SIAM J. OPTIM.
Vol. 29, No. 1, pp. 743–766

A SEQUENTIAL OPTIMALITY CONDITION RELATED TO THE QUASI-NORMALITY CONSTRAINT QUALIFICATION AND ITS ALGORITHMIC CONSEQUENCES*

ROBERTO ANDREANI†, NADIA S. FAZZIO‡, MARIA L. SCHUVERDT‡, AND LEONARDO D. SECCHIN§

Abstract. In the present paper, we prove that the augmented Lagrangian method converges to KKT points under the quasi-normality constraint qualification, which is associated with the external penalty theory. An interesting consequence is that the Lagrange multiplier estimates computed by the method remain bounded in the presence of the quasi-normality condition. In order to establish a more general convergence result, 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. PAKKT points are substantially better than points obtained by the classical approximate KKT (AKKT) condition, which has been used to establish theoretical convergence results for several methods. In particular, we present a simple problem with complementarity constraints such that all its feasible points are AKKT, while only the solutions and a pathological point are PAKKT. This shows the efficiency of the methods that reach PAKKT points, particularly the augmented Lagrangian algorithm, in such problems. We also provide the appropriate strict constraint qualification associated with the PAKKT sequential optimality condition, called PAKKT-regular, and we prove that it is strictly weaker than both quasi-normality and the cone continuity property. PAKKT-regular connects both branches of these independent constraint qualifications, generalizing all previous theoretical convergence results for the augmented Lagrangian method in the literature.

Key words. augmented Lagrangian methods, global convergence, constraint qualifications, quasi-normality, sequential optimality conditions

AMS subject classifications. 90C46, 90C30, 65K05

DOI. 10.1137/17M1147330

1. Introduction.

We will consider the general constrained nonlinear problem

$$\text{(P)} \qquad \min_x f(x) \text{ subject to } x \in X,$$

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

$$X = \{x \mid h(x) = 0,\ g(x) \le 0\},$$

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

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.

*Received by the editors September 12, 2017; accepted for publication (in revised form) December 26, 2018; published electronically March 5, 2019.
http://www.siam.org/journals/siopt/29-1/M114733.html
Funding: 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 Sérgio Buarque de Holanda, 651, 13083-859, Campinas, São Paulo, Brazil (andreani@ime.unicamp.br).
‡CONICET, Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina (nadiafazzio@gmail.com, schuverd@mate.unlp.edu.ar).
§Department of Applied Mathematics, Federal University of Espírito Santo, Rodovia BR 101, Km 60, 29932-540, São Mateus, Espírito Santo, Brazil (leonardo.secchin@ufes.br).

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 that are considerably easier to solve. In each subproblem, with a fixed 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 parameter ρ and the multiplier estimates are updated, and a new iteration starts. Specifically, we consider the Powell–Hestenes–Rockafellar (PHR) augmented Lagrangian function that, for problem (P), takes the form

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

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, other functions were also employed; see, for example, [11, 18] and references therein. The choice of (1) is justified by its intrinsic relation to the external penalty theory, where quasi-normality, a very general constraint qualification proposed by Hestenes [22], plays an important role. This is the central issue in this work.

In the last few years, special attention has been devoted to so-called sequential optimality conditions for nonlinear constrained optimization (see, for example, [3, 7, 9, 8, 15, 25]). 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 therein). 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 conditions 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

$$\lim_{k\to\infty} x^k = x^*, \tag{2a}$$

$$\lim_{k\to\infty} \|\nabla_x L(x^k, \lambda^k, \mu^k)\| = 0, \tag{2b}$$

$$\lim_{k\to\infty} \|\min\{-g(x^k), \mu^k\}\| = 0, \tag{2c}$$

where $L(\cdot)$ is the usual Lagrangian function associated with (P). Such kinds 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). In the last few years, it has been proved that AKKT points are stationary points under different constraint qualifications such as *constant positive linear dependence* (CPLD) [6, 27], *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 the AKKT-regular CQ [9]) is the least stringent constraint qualification with this property [8]. All these CQs deal with rank and/or positive linear dependence assumptions. Pseudonormality and quasi-normality CQs [22, 12], on the contrary, have a different nature. They impose a control on the multipliers around the point of interest. In this paper we define a new constraint qualification, called *PAKKT-regular*, that unifies both types of such CQs. In other words, we prove that PAKKT-regular is strictly weaker than both quasi-normality and CCP; see Figure 4. The interest in this new CQ is that it is associated with the convergence of a practical augmented Lagrangian method, as we will discuss later.

Despite the clear similarities between the PHR augmented Lagrangian and the pure external penalty methods, an interesting topic that is still unsolved is the convergence of the augmented Lagrangian method under the quasi-normality CQ. The authors of [8] show that CCP has no relation to quasi-normality or, in particular, that there are examples where quasi-normality and the AKKT condition hold but KKT does not. In this sense, it is not possible to answer the proposed question using only the AKKT condition. Indeed, it is surprising to note that the PHR augmented Lagrangian method, which uses the quadratic penalty-like function (1), naturally generates AKKT points, but it is not trivial to understand how it handles the sign of the multipliers, as performed by the external penalty method, which generates multiplier estimates with the same sign as their corresponding constraints.

In the present paper we address the previous question by proving that the PHR augmented Lagrangian method converges under the new PAKKT-regular CQ (and, consequently, under quasi-normality). To the best of our knowledge, this is the first time it has been proved that a practical algorithm converges under quasi-normality. In particular, we show that their feasible accumulation points are in fact stronger than AKKT points. We call these points *positive approximate Karush–Kuhn–Tucker* (PAKKT), for which PAKKT-regular is the least stringent associated constraint qualification that ensures KKT. More generally, the theoretical convergence of every method that generate AKKT points can be improved if we are able to show that the method actually generates PAKKT points. As an illustration, let us consider the problem

$$\min_x \, (x_1 - 1)^2 + (x_2 - 1)^2 \quad \text{subject to} \quad x_1, x_2 \geq 0, \quad x_1 x_2 \leq 0, \tag{3}$$

for which $(1, 0)$ and $(0, 1)$ are the unique solutions. We affirm that every feasible point $\bar{x}$ is AKKT. In fact, we can suppose without loss of generality that $\bar{x}_2 = 0$. If $\bar{x}_1 > 0$, then it is straightforward to verify that $\bar{x}$ is an AKKT point with the sequences defined by $x^k = (\bar{x}_1, 2(1 - \bar{x}_1)/k)$ and $\mu^k = (0, k\bar{x}_1 - 2, k)$ for all $k \geq 2/\bar{x}_1$. If $\bar{x}_1 = 0$, it is sufficient to take $x^k = (2/k, 2/k)$ and $\mu^k = (0, 0, k)$ for all $k \geq 1$. However, only $(1, 0)$, $(0, 1)$, and $(0, 0)$ are PAKKT points because, in particular, all feasible points, excluding the origin, satisfy the quasi-normality CQ. Thus, while a pure "AKKT method" may reach any point, a method that ensures convergence to PAKKT points will avoid most of the undesirable ones. Problem (3) belongs to the class of *mathematical programs with complementarity constraints* (MPCCs). We make further comments about these problems in subsection 4.1, revisiting and extending previously known results about the convergence of augmented Lagrangian methods [10, 23]. Furthermore, we prove that the sequence of Lagrange multiplier estimates generated by the method applied to the general problem (P) is bounded whenever the quasi-normality condition holds at the accumulation point. This is particularly true for MPCCs under the so-called MPCC–quasi-normality CQ.

The key to defining the PAKKT condition is to take into account the sign of Lagrange multipliers as in the Fritz John necessary conditions described in [12] (see also [22]). Specifically, we rely on the following result.

THEOREM 1.1 (see [12, Proposition 3.3.5]). *Let x^* be a local minimizer of 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*

1. $\sigma \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0$,
2. $(\sigma, \lambda, \mu) \neq 0$,
3. *in every neighborhood $\mathcal{B}(x^*)$ of x^* there is an $x \in \mathcal{B}(x^*)$ 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 $\mu_j > 0$.*

We will refer to Theorem 1.1 as the enhanced Fritz John (necessary) conditions. We call a point x^* that fulfills all the three items of Theorem 1.1 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, 31].

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 other sequential optimality conditions in the literature. Relations of PAKKT-regular with other known constraint qualifications are included in section 3. In section 4 we present the global convergence of the augmented Lagrangian method using the PAKKT-regular CQ. The strength of the PAKKT condition is discussed in subsection 4.1, and the boundedness of the dual sequences generated by the augmented Lagrangian method is treated in subsection 4.2. Conclusions and lines for future research are given in section 5.

Notation.

- $\mathbb{R}_+ = \{t \in \mathbb{R} \mid t \geq 0\}$, $\|\cdot\|$ denotes an arbitrary vector norm, $\|\cdot\|_\infty$ the supremum norm, and $\|\cdot\|_2$ the Euclidean norm.
- v_i is the ith component of the vector v.
- For all $y \in \mathbb{R}^n$, $y_+ = (\max\{0, y_1\}, \ldots, \max\{0, y_n\})$.
- If $K = \{k_0, k_1, k_2, \ldots\} \subset \mathbb{N}$ $(k_{j+1} > k_j)$, we write $\lim_{k \in K} y^k = \lim_{j \to \infty} y^{k_j}$. In particular, $\lim_k y^k = \lim_{k \in \mathbb{N}} y^k$.
- If $\{\gamma_k\} \subset \mathbb{R}$, $\gamma_k > 0$, and $\gamma_k \to 0$, we write $\gamma \downarrow 0$.
- We define the "sign function" $\operatorname{sgn} a$, putting $\operatorname{sgn} a = 1$ if $a > 0$ and $\operatorname{sgn} a = -1$ if $a < 0$. We have $\operatorname{sgn}(a \cdot b) = \operatorname{sgn} a \cdot \operatorname{sgn} b$.

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*

$$\lim_k x^k = x^*, \tag{4}$$

$$\lim_k \|\nabla_x L(x^k, \lambda^k, \mu^k)\| = 0, \tag{5}$$

$$\lim_k \|\min\{-g(x^k), \mu^k\}\| = 0, \tag{6}$$

$$\lambda_i^k h_i(x^k) > 0 \text{ if } \lim_k \frac{|\lambda_i^k|}{\delta_k} > 0, \tag{7}$$

$$\mu_j^k g_j(x^k) > 0 \text{ if } \lim_k \frac{\mu_j^k}{\delta_k} > 0, \tag{8}$$

where $\delta_k = \|(1, \lambda^k, \mu^k)\|_\infty$. *In this case,* $\{x^k\}$ *is called a* PAKKT sequence.

Expressions (4)–(6) are related to the KKT conditions, and they are used in the approximate KKT (AKKT) optimality condition presented in the introduction. Expressions (7) and (8) 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 (7) and (8) give an adequate counterpart for the third item of Theorem 1.1 in the sequential case. As always $|\lambda_i^k|/\delta_k, \mu_j^k/\delta_k \in [0, 1]$, and 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 (7) and (8) is not. The next example shows that the complementary slackness at the limit x^* may fail without condition (6) if $\{\delta_k\}$ is unbounded.

Example 1. Let us consider the problem

$$\min_x \; -x_1 + x_2 \quad \text{subject to} \quad x_2^2 = 0, \quad x_1 - 1 \le 0,$$

for which

$$\nabla_x L(x, \lambda, \mu) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 2x_2 \end{bmatrix} + \mu \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The point $x^* = (0, 0)$ satisfies (4), (5), (7), and (8) with, for example, the sequences defined by $x^k = (0, -1/k)$, $\lambda^k = k/2$, and $\mu^k = 1$ for all $k \ge 1$. In fact, we have $x^k \to (0, 0)$, $\nabla_x L(x^k, \lambda^k, \mu^k) = 0$, $\lambda^k (x_2^k)^2 = 1/(2k) > 0$ with $\lim_k |\lambda^k|/\delta_k = 1$, and $\lim_k \mu^k/\delta_k = 0$. But, taking a subsequence if necessary, any sequence satisfying (4), (5), (7), and (8) is such that $|\lambda^k| \to \infty$ and $\mu^k \to 1$. Thus $\lim_k \mu^k/\delta_k = 0$, but $\min\{-(x_1^k - 1), \mu^k\} \to 1$.

THEOREM 2.2. *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

$$\min_x f(x) + 1/2\|x - x^*\|_2^2 \quad \text{subject to} \quad h(x) = 0, \quad g(x) \le 0, \quad \|x - x^*\| \le \alpha$$

for a certain $\alpha > 0$. Let x^k be a global minimizer of the penalized problem

$$\min_x f(x) + 1/2\|x - x^*\|_2^2 + \frac{\rho_k}{2}\left[\|h(x)\|_2^2 + \|g(x)_+\|_2^2\right] \quad \text{subject to} \quad \|x - x^*\| \le \alpha,$$

$\rho_k > 0$, 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 (4) is satisfied. We have $\|x^k - x^*\| < \alpha$ for all k sufficiently large (let us say for all $k \in K$), and from the optimality conditions of the penalized problem we obtain

$$\lim_{k \in K} \nabla_x L(x^k, \lambda^k, \mu^k) = \lim_{k \in K} \left[\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\right] = -\lim_{k \in K}(x^k - x^*) = 0,$$

where, for each $k \in K$,

$$\lambda^k = \rho_k h(x^k) \quad \text{and} \quad \mu^k = \rho_k g(x^k)_+ \ge 0.$$

Therefore (5) and (6) are satisfied. If $\mu_i^k > 0$, $k \in K$, then $g_i(x^k) > 0$ and hence $\mu_i^k g_i(x^k) = \rho_k [g_i(x^k)]^2 > 0$. Analogously, if $\lambda_i^k \neq 0$, $k \in K$, then $h_i(x^k) \neq 0$ and hence $\lambda_i^k h_i(x^k) = \rho_k [h_i(x^k)]^2 > 0$. Thus (7) and (8) are fulfilled, independently of the limits of the dual sequences. ∎

We say that SCQ is a *strict constraint qualification* for the sequential optimality condition A if

A + SCQ implies KKT

(see [15]). Since all sequential optimality conditions are satisfied in any local minimizer independently of the fulfillment of CQs, an SCQ is in fact a constraint qualification. The reciprocal is not true. For instance, Abadie's CQ [1] or quasi-normality [22] 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 far as SCQ is less stringent (weaker). In [8], the authors presented the weakest strict constraint qualification associated with AKKT, called the *cone continuity property* (CCP). Recently [9], CCP was renamed as "AKKT-regular" and the weakest SCQs related to the SAKKT [21], CAKKT [7], and AGP [25] conditions were established.

In this section, we provide the weakest SCQ for the 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

$$\begin{aligned} &K_+(x,\alpha,\beta) \\ &= \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in I_g(x^*)} \mu_j \nabla g_j(x) \;\middle|\; \begin{array}{l} \lambda_i h_i(x) \geq \alpha \text{ if } |\lambda_i| > \beta \|(1,\lambda,\mu)\|_\infty, \\ \mu_j g_j(x) \geq \alpha \text{ if } \mu_j > \beta \|(1,\lambda,\mu)\|_\infty, \\ \lambda \in \mathbb{R}^m,\ \mu_j \in \mathbb{R}_+,\ j \in I_g(x^*) \end{array} \right\}. \end{aligned}$$

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

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

$$\limsup_{z \to z^*} \Gamma(z) = \{ y^* \in \mathbb{R}^q \mid \exists (z^k, y^k) \to (z^*, y^*) \text{ with } y^k \in \Gamma(z^k) \} \tag{9}$$

(see [28]). We say that Γ is *outer semi-continuous* at z^* if $\limsup_{z\to z^*} \Gamma(z) \subset \Gamma(z^*)$. We define the PAKKT-regular condition, which imposes an outer semi-continuity–like condition on the multifunction $(x,\alpha,\beta) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+ \rightrightarrows K_+(x,\alpha,\beta)$. Analogously to (9), we consider the following set:

$$\begin{aligned} &\limsup_{x\to x^*,\ \alpha\downarrow 0,\ \beta \downarrow 0} K_+(x,\alpha,\beta) \\ &\quad = \{ y^* \in \mathbb{R}^n \mid \exists (x^k, y^k) \to (x^*, y^*),\ \alpha_k \downarrow 0,\ \beta_k \downarrow 0 \text{ with } y^k \in K_+(x^k, \alpha_k, \beta_k) \}. \end{aligned}$$

Definition 2.3. *We say that $x^* \in X$ satisfies the* PAKKT-regular *condition if*

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

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

Theorem 2.4. *If x^* is a PAKKT point that fulfills the PAKKT-regular condition, then x^* is a KKT point. Conversely, 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 $x^k \to x^*$, (6)–(8) hold, and $\nabla f(x^k) + \omega^k \to 0$, where

$$\omega^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k).$$

By (6) we can suppose without loss of generality that $\mu_j^k = 0$ 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 $I_+ = \{i \in \{1, \ldots, m\} \mid \lim_k |\lambda_i^k|/\delta_k > 0\}$ and $J_+ = \{j \in I_g(x^*) \mid \lim_k \mu_j^k/\delta_k > 0\}$, and for each k we take

$$\alpha_k = \min\left\{\frac{1}{k},\ \min_{i \in I_+}\{\lambda_i^k h_i(x^k)\},\ \min_{j \in J_+}\{\mu_j^k g_j(x^k)\}\right\}$$

and

$$\beta_k = \max\left\{\frac{1}{k},\ \max_{i \notin I_+} \frac{|\lambda_i^k|}{\delta_k},\ \max_{j \notin J_+} \frac{\mu_j^k}{\delta_k}\right\} + \frac{1}{k}.$$

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

$$\begin{aligned} -\nabla f(x^*) = \lim_k \omega^k \in \limsup_k K_+(x^k, \alpha_k, \beta_k) \subset \limsup_{x \to x^*,\ \alpha \downarrow 0,\ \beta \downarrow 0} K_+(x, \alpha, \beta) \\ \subset K_+(x^*, 0, 0), \end{aligned}$$

that is, x^* is a KKT point. This proves the first statement.

Now let us show the converse. Let $w^* \in \limsup_{x \to x^*,\ \alpha \downarrow 0,\ \beta \downarrow 0} K_+(x, \alpha, \beta)$. 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 $\lambda^k \in \mathbb{R}^m$ and $\mu^k \in \mathbb{R}_+^{|I_g(x^*)|}$ such that

$$\omega^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in I_g(x^*)} \mu_j^k \nabla g_j(x^k). \tag{10}$$

We define $f(x) = -(\omega^*)^t x$. If $\lim_k |\lambda_i^k|/\delta_k > 0$, then $|\lambda_i^k| > \beta_k \delta_k$ for all k sufficiently large (the same happens with μ). In other words, the control over the sign of the multipliers performed by (7) and (8) 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 hypothesis, x^* is a KKT point, and hence $\lim_k \omega^k = -\nabla f(x^*) \in K_+(x^*, 0, 0)$. This concludes the proof. ∎

As a consequence of Theorems 2.2 and 2.4, it follows that any minimizer of (P) satisfying the PAKKT-regular condition is a KKT point. Equivalently, we obtain the next result.

COROLLARY 2.5. *PAKKT-regular is a constraint qualification.*

As expected, every KKT point is a PAKKT point (see Lemma 2.6 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 2.6. *Every KKT point is a PAKKT point.*

Proof. Let x^* be a KKT point with associated multipliers λ and μ. We will show that there is a PAKKT sequence associated with x^*. In fact, let us consider the sets I and J of indexes of nonzero multipliers λ_i and μ_j, respectively. If $I = J = \emptyset$, then a PAKKT sequence is simply $x^k = x^*$ with $\lambda^k = 0$ and $\mu^k = 0$ for all k. We then suppose that at least one of the sets I and J are nonempty. By [4, Lemma 1] there are sets $\mathcal{I} \subset I$, $\mathcal{J} \subset J$, and vectors $\hat{\lambda}_{\mathcal{I}}$, $\hat{\mu}_{\mathcal{J}}$ such that

$$\nabla f(x^*) + \sum_{i \in \mathcal{I}} \hat{\lambda}_i \nabla h_i(x^*) + \sum_{j \in \mathcal{J}} \hat{\mu}_j \nabla g_j(x^*) = 0,$$

$\hat{\lambda}_i \neq 0$ for all $i \in \mathcal{I}$, $\hat{\mu}_j > 0$ for all $j \in \mathcal{J}$, and the set of corresponding gradients

$$\{\nabla h_i(x^*), \nabla g_j(x^*) \mid i \in \mathcal{I},\, j \in \mathcal{J}\}$$

is linearly independent. In particular, taking $\hat{\lambda}_i = 0$, $i \notin \mathcal{I}$, and $\hat{\mu}_j = 0$, $j \notin \mathcal{J}$, the vectors $\hat{\lambda}$ and $\hat{\mu}$ are KKT multipliers for x^*. We then define the constant sequences

$$\lambda^k = \hat{\lambda} \quad \text{and} \quad \mu^k = \hat{\mu}$$

for all k. Note that the only multipliers that are taken into account in expressions (7) and (8) are those with indexes in $\mathcal{I}$ and $\mathcal{J}$.

Now, we will show that there exists a (PAKKT) sequence $\{x^k\}$, converging to x^*, such that (7) and (8) are satisfied for $\{\lambda^k = \hat{\lambda}\}$ and $\{\mu^k = \hat{\mu}\}$. To do this, we consider for $\gamma > 0$, $i \in \mathcal{I}$, and $j \in \mathcal{J}$ the sets

$$\mathcal{H}_i^\gamma = \left\{ y \in \mathbb{R}^n \backslash \{x^*\} \;\middle|\; \begin{array}{ll} \nabla h_i(x^*)^t \dfrac{y - x^*}{\|y - x^*\|_2} \geq \gamma & \text{if } \hat{\lambda}_i > 0, \\ \nabla h_i(x^*)^t \dfrac{y - x^*}{\|y - x^*\|_2} \leq -\gamma & \text{if } \hat{\lambda}_i < 0 \end{array} \right\}$$

and

$$\mathcal{G}_j^\gamma = \left\{ y \in \mathbb{R}^n \backslash \{x^*\} \;\middle|\; \nabla g_j(x^*)^t \frac{y - x^*}{\|y - x^*\|_2} \geq \gamma \right\}.$$

Note that these sets are nonempty for all $\gamma > 0$ small enough, since $\nabla g_j(x^*) + x^* \in \mathcal{G}_j^\gamma$ for all $0 < \gamma \leq \|\nabla g_j(x^*)\|_2$ (analogously for $\mathcal{H}_i^\gamma$). We affirm that for each $j \in \mathcal{J}$ and $\gamma > 0$ there is an open ball $\mathcal{B}_j(x^*)$ with radius $\omega_j(\gamma) > 0$, centered at x^*, such that

$$g_j(y) > 0 \quad \forall y \in \mathcal{G}_j^\gamma \cap \mathcal{B}_j(x^*). \tag{11}$$

In fact, by the smoothness of function g_j, we can write

$$g_j(y) = g_j(x^*) + \nabla g_j(x^*)^t (y - x^*) + r_j(y - x^*),$$

where $r_j(y - x^*)/\|y - x^*\|_2 \to 0$ when $y \to x^*$. As $j \in \mathcal{J} \subset I_g(x^*)$, it follows that

$$\frac{g_j(y)}{\|y - x^*\|_2} = \nabla g_j(x^*)^t \frac{y - x^*}{\|y - x^*\|_2} + \frac{r_j(y - x^*)}{\|y - x^*\|_2} \geq \frac{\gamma}{2} > 0$$

for all $y \in \mathcal{G}_j$ close to x^*. Analogously, for each $i \in \mathcal{I}$ and $\gamma > 0$, there is an open ball $\mathcal{B}_i(x^*)$ with radius $\omega_i(\gamma) > 0$, centered at x^*, such that

$$h_i(y) \begin{cases} > 0 & \text{if } \hat{\lambda}_i > 0, \\ < 0 & \text{if } \hat{\lambda}_i < 0 \end{cases} \qquad \forall y \in \mathcal{H}_i^\gamma \cap \mathcal{B}_i(x^*). \tag{12}$$

Finally, it is sufficient to show that for some $\gamma > 0$, the intersection

$$\left(\bigcap_{i \in \mathcal{I}} \mathcal{H}_i^\gamma \right) \cap \left(\bigcap_{j \in \mathcal{J}} \mathcal{G}_j^\gamma \right) \tag{13}$$

contains points arbitrarily close to x^*, since in this case any sequence $\{x^k\}$ in this intersection converging to x^* will be a PAKKT sequence associated with the constant sequences $\{\lambda^k = \hat{\lambda}\}$ and $\{\mu^k = \hat{\mu}\}$, by (11) and (12) (or taking a subsequence of $\{x^k\}$ with k large enough if necessary); see Figure 1. Let us consider the set

$$X' = \{x \mid h_i(x) \geq 0,\ h_l(x) \leq 0,\ g_j(x) \geq 0,\ i \in \mathcal{I}_+,\ l \in \mathcal{I}_-,\ j \in \mathcal{J}\},$$

where $\mathcal{I}_+ = \{i \in \mathcal{I} \mid \hat{\lambda}_i > 0\}$ and $\mathcal{I}_- = \{i \in \mathcal{I} \mid \hat{\lambda}_i < 0\}$. With respect to X', the linear independence constraint qualification holds at x^* and therefore the Mangasarian–Fromovitz CQ also holds at x^*. Thus, there is a unitary d such that

$$\nabla h_i(x^*)^t d > 0, \quad \nabla h_l(x^*)^t d < 0, \quad \nabla g_j(x^*)^t d > 0, \quad i \in \mathcal{I}_+,\ l \in \mathcal{I}_-,\ j \in \mathcal{J}.$$

Taking

$$\gamma = \min \left\{ \nabla h_i(x^*)^t d,\ -\nabla h_l(x^*)^t d,\ \nabla g_j(x^*)^t d \mid i \in \mathcal{I}_+,\ l \in \mathcal{I}_-,\ j \in \mathcal{J} \right\} > 0$$

and $y \in \mathbb{R}^n$ such that $d = (y - x^*)/\|y - x^*\|_2$, we conclude that this y belongs to the intersection (13). Since such a y can be taken arbitrarily close to x^*, the proof is complete. ∎

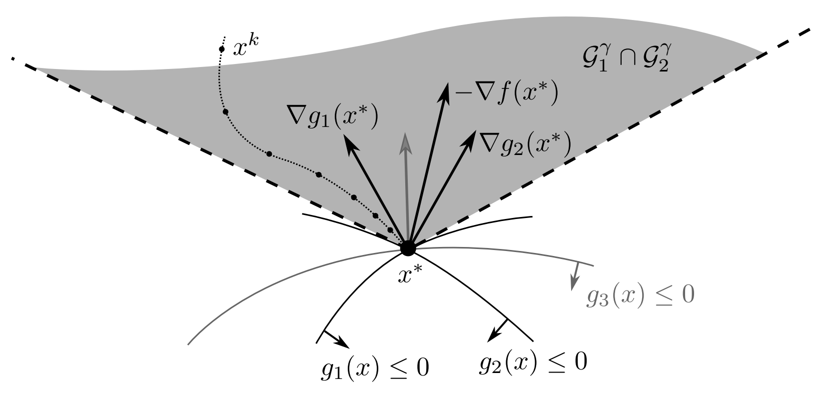


FIG. 1. *Geometry for Lemma 2.6. The constraint $g_3(x) \leq 0$ is redundant, and the gradients of the other active constraints at x^* are linearly independent. The sequence $\{x^k\}$ belongs to the set $\mathcal{G}_1^\gamma \cap \mathcal{G}_2^\gamma$ and converges to x^* from the outside of the feasible set.*

3. Relations.

3.1. Relations with other sequential optimality conditions. In this subsection, we establish the relations between the PAKKT condition and other sequential optimality conditions in the literature. In addition to the AKKT condition presented in the introduction, we consider the following ones.

- We say that $x^* \in X$ 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 (4), (5) hold and, for all $i = 1, \dots, m$ and $j = 1, \dots, p$,

$$\lim_k \lambda_i^k h_i(x^k) = 0 \quad \text{and} \quad \lim_k \mu_j^k g_j(x^k) = 0.$$

  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 X of (P) at x,

$$\Omega(x) = \left\{ z \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} g_j(x) + \nabla g_j(x)^t (z - x) \le 0 & \text{if } g_j(x) < 0, \\ \nabla g_j(x)^t (z - x) \le 0 & \text{if } g_j(x) \ge 0, \\ \nabla h(x)^t (z - x) = 0 & \end{array} \right\}.$$

  We define the *approximate gradient projection* by $d(x) = P_{\Omega(x)}(x - \nabla f(x)) - x$, where $P_C(\cdot)$ denotes the orthogonal projection onto the closed and convex set C. We say that $x^* \in X$ is an *approximate gradient projection* (AGP) [25] 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, 25]. 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 for analyzing accumulation points of inexact restoration techniques [26].

Since the AKKT condition is exactly PAKKT without expressions (7) and (8), 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 2 (a CAKKT point may not be PAKKT). Let us consider the problem

$$\min_x \frac{(x_1 - 1)^2}{2} + \frac{(x_2 + 1)^2}{2} \quad \text{subject to} \quad x_1 x_2 \le 0,$$

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 $k \ge 1$. Furthermore, any CAKKT sequence $\{x^k\}$ converging to the origin with associated $\{\mu^k\}$ satisfies $\lim_k \mu^k x_1^k = -1$ and $\lim_k \mu^k x_2^k = 1$. Thus, for all k sufficiently large we have $\operatorname{sgn} \mu^k = 1$, $\operatorname{sgn} x_1^k = -1$, and $\operatorname{sgn} x_2^k = 1$, and hence $\operatorname{sgn}(\mu^k x_1^k x_2^k) = -1$. That is, the origin is not a PAKKT point.

Example 3 (a PAKKT point may not be AGP). Let us consider the problem

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

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

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

Thus, given a sequence $\{x^k\}$ converging to $(0,-1)$, the set $\Omega(x^k)$ tends to a vertical semi-line on the y-axis if $x_1^k \neq 0$ or it is the semi-plane $z_1 \geq 0$ otherwise. As $x^k - \nabla f(x^k) = (x_1^k, x_2^k - 1) \to (0, -2)$, we always have

$$\|d(x^k)\|_\infty = \|P_{\Omega(x^k)}(x^k - \nabla f(x^k)) - x^k\|_\infty \to 1.$$

Then $(0,-1)$ is not an AGP point.

In particular, as every CAKKT point is AGP (and thus AKKT), Example 2 also shows that there exist AGP and AKKT points that are not PAKKT points. In the same way, Example 3 implies that there exist PAKKT points that are not CAKKT. Figure 2 summarizes the relationships between all sequential optimality conditions discussed here.

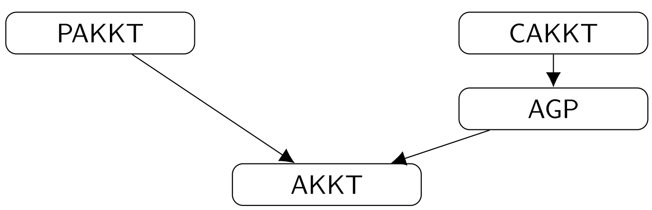


FIG. 2. *Relations between PAKKT and other sequential optimality conditions. An arrow between two sequential conditions means that one is strictly stronger than the other. Note that PAKKT is independent of the AGP and CAKKT conditions.*

3.2. Relations between PAKKT-regular and other known CQs. In section 2, we demonstrated that PAKKT-regular is a constraint qualification (see Corollary 2.5). 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 the cone continuity property, CCP) [8]. Defining the cone

$$K_+(x, 0, \infty) = \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in I_g(x^*)} \mu_j \; g_j(x) \;\middle|\; \lambda \in \mathbb{R}^m, \; \mu_j \in \mathbb{R}_+, \; j \in I_g(x^*) \right\},$$

we say that $x^* \in X$ satisfies the AKKT-regular condition if the multifunction $x \in \mathbb{R}^n \rightrightarrows K_+(x, 0, \infty)$ is outer semi-continuous at x^*, that is, if

$$\limsup_{x \to x^*} K_+(x, 0, \infty) \subset K_+(x^*, 0, \infty).$$

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 a point of (P). These observations are the key to proving the next result.

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

Proof. If x^* satisfies the AKKT-regular condition, then

$$\limsup_{x\to x^*,\,\alpha\downarrow 0,\,\beta\downarrow 0} K_+(x,\alpha,\beta) \subset \limsup_{x\to x^*} K_+(x,0,\infty) \subset K_+(x^*,0,\infty) = K_+(x^*,0,0),$$

which completes the proof. ◻

A natural constraint qualification associated with EFJ points is quasi-normality (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 quasi-normality CQs are connected. In the next part, we discuss this relation.

DEFINITION 3.3 (see [22]). *We say that $x^* \in X$ satisfies the* quasi-normality *constraint qualification if there are no $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}_+^{|I_g(x^*)|}$ such that*

1. $\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j\in I_g(x^*)} \mu_j \nabla g_j(x^*) = 0$,
2. $(\lambda,\mu) \neq 0$, *and*
3. *in every neighborhood $\mathcal{B}(x^*)$ of x^* there is a $x \in \mathcal{B}(x^*)$ 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 3.4. *Quasi-normality implies PAKKT-regular.*

Proof. We suppose that x^* is not PAKKT-regular. Then there exists

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

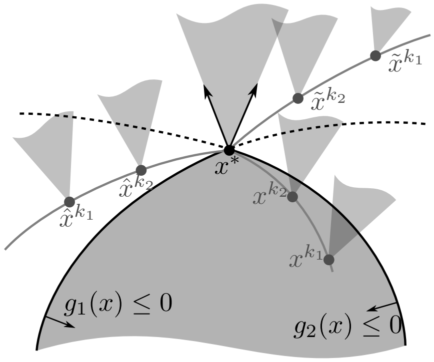
Let us take $\{x^k\} \subset \mathbb{R}^n$, $\{\omega^k\} \subset \mathbb{R}^n$, $\{\alpha_k\} \subset \mathbb{R}$, and $\{\beta_k\} \subset \mathbb{R}$ such that $x^k \to x^*$, $\omega^k \to \omega^*$, $\alpha_k \downarrow 0$, $\beta_k \downarrow 0$, and $\omega^k \in K_+(x^k,\alpha_k,\beta_k)$ for all $k \geq 1$, where ω^k is as in (10). We define $\tilde{\delta}_k = \|(\lambda^k,\mu^k)\|_\infty$. The sequence $\{\tilde{\delta}_k\}$ 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 (10) by $\tilde{\delta}_k$ and taking the limit, we obtain

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

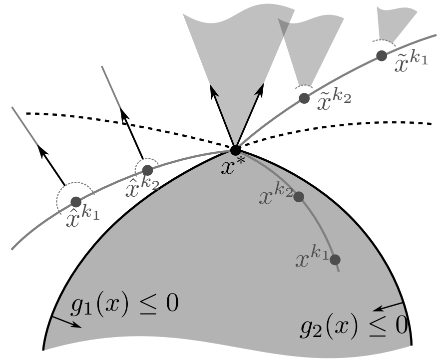
where $(\lambda^*,\mu^*) \neq 0$. Given a neighborhood $\mathcal{B}(x^*)$ of x^*, we have for some k large enough $x^k \in \mathcal{B}(x^*)$ and $\operatorname{sgn}(\lambda_i^* h_i(x^k)) = \operatorname{sgn}(\lambda_i^k h_i(x^k)) = 1$ whenever $\lambda_i^* \neq 0$ (note that $\lim_k \lambda_i^k/\tilde{\delta}_k = \lambda_i^* \neq 0$ implies $|\lambda_i^k| > \beta_k\tilde{\delta}_k = \beta_k\|(1,\lambda^k,\mu^k)\|_\infty$ for all k sufficiently large). The same happens with μ^*. Hence x^* does not satisfy the quasi-normality CQ, which completes the proof. ◻

As we already mentioned, AKKT-regular and quasi-normality are independent constraint qualifications [8]. Then, from Theorems 3.2 and 3.4 we conclude that PAKKT-regular does not imply either of these two conditions. A geometric comparison between AKKT-regular and PAKKT-regular CQs is given in Figure 3.

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^*



(a) *AKKT-regular*



(b) *PAKKT-regular*

FIG. 3. *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, $\{\hat{x}^k\}$ 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 nonnegativity, and at all points of the sequences, both gradients ∇g_1 and ∇g_2 take place (part* (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 α is not sufficiently large, where a and b depend on α and β. Part* (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 $K_+(\cdot, \alpha, 0) = \{0\}$; for $\{\hat{x}^k\}$ only ∇g_1 can be present; and for $\{\tilde{x}^k\}$ both gradients may constitute the set $K_+(\cdot, \alpha, 0)$.*

by $\mathcal{T}(x^*)$, and its linearization by $\mathcal{L}(x^*)$. Furthermore, C° will denote the polar of the set C. Recall that Abadie's CQ consists of the equality $\mathcal{T}(x^*) = \mathcal{L}(x^*)$.

LEMMA 3.5 (see [8, Lemma 4.3]). *For each $x^* \in X$ and $\omega^* \in \mathcal{T}^\circ(x^*)$, 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. *$\omega^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k)$ converges to ω^*,*
2. *$\lambda^k = kh(x^k)$ and $\mu^k = kg(x^k)_+$.*

THEOREM 3.6. *PAKKT-regular implies Abadie's CQ.*

Proof. The multipliers in item 2 of Lemma 3.5 have the same sign of their corresponding constraints for all k. Thus the proof follows the same arguments used in [8, Theorem 4.4], taking appropriate sequences $\{\alpha_k\}$ and $\{\beta_k\}$. ∎

Example 4 (Abadie's CQ does not imply PAKKT-regular). Let us consider the constraints

$$g_1(x) = x_2 - x_1^2, \quad g_2(x) = -x_2 - x_1^2, \quad g_3(x) = x_2 - x_1^5,$$
$$g_4(x) = -x_2 - x_1^5, \quad \text{and} \quad g_5(x) = -x_1.$$

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

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

and, for all k, $\mu_1^k = \mu_2^k = k/4 < k^2 = \beta_k\delta_k$, $\mu_5^k = 0 < \beta_k\delta_k$, and $\mu_3^k = \mu_4^k = k^3 > \beta_k\delta_k$ with $\mu_3^k g_3(x^k) = \mu_4^k g_4(x^k) = 1/k^2 \geq \alpha_k$. Therefore $\omega^k \in K_+(x^k, \alpha_k, \beta_k)$ for all k, but $\omega^* \notin K_+(x^*, 0, 0) = \{(-x_1, x_2) \mid x_1 \in \mathbb{R}_+,\ x_2 \in \mathbb{R}\}$. That is, the origin does not satisfy the PAKKT-regular CQ.

Figure 4 shows the relations between CQs. We emphasize that PAKKT-regular unifies the branches from the independent AKKT-regular and quasi-normality CQs under the augmented Lagrangian convergence theory, as we will see in the next section.

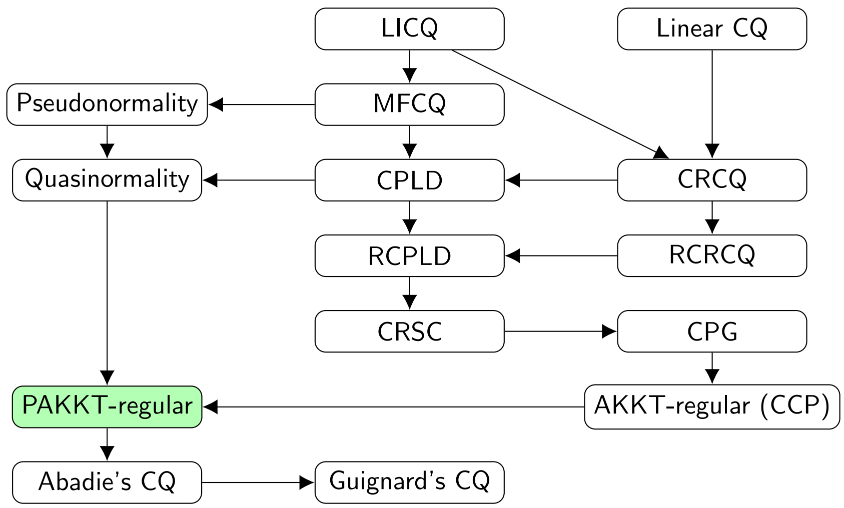


FIG. 4. *Updated landscape of constraint qualifications. The arrows indicate logical implications. Note that the independent AKKT-regular and quasi-normality CQs imply PAKKT-regular, which in turn is strictly stronger than Abadie's CQ.*

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].

The vector V^k is responsible for measuring infeasibility and noncomplementarity with respect to the inequality constraints. If the PHR augmented Lagrangian (1) is used, then

$$V^k = \max\left\{ g(x^k), -\frac{\bar{\mu}^k}{\rho_k} \right\}, \tag{16}$$

according to [15]. In this paper, we only deal with the augmented Lagrangian (1) (and then we always have (16)), but we note that the general form (15) is also appropriate for the case when a nonquadratic penalty augmented Lagrangian function is employed, as in [18, 29].

It is known that when Algorithm 1 does not stop by failure, it generates an AKKT sequence for problem (P) if its limit point is feasible (see [15]). In particular, every feasible accumulation 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 accumulation point $x^* \in X$ generated by Algorithm 1 is a PAKKT point.*

Algorithm 1 Augmented Lagrangian method.

Let $x^1 \in \mathbb{R}^n$ be an arbitrary initial point. The given parameters for the execution of the algorithm are as follows: $\tau \in [0,1)$, $\gamma > 1$, $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, and $\rho_1 > 0$. Also, let $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\bar{\mu}^1 \in [0, \mu_{\max}]^p$ be the initial Lagrange multipliers estimates. Finally, $\{\varepsilon_k\} \subset \mathbb{R}_+$ is a sequence of tolerance parameters such that $\lim_k \varepsilon_k = 0$. Initialize $k \leftarrow 1$.

Step 1. (Solve the subproblem.) Compute (if possible) $x^k \in \mathbb{R}^n$ such that

$$\|\nabla_x L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)\| \le \varepsilon_k. \tag{14}$$

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

Step 2. (Estimate new multipliers.) Compute

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k) \quad \text{and} \quad \mu^k = \left(\bar{\mu}^k + \rho_k g(x^k)\right)_+.$$

Step 3. (Update the penalty parameter.) Define

$$V^k = \frac{\mu^k - \bar{\mu}^k}{\rho_k}. \tag{15}$$

If $k = 1$ or

$$\max\left\{\|h(x^k)\|, \|V^k\|\right\} \le \tau \max\left\{\|h(x^{k-1})\|, \|V^{k-1}\|\right\},$$

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

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

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

Proof. Let $\{x^k\}$, $\{\bar{\lambda}^k\}$, $\{\bar{\mu}^k\}$, and $\{\rho_k\}$ be sequences generated by Algorithm 1 and let x^* be a feasible accumulation point of $\{x^k\}$. We can suppose, taking a subsequence if necessary, that $\lim_k x^k = x^*$. By (14) we have

$$\nabla_x L(x^k, \lambda^k, \mu^k) = \nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k \to 0, \tag{17}$$

where

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k) \quad \text{and} \quad \mu^k = \left(\bar{\mu}^k + \rho_k g(x^k)\right)_+. \tag{18}$$

With the sequence $\{(x^k, \lambda^k, \mu^k)\}$, (17) implies (5). If $\rho_k \to \infty$, then $\mu_j^k = 0$ whenever $g_j(x^*) < 0$ and k is sufficiently large. If $\{\rho_k\}$ is bounded, then $\lim_k V^k = 0$, and thus $\lim_k \mu_j^k = 0$ whenever $g_j(x^*) < 0$. Therefore $\lim_k \mu_j^k = 0$, and (6) holds.

Define $\delta_k = \|(1, \lambda^k, \mu^k)\|_\infty$ as in the PAKKT definition. If $\{\delta_k\}$ is unbounded, then we can suppose, taking a subsequence if necessary, that, in addition to the unboundedness of $\{\delta_k\}$, $\lim_k \lambda_i^k/\delta_k$ exists for all i and $\lim_k \mu_j^k/\delta_k$ exists for all j. Then, from the boundedness of $\{\bar{\lambda}^k\}$ we have that, for each i such that $\lim_k \lambda_i^k/\delta_k \neq 0$,

$$0 \neq \lim_k \frac{\lambda_i^k}{\delta_k} = \lim_k \left[\frac{\bar{\lambda}_i^k}{\delta_k} + \frac{\rho_k h_i(x^k)}{\delta_k}\right] = \lim_k \frac{\rho_k h_i(x^k)}{\delta_k} \quad \Rightarrow \quad \lambda_i^k h_i(x^k) > 0 \ \forall k \geq k_i,$$

for some $k_i \geq 1$. Then (7) is satisfied on a subsequence of $\{(x^k, \lambda^k, \mu^k)\}$ initializing from the index $\max_i k_i$. Condition (8) is obtained in the same way for the indexes j such that $g_j(x^*) = 0$. Now, if $g_j(x^*) < 0$, then (6) implies $\lim_k \mu_j^k/\delta_k \leq \lim_k \mu_j^k = 0$, and these indexes j do not violate (8). Therefore, we have shown that $\{x^k\}$ is a PAKKT sequence, and thus that x^* is a PAKKT point when $\{\delta_k\}$ is unbounded.

If $\{\delta_k\}$ is bounded, then, taking the limit in (17) on an appropriate subsequence, we conclude that x^* is KKT, and then a PAKKT point by Lemma 2.6 (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 accumulation point x^* there is a corresponding PAKKT sequence, but the generated sequence $\{x^k\}$ does not necessarily have a subsequence with this property. Specifically, when the dual sequence $\{\delta_k\}$ is unbounded, $\{x^k\}$ has a PAKKT subsequence associated with x^*. But in the proof of the above theorem we do not have any guarantee that the sequence generated by the method has a PAKKT subsequence if $\{\delta_k\}$ is bounded. Of course, it is not a problem because in the last case the limit point is already a KKT point. The next example shows that this situation may occur.

Example 5. 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 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 step 4:

$$\begin{aligned} \bar{\lambda}_i^{k+1} &= \min\{\, \lambda_{\max}\,,\ \max\{\lambda_{\min}, \bar{\lambda}_i^k + \rho_k h_i(x^k)\}\,\}, \quad i = 1, \ldots, m, \\ \bar{\mu}_j^{k+1} &= \min\{\, \mu_{\max}\,,\ \max\{0, \bar{\mu}_j^k + \rho_k g_j(x^k)\}\,\}, \quad j = 1, \ldots, p. \end{aligned} \tag{19}$$

This rule corresponds to projecting the estimates λ^k and μ^k 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 a convergent subsequence generated by Algorithm 1 is a PAKKT sequence, as the next example illustrates.

Example 6. Let us consider the same problem as in Example 5:

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

The origin is the global minimizer and it satisfies the well-known linear independence constraint qualification. Although it is a KKT point (and then a PAKKT point by Lemma 2.6), 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 $\nabla_x L_{\rho_k}(x_k, \bar{\mu}_k) = 1 - (\bar{\mu}_k - 1/(k+1)^2)_+ \to 0$ whenever $\bar{\mu}_k \to 1$. For a fixed $\tau \in (0,1)$, we also have $|V_{k-1}| = 1/(k^2\rho_{k-1}) > \tau/((k+1)^2\rho_k) = \tau|V_k|$ for all k large enough. If we initialize x_0 sufficiently close 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

$$\bar{\mu}_{k+1} = \left[\bar{\mu}_k - \frac{1}{(k+1)^2}\right]_+ = \cdots = \bar{\mu}_0 - \sum_{i=1}^{k+1} \frac{1}{i^2}. \tag{20}$$

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 expression (20) makes sense. We also note that, as $\rho_{k+1} = \gamma\rho_k$ for all k, the iterate x_k only depends on ρ_0 and γ, avoiding cyclic definitions. This concludes the discussion.

In [21] the sequential complementarity (6) is exchanged for 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 $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, and $\{\mu^k\} \subset \mathbb{R}^p_+$ such that (4) and (5) hold and, for all k,

$$g_j(x^k) < 0 \quad \Rightarrow \quad \mu_j^k = 0. \tag{21}$$

The authors of [21] 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 can generate a non-SAKKT sequence. However, by [7, Theorem 4.1] and [21, Theorem 4.3], every CAKKT point is also SAKKT. Furthermore, when a very weak generalized Lojasiewicz inequality on the sum-of-squares infeasibility measure holds, Algorithm 1 generates CAKKT points [7] (the previous example satisfies this assumption). Thus, although Algorithm 1 does not always generate SAKKT sequences, it reaches SAKKT points.

It is already known in the literature that the augmented Lagrangian method (Algorithm 1) converges to KKT points under the AKKT-regular constraint qualification [8]. Theorem 4.1 asserts that Algorithm 1 reaches PAKKT points, and then by Theorem 2.4 it converges to a KKT point under the less stringent PAKKT-regular CQ. The next result is a direct consequence of Theorem 2.4, 3.4, 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 quasi-normality CQ.

COROLLARY 4.2. *Let x^* be a feasible accumulation point of* (P) *generated by Algorithm* 1. *If x^* satisfies PAKKT-regular (and thus quasi-normality), then x^* is a KKT point.*

4.1. Strength of the PAKKT condition: AKKT vs. PAKKT methods. As already mentioned in the introduction, sequential optimality conditions have been used to strengthen the theoretical convergence results of several methods; see [2, 3, 4, 7, 8, 9, 15, 25] and references therein. The goal is to prove that a specific method achieves, say, AKKT points. Thus, the theoretical convergence of this "AKKT method" is established under any constraint qualification that ensures that an AKKT point is actually KKT. In this sense, we improve the previous convergence results for the PHR augmented Lagrangian method by proving that it reaches PAKKT points (see Theorem 4.1 and Corollary 4.2).

The aim of this section is to illustrate how "PAKKT methods" may have significantly better theoretical convergence results than pure AKKT methods. For this purpose, we consider so-called *mathematical programs with complementarity constraints* (MPCC). It is worth noticing that, although we make some progress, we do not intend to extend existing convergence results for MPCC. In fact, it is known that this type of result was obtained for specific methods, in particular Algorithm 1 [10, 23]. On the other hand, our approach focuses on sequential optimality conditions, which do not depend on the method considered. We emphasize that the AKKT condition is naturally associated with stopping criteria adopted by practical algorithms, since it consists of a simple and computable inexact version of the KKT conditions (see (2)). Our theory ensures that methods based on external penalty approaches generate more than AKKT points, and then it helps to clarify why algorithms such as Algorithm 1 have good behaviour in MPCCs.

Firstly, we note that PAKKT methods are not worst than the AKKT methods, since every PAKKT point is an AKKT one (Theorem 3.1). Now, let us recall the MPCC (3) of the introduction:

$$\min_x \, (x_1-1)^2+(x_2-1)^2 \quad \text{subject to} \quad x_1, x_2 \geq 0, \quad x_1x_2 \leq 0. \tag{22}$$

We have shown that every feasible point of this problem is AKKT. This means that, at least in theory, a generic AKKT method can converge to an arbitrary feasible point. On the other hand, it is straightforward to verify that only the global solutions $(1,0)$ and $(0,1)$ and the point $(0,0)$ are PAKKT. In other words, a generic PAKKT method has a drastically better theoretical convergence guarantee than the pure AKKT methods when applied to (22). This example shows that it is not possible to establish any reasonable convergence theory to the classical stationarity concepts for MPCC (strong, Mordukhovich, Clarke, or even weak stationarity) using the AKKT condition.

Let us analyze the previous problem in a more comprehensive way. The mathematical program with complementarity constraints can be generically stated as

$$\begin{aligned} &\min_x f(x) \\ \text{(MPCC)} \qquad &\text{subject to } \; h(x)=0, \;\; g(x)\leq 0, \\ &\qquad\qquad\quad H(x)\geq 0, \;\; G(x)\geq 0, \;\; H_i(x)G_i(x)\leq 0, \; i=1,\dots,s, \end{aligned}$$

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 s inequality constraints, which ensure that G and H are complementary, are called *complementarity constraints.*

MPCCs constitute an important class of optimization problems, and there is an extensive literature devoted to them (see, for example, [17] and references therein). They are highly degenerate problems. For instance, even with the simple constraints $x_1, x_2 \geq 0$ and $x_1x_2 \leq 0$, it is straightforward to show the AKKT-regular CQ is not satisfied at any feasible point (in particular, even Abadie's CQ fails at the origin). More generally, we can expect, as in problem (22), that every feasible point is AKKT in a wide variety of instances, since the complementary constraints of a generic MPCC can be rewritten in this simple form by inserting slack variables. On the other hand, when all the gradients of the active constraints, excluding the complementarity one, at a feasible point x^* are linearly independent (a condition known as MPCC-LICQ [30]) and when *lower level strict complementarity* is satisfied at x^*, that is, when $G_i(x^*)>0$ or $H_i(x^*)>0$ for each $i=1,\dots,s$, the quasi-normality CQ holds. In fact, if $(\lambda, \mu, \gamma^H, \gamma^G, \gamma^0) \in \mathbb{R}^m \times \mathbb{R}^p_+ \times \mathbb{R}^{3s}_+$ is such that

$$\nabla h(x^*)\lambda + \nabla g(x^*)\mu - \nabla H(x^*)\gamma^H - \nabla G(x^*)\gamma^G + \sum_{i=1}^{s} \gamma_i^0 \left(\nabla H_i(x^*)G_i(x^*) + \nabla G_i(x^*)H_i(x^*)\right) = 0, \tag{23}$$

then, by the linear independence hypothesis, $\lambda = 0$, $\mu = 0$, and $\gamma_i^H - \gamma_i^0 G_i(x^*) = \gamma_i^G - \gamma_i^0 H_i(x^*) = 0$ for all $i = 1, \ldots, s$. If $\gamma^0 = 0$, then $\gamma^H = \gamma^G = 0$, and thus quasi-normality holds at x^*. If otherwise $\gamma_j^0 > 0$ for some fixed j, then lower level strict complementarity ensures that $\gamma_j^H = \gamma_j^0 G_j(x^*) > 0$ or $\gamma_j^G = \gamma_j^0 H_j(x^*) > 0$. We can suppose without loss of generality that $\gamma_j^H > 0$, and consequently $H_j(x^*) = 0$, $G_j(x^*) > 0$. By the continuity of G, we have $G_j(x) > 0$ for all x in a neighborhood $\mathcal{B}(x^*)$ of x^*. If $\gamma_j^H(-H_j(x)) = -\gamma_j^0 G_j(x^*)H_j(x) > 0$, then $H_j(x) < 0$ and $\gamma_j^0(H_j(x)G_j(x)) < 0$ for every $x \in \mathcal{B}(x^*)$. This contradicts the third condition of Definition 3.3, and thus x^* fulfills the quasi-normality CQ (and consequently PAKKT-regular) as we want to prove. We conclude that, under the two hypotheses made on x^*, any PAKKT method converges to a KKT point x^* of (MPCC).

The same previous conclusion about the convergence of PAKKT methods in MPCCs can be obtained if we replace MPCC-LICQ by the much weaker MPCC–quasi-normality condition defined in [24]. This condition is an adaptation of the quasi-normality CQ to the MPCC setting. For a feasible x^* for (MPCC) we define the sets of indexes

$$I_H(x^*) = \{i \mid H_i(x^*) = 0\},\ I_G(x^*) = \{i \mid G_i(x^*) = 0\},\ \text{and}\ I_0(x^*) = I_H(x^*) \cap I_G(x^*).$$

Note that $I_H(x^*) \cup I_G(x^*) = \{1, \ldots, s\}$, and that lower level strict complementarity at x^* consists of imposing $I_0(x^*) = \emptyset$.

DEFINITION 4.3. *We say that a feasible point x^* conforms to the* MPCC–quasi-normality CQ *if there is no $(\lambda, \mu, \gamma^H, \gamma^G) \in \mathbb{R}^m \times \mathbb{R}^p_+ \times \mathbb{R}^{2s}$ such that*

1. $\nabla h(x^*)\lambda + \nabla g(x^*)\mu - \nabla H(x^*)\gamma^H - \nabla G(x^*)\gamma^G = 0$;
2. $\mu_i = 0\ \forall i \notin I_g(x^*)$, $\gamma_i^H = 0\ \forall i \in I_G(x^*)\backslash I_H(x^*)$, $\gamma_i^G = 0\ \forall i \in I_H(x^*)\backslash I_G(x^*)$ *and either* $\gamma_i^H > 0$, $\gamma_i^G > 0$ *or* $\gamma_i^H \gamma_i^G = 0\ \forall i \in I_0(x^*)$;
3. $(\lambda, \mu, \gamma^H, \gamma^G) \neq 0$;
4. *in every neighborhood $\mathcal{B}(x^*)$ of x^* there is an $x \in \mathcal{B}(x^*)$ such that*
   - $\lambda_i h_i(x) > 0$ *for all i with* $\lambda_i \neq 0$,
   - $\mu_i g_i(x) > 0$ *for all i with* $\mu_i > 0$,
   - $-\gamma_i^H H_i(x) > 0$ *for all i with* $\gamma_i^H \neq 0$, *and*
   - $-\gamma_i^G G_i(x) > 0$ *for all i with* $\gamma_i^G \neq 0$.

LEMMA 4.4. *Under lower level strict complementarity, MPCC–quasi-normality implies the standard quasi-normality CQ for* (MPCC).

Proof. Suppose that the standard quasi-normality CQ is not satisfied at x^*. In other words, there are a nonzero vector $(\lambda, \mu, \gamma^H, \gamma^G, \gamma^0) \in \mathbb{R}^m \times \mathbb{R}^p_+ \times \mathbb{R}^{3s}_+$ and a sequence $\{x^k\} \subset \mathbb{R}^n$ converging to x^* such that (23) holds,

$$\mu_i = 0\ \forall i \notin I_g(x^*), \quad \gamma_i^H = 0\ \forall i \notin I_H(x^*), \quad \gamma_i^G = 0\ \forall i \notin I_G(x^*), \tag{24}$$

and, for all k,

$$\lambda_i h_i(x^k) > 0 \text{ whenever } \lambda_i \neq 0, \tag{25a}$$

$$\mu_i g_i(x^k) > 0 \text{ whenever } \mu_i > 0, \tag{25b}$$

$$-\gamma_i^H H_i(x^k) > 0 \text{ whenever } \gamma_i^H > 0, \tag{25c}$$

$$-\gamma_i^G G_i(x^k) > 0 \text{ whenever } \gamma_i^G > 0, \text{ and} \tag{25d}$$

$$\gamma_i^0 H_i(x^k) G_i(x^k) > 0 \text{ whenever } \gamma_i^0 > 0. \tag{25e}$$

We note that, by the lower level strict complementarity assumption, $I_0(x^*) = \emptyset$, $i \notin I_H(x^*)$ iff $i \in I_G(x^*) = I_G(x^*) \backslash I_H(x^*)$, and $i \notin I_G(x^*)$ iff $i \in I_H(x^*) = I_H(x^*) \backslash I_G(x^*)$. That is, condition 2 of Definition 4.3 can be rewritten exactly as (24). Thus, if $\gamma_i^0 = 0$ for all $i = 1, \ldots, s$, then the vector $(\lambda, \mu, \gamma^H, \gamma^G)$ and the sequence $\{x^k\}$ fulfill the four items of Definition 4.3. In this case MPCC–quasi-normality does not hold at x^*, and the statement is proved.

From now on, we suppose that $\gamma_j^0 > 0$ for a fixed index j. We can suppose without loss of generality that $j \in I_H(x^*)$, since the $j \in I_G(x^*)$ case is analogous. Equation (23) can be rewritten as

$$\nabla h(x^*)\lambda + \nabla g(x^*)\mu - \nabla H(x^*)\tilde{\gamma}^H - \nabla G(x^*)\tilde{\gamma}^G = 0,$$

where, for all $i = 1, \ldots, s$,

$$\tilde{\gamma}_i^H = \gamma_i^H - \gamma_i^0 G_i(x^*) \quad \text{and} \quad \tilde{\gamma}_i^G = \gamma_i^G - \gamma_i^0 H_i(x^*). \tag{26}$$

We affirm that $\tilde{\gamma}_j^H \neq 0$. In fact, suppose by contradiction that $\tilde{\gamma}_j^H = 0$. By lower level strict complementarity at x^*, we have $j \notin I_G(x^*)$ and then

$$\gamma_j^0 G_j(x^k) > 0 \tag{27}$$

for all k sufficiently large, let us say, for all $k \in K$. By (26), we also have $\gamma_j^H = \gamma_j^0 G_j(x^*) > 0$. From condition (25c) we then obtain

$$H_j(x^k) < 0 \tag{28}$$

for all k. Therefore, from (27) and (28) we conclude that $\gamma_j^0 H_j(x^k) G_j(x^k) < 0$ for all $k \in K$, contradicting (25e). Thus $\tilde{\gamma}_j^H \neq 0$. In particular, the whole vector of multipliers $(\lambda, \mu, \tilde{\gamma}^H, \tilde{\gamma}^G)$ is nonzero.

Now, we note that (24) and lower level strict complementarity imply $\tilde{\gamma}_i^H = 0 \ \forall i \notin I_H(x^*)$ and $\tilde{\gamma}_i^G = 0 \ \forall i \notin I_G(x^*)$. Furthermore, let us consider a nonzero multiplier $\tilde{\gamma}_i^H = \gamma_i^H - \gamma_i^0 G_i(x^*)$, $i \in I_H(x^*)$. If $\tilde{\gamma}_i^H > 0$, then necessarily $\gamma_i^H > 0$. By (25c), $H_i(x^k) < 0$ for all k and then $-\tilde{\gamma}_i^H H_i(x^k) > 0$ for all k. If otherwise $\tilde{\gamma}_i^H < 0$, then $\gamma_i^0 > 0$. Since $i \notin I_G(x^*)$, we have $G_i(x^k) > 0$ for all k large enough. Then, condition (25e) implies $H_i(x^k) > 0$ for these indexes k, and consequently $-\tilde{\gamma}_i^H H_i(x^k) > 0$ for all k large enough. Therefore, the nonzero vector $(\lambda, \mu, \tilde{\gamma}^H, \tilde{\gamma}^G)$ and the sequence $\{x^k\}$ (passing to a subsequence if necessary) violate the MPCC–quasi-normality condition. This concludes the proof. ∎

THEOREM 4.5. *Let x^* be a feasible accumulation point of* (MPCC) *generated by a generic PAKKT method (in particular, Algorithm* 1*). Suppose that x^* satisfies MPCC–quasi-normality and lower level strict complementarity. Then x^* is a KKT point of* (MPCC).

Proof. This is a direct consequence of Lemma 4.4, and Theorems 2.4 and 3.4. ∎

Theorem 4.5 extends Theorem 3.2 of [23] (see also [10]) in the case where lower level strict complementarity holds. This previous result deals exclusively with augmented Lagrangian methods and was obtained assuming MPCC-LICQ, a much more stringent condition than MPCC–quasi-normality. On the contrary, in addition to allowing us to change MPCC-LICQ to the less stringent MPCC–quasi-normality constraint qualification, the PAKKT condition enables us to prove a result not exclusively valid for Algorithm 1, but for any PAKKT method.

When we deal with a "linear MPCC," i.e., an MPCC where all the functions h, g, H, and G are affine, MPCC–quasi-normality is always satisfied (see, for instance, [19]). In particular, Theorem 4.5 can be applied to this important class of problems.

COROLLARY 4.6. *Suppose that h, g, H, and G are affine functions. If x^* is a feasible accumulation point of* (MPCC) *generated by a generic PAKKT method (in particular, Algorithm* 1*) that satisfies lower level strict complementarity, then x^* is a KKT point of* (MPCC).

4.2. Boundedness of the dual sequences generated by the method. An important consequence of the above discussion is that the sequences of Lagrange multiplier estimates associated with a PAKKT sequence are bounded whenever quasi-normality holds at the limit point x^*. More specifically, we have the following result.

THEOREM 4.7. *Let $\{x^k\}$ be a PAKKT sequence for* (P) *with associated dual sequence $\{(\lambda^k, \mu^k)\}$. If $x^* = \lim_k x^k$ satisfies quasi-normality, then the dual sequence is bounded.*

Proof. Let us define $\delta_k = \|(1, \lambda^k, \mu^k)\|_\infty$ for all k, and suppose that $\{\delta_k\}$ is unbounded. By (7) and (8) we have

$$\lim_k \frac{\lambda_i^k}{\delta_k} \neq 0 \quad \Rightarrow \quad \lambda_i^k h_i(x^k) > 0 \ \forall k \tag{29}$$

and

$$\lim_k \frac{\mu_j^k}{\delta_k} \neq 0 \quad \Rightarrow \quad \mu_j^k g_j(x^k) > 0 \ \forall k, \tag{30}$$

for all $i = 1, \dots, m$ and $j = 1, \dots, p$. By the definition of δ_k, at least one of the left-hand side limits of (29) or (30) is nonzero. Furthermore, condition (5) guarantees

$$\nabla_x L(x^k, \lambda^k, \mu^k) = \nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k \to 0. \tag{31}$$

Thus, dividing (31) by δ_k and taking the limit, the quasi-normality condition at x^* is violated. This concludes the proof. ∎

Theorem 4.7 says that any PAKKT method generates a bounded sequence of Lagrange multiplier estimates under quasi-normality. In particular, the dual sequences generated by Algorithm 1, namely $\{\lambda^k = \bar{\lambda}^k + \rho_k h(x^k)\}$ and $\{\mu^k = (\bar{\mu}^k + \rho_k g(x^k))_+\}$ (see step 2), are bounded under quasi-normality.

COROLLARY 4.8. *Let $\{x^k\}$ be the sequence generated by Algorithm* 1 *applied to* (P). *Suppose that $x^* \in X$ is a feasible accumulation point of $\{x^k\}$ with associated infinite index set K, i.e., $\lim_{k\in K} x^k = x^*$. Also, suppose that x^* satisfies quasi-normality. Then the associated dual sequences $\{\lambda^k\}_{k\in K}$ and $\{\mu^k\}_{k\in K}$ computed in step* 2 *are bounded.*

Proof. By the proof of Theorem 4.1, if the dual sequence $\{(\lambda^k, \mu^k)\}_{k\in K}$ generated by Algorithm 1 is unbounded, then the method in fact generates a PAKKT sequence. Thus, the statement follows from Theorem 4.7. ∎

The boundedness of the multiplier estimates has a practical consequence for the stability of Algorithm 1. In fact, we can expect, at least in theory, that, under the quasi-normality CQ, the multipliers generated by the algorithm do not grow indefinitely, avoiding numerical instabilities. In particular, the boundedness of μ^k in (15) tends to enforce the nullity of V^k if $\mu_{\max}$ is sufficiently large to encompass all μ^k's, leading to the choice $\rho_{k+1} = \rho_k$ in step 3. It is worth noticing that the interior point method implemented in the popular IPOPT package tends to generate unbounded multipliers even for linear problems of the NETLIB collection [20].

Evidently, Corollary 4.8 (and more generally Theorem 4.7) is valid for any CQ stronger than quasi-normality, like pseudonormality, LICQ, MFCQ, CPLD, CRCQ, and linear constraints (see Figure 4). It is surprising that the authors of [6] did not prove this statement for CPLD. Actually, we did not establish this result in the first publicly available version of this paper either, although it is a natural consequence of our theory. Very recently, and citing this earlier version, Bueno, Haeser, and Rojas reported this fact in a more general context, namely, of the generalized Nash equilibrium problems (see [16, Theorem 6.3]). In this sense, it is not the first time that Corollary 4.8 appears in the literature.

Finally, it is worth mentioning that, under the MPCC–quasi-normality condition, Theorem 4.7 and Lemma 4.4 imply the boundedness of MPCC multiplier estimates (26) generated by Algorithm 1 when applied to MPCCs. To the best of our knowledge, this type of result for augmented Lagrangian methods has not been previously established in the literature. We summarize this below.

COROLLARY 4.9. *Let $\{x^k\}$ be the sequence generated by Algorithm* 1 *applied to* (MPCC). *Suppose that $x^* \in X$ is a feasible accumulation point of $\{x^k\}$ with associated infinite index set K, and for which MPCC–quasi-normality is satisfied. Then the associated sequence of the usual Lagrange multipliers $\{(\lambda^k, \mu^k, \gamma^{H,k}, \gamma^{G,k}, \gamma^{0,k})\}_{k\in K}$ computed in step* 2 *is bounded. In particular, the associated sequence of MPCC multipliers $\{(\lambda^k, \mu^k, \tilde{\gamma}^{H,k}, \tilde{\gamma}^{G,k})\}_{k\in K}$, where*

$$\tilde{\gamma}_i^{H,k} = \gamma_i^{H,k} - \gamma_i^{0,k} G_i(x^k) \quad \textit{and} \quad \tilde{\gamma}_i^{G,k} = \gamma_i^{G,k} - \gamma_i^{0,k} H_i(x^k)$$

for all k and $i = 1, \ldots, s$, is bounded.

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 [22]. This control is related to the external penalty theory and, therefore, it brings the quasi-normality constraint qualification into play. 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 quasi-normality constraint qualification, a new result in the literature. In addition, we were able to prove the boundedness of Lagrange multiplier estimates generated by the method.

We also provided the strict constraint qualification related to the PAKKT optimality condition, called PAKKT-regular, and we proved that it is less stringent than both quasi-normality and the cone continuity property (see [8]). 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) points defined in [3], which had been used to analyze the convergence of this popular technique [3, 15]. In this sense, we highlighted the class of degenerate mathematical programs with complementarity constraints (MPCCs) for which the gap between AKKT and PAKKT points is drastic. Some new results in the MPCC context were also obtained. Furthermore, we presented the relationship between PAKKT and other known sequential optimality conditions in the literature.

From a practical point of view, the fact that the PAKKT condition is defined independently of a particular method is very important for generalizing convergence properties of existent algorithms, and this will be a topic for a future work. In particular, as PAKKT describes the behavior of the classical external penalty approach, it is reasonable to expect that other methods based on this technique will be able to achieve PAKKT points.

REFERENCES

[1] J. Abadie, *On the Kuhn–Tucker theorem*, in Nonlinear Programming, J. Abadie, ed., John Wiley, New York, 1967, pp. 21–36.

[2] 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, https://doi.org/10.1137/060654797.

[3] R. Andreani, G. Haeser, and J. M. Martínez, *On sequential optimality conditions for smooth constrained optimization*, Optimization, 60 (2011), pp. 627–641, https://doi.org/10.1080/02331930903578700.

[4] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, *A relaxed constant positive linear dependence constraint qualification and applications*, Math. Program., 135 (2012), pp. 255–273, https://doi.org/10.1007/s10107-011-0456-0.

[5] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, *Two new weak constraint qualifications and applications*, SIAM J. Optim., 22 (2012), pp. 1109–1135, https://doi.org/10.1137/110843939.

[6] R. Andreani, J. M. Martínez, and M. L. Schuverdt, *On the relation between constant positive linear dependence condition and quasinormality constraint qualification*, J. Optim. Theory Appl., 125 (2005), pp. 473–485, https://doi.org/10.1007/s10957-004-1861-9.

[7] R. Andreani, J. M. Martínez, and B. F. Svaiter, *A new sequential optimality condition for constrained optimization and algorithmic consequences*, SIAM J. Optim., 20 (2010), pp. 3533–3554, https://doi.org/10.1137/090777189.

[8] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva, *A cone-continuity constraint qualification and algorithmic consequences*, SIAM J. Optim., 26 (2016), pp. 96–110, https://doi.org/10.1137/15M1008488.

[9] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva, *Strict constraint qualifications and sequential optimality conditions for constrained optimization*, Math. Oper. Res., 43 (2018), pp. 693–1050, https://doi.org/10.1287/moor.2017.0879.

[10] R. Andreani, L. D. Secchin, and P. J. S. Silva, *Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints*, SIAM J. Optim., 28 (2018), pp. 2574–2600, https://doi.org/10.1137/17M1125698.

[11] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, Nashua, NH, 1996.

[12] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Nashua, NH, 1999.

[13] D. P. Bertsekas, A. E. Ozdaglar, and P. Tseng, *Enhanced Fritz John conditions for convex programming*, SIAM J. Optim., 16 (2006), pp. 766–797, https://doi.org/10.1137/040613068.

[14] E. G. Birgin, R. A. Castillo, and J. M. Martínez, *Numerical comparison of augmented Lagrangian algorithms for nonconvex problems*, Comput. Optim. Appl., 31 (2005), pp. 31–55, https://doi.org/10.1007/s10589-005-1066-7.

[15] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM, Philadelphia, PA, 2014, https://doi.org/10.1137/1.9781611973365.

[16] L. F. Bueno, G. Haeser, and F. N. Rojas, *Optimality Conditions and Constraint Qualifications for Generalized Nash Equilibrium Problems and Their Practical Implications*, Technical report, http://www.optimization-online.org/DB_HTML/2017/12/6386.html, 2017.

[17] S. Dempe, *Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints*, Optimization, 52 (2003), pp. 333–359, https://doi.org/10.1080/0233193031000149894.

[18] N. Echebest, M. D. Sánchez, and M. L. Schuverdt, *Convergence results of an augmented Lagrangian method using the exponential penalty function*, J. Optim. Theory Appl., 168 (2016), pp. 92–108, https://doi.org/10.1007/s10957-015-0735-7.

[19] L. Guo and G.-H. Lin, *Notes on some constraint qualifications for mathematical programs with equilibrium constraints*, J. Optim. Theory Appl., 156 (2013), pp. 600–616, https://doi.org/10.1007/s10957-012-0084-8.

[20] G. Haeser, O. Hinder, and Y. Ye, *On the Behavior of Lagrange Multipliers in Convex and Non-Convex Infeasible Interior Point Methods*, Technical report, http://www.optimization-online.org/DB_HTML/2017/07/6139.html, 2017.

[21] G. Haeser and M. L. Schuverdt, *On approximate KKT condition and its extension to continuous variational inequalities*, J. Optim. Theory Appl., 149 (2011), pp. 528–539, https://doi.org/10.1007/s10957-011-9802-x.

[22] M. R. Hestenes, *Otimization Theory: The Finite Dimensional Case*, John Wiley, New York, 1975.

[23] A. F. Izmailov, M. V. Solodov, and E. I. Uskov, *Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints*, SIAM J. Optim., 22 (2012), pp. 1579–1606, https://doi.org/10.1137/120868359.

[24] C. Kanzow and A. Schwartz, *Mathematical programs with equilibrium constraints: Enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results*, SIAM J. Optim., 20 (2010), pp. 2730–2753, https://doi.org/10.1137/090774975.

[25] 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, https://doi.org/10.1023/A:1024791525441.

[26] J. M. Martínez and E. A. Pilotta, *Inexact Restoration Methods for Nonlinear Programming: Advances and Perspectives*, Springer, Boston, MA, 2005, pp. 271–291, https://doi.org/10.1007/0-387-24255-4_12.

[27] 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, https://doi.org/10.1137/S1052623497326629.

[28] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, 1st ed., Grundlehren Math. Wiss. 317, Springer, Berlin, Heidelberg, 1998, https://doi.org/10.1007/978-3-642-02431-3.

[29] M. D. Sánchez and M. L. Schuverdt, *A Second Order Convergence Augmented Lagrangian Method Using Non-Quadratic Penalty Functions*, Opsearch, submitted.

[30] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22, https://doi.org/10.1287/moor.25.1.1.15213.

[31] J. J. Ye and J. Zhang, *Enhanced Karush–Kuhn–Tucker condition and weaker constraint qualifications*, Math. Program., 139 (2013), pp. 353–381, https://doi.org/10.1007/s10107-013-0667-7.