On the convergence of augmented Lagrangian strategies for nonlinear programming[*]

R. Andreani[†]   A. R. V. Cárdenas[‡]   A. Ramos [‡]

A. A. Ribeiro[‡]   L. D. Secchin[§]

March 28, 2020 (revised Dec 02, 2020)

Abstract

Augmented Lagrangian algorithms are very popular and successful methods for solving constrained optimization problems. Recently, the global convergence analysis of these methods has been dramatically improved by using the notion of sequential optimality conditions. Such conditions are necessary for optimality, regardless of the fulfillment of any constraint qualifications, and provide theoretical tools to justify stopping criteria of several numerical optimization methods. Here, we introduce a new sequential optimality condition stronger than the previous stated in the literature. We show that a well-established safeguarded Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian algorithm generates points that satisfy the new condition under a Lojasiewicz-type assumption, improving and unifying all the previous convergence results. Furthermore, we introduce a new primal-dual augmented Lagrangian method capable of achieving such points without the Lojasiewicz hypothesis. We then propose a hybrid method in which the new strategy acts to help the safeguarded PHR method when it tends to fail. We show by preliminary numerical tests that all the problems already successfully solved by the safeguarded PHR method remain unchanged, while others where the PHR method failed, are now solved with an acceptable additional computational cost.

Keywords: Nonlinear optimization, Augmented Lagrangian methods, Optimality conditions, Approximate KKT conditions, Stopping criteria.

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

[†]Department of Applied Mathematics, University of Campinas, Rua Sérgio Buarque de Holanda, 651, 13083-859, Campinas, SP, Brazil. E-mail: andreani@unicamp.br

[‡]Department of Mathematics, Federal University of Paraná, 81531-980, Curitiba, PR, Brazil. E-mail: ariel.rv68@gmail.com, albertoramos@ufpr.br, ademir.ribeiro@ufpr.br.

[§]Department of Applied Mathematics, Federal University of Espírito Santo, Rodovia BR 101, Km 60, 29932-540, São Mateus, ES, Brazil. E-mail: leonardo.secchin@ufes.br

1

1 Introduction

In this work, we deal with constrained optimization problems of the form

$$\text{minimize } f(x) \text{ subject to } h(x) = 0, \ \ g(x) \leq 0, \qquad \text{(NLP)}$$

where the functions $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions.

Numerical methods for solving (NLP) are often iterative, and include stopping criteria that indicate when the current iterate is close to a solution. In practice, several optimization algorithms test the approximate fulfillment of the Karush-Kuhn-Tucker (KKT) conditions. Such practice is theoretically justified by the fact that every local minimizer of (NLP) is a limit point of certain sequences satisfying approximately the KKT conditions with tolerances going to zero [3, 5, 12, 28]. It is worth noting that such approximations are possible even at local minimizers where KKT fails. Thus, in addition to encompassing natural numerical approximations, such a strategy allows us to describe degenerate minimizers. This idea leads to the notion of sequential optimality condition which we discuss in detail in this work.

The KKT conditions may be approximated in different ways. One of the most popular is to require that, for some sequences $\{x^k\} \to x^*$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$ and $\{\varepsilon_k\} \to 0$, we have

$$\|\nabla L(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k \quad \text{and} \quad |\min\{-g_j(x^k), \mu_j^k\}| \leq \varepsilon_k, \ \forall j, \ \forall k, \qquad (1)$$

where L stands for the Lagrangian function. Condition (1) is known as *approximate KKT* (AKKT) [5, 17]. Such condition was useful to analyze the global convergence of several methods for solving (NLP), see [3, 7] and references therein. However, for some situations, (1) may lead to accept spurious candidates as a possible solution: consider the problem

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

Here, the only minimizers are $(1, 0)$ and $(0, 1)$, but *any* feasible point of (2) is the limit of a sequence $\{x^k\}$ satisfying (1), see [3]. Thus, in theory, numerical methods for solving (2) that use stopping criteria based on (1) may accept any feasible point as a solution. As inexactness is a natural (perhaps inevitable) issue in the numerical world, the way that we approximate stationarity becomes an important question in the study of the theoretical convergence of practical methods for solving (NLP). In particular, this issue is treated very recently in [8] for a class of problems that includes (2).

Such observations lead to the search for sequential optimality conditions that are stronger than (1), to avoid spurious points as much as possible. As (1), As (1), we do not require any constraint qualification (CQ), that is, such conditions are truly necessary for optimality. Also, they imply KKT conditions under weaker CQs, and usually provide stopping criteria for different methods. This property makes the sequential optimality conditions a useful tool for the improvement of global convergence analysis of several NLP methods, including augmented Lagrangian, sequential quadratic programming, interior-point and inexact-restoration methods. See [6, 7, 10, 11, 28]. Furthermore, such concepts have been extended beyond standard nonlinear programming as mathematical

programs with complementary constraints [8, 29], semidefinite programming [9], nonsmooth optimization [25] and multiobjective optimization [20].

Among the sequential optimality conditions for NLP, we mention the *positive approximate KKT* (PAKKT) [3] and the *complementary approximate KKT* (CAKKT) [12] (see Definition 1). Both conditions improve the convergence analysis of augmented Lagrangian (AL) methods and, since they are independent to each other, they capture different features of the minimizers. Using the PAKKT condition, it was proved that accumulation points generated by the safeguarded Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian method are KKT under the quasinormality CQ, see [3]. On the other hand, it was proved that this method ensures CAKKT sequences under an additional hypothesis, namely, that the quadratic measure of infeasibility associated with (NLP) satisfies a generalized Lojasiewicz (GL) inequality at the limit point [12] (see Section 3 for the definition).

In this work we show that besides the safeguarded PHR AL method reaches PAKKT and CAKKT points, the associated sequences have superior properties. We start with a curious fact: there are situations where x^* is both PAKKT and CAKKT point, but there is no sequence that carries both properties simultaneously; on the other hand, the sequences generated by the algorithm ensure these two qualities. In other words, the algorithm reaches points with superior properties than those guaranteed by previous results. To unify these convergence results under the framework of sequential optimality conditions, we define a new one that we call *positive complementary approximate KKT* (PCAKKT), see Definition 2. Of course, PCAKKT is independent of algorithms, and potentially can be used to prove convergence for other optimization methods. Furthermore, motivated by the PCAKKT condition, we propose a new primal-dual augmented Lagrangian method, one that employs a new augmented Lagrangian function, defined below, in their subproblems.

Given $\rho, \nu > 0$, $\bar{\lambda} \in \mathbb{R}^m$ and $\bar{\mu} \in \mathbb{R}^p_+$, the proposed primal-dual augmented Lagrangian function is

$$L_{\rho,\nu,\bar{\lambda},\bar{\mu}}(x, \lambda^a, \mu^a) := f(x) + \frac{\rho}{2} \left\| h(x) + \frac{\bar{\lambda}}{\rho} \right\|^2 + \frac{\rho}{2} \left\| \left(g(x) + \frac{\bar{\mu}}{\rho} \right)_+ \right\|^2 \tag{3a}$$

$$+ \frac{\rho}{2} \left\| h(x) + \frac{\bar{\lambda} - \lambda^a}{\rho} \right\|^2 + \frac{\rho}{2} \left\| \left(g(x) + \frac{\bar{\mu} - \mu^a}{\rho} \right)_+ \right\|^2 + \frac{\rho}{2} \left\| \frac{\mu^a}{\rho} \right\|^2 \tag{3b}$$

$$+ \frac{\nu}{2} \sum_{i=1}^m \left(\lambda_i^a h_i(x) \right)^2 + \frac{\nu}{2} \sum_{j=1}^p \left(\mu_j^a g_j(x) \right)_+^2, \tag{3c}$$

where $\lambda^a \in \mathbb{R}^m$, $\mu^a \in \mathbb{R}^p_+$, and $\| \cdot \|$ stands for the Euclidean norm. Here, λ^a and μ^a can be interpreted as estimates of the Lagrange multiplier vectors for the constraints $h(x) = 0$ and $g(x) \leq 0$, respectively. The vectors $\bar{\lambda}$ and $\bar{\mu}$ play the role of safeguarded multipliers as in the safeguarded PHR AL method. At each iteration, we solve approximately the problem of minimizing $L_{\rho,\nu,\bar{\lambda},\bar{\mu}}(x, \lambda^a, \mu^a)$ subject to $\mu^a \geq 0$ (this justifies the name "primal-dual"). Observe that the term in the right side of (3a) corresponds to the PHR augmented Lagrangian function used in many successfully numerical methods as LANCELOT [19] and ALGENCAN [2, 17]. In turn, the terms (3a)–(3b) correspond to the stabilized primal-dual augmented Lagrangian function presented in [24] for equality con-

3

straints used to develop a sequential quadratic programming (SQP) method [22]. For more details of this primal-dual augmented Lagrangian function, see [22, 24]. Finally, we consider the additional term (3c), which aims to control the fulfillment of the complementary condition. To the best of our knowledge, it is new in the literature.

The proposed method employs rules to control the growth of parameters ρ and ν in (3). The idea is to increase ρ (respectively ν) only if the feasibility (respectively complementarity) is not improved between consecutive iterations. Thus, when everything "goes well", we can expect to reduce the numerical instabilities associated with a large parameter. This type of rule is used successfully in the ALGENCAN package [17]. We show that the new primal-dual method is capable of recovering PCAKKT points in the case that only one of ρ or ν remains bounded. In particular, the term (3c) is treated separately from the others, and there is, at least theoretically, the possibility to reach PCAKKT (and so CAKKT) points without additional assumptions on the problem, even when ρ tends to infinity. We recall that to obtain CAKKT points, the safeguarded PHR AL method need the validity of GL inequality, already mentioned.

Preliminary numerical tests were performed. Although the proposed method has good theoretical convergence results, their subproblems are more challenging than those of PHR AL: they involve minimization in both primal and dual variables. Furthermore, state-of-the-art solvers such as ALGENCAN are effective on a wide variety of problems. However, as expected, they occasionally fail. Thus, hybrid strategies that try to overcome the difficulties encountered by traditional algorithms are reasonable. For instance, in [14], the authors proposed a hybrid second-order AL algorithm to solve a class of degenerate problems. This algorithm employs second-order information only when first order stationarity tends to fail, which leads to better results in some cases. In this sense, we propose a hybridization of the safeguarded AL PHR with the proposed primal-dual method. Our preliminary tests indicate that the hybrid strategy leads to an improvement in convergence for some cases, while previously successfully solved problems are maintained. Although the overall additional computational cost with primal-dual iterations is not prohibitive, we believe that a specialized and optimized implementation can perform much better. In fact, we solve primal-dual subproblems using the standard inner solver of ALGENCAN package, called GENCAN [15], which is an active-set algorithm with spectral gradients, and it is optimized to handle the PHR augmented Lagrangian function, not (3).

1.1 Contributions and organization of the paper

We summarize our main contributions in the three topics below, covered in separate sections throughout the paper:

- In Section 2, *we present our new PCAKKT sequential optimality condition* motivated by the fact that common PAKKT and CAKKT sequences are stronger than requiring these properties at limit points only. The relations of PCAKKT with other conditions from the literature are presented, as well as the least stringent CQ that ensures that a PCAKKT point is KKT. We will show that this new CQ, called PCAKKT-regular, is strictly implied by all other known CQs from the literature associated with the convergence of algorithms;

- In Section 3, *we prove that the well-studied safeguarded PHR augmented Lagrangian method reaches PCAKKT points* under a mild hypothesis known from the literature. Thus, we improve previous results regarding this method;

- In Section 4, *we present a new primal-dual augmented Lagrangian* method that employs (3). In particular, we show that the sequences generated by this method are PCAKKT under very mild assumptions. In this sense, the new method has stronger convergence properties than others from the literature. Section 5 is devoted to numerical experience. We describe the hybrid strategy in detail, discussing how we can use the primal-dual iterations to improve the well established PHR AL method. We report instances where such improvements have been observed.

Finally, Section 6 presents our conclusions and possibilities for future work.

1.2 Notation and terminology

Our notation is standard in optimization and variational analysis. The symbols $\|\cdot\|$ and $\|\cdot\|_\infty$ stand for the Euclidean and sup norms, respectively. We set $\beta_+ := \max\{0, \beta\}$ ($\beta \in \mathbb{R}$) and $z_+ = ((z_1)_+, \ldots, (z_n)_+)$ ($z \in \mathbb{R}^n$). If $y, z \in \mathbb{R}^n$ then $y * z = (y_1 z_1, \ldots, y_n z_n) \in \mathbb{R}^n$ is the Hadamard product between y and z. The orthogonal projection of $z \in \mathbb{R}^n$ onto the closed convex set C is denoted by $P_C(z)$. The symbol $\beta \downarrow 0$ means that $\beta \geq 0$ and $\beta \to 0$, while $\beta \downarrow 0^+$ stands for $\beta > 0$ and $\beta \to 0$. The Lagrangian function associated with (NLP) is

$$L(x, \lambda, \mu) := f(x) + h(x)^T \lambda + g(x)^T \mu,$$

where $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ are the dual variables. The set of indexes of active inequality constraints is denoted by $I_g(x) = \{j \in \{1, \ldots, p\} \mid g_j(x) = 0\}$. Given a function q, $\nabla_z q$ is its gradient with respect to z.

For a given set-valued mapping $\mathcal{K} : \mathbb{R}^s \rightrightarrows \mathbb{R}^n$, the *sequential Painlevé-Kuratowski outer/upper limit* of $\mathcal{K}(z)$ as $z \to z^*$ [30] is defined as the set

$$\limsup_{z \to z^*} \mathcal{K}(z) = \{y^* \in \mathbb{R}^n \mid \exists (z^k, y^k) \to (z^*, y^*) \text{ with } y^k \in \mathcal{K}(z^k), \quad \forall k \in \mathbb{N}\}.$$

2 A new sequential optimality condition

In this section we define the proposed sequential optimality condition, called *Positive Complementary Approximate* KKT (PCAKKT) condition. As every reasonable sequential condition, (i) it is necessary for optimality independently of the fulfillment of any CQ; (ii) it implies optimality conditions of the form "KKT or not-CQ" for some CQs; and (iii) there are numerical methods for solving (NLP) that generate sequences of iterates whose accumulation points satisfy it. In this section, we show that the PCAKKT condition fulfills (i) and (ii). Sections 3 and 4 are devoted to treat the third property.

2.1 The new optimality condition and its relation to other ones from the literature

Here, we show that PCAKKT is an optimality condition and derive some important properties. First we recall the definitions of some useful sequential
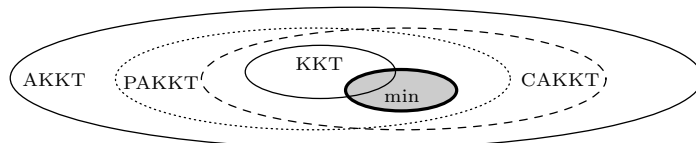
Figure 1: Known sequential optimality conditions from literature. Some minimizers are not KKT, but they all satisfy any sequential optimality condition. AKKT is the least stringent, while the more stringent conditions PAKKT and CAKKT are independent of each other.

optimality conditions from the literature. They differ, essentially, in how complementarity is approximated.

Definition 1. *Let x^* be a feasible point for* (NLP). *Suppose that there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$ and $\{\mu^k\} \subset \mathbb{R}^p_+$ such that*

$$\lim_k x^k = x^* \quad and \quad \lim_k \|\nabla_x L(x^k, \lambda^k, \mu^k)\| = 0. \tag{4}$$

We say that $\{x^k\}$ is

(i) *[5] an* Approximate KKT *(AKKT) sequence if, additionally to* (4),

$$\lim_k \| \min\{-g(x^k), \mu^k\}\| = 0. \tag{5}$$

In this case, the limit x^* is an AKKT *point;*

(ii) *[12] a* Complementary Approximate KKT *(CAKKT) sequence if, additionally to* (4), *we have*

$$\lim_k c_k = 0 \quad where \quad c_k := \sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j=1}^p |\mu_j^k g_j(x^k)|, \quad \forall k. \tag{6}$$

In this case, x^* is a CAKKT *point;*

(iii) *[3] a* Positive Approximate KKT *(PAKKT) sequence if, additionally to* (4), *condition* (5) *holds and*

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

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

Clearly, condition (6) implies (5), and thus CAKKT implies AKKT. It is clear that PAKKT also implies AKKT, but it is known that CAKKT and PAKKT are independent of each other [3]. All these implications, illustrated in Figure 1, are strict.

An interesting issue is the following: Suppose that x^* is simultaneously CAKKT and PAKKT point. This means that there is a CAKKT *sequence* converging to x^* and a PAKKT *sequence* also converging to x^*. The question is whether there is a *common sequence* that characterizes x^* as CAKKT and PAKKT point. Contrary to what we might expect, it is not true in general. Curiously, it is possible that a point x^* is characterized by two distinct sequences $\{\tilde{x}^k\}$ and $\{\bar{x}^k\}$, one CAKKT and other PAKKT, without the existence of a *common* sequence. Example 1 illustrates this situation.

Example 1. *Let us consider the problem*

$$\text{minimize } \frac{(x_1 - 1)^2}{2} + \frac{(x_2 + 1)^2}{2} \text{ subject to } x_1^3 x_2^3 \le 0, \ x_2 \le 0, \ e^{x_1} \sin^2 x_2 \le 0.$$

The gradient of the Lagrangian function is

$$\nabla_x L(x, \mu) = \begin{bmatrix} x_1 - 1 \\ x_2 + 1 \end{bmatrix} + \mu_1 \begin{bmatrix} 3x_1^2 x_2^3 \\ 3x_1^3 x_2^2 \end{bmatrix} + \mu_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \mu_3 \begin{bmatrix} e^{x_1} \sin^2 x_2 \\ 2e^{x_1} \sin x_2 \cos x_2 \end{bmatrix}. \quad (8)$$

Clearly, $x^* = (0,0)$ is not a local minimizer and the KKT conditions fail at x^*. On the other hand, it is easy to verify that x^* is a CAKKT point with the sequences defined by $\tilde{x}^k := (-1/k, 1/k)$, $\tilde{\mu}^k := (k^5/3, 0, 0)$ for all $k \ge 1$. Furthermore, x^* is a PAKKT point by taking the sequences

$$\bar{x}^k := (1/k, -1/k), \quad \bar{\mu}^k := (0, -1 + 2[\tan(1/k)]^{-1}, [e^{1/k} \sin^2(1/k)]^{-1}).$$

In fact, we have $\bar{\mu}_3^k(e^{\bar{x}_1^k} \sin^2 \bar{x}_2^k) > 0$, $\forall k \in \mathbb{N}$, and $\lim_k \bar{\mu}_2^k / \|(1, \bar{\mu}^k)\|_\infty = 0$.

Note that the nature of these two sequences is distinct: the CAKKT sequence is not PAKKT since $\tilde{\mu}_1^k$ is the unique multiplier that tends to infinity and $\tilde{\mu}_1^k(\tilde{x}_1^k)^3(\tilde{x}_2^k)^3 < 0$ for all $k \in \mathbb{N}$. We will show that this behaviour occurs for any CAKKT sequence. That is, a CAKKT sequence can never be PAKKT.

Let $\{x^k\}$ be a CAKKT sequence with associated multiplier sequence $\{\mu^k\}$. Related to the third constraint, we have $\lim_k \mu_3^k(e^{x_1^k} \sin^2 x_2^k) = 0$. Now, from (8) and $\lim_k \nabla_x L(x^k, \mu^k) = 0$, it follows that $\lim_k 3\mu_1^k(x_1^k)^2(x_2^k)^3 = 1$. Thus, we can suppose without loss of generality that $x_2^k > 0$, $\forall k$. Moreover, as $\lim_k x^k = x^* = (0,0)$, we have

$$\mu_1^k \to \infty \quad \text{and} \quad 0 < \mu_1^k(x_2^k)^2 \to \infty. \quad (9)$$

We continue by proving that $\lim_k \mu_3^k / \mu_1^k = 0$ and $\lim_k \mu_2^k / \mu_1^k = 0$. For the first limit, we divide the first row of (8) by $\mu_1^k(x_2^k)^2$ and take the limit. Thus, using (9) and $\lim_k \nabla_x L(x^k, \mu^k) = 0$, we obtain

$$\frac{x_1^k - 1}{\mu_1^k(x_2^k)^2} + 3(x_1^k)^2 x_2^k - \frac{\mu_3^k}{\mu_1^k} e^{x_1^k} \left(\frac{\sin x_2^k}{x_2^k} \right)^2 \to 0.$$

Since $\lim_k x^k = (0,0)$, we conclude that $\lim_k \mu_3^k / \mu_1^k = 0$. Analogously, dividing the second row of (8) by μ_1^k, taking the limit and using (9), we obtain

$$\frac{x_2^k + 1}{\mu_1^k} + 3(x_1^k)^3(x_2^k)^2 + \frac{\mu_2^k}{\mu_1^k} + 2\frac{\mu_3^k}{\mu_1^k} e^{x_1^k} \sin x_2^k \cos x_2^k \to 0.$$

So, as $\lim_k \mu_3^k / \mu_1^k = 0$ and $\lim_k x^k = 0$, we get $\lim_k \mu_2^k / \mu_1^k = 0$.

Thus, since $\lim_k \mu_2^k / \mu_1^k = \lim_k \mu_3^k / \mu_1^k = 0$, condition (7) of the PAKKT definition does not take into account the multipliers $\{\mu_2^k\}$ and $\{\mu_3^k\}$. Furthermore, we have $\|(1, \mu^k)\|_\infty = \mu_1^k$ for all k sufficiently large, giving $\lim_k \mu_1^k / \|(1, \mu^k)\|_\infty = 1$. To see that $\{x^k\}$ is not a PAKKT sequence, note that $\mu_1^k(x_1^k)^3(x_2^k)^2 < 0$ for all k large enough, because otherwise, conditions $\mu^k \ge 0$ and $x_2^k > 0$, $\forall k$, would imply that the second row of (8) would be greater than 1 for all k large enough, contradicting $\lim_k \nabla_x L(x^k, \mu^k) = 0$. $\qquad \square$

The difference between considering points and sequences when dealing with PAKKT and CAKKT conditions, as illustrated by Example 1, motivates us to define a new sequential optimality condition. This new condition consists exactly in the existence of a sequence $\{x^k\}$ converging to x^* which is *simultaneously* CAKKT and PAKKT.

Definition 2. *We say that a feasible point x^* for* (NLP) *is a* Positive Complementary Approximate KKT *(PCAKKT) 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), (6) *and* (7) *hold. In this case, $\{x^k\}$ is called a* PCAKKT *sequence.*

From Definition 2, it is obvious that every PCAKKT point is a CAKKT one. Furthermore, it is also a PAKKT point since (6) implies (5). We stress that these implications are strict (see Figure 2). In particular, the origin in Example 1 is not a PCAKKT point. That is, the PCAKKT condition is more than the fulfillment of the CAKKT and PAKKT conditions simultaneously.

The next theorem says that PCAKKT is a legitimate necessary optimality condition. It can be proved using the external penalty theory, as in [12, Theorem 3.3] and [3, Theorem 2.2]. The adaptation of these results to our case is straightforward and therefore we omit it.

Theorem 1. *PCAKKT is a necessary optimality condition for* (NLP), *that is, every local minimizer of* (NLP) *is a PCAKKT point.*

It is known that a KKT point is PAKKT [3, Lemma 2.6]. Moreover, a KKT point x^* with multiplier vector (λ^*, μ^*) is trivially CAKKT taking the constant sequences $\{x^k := x^*\}$ and $\{(\lambda^k, \mu^k) := (\lambda^*, \mu^*)\}$. Next we show that, as expected, every KKT point is PCAKKT.

Theorem 2. *Every KKT point x^* of* (NLP) *is PCAKKT.*

Proof. We will show that there is a PCAKKT sequence associated with x^*. By the proof of [3, Lemma 2.6], we can find a PAKKT sequence $\{x^k\}$ converging to x^* such that the corresponding sequence of multipliers is bounded. The boundedness of the multipliers implies that $\{x^k\}$ is also a CAKKT sequence with the same sequence of multipliers. Thus, x^* is a PCAKKT point. $\qquad\square$

It is worth mentioning that PCAKKT sequences, since they are simultaneously a PAKKT and CAKKT sequence, inherit all good properties of the these conditions. For instance, we may highlight two of them already presented in the literature:

- PCAKKT sequences have bounded dual sequences under the quasinormality CQ [3, Theorem 4.7];

- PCAKKT sequences are sufficient for global optimality in convex problems [12, Theorem 4.2].

Another known sequential optimality condition is the *approximate gradient projection* (AGP) condition introduced in [28], which is useful in the convergence analysis of inexact restoration algorithms [13, 21, 27].
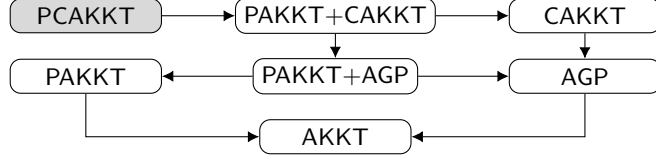
Figure 2: Landscape of sequential optimality conditions for constrained optimization. An arrow indicate a strict implication between two conditions.

Definition 3. *We say that a feasible x^* for* (NLP) *is an* AGP *point if there is a sequence $\{x^k\} \subset \mathbb{R}^n$ converging to x^* such that*

$$P_{\Omega(x^k)}(-\nabla f(x^k)) \to 0,$$

where $P_{\Omega(x^k)}$ is the orthogonal projection onto

$$\Omega(x^k) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla h_i(x^k)^T d = 0, & i = 1, \ldots, m \\ \min\{0, g_j(x^k)\} + \nabla g_j(x^k)^T d \le 0, & j \in I_g(x^*) \end{array} \right\}.$$

In this case, $\{x^k\}$ is called an AGP *sequence.*

Inspired by Example 1, we can establish the relationship between PCAKKT and "PAKKT + AGP" (saying that a point is simultaneously PAKKT and AGP makes sense because these conditions are independent to each other [3]). Indeed, as CAKKT implies AGP [12], Example 1 also shows that PCAKKT is stronger than "PAKKT+AGP". Figure 2 summarizes all the relations discussed here.

We finish this subsection with an alternative definition of the AGP condition, which will be useful for future discussions, especially for the convergence analysis of our primal-dual augmented Lagrangian method presented in Section 4. A similar statement was obtained in [4] for inequality constraints only.

Theorem 3. *Let x^* be a feasible point of* (NLP). *Then, the AGP condition holds at x^* iff there exist 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 $\lim_k \mu_j^k \min\{0, g_j(x^k)\} = 0$ for all $j \in I_g(x^*)$.*

Proof. Assume that 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 $\lim_k \mu_j \min\{0, g_j(x^k)\} = 0$, $j \in I_g(x^*)$. Define $d^k := P_{\Omega(x^k)}(-\nabla f(x^k))$, which is the unique solution of

$$\text{minimize} \quad \frac{1}{2}\| - \nabla f(x^k) - d\|^2 \quad \text{subject to} \quad d \in \Omega(x^k). \tag{10}$$

Since $0 \in \Omega(x^k)$, we have $\|\nabla f(x^k) + d^k\|^2 \le \|\nabla f(x^k)\|^2$, which implies $\|d^k\|^2 \le -2\nabla f(x^k)^T d^k$. On the other hand, multiplying the expression

$$\nabla_x L(x^k, \lambda^k, \mu^k) = \nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k$$

by d^k we obtain, since $d^k \in \Omega(x^k)$ and $\mu_j^k = 0$, $j \notin I_g(x^*)$,

$$\|d^k\|^2 \le -2(d^k)^T \nabla f(x^k) \le -2(d^k)^T \nabla_x L(x^k, \lambda^k, \mu^k) - 2 \sum_{j \in I_g(x^*)} \mu_j^k \min\{0, g_j(x^k)\}$$

9

for all k large enough. Taking the limit we have $d^k \to 0$, and thus AGP condition holds at x^*. The converse follows from the KKT conditions for the problem (10) at d^k (note that KKT conditions hold since d^k is a minimizer and $\Omega(x^k)$ is defined by linear constraints). $\qquad\square$

2.2 Strength of the new sequential optimality condition

In this subsection, we are interested in the sufficient assumptions that ensure that a PCAKKT point is actually a KKT one for every smooth problem (NLP). We refer to the least stringent of such assumptions by *strict constraint qualification* (SCQ). Note that, in view of Theorem 1, the SCQ associated with the sequential optimality condition PCAKKT is a constraint qualification. Inspired by the SCQs for AKKT, CAKKT and PAKKT, namely, AKKT-regular (also known as Cone Continuity Property – CCP) [10, 11], CAKKT-regular [11] and PAKKT-regular [3] respectively, we provide in the sequel the SCQ associated to PCAKKT, that we will call *PCAKKT-regular*.

First, note that the KKT conditions hold at the feasible point x^* if, and only if, $-\nabla f(x^*) \in K(x^*)$, where $K(x)$ is the convex closed cone defined as

$$K(x) := \{R(x, \lambda, \mu) \mid (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+, \ \mu_j = 0 \text{ for } j \notin I_g(x^*)\}$$

(x^* will be clear in the context), where

$$R(x, \lambda, \mu) := \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j=1}^p \mu_j \nabla g_j(x).$$

Now, we turn our attention to the SCQ for PCAKKT. Fixed x^*, we define for given $x \in \mathbb{R}^n$, $\alpha, \beta, \sigma \geq 0$, the set

$$K^{\mathrm{PC}}(x, \alpha, \beta, \sigma) := \{R(x, \lambda, \mu) \mid (\lambda, \mu) \in M(x, \alpha, \beta, \sigma)\},$$

where $M(x, \alpha, \beta, \sigma)$ is the set of all $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that $\mu_j = 0$ for all $j \notin I_g(x^*)$ and

$$\lambda_i h_i(x) \geq \alpha \quad \text{if} \quad |\lambda_i| > \beta\|(1, \lambda, \mu)\|_\infty, \tag{11a}$$

$$\mu_j g_j(x) \geq \alpha \quad \text{if} \quad \mu_j > \beta\|(1, \lambda, \mu)\|_\infty \text{ and } j \in I_g(x^*), \tag{11b}$$

$$\sum_{i=1}^m |\lambda_i h_i(x)| + \sum_{j \in I_g(x^*)} |\mu_j g_j(x)| \leq \sigma. \tag{11c}$$

The set K^{PC} mimics the shape of multipliers in the PCAKKT definition, and it is not difficult to see that $K^{\mathrm{PC}}(x^*, 0, 0, 0) = K(x^*)$. To define our PCAKKT-regular condition, consider the set

$$\limsup_{x \to x^*, \, \alpha \downarrow 0^+, \, \beta \downarrow 0, \, \sigma \downarrow 0} K^{\mathrm{PC}}(x, \alpha, \beta, \sigma) = \left\{ \omega \in \mathbb{R}^n \ \middle| \ \begin{array}{l} \exists (x^k, \omega^k) \to (x^*, \omega), \ \alpha_k \downarrow 0^+, \ \beta_k \downarrow 0, \ \sigma_k \downarrow 0 \\ \text{with } \omega^k \in K^{\mathrm{PC}}(x^k, \alpha_k, \beta_k, \sigma_k), \forall k \in \mathbb{N} \end{array} \right\}.$$

Definition 4. *A feasible point x^* for* (NLP) *satisfies the* PCAKKT-regular *condition if*

$$\limsup_{x \to x^*, \, \alpha \downarrow 0^+, \, \beta \downarrow 0, \, \sigma \downarrow 0} K^{PC}(x, \alpha, \beta, \sigma) \subset K^{PC}(x^*, 0, 0, 0).$$
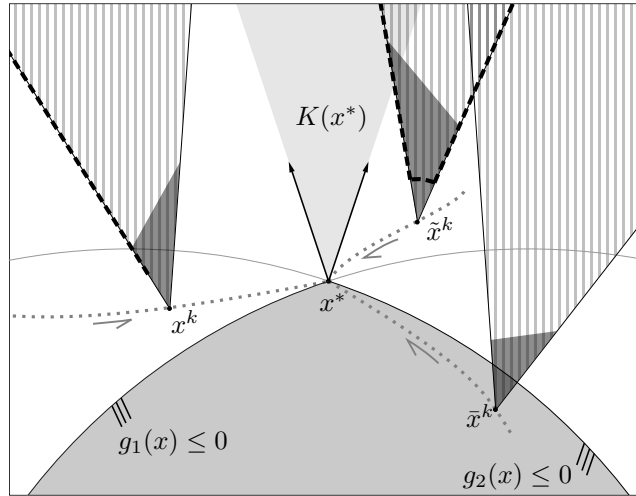
Figure 3: Geometric interpretation of PCAKKT-regularity. Consider the point $x^* = (0,0)$ and the constrained set given by the functions $g_1(x) = (x_1-2)^2+(x_2+6)^2-40 \leq 0$ and $g_2(x) = (x_1+2)^2+(x_2+6)^2-40 \leq 0$. Three different sequences converge to $x^* = (0,0)$: $\{x^k\}$, which violates the first constraint and satisfy the second; $\{\tilde{x}^k\}$, which violates both; and $\{\bar{x}^k\}$, which satisfies both. $K(\cdot)$, $K^P(\cdot,\alpha,\beta)$ and $K^C(\cdot,\sigma)$ are represented, respectively, by the regions filled with lines, by the regions delimited by strong dashed lines, and by the shaded areas (here $\beta = 0$ for simplicity); we have $K^{PC}(\cdot,\alpha,\beta,\sigma) \subset K^P(\cdot,\alpha,\beta) \cap K^C(\cdot,\sigma)$. All of these sets converge to the limit cone $K(x^*)$: K takes into account both gradients $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$, those of the active constraints at x^*; K^P considers only those gradients associated with violated constraints at intermediate points of the sequences (∇g_1 for $\{x^k\}$; ∇g_1 and ∇g_2 for $\{\tilde{x}^k\}$; and none for $\{\bar{x}^k\}$); and K^C is "truncated" in a fashion that depends on the magnitude of σ and how close the constraints are to zero. The set K^{PC} encompasses these two properties.

PCAKKT-regularity is inspired by the related conditions CAKKT-regular [11] and PAKKT-regular [3]. Each of them consists in an outer semicontinuity-like condition at x^*, just like that of Definition 4, of the sets

$$K^C(x,\sigma) := \{R(x,\lambda,\mu) \mid (11c),\ \mu_j = 0 \text{ for } j \notin I_g(x^*)\} \quad \text{and}$$

$$K^P(x,\alpha,\beta) := \{R(x,\lambda,\mu) \mid (11a),(11b),\ \mu_j = 0 \text{ for } j \notin I_g(x^*)\},$$

respectively. See [3, 11] for details. Note that $K^C(x^*,0) = K^P(x^*,0,0) = K(x^*)$ and $K^{PC}(x,\alpha,\beta,\sigma) \subset K^P(x,\alpha,\beta) \cap K^C(x,\sigma)$. Figure 3 gives a geometric view of these conditions.

Next we prove that PCAKKT-regular is the weakest SCQ for PCAKKT, in the sense that it is for PCAKKT just as Guignard's CQ is for KKT.

Theorem 4. *Every PCAKKT point that satisfies PCAKKT-regularity is KKT. Reciprocally, if a PCAKKT point x^* is also KKT, for every smooth objective function f, then x^* satisfies the PCAKKT-regularity condition.*

Proof. Let x^* be a PCAKKT point of (NLP) with the corresponding sequences

11

$\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$ and $\{\mu^k\} \subset \mathbb{R}^p_+$. Assume that $\mu^k_j = 0$, $\forall j \notin I_g(x^*)$. Thus, in view of (4), we have $\lim_k(\nabla f(x^k) + \omega^k) = 0$ where $\omega^k := R(x^k, \lambda^k, \mu^k)$. Let us show that $\omega^k \in K^{\mathrm{PC}}(x^k, \alpha_k, \beta_k, \sigma_k)$ for some α_k, β_k, $\sigma_k \geq 0$ converging to zero. Define $\delta_k := \|(1, \lambda^k, \mu^k)\|_\infty$, for $k \in \mathbb{N}$, and the sets $I_+ := \{i \in \{1, \ldots, m\} \mid \lim_k |\lambda^k_i|/\delta_k > 0\}$, $J_+ := \{j \in \{1, \ldots, p\} \mid \lim_k \mu^k_j/\delta_k > 0\}$ (these sets are well-defined and independent of k, after taking a subsequence if necessary). For all k, define

$$\alpha_k := \min\left\{1/k\,,\ \min_{i \in I_+}\{\lambda^k_i h_i(x^k)\}\,,\ \min_{j \in J_+}\{\mu^k_j g_j(x^k)\}\right\}, \tag{12}$$

$$\beta_k := \max\left\{1/k\,,\ \max_{i \notin I_+} |\lambda^k_i|/\delta_k\,,\ \max_{j \notin J_+} \mu^k_j/\delta_k\right\} \qquad \text{and} \qquad \sigma_k := c_k, \tag{13}$$

where c_k is as in (6). From conditions (6) and (7), we have $\alpha_k \downarrow 0^+$ and $\sigma_k \downarrow 0$. Furthermore, by the definitions of I_+ and J_+, we have $\beta_k \downarrow 0$. Given $k \in \mathbb{N}$, consider an index i such that $|\lambda^k_i| > \beta_k \delta_k$. Since $\delta_k > 0$, from (13) we obtain $|\lambda^k_i|/\delta_k > \beta_k \geq \max_{t \notin I_+} |\lambda^k_t|/\delta_k$ and thus $i \in I_+$. Using (12), we conclude that $\lambda^k_i h_i(x^k) \geq \alpha_k$, which means that (11a) is satisfied for these sequences. Analogously we show that such sequences also satisfy (11b). Since (11c) is immediate, we conclude that $\omega^k \in K^{\mathrm{PC}}(x^k, \alpha_k, \beta_k, \sigma_k)$. Therefore, using the hypothesis that x^* is PCAKKT-regular, we obtain

$$-\nabla f(x^*) = \lim_k \omega^k \in \limsup_{x \to x^*,\, \alpha \downarrow 0^+,\, \beta \downarrow 0,\, \sigma \downarrow 0} K^{\mathrm{PC}}(x, \alpha, \beta, \sigma) \subset K^{\mathrm{PC}}(x^*, 0, 0, 0).$$

But, since $K^{\mathrm{PC}}(x^*, 0, 0, 0) = K(x^*)$, we conclude that x^* is a KKT point.

Conversely, let x^* be a feasible point such that whenever x^* is PCAKKT point for some objective function then the KKT conditions hold at x^*. Here, we will show that PCAKKT-regularity holds at x^*. For this purpose, take $\omega \in \limsup_{x \to x^*,\, \alpha \downarrow 0^+,\, \beta \downarrow 0,\, \sigma \downarrow 0} K^{\mathrm{PC}}(x, \alpha, \beta, \sigma)$. Then, there exist sequences $x^k \to x^*$, $\omega^k \to \omega$, $\alpha_k \downarrow 0^+$, $\beta_k \downarrow 0$ and $\sigma_k \downarrow 0$ such that $\omega^k \in K^{\mathrm{PC}}(x^k, \alpha_k, \beta_k, \sigma_k)$. In turn, there are sequences $\{\lambda^k\} \subset \mathbb{R}^m$ and $\{\mu^k_j\} \subset \mathbb{R}_+$, $j \in I_g(x^*)$, such that $\omega^k = R(x^k, \lambda^k, \mu^k)$ and (11a)–(11c) hold. Define $\mu^k_j = 0$ for all $j \notin I_g(x^*)$ and $k \in \mathbb{N}$, and take $f(x) = -\omega^T x$. We claim that x^* is a PCAKKT point for this f. Indeed, we have immediately (4). Moreover, (6) follows from (11c) and the fact that $\mu^k_j = 0$ for $j \notin I_g(x^*)$. Finally, if $\lim_k |\lambda^k_i|/\delta_k > 0$, we have $|\lambda^k_i| > \beta_k \delta_k$ for all k sufficiently large. So, by (11a), we conclude that $\lambda^k_i h_i(x^k) \geq \alpha_k > 0$. The same reasoning shows that $\mu^k_i g_i(x^k) > 0$ if $\lim_k \mu^k_i/\delta_k > 0$. Therefore (7) holds, and x^* is PCAKKT. We then conclude that x^* is a KKT point and thus $\omega = -\nabla f(x^*) \in K(x^*) = K^{\mathrm{PC}}(x^*, 0, 0, 0)$, which implies that the PCAKKT-regular condition holds at x^*. $\qquad \square$

By Theorems 1 and 4, it follows that every local minimizer of (NLP) that is PCAKKT-regular is actually a KKT point, that is, we have the following result.

Corollary 1. *The PCAKKT-regular condition is a constraint qualification.*

By [11, Theorem 2], CAKKT-regularity holds at x^* iff for every continuously differentiable objective function for which x^* is CAKKT, we have that the KKT conditions hold at x^*. Since PCAKKT implies CAKKT, and using the characterization given by Theorem 4, we conclude that CAKKT-regularity implies PCAKKT-regularity. Using a similar reasoning and [3, Theorem 2.4], we also

have that PAKKT-regularity implies PCAKKT-regularity. These implications are strict, since the sequential optimality condition PCAKKT strictly implies each of conditions CAKKT and PAKKT. See Figure 4.

To complete the landscape of CQs known in the literature, we will show that PCAKKT-regularity is stronger than the Abadie's CQ. We say that the Abadie's CQ holds at a feasible x^* if the tangent cone to the feasible set $\mathcal{F}$ of (NLP) at x^* given by

$$\mathcal{T}(x^*) := \{d \in \mathbb{R}^n \mid \text{ there exist } t_k \downarrow 0, \ d_k \to d \text{ with } x^* + t_k d_k \in \mathcal{F}, k \in \mathbb{N}\},$$

coincides with its linearization cone

$$\mathcal{L}(x^*) := \{d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0, \forall i, \ \nabla g_j(x^*)^T d \leq 0, j \in I_g(x^*)\}.$$

Theorem 5. *PCAKKT-regularity implies Abadie's CQ.*

Proof. The statement can be obtained by similar arguments of the proof of [11, Theorem 6], which uses [11, Lemma 2]. We note that the proof of this lemma provides multipliers defined by $\lambda^k = kh(x^k)$ and $\mu^k = kg(x^k)_+$ (they also appear in [10, Lemma 4.3]). Thus, λ^k and μ^k have the same sign of their corresponding constraints, and furthermore, $\mu_j^k = 0$ for all k large enough whenever $j \notin I_g(x^*)$, and $\omega^k \in K^{\text{PC}}(x^k, \alpha_k, \beta_k, \sigma_k)$ for the sequences $\alpha_k \downarrow 0^+$ and $\beta_k \downarrow 0$ defined by (12) and (13). Note that we can suppose that $\alpha_k > 0$ for all k since it is possible to extract a subsequence of $\{x^k\}$ so that $\lambda_i^k = kh_i(x^k) \neq 0, \forall i \in I_+$ and $\mu_j^k = kg_j(x^k)_+ > 0, \forall j \in J_+$. $\qquad \square$

The implication in Theorem 5 is strict, as the next example shows.

Example 2 (Abadie's CQ does not imply PCAKKT-regularity)**.** *Consider the point $x^* = (0, 0)$, and the inequality constraints of Example 4 of [3]*

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

It was shown in [3] that Abadie's CQ holds at x^*, and that $K(x^*) = \mathbb{R}_- \times \mathbb{R}$. To see that x^* is not PCAKKT-regular, consider $\omega^* := (1, 0) \notin K(x^*)$. For all $k \geq 1$, define the sequences $x^k := (-1/k, 0)$, $\mu^k := (k/4, k/4, k^3, k^3, 0)$, $\alpha_k := 1/k^2$, $\beta_k := 1/k$, $\sigma_k := \sum_{j \in I_g(x^*)} |\mu_j^k g_j(x^k)| = (1/2k) + (2/k^2)$, and $\omega^k := \sum_{j=1}^5 \mu_j^k \nabla g_j(x^k)$. Straightforward calculations show that $\omega^k \in K^{PC}(x^k, \sigma_k, \alpha_k, \beta_k)$ for all k; $\alpha_k, \beta_k, \sigma_k \to 0$ and $\omega^k \to \omega^*$. Thus, PCAKKT-regularity fails at x^*.

Figure 4 shows some relations between several CQs in the literature. Note the unifying role of the PCAKKT-regular condition. For other CQs considered in the figure, see [11] and references therein. Since PCAKKT-regularity is less stringent than P/CAKKT-regularity, we can use it to establish an algorithm with better convergence properties than others from the literature, by proving that such an algorithm generates PCAKKT points. We dedicate the rest of the paper to the study of algorithms with this property.
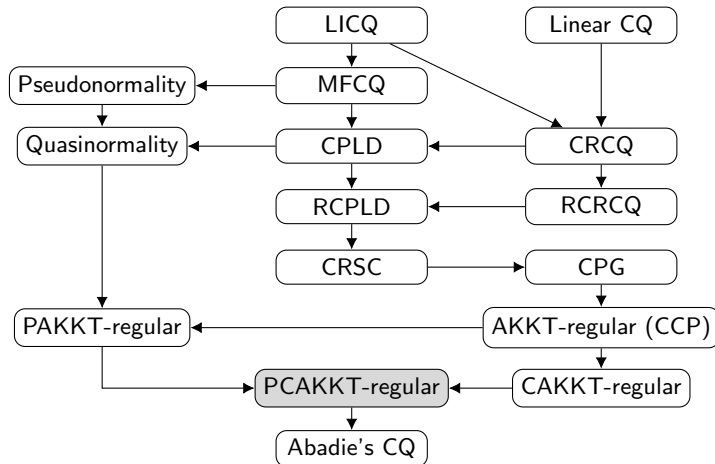
Figure 4: Relations between the several CQs in the literature. An arrow indicate a logical strict implication between two CQs.

3 Convergence of the safeguarded PHR AL method using PCAKKT

In recent years, the global convergence analysis of AL methods has been dramatically improving by the use of sequential optimality conditions and weak CQs, see [1, 2, 3, 12, 17] and references therein. Here, we show that the new sequential optimality condition can be useful in the global convergence of the augmented Lagrangian method proposed in [2] (see Algorithm 1 below). As done in [12], we use the following generalization of the Lojasiewicz inequality: we say that a continuously differentiable function $\Phi : \mathbb{R}^n \to \mathbb{R}$ satisfies the *generalized Lojasiewicz* (GL) inequality at x^* if there is an open neighbourhood $\mathcal{B}(x^*)$ and a continuous function $\varphi : \mathcal{B}(x^*) \to \mathbb{R}$ such that $\lim_{x \to x^*} \varphi(x) = 0$ and, for all $x \in \mathcal{B}(x^*)$, $|\Phi(x) - \Phi(x^*)| \leq \varphi(x)\|\nabla\Phi(x)\|$. This condition roughly says that, in the case of $\nabla\Phi(x^*) = 0$, the functional value $\Phi(x)$ approaches $\Phi(x^*)$ faster than its gradient vanishes when x converges to x^*. It is worth mentioning that the GL condition is a generalization of the Lojasiewicz inequality, which corresponds to choosing $\varphi(x) = c|\Phi(x) - \Phi(x^*)|^{1-\theta}$ for certain constants $c > 0$ and $\theta \in (0, 1)$. For further discussion and examples, see [12, 18] and references therein.

Now, we will proceed to analyze Algorithm 1 below for solving (NLP). It makes use of the PHR augmented Lagrangian function

$$L_{\rho,\bar{\lambda},\bar{\mu}}^{\mathrm{PHR}}(x) := f(x) + \frac{\rho}{2} \left\| h(x) + \frac{\bar{\lambda}}{\rho} \right\|^2 + \frac{\rho}{2} \left\| \left(g(x) + \frac{\bar{\mu}}{\rho} \right)_+ \right\|^2. \qquad (14)$$

Theorem 6. *Let x^* be a feasible accumulation point of the sequence $\{x^k\}$ generated by Algorithm 1. Suppose that the "measure of infeasibility"*

$$\Phi(x) := \|h(x)\|^2 + \|g(x)_+\|^2$$

satisfies the GL inequality at x^*. Then, x^* is a PCAKKT point, and therefore a KKT one under the PCAKKT-regularity condition.

14

Algorithm 1 (Safeguarded) PHR augmented Lagrangian method

Let $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $\rho_1 > 0$, $\tau \in (0,1)$. Let $\{\varepsilon_k\} \subset \mathbb{R}_+$ be a sequence of positive scalars with $\lim \varepsilon_k = 0$. Choose $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\bar{\mu}^1 \in [0, \mu_{\max}]^p$. Initialize with $k \leftarrow 1$.

Step 1 (Solving the subproblems). Find an approximate minimizer x^k of $L_{\rho_k, \bar{\lambda}^k, \bar{\mu}^k}^{\mathrm{PHR}}(x)$, i.e., compute a point x^k satisfying $\|\nabla_x L_{\rho_k, \bar{\lambda}^k, \bar{\mu}^k}^{\mathrm{PHR}}(x^k)\|_\infty \leq \varepsilon_k$.

Step 2 (Update the penalty parameter). Define

$$V_k := \max\{ \|h(x^k)\|_\infty , \| \min\{-g(x^k), \bar{\mu}^k/\rho_k\}\|_\infty \}.$$

If $k > 1$ and $V_k \leq \tau V_{k-1}$, set $\rho_{k+1} := \rho_k$. Otherwise, take $\rho_{k+1} \geq \gamma \rho_k$.

Step 3 (Estimate new projected multipliers). Choose $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^{k+1} \in [0, \mu_{\max}]^p$, $k \leftarrow k + 1$ and go to Step 1.

Proof. The proof follows from the results in [3, 12]. Let $\{\lambda^k\}$ and $\{\mu^k\}$ be the associated dual sequences generated by Algorithm 1. When $\{(\lambda^k, \mu^k)\}$ has a bounded subsequence, we may take a subsequence such that λ^k and μ^k converge to λ and μ, respectively. Thus, x^* is a KKT point, and hence x^* is a PCAKKT by Theorem 2.

Now, suppose that $\{(\lambda^k, \mu^k)\}$ does not have a bounded subsequence. From [3, Theorem 4.1], there is a subsequence $\{x^k\}_{k \in K}$ which is a PAKKT sequence. Furthermore, by the proof of [12, Theorem 5.1], we can extract a further subsequence of $\{(x^k, \lambda^k, \mu^k)\}_{k \in K}$ conforming the definition of CAKKT. Thus, the final subsequence satisfies the requirements of the PCAKKT condition. $\qquad \square$

Remark. *Algorithm 1 resembles the external penalty method when the penalty parameter goes to infinity since its subproblems use the bounded multipliers estimates computed in Step 3. However, when we are able to choose the projected multipliers $\bar{\lambda}^{k+1}$ and $\bar{\mu}^{k+1}$ as the real estimates given by gradient of (14), $\bar{\lambda}^k + \rho_k h(x^k)$ and $[\bar{\mu}^k + \rho_k g(x^k)]_+$, respectively; Algorithm 1 behaves like the classical augmented Lagrangian algorithm. Therefore, to avoid truncating the multipliers, in practice, it is common to project the multiplier estimates into a large bounded set. For instance, the projected multipliers can be taken as $\bar{\lambda}^{k+1} = P_{[\lambda_{min}, \lambda_{max}]^m}(\bar{\lambda}^k + \rho_k h(x^k))$ and $\bar{\mu}^{k+1} = P_{[0,\mu_{max}]^p}([\bar{\mu}^k + \rho_k g_j(x^k)]_+)$, where $\lambda_{min} = -10^{30}$ and $\lambda_{max} = \mu_{max} = 10^{30}$. Thus, from the practical point of view, safeguards are not a limitation, on the contrary, they can be beneficial [26]. It is worth noting that this strategy and parameters for updating projected multipliers in Step 3 are used in our tests. See Section 5 for details. We leave the projected multipliers in Step 3 of Algorithm 1 free because the theory presented here covers any choice.*

Theorem 6 provides the strongest result about the global convergence of the AL method that we are aware of. Furthermore, Example 1 says that Theorem 6 implies more than the mere fulfillment of the PAKKT and CAKKT conditions simultaneously. Thus, Theorem 6 improves and unifies the convergence results of [3] and [12], under the GL assumption. Anyway, PCAKKT unifies two

branches of the sequential optimality conditions concerning the convergence of Algorithm 1, one related to the approximate fulfillment of the (enhanced) Fritz-John condition (i.e., PAKKT), and the other related to the approximate fulfillment of the KKT conditions (CAKKT). See Figure 2.

The fulfillment of the GL inequality of the infeasibility measure $\Phi(x)$ is a very general property and it is satisfied for a broad family of mappings which encompass analytic and semi-algebraic functions, see [18] and references therein. Besides the applicability of the assumptions and due to the possibility of the (P)CAKKT condition avoids undesirable non-minimizers, it is natural to ask for general-purpose methods for solving (NLP) with such convergence properties without imposing the GL inequality. Following this line of research, an interesting new method with good properties is presented in [23]. The method consists of the minimization of a shifted primal-dual penalty-barrier merit function, and their subproblems are solving by an interesting modification of Newton's method. The convergence analysis is done by means of a sequential optimality condition, that the authors also called complementary AKKT (CAKKT). However, although the proposed optimality condition is inspired by the CAKKT definition originally presented in [12] (see Definition 1), it is weaker than that. Indeed, to fit the formulation to the barrier method considered, the authors consider the problem (NLP) with only inequality constraints, which, after inserting slack variables, takes the form

$$\text{minimize}_{x,s} \; f(x) \;\; \text{subject to} \;\; g(x) = s, \;\; s \leq 0.$$

Therefore, they define their sequential optimality condition using this problem in the following way: a feasible point (x^*, s^*) satisfies the CAKKT condition (in the sense of [23, Definition 4.1]) if there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{s^k\} \subset \mathbb{R}^p$, $\{\mu^k\} \subset \mathbb{R}^p$ and $\{z^k\} \subset \mathbb{R}^p_+$ such that $x^k \to x^*$, $s^k \to s^* = g(x^*)$,

$$\nabla f(x^k) + \nabla g(x^k)\mu^k \to 0, \quad \mu^k - z^k \to 0 \quad \text{and} \quad z_j^k s_j^k \to 0, \; \forall j \in I_g(x^*). \quad (15)$$

In the sequel, we show that not only (15) is weaker than the original CAKKT condition (Definition 1) for problem (NLP) with inequality constraints only, but it is actually strictly weaker than the less stringent AGP condition (Definition 3). See Figure 2.

Indeed, if x^* is an AGP point for (NLP) with only inequality constraints, by Theorem 3 there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\mu^k\} \subset \mathbb{R}^p_+$ such that $x^k \to x^*$, $\lim_k \nabla_x L(x^k, \mu^k) = 0$ and $\lim_k \mu_j^k \min\{0, g_j(x^k)\} = 0$. Then, (15) holds by choosing $z^k := \mu^k \geq 0$ and $s^k := \min\{0, g(x^k)\}$. That is, AGP implies (15). Secondly, this implication is strict, as the next example shows.

Example 3. *Consider the bidimensional problem*

$$minimize \;\; \frac{1}{2}(x_2 - 2)^2 \quad subject \; to \;\; -x_1 \leq 0, \;\; x_1 x_2 \leq 0.$$

We affirm that $x^* = (0, 1)$ is not an AGP point. Otherwise, by Theorem 3, there would be sequences $\{x^k\} \subset \mathbb{R}^2$ and $\{\mu^k\} \subset \mathbb{R}^2_+$ such that $x^k \to (0, 1)$, $\mu_1^k \min\{0, -x_1^k\} \to 0$, $\mu_2^k \min\{0, x_1^k x_2^k\} \to 0$ and

$$\nabla_x L(x^k, \mu^k) = \begin{bmatrix} 0 \\ x_2^k - 2 \end{bmatrix} + \mu_1^k \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \mu_2^k \begin{bmatrix} x_2^k \\ x_1^k \end{bmatrix} \to 0. \quad (16)$$

In this case, as $x_2^k \to 1$, we have $\mu_2^k x_1^k \to 1$, which in turn implies $\mu_2^k x_1^k x_2^k \to 1$ and $x_1^k > 0$ for all k large enough. Now, multiplying the first row of (16) by x_1^k and taking the limit, we get $\mu_1^k x_1^k \to 1$, which contradicts $\mu_1^k \min\{0, -x_1^k\} \to 0$.

To prove that (15) holds at x^*, consider the sequences $\bar{x}^k := (-1/k, 1)$, $\bar{z}^k := \bar{\mu}^k := (k, k) \geq 0$, $\bar{s}^k := (1/k^2, 1/k^2)$. Clearly, $\bar{x}^k \to x^* = (0, 1)$ and $\bar{s}^k \to s^* = (0, 0)$. Finally, it is straightforward to verify that, for each k, $\nabla_x L(\bar{x}^k, \bar{\mu}^k) = 0$, $\bar{\mu}^k - \bar{z}^k = 0$ and $\bar{z}_j^k \bar{s}_j^k = 1/k \to 0, \forall j$. $\qquad\square$

Supported by the PCAKKT condition, we propose in the next section a new method based on the augmented Lagrangian function (3).

4   A new shifted primal-dual method

In this section, we present our new augmented Lagrangian method and its convergence properties. We consider the penalty-like augmented Lagrangian function (3) that carries the complementarity, bringing it to the minimization phase of the algorithm. Deriving (3), we obtain

$$\nabla_x L_{\rho,\nu,\bar{\lambda},\bar{\mu}}(x, \lambda^a, \mu^a) = \nabla f(x) + \nabla h(x)\lambda + \nabla g(x)\mu; \qquad (17)$$

$$\nabla_{\lambda^a} L_{\rho,\nu,\bar{\lambda},\bar{\mu}}(x, \lambda^a, \mu^a) = \nu[\lambda^a * h(x)] * h(x) - \left[h(x) + \frac{\bar{\lambda} - \lambda^a}{\rho}\right];$$

$$\nabla_{\mu^a} L_{\rho,\nu,\bar{\lambda},\bar{\mu}}(x, \lambda^a, \mu^a) = \nu[\mu^a * g(x)]_+ * g(x) - \left[g(x) + \frac{\bar{\mu} - \mu^a}{\rho}\right]_+ + \frac{\mu^a}{\rho}.$$

where the associated Lagrange multipliers in (17) are given by

$$\lambda = [\rho h(x) + \bar{\lambda}] + [\rho h(x) + \bar{\lambda} - \lambda^a] + \nu[\lambda^a * h(x)] * \lambda^a,$$

$$\mu = [\rho g(x) + \bar{\mu}]_+ + [\rho g(x) + \bar{\mu} - \mu^a]_+ + \nu[\mu^a * g(x)]_+ * \mu^a \geq 0.$$

We present our method in Algorithm 2. The kth iteration consists of finding an approximate solution (x, λ^a, μ^a) for the problem of minimizing $L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x, \lambda^a, \mu^a)$ subject to $\mu^a \geq 0$. It is straightforward to verify that the KKT conditions can be written using the so-called *projected gradient* $\mathcal{G}_P$, given by

$$\mathcal{G}_P^k(x, \lambda^a, \mu^a) := P_\Omega\left((x, \lambda^a, \mu^a) - \nabla L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x, \mu^a)\right) - (x, \lambda^a, \mu^a),$$

where $P_\Omega(z)$ is the orthogonal projection of z onto $\Omega := \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p$. Indeed, the KKT conditions can be written as $\mathcal{G}_P^k(x, \lambda^a, \mu^a) = 0$, that is,

$$\|\nabla_x L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x, \lambda^a, \mu^a)\|_\infty = 0, \quad \|\nabla_{\lambda^a} L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x, \lambda^a, \mu^a)\|_\infty = 0 \quad \text{and}$$

$$\left\|\left[\mu^a - \nabla_{\mu^a} L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x, \lambda^a, \mu^a)\right]_+ - \mu^a\right\|_\infty = 0.$$

Let us highlight some aspects of Algorithm 2:

- Differently from other augmented Lagrangian methods (for instance, that of Section 3), in Step 1 we compute a primal-dual pair instead of only x^k. On the other hand, the (bounded) estimate multipliers used in safeguarded methods are present. Although there is no guarantee that $\{\lambda^{a,k}\}$

Algorithm 2 Primal-dual augmented Lagrangian method

Let $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $\tau, a, \theta \in (0,1)$, $\{M_k\} \subset \mathbb{R}_+$ be a bounded sequence and $\{\varepsilon_k\} \subset \mathbb{R}_+$ be a sequence such that $\lim_{k\to\infty} \varepsilon_k = 0$.
Take $\hat{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\hat{\mu}^1 \in [0, \mu_{\max}]^p$, $\tau_1 > 0$, $\rho_1 > 0$, $\nu_1 > 1$, $\zeta_1^{max} \in (0,1)$. Initialize with $k \leftarrow 1$, $\bar{\lambda}^1 := \hat{\lambda}^1/\nu_1$ and $\bar{\mu}^1 := \hat{\mu}^1/\nu_1$.

Step 1 (Solving the subproblems). Find an approximate minimizer $(x^k, \lambda^{a,k}, \mu^{a,k})$ of $L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(\cdot)$, satisfying $\mu^{a,k} \geq 0$ and

$$\|\nabla_x L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k, \lambda^{a,k}, \mu^{a,k})\|_\infty \leq \varepsilon_k, \tag{18a}$$

$$\|\nabla_{\lambda^a} L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k, \lambda^{a,k}, \mu^{a,k})\|_\infty \leq \frac{M_k}{\rho_k \nu_k}, \tag{18b}$$

$$\left\| \left[\mu^{a,k} - \nabla_{\mu^a} L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k, \lambda^{a,k}, \mu^{a,k})\right]_+ - \mu^{a,k} \right\|_\infty \leq \frac{M_k}{\rho_k \nu_k}. \tag{18c}$$

Step 2 (Update penalty parameters). Define

$$V_k := \max\{ \|h(x^k)\|_\infty, \|\min\{-g(x^k), \bar{\mu}^k/\rho_k\}\|_\infty \},$$
$$C_k := \max\{ \|\lambda^{a,k} * h(x^k)\|_\infty, \|[\mu^{a,k} * g(x^k)]_+\|_\infty \}, \text{ and}$$
$$\zeta_k := \max\{ V_k, C_k \}.$$

If $k > 1$ and $\zeta_k \leq \min\{a/\nu_k, \zeta_k^{max}\}$ then set $(\rho_{k+1}, \nu_{k+1}) := (\rho_k, \nu_k)$, choose $\zeta_{k+1}^{max} \leq \theta \zeta_k^{max}$ and go to Step 3. Otherwise, set $\zeta_{k+1}^{max} := \zeta_k^{max}$ and

(i) if $V_k \leq \tau V_{k-1}$, set $\rho_{k+1} := \rho_k$. Otherwise, choose $\rho_{k+1} \geq \gamma \rho_k$;

(ii) if $C_k \leq \tau C_{k-1}$, set $\nu_{k+1} := \nu_k$. Otherwise, choose $\nu_{k+1} \geq \nu_k + a$.

Step 3 (Estimate new projected multipliers). Choose $\hat{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\hat{\mu}^{k+1} \in [0, \mu_{\max}]^p$. Set $\bar{\lambda}^{k+1} := \hat{\lambda}^{k+1}/\nu_{k+1}$, $\bar{\mu}^{k+1} := \hat{\mu}^{k+1}/\nu_{k+1}$. Take $k \leftarrow k+1$ and go to Step 1.

---

or $\{\mu^{a,k}\}$ are bounded, the regularization terms in $L_{\rho, \nu, \bar{\lambda}, \bar{\mu}}(\cdot)$, as well as (18b) and (18c), tend to control the growth of these sequences. We also observe that requirements similar to (18b) and (18c) were used in [22] in the context of stabilized SQP methods. In this sense, Algorithm 2 combines these two different strategies;

- Condition (18) can be theoretically achieved by any box-constrained minimization algorithm, since $\Omega = \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p$ is a box (of course, even if x is in a box, the resulting constraints, after adding $\mu^{a,k} \geq 0$, are still a box). One of them is the active-set strategy with spectral gradients known as GENCAN [15]. GENCAN is used in the PHR AL method ALGENCAN [2, 17], which has a mature and robust implementation provided by the TANGO project (www.ime.usp.br/~egbirgin/tango). We use GENCAN/ALGENCAN codes in our implementations and numerical tests (Section 5);

- We update the parameters ρ and ν according to the behaviour of the feasibility and the complementarity measures given by V_k and C_k respectively.

Each parameter is increased to emphasize the respective measure in the subsequent iteration. Note that the increment rule for ρ is more aggressive than that for ν. That reflects a preference for feasibility over complementarity in the algorithm. It is worth noting that V_k and the update rule for ρ are the same as the PHR AL method (see Algorithm 1). Furthermore, ν remains unchanged if the complementarity measure C_k has sufficiently decreased. In this sense, Algorithm 2 tries to mimic the behaviour of the PHR augmented Lagrangian method when the CAKKT-like complementarity measure decrease adequately;

- The term ζ_k can be interpreted as a measure of the feasibility and the fulfillment of the complementary term. Furthermore, if in some iteration we have $\|\nabla_x L_{\rho_k, \nu_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k, \lambda^{a,k}, \mu^{a,k})\|_\infty = 0$ and $\zeta_k = 0$, then x^k is a KKT point for (NLP);

- In theory, the choice of $\hat{\lambda}^{k+1}$ and $\hat{\mu}^{k+1}$ in Step 3 is free. However, following Remark 3, a practical choice is to project the multipliers estimates given by the gradient of the augmented Lagrangian function (3), obtained after solving (18), onto the boxes $[\lambda_{\min}, \lambda_{\max}]^m$ and $[0, \mu_{\max}]^p$. We use this strategy in our numerical tests (Section 5).

4.1 Global convergence analysis

In this subsection, we present the convergence results for Algorithm 2. To establish convergence to PCAKKT points, we deal separately with the generation of PAKKT and CAKKT sequences. Thus, once we have established these two results, we can state our main convergence result.

By (17) and (18a), the Lagrange multipliers computed by Algorithm 2 are

$$\lambda^k = [\rho_k h(x^k) + \bar{\lambda}^k] + [\rho_k h(x^k) + \bar{\lambda}^k - \lambda^{a,k}] + \nu_k[\lambda^{a,k} * h(x^k)] * \lambda^{a,k}, \quad (19a)$$

$$\mu^k = [\rho_k g(x^k) + \bar{\mu}^k]_+ + [\rho_k g(x^k) + \bar{\mu}^k - \mu^{a,k}]_+ + \nu_k[\mu^{a,k} * g(x^k)]_+ * \mu^{a,k}, \quad (19b)$$

where $(x^k, \lambda^{a,k}, \mu^{a,k})$ is the current iterate. From now on, λ^k and μ^k will always refer to (19). Also, for the sake of simplicity, we proceed supposing that

Assumption A: x^* is an accumulation point of the sequence $\{x^k\}$ generated by Algorithm 2. In this case, we assume that $\lim_{k \in K} x^k = x^*$, where $K \subset \mathbb{N}$.

4.1.1 Auxiliary results

For simplicity, during this subsection we will write $h_i^k := h_i(x^k)$ and $g_j^k := g_j(x^k)$. Given $j \in I_g(x^*)$, we split K into the following disjoint subsets:

$$K_1 = \{k \in K \mid \rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k} < 0, \quad g_j^k < 0\}; \quad (20a)$$

$$K_2 = \{k \in K \mid \rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k} \geq 0, \quad g_j^k < 0\}; \quad (20b)$$

$$K_3 = \{k \in K \mid \rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k} < 0, \quad g_j^k \geq 0\}; \quad (20c)$$

$$K_4 = \{k \in K \mid \rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k} \geq 0, \quad g_j^k \geq 0\}. \quad (20d)$$

These subsets will be useful for subsequent analysis.

The next result says that the estimate (19b) for the Lagrange multiplier vector associated with inequality constraints has a null component whenever the correspondent constraint is inactive at the limit point x^*. That is, in this case Algorithm 2 computes correctly the final multiplier (if it exists), and the complementarity related to inactive constraints is satisfied exactly. The same property is verified in the PHR augmented Lagrangian method (Algorithm 1) [17, Theorem 4.1].

Lemma 1. *If $g_j(x^*) < 0$ then $\mu_j^k = 0$ for all $k \in K$ sufficiently large (here, x^* is not necessarily feasible).*

Proof. If $\{\rho_k\}$ is unbounded then, by the boundedness of $\{\bar{\mu}^k\}$, $\rho_k g_j^k + \bar{\mu}_j^k \le 0$ for all $k \in K$ large enough; for these k's, we have $\mu_j^k = 0$ since $\mu_j^{a,k} \ge 0$ and $g_j^k \le 0$. If $\{\rho_k\}$ is bounded then, by Step 2, $\lim_k V_k = 0$ and thus $\lim_{k \in K} \bar{\mu}^k = 0$. As $g_j^k \le g_j(x^*)/2 < 0$ for all k large, (19b) implies that $\mu_j^k = 0$ for these indexes k. $\qquad\square$

The first convergence result by means of a sequential optimality condition is stated in Lemma 3 below. As the safeguarded PHR augmented Lagrangian method (Algorithm 1), see [3], our new algorithm generates PAKKT sequences whenever the multipliers estimates are unbounded. In view of Example 1, it is important to guarantee PAKKT *sequences* in order to state the convergence to PCAKKT points, our main objective. The case where multipliers estimates form a bounded sequence (or at least have a bounded subsequence) is trivial, since in this case x^* is actually a KKT point, and thus PAKKT [3, Lemma 2.6]. So, this case is left to the main and more general result involving the PCAKKT condition.

Before we relate Algorithm 2 to PAKKT sequences, we need the following auxiliary technical result.

Lemma 2. *For all k and $i = 1, \ldots, m$, and $j = 1, \ldots, p$, we have*

(a) $\left| \lambda_i^{a,k} - \frac{\bar{\lambda}_i^k + \rho_k h_i^k}{1 + \nu_k \rho_k (h_i^k)^2} \right| \le \frac{M_k}{\nu_k (1 + \nu_k \rho_k (h_i^k)^2)}$;

(b) $\left| \left[\mu_j^{a,k} + \left(g_j^k + \frac{\bar{\mu}_j^k - \mu_j^{a,k}}{\rho_k} \right)_+ - \frac{\mu_j^{a,k}}{\rho_k} - \nu_k \mu_j^{a,k} (g_j^k)_+^2 \right]_+ - \mu_j^{a,k} \right| \le \frac{M_k}{\rho_k \nu_k}$;

(c) $\{\nu_k \lambda_i^{a,k} h_i^k\}_{k \in K}$ *and* $\{\nu_k \mu_g^{a,k} [g_j^k]_+\}_{k \in K}$ *are bounded.*

Proof. Items a and b follow directly from (18b) and (18c). To prove item c, we multiply item a by $\nu_k |h_i^k|$, obtaining

$$\left| \nu_k \lambda_i^{a,k} |h_i^k| - \frac{\nu_k \bar{\lambda}_i^k |h_i^k| + \nu_k \rho_k h_i^k |h_i^k|}{1 + \nu_k \rho_k (h_i^k)^2} \right| \le \frac{M_k |h_i^k|}{1 + \nu_k \rho_k (h_i^k)^2}. \qquad (21)$$

From the boundedness of $\{M_k\}$, the right hand side of the above inequality remains bounded. Also, from the fact that $\nu_k \bar{\lambda}_i^k = \hat{\lambda}_i^k$ remains on a compact set, $\{\nu_k \bar{\lambda}_i^k |h_i^k|\}_{k \in K}$ is bounded. Using (21) and triangle inequality, we have

$$\nu_k |\lambda_i^{a,k} h_i^k| \le \frac{\nu_k \bar{\lambda}_i^k |h_i^k|}{1 + \nu_k \rho_k (h_i^k)^2} + \frac{\nu_k \rho_k (h_i^k)^2}{1 + \nu_k \rho_k (h_i^k)^2} + \frac{M_k |h_i^k|}{1 + \nu_k \rho_k (h_i^k)^2}.$$

From the last expression, we see that $\{\nu_k \lambda_i^{a,k} h_i^k\}_{k \in K}$ is a bounded sequence. Now we treat item c for the inequality case. If $j \notin I_g(x^*)$, the result is trivial. Suppose that $j \in I_g(x^*)$ and split the set K into disjoint sets as in (20). By the definition of K_1 and K_2, the sequence $\{\nu_k \mu_g^{a,k} [g_j^k]_+ = 0\}_{k \in K_1 \cup K_2}$ is trivially bounded. For all $k \in K_3$, item b takes the form

$$\left| \left[\mu_j^{a,k} - \frac{\mu_j^{a,k}}{\rho_k} - \nu_k \mu_j^{a,k} g_j(x_k)^2 \right]_+ - \mu_j^{a,k} \right| \leq \frac{M_k}{\rho_k \nu_k}.$$

If the expression between brackets are non-positive then $\mu_j^{a,k} \leq M_k / [\rho_k \nu_k]$; and if it is positive then $\mu_j^{a,k} \leq M_k / [\nu_k(1 + \nu_k \rho_k (g_j^k)^2)]$. Multiplying both previous inequalities by $\nu_k g_j^k \geq 0$, we have that $\{\nu_k \mu_j^{a,k} g_j^k\}_{k \in K_3}$ is bounded. Finally, if $k \in K_4$, we multiply item b by $\rho_k / (1 + \nu_k \rho_k (g_j^k)^2)$ to obtain an analogous expression to item a. We then proceed as the equality case, multiplying it by $\nu_k g_j^k \geq 0$ and passing the limit over K_4. This concludes the proof. $\qquad \square$

Lemma 3. *Suppose that x^* is feasible, and that the sequence of Lagrange multipliers estimates $\{(\lambda^k, \mu^k)\}_{k \in K}$ is unbounded. Then $\{x^k\}_{k \in K}$ admits a PAKKT subsequence.*

Proof. By Step 2, condition (4) of the definition of PAKKT is naturally satisfied. Since the sequence $\{\delta_k := \|(1, \lambda^k, \mu^k)\|_\infty\}_{k \in K}$ is unbounded, we may assume, after taking a subsequence if necessary, that $\delta_k \to \infty$ and the bounded sequences $\{\lambda^k / \delta_k\}_{k \in K}$ and $\{\mu^k / \delta_k\}_{k \in K}$ converge.

Let us consider the case where $\lim_{k \in K} |\lambda_i^k| / \delta_k > 0$ for a given index i. From Lemma 2, $\delta_k \to \infty$ and the boundedness of $\{M_k\}$, we obtain

$$\lim_{k \in K} \left[\frac{\lambda_i^{a,k}}{\delta_k} - \frac{\rho_k h_i^k}{\delta_k (1 + \nu_k \rho_k (h_i^k)^2)} \right] = 0. \tag{22}$$

Then, by (19a), (22) and the boundedness of $\{\bar{\lambda}^k\}$,

$$0 \neq \lim_{k \in K} \frac{\lambda_i^k}{\delta_k} = \lim_{k \in K} \frac{h_i^k}{\delta_k} \left[\frac{\rho_k + 2\nu_k \rho_k^2 (h_i^k)^2}{1 + \nu_k \rho_k (h_i^k)^2} + \nu_k (\lambda^{a,k})^2 \right].$$

The expression between the brackets are positive for all $k \in K$. Thus, λ_i^k and h_i^k have the same sign. That is, $\lambda_i^k h_i^k > 0$ for all $k \in K$, as required by the PAKKT definition (see Definition 1).

Now, suppose that $\lim_{k \in K} \mu_j^k / \delta_k > 0$, that is,

$$\lim_{k \in K} \left[(\rho_k g_j^k + \bar{\mu}_j^k)_+ + \nu_k (\mu_j^{a,k} g_j^k)_+ \mu_j^{a,k} + (\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k})_+ \right] / \delta_k > 0.$$

Thus, at least one of the three terms in the above sum is bounded below by a positive scalar for all $k \in K$ large enough. If that hold for any of the two first terms, we trivially have $g_j^k > 0$ for all $k \in K$ large enough (remember that $\mu_j^{a,k} \geq 0$ for all k). Suppose now that the mentioned property occurs for the third term, that is,

$$(\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k})_+ / \delta_k \geq c, \ \forall k \in K \text{ large enough, for some } c > 0. \tag{23}$$

In this case, the set $K_2 \cup K_4$ is infinite (see (20)), which enable us to consider from now on, taking a subsequence if necessary, all indexes k in this set. If K_4 is finite then $k \in K_2$ for all k large enough, which implies the boundedness of $\{[\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k}]_+\}$, contradicting (23) (recall that $\delta_k \to \infty$). We then conclude that K_4 is infinite, which in turn allow us to assume that $k \in K_4$, $\forall k$. Therefore, $g_j^k \geq 0$, $\forall k$. If $g_j^k = 0$ for infinitely many indexes k, we would have $c \leq \rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k} = \bar{\mu}_j^k - \mu_j^{a,k}$, which implies the boundedness of $\{\mu_j^{a,k}\}$. But this contradicts (23) since $\delta_k \to \infty$. Thus, $g_j^k > 0$ for all k large enough.

Repeating the above argument for all indexes i and j, and taking successive subsequences, we achieve a PAKKT subsequence as we want. $\qquad \square$

Now we turn our attention to the generation of CAKKT sequences by Algorithm 2. The next auxiliary result states that Algorithm 2 asymptotically fulfils the CAKKT-like complementarity on subsequences over K_1 or K_2.

Lemma 4. *Assume that x^* is feasible. Then $\lim_{k \in K_1 \cup K_2} \mu_j^k g_j^k = 0$ for all $j \in I_g(x^*)$.*

Proof. Consider the partition of K given by (20). In the sequel, we assume implicit that each set K_ℓ is infinite whenever a limit is considered (otherwise there is nothing to do). From (19b), recall that

$$\mu_j^k = [\rho_k g_j^k + \bar{\mu}_j^k]_+ + [\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k}]_+ + \nu_k (\mu_j^{a,k})^2 [g_j^k]_+.$$

Let us analyze the proper limit in each subset.

Subsequences over K_1. As $g_j^k < 0$ for $k \in K_1$, we have $[\rho_k g_j^k + \bar{\mu}_j^k]_+ \leq \bar{\mu}_j^k$, and thus $\{[\rho_k g_j^k + \bar{\mu}_j^k]_+\}_{k \in K_1}$ is bounded. From $\lim_{k \in K_1} g_j^k = 0$, we obtain $\mu_j^k g_j^k = [\rho_k g_j^k + \bar{\mu}_j^k]_+ g_j^k \to_{K_1} 0$.

Subsequences over K_2. As the above case, $\{[\rho_k g_j^k + \bar{\mu}_j^k]_+\}_{k \in K_2}$ is bounded. From $\mu_j^{a,k} \geq 0$ we have $[\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k}]_+ \leq [\rho_k g_j^k + \bar{\mu}_j^k]_+$. Thus, $\mu_j^k g_j^k = [\rho_k g_j^k + \bar{\mu}_j^k]_+ g_j^k + [\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k}]_+ g_j^k \to_{K_2} 0$.

Therefore we conclude that $\lim_{k \in K_1 \cup K_2} \mu_j^k g_j^k = 0$. $\qquad \square$

Let us define the set

$$\underline{K} := \left\{ k \in K \ \middle| \ \zeta_k \leq \min\left\{ \frac{a}{\nu_k}, \zeta_k^{max} \right\} \right\}. \tag{24}$$

Note that $\underline{K} \subset K$ and, in view of Assumption A, x^* is the *unique* limit point of $\{x^k\}_{k \in \underline{K}}$ whenever $\underline{K}$ is infinite. The set $\underline{K}$ is related to the successful iterates of Step 2 of Algorithm 2, for which both parameters ρ and ν remain unchanged.

In the following two lemmas, we analyze the generation of CAKKT sequences by Algorithm 2.

Lemma 5. *If $\underline{K}$ is infinite, then $\{x^k\}_{k \in \underline{K}}$ is a CAKKT sequence.*

Proof. Since $\underline{K}$ is infinite, Step 2 of Algorithm 2 implies that $\lim_k \zeta_k^{max} = 0$ and hence $\lim_{k \in \underline{K}} V_k = \lim_{k \in \underline{K}} C_k = 0$. Furthermore, from the definition of V_k, the point x^* is feasible for (NLP). To obtain the desired CAKKT sequence, firstly observe that condition (4) is naturally satisfied with the approximate multipliers λ^k and μ^k defined by (19). Then it remains to prove condition (6).

We start by proving that $\lim_{k \in \underline{K}} \lambda_i^k h_i^k = 0$ *for every* $i = 1, \ldots, m$. Take $i \in \{1, \ldots, m\}$. Then, we have $|\lambda^{a,k} h_i^k| \leq C_k \to_{\underline{K}} 0$, and from $C_k \leq \zeta_k \leq \min\{a/\nu_k, \zeta_k^{max}\}$, we also have $|\nu_k \lambda^{a,k} h_i^k| \leq \nu_k C_k \leq a$ for all $k \in \underline{K}$. Thus $\lim_{k \in \underline{K}} \nu_k (\lambda^{a,k} h_i^k)^2 = 0$. From (21) and the fact that $\{\nu_k \bar{\lambda}_i^k = \hat{\lambda}_i^k\}$ is in a compact set, we have $\limsup_{k \in \underline{K}} \frac{\nu_k \rho_k (h_i^k)^2}{1 + \nu_k \rho_k (h_i^k)^2} \leq a < 1$. Hence $\{\nu_k \rho_k (h_i^k)^2\}_{k \in \underline{K}}$ is bounded, which implies $\lim_{k \in \underline{K}} \rho_k (h_i^k)^2 = 0$. In summary, we have $\lim_{k \in \underline{K}} \rho_k (h_i^k)^2 = \lim_{k \in \underline{K}} \lambda_i^{a,k} h_i^k = \lim_{k \in \underline{K}} \nu_k (\lambda_i^{a,k} h_i^k)^2 = 0$, which, by (19a), imply $\lambda_i^k h_i^k \to_{k \in \underline{K}} 0$. That is, the approximate CAKKT-like complementary holds for equality constraints.

Now we proceed by showing that $\lim_{k \in \underline{K}} \mu_j^k g_j^k = 0$ *for all* $j = 1, \ldots, p$. Fix an index $j \in \{1, \ldots, p\}$. If $g_j(x^*) < 0$, then Lemma 1 ensures that $\mu_j^k = 0$ for all k large enough, and $\lim_{k \in \underline{K}} \mu_j^k g_j^k = 0$ trivially holds. Thus, assume that $g_j(x^*) = 0$ and split the set K into the four disjoint sets K_1, K_2, K_3 and K_4 as (20). This induces the partition of $\underline{K}$ into the sets $\underline{K}_\ell := K_\ell \cap \underline{K}$. In the sequel, we suppose implicitly that each of these $\underline{K}_\ell$ is infinite whenever a limit is considered (otherwise there is nothing to do), and then we will prove that $\lim_{k \in \underline{K}_\ell} \mu_j^k g_j^k = 0$, $\ell = 1, \ldots, 4$. From Lemma 4, $\lim_{k \in K_1 \cup K_2} \mu_j^k g_j^k = 0$, so we only need to analyze the sequences over $\underline{K}_3$ and $\underline{K}_4$.

From (19b), recall that

$$\mu_j^k = [\rho_k g_j^k + \bar{\mu}_j^k]_+ + [\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k}]_+ + \nu_k (\mu_j^{a,k})^2 [g_j^k]_+.$$

So, as $C_k \leq \min\{a/\nu_k, \zeta_k^{max}\}$ for every $k \in \underline{K}$,

$$\mu_j^{a,k}[g_j^k]_+ \to_{\underline{K}} 0, \quad |\nu_k \mu_j^{a,k}[g_j^k]_+| \leq a \quad \text{and} \quad \nu_k (\mu_j^{a,k}[g_j^k]_+)^2 \to_{\underline{K}} 0. \tag{25}$$

○ *Subsequences over* $\underline{K}_3$. Multiplying the inequality $\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k} < 0$ by $g_j^k \geq 0$ and using the first limit in (25), we get $\lim_{k \in \underline{K}_3} \rho_k (g_j^k)^2 = 0$ and therefore, using the last limit in (25), $\mu_j^k g_j^k = [\rho_k g_j^k + \bar{\mu}_j^k]_+ g_j^k + \nu_k (\mu_j^{a,k})^2 [g_j^k]_+^2 \to_{k \in \underline{K}_3} 0$.

○ *Subsequences over* $\underline{K}_4$. For every $k \in \underline{K}_4$, we have

$$\mu_j^k = [\rho_k g_j^k + \bar{\mu}_j^k] + [\rho_k g_j^k + \bar{\mu}_j^k - \mu_j^{a,k}] + \nu_k (\mu_j^{a,k})^2 g_j^k.$$

Note this μ_j^k has the same shape of the Lagrange multiplier estimate λ_i^k for equality constraints, since all their terms are nonnegative (see (19)). Thus, using similar arguments to the equality case, and having in mind item b of Lemma 2, the result is valid for μ_j^k and $k \in \underline{K}_4$.

Finally, we observe that all the arguments are valid for all $k \in \underline{K}$ sufficiently large, and hence $\{x^k\}_{k \in \underline{K}}$ is a CAKKT sequence, concluding the proof. $\qquad \square$

Lemma 6. *Suppose that* x^* *is feasible and* $\underline{K}$ *is finite. If the nondecreasing sequence* $\{\min\{\rho_k, \nu_k\}\}$ *is bounded then* $\{x^k\}_{k \in K}$ *is a CAKKT sequence.*

Proof. It is sufficient to show that $\lim_{k \in K} \lambda_i^k h_i^k = 0$, $\forall i$, and $\lim_{k \in K} \mu_j^k g_j^k = 0$, $\forall j$. We will only prove the statement for equality constraints; for inequalities with $j \in I_g(x^*)$ the proof is similar, and for those where $j \notin I_g(x^*)$, the result follows from Lemma 1.

Take an index $i \in \{1, \dots, m\}$. Let us recall that, from (19a),

$$\lambda_i^k h_i^k = 2\rho_k (h_i^k)^2 + 2\bar{\lambda}_i^k h_i^k - \lambda_i^{a,k} h_i^k + \nu_k (\lambda_i^{a,k} h_i^k)^2.$$

When $\underline{K}$ is finite, the sequence $\{\zeta_k^{max}\}$ is updated only by a finite number of steps, which implies $\zeta_k^{max} = \zeta^{max}$ for all k sufficiently large. Now, if $\{\max\{\rho_k, \nu_k\}\}$ is bounded, say, by $A > 0$, then $\zeta_k > \min\{a/A, \zeta^{max}\} > 0$ for every k large enough. On the other hand, the boundedness of $\{\max\{\rho_k, \nu_k\}\}$ and the Step 2 of Algorithm 2 imply that $\lim_{k \in K} V_k = \lim_{k \in K} C_k = 0$. Hence $\lim_{k \in K} \zeta_k = 0$, which leads us to a contradiction. Thus, $\max\{\rho_k, \nu_k\} \to \infty$. With this, and in view of the hypotheses, it is enough to consider the following two cases:

○ $\{\rho_k\}$ *bounded and* $\{\nu_k\}$ *unbounded.* Clearly, $\lim_{k \in K} \rho_k (h_i^k)^2 = 0$. From the boundedness of $\{\nu_k \lambda_i^{a,k} h_i^k\}_{k \in K}$ (Lemma 2, item c), we conclude that $\lim_{k \in K} \lambda_i^{a,k} h_i^k = \lim_{k \in K} \nu_k (\lambda_i^{a,k} h_i^k)^2 = 0$. As a consequence, $\lim_{k \in K} \lambda_i^k h_i^k = 0$.

○ $\{\rho_k\}$ *unbounded and* $\{\nu_k\}$ *bounded.* From Step 2 of Algorithm 2, $\lim_{k \in K} \lambda_i^{a,k} h_i^k = 0$ and hence $\lim_{k \in K} \nu_k \lambda_i^{a,k} h_i^k = 0$. Together with (21), we get $\lim_{k \in K} \nu_k \rho_k (h_i^k)^2 = 0$ and so $\lim_{k \in K} \rho_k (h_i^k)^2 = 0$. Therefore, we have $\lim_{k \in K} \lambda_i^k h_i^k = 0$. □

4.1.2 Main convergence results

Next we present the main convergence result for Algorithm 2.

Theorem 7. *We have the following:*

(a) *If the set $\underline{K}$ defined in (24) is infinite then every accumulation point x^* of $\{x^k\}_{k \in \underline{K}}$ is a PCAKKT point. Thus, if additionally PCAKKT-regularity holds at x^*, then x^* is a KKT point of* (NLP).

(b) *If $\underline{K}$ is finite then every accumulation point of $\{x^k\}_{k \in K}$ is*

    (b1) *a PCAKKT point, whenever $\{\min\{\rho_k, \nu_k\}\}$ is bounded. In this case, x^* is a KKT point if it conforms to the PCAKKT-regular condition;*

    (b2) *at least a PAKKT and AGP point simultaneously, in the case that $\{\min\{\rho_k, \nu_k\}\}$ is unbounded.*

Proof. First, note that if x^* is feasible and the sequence of Lagrange multipliers estimates $\{(\lambda^k, \mu^k)\}_{k \in K}$ is bounded (or at least has a bounded subsequence), then by Step 1 of Algorithm 2, the point x^* satisfies the KKT conditions. Then all items follows from Theorem 2 and the implications of Figure 2. Thus, suppose then that $\{(\lambda^k, \mu^k)\}_{k \in K}$ is an unbounded sequence.

Item a: Applying Lemma 5, we have that $\{x^k\}_{k \in \underline{K}}$ is a CAKKT sequence. Then, applying Lemma 3 on such sequence we conclude that x^* is PCAKKT. The second statement follows from Theorem 4.

Item b1: Follows from Lemmas 3 and 6, and Theorem 4.

Item b2: By Lemma 3, we get that x^* is a PAKKT point. To show that x^* is an AGP point, it is enough to show, in view of Theorem 3, that $\lim_{k \in K} \mu_j^k \min\{0, g_j(x^k)\} = 0$, $j \in I_g(x^*)$. This follows from Lemma 4, since $\min\{0, g_j(x^k)\} = 0$ for all $k \in K_3 \cup K_4$. □

From Theorem 7, we see that Algorithm 2 can reach KKT points under the PCAKKT-regular condition. This is a strong convergence result for an implementable algorithm obtained by means of a sequential optimality condition. As other ones, the PCAKKT condition is independent of a specific algorithm, and then it allows us to idealize other algorithms with the same convergence status. We stress that Algorithm 2 and possibly others, whenever they generate PCAKKT points, enjoy the good properties of the P/CAKKT sequences (such as sufficiency for global optima under convexity and the boundedness of Lagrange multipliers under quasinormality – see the discussion after Theorem 2). With respect to Theorem 7, the case (b2) is very pathological in the sense that it occurs only when the first test in Step 2 fails for all k sufficiently large ($\underline{K}$ finite), and both parameters ρ and ν go to infinity ($\min\{\rho_k, \nu_k\} \to \infty$). Even in this case, we are able to prove convergence to "PAKKT+AGP" points. Although this is weaker than the PCAKKT concept (see Example 1), the accumulation point x^* is a KKT under one of the mild CQs PAKKT-regular, defined in [3] (see discussion after Definition 4), or AGP-regular, defined in [11]. So, to the best of our knowledge, Theorem 7 is the strongest result for an augmented Lagrangian strategy.

Finally, we show that Algorithm 2 always reaches stationary points of the infeasibility problem

$$\min_x \Phi(x) = \|h(x)\|^2 + \|g(x)_+\|^2. \tag{26}$$

In this sense, x^* is the point with "minimal infeasibility". This is a desirable property, specially when we deal with infeasible problems.

Theorem 8. *The point x^* is KKT for (26).*

Proof. If x^* is feasible for (NLP) there is nothing to do. Suppose that x^* is not feasible. In this case, $\underline{K}$ is finite and $\rho_k \to \infty$. From (18a) we obtain

$$\frac{\nabla f(x^k)}{\rho_k} + \nabla h(x^k) \left[\frac{\lambda^k}{\rho_k}\right] + \nabla g(x^k) \left[\frac{\mu^k}{\rho_k}\right] \to 0. \tag{27}$$

We will analyze the asymptotic behaviour of $\{\lambda^k/\rho_k\}_{k \in K}$ and $\{\mu^k/\rho_k\}_{k \in K}$.

Sequence $\{\lambda^k/\rho_k\}_{k \in K}$**.** Take $i \in \{1, \ldots, m\}$. Dividing the expression in item a of Lemma 2 by ρ_k, we get, for all $k \in K$,

$$\left| \frac{\lambda_i^{a,k}}{\rho_k} - \frac{\bar{\lambda}_i^k}{\rho_k(1 + \nu_k \rho_k h_i(x^k)^2)} - \frac{h_i(x^k)}{1 + \nu_k \rho_k h_i(x^k)^2} \right| \leq \frac{M_k}{\nu_k \rho_k(1 + \nu_k \rho_k h_i(x^k)^2)}.$$

From the boundedness of $\{M_k\}$ and $\{\bar{\lambda}_i^k\}$, and from $\rho_k \to \infty$, the second term within modulus and the right hand side of the inequality tend to zero. The third term within the modulus also tends to zero independently if $h_i(x^k)$ vanishes or not. Thus $\lim_{k \in K} \lambda_i^{a,k}/\rho_k = 0$.

From Lemma 2, item c, $\{\nu_k \lambda_i^{a,k} |h_i(x^k)|\}_{k \in K}$ is a bounded sequence independently if $h_i(x^k)$ vanishes or not. So, dividing (19a) by ρ_k and using the boundedness of $\{\bar{\lambda}_i^k\}_{k \in K}$, we obtain

$$\frac{\lambda_i^k}{\rho_k} = 2 \left[h_i(x^k) + \frac{\bar{\lambda}_i^k}{\rho_k}\right] - \frac{\lambda_i^{a,k}}{\rho_k} + \frac{\lambda_i^{a,k}}{\rho_k}(\nu_k \lambda_i^{a,k} h_i(x^k)) \to_K 2h_i(x^*).$$

Sequence $\{\mu^k/\rho_k\}_{k\in K}$**.** We will show in an analogous way that $\lim_{k\in K}\mu^k/\rho_k = 2[g(x^*)]_+$. Take $j\in\{1,\ldots,p\}$. Dividing (19b) by ρ_k we obtain, for all $k\in K$,

$$\frac{\mu_j^k}{\rho_k} = \left[g_j(x^k)+\frac{\bar\mu_j^k}{\rho_k}\right]_+ + \left[g_j(x^k)+\frac{\bar\mu_j^k}{\rho_k}-\frac{\mu_j^{a,k}}{\rho_k}\right]_+ + \frac{\mu_j^{a,k}}{\rho_k}\nu_k(\mu_j^{a,k}g_j(x^k))_+. \quad (28)$$

Now, split the set K into four disjoint sets K_1, K_2, K_3 and K_4 as (20). We will show that for each of these subsets, $\lim_{k\in K_\ell}\mu_j^k/\rho_k = 2[g_j(x^*)]_+$ whenever K_ℓ is an infinite subset of K, $\ell = 1,\ldots,4$. Thus, without loss of generality, we assume that each of these subsets is infinite.

Subsequence over K_1. By (28) and the boundedness of $\{\bar\mu_j^k\}$, we have $\mu_j^k/\rho_k = [g_j(x^k)+\bar\mu_j^k/\rho_k]_+ \to_{K_1} [g_j(x^*)]_+ = 0 = 2[g_j(x^*)]_+$.

Subsequence over K_2. Here, $\{[\rho_k g_j(x^k)+\bar\mu_j^k-\mu_j^{a,k}]_+\}_{k\in K_2}$ is bounded. Thus, from (28) and the boundedness of $\{\bar\mu_j^k\}$,

$$\frac{\mu_j^k}{\rho_k} = \left[g_j(x^k)+\frac{\bar\mu_j^k}{\rho_k}\right]_+ + \frac{[\rho_k g_j(x^k)+\bar\mu_j^k-\mu_j^{a,k}]_+}{\rho_k} \to_{K_2} 0 = 2[g_j(x^*)]_+.$$

Subsequence over K_3. From Lemma 2, item c, $\{\nu_k\mu_j^{a,k}g_j(x^k)\}_{k\in K_3}$ is bounded independently if $g_j(x^k)$ vanishes or not. Now, observe that taking the limit over K_3 in the inequality $g_j(x^k)+\bar\mu_j^k/\rho_k-\mu_j^{a,k}/\rho_k\le 0$, obtained from K_3, we get $g_j(x^*) = 0$. Thus, using (28) we have

$$\frac{\mu_j^k}{\rho_k} = \left[g_j(x^k)+\frac{\bar\mu_j^k}{\rho_k}\right]_+ + \frac{\mu_j^{a,k}}{\rho_k}(\nu_k\mu_j^{a,k}g_j(x^k)) \to_{K_3} 0 = 2[g_j(x^*)]_+.$$

Subsequence over K_4. In this case, μ_j^k/ρ_k has the same shape of λ_i^k/ρ_k. Thus we conclude that $\lim_{k\in K_4}\mu_j^k/\rho_k = 2[g_j(x^*)]_+$ in an analogous way that for equality constraints (here, in fact, maybe $g_j(x^*)\ne 0$).

Finally, expression (27) together the previous cases imply $\nabla h(x^*)h(x^*) + \nabla g(x^*)[g(x^*)]_+ = 0$, which says that x^* is a KKT point of (26). $\qquad\square$

5   Numerical experience

The tests were run on an Intel(R) Xeon(R) Silver 4114 CPU 2.20GHz, under the Ubuntu 18.04.4 operating system. We implemented Algorithm 2 in Fortran 90, adapting the ALGENCAN 3.1.1 package provided freely by the TANGO project. We compiled all the code with GNU Fortran 7.5.0 using the "O3" flag. ALGENCAN 3.1.1 is an implementation of Algorithm 1 with some improvements made over time (see [17] and references therein). One of these improvements is an acceleration process which consists of switching to a Newtonian strategy at the final (outer) steps of the minimization process. But as we consider AL strategies, we have disabled this feature.

Our aim here is not to compare Algorithms 1 and 2 to each other. Instead, we see Algorithm 2 as a complement to its classical counterpart, a strategy that tries to overcome difficulties of Algorithm 1. We then consider a hybridization of the

methods, based on Algorithm 1, where a primal-dual iteration of Algorithm 2 is applied when it is needed to force the fulfillment of the complementarity condition. This is reasonable since (i) ALGENCAN performs very well on a variety of test-problems [2, 14, 16]; and (ii) the subproblems of Algorithm 2 are (probably) more difficult to handle numerically than those of Algorithm 1, since they involve minimizing the augmented Lagrangian (3) in both primal and dual variables. Thus, solving these subproblems efficiently maybe requires specialized algorithms.

We adopt the next rules to switch to a primal-dual iteration of Algorithm 2:

1. A primal-dual iteration is applied if the stopping criterion of Algorithm 1 for success was fulfilled, but CAKKT complementarity seems to be not satisfied. Specifically, we decide to apply a primal-dual iteration if

$$\|\nabla_x L^{\mathrm{PHR}}_{\rho_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k)\|_\infty \leq \varepsilon_{\mathrm{opt}}, \quad V_k \leq \varepsilon_{\mathrm{opt}},$$
$$\max\{\|\lambda^k * h(x^k)\|_\infty, \|\mu^k * g(x^k)\|_\infty\} > \sqrt{\varepsilon_{\mathrm{opt}}}, \tag{29}$$

where $\varepsilon_{\mathrm{opt}}$ is the ALGENCAN's tolerance for optimality. This criterion aims to force CAKKT complementarity, since Algorithm 2 is, theoretically, more likely to achieve it than Algorithm 1;

2. Analogously to the previous item, we switch to a primal-dual iteration when a stationary point of the infeasibility (problem (26)) was achieved;

3. When a primal-dual iteration is performed and passes the first test of Step 2, that is, when $\zeta_k \leq \min\{a/\nu_k, \zeta_k^{max}\}$, the next iteration is also primal-dual. Therefore, we maintain the minimization on the primal and dual variables whenever both penalty parameters ρ and ν remain unchanged. Remember that, by item a of Theorem 7, such situation is related to the convergence to PCAKKT points;

4. A primal-dual iteration is chosen when $\rho \geq 10^5$. This criterion is applied only once;

5. At every iteration, we compute the relative displacement of the primal iterate $\Delta x^k = \|x^{k-1} - x^k\|_\infty / \max\{1, \|x^k\|_\infty\}$. When $\Delta x^k \leq \varepsilon_k^{1/4}$, we consider that the primal iterate does not move substantially from iteration $k-1$ to k. If this happens during consecutive iterations $k, k+1, \ldots, k+p$, we allow strategies 1 and 2 to be applied only once throughout these iterations. On the other hand, the chance of applying these strategies are renewed whenever $\Delta x^k > \varepsilon_k^{1/4}$. In particular, if $\Delta x^k \leq \varepsilon_k^{1/4}$ and the iteration k is primal-dual, a new primal-dual iteration is prohibited at the next iteration $k+1$. Thus we allow a primal-dual iteration only when the method has progressed since the last use of Algorithm 2.

After a primal-dual iteration, we go back to PHR iterations of Algorithm 1 if none of the above situations are verified. Furthermore, in a primal-dual iteration we execute the test ii of Step 2 (of Algorithm 2) only if $V_k \leq \sqrt{\varepsilon_{\mathrm{opt}}}$. That is, the penalty parameter ν can only be increased after a "sufficient" fulfillment of the (AKKT-type) complementarity $\min\{-g(x^k), \bar{\mu}^k/\rho_k\} \approx 0$. Note that this expression is used in the PHR augmented Lagrangian method (Algorithm 1) to attest approximate complementarity, which is enough, by Theorem 6, to

ensures PCAKKT points under the GL inequality hypothesis. Thus, we can expect CAKKT complementarity frequently. Then, our strategy aims to give measure V_k a chance to achieve CAKKT points before to increment ν. The reader may note that our convergence theory for Algorithm 2 remains valid with this modification, specifically Lemma 5.

Another issue is how to initialize $\lambda^{a,k}$ and $\mu^{a,k} \geq 0$ when a primal-dual iteration is set immediately after an iteration of Algorithm 1. In this case, we compute $\lambda^{a,k}$ and $\mu^{a,k} \geq 0$ so that (19) equals to their first terms $\rho_k h(x^k) + \bar{\lambda}^k$ and $[\rho_k g(x^k) + \bar{\lambda}^k]_+$. The reason is that the PHR augmented Lagrangian function (14) gives these multipliers estimates, and then we try to take advantage of the minimization process already done. If it is not possible to compute such $\lambda_i^{a,k}$ and $\mu_j^{a,k} \geq 0$, we set them as zero.

For Algorithm 2, we set $a = 0.99$, $\theta = 0.1$, $M_k \equiv 10^3$, $\nu_1 = 1.0$ and $\zeta_1^{\max} = a/\nu_1$. All other parameters are initialized as ALGENCAN's default values (in particular, $\lambda_{\min} = -10^{30}$ and $\lambda_{\max} = \mu_{\max} = 10^{30}$, see Remark 3). We consider 241 constrained test problems from Hock & Schittkowski' and Maros & Meszaros' libraries, both available from CUTEst. In 29 of them (12.03%), at least one primal-dual iteration was employed. Of the total problems considered, ALGENCAN (Algorithm 1) did not declare convergence in 25 (10.37%), and the hybrid strategy was capable of recovering optimality in 3 of them (12%), thus declaring convergence. In all other problems, the hybrid strategy had the same result of ALGENCAN (primal-dual iterations never were applied or they were not able to induce the success of the minimization process as a whole). Table 1 presents the problems with different result between the two strategies. Columns "Problem", "it", "st", "feas" and "cpl (29)" mean, respectively, the problem name, final number of outer iterations (/number of primal-dual iterations), status (0 = success; 1 = converges to stationary point of infeasibility; 2 = stops with huge $\max\{\rho, \nu\}$; 3 = the maximum number of iterations was achieved), final sup-norm violation of constraints and the final CAKKT-type complementarity measure like in (29). The other columns "f", "$\|\nabla L\|$", "$\|\nabla L^{\mathrm{PHR}}\|$" and "$V_k$" contain the final value of each quantity.

Hybrid strategy (Algorithm 1+Algorithm 2)

Problem	st	it	f	$\|\nabla L\|$	feas	V_k	cpl (29)
HS56	0	7/1	-3,46e+00	4,89e-08	8,07e-07	1,22e-07	1,16e-06
QE226	0	29/1	2,13e+02	4,46e-12	2,31e-07	2,31e-07	2,04e-07
QSHARE1B	0	42/1	7,20e+05	1,49e-07	2,56e-09	9,09e-12	2,56e-11

ALGENCAN (Algorithm 1)

Problem	st	it	f	$\|\nabla L^{\mathrm{PHR}}\|$	feas	V_k
HS56	2	27	-1,57e+00	3,62e+18	1,71e-02	2,73e-03
QE226	3	50	3,45e+02	4,14e+00	2,89e+01	2,60e-01
QSHARE1B	3	50	7,20e+05	4,55e-02	1,59e-05	1,83e-08

Table 1: Computational results. In the table, we present test problems where ALGENCAN fails and where iterations of Algorithm 2 recover optimality.

We highlight some observed aspects. First, few primal-dual iterations were

required. In fact, we can expect this since, as we already mentioned, (i) the overall behaviour of Algencan is good and, of course, (ii) we establish the rules 1–5 above in order to apply a primal dual iteration only when Algencan seems to fail or when it converges to a "poor" point. The second aspect is related to the CAKKT-type complementarity achievement. We observe that, although neither Algorithms 2 nor 1 explicitly requires this type of complementarity with "real" multipliers estimates (see (19)), it was achieved frequently by the hybrid strategy (in particular, in the problems of Table 1). It is interesting to observe that a primal-dual iteration, even when applied in an intermediate stage of the minimization process, reduces the CAKKT-like complementarity measure (29) substantially.

Finally, we stress that our proposal, at least from the practical point of view, can be viewed as a strategy to improve the effectiveness of augmented Lagrangian methods, especially when the complementarity is considered important. In fact, the quality of the primal solution obtained by Algencan is often good. In this sense, an important issue is the amount of additional computational cost that the primal-dual iterations bring. In our tests, we limit the number of inner-iterations (those performed by Gencan) in a primal-dual iteration to the maximum of necessary iterations for solving PHR subproblems so far. We compare computational times of the hybrid strategy against the standard Algencan in the following way: for each problem P, we take the arithmetic means $T_{\text{hybrid}}(P)$ and $T_{\text{Algencan}}(P)$ of the times, on the runs required to obtain a minimum of 10 seconds; we then take the geometric mean of $T_{\text{hybrid}}(P)/T_{\text{Algencan}}(P)$ over all problems. This provides a factor that globally measures the execution time of the hybrid algorithm in relation to Algencan. The average increase in computing time was only 3% of the hybrid strategy compared to Algencan. That is, primal-dual iterations can be useful to improve convergence without spending much more time. Nevertheless, as we already mentioned in the introduction, we believe that this additional time can be reduced even more if a specialized inner-solver for Algorithm 2' subproblems is developed, instead of using Gencan purely. In fact, when we look only at the problems in which primal-dual was used, the increase in time was about 25%.

6 Conclusions

Recent progress has been made on the convergence of the (safeguarded) Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian method (Algorithm 1) by means of the so-called sequential optimality conditions: the positive approximate KKT (PAKKT) [3] and the complementary KKT (CAKKT) [12]. Each of them describes a property of the sequences generated by the method. The first deals with the sign between constraints and dual variables, a property related to the enhanced Fritz-John optimality conditions; and the second deals with the way that KKT complementarity is (sequentially) achieved. We proposed a new sequential optimality condition that unify these characteristics, called *positive complementary approximate KKT* (PCAKKT). By Example 1, the PCAKKT condition is not the simple fulfillment of PAKKT and CAKKT conditions. Thus, the PCAKKT condition really leads to an improvement in the convergence of methods over all previous (first order) conditions. In particular, we showed that the safeguarded PHR augmented Lagrangian method also generates PCAKKT

points under a GL inequality (which is also used in [12]). As a consequence, this method reaches KKT points under a new CQ, related to PCAKKT, called PCAKKT-regularity. To the best of our knowledge, PCAKKT-regularity is the strongest CQ associated with convergence of a practical method (see Figure 4). A novel primal-dual augmented Lagrangian algorithm is also presented (Algorithm 2). We showed that it achieves PCAKKT points under mild assumptions. In particular, we provide strong convergence results, showing that in some situations the Lojasiewicz-type assumption is not necessary.

Computational tests were performed. We observed that the inner solver GENCAN presented difficulties in dealing with subproblems of Algorithm 2 when compared with its behaviour on Algorithm 1' subproblems. In our experience, GENCAN needs too many iterations to achieve a satisfactory optimality tolerance, and in many cases it was not able to reach it. In fact, the subproblems of Algorithm 2 are more challenging than those of Algorithm 1. We believe that a specialized solver, which exploits the structure derived from our primal-dual augmented Lagrangian function (3), is necessary to implement a competitive algorithm. Thus, a detailed investigation on how we can efficiently minimize the function (3), perhaps using second-order information, may be addressed in a future work. For instance, very recently Gill, Kungurtsev and Robinson [23] propose a shifted primal-dual penalty-barrier method based on a primal-dual augmented Lagrangian similar to (3), that also brings the multipliers estimates λ^a and μ^a and complementarity terms to the minimization of the subproblems. To solve the correspondent subproblems, the authors propose a Newtonian strategy carefully adapted to their necessity. Anyway, the preliminary computational results presented in this paper suggest that the primal-dual augmented Lagrangian (3) may be useful to improve the practical behaviour of the classical variant. In this sense, besides establishing adequate solvers for inner problems, rules for hybridizing the two strategies, such as those used in our tests, should be investigated.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 111(1):5–32, 2008.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309, 2007.

[3] R. Andreani, N. Fazzio, M. Schuverdt, and L. Secchin. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM Journal on Optimization*, 29(1):743–766, 2019.

[4] R. Andreani, W. Gomez, G. Haeser, L. Mito, and A. Ramos. On optimality conditions for nonlinear conic programming. *Optimization Online*, 2020.

[5] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011.

[6] R. Andreani, G. Haeser, A. Ramos, and P. J. S. Silva. A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA Journal of Numerical Analysis*, 37(4):1902–1929, 2017.

[7] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012.

[8] R. Andreani, G. Haeser, L. D. Secchin, and P. J. S. Silva. New sequential optimality conditions for mathematical problems with complementarity constraints and algorithmic consequences. *SIAM Journal on Optimization*, 29(4):3201–3230, 2019.

[9] R. Andreani, G. Haeser, and D. S. Viana. Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, 180:203–235, 2020.

[10] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016.

[11] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, 43(3):693–717, 2018.

[12] R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554, 2010.

[13] R. Andreani, V. A. Ramirez, S. A. Santos, and L. D. Secchin. Bilevel optimization with a multiobjective problem in the lower level. *Numerical Algorithms*, 81(3):915–946, 2019.

[14] 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 Journal on Optimization*, 28(3):2574–2600, 2018.

[15] E. G. Birgin and J. M. Martínez. Large-scale active-set box-constrained optimization method with spectral projected gradients. *Computational Optimization and Applications*, 23(1):101–125, 2002.

[16] E. G. Birgin and J. M. Martínez. Augmented Lagrangian method with nonmonotone penalty parameters for constrained optimization. *Computational Optimization and Applications*, 51(3):941–965, 2012.

[17] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[18] J. Bolte, A. Daniilidis, and A. Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

[19] A. R. Conn, N. I. M. Gould, and P. L. Toint. *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization*, volume 17 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, 1992.

[20] M. Feng and S. Li. An approximate strong KKT condition for multiobjective optimization. *TOP*, 26(3):489–509, 2018.

[21] A. Fischer and A. Friedlander. A new line search inexact restoration approach for nonlinear programming. *Computational Optimization and Applications*, 46(2):333–346, 2009.

[22] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A stabilized SQP method: global convergence. *IMA Journal of Numerical Analysis*, 37(1):407–443, 2016.

[23] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A shifted primal-dual penalty-barrier method for nonlinear optimization. *SIAM Journal on Optimization*, 30(2):1067–1093, 2020.

[24] P. E. Gill and D. P. Robinson. A primal-dual augmented Lagrangian. *Computational Optimization and Applications*, 51(1):1–25, 2012.

[25] E. S. Helou, S. A. Santos, and L. E. A. Simões. A new sequential optimality condition for constrained nonsmooth optimization. *to appear in SIAM Journal on Optimization*, 2020.

[26] C. Kanzow and D. Steck. An example comparing the standard and safeguarded augmented Lagrangian methods. *Operations Research Letters*, 45(6):598–603, 2017.

[27] J. M. Martínez and E. A. Pilotta. Inexact-restoration algorithm for constrained optimization. *Journal of Optimization Theory and Applications*, 104(1):135–163, 2000.

[28] J. M. Martínez and B. F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, 118(1):117–133, 2003.

[29] A. Ramos. Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. *Optimization Methods and Software*, pages 1–37, 2019.

[30] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1 edition, 1998.