A sequential optimality condition related to the quasinormality constraint qualification and its algorithmic consequences

R. Andreani † N. S. Fazzio ‡ M. L. Schuverdt § L. D. Secchin ¶

September 5, 2017

Abstract

In the present paper, we prove that the augmented Lagrangian method converges to KKT points under the quasinormality constraint qualification, which is associated with the external penalty theory. For this purpose, a new sequential optimality condition for smooth constrained optimization, called PAKKT, is defined. The new condition takes into account the sign of the dual sequence, constituting an adequate sequential counterpart to the (enhanced) Fritz-John necessary optimality conditions proposed by Hestenes, and later extensively treated by Bertsekas. We also provided the appropriate strict constraint qualification associated with the PAKKT sequential optimality condition and we prove that it is strictly weaker than both quasinormality and cone continuity property. This generalizes all previous theoretical convergence results for the augmented Lagrangian method in the literature.

Key words: Augmented Lagrangian methods, global convergence, constraint qualifications, quasinormality.

1 Introduction

We will consider the general constrained nonlinear problem

\[ \min_{x} f(x) \text{ subject to } x \in X \]  \hspace{1cm} (P)

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) and \( X \) is the feasible set composed of equality and inequality constraints of the form

\[ X = \{ x \mid h(x) = 0, g(x) \leq 0 \}, \]

with \( h : \mathbb{R}^n \rightarrow \mathbb{R}^m \) and \( g : \mathbb{R}^n \rightarrow \mathbb{R}^p \). We assume that the functions \( f, h \) and \( g \) are continuously differentiable in \( \mathbb{R}^n \). Given \( x^* \in X \) we denote by \( I_g(x^*) = \{ j \in \{1, \ldots, p \} \mid g_j(x^*) = 0 \} \) the set of active inequality constraints at \( x^* \).

Several of the more traditional nonlinear programming methods are iterative: given an iterate \( x^k \), they try to find a better approximation \( x^{k+1} \) of the solution. In this paper we consider the augmented Lagrangian method, a popular technique in constrained optimization. The classical augmented Lagrangian method uses an iterative sequence of subproblems considerably easier to solve. In each subproblem, fixed a penalty parameter \( \rho > 0 \) and Lagrange multiplier estimates \( \lambda \in \mathbb{R}^m \) and \( \mu \in \mathbb{R}^p, \mu \geq 0 \), an augmented Lagrangian function is approximately minimized. Once the approximate solution is found, the penalty

---

*This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (Grant 2013/05475-7) and CNPq (Grant 303013/2013-3).
†Institute of Mathematics, University of Campinas, Rua Sergio Buarque de Holanda, 651, 13083-859, Campinas, São Paulo, Brazil. Email: andreani@ime.unicamp.br
‡CONICET, Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina. Email: nadiafazzio@gmail.com
§CONICET, Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina. Email: schuverd@mate.unlp.edu.ar.
¶Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, Espírito Santo, Brazil. Email: leonardo.secchin@ufes.br.
parameter $\rho$ and the multipliers estimates are updated, and a new iteration starts. Specifically, we consider the Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian function that, for the problem (P), takes the form

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

where $x \in \mathbb{R}^n$, $\rho > 0$, $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$, $\mu \geq 0$. This is the most widely used augmented Lagrangian function in practical implementations (see [14] for a numerical comparison between several of them). However, others functions were also employed, see for example [11, 17] and references therein. The choice of (1) is justified by its intrinsic relation to the external penalty theory, where the quasinormality, a very general constraint qualification proposed by Hestenes [19], plays an important role. This is the central issue in this work.

In the last years, special attention has been devoted to so-called sequential optimality conditions for nonlinear constrained optimization (see for example [3, 7, 8, 9, 15, 21]). They are related to the stopping criteria of algorithms, and aim to unify their theoretical convergence results. In particular, they have been used to study the convergence of the augmented Lagrangian method (see [15] and references there in). An important feature of sequential optimality conditions is that they are necessary for optimality: a local minimizer of (P) verifies such a condition independently of the fulfillment of any constraint qualification. One of the most popular sequential optimality condition is the Approximate Karush-Kuhn-Tucker (AKKT) condition, defined in [3]. We say that $x^* \in X$ satisfies the AKKT condition if there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$ and $\{\mu^k\} \subset \mathbb{R}^p$, $\mu^k \geq 0$, such that

$$\lim_{k \to \infty} x^k = x^*,$$
$$\lim_{k \to \infty} \|\nabla L(x^k, \lambda^k, \mu^k)\| = 0,$$
$$\lim_{k \to \infty} \min \{ -g(x^k), \mu^k \} = 0.$$

Such kind of points $x^*$ will be called AKKT points and $\{x^k\}$ an AKKT sequence.

Of course, when it is proved that an AKKT point is in fact a KKT point under a certain constraint qualification (CQ), all the algorithms that reach AKKT points, such as the augmented Lagrangian method [15], have their theoretical convergence automatically established with the same CQ (this is exactly what we mean when we say that a sequential optimality condition unifies convergence results). During the last years, it has been proved that AKKT points are stationary points under different constraint qualifications such as the constant positive linear dependence (CPLD) [6, 23], the relaxed constant positive linear dependence (RCPLD) [4] and the constant positive generator (CPG) [5]. Finally, it was shown that the cone continuity property (CCP) (also called AKKT-regular CQ [8]) is the weakest constraint qualification with this property [9].

Despite the clear similarities between the PHR augmented Lagrangian and the pure external penalty methods, an interesting topic that still unsolved is the convergence of the augmented Lagrangian method under the quasinormality CQ. The authors of [9] show that CCP has no relation with quasinormality or, in particular, that there are examples where quasinormality and the AKKT condition hold but KKT does not. In this sense, the AKKT condition can not give an answer to the proposed question. Indeed, it is surprising to note that the PHR augmented Lagrangian method, which uses the quadratic penalty-like function (1), naturally generates AKKT points while it is not trivial to understand how it handles the sign of the multipliers, as performed by the external penalty method.

In the present paper we prove that the augmented Lagrangian method converges under quasinormality showing that the limit points generated by the method are in fact stronger than AKKT points. We call these points Positive Approximate Karush-Kuhn-Tucker (PAKKT). The key is to take into account the sign of Lagrange multipliers as in the Fritz-John necessary conditions described in [12] (see also [19]). Specifically, we rely on the following result:

**Theorem 1.1** ([12, Proposition 3.3.5]). Let $x^*$ be a local minimizer of the problem (P). Then there are $\sigma \in \mathbb{R}^+_+$, $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$, $\mu \geq 0$, such that...
\[ \sigma \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0; \]

2. \( (\sigma, \lambda, \mu) \neq 0; \)

3. In every neighborhood \( B(x^*) \) of \( x^* \) there is an \( x \in B(x^*) \) such that \( \lambda_i h_i(x) > 0 \) for all \( i \) with \( \lambda_i \neq 0 \) and \( \mu_j g_j(x) > 0 \) for all \( j \) with \( \mu_j > 0 \).

From now on we will refer to Theorem 1.1 by enhanced Fritz-John (necessary) conditions. We call a point \( x^* \) that fulfills all the three items of Theorem 1.1 by an Enhanced Fritz-John (EFJ) point. We note that an EFJ point is a KKT point whenever \( \sigma > 0 \). Conditions (1.) and (2.) symbolize the classical Fritz-John result about stationary points. Condition (3.) stands for the existence of sequences which connect the sign of the multiplier with the sign of the associated constraint in a neighborhood of the stationary point. Enhanced Fritz-John conditions were used previously to generalize some classical results [13, 27].

In addition to show that the augmented Lagrangian method reaches PAKKT points, we also provide the weakest constraint qualification that ensures that a PAKKT point is KKT, which we call \( \text{PAKKT-regular} \). We prove that the new CQ is weaker than both quasinormality and CCP (AKKT-regular) constraint qualifications. This result generalizes all previous theorems about the convergence of the PHR augmented Lagrangian method. Furthermore, to the best of our knowledge, this is the first time it has been proved that a practical algorithm converges under the quasinormality CQ. We also make a correspondence between the theory presented here and previous convergence results for mathematical programs with complementarity constraints [10, 20].

This paper is organized as follows. In section 2 we describe the new sequential optimality condition PAKKT and its associated strict constraint qualification PAKKT-regular. In section 3 we establish the relationship between PAKKT and others sequential optimality conditions in the literature. Relations of PAKKT-regular with other known constraint qualifications are also included in section 3. In section 4 we present the global convergence of the augmented Lagrangian method using the PAKKT-regular CQ. Conclusions and lines for future research are given in section 5.

Notation:
- \( \mathbb{R}_+ = \{ t \in \mathbb{R} | t \geq 0 \} \), \( \| \cdot \| \) denotes an arbitrary vector norm and \( \| \cdot \|_\infty \) the supremum norm.
- \( v_i \) is the \( i \)-th component of the vector \( v \).
- For all \( y \in \mathbb{R}^n \), \( y_+ = (\max\{0, y_1\}, \ldots, \max\{0, y_n\}) \).
- If \( K = \{k_0, k_1, k_2, \ldots \} \subset \mathbb{N} \) and \( k_{j+1} > k_j \), we denote \( \lim_{k \in K} y^k = \lim_{j \to \infty} y^{k_j} \). In particular, \( \lim_{k \in \mathbb{N}} y^k \).
- If \( \{ \gamma_k \} \subset \mathbb{R} \), \( \gamma_k > 0 \), and \( \gamma_k \to 0 \), we write \( \gamma \downarrow 0 \).
- We define the “sign function” \( \text{sgn} a \) putting \( \text{sgn} a = 1 \) if \( a > 0 \) and \( \text{sgn} a = -1 \) if \( a < 0 \). We have \( \text{sgn} (a \cdot b) = \text{sgn} a \cdot \text{sgn} b \).

2 The positive approximate Karush-Kuhn-Tucker condition

In this section we define the positive approximate Karush-Kuhn-Tucker condition and we show that it is a genuine necessary optimality condition.

**Definition 2.1.** We say that \( x^* \in X \) is a Positive Approximate KKT (PAKKT) point if there are
sequences \( \{x^k\} \subset \mathbb{R}^n \), \( \{\lambda^k\} \subset \mathbb{R}^m \) and \( \{\mu^k\} \subset \mathbb{R}^p \) such that
\[
\lim_{k \to \infty} x^k = x^*, \quad (2)
\]
\[
\lim_{k \to \infty} \|\nabla_x L(x^k, \lambda^k, \mu^k)\| = 0, \quad (3)
\]
\[
\lim_{k \to \infty} \|\min \{-g(x^k), \mu^k\}\| = 0, \quad (4)
\]
\[
\lambda^k h_i(x^k) > 0 \text{ if } \lim_{k \to \infty} \frac{\lambda^k_i}{\delta_k} > 0, \quad (5)
\]
\[
\mu^k_i g_j(x^k) > 0 \text{ if } \lim_{k \to \infty} \frac{\mu^k_i}{\delta_k} > 0, \quad (6)
\]
where \( \delta_k = \|(1, \lambda^k, \mu^k)\|_\infty \). In this case, \( \{x^k\} \) is called a PAKKT sequence.

The expressions (2)–(4) are related to the KKT conditions, and they are used in the Approximate KKT (AKKT) optimality condition presented in the introduction. The expressions (5) and (6) aim to control the sign of Lagrange multipliers, justifying the name of our new condition. They are related to the enhanced Fritz-John necessary optimality conditions described in the introduction (Theorem 1.1). We will see that (5) and (6) give an adequate counterpart for the third item of Theorem 1.1 in the sequential case. As always \( |\lambda^k_i|/\delta_k, \mu^k_i/\delta_k \in [0, 1] \) we can suppose, taking a subsequence if necessary, that these limits exist. It is important to note that item (3.) of Theorem 1.1 is sufficient for complementary slackness, but the sequential counterpart (5) and (6) is not. The next example shows that the complementary slackness at the limit \( x^* \) may fail without condition (4) if \( \{\delta_k\} \) is unbounded.

**Example 2.1.** Let us consider the problem
\[
\min_x -x_1 + x_2 \quad \text{subject to} \quad x_2^2 = 0, \quad x_1 - 1 \leq 0,
\]
for which
\[
\nabla_x L(x, \lambda, \mu) = \begin{bmatrix} -1 & 0 \\ 1 & 2x_2 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \mu \begin{bmatrix} 0 \\ 1 \end{bmatrix}.
\]
The point \( x^* = (0, 0) \) satisfies (2), (3), (5) and (6) with, for example, the sequences defined by \( x^k = (0, -1/k), \lambda^k = k/2 \) and \( \mu^k = 1 \) for all \( k \geq 1 \). In fact, we have \( x^k \to (0, 0), \nabla_x L(x^k, \lambda^k, \mu^k) = 0, \lambda^k(x^k)^2 = 1/(2k) > 0 \) with \( \lim |\lambda^k|/\delta_k = 1 \), and \( \lim \mu^k/\delta_k = 0 \). But, taking a subsequence if necessary, any sequence satisfying (2), (3), (5) and (6) is such that \( |\lambda^k| \to \infty \) and \( \mu^k \to 1 \). Thus \( \lim \mu^k/\delta_k = 0 \), but \( \min \{-x_1 - 1, \mu^k\} \to 1 \).

**Theorem 2.1.** PAKKT is a necessary optimality condition.

**Proof.** Let \( x^* \) be a local minimizer of (P). Then \( x^* \) is the unique global minimizer of the problem
\[
\min_x f(x) + 1/2\|x - x^*\|^2 \quad \text{subject to} \quad h(x) = 0, \quad g(x) \leq 0, \quad \|x - x^*\| \leq \alpha
\]
for a certain \( \alpha > 0 \). Let \( x^k \) be a global minimizer of the penalized problem
\[
\min_x f(x) + 1/2\|x - x^*\|^2 + \frac{\rho_k}{2} \left[ \|h(x)\|^2 + \|g(x)\|^2 \right] \quad \text{subject to} \quad \|x - x^*\| \leq \alpha,
\]
where \( \rho_k > 0 \), which exists by the continuity of the objective function and compactness of the feasible set. We suppose that \( \rho_k \to \infty \). From the external penalty theory, \( x^k \to x^* \) and thus (2) is satisfied. We have \( \|x^k - x^*\| < \alpha \) for all \( k \) sufficiently large (let us say for all \( k \in K \)), and from optimality conditions of the penalized problem we obtain
\[
\lim_{k \to \infty} \nabla_x L(x^k, \lambda^k, \mu^k) = \lim_{k \to \infty} \left[ \nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k \right] = -\lim_{k \to \infty} (x^k - x^*) = 0,
\]
where, for each \( k \in K \),
\[
\lambda^k = \rho_k h(x^k) \quad \text{and} \quad \mu^k = \rho_k g(x^k) \geq 0.
\]
Therefore (3) and (4) are satisfied. If \( \lambda^k_i > 0, k \in K \), then \( g_i(x^k) > 0 \) and hence \( \mu^k_i g_i(x^k) = \rho_k |g_i(x^k)|^2 > 0 \). Analogously, if \( \lambda^k_i \neq 0, k \in K \), then \( h_i(x^k) \neq 0 \) and hence \( \lambda^k_i h_i(x^k) = \rho_k |h_i(x^k)|^2 > 0 \). Thus (5) and (6) are fulfilled, independently of the limits of the dual sequences. \( \square \)
We say that SCQ is a strict constraint qualification for the sequential optimality condition A if

\[ A + \text{SCQ implies KKT} \]

(see [15]). Since all sequential optimality conditions are satisfied in any local minimizer independently of fulfillment of CQs, a SCQ is in fact a constraint qualification. The reciprocal is not true. For instance, Abadie’s CQ [1] or quasinormality [19] are CQs that are not SCQs for the AKKT sequential optimality condition. On the other hand, the strict constraint qualification SCQ provides a measure of the quality of the sequential optimality condition A. Specifically, A is better as SCQ is less stringent (weaker).

In [9], the authors presented the weakest strict constraint qualification associated with AKKT, called Cone Continuity Property (CCP). In a recent report [8], CCP was renamed to “AKKT-regular” and the weakest SCQs related to SAKKT [18], CAKKT [7] and AGP [21] conditions were established.

In this section, we provide the weakest SCQ for PAKKT condition, which we call PAKKT-regular. For this purpose, we define for each \( x^* \in X \) and \( x \in \mathbb{R}^n, \alpha, \beta \geq 0 \), the set

\[ K_+ (x, \alpha, \beta) = \left\{ \sum_{i=1}^{m} \lambda_i \vartriangledown h_i(x) + \sum_{j \in I_j (x^*)} \mu_j \vartriangledown g_j(x) \bigg| \begin{array}{l} \lambda_i h_i(x) \geq \alpha \text{ if } |\lambda_i| \geq |(1, \lambda, \mu)|_{\infty} \\ \mu_j g_j(x) \geq \alpha \text{ if } |\mu_j| \geq |(1, \lambda, \mu)|_{\infty} \end{array} \right\}. \]

Note that the KKT conditions for (P) can be written as \(-\nabla f(x^*) \in K_+ (x^*, 0, 0)\).

Given a multifunction \( \Gamma : \mathbb{R}^q \Rightarrow \mathbb{R}^q \), the sequential Painlevé–Kuratowski outer/upper limit of \( \Gamma(z) \) as \( z \to z^* \) is denoted by

\[ \limsup_{z \to z^*} \Gamma(z) = \{ y^* \in \mathbb{R}^q \mid \exists (z^k, y^k) \to (z^*, y^*) \text{ with } y^k \in \Gamma(z^k) \} \]

[24]. We say that \( \Gamma \) is outer semicontinuous at \( z^* \) if \( \limsup_{z \to z^*} \Gamma(z) \subset \Gamma(z^*) \). We define the PAKKT-regular condition imposing an outer semicontinuity-like on the multifunction \((x, \alpha, \beta) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+ \Rightarrow K_+(x, \alpha, \beta)\). Analogously to (7), we consider the following set:

\[ \limsup_{x \to x^*, \alpha \downarrow 0, \beta \downarrow 0} K_+(x, \alpha, \beta) = \{ y^* \in \mathbb{R}^n \mid \exists (z^k, y^k) \to (x^*, y^*), \alpha_k \downarrow 0, \beta_k \downarrow 0 \text{ with } y^k \in K_+(x^k, \alpha_k, \beta_k) \}. \]

**Definition 2.2.** We say that \( x^* \in X \) satisfies the PAKKT-regular condition if

\[ \limsup_{x \to x^*, \alpha \downarrow 0, \beta \downarrow 0} \Gamma^*(x, \alpha, \beta) \subset K_+(x^*, 0, 0). \]

Next we prove the main result of this section, which guarantees that PAKKT-regular is the weakest SCQ for the sequential optimality condition.

**Theorem 2.2.** If \( x^* \) is a PAKKT point that fulfills the PAKKT-regular condition then \( x^* \) is a KKT point. Reciprocally, if for every continuously differentiable function \( f \) the PAKKT point \( x^* \) is also KKT, then \( x^* \) satisfies the PAKKT-regular condition.

**Proof.** If \( x^* \) is a PAKKT point, there are sequences \( \{x^k\} \subset \mathbb{R}^n, \{\lambda^k\} \subset \mathbb{R}^m \) and \( \{\mu^k\} \subset \mathbb{R}_+^p, k \geq 1 \), such that \( x^k \to x^* \), (4)–(6) hold, and \( \nabla f(x^k) + \omega^k \to 0 \) where

\[ \omega^k = \sum_{i=1}^{m} \lambda^k_i \vartriangledown h_i(x^k) + \sum_{j=1}^{p} \mu^k_j \vartriangledown g_j(x^k). \]

By (4) we can suppose without loss of generality that \( \mu^k_j = 0 \) whenever \( j \notin I_g(x^*) \). As in the PAKKT definition, we consider \( \delta_k = |(1, \lambda^k, \mu^k)|_{\infty} \). Let us define the sets \( I_+ = \{ i \in \{1, \ldots, m\} \mid \lim_k |\lambda_i|/\delta_k > 0 \} \) and \( J_+ = \{ j \in I_g(x^*) \mid \lim_k \mu_j/\delta_k > 0 \} \), and for each \( k \) we take

\[ \alpha_k = \min \left\{ \frac{1}{k}, \min_{i \in I_+} \{ \lambda^k_i h_i(x^k) \}, \min_{j \in J_+} \{ \mu^k_j g_j(x^k) \} \right\} \]
and
\[ \beta_k = \max \left\{ \frac{1}{k}, \max_{j \in J_k} \frac{|\lambda^i_k|}{\delta_k}, \max_{j \in J_k} \frac{\mu_j^k}{\delta_k} \right\} + 1, \]

We note that \( \alpha_k \downarrow 0, \beta_k \downarrow 0 \) and \( \omega^k \in K_+(x^k, \alpha_k, \beta_k) \) for all \( k \) large enough. As \( x^* \) fulfills the PAKKT-regular condition, we have
\[ -\nabla f(x^*) = \lim_{k} \omega^k \in \limsup_{k} K_+(x^k, \alpha_k, \beta_k) \subset \limsup_{x \to x^*} K_+(x, \alpha, \beta) \subset K_+(x^*, 0, 0), \]
that is, \( x^* \) is a KKT point. This proves the first statement.

Now let us show the reciprocal. Let \( w^* \in \limsup_{x \to x^*} x^*, \alpha \downarrow 0, \beta \downarrow 0 \) be a KKT point. Then there are sequences \( \{x^k\} \subset \mathbb{R}^n, \{\omega^k\} \subset \mathbb{R}^n, \{\alpha_k\} \subset \mathbb{R} \) and \( \{\beta_k\} \subset \mathbb{R} \) such that \( x^k \to x^*, \omega^k \to \omega^*, \alpha_k \downarrow 0, \beta_k \downarrow 0 \) and \( \omega^k \in K_+(x^k, \alpha_k, \beta_k) \) for all \( k \). Furthermore, for each \( k \) there are \( \lambda^i_k \in \mathbb{R}^m \) and \( \mu^k_j \in \mathbb{R}_+ | l_j(x^k) | \) such that
\[ \omega^k = \sum_{i=1}^m \lambda^i_k \nabla h_i(x^k) + \sum_{j \in I_k(x^*)} \mu^k_j \nabla g_j(x^k). \tag{8} \]

We define \( f(x) = -(\omega^*)^t x \). If \( \lim_{k} |\lambda^i_k|/\delta_k > 0 \) then \( \lambda^i_k \geq \beta_k \delta_k \) for all \( k \) sufficiently large (the same happens with \( \mu \)). In other words, the control over the sign of the multipliers performed by (5) and (6) is encapsulated in the expression \( \omega^k \in K_+(x^k, \alpha_k, \beta_k) \). Therefore, as \( -\nabla f(x^k) + \omega^k = -\omega^* + \omega^k \to 0 \), we conclude that \( x^* \) is a PAKKT point. By hypotheses \( x^* \) is a KKT point, and hence \( \lim_k \omega^k = -\nabla f(x^*) \in K_+(x^*, 0, 0) \). This concludes the proof. \( \square \)

As a consequence of Theorems 2.1 and 2.2, it follows that any minimizer of (P) satisfying the PAKKT-regular condition is a KKT point. Equivalently, we obtain the next result.

**Corollary 2.1.** PAKKT-regular is a constraint qualification.

As expected, every KKT point is a PAKKT point (Lemma 2.1 below). However, an observation must be taken into account: consider, for example, the two constraints \( g_1(x) = x \leq 0, g_2(x) = -x \leq 0 \), and the constant objective function \( f(x) = 1 \). The origin is a KKT point with multipliers \( \mu_1 = \mu_2 = 1 \). In this case, any point \( x \neq 0 \) near the origin satisfies \( \mu_1 g_1(x) < 0 \) or \( \mu_2 g_2(x) < 0 \). In order words, this situation is not suitable for the PAKKT condition. But note that the Lagrange multipliers are not unique in this example. Fortunately, a KKT point always admits Lagrange multipliers with adequate signs for the PAKKT condition.

**Lemma 2.1.** Every KKT point is a PAKKT point.

*Proof.* Let \( x^* \) be a KKT point. Then \( x^* \) is an EFJ point with \( \sigma = 1 \) [12, Exercise 3.3.10(b)]. Thus, there are associated Lagrange multipliers \( \lambda^i_*, \mu^j_* \) with \( \mu^j_* \geq 0 \) and
\[ \nabla f(x^*) + \sum_{j \in I} \lambda^i_* \nabla h_i(x^*) + \sum_{i \in J} \mu^j_* \nabla g_j(x^*) = 0, \tag{9} \]
where
\[ I = \{i \in \{1, \ldots, m\} \mid \lambda^i_* \neq 0\}, \quad J = \{j \in I_g(x^*) \mid \mu^j_*> 0\}, \]
and such that in all neighborhood \( B(x^*) \) of \( x^* \), there is a \( x \in B(x^*) \) such that \( \lambda^i_* h_i(x) > 0 \) whenever \( i \in I \), and \( \mu^j_* g_j(x) > 0 \) whenever \( j \in J \). Therefore, taking \( (\lambda^i_k, \mu_j^k) = (\lambda^i_*, \mu^j_*) \) and all others \( (\lambda^i_k, \mu^j_k) \) equal to zero, there is a sequence \( \{x^k\} \) converging to \( x^* \) that fulfills the PAKKT definition. \( \square \)

### 3 Relations

#### 3.1 Relations with others sequential optimality conditions

In this subsection, we establish the relations between the PAKKT condition and others sequential optimality conditions in the literature. In addition to the AKKT condition presented in the introduction, we consider the following ones:
• We say that $x^* \in \mathbb{X}$ is a Complementary Approximate KKT (CAKKT) [7] point if there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$ and $\{\mu^k\} \subset \mathbb{R}^p_+$ such that (2), (3) hold and, for all $i = 1, \ldots, m$ and $j = 1, \ldots, p$,

\[
\lim_k \lambda^k_i h_i(x^k) = 0 \quad \text{and} \quad \lim_k \mu^k_j g_j(x^k) = 0.
\]

In this case, $\{x^k\}$ is called a CAKKT sequence.

• For each $x \in \mathbb{R}^n$, let us consider the linear approximation of the feasible set $\mathcal{X}$ at $x$

\[
\Omega(x) = \left\{ z \in \mathbb{R}^n \left| \begin{array}{l}
g_j(x) + \nabla g_j(x)^T(z - x) \leq 0 \quad \text{if} \quad g_j(x) < 0 \\
\nabla g_j(x)^T(z - x) \leq 0 \quad \text{if} \quad g_j(x) \geq 0 \\
\n\nabla h(x)^T(z - x) = 0
\end{array} \right. \right\}.
\]

We define the approximate gradient projection by $d(x) = P_{\Omega(x)}(x - \nabla f(x)) - x$, where $P_{\Omega(\cdot)}$ denotes the orthogonal projection onto the closed and convex set $\Omega$. We say that $x^* \in \mathbb{X}$ is an Approximate Gradient Projection (AGP) [7] point if there is a sequence $\{x^k\} \subset \mathbb{R}^n$ converging to $x^*$ such that $d(x^k) \to 0$.

Both of the above conditions have been proved to be sequential optimality conditions [7, 21]. Also, it is known that CAKKT is strictly stronger than AGP [7], which in turn is strictly stronger than AKKT [3]. CAKKT sequences are generated by the augmented Lagrangian method of the next section with an additional hypothesis that the sum-of-squares infeasibility measure satisfies a generalized Lojasiewicz inequality (see [7] for details). On the other hand, AGP sequences are useful to analyze limit points of inexact restoration techniques [22].

Since AKKT condition is exactly PAKKT without expressions (5) and (6), the next result is trivial.

**Theorem 3.1.** Every PAKKT sequence is also an AKKT sequence. In particular, every PAKKT point is an AKKT point.

**Example 3.1** (a CAKKT point may not be PAKKT). Let us consider the problem

\[
\min_x \left( \frac{(x_1 - 1)^2}{2} + \frac{(x_2 + 1)^2}{2} \right) \quad \text{subject to} \quad x_1x_2 \leq 0,
\]

for which

\[
\nabla_x L(x, \mu) = \begin{bmatrix} x_1 - 1 \\ x_2 + 1 \end{bmatrix} + \mu \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}.
\]

The origin is a CAKKT point with the sequences defined by $x^k = (-1/k, 1/k)$ and $\mu^k = k$ for all $k \geq 1$. Furthermore, any CAKKT sequence $\{x^k\}$ converging to the origin with associated $\{\mu^k\}$ satisfies $\lim_k \mu^k x^k_1 = -1$ and $\lim_k \mu^k x^k_2 = 1$. Thus, for all $k$ sufficiently large we have $\text{sgn}(\mu^k) = 1$, $\text{sgn}(x^k_1) = -1$ and $\text{sgn}(x^k_2) = 1$, and hence $\text{sgn}(\mu^k x^k_1 x^k_2) = -1$. That is, the origin is not a PAKKT point.

**Example 3.2** (a PAKKT point may not be AGP). Let us consider the problem

\[
\min_x x_2 \quad \text{subject to} \quad x_1^2x_2 = 0, \quad -x_1 \leq 0.
\]

The point $(0, -1)$ is PAKKT with the sequences defined by $x^k = (1/k, -1)$, $\lambda^k = -k^2$ and $\mu^k = 2k$ for all $k \geq 1$. By straightforward calculations we obtain

\[
\Omega(x_1, x_2) = \left\{ (z_1, z_2) \in \mathbb{R}^2 \left| \begin{array}{l}
z_1 \geq \min \{0, x_1\} \\
(2x_1x_2)z_1 + (x_1^2)z_2 = 3x_1^2x_2
\end{array} \right. \right\}.
\]

Thus, given a sequence $\{x^k\}$ converging to $(0, -1)$, the set $\Omega(x^k)$ tends to a vertical semi-line within the $y$-axis if $x^k_1 \neq 0$ or it is the semi-plane $z_1 \geq 0$ otherwise. As $x^k - \nabla f(x^k) = (x^k_1, x^k_2 - 1) \to (0, -2)$, we always have $\|d(x^k)\|_{\infty} = \|P_{\Omega(x^k)}(x^k - \nabla f(x^k)) - x^k\|_{\infty} \to 1$. Thus $0, -1$ is not an AGP point.

In particular, as every CAKKT point is AGP (and thus AKKT), Example 3.1 also shows that exist AGP and AKKT points that are not PAKKT points. In the same way, Example 3.2 implies that exist PAKKT points that are not CAKKT. Figure 1 summarizes the relationship between all sequential optimality conditions discussed here.
3.2 Relations between PAKKT-regular and others known CQs

In the section 2, we demonstrated that PAKKT-regular is a constraint qualification (Corollary 2.1). Now we discuss the relationship between PAKKT-regular and other CQs in the literature, giving an updated landscape of various CQs.

We already mentioned that the weakest strict constraint qualification for the AKKT sequential optimality condition is the AKKT-regular CQ (also called cone continuity property – CCP) [9]. Defining the cone
\[ K(x, \alpha, \beta) = \left\{ \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j \in I_g(x^*)} \mu_j g_j(x) \mid \lambda \in \mathbb{R}^m, \mu_j \in \mathbb{R}^+, j \in I_g(x^*) \right\}, \]
we say that \( x^* \in X \) satisfies the AKKT-regular condition if the multifunction \( x \in \mathbb{R}^n \Rightarrow K(x) \) is outer semicontinuous at \( x^* \), that is, if
\[ \limsup_{x \to x^*, \alpha \downarrow 0, \beta \downarrow 0} K(x, \alpha, \beta) \subset K(x^*, 0, \infty). \]
Note that \( K(x, \alpha, \beta) \subset K(x, 0, \infty) \) for all \( x \in \mathbb{R}^n \) and \( \alpha, \beta > 0 \). Furthermore \( K(x, 0, \infty) = K(x, 0, 0) \) whenever \( x \) is feasible for (P). These observations are the key to prove the next result.

**Theorem 3.2.** AKKT-regular implies PAKKT-regular.

**Proof.** If \( x^* \) satisfies the AKKT-regular condition then
\[ \limsup_{x \to x^*, \alpha \downarrow 0, \beta \downarrow 0} K(x, \alpha, \beta) \subset \limsup_{x \to x^*} K(x, 0, \infty) \subset K(x^*, 0, \infty) = K(x^*, 0, 0), \]
completing the proof. \( \square \)

A natural constraint qualification associated with EFJ points is quasinormality (see for example [12]). Since PAKKT is an optimality condition that translates the sign control in the EFJ points to the sequential world, it is reasonable that PAKKT-regular and quasinormality CQs are connected. In the next, we discuss this relation.

**Definition 3.1** ([19]). We say that \( x^* \in X \) satisfies the quasinormality constraint qualification if there are no \( \lambda \in \mathbb{R}^m \) and \( \mu \in \mathbb{R}^{|I_g(x^*)|} \) such that
1. \[ \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in I_g(x^*)} \mu_j g_j(x^*) = 0; \]
2. \( (\lambda, \mu) \neq 0 \); and
3. in every neighborhood \( B(x^*) \) of \( x^* \) there is an \( x \in B(x^*) \) such that \( \lambda_i h_i(x) > 0 \) for all \( i \) with \( \lambda_i \neq 0 \) and \( \mu_j g_j(x) > 0 \) for all \( j \) with \( \mu_j > 0 \).

**Theorem 3.3.** Quasinormality implies PAKKT-regular.
Proof. We suppose that \( x^* \) is not PAKKT-regular. Then exists

\[
w^* \in \left( \limsup_{x \to x^*, \alpha \downarrow 0, \beta \downarrow 0} K_+(x, \alpha, \beta) \right) \setminus K_+(x^*, 0, 0).
\]

Let us take \( \{x^k\} \subset \mathbb{R}^n, \{\omega^k\} \subset \mathbb{R}^n, \{\alpha_k\} \subset \mathbb{R} \) and \( \{\beta_k\} \subset \mathbb{R} \) such that \( x^k \to x^*, \omega^k \to \omega^* \), \( \alpha_k \downarrow 0 \), \( \beta_k \downarrow 0 \) and \( \omega^k \in K_+(x^k, \alpha_k, \beta_k) \) for all \( k \geq 1 \), where \( \omega^k \) is as in (8). We define \( \delta_k = \| (\lambda_k^*, \mu_k^*) \|_\infty \). The sequence \( \{\delta_k\} \) is unbounded because, otherwise, we would have \( \omega^* \in K_+(x^*, 0, 0) \) since \( \omega^k \in K_+(x^k, \alpha_k, \beta_k) \) for all \( k \). Thus dividing (8) by \( \delta_k \) and taking the limit we obtain

\[
\sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j \in I_g(x^*)} \mu_j^* \nabla g_j(x^*) = 0
\]

where \((\lambda^*, \mu^*) \neq 0\). Given a neighborhood \( B(x^*) \) of \( x^* \), we have for some \( k \) large enough \( x^k \in B(x^*) \) and \( sgn (\lambda_i^* h_i(x^k)) = sgn (\lambda_i^* h_i(x^k)) = 1 \) whenever \( \lambda_i^* \neq 0 \) (note that \( \lim_k \lambda_i^*/\delta_k = \lambda_i^* \neq 0 \) implies \( |\lambda_i^*| \geq \beta_k \delta_k = \beta_k \| (1, \lambda_k^*, \mu_k^*) \|_\infty \) for all \( k \) sufficiently large). The same happens with \( \mu^* \). Hence \( x^* \) does not satisfy the quasinormality CQ, which completes the proof.

As we already mentioned, AKKT-regular and quasinormality are independent constraint qualifications [9]. Then, from Theorems 3.2 and 3.3 we conclude that PAKKT-regular does not imply any of these two conditions. A geometric comparison between AKKT-regular and PAKKT-regular CQs is given in Figure 2.

![Figure 2: Geometric interpretation of AKKT-regular and PAKKT-regular conditions. The inequality constraints \( g_1(x) \leq 0 \) and \( g_2(x) \leq 0 \) are active at \( x^* \). The sequence \( \{x^k\} \) is feasible with respect to the two constraints, \( \{\tilde{x}^k\} \) is feasible only with respect to the second one and \( \{\tilde{x}^k\} \) is infeasible with respect to both constraints. In the cones \( K_+(\cdot, 0, \infty) \) of AKKT-regular there are no restrictions on the sign of the multipliers beyond non-negativity, and at all points of the sequences, both gradients \( \nabla g_1 \) and \( \nabla g_2 \) take place (figure (a)). The sets \( K_+(\cdot, \alpha, \beta) \) are possibly “discontiguous” in the following sense: if \( 0 \neq z \in K_+(x, \alpha, \beta) \) then it is possible that \( \gamma z \notin K_+(x, \alpha, \beta) \) for all \( \gamma \in [a, b] \) whenever \( a \) is not sufficiently large, where \( a \) and \( b \) depend on \( \alpha \) and \( \beta \). Figure (b) illustrates the case where \( \beta = 0 \). In this situation, the multipliers related to strict satisfied constraints vanish due to the sign control in the PAKKT-regular condition: for the sequence \( \{x^k\} \), both multipliers are null and then \( K_+(\cdot, \alpha, 0) = \emptyset \); for \( \{\tilde{x}^k\} \) only \( \nabla g_1 \) can be present; and for \( \{\tilde{x}^k\} \) both gradients may compose the set \( K_+(\cdot, \alpha, 0) \).](figure.png)

In order to provide a complete relationship between PAKKT-regular and other known constraint qualifications, we will now prove that PAKKT-regular is stronger than Abadie’s CQ [1]. We denote the tangent cone to the feasible set \( X \) of (P) at \( x^* \) by \( T(x^*) \), and its linearization by \( L(x^*) \). Furthermore, \( C^\circ \) will denote the polar of the set \( C \). Recall that Abadie’s CQ consists of the equality \( T(x^*) = L(x^*) \).
Lemma 3.1 ([9, Lemma 4.3]). For each \( x^* \in X \) and \( \omega^* \in \mathcal{T}^o(x^*) \), there are sequences \( \{x^k\} \subset \mathbb{R}^n \), \( \{\lambda^k\} \subset \mathbb{R}^m \) and \( \{\mu^k\} \subset \mathbb{R}_+^p \) such that \( x^k \to x^* \),

1. \( \omega^k = \sum_{i=1}^{m} \lambda^k_i \nabla h_i(x^k) + \sum_{j=1}^{p} \mu^k_j \nabla g_j(x^k) \) converges to \( \omega^* \),

2. \( \lambda^k = kh(x^k) \) and \( \mu^k = kg(x^k)_+ \).

Theorem 3.4. \( \text{PAKKT-regular} \) implies \( \text{Abadie’s CQ} \).

Proof. The multipliers in the item (2.) of Lemma 3.1 have the same sign of their corresponding constraints for all \( k \). Thus the proof follows the same arguments used in [9, Theorem 4.4], taking appropriate sequences \( \{\alpha_k\} \) and \( \{\beta_k\} \).

Example 3.3 (Abadie’s CQ does not imply \( \text{PAKKT-regular} \)). Let us consider the constraints

\[
\begin{align*}
g_1(x) &= x_2 - x_1^2, & g_2(x) &= -x_2 - x_1^2, & g_3(x) &= x_2 - x_1^5, \\
g_4(x) &= -x_2 - x_5^5 & g_5(x) &= -x_1.
\end{align*}
\]

All these constraints are active at the point \( x^* = (0,0) \), which fulfills the Abadie’s CQ since \( \mathcal{T}(x^*) = \mathcal{L}(x^*) = \{(x_1,0) | x_1 \in \mathbb{R}_+\} \). We affirm that \( \text{PAKKT-regular} \) does not hold at \( x^* \). In fact, consider the vector \( \omega^* = (1,0) \). With the sequences defined by \( x^k = (-1/k,0) \), \( \mu^k = (k/4,k/4,k^3,0) \), \( \alpha_k = 1/k^2 \) and \( \beta_k = 1/k \) for all \( k \geq 1 \) we have \( \alpha_k \downarrow 0 \), \( \beta_k \downarrow 0 \), \( \delta_k = \|(1,\mu^k)\|_\infty = k^3 \),

\[
\omega^k = \sum_{j=1}^{5} \mu^k_j \nabla g_j(x^k) = \begin{bmatrix}
1/2 \\
k/4
\end{bmatrix} + \begin{bmatrix}
1/2 \\
-k/4
\end{bmatrix} + \begin{bmatrix}
-5/k \\
k^3
\end{bmatrix} \to \begin{bmatrix}
1 \\
0
\end{bmatrix} = \omega^*
\]

and, for all \( k \), \( \mu^k_1 = \mu^k_2 = k/4 < k^2 = \beta_k \delta_k \), \( \mu^k_3 = 0 < \beta_k \delta_k \) and \( \mu^k_4 = \mu^k_5 = k^3 \geq \beta_k \delta_k \) with \( \mu^k_1 g_3(x^k) = \mu^k_2 g_4(x^k) = 1/k^2 \geq \alpha_k \). Therefore \( \omega^k \in K_+(x^k,\alpha_k,\beta_k) \) for all \( k \), but \( \omega^* \notin K_+(x^*,0,0) = \{(-x_1,x_2) | x_1 \in \mathbb{R}_+, x_2 \in \mathbb{R}\} \). That is, the origin does not satisfy the \( \text{PAKKT-regular CQ} \).

Figure 3 shows the relations between CQs. We emphasize that \( \text{PAKKT-regular} \) unifies the branches from the independent CQs \( \text{AKKT-regular} \) and \( \text{quasinormality} \) under the augmented Lagrangian convergence theory, as we will see in the next section.

![Updated landscape of constraint qualifications](image)

Figure 3: Updated landscape of constraint qualifications. The arrows indicate logical implications. Note that the independent CQs \( \text{AKKT-regular} \) and \( \text{quasinormality} \) imply \( \text{PAKKT-regular} \), which in turn is strictly stronger than \( \text{Abadie’s CQ} \).
4 Global convergence of the augmented Lagrangian method using the PAKKT-regular constraint qualification

Next we present the augmented Lagrangian algorithm proposed to solve (P) [15].

Algorithm 1 Augmented Lagrangian method

Let \( x^0 \in \mathbb{R}^n \) be an arbitrary initial point. The given parameters for the execution of the algorithm are: \( \tau \in [0,1), \gamma > 1, \lambda_{\text{min}} < \lambda_{\text{max}}, \mu_{\text{max}} > 0 \) and \( \rho_0 > 0 \). Also, let \( \tilde{\lambda}_1 \in [\lambda_{\text{min}}, \lambda_{\text{max}}]^m \) and \( \tilde{\mu}_1 \in [0, \mu_{\text{max}}]^p \) be the initial Lagrange multipliers estimates. Finally, \( \{\varepsilon_k\} \subset \mathbb{R}_+ \) is a sequence of tolerance parameters such that \( \lim k \varepsilon_k = 0 \).

**Step 1.** (Solving the subproblem) Compute (if possible) \( x^k \in \mathbb{R}^n \) such that
\[
\|\nabla_x L_{\rho_k}(x^k, \tilde{\lambda}^k, \tilde{\mu}^k)\| \leq \varepsilon_k.
\] (10)

If it is not possible, stop the execution of the algorithm declaring failure.

**Step 2.** (Estimate new multipliers) Compute
\[
\lambda^{k+1} = \tilde{\lambda}^k + \rho_k h(x^k) \quad \text{and} \quad \mu^{k+1} = (\tilde{\mu}^k + \rho_k g(x^k))^+.
\]

**Step 3.** (Update the penalty parameter) Define
\[
V^k = \frac{\mu^{k+1} - \tilde{\mu}^k}{\rho_k}.
\] (11)

If \( k = 1 \) or
\[
\max \left\{ \|h(x^k)\|, \|V^k\| \right\} \leq \tau \max \left\{ \|h(x^{k-1})\|, \|V^{k-1}\| \right\},
\]
choose \( \rho_{k+1} \geq \rho_k \). Otherwise, define \( \rho_{k+1} = \gamma \rho_k \).

**Step 4.** (Update multipliers estimates) Compute \( \tilde{\lambda}^{k+1} \in [\lambda_{\text{min}}, \lambda_{\text{max}}]^m \) and \( \tilde{\mu}^{k+1} \in [0, \mu_{\text{max}}]^p \).

**Step 5.** (Begin a new iteration) Set \( k \leftarrow k + 1 \) and go to Step 1.

Vector \( V^k \) is responsible for measuring infeasibility and non-complementarity with respect to the inequality constraints. If the PHR augmented Lagrangian (1) is used then
\[
V^k = \max \left\{ g(x^k), -\frac{\tilde{\mu}^k}{\rho_k} \right\},
\] (12)
according to [15]. In this paper, we only deal with the augmented Lagrangian (1) (and then we always have (12)), but we note that the general form (11) is also adequate to the case when a non-quadratic penalty augmented Lagrangian function is employed, as in [17, 25].

It is known that when Algorithm 1 does not stop by failure, it generates an AKKT sequence for the problem (P) if its limit is feasible (see [15]). In particular, every feasible limit point of this algorithm is an AKKT point. Next we prove that it also reaches the stronger PAKKT points (from now on, we suppose that the method generates an infinite primal sequence).

**Theorem 4.1.** Every feasible limit point \( x^* \in X \) generated by Algorithm 1 is a PAKKT point.

**Proof.** Let \( \{x^k\}, \{\lambda^k\}, \{\tilde{\mu}^k\} \) and \( \{\rho_k\} \) be sequences generated by Algorithm 1 and \( x^* \) a feasible limit point of \( \{x^k\} \). By (10) we have
\[
\nabla_x L(x^k, \lambda^k, \mu^k) = \nabla f(x^k) + \nabla h(x^k) \lambda^k + \nabla g(x^k) \mu^k \to 0,
\] (13)
\[ \lambda^k = \lambda^k + \rho_k h(x^k) \quad \text{and} \quad \mu^k = (\overline{\mu}^k + \rho_k g(x^k))^+. \]  

With the sequence \( \{(x^k, \lambda^k, \mu^k)\} \), (13) implies (3). If \( \rho_k \to \infty \) then \( \mu_j^k = 0 \) whenever \( g_j(x^*) < 0 \) and \( k \) is sufficiently large. If \( \{\rho_k\} \) is unbounded then \( \lim_k V^k = 0 \), and thus \( \lim_k \mu_j^k = 0 \) whenever \( g_j(x^*) < 0 \). Therefore \( \lim_k \overline{\mu}_j^k = 0 \), and (4) holds.

Let us define \( \delta_k = \|1, \lambda^k, \mu^k\|_\infty \) as in the PAKKT definition. If \( \{\delta_k\} \) is unbounded, from the boundness of \( \{\lambda^k\} \) we have that, for each \( i \) such that \( \lim \lambda_i^k/\delta_k \neq 0 \),

\[
0 \neq \lim_{k} \frac{\lambda_i^k}{\delta_k} = \lim_{k} \left[ \frac{\lambda_i^k}{\delta_k} + \frac{\rho_k h_i(x^k)}{\delta_k} \right] = \lim_{k} \frac{\rho_k h_i(x^k)}{\delta_k} = \lambda_i^k h_i(x^k) > 0, \forall k \geq k_i,
\]

for some \( k_i \geq 1 \). Then (5) is satisfied on a subsequence of \( \{(x^k, \lambda^k, \mu^k)\} \) initializing from the index \( \max_i k_i \). The condition (6) is obtained in the same way for the indices \( j \) such that \( g_j(x^*) = 0 \). Now, if \( g_j(x^*) < 0 \) then (4) implies \( \lim_k \mu_j^k/\delta_k \leq \lim_k \mu_j^k = 0 \), and these indices \( j \) do not violate (6). Therefore, we showed that \( x^* \) is a PAKKT point when \( \{\delta_k\} \) is unbounded.

If \( \{\delta_k\} \) is bounded then, dividing (13) by \( \delta_k \) and taking the limit, we conclude that \( x^* \) is KKT, and then a PAKKT point by Lemma 2.1 (not necessarily with the same primal-dual sequence generated by the method).

It is important to note that Algorithm 1 generates PAKKT points, but not necessarily PAKKT sequences. That is, we claim that for each feasible limit point \( x^* \) there is a correspondent PAKKT sequence, but not necessarily the generated sequence \( \{x^k\} \) is one of them. Specifically, when the dual sequence \( \{\delta_k\} \) is unbounded, \( \{x^k\} \) is in fact a PAKKT sequence associated with \( x^* \). But in the proof of the above theorem we do not have any guarantee that the sequence generated by the method is a PAKKT sequence if \( \{\delta_k\} \) is bounded. Of course, it is not a trouble because in the last case the limit point is already a KKT point. The next example shows that this situation may occur.

**Example 4.1.** Let us consider the minimization of \( f(x) = x \) subject to \(-x \leq 0 \). Then \( \nabla x L_{\rho_k}(x_k, \mu_k) = 0 \) iff \( x_k = (\mu_k - 1)/\rho_k \). If we always choose \( \bar{\mu}_{k+1} = 2 \) in the Step 4, we will have \( (\bar{\mu}_k - \rho_k x_k)_+(-x_k) = -1/\rho_k < 0 \) for all \( k \geq 1 \).

Practical implementations of Algorithm 1 adopt the following updating rule for the Lagrange multipliers in the Step 4:

\[
\lambda_i^{k+1} = \min \{ \lambda_{\max}, \max \{\lambda_{\min}, \lambda_i^k + \rho_k h_i(x^k)\}\}, \quad i = 1, \ldots, m,
\]

\[
\mu_j^{k+1} = \min \{ \mu_{\max}, \max \{0, \mu_j^k + \rho_k g_j(x^k)\}\}, \quad j = 1, \ldots, p.
\]  

This rule corresponds to projecting the estimates \( \lambda^{k+1} \) and \( \mu^{k+1} \) from Step 2 onto the boxes \([\lambda_{\min}, \lambda_{\max}]^m\) and \([0, \mu_{\max}]^p\), respectively. It is used, for example, in the implementation of the so-called augmented Lagrangian method ALGENCAN [2] provided by TANGO project (www.ime.usp.br/~egbirgin/tango). Even with this updating rule, there is no guarantee that the sequence generated by Algorithm 1 is a PAKKT sequence, as the next example illustrates.

**Example 4.2.** Let us consider the same problem of Example 4.1

\[
\min_x x \quad \text{subject to} \quad -x \leq 0.
\]

The origin is the global maximizer and it satisfies the well known linear independence constraint qualification. Although it is a KKT point (and then a PAKKT point by Lemma 2.1), Algorithm 1 may converges to the origin with a non-PAKKT primal sequence. In fact, consider the sequence defined by \( x_k = 1/(k+1)^2/\rho_k \), \( k \geq 0 \), which converges to the origin. We have \( \nabla x L_{\rho_k}(x_k, \mu_k) = 1 - (\mu_k - 1/(k+1)^2) \to 0 \) whenever \( \mu_k \to 1 \). Fixed \( \tau \in (0, 1) \), we also have \( |V_{k-1}| = 1/(k^2 \rho_{k-1}) > \tau/(k+1)^2 \rho_k \) for all \( k \) large enough. If we initialize \( x_0 \) sufficiently closed to the origin, we can suppose without loss of generality
that this occurs for all \( k \geq 0 \). Thus, as \( x_k > 0 \), we have \( \bar{\mu}_{k+1} = \mu_{k+1} = (\bar{\mu}_k - \rho_k x_k)_+ \), for all \( k \geq 1 \) and then

\[
\bar{\mu}_{k+1} = \left[ \bar{\mu}_k - \frac{1}{(k+1)^2} \right]_+ = \cdots = \bar{\mu}_0 - \sum_{i=1}^{k+1} \frac{1}{i^2}.
\]

(16)

The above series is convergent as \( k \to \infty \), and hence it is possible to choose \( \bar{\mu}_0 > 0 \) so that \( \lim_k \bar{\mu}_{k+1} = 1 \) as we wanted. In this case \( \bar{\mu}_k > 0 \) for all \( k \), and then the expression (16) makes sense. We also note that, as \( \rho_{k+1} = \gamma \rho_k \) for all \( k \), the iterate \( x_k \) only depends on \( \rho_0 \) and \( \gamma \), avoiding cyclic definitions. This concludes the discussion.

In [18] the sequential complementarity (4) is changed by a more stringent condition, resulting in the so-called strong AKKT notion. We say that \( x^* \in X \) is a Strong AKKT (SAKKT) point if there are sequences \( \{x^k\} \subseteq \mathbb{R}^n \), \( \{\lambda^k\} \subseteq \mathbb{R}^m \) and \( \{\mu^k\} \subseteq \mathbb{R}^p \) such that (2), (3) hold and, for all \( k \),

\[
g_j(x^k) < 0 \quad \Rightarrow \quad \mu^k_j = 0.
\]

(17)

The authors of [18] also present some relations between SAKKT and AKKT, but no result linking SAKKT points to practical algorithms. The previous example shows, in particular, that Algorithm 1 may generates a non-SAKKT sequence.

It was already known in the literature that the augmented Lagrangian method (Algorithm 1) converges to KKT points under the AKKT-regular constraint qualification [9]. Theorem 4.1 asserts that Algorithm 1 reaches PAKKT points, and then by Theorem 2.2 it converges to a KKT point under the weaker CQ PAKKT-regular. The next result is a direct consequence of Theorems 2.2, 3.3 and 4.1. To the best of our knowledge, it is the first time it has been proved that a practical algorithm for general nonlinear constrained optimization converges to KKT points under the quasinormality CQ.

**Corollary 4.1.** Let \( x^* \) be a feasible limit point of (P) generated by Algorithm 1. If \( x^* \) satisfies the quasinormality constraint qualification then \( x^* \) is a KKT point.

One problem in which the quasinormality CQ plays an interesting role is the Mathematical Program with Complementarity Constraints (MPCC), which is stated as

\[
\min_x f(x) \text{ subject to } h(x) = 0, \ g(x) \leq 0, \ H(x) \geq 0, \ G(x) \geq 0, \ H(x)^T G(x) \leq 0,
\]

where \( f : \mathbb{R}^n \to \mathbb{R}, \ h : \mathbb{R}^n \to \mathbb{R}^m, \ g : \mathbb{R}^n \to \mathbb{R}^p \) and \( H, G : \mathbb{R}^n \to \mathbb{R}^s \). We suppose that all these functions are continuously differentiable. The last inequality constraint, which ensures that \( G \) and \( H \) are complementary, is called complementarity constraint. These problems constitute an important class of optimization problems, and there is an extensive literature about them (see for example [16] and references there in). They are high degenerate problems. For instance, even with the simple constraints \( x_1, x_2 \geq 0 \) and \( x_1 x_2 \leq 0 \), the AKKT-regular CQ is not satisfied at any feasible point (in particular, even Abadie’s CQ fails at the origin). On the other hand, when all the gradients of the active constraints, excluding the complementarity constraint, at a feasible point \( x^* \) are linearly independent (a condition known as MPCC-LICQ [26]) and when the lower level strict complementarity is satisfied at \( x^* \), that is, when \( G_i(x^*) > 0 \) or \( H_i(x^*) > 0 \) for each \( i \), the quasinormality CQ holds. In fact, if \( (\lambda, \mu, \gamma) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}_+^{2s+1} \) is such that

\[
\nabla h(x^*)^T \lambda + \nabla g(x^*)^T \mu - \nabla H(x^*)^T \gamma_H - \nabla G(x^*)^T \gamma_G + \gamma_0 \sum_{i=1}^s (\nabla H_i(x^*)^T G_i(x^*) + \nabla G_i(x^*)^T H_i(x^*)) = 0
\]

then, by the linear independence hypothesis, \( \lambda = 0, \mu = 0 \) and \( \gamma_0 G(x^*) - \gamma_H = \gamma_0 H(x^*) - \gamma_G = 0 \). If \( \gamma_0 = 0 \) then \( \gamma_H = \gamma_G = 0 \), and thus quasinormality holds at \( x^* \). If otherwise \( \gamma_0 > 0 \) then the lower level strict complementarity ensures that \( \gamma_H^i = \gamma_0 G_i(x^*) > 0 \) or \( \gamma_G^i = \gamma_0 H_i(x^*) > 0 \) for each \( i = 1, \ldots, s \). We can suppose without loss of generality that \( \gamma_H^i > 0 \) for all \( i \), and consequently \( H(x^*) = 0, \ G(x^*) > 0 \). By the continuity of \( G \), we have \( G(x) > 0 \) for all \( x \) in a neighborhood \( B(x^*) \) of \( x^* \). If \( \gamma_0^i (H_i(x)) = -\gamma_0 G_i(x^*) H_i(x) > 0 \) for all \( i \) then \( H(x) < 0 \) and \( \gamma_0 (H(x)^T G(x)) < 0 \) for every \( x \in B(x^*) \). This contradicts the third condition of Definition 3.1, and thus \( x^* \) fulfills the quasinormality
CQ (and consequently PAKKT-regular) as we want to prove. We conclude by the previous discussion that Corollary 4.1 implies, with the two hypotheses made on \( x^* \), that Algorithm 1 converges to a KKT point \( x^* \) of MPCC. This result was previously obtained directly, without the help of quasinormality, in [20] (see also the recent report [10]). Actually, the same conclusion can be obtained analogously if we change MPCC-LICQ to the weaker condition MPCC-Mangasarian-Fromovitz CQ (MPCC-MFCQ) defined in [26].

5 Conclusions and future work

A new sequential optimality condition, called Positive Approximate KKT (PAKKT), is defined in the present work. The main goal of this new condition is to take into account the control of the dual sequence inspired in the enhanced Fritz-John optimality conditions developed by Hestenes [19]. This control is related to the external penalty theory and, therefore, it brings the quasinormality constraint qualification to the context. As the Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian method has an intrinsic connection with the pure external penalty method, we were able to prove that this practical method converges to KKT points under the quasinormality constraint qualification, a new result in the literature.

We also provided the strict constraint qualification related to the PAKKT optimality condition, called PAKKT-regular, and we proved that it is weaker than quasinormality and the cone continuity property (see [9]). As a consequence, we generalized all previous theoretical convergence results for the PHR augmented Lagrangian method. In fact, we proved that this method reaches the new PAKKT points. These points are stronger than the Approximate KKT (AKKT) notion defined in [3], which had been used to analyse the convergence of this popular technique [3, 15]. Furthermore, we presented the relationship between PAKKT and other known sequential optimality conditions in the literature.

From the practical point of view, the fact that the PAKKT condition is defined independently of a particular method is very important to generalize convergence properties of existent algorithms, and this will be a topic for a future work.

References


