}$$

$$\lim_{k\to\infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| = 0, \tag{12}$$

and

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

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,

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

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. In addition to (11), (12), and (14), CAKKT requires that

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

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 AGP optimality condition introduced by Martínez and Svaiter in [24]. We say that the feasible point x^* satisfies AGP if there exists a sequence $\{x^k\}$ such that

$$\lim_{k\to\infty} \|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\| = 0, \tag{16}$$

where

$$\begin{aligned} \Omega_k = \{x \in \mathbb{R}^n \mid \nabla h(x^k)^T(x - x^k) = 0, \ \min\{g_i(x^k), 0\} \\ + \nabla g_i(x^k)^T(x - x^k) \le 0, i = 1, \dots, p\}. \end{aligned} \tag{17}$$

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 LAGP condition is satisfied. Properties of LAGP 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 problem such as (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 AKKT conditions. This suggests that, in the vicinity of a CAKKT point, we have better chances of being close to a local minimizer than we do in the vicinity of a point that satisfies alternative AKKT conditions. 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 derives from the 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. 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 CAKKT condition for problem (2) if

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

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

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

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

and

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

Points that satisfy the CAKKT condition will be called "CAKKT points." If x^* satisfies the KKT conditions then it 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 nonnegativity of μ^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 $\lim_{k\to\infty} x^k = x^*$,* (18) *and* (19) *hold, and, in addition, there exists a nonnegative sequence ε_k that tends to zero such that*

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

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

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 $\lim_{k\to\infty} x^k = x^*$,* (18) *and* (19) *hold, and, in addition,*

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

Proof. Assume that x^* satisfies CAKKT and let x^k, λ^k, μ^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

$$\lim_{k\to\infty} \mu_i^k = 0. \tag{22}$$

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 μ_i^k by $\tilde{\mu}_i^k$. Moreover, by (22) and the boundedness of $\nabla g_i(x^k)$, (18) also holds replacing μ_i^k by $\tilde{\mu}_i^k$. This completes the proof. ∎

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 stop and declare "convergence" when, for suitable multipliers $\lambda^k \in \mathbb{R}^m$, $\mu^k \in \mathbb{R}^p_+$,

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

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

and

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

In the nonlinear programming software Algencan [1][1] and other augmented Lagrangian algorithms [11] the convergence stopping criterion is given by (23), (24), and

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

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

$$f(x_1, x_2) = \frac{(x_2 - 2)^2}{2} \tag{27}$$

and

$$h_1(x_1, x_2) = x_1, h_2(x_1, x_2) = x_1 x_2. \tag{28}$$

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,

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

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, 24] for analyzing different optimality conditions.

THEOREM 3.3. *Let x^* be a local minimizer of* (2). *Then, x^* satisfies CAKKT.*

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

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

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

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

By the compactness of $B(x^*, \delta)$, x^k is well defined for all k. Clearly,

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

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

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

Thus,

$$\lim_{k \to \infty} \left[\sum_{i=1}^m \rho_k h_i(x^k)^2 + \sum_{i=1}^p \rho_k g_i(x^k)_+^2 \right] = 0. \tag{32}$$

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

Define

$$\lambda^k = \rho_k h(x^k), \;\; \mu^k = \rho_k g(x^k)_+. \tag{33}$$

Then, by (32) and (33),

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

Thus, (19) follows from (33) 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 (33), 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, (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 fufillment 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, 14, 27]. 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 are in the proximity of a CAKKT point have more chances to be close to local minimizers than points that approximately fulfill 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 AKKT 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 (CPLD) constraint qualification.[2] The CPLD condition was introduced in

[2]A feasible point x^* is said to satisfy the CPLD condition if the existence of linear dependent gradients of active constraints, with nonnegative coefficients associated with inequalities, implies that the same gradients are linearly dependent in a neighborhood of x^* [4, 26].

[26] 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 [21]) the property

$$\text{KKT or not-CPLD} \tag{35}$$

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 use a different sequential optimality condition introduced in [24], analyzed in [2], and employed in several algorithmically oriented papers ([13, 14, 15, 22, 23] and others).

Recall that a feasible point x^* satisfies the AGP property if there exists a sequence $\{x^k\}$ that converges to x^* and satisfies

$$\lim_{k\to\infty} \|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\| = 0, \tag{36}$$

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

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

and

$$\nabla g(x^k)^T(x - x^k) + g(x^k)_- \le 0. \tag{38}$$

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 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].

We now show 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

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

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

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

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

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

The objective function of (39) is a strictly convex quadratic and Ω_k is defined by linear constraints. Since $x^k \in \Omega_k$ one has that Ω_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 (39) are linear, the KKT conditions are fulfilled at y^k. Therefore, there exist $\{\widehat{\lambda}^k\} \subset \mathbb{R}^m$, $\{\widehat{\mu}^k\} \subset \mathbb{R}^p_+$ such that

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

$$\nabla h_i(x^k)^T(y^k - x^k) = 0, i = 1, \ldots, m, \tag{41}$$

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

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

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

and

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

By (41), (42), (43), and (44), premultiplying (40) by $(y^k - x^k)^T$, we obtain

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

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

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

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

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

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 (41),

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

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

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,

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

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

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

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

Therefore, by (47),

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

By (19), $\lim_{k\to\infty} |\mu_i^k g_i(x^k)| = 0$ for all i, so the sequence $\{\|y^k - x^k\|\}$ is bounded and, taking limits in both sides of (48), 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.

Finally, if x^* satisfies the CPLD constraint qualification, the fulfillment of KKT follows from the fulfillment of AGP [2, 14, 27]. ∎

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) and (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

$$\lim_{k\to\infty} x_1^{(k)} = 0, \ \lim_{k\to\infty} x_2^{(k)} = 1 \tag{49}$$

such that, by (18),

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

and, by (19),

$$\lim_{k\to\infty} \lambda_1^{(k)} x_1^{(k)} = 0 \tag{51}$$

and

$$\lim_{k\to\infty} \lambda_2^{(k)} x_1^{(k)} x_2^{(k)} = 0. \tag{52}$$

By (50) we have

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

and

$$\lim_{k\to\infty} x_2^{(k)} + \lambda_2^{(k)} x_1^{(k)} = 2. \tag{53}$$

By (49) and (53), 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 (52). Therefore, a sequence satisfying (49), (50), (51), and (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 $k \in \mathbb{N}$,

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

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

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

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

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

Thus,

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

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

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.

Our results in this section make use a generalization of the Lojasiewicz inequality [5, 9, 19, 20].

One says that the continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}$ satisfies the Lojasiewicz inequality [5, 19, 20] at $\bar{x}$ if there exists $\delta > 0$, $\theta \in (0,1)$, $c > 0$ such that, for all $x \in B(\bar{x}, \delta)$,

$$|F(x) - F(\bar{x})|^\theta \leq c\|\nabla F(x)\|. \tag{54}$$

The properties of functions that satisfy this inequality have been studied in several recent papers in connection with minimization methods, complexity theory, asymptotic analysis of partial differential equations, and tame optimization [5, 6, 9, 10, 17]. Smooth functions satisfy this inequality under fairly weak conditions. For example, analytic functions fulfill the Lojasiewicz inequality [9, 20].

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

$$|F(x) - F(\bar{x})| \leq \varphi(x)\|\nabla F(x)\|. \tag{55}$$

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

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

$F(0) = 0$.

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

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

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

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

Therefore, F does not satisfy the Lojasiewicz inequality.

If the functions $h_1(x)^2, \dots, h_m(x)^2, g_1(x)^2, \dots, g_p(x)^2$ satisfy GL at $\bar{x}$ with respect to the neighborhood $B(\bar{x}, \delta)$, it is easy to see that the function defined by $\Phi(x) = \sum_{i=1}^{m} h_i(x)^2 + \sum_{i=1}^{p} g_i(x)_+^2$ satisfies GL at $\bar{x}$ in the same neighborhood. Moreover, sums, products, quotients with non-null denominators, and compositions of analytic functions are analytic [18]. As a consequence of these two facts, the GL condition is satisfied at every feasible point of a constrained optimization problem (2), provided that the functions h_i, g_j are analytic. In fact, much weaker sufficient smoothness conditions guarantee the fulfillment of this property [9].

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

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0,\ \ g(x) \le 0,\ \ \underline{h}(x) = 0,\ \ \underline{g}(x) \le 0, \tag{56}$$

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

We say that $\underline{h}(x) = 0, \underline{g}(x) \le 0$ are "lower-level constraints." These constraints are usually simpler than the "upper-level constraints" $h(x) = 0, g(x) \le 0$, 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 $\ell \le x \le u$.

For all $x \in \mathbb{R}^n$, we define the "upper-level infeasibility" by

$$\Phi(x) = \|h(x)\|_2^2 + \|g(x)_+\|_2^2. \tag{57}$$

Note that, for all $x \in \mathbb{R}^n$,

$$\nabla\Phi(x) = 2\left[\sum_{i=1}^{m} h_i(x)\nabla h_i(x) + \sum_{i=1}^{p} g_i(x)_+ \nabla g_i(x)\right]. \tag{58}$$

For all $x \in \mathbb{R}^n$, $\bar{\lambda} \in \mathbb{R}^m$, $\bar{\mu} \in \mathbb{R}^p_+$, $\rho > 0$, the "displaced upper-level infeasibility" will be defined by

$$\Phi_{\bar{\lambda},\bar{\mu},\rho}(x) = \left[\left\|h(x) + \frac{\bar{\lambda}}{\rho}\right\|_2^2 + \left\|\left(g(x) + \frac{\bar{\mu}}{\rho}\right)_+\right\|_2^2\right]. \tag{59}$$

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

$$L_{\bar{\lambda},\bar{\mu},\rho}(x) = f(x) + \rho\Phi_{\bar{\lambda},\bar{\mu},\rho}(x). \tag{60}$$

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, 26] are KKT points.

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

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 $x^k \in \mathbb{R}^n$, $\underline{\lambda}^k \in \mathbb{R}^{\underline{m}}$, $\underline{\mu}^k \in \mathbb{R}_+^{\underline{p}}$ such that

$$\|\nabla L_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k) + \nabla \underline{h}(x^k)\underline{\lambda}^k + \nabla \underline{g}(x^k)\underline{\mu}^k\| \le \varepsilon_k, \tag{61}$$

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

and

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

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

$$V_i^k = \max\left\{g_i(x^k), \frac{-\bar{\mu}_i^k}{\rho_k}\right\}. \tag{62}$$

If $k = 1$ or

$$\max\{\|h(x^k)\|, \|V^k\|\} \le r \max\{\|h(x^{k-1})\|, \|V^{k-1}\|\} \tag{63}$$

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

Remarks.

- In general, we define $\rho_{k+1} = \rho_k$ if (63) holds and $\rho_{k+1} = \gamma\rho_k$ otherwise [1]. Here we prefer to use the more general form, in which $\rho_{k+1} \ge \rho_k$ and $\rho_{k+1} \ge \gamma\rho_k$ 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 relaxed GL assumption (64), 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 $\delta > 0$, $\varphi : B(x^*, \delta) \to \mathbb{R}$, $\lim_{x \to x^*} \varphi(x) = 0$ such that, for all $x \in B(x^*, \delta)$,*

$$|\Phi(x) - \Phi(x^*)| \le \varphi(x)\|\nabla\Phi(x)\|, \tag{64}$$

where Φ is defined by (57). *Then, x^* satisfies CAKKT.*

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

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k), \quad \mu^k = (\bar{\mu}^k + \rho_k g(x^k))_+. \tag{65}$$

By (60), (61), and (65) 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 (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 λ^k by $\underline{\lambda}^k$, μ^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 $\{\underline{\lambda}^k, \underline{\mu}^k\}$.

Consider now the case in which $\{\rho_k\}$ is bounded above. By (65), we have that $\{\lambda^k\}$ and $\{\mu^k\}$ are bounded. Moreover, by the choice of ρ_{k+1}, defining V^k as in (62), 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 (65), $\mu_i^k = 0$ 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 (19) in the case that $\{\rho_k\}$ is bounded.

Let us consider now the case in which $\lim_{k\to\infty} \rho_k = \infty$. By (60), (61), the continuity of $f, \nabla \underline{h}, \nabla \underline{g}$, and the boundedness of $\{\underline{\lambda}^k\}, \{\underline{\mu}^k\}$, we have that there exists $M > 0$ such that

$$\rho_k \|\nabla \Phi_{\bar{\lambda}^k, \bar{\mu}^k, \rho_k}(x^k)\| \le M \tag{66}$$

for all k.

Now, for all $x \in \mathbb{R}^n$, we have

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

Therefore, by (66),

$$\|\rho_k(\sum_{i=1}^m h_i(x^k)\nabla h_i(x^k) + \sum_{i=1}^p g_i(x^k)_+ \nabla g_i(x^k))$$

$$+\sum_{i=1}^{m}\bar{\lambda}_i^k\nabla h_i(x^k)+\sum_{i=1}^{p}[(\rho_k g_i(x^k)+\bar{\mu}_i^k)_+-\rho_k g_i(x^k)_+]\nabla g_i(x^k)\|\le 2M$$

for all k.

Thus,

$$\|\rho_k(\sum_{i=1}^{m}h_i(x^k)\nabla h_i(x^k)+\sum_{i=1}^{p}g_i(x^k)_+\nabla g_i(x^k))\|$$

$$\le\|\sum_{i=1}^{m}\bar{\lambda}_i^k\nabla h_i(x^k)+\sum_{i=1}^{p}[(\rho_k g_i(x^k)+\bar{\mu}_i^k)_+-\rho_k g_i(x^k)_+]\nabla g_i(x^k)\|+2M$$

for all k.

Note that

$$|(\rho_k g_i(x^k)+\bar{\mu}_i^k)_+-\rho_k g_i(x^k)_+|\le\bar{\mu}_i^k.$$

Then, by (58), the boundedness of $\{x^k\}$, $\{\bar{\lambda}^k\}$, and $\{\bar{\mu}^k\}$, and the continuity of ∇h and ∇g, there exists $M_1>0$ such that

$$\|\rho_k\nabla\Phi(x^k)\|=2\|\rho_k(\sum_{i=1}^{m}h_i(x^k)\nabla h_i(x^k)+\sum_{i=1}^{p}g_i(x^k)_+\nabla g_i(x^k))\|\le M_1$$

for all k.

Since x^* is feasible, we have that $\Phi(x^*)=0$. Therefore, by (64),

$$|\rho_k\Phi(x^k)|\le\varphi(x^k)\|\rho_k\nabla\Phi(x^k)\|\le M_1\varphi(x^k)$$

for all k. Since $\lim_{k\to\infty}\varphi(x^k)=0$ this implies that

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

Thus, for all $i=1,\dots,m$, $j=1,\dots,p$,

$$\lim_{k\to\infty}\rho_k h_i(x^k)^2=0,\quad\lim_{k\to\infty}\rho_k g_i(x^k)_+^2=0.$$

But, since x^* is feasible, we have that $\lim_{k\to\infty}\|h(x^k)\|=\lim_{k\to\infty}\|g(x^k)_+\|=0$. Therefore, by the boundedness of $\{\bar{\lambda}^k\}$ and $\{\bar{\mu}^k\}$,

$$\lim_{k\to\infty}|\lambda_i^k h_i(x^k)|=\lim_{k\to\infty}|[\bar{\lambda}_i^k+\rho_k h_i(x^k)]h_i(x^k)|=0$$

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

$$\lim_{k\to\infty}|\mu_i^k g_i(x^k)_+|=\lim_{k\to\infty}|(\bar{\mu}_i^k+\rho_k g_i(x^k))_+g_i(x^k)_+|=0$$

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

If $g_i(x^*)<0$ one has that $g_i(x^k)<g_i(x^*)/2<0$ for k large enough. Since $\rho_k\to\infty$ we have that $\bar{\mu}_i^k+\rho_k g_i(x^k)<0$ and, so, $\mu_i^k=0$ for k large enough.

Assume, finally, that $g_i(x^*)=0$ and there exists an infinite set of indices K such that $g_i(x^k)<0$. In this case,

$$\mu_i^k=(\bar{\mu}_i^k+g_i(x^k))_+\le\bar{\mu}_i^k$$

for all $k\in K$. Therefore, by the boundedness of $\{\bar{\mu}_i^k\}$ we also have that $\lim_{k\in K}\mu_i^k g_i(x^k)=0$. This completes the proof. ∎

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

Consider the problem

$$\text{Minimize } x \ \text{ subject to } \ h(x) = 0, \tag{67}$$

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 $\rho_k \to \infty$, 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

$$h'(x) = x^2 \left[4x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)\right] \tag{68}$$

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

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

$$R(x) = 4x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) + x^3 \tag{69}$$

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

$$z_k = \frac{1}{2k\pi}, \ y_k = \frac{1}{(2k + 1/2)\pi}. \tag{70}$$

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

$$R(x_k) = 0. \tag{71}$$

By (70) we have that $y_k \to 0, z_k \to 0$. Therefore, since $x_k \in (y_k, z_k)$, $\lim_{k\to\infty} x_k = 0$. Moreover, since $x_k \in (y_k, z_k)$ we have that

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

and, by (70),

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

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

$$\sin\left(\frac{1}{x_k}\right) > 0. \tag{72}$$

Since $x_k \to 0$, we have that $\lim_{k\to\infty} x_k \sin(1/x_k) = 0$. Then, by (69) and (71),

$$\lim_{k\to\infty} \cos\left(\frac{1}{x_k}\right) = 0. \tag{73}$$

By (72) and (73), we have that

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

Therefore, for k large enough,

$$\sin\left(\frac{1}{x_k}\right) > \frac{1}{2}. \tag{74}$$

By (68) and (71),

$$h'(x_k) = -x_k^5 \tag{75}$$

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

Now, for k large enough, define

$$\rho_k = \frac{-1}{h(x_k)h'(x_k)}. \tag{76}$$

By (75), 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 (74), ρ_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

$$\rho_{k+1} \geq \gamma \rho_k \tag{77}$$

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

Now, by (60), 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 (76),

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

This means that the sequence $\{x_k\}$ is defined by the application of Algorithm 5.1 to the problem (67) with $\bar{\lambda}^k = 0$ for all k and the penalty parameters given by (76).

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

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

Therefore, by (75),

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

and, so,

$$\lim_{k\to\infty} \lambda_k x_k^5 = 1. \tag{78}$$

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 (74) and (78), since $x_k \to 0$, we deduce that

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

Therefore, (19) does not hold.

6. Conclusions.

In the first section of this paper we motivated the introduction of a new strong AKKT condition using a simple example that shows that points that approximately 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. More precisely, we endorse the position that the stopping criterion based on CAKKT should be tested whenever one is using an algorithm that guaranteedly produces CAKKT sequences. As a consequence, the *theoretical* question about the CAKKT behavior of optimization algorithms has a clear *practical* relevance.

Little 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 tend to infinity. Many practical problems may have this characteristic and, thus, algorithms should be equipped with adequate procedures to deal with such 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 with 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 no 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

$$\nabla f(x) + \nabla h(x)\lambda = 0, \quad h(x) = 0. \tag{79}$$

In fact (79) is a square nonlinear system and variations of Newton's method are often successful for solving it. Now, (79) is obviously equivalent to

$$\nabla f(x) + \nabla h(x)\lambda = 0, \; h(x) = 0, \; \lambda_i h_i(x) = 0, i = 1, \ldots, m, \tag{80}$$

but equivalence disappears when, roughly speaking, we admit that multipliers may tend to infinity. In this case, the approximate fulfillment of (79) corresponds to the AKKT condition, but the approximate fulfillment of (80) 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, in some cases, the rectangular form (80).

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 [25]. 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 (80) 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, λ_k) such that $x_k \to 0, \lambda_k \to \infty, \lambda_k h(x_k) \to 0$ 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) is not.

Motivated by the example above, we studied the practical behavior of Newton's method in KKT systems corresponding to other problems of the form

$$\text{Minimize } f(x) \text{ subject to } h_i(x)^2 = 0, i = 1, \ldots, m.$$

Since the gradients of all the constraints vanish at feasible points, KKT conditions hold at a solution only in the improbable situation that the gradient of f also vanishes. Surprisingly, in many cases, Newton's method behaved as in the first one-dimensional example: (a) convergence to the primal solution ocurred; (b) the Lagrangian residual corresponding to (18) did not tend to zero; (c) the product of each multiplier approximation with the corresponding feasibility residual converged to zero. Several open questions are suscitated by these observations. One one hand, it should be interesting to discover sufficient theoretical conditions under which Newton's method necessarily exhibits such behavior. On the other hand, variations of the Newton sequence that possibly guarantee CAKKT verification should be studied and their practical efficiency should be compared with existing versions of the Newton-SQP method.

Acknowledgments. We are indebted to two anonymous referees whose careful comments and insights helped us a lot to improve the present paper and to Prof. Jerome Bolte, who provided us with useful remarks about the properties of Lojasiewicz-type inequalities. We also acknowledge Prof. Robert Gower, who corrected many language errors.

REFERENCES

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, *On augmented Lagrangian methods with general lower-level constraints*, SIAM J. Optim., 18 (2007), pp. 1286–1309.

[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*, J. Global Optim., 43 (2009), pp. 1–10.

[4] R. Andreani, J. M. Martínez, and M. L. Schuverdt, *On the relation between the constant positive linear dependence condition and quasinormality constraint qualification*, J. Optim. Theory Appl., 125 (2005), pp. 473–485.

[5] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program., 116 (2009), pp. 5–16.

[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*, Math. Oper. Res., 35 (2010), pp. 438–457.

[7] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., 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*, Math. Program., 125 (2010), pp. 139–162.

[9] J. Bolte, A. Daniilidis, and A. S. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2007), pp. 1205–1223.

[10] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, *Characterizations of Lojasiewicz inequalities and applications: Subgradient flows, talweg, convexity*, Trans. Amer. Math. Soc., 362 (2010), pp. 3319–3363.

[11] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust-Region Methods*, MPS/SIAM Ser. Optim. 1, 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*, Comput. Optim. Appl., 46 (2010), pp. 333–346.

[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 J. Sci. Comput., 31 (2009), pp. 1628–1652.

[15] C. C. Gonzaga, E. W. Karas, and M. Vanti, *A globally convergent filter method for nonlinear programming*, SIAM J. Optim., 14 (2004), pp. 646–669.

[16] M. R. Hestenes, *Optimization Theory: The Finite Dimensional Case*, Wiley, New York, 1975.

[17] A. D. Ioffe, *An invitation to tame optimization*, SIAM J. Optim., 19 (2009), pp. 1894–1917.

[18] S. G. Krantz and H. G. Parks, *A Primer on Real Analytic Functions*, 2nd ed., Birkhäuser Boston, Cambridge, Mass., 2002.

[19] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Ann. Inst. Fourier (Grenoble), 48 (1998), pp. 769–783.

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

[21] O. L. Mangasarian and S. Fromovitz, *The Fritz-John necessary optimality conditions in presence of equality and inequality constraints*, J. Math. Anal. Appl., 17 (1967), pp. 37–47.

[22] J. M. Martínez, *Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming*, J. Optim. Theory Appl., 111 (2001), pp. 39–58.

[23] J. M. Martínez and E. A. Pilotta, *Inexact restoration algorithms for constrained optimization*, J. Optim. Theory Appl., 104 (2000), pp. 135–163.

[24] J. M. Martínez and B. F. Svaiter, *A practical optimality condition without constraint qualifications for nonlinear programming*, J. Optim. Theory Appl., 118 (2003), pp. 117–133.

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

[26] L. Qi and Z. Wei, *On the constant positive linear dependence condition and its application to SQP methods*, SIAM J. Optim., 10 (2000), pp. 963–981.

[27] 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.