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 May 18, 2020)

Abstract

Augmented Lagrangian algorithms are very popular and successful methods for solving constrained optimization problems. Recently, the global convergence analysis of these methods have been dramatically improved by using the notion of the sequential optimality conditions. Such conditions are optimality conditions independently 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 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 primal-dual augmented Lagrangian variant capable of achieving such points without the Lojasiewicz hypothesis. We then propose a hybrid method in which the new strategy acts to help the classical PHR variant when it tends to fail. We show by preliminary numerical tests that, in fact, all the problems already successfully solved by the pure PHR method remain unchanged, while others where it 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

1 Introduction

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

\[
\text{minimize } f(x) \text{ subject to } h(x) = 0, \quad g(x) \leq 0, \quad (\text{NLP})
\]

*This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPES (grant 116/2019), FAPESP (grants 2013/05475-7, 2017/18308-2) and CNPq (grants 301888/2017-5, 309437/2016-4, 438185/2018-8, 307270/2019-0).

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

Numerical methods for solving (NLP) are often iterative, and include stopping criteria that indicate when the current iterate is close to a solution candidate. 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 an accumulation point of certain sequences satisfying approximately the KKT conditions with tolerances going to zero [3, 5, 12, 27]. It is interesting to note that this property holds independently of the fulfillment of a constraint qualification (CQ) even at local minimizers where KKT fails. Thus this point of view, besides to give a theoretical treatment to the natural numerical approximations, allows a correct identification of these solutions. This idea leads to the notion of sequential optimality conditions, which we discuss in detail in this work.

The KKT conditions may be approximate by different ways. One of the most popular way is requiring 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

$$\|\nabla L(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k, \quad \min \{-g_j(x^k), \mu^k_j\} \leq \varepsilon_k, \quad \forall j$$

(1)

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

$$\text{minimize } (x_1 \! - \! 1)^2 + (x_2 \! - \! 1)^2 \text{ subject to } x_1 \geq 0, \ x_2 \geq 0, \ x_1x_2 \leq 0.$$  (2)

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 that use stopping criteria based on (1) may accept any feasible point as solution. As inexactness is a natural (perhaps inevitable) issue in the numerical world, the way that we approximate exact stationarity becomes an important question in the study of the theoretical convergence of practical methods. 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), avoiding spurious points as much as possible. As (1), no CQ is required on minimizers, 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, 27]. Furthermore, such concepts have been extended beyond standard nonlinear programming as mathematical programs with complementary constraints [8, 28], semidefinite programming [9], nonsmooth optimization [25] and multiobjective optimization [20].

Among the sequential optimality conditions for NLP, we mention the positive approximate KKT points (PAKKT) [3], and the complementary approximate KKT (CAKKT) [12] (see Definition 1). Both condition have been used to 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 has been shown 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, their sequences have superior properties. We start with a curious fact: there are situations where a point $x^*$ is simultaneously PAKKT and CAKKT, but does not exist a common sequence that carries both properties; 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. In order 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 then it can be potentially used to prove convergence for other 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.

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

$$L_{\rho, \nu, \bar{\lambda}, \bar{\mu}}(x, \lambda^a, \mu^a) := f(x) + \frac{\rho}{2} \left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \frac{\rho}{2} \left\| \left( g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \nu \sum_{i=1}^{m} (\lambda^a_i h_i(x))^2 + \nu \sum_{j=1}^{p} (\mu^a_j g_j(x))^2, \quad (3a)$$

where $\lambda^a \in \mathbb{R}^m$ and $\mu^a \in \mathbb{R}^p_+$, and $\| \cdot \|$ stands for the Euclidean norm. Here, $\lambda^a$ and $\mu^a$ can be interpreted as estimates of the Lagrange multiplier vectors for the constraints $h(x) = 0$ and $g(x) \leq 0$ respectively. The vectors $\lambda$ and $\bar{\mu}$ play the role of safeguarded multipliers as in the PHR AL method. At each iteration, an approximate stationary point $(x, \lambda^a, \mu^a)$ of the subproblem minimize $L_{\rho, \nu, \bar{\lambda}, \bar{\mu}}(x, \lambda^a, \mu^a)$ subject to $\mu^a \geq 0$ is computed (this justifies the name “primal-dual”). As the reader has noted, the term in the right side of (3a) corresponds to the classical PHR augmented Lagrangian function, which has been used in many successfully numerical methods as LANCELOT [19] and ALGENCAN [2, 17]. In turn, the term (3a) plus (3b) corresponds to the stabilized primal-dual augmented Lagrangian function presented in [24] (when only equality constraints are considered). It was used to provide a globalized version of the regularized 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
The proposed method employs rules to control the growth of penalty parameters $\rho$ and $\nu$ in (3). The idea is to increase $\rho$ (respectively $\nu$) only if feasibility (respectively complementarity) was not improved between consecutive iterations. Thus, when everything “goes well”, we can expect that numerical instabilities are minimized. This type of rule is used successfully in the Algencan package [17]. We show that the new primal-dual method is capable to recover PCAKKT points in the case that only one of $\rho$ or $\nu$ 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 even when $\rho$ tends to infinite, and without additional assumptions on the problem. We recall that convergence results for the safeguarded PHR AL method need the validity of GL inequality, already mentioned above.

Preliminary numerical tests were performed. Although the proposed method has good theoretical convergence, their subproblems are more challenging than those of PHR AL: in fact, they involve the minimization in both primal and dual variables. Furthermore, state of the art solvers as Algencan are effective on a wide variety of problems. However, as expected, they eventually fail. Thus, it is reasonable to propose alternative hybrid strategies to overcome difficulties encountered by traditional algorithms. For instance, a second-order AL algorithm was considered in [14] to solve a degenerate class of 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, namely Gencan [15]. Gencan 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 observation that common PAKKT and CAKKT sequences are stronger than requiring these properties at the limit point only. The relation of PCAKKT with other conditions from the literature are also presented, as well as the less stringent CQ that ensures that a PCAKKT point is actually KKT. It will be shown 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 CAKKT under mild assumptions. In this sense, the new method has stronger convergence properties than others from the literature. Section 5 is devoted to discuss numerical experience. We describe the hybrid strategy detailed, discussing how primal-dual iterations can be used to improve the well established PHR AL method. Instances where improvements are observed are reported.

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

1.2 Notation and terminology

Our notation is standard in optimization and variational analysis. \(\|\cdot\|\) and \(\|\cdot\|_\infty\) stands 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 \(y \ast z = (y_1z_1, \ldots, y_nz_n) \in \mathbb{R}^n\) is the Hadamard product between \(y\) and \(z\). 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 \(I_g(x) = \{j \in \{1, \ldots, p\} \mid g_j(x) = 0\}\). Given a function \(q, \nabla_z q\) is the gradient of \(q\) with respect to \(z\).

For a given set-valued mapping \(K : \mathbb{R}^s \rightrightarrows \mathbb{R}^n\), the sequential Painlevé-Kuratowski outer/upper limit of \(K(z)\) as \(z \to z^*\) [29] is defined as the set

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

2 A new sequential optimality condition

In this section we define a new sequential optimality condition, called Positive Complementary Approximate KKT (PCAKKT) condition. As every reasonable sequential condition, (i) it is a necessary optimality condition 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 the two first properties. 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 will show that PCAKKT is an optimality condition and derive some important properties. First we recall the definitions of some useful sequential optimality conditions from the literature. They differ, essentially, in the way that complementarity is approximated.
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.

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
\[
\lim_k x^k = x^* \quad \text{and} \quad \lim_k \|\nabla x L(x^k, \lambda^k, \mu^k)\| = 0. \tag{4}
\]
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, $x^*$ is an AKKT point;

(ii) [12] a Complementary Approximate KKT (CAKKT) sequence if, additionally to (4), condition (5) holds and
\[
\lim c_k = 0 \quad \text{where} \quad c_k := \sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j=1}^p |\mu_j^k g_j(x^k)|. \tag{6}
\]
for each $k$. In this case, $x^*$ is a CAKKT point;

(iii) [3] a Positive Approximate KKT (PAKKT) sequence if, additionally to (4), condition (5) holds and
\[
\lambda_i^k h_i(x^k) > 0 \, \text{if} \lim_k |\lambda_i^k|/\delta_k > 0, \quad \mu_j^k g_j(x^k) > 0 \, \text{if} \lim_k \mu_j^k/\delta_k > 0, \tag{7}
\]
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: let us consider a point $x^*$ that is simultaneously CAKKT and PAKKT. 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. 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 guarantee that there is a common sequence. Example 1 illustrates this situation.
Example 1. Let us consider the problem

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

The gradient of the Lagrangian function is

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

$$\bar{x}^k := (-1/k, 1/k), \ \bar{\mu}^k := (k^3/3, 0) \text{ for all } k \geq 1.$$

Furthermore, $x^*$ is a PAKKT point by taking the sequences

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

In fact, we have $\mu_2^k (\tilde{x}^k \sin^2 \tilde{x}_2^k) > 0$, $\forall k \in \mathbb{N}$, and $\lim_k \mu_2^k/(1, \mu_2^k) = 0$.

Note that the nature of these two sequences is distinct: the CAKKT sequence is not PAKKT since $\bar{\mu}^k_2$ is the unique multiplier that tends to infinity and $\bar{\mu}_2^k (x_1^3 \tilde{x}_2^3) < 0$ 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 $\lim_k \mu_2^k (x_1^3 \sin^2 x_2^k) = 0$. Now, from (8) and $\lim_k \nabla_x L(x^k, \mu^k) = 0$, it follows that $\lim_k 3\mu_2^k (x_1^3 \tilde{x}_2^3) = 0$. Taking a subsequence if necessary, 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 in particular that

$$\mu_1^k \to \infty \text{ and } 0 < \mu_1^k (x_2^k)^2 \to \infty. \hspace{1cm} (9)$$

We continue by proving that $\lim_k \mu_2^k/\mu_1^k = 0$ and $\lim_k \mu_3^k/\mu_1^k = 0$. 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 the fact that $\lim_k \nabla_x L(x^k, \mu^k) = 0$, we obtain

$$\frac{x^k_1 - 1}{\mu_1^k (x_2^k)^2} + 3(x_1^k)^2 x_2^k - \frac{\mu_2^k}{\mu_1^k} e^{x_1} \left(\sin \frac{x_2^k}{x_2^k}\right)^2 \to 0.$$

Since $\lim_k x^k = (0, 0)$, we conclude that $\lim_k \mu_2^k/\mu_1^k = 0$. Analogously, dividing the second row of (8) by $\mu_1^k$, taking the limit and using (9), we obtain

$$\frac{x^k_2 + 1}{\mu_1^k} + 3(x_1^k)^3 (x_2^k)^2 + \frac{\mu_2^k}{\mu_1^k} + 2\frac{\mu_3^k}{\mu_1^k} e^{x_1} \sin x_2 \cos x_2 \to 0.$$

So, as $\lim_k \mu_2^k/\mu_1^k = 0$ 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 $\|1, \mu_2^k\| = \mu_1^k$ for all sufficiently large $k$, giving $\lim_k \mu_2^k/(1, \mu_2^k) = 1$. 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 $\mu^k \geq 0$ 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 $\lim_k \nabla_x L(x^k, \mu^k) = 0$. \hfill \Box
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 a CAKKT and PAKKT sequence.

**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. That is, the PCAKKT condition is not the mere fulfilment 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 \((\lambda^*, \mu^*)\) 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 have to show that there is a PCAKKT sequence associated with \(x^*\). The multipliers sequence in the proof of [3, Lemma 2.6] is bounded. Then, the corresponding PAKKT sequence \(\{x^k\}\) is also a CAKKT one with the same sequence of multipliers.

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 [27], which is useful in the convergence analysis of algorithms based on inexact restoration technique [13, 21, 26].

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

\[
\text{proj}_{\Omega(x^*)}(-\nabla f(x^k)) \to 0,
\]

8
where $\text{proj}_{\Omega(x^*)}$ is the orthogonal projection onto

$$\Omega(x^*) := \left\{ d \in \mathbb{R}^n \mid \nabla h_i(x^k)^T d = 0, \quad i = 1, \ldots, m \
\quad \min\{0, g_j(x^k)\} + \nabla g_j(x^k)^T d \leq 0, \quad j \in I_g(x^*) \right\}.$$  

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

Inspired by Example 1, we can similarly discuss the relationship between PCAKKT and “PAKKT+AGP” (saying that a point is simultaneously PCAKKT and Ω(x∗) is defined by linear constraints). Indeed, as CAKKT implies AGP [12], Example 1 also shows that PCAKKT is stronger than “PAKKT+AGP”. Figure 2 summarizes all relations discussed here.

We finish this subsection with an alternative definition of the AGP condition. It 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\} $\subset \mathbb{R}^n$, \{λ^k\} $\subset \mathbb{R}^m$ and \{µ^k\} $\subset \mathbb{R}_+$ such that (4), (5) hold, and $\lim_k µ^k_j \min\{0, g_j(x^k)\} = 0$ for all $j \in I_g(x^*)$.

**Proof.** Assume that there are sequences \{x^k\} $\subset \mathbb{R}^n$, \{λ^k\} $\subset \mathbb{R}^m$ and \{µ^k\} $\subset \mathbb{R}_+$ such that (4) and (5) hold and $\lim_k µ^k_j \min\{0, g_j(x^k)\} = 0$, $j \in I_g(x^*)$. Define $d^k := \text{proj}_{\Omega(x^*)}(−\nabla f(x^k))$, which is the unique solution of

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

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

$$\nabla x L(x^k, λ^k, µ^k) = \nabla f(x^k) + \nabla h(x^k)λ^k + \nabla g(x^k)µ^k$$

by $d^k$ we obtain, since $d^k \in \Omega(x^k)$ and $µ^k_j = 0$, $j \notin I_g(x^*)$, for all $k$ large enough,

$$\|d^k\|^2 \leq −(d^k)^T \nabla f(x^k) \leq −(d^k)^T \nabla x L(x^k, λ^k, µ^k) = \sum_{j \in I_g(x^*)} µ^k_j \min\{0, g_j(x^k)\}$$

for all $k$ large enough. Taking the limit we get that $d^k \to 0$, 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 at $d^k$ since it is a minimizer and $\Omega(x^k)$ is defined by linear constraints.  \(\square\)
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 KKT for every smooth problem (NLP). We refer to the less 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 the AKKT, CAKKT and PAKKT conditions, 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*.

Let us recall the SCQs associated with CAKKT and PAKKT conditions. First, note that the KKT conditions hold at the feasible point $x^*$ for the objective function $f$ if, and only if, $-\nabla f(x^*) \in K(x^*)$, where $K(x)$ is the convex closed cone defined as

$$K(x) := \{ R(x, \lambda, \mu) \mid (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \},$$

where

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

($x^*$ will be clear in the context). 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

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

where $M(x, \alpha, \beta, \sigma)$ is the set of all $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that

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

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

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

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

$$\limsup_{x \to x^*, \alpha, \beta, \sigma > 0} K^{PC}(x, \alpha, \beta, \sigma) = \{ \omega \in \mathbb{R}^n \mid \exists (x^k, \omega^k) \to (x^*, \omega), \alpha_k \downarrow 0^+, \beta_k \downarrow 0, \sigma_k \downarrow 0 \text{ with } \omega^k \in K^{PC}(x^k, \alpha_k, \beta_k, \sigma_k), \forall k \in \mathbb{N} \}.$$  

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

$$\limsup_{x \to x^*, \alpha, \beta, \sigma > 0} K^{PC}(x, \alpha, \beta, \sigma) \subset K^{PC}(x^*, 0, 0, 0).$$

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

$$K^C(x, \sigma) := \{ R(x, \lambda, \mu) \mid (11c) \}, \quad K^P(x, \alpha, \beta) := \{ R(x, \lambda, \mu) \mid (11a), (11b) \},$$

10
Figure 3: Geometric interpretation of PCAKKT-regularity. There are two inequality constraints \( g_i(x) \leq 0, i = 1, 2 \). Three different sequences converges to \( x^* \): \( \{ x^k \} \) that violates the first constraint and satisfy the second; \( \{ \tilde{x}^k \} \) that violates both; and \( \{ \bar{x}^k \} \), that satisfies both. \( K(\cdot), K^P(\cdot, \alpha, \beta) \) and \( K^C(\cdot, \sigma) \) are represented, respectively, by small points, regions delimited by strong dashed lines, and by 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 are converging into limit cone \( K(x^*) \): \( K \) takes into account all the gradients of active constraints at \( x^* \); \( K^P \) considers only those gradients associated with violate constraints at intermediate points of the sequences; and \( K^C \) is “truncated” in a fashion that depends on the magnitude of \( \sigma \) and the measure of violation of each constraint. The set \( K^{PC} \) encompasses these two properties.

respectively. See [3, 11] for details. The reader may 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 of 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 for every smooth function \( f \) having \( x^* \) as a PCAKKT point it is also a KKT point, then PCAKKT-regular condition holds at \( x^* \).

**Proof.** Let \( x^* \) be a PCAKKT point of (NLP) with the corresponding sequences \( \{ x^k \} \subset \mathbb{R}^n, \{ \lambda^k \} \subset \mathbb{R}^m \) and \( \{ \mu^k \} \subset \mathbb{R}_+^p \). Assume that \( \mu^k_j = 0, \forall j \notin I_+(x^*) \).

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 \( \omega^k \in K^{PC}(x^k, \alpha_k, \beta_k, \sigma_k) \) for some \( \alpha_k, \beta_k, \sigma_k \geq 0 \) converging to zero. Define \( \delta_k := \| (1, \lambda^k, \mu^k) \|_\infty \), for \( k \in \mathbb{N} \), and the sets \( I_+ := \{ i \in \{ 1, \ldots, m \} \mid \lim_k |\lambda^k_i|/\delta_k > 0 \} \) and \( J_+ := \{ j \in \{ 1, \ldots, p \} \mid \lim_k \mu^k_j/\delta_k > 0 \} \) (these sets are well-defined and independent of \( k \), after taking a subsequence if necessary). For all \( k \), define

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

(12)

\[
\beta_k := \max \left\{ 1/k, \max_{i \in I_+} |\lambda^k_i|/\delta_k, \max_{j \notin J_+} \mu^k_j/\delta_k \right\} \quad \text{and} \quad \sigma_k := c_k,
\]

(13)

where \( c_k \) is as in (6). From conditions (6) and (7), we have that \( \alpha_k \downarrow 0^+ \) and \( \beta_k \downarrow 0 \). Furthermore, by the definition of \( I_+ \) and \( J_+ \), we have \( \beta_k \downarrow 0 \). Given
\( k \in \mathbb{N}, \) consider an index \( i \) such that \( |\lambda_i^k| > \beta_k \delta_k \). Since \( \delta_k > 0 \), using (13), we obtain \( |\lambda_i^k|/\delta_k > \beta_k \geq \max_{i \in I_k} |\lambda_i^k|/\delta_k \) and thus \( i \) must be in \( I_k \). Using (12), we conclude that \( \lambda_i^k \) are sufficiently large. So, by (11a), we conclude that PCAKKT-regularity implies PAKKT-regularity. These implications are strict, since the sequential optimality condition PCAKKT strictly implies CAKKT-regularity. Using a similar reasoning and [3, Theorem 2.4], we also have that PCAKKT-regularity implies PAKKT-regularity. The implications are strict, since the sequential optimality condition PCAKKT strictly implies each of conditions CAKKT and PAKKT. See Figure 4.

In order to complete the landscape of constraint qualifications known in the literature, we will show next 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), 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

\[
\mathcal{L}(x^*) := \{ d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0, \forall i, \nabla g_j(x^*)^T d \leq 0, j \in I_g(x^*) \}.
\]
Theorem 5. PCKKT-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 by the proof of this lemma that it provides multipliers sequences with the form \( \lambda^k = kh(x^k) \) and \( \mu^k = kg_i(x^k)_+ \) (which also appear in [10, Lemma 4.3]). Thus, \( \lambda^k \) and \( \mu^k \) have the same sign of their corresponding constraints, and furthermore, \( \mu^k_j = 0 \) for all \( k \) large enough, whenever \( j \notin J_0(x^*) \), and \( \omega^k \in K_{PC}(x^k, \alpha_k, \beta_k, \sigma_k) \) 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 we can extract a subsequence of \( \{x^k\} \) in order to \( \lambda^k = kh_i(x^k) \neq 0 \), for all \( i \in I_+ \), and \( \mu^k_i = kg_j(x^k)_+ > 0 \), for all \( j \in J_+ \).

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

Example 2 (Abadie’s CQ does not imply PCKKT-regularity). Consider the inequality constraints of Example 4 of [3]

\[
\begin{align*}
g_1(x) &= -x_1^2 + x_2, \\
g_2(x) &= -x_1^2 - x_2, \\
g_3(x) &= -x_1^5 + x_2, \\
g_4(x) &= -x_1^7 - x_2 \quad \text{and} \quad g_5(x) = -x_1,
\end{align*}
\]

and the point \( x^* = (0,0) \). 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 PCKKT-regular, consider \( \omega^* := (1,0) \notin K(x^*) \). Now, for all \( k \geq 1 \), define the sequences \( x^k := (-1/k, 0) \), \( \mu^k := (k^4/4, k^4/4, k^3, k^3, 0) \), \( \alpha_k := 1/k^2 \), \( \beta_k := 1/k \), \( \sigma_k := \sum_{j \in I_0(x^k)} \mu^k_j g_j(x^k) = (1/2k) + (2/k^2) \), and \( \omega^k := \sum_{j=1}^{5} \mu^k_j \nabla g_j(x^k) \). 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, PCKKT-regularity fails at \( x^* \).

Figure 4 shows some relations between several CQs in the literature. Note the unifying role of the PCKKT-regular condition. For others CQs considered in the figure, see [11] and references therein. Since PCKKT-regularity is less stringent than \( P/CAKKT \)-regularity, we can use it to establish an algorithm with stronger convergence properties than others from literature, by proving that such algorithm generates PCKKT points. The rest of the paper is devoted to study algorithms with this property.

3 Convergence of the safeguarded PHR AL method using PCKKT

In the last years, the global convergence analysis of AL methods has been dramatically improved 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 augmented Lagrangian method proposed in [2] (see Algorithm 1 below). As CAKKT, we use a generalization of the Lojasiewicz inequality introduced in [12]; 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 \( B(x^*) \) and a continuous function \( \varphi : B(x^*) \to \mathbb{R} \) such that \( \lim_{x \to x^*} \varphi(x) = 0 \) and, for all \( x \in B(x^*) \), we have that \( |\Phi(x) - \Phi(x^*)| \leq \varphi(x)\|\nabla \Phi(x)\| \).
Now, we will proceed to analyze Algorithm 1 below for solving (NLP). It makes use of the PHR augmented Lagrangian function

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

**Algorithm 1** 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 $\lim \varepsilon_k = 0$. Let $\lambda^1 \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\mu^1 \in [0, \mu_{\max}]^p$. Initialize with $k \leftarrow 1$.

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

**Step 2 (Update the penalty parameter).** Define

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

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 $\lambda^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$, $\mu^{k+1} \in [0, \mu_{\max}]^p$, $k \leftarrow k + 1$ and go to Step 1.

**Theorem 6.** Let $x^*$ be a feasible accumulation point of the sequence $\{x^k\}$ generated by Algorithm 1, let us say, $\lim_{k \in K} x^k = x^*$. 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 a KKT point under the PCAKKT-regularity condition.

In the case that $\{\|\lambda^k, \mu^k\|_\infty\}_{k \in K}$ is unbounded, then $\{x^k\}$ admits a subsequence which is a PCAKKT sequence.
Proof. The proof follows from the results in [3, 12]. When \( \{\|(1, \lambda_k^k, \mu_k^k)\|_\infty\}_{k \in K} \) is a bounded sequence, after taking a further subsequence, we may assume that \( \lambda_k^k \) and \( \mu_k^k \) converge to \( \lambda \) and \( \mu \) respectively. Thus, \( x^* \) is a KKT point, and hence \( x^* \) is a PCAKKT from Theorem 2.

Now, suppose that \( \{\|(1, \lambda_k^k, \mu_k^k)\|_\infty\}_{k \in K} \) is a unbounded sequence. From [3, Theorem 4.1], there is a subsequence of \( \{x_k^k\}_{k \in K} \) which is a PAKKT sequence. Furthermore, by the proof of [12, Theorem 5.1], we can extract a further subsequence of \( \{(x_k^k, \lambda_k^k, \mu_k^k)\}_{k \in K} \) conforming the definition of CAKKT. Thus, the final subsequence satisfies the requirements of the PCAKKT condition.

Theorem 6 provides the strongest result about the global convergence of the augmented Lagrangian method that we are aware of. Furthermore, Example 1 says that Theorem 6 implies more than the mere fulfilment 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 unify 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 (PAKKT), and other related with the approximate fulfillment of the KKT conditions (CAKKT). See Figure 2.

The fulfilment of the GL inequality of the measure of infeasibility \( \Phi(x) \) is a very general property and it is satisfied for a broad family of mappings with encompasses 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 avoid undesirable non-minimizers, it is natural to to ask for general purpose methods for solving (NLP) with such convergence properties without imposing the GL inequality. Following this line of research, a very interesting new method with nice properties is presented in [23]. It is based on the minimization of a shifted primal-dual penalty-barrier merit function, and their subproblems are solving by an interesting modification of the 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 in the CAKKT definition originally presented in [12] (see Definition 1), it is weaker than that. Indeed, the authors consider problem (NLP) with only inequality constraints, which, after the insertion of slack variables, takes the form

\[
\text{minimize}_{x,s} \ f(x) \ \text{subject to} \ g(x) = s, \ s \leq 0
\]

(this is adequate to the barrier method considered). Therefore, they define their sequential optimality condition using this problem [23, Definition 4.1], in the following way: a feasible point \((x^*, s^*)\) satisfies the CAKKT condition (in the sense of [23]) 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 \( x^k \to x^*, \ s^k \to s^* = g(x^*), \ \nabla f(x^k) + \nabla g(x^k) \mu^k \to 0, \ \mu^k - z^k \to 0, \ z_j^k s_j^k \to 0, \ \forall j \in I_g(x^*). \) (15)

In the sequel, we will show that not only (15) is weaker than the original CAKKT condition (Definition 1) for the problem (NLP) with inequality constraints only, but it actually strictly weaker than the less stringent AGP condition (Definition 3). See Figure 2.
Firstly, if \( x^* \) is an AGP point for (NLP) with only inequality constraints, by Theorem 3 there exist sequences \( \{x^k\} \subset \mathbb{R}^n \), \( \{\mu^k\} \subset \mathbb{R}^m_+ \) such that \( x^k \to x^* \), \( \lim k \nabla_x L(x^k, \mu^k) = 0 \), and \( \lim \mu^k \min_0 \{g_j(x^k)\} = 0 \). 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

\[
\begin{align*}
\text{minimize} & \quad 1/2(x_2 - 2)^2 \quad \text{subject to} \quad -x_1 \leq 0, \quad x_1 x_2 \leq 0.
\end{align*}
\]

Let us see that \( x^* = (0,1) \) is not an AGP point. Otherwise, by Theorem 3, there must exist sequences \( \{x^k\} \subset \mathbb{R}^2 \) and \( \{\mu^k\} \subset \mathbb{R}^2_+ \) such that \( x^k \to (0,1) \), \( \mu^k \min_0 \{x_1^k x_2^k\} \to 0 \), \( \mu^k_2 \min_0 \{x_1^k x_2^k\} \to 0 \) and

\[
\nabla_x L(x^k, \mu^k) = \begin{bmatrix} 0 \\ x_2^k - 2 \end{bmatrix} + \mu^k_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \mu^k_2 \begin{bmatrix} x_2^k \\ x_1^k \end{bmatrix} \to 0. \tag{16}
\]

As \( x_2^k \to 1 \), we have \( \mu^k_2 x_1^k \to 1 \), which in turn implies \( \mu^k_2 x_1^k x_2^k \to 1 \) and \( x_1^k \to \infty \) for all \( k \) large enough. Now, multiplying the first row of (16) by \( x_1^k \) and taking the limit, we get \( \mu^k_1 x_1^k \to 1 \), which contradicts \( \mu^k_1 \min_0 \{x_1^k x_2^k\} \to 0 \).

To see that condition (15) holds at \( x^* \), consider the sequences \( \bar{x}^k := (-1/k^2, 1), \bar{z}^k := \bar{\mu}^k := (k, k) \geq 0, \bar{s}^k := (1/k^2, 1/k^2) \). 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 \( k \), \( \nabla_x L(\bar{x}^k, \bar{\mu}^k) = 0, \bar{\mu}^k_2 \bar{z}^k = 0 \) and \( \bar{x}^k \bar{s}^k = 1/k \) for all \( k \).

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

## 4 A new shifted primal-dual method

In this section, we present our new augmented Lagrangian method and their convergence properties. We consider a penalty-like augmented Lagrangian function that carries the complementarity, bringing it to the minimization phase of the algorithm. Specifically, our method employs the function (3). It is straightforward to verify that

\[
\nabla_x L_{\rho, \nu, \lambda, \mu}(x, \lambda^a, \mu^a) = \nabla f(x) + \nabla h(x) \lambda + \nabla g(x) \mu, \tag{17}
\]

\[
\nabla_{\lambda^a} L_{\rho, \nu, \lambda, \mu}(x, \lambda^a, \mu^a) = \nu [\lambda^a \ast h(x)] \ast h(x) - [h(x) + (\bar{\lambda} - \lambda^a)/\rho],
\]

\[
\nabla_{\mu^a} L_{\rho, \nu, \lambda, \mu}(x, \lambda^a, \mu^a) = \nu [\mu^a \ast g(x)]_+ \ast g(x) - [g(x) + (\bar{\mu} - \mu^a)/\rho]_+ + \mu^a / \rho.
\]

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

\[
\lambda = [\rho h(x) + \bar{\lambda}] + [\rho h(x) + \bar{\lambda} - \lambda^a] \ast [\mu^a \ast h(x)] \ast \lambda^a,
\]

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

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

\[
\mathcal{G}_P(x, \lambda^a, \mu^a) := P_{\Omega} ((x, \lambda^a, \mu^a) - \nabla L_{\rho, \nu, \lambda, \mu}(x, \mu^a)) - (x, \lambda^a, \mu^a),
\]

16
where \(P_2[z]\) denotes 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 \(G^k_\rho(x, \lambda^a, \mu^a) = 0\), that is,
\[
\|\nabla_\lambda L_{\rho_k, v_k, \lambda^a, \mu^a}(x, \lambda^a, \mu^a)\|_\infty = 0, \quad \|\nabla_\lambda L_{\rho_k, v_k, \lambda^a, \mu^a}(x, \lambda^a, \mu^a)\|_\infty = 0 \quad \text{and} \quad \left\|\mu^a - \nabla_\mu L_{\rho_k, v_k, \lambda^a, \mu^a}(x, \lambda^a, \mu^a)\right\|_\infty = 0.
\]

**Algorithm 2 Primal-dual augmented Lagrangian method**

Let \(\lambda_{\text{min}} < \lambda_{\text{max}}, \mu_{\text{max}} > 0, \gamma > 1, \tau, a, \theta \in (0, 1), \{M_k\} \subset \mathbb{R}_+\) be a bounded sequence and \(\{\varepsilon_k\} \subset \mathbb{R}_+\) be a sequence such that \(\lim_{k \to \infty} \varepsilon_k = 0\).

Take \(\lambda^1 \in [\lambda_{\text{min}}, \lambda_{\text{max}}]^m, \mu^1 \in [0, \mu_{\text{max}}]^p, \tau_1 > 0, \rho_1 > 0, \nu_1 > 1, \zeta_1^{\text{max}} \in (0, 1)\). Initialize with \(k \leftarrow 1, \lambda^1 := \tilde{\lambda}^1/\rho_1\) and \(\mu^1 := \tilde{\mu}^1/\nu_1\).

**Step 1 (Solving the subproblems).** Find an approximate minimizer \((x^k, \lambda^{a,k}, \mu^{a,k})\) of \(L_{\rho_k, v_k, \lambda^a, \mu^a}(\cdot)\), satisfying \(\mu^{a,k} \geq 0\) and
\[
\begin{align}
\|\nabla_\lambda L_{\rho_k, v_k, \lambda^a, \mu^a}(x^k, \lambda^{a,k}, \mu^{a,k})\|_\infty &\leq \varepsilon_k, \quad (18a) \\
\|\nabla_\lambda L_{\rho_k, v_k, \lambda^a, \mu^a}(x^k, \lambda^{a,k}, \mu^{a,k})\|_\infty &\leq \frac{M_k}{\rho_k v_k}, \quad (18b) \\
\left\|\mu^{a,k} - \nabla_\mu L_{\rho_k, v_k, \lambda^a, \mu^a}(x^k, \lambda^{a,k}, \mu^{a,k})\right\|_\infty &\leq \frac{M_k}{\rho_k v_k}. \quad (18c)
\end{align}
\]

**Step 2 (Update penalty parameters).** Define
\[
V_k := \max\{\|h(x^k)\|_\infty, \|\min\{-g(x^k), \mu^k/\rho_k\}\|_\infty\},
C_k := \max\{\|\lambda^{a,k} * h(x^k)\|_\infty, \|\mu^{a,k} * g(x^k)\|_\infty\}, \quad \text{and}
\zeta_k := \max\{V_k, C_k\}.
\]

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

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

Let us highlight some aspects of Algorithm 2:

• Different 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 \(\{\lambda^{a,k}\}\) or \(\{\mu^{a,k}\}\) are bounded, the regularization terms in \(L_{\rho_k, v_k, \lambda^a, \mu^a}(\cdot)\), as well as (18b) and (18c), tend to control the grow 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 on the subproblems still be a box). One of them is the active-set strategy with spectral gradients known as GENCAN [15]. GENCAN is used in the PHR augmented Lagrangian 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);

- The parameters \( \rho \) and \( \nu \) are updated according to the behaviour of the feasibility and the complementary measures, given by \( V_k \) and \( C_k \) respectively. Each parameter is increased to emphasize the respectively measure in subsequent iteration. Note that the increment rule for \( \rho \) is more aggressive than that for \( \nu \). This reflects a preference for feasibility over complementarity in the algorithm. It is worth noting that \( V_k \) and the update rule for \( \rho \) are the same of the PHR augmented Lagrangian method (see Algorithm 1). Furthermore, \( \nu \) remains unchanged if the complementarity measure \( C_k \) decrease satisfactory. In this sense, Algorithm 2 tries to mimics the behaviour of PHR augmented Lagrangian method when the CAKKT-like complementarity goes well;

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

### 4.1 Global convergence analysis

In this subsection, we present the convergence results for Algorithm 2. In order 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 are able to state our main convergence result.

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

\[
\lambda^k = [\rho_k h(x^k) + \lambda^{k,0}] + [\rho_k h(x^k) + \lambda^{k,-} - \lambda^{a,k}] + \nu_k [\lambda^{a,k} * h(x^k)] * \lambda^{a,k}, \quad (19a)
\]

\[
\mu^k = [\rho_k g(x^k) + \mu^{k,0}] + [\rho_k g(x^k) + \mu^{k,-} - \mu^{a,k}] + \nu_k [\mu^{a,k} * g(x^k)] * \mu^{a,k}, \quad (19b)
\]

where \((x^k, \lambda^{a,k}, \mu^{a,k})\) is the current iterate. From now on, \( \lambda^k \) and \( \mu^k \) 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 \( K \subset \mathbb{N} \). Note that \( x^* \) is the unique limit point of \( \{x^k\}_{k \in K} \).
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 $j \in I_\rho(x^*)$, we split $K$ into the following disjoint subsets:

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

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

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

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

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 are 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 $\{\bar{\mu}_j^k\}$, $\rho_k g_j^k + \bar{\mu}_j^k \leq 0$ for all $k \in K$ large enough; for these $k$’s, we have $\mu_j^k = 0$ since $\mu_j^{a,k} \geq 0$ and $g_j^k \leq 0$. If $\{\rho_k\}$ is bounded then, by step 2, $\lim_k V_k = 0$ and thus $\lim_{k \in K} \bar{\mu}_j^k = 0$. As $g_j^k \leq g_j(x^*)/2 < 0$ for all $k$ large large, (19b) implies that $\mu_j^k = 0$ for these indexes $k$.

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 KKT, 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

$$\begin{align*}
(a) \quad \left| \chi_i^{a,k} - \frac{\chi_i^{a,k} + \rho_k h_i^k}{1 + \nu_k h_i^k} \right| & \leq \frac{M_h}{\nu_k (1 + \nu_k h_i^k)^2}; \\
(b) \quad \left| \left[ \mu_j^{a,k} + \left( \frac{\mu_j^{a,k} - \mu_j^{a,k}}{\rho_k} \right)_+ - \frac{\mu_j^{a,k}}{\rho_k} - \nu_k h_j^k (g_j^k)_+ \right] \right| & \leq \frac{M_h}{\rho_k \rho_k}; \\
(c) \quad \{ \nu_k \chi_i^{a,k} h_i^k \}_{k \in K} \text{ and } \{ \nu_k \mu_j^{a,k} (g_j^k)_+ \}_{k \in K} \text{ are bounded.}
\end{align*}$$
Proof. Items a and b can be proven directly from (18b) and (18c). In order to prove item c, we multiply item a by $\nu_k |h^k_i|$ obtaining

$$\left| \nu_k \lambda^a_i |h^k_i| = \frac{\nu_k \lambda^k_i |h^k_i| + \nu_k \rho_k h^k_i |h^k_i|}{1 + \nu_k \rho_k (h^k_i)^2} \right| \leq \frac{M_k |h^k_i|}{1 + \nu_k \rho_k (h^k_i)^2}. \quad (21)$$

From the boundedness of $\{M_k\}$, the right hand side of the above inequality remains bounded. Also, from the fact that $\nu_k \lambda^k_i = \hat{\lambda}^k_i, \forall k$, is in a compact set, $\{\nu_k \lambda^k_i |h^k_i|\}_{k \in K}$ is bounded. Since $\limsup_{k \in K} \frac{\nu_k \rho_k (h^k_i)^2}{1 + \nu_k \rho_k (h^k_i)^2} \leq 1$, $\{\nu_k \lambda^a_i h^k_i\}_{k \in K}$ must be bounded.

Now we treat item c for the inequality case. If $j \notin I_{\gamma}(x^*)$, the result is trivial. Suppose that $j \in I_{\gamma}(x^*)$ and split the set $K$ into disjoint sets as in (20). By the definition of $K_1$ and $K_2$, the sequence $\{\nu_k \mu_j^{a,k} |g_j^k|\}_{j \in K_1 \cup K_2}$ is trivially bounded. For all $k \in K_3$, item b takes the form

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

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 $\mu_j^{a,k} \leq M_k/\left[\nu_k(1 + \nu_k \rho_k (g_j^k)^2)\right]$. 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. \hfill \Box

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) is naturally satisfied. As the sequence $\{\delta_k := (\lambda^k, \mu^k)\}_{k \in K}$ has an unbounded subsequence, we may assume, 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 $\lim_{k \in K} |\lambda^k_i|/\delta_k > 0$ for a given index $i$. From Lemma 2, $\delta_k \to \infty$ and the boundedness of $\{M_k\}$ we obtain

$$\lim_{k \in K} \lambda^{a,k}_i \frac{\lambda^k_i}{\delta_k} = 0. \quad (22)$$

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

$$0 \neq \lim_{k \in K} \frac{\lambda^k_i}{\delta_k} = \lim_{k \in K} \frac{h^k_i}{\delta_k} \left[ \frac{\rho_k + 2 \nu_k \rho_k^2 (h^k_i)^2}{1 + \nu_k \rho_k (h^k_i)^2} + \nu_k (\lambda^a_i)^2 \right].$$

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

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

$$\lim_{k \in K} \left[ (\rho_k g_j^k + \hat{\mu}^{a,k}_j)_+ + \nu_k (\mu_j^{a,k} g_j^k)_+ \mu_j^{a,k} + (\rho_k g_j^k + \hat{\mu}^{a,k}_j)_+ \right] / \delta_k > 0.$$
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 fixed $c > 0$. 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). We then conclude that $K_4$ if infinite, which in turn allow us to assume that $k \in K_4$ for all $k$. Therefore, $g_j^k \geq 0$ for all $k$. But, if $g_j^k = 0$ for infinitely 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.

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, for all $j \in I_j(x^*)$, we have that $\lim_{k \in K_1 \cup K_2} \mu_j^k g_j^k = 0$.

**Proof.** Consider the partition of $K$ given by (20). In the sequel, we assume implicit that each of these $K_i$ 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 $K_1$.** The nonexpansiveness of the projection operator implies that $|[\rho_k g_j^k + \bar{\mu}_j^k]^+ - [\rho_k g_j^k]^+| \leq \bar{\mu}_j^k$, 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 get $\mu_j^k g_j^k = [\rho_k g_j^k + \bar{\mu}_j^k]^+ g_j^k \to K_0, 0$.

**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 $\lim_{k \in K_1 \cup K_2} \mu_j^k g_j^k = 0$.

Let us define the set

$$K := \left\{ k \in K \mid \zeta_k \leq \min \left\{ \frac{a}{\nu_k}, \nu_{\max} \right\} \right\}.$$

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

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

21
Lemma 5. Assume that $K$ is infinite. Then the entire sequence $\{x^k\}_{k \in K}$ (or at least one starting from a fixed index $k_0 \in K$) is a CAKKT sequence.

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

We start by proving that $\lim_{k \in K} \lambda^k h_i^k = 0$ for every $i = 1, \ldots, m$. Take $i \in \{1, \ldots, m\}$. We have $|\lambda^{a,k} h_i^k| \leq C_k \rightarrow_k 0$, and from $C_k \leq \zeta_k \leq \min\{a/v, \zeta^{\max}\}$, we also have $|\nu_k \lambda^{a,k} h_i^k| \leq \nu_k C_k \leq a$ for all $k \in K$.

Thus $\lim_{k \in K} \nu_k (\lambda^{a,k} h_i^k)^2 = 0$. From (21) and the fact that $\nu_k \lambda^k = \lambda^k$, $\forall k$, is in a compact set, we have $\limsup_{k \in K} \nu_k \rho_k (h_i^k)^2 \leq a < 1$. Hence $\{\nu_k \rho_k (h_i^k)^2\}_{k \in K}$ is bounded, which implies $\lim_{k \in K} \rho_k (h_i^k)^2 = 0$. In summary, we have $\lim_{k \in K} \rho_k (h_i^k)^2 = 0$, $\lim_{k \in K} \lambda^{a,k} h_i^k = 0$ and $\lim_{k \in K} \nu_k (\lambda^{a,k} h_i^k)^2 = 0$, which, by (19a), imply $\lambda^k h_i^k \rightarrow_k 0$. That is, the approximate CAKKT-like complementary condition for the equality constraints.

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

From (19b), recall that

$$
\mu^k_j = [\rho_k g_j^k + \mu^k_j]_+ + [\rho_k g_j^k + \mu^k_j - \mu^{a,k} j]_+ + \nu_k (\mu^{a,k} j)^2 |g_j^k|_+.
$$

and from $C_k \leq \min\{a/v_k, \zeta^{\max}\}$ for every $k \in K$, we have

$$
\mu^{a,k} j |g_j^k|_+ \rightarrow_k 0, \quad |\nu_k \mu^{a,k} j |g_j^k|_+| \leq a \quad \text{and} \quad \nu_k (\mu^{a,k} j |g_j^k|_+)^2 \rightarrow_k 0. \tag{25}
$$

Subsequences over $K_3$. Multiplying the inequality $\rho_k g_j^k + \mu^k_j - \mu^{a,k} j < 0$ by $g_j^k \geq 0$ and using the first limit in (25), we get $\lim_{k \in K_3} \rho_k (g_j^k)^2 = 0$ and therefore, using the last limit in (25), $\mu^k_j g_j^k = [\rho_k g_j^k + \mu^k_j]_+ |g_j^k|_+ + \nu_k (\mu^{a,k} j)^2 |g_j^k|_+^2 \rightarrow_k 0$.

Subsequences over $K_4$. Here, for every $k \in K_4$, we have

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

Note this $\mu^k_j$ has the same shape of the Lagrange multiplier estimate $\lambda^k$ 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 $\mu^k_j$ and $k \in K_4$.

Finally, we observe that all the arguments are valid for all $k \in K$ sufficiently large, and hence there is an index $k_0 \in K$ such that the entire sequence $\{x^k\}_{k \in K, k \geq k_0}$ is CAKKT, concluding the proof.

\[\square\]
Lemma 6. Suppose that $x^*$ is feasible and $K$ is finite. If the nondecreasing sequence $\{\min\{\rho_k, \nu_k\}\}$ is bounded then the entire sequence $\{x^k\}_{k \in K}$ (or at least one starting from a fixed index $k_0 \in K$) is a CAAKT sequence.

Proof. It is sufficient to show that $\lim_{k \in K} \lambda_i^k h^k_i = 0$, $\forall i$, and $\lim_{k \in K} \mu_i^k g^k_i = 0$, $\forall j$. We only prove the statement for equality constraints; for the inequalities with $j \in I_g(x^*)$ the proof is similar (the case where $j \not\in I_g(x^*)$ follows from Lemma 1).

Take an index $i \in \{1, \ldots, m\}$. Let us recall that, from (10a),

$$\lambda_i^k h_i^k = 2\rho_k (h_i^k)^2 + 2\lambda_i^k h_i^k - \lambda_i^k h_i^k + \nu_k (\lambda_i^k h_i^k)^2.$$

When $K$ is finite, the sequence $\{\zeta_i^k\}$ is updated only by a finite number of steps, which implies $\zeta_i^k = \zeta_i^k$ for all $k$ sufficiently large. Now, if $\{\max\{\rho_k, \nu_k\}\}$ is bounded, let says by $A > 0$, then $\zeta_i^k > \min\{a/A, \zeta_i^k\}$ 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 $\lim_{k \in K} \zeta_i^k = 0$, which lead 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 $\{\nu_k \lambda_i^k h_i^k\}_{k \in K}$ (Lemma 2, item c), we conclude that $\lim_{k \in K} \lambda_i^k h_i^k = \lim_{k \in K} \nu_k (\lambda_i^k h_i^k)^2 = 0$. 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, $\lim_{k \in K} \lambda_i^k h_i^k = 0$ and hence $\lim_{k \in K} \nu_k (\lambda_i^k h_i^k)^2 = 0$. 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 $\lim_{k \in K} \lambda_i^k h_i^k = 0$. \(\square\)

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

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

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

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

Proof. Firstly note that if $x^*$ is feasible and the sequence of Lagrange multipliers estimates $\{\{\lambda_i^k, \mu_i^k\}_{k \in K}$ is bounded (or at least has a bounded subsequence), then by Step 1 of Algorithm 2, $x^*$ is KKT. Then all items follows from Theorem 2 and the implications of Figure 2. Suppose then that such sequence is unbounded.

Item a: Applying Lemma 5, we have that the entire sequence $\{x^k\}_{k \in K}$, possibly starting from an fixed index, is a CAAKT sequence. Then, applying Lemma 3
on such sequence we get 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. In order to show that $x^*$ is an AGP point, it is enough to show, in view of Theorem 3, that

$$\lim_{k \in K} \mu_k \min \{0, g_j(x^k)\} = 0, \quad j \in I_g(x^*).$$

But since $\min \{0, g_j(x^k)\} = 0$ for all $k \in K_3 \cup K_4$, the result follows directly from Lemma 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 achieves 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 ($K$ finite), and both parameters $\rho$ and $\nu$ 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 point 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 will show that Algorithm 2 always reaches stationary points of the infeasibility problem

$$\min_x \Phi(x) = \|h(x)\|^2 + \|g(x)\|^2.$$  \hspace{1cm} (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, $K$ is finite and $\rho_k \to \infty$. From (18a) we obtain

$$\nabla f(x^k)/\rho_k + \nabla h(x^k)[\lambda^k/\rho_k] + \nabla g(x^k)[\mu^k/\rho_k] \to 0.$$ \hspace{1cm} (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 $\rho_k$, we get, $\forall k \in K$, that

$$\left| \lambda_i^{a,k} - \frac{\lambda_i^k}{\rho_k} \right| = \frac{h_i(x^k)}{\rho_k (1 + \nu_k \rho_k h_i(x^k))} \leq \frac{M_k}{\nu_k \rho_k (1 + \nu_k \rho_k h_i(x^k))}.$$  \hspace{1cm} (28)

From the boundedness of $\{M_k\}$ and $\{\lambda_i^k\}$, and from $\rho_k \to \infty$, the second term and the right hand side tend to zero. The third term also tends to zero independently if $h_i(x^k)$ vanishes or not. Thus $\lim_{k \in K} \lambda_i^{a,k}/\rho_k = 0$.  \hspace{2cm} (24)
From Lemma 2, item c, \( \{v_k \lambda_k^{a,k} | h_i(x_k)\} \) is a bounded sequence independently if \( h_i(x_k) \) vanishes or not. So, dividing (19a) by \( \rho_k \) and using the boundedness of \( \{\lambda_k\}_{k \in K} \), we obtain

\[
\frac{\lambda_k}{\rho_k} = 2 \left[ h_i(x_k) + \frac{\lambda_k}{\rho_k} \right] = \frac{\lambda_k^{a,k}}{\rho_k} + \frac{\lambda_k^{a,k}}{\rho_k} (v_k \lambda_k^{a,k} h_i(x_k)) \to K 2 h_i(x^*).
\]

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

\[
\frac{\mu_j^k}{\rho_k} = g_j(x^k) + \frac{\mu_j^k}{\rho_k} + g_j(x^k) - \frac{\mu_j^k}{\rho_k} + \frac{\mu_j^k}{\rho_k} \nu_k g_j(x^k) + \nu_k g_j(x^k) + \nu_k g_j(x^k). \quad (28)
\]

Now, split the set \( K \) into four disjoint sets \( K_1, K_2, K_3 \) and \( K_4 \) as (20). We will show that in each partition, \( \lim_{k \in K_i} \mu_k^k/\rho_k = 2[g_j(x^*)]_+ \) whenever \( K_i \) 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 \( \{\mu_j^k\} \), we see that \( \mu_j^k/\rho_k = [g_j(x^k) + \mu_j^k/\rho_k]_+ \to K_1 [g_j(x^*)]_+ = 0 = 2[g_j(x^*)]_+ \).

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

\[
\frac{\mu_j^k}{\rho_k} = g_j(x^k) + \frac{\mu_j^k}{\rho_k} + [\rho_k g_j(x^k) + \mu_j^k - \mu_j^a(x^k)]_+ \to K_2 0 = 2[g_j(x^*)]_+.
\]

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

\[
\frac{\mu_j^k}{\rho_k} = g_j(x^k) + \frac{\mu_j^k}{\rho_k} + \mu_j^a(x^k) g_j(x^k) \to K_3 0 = 2[g_j(x^*)]_+.
\]

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

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).

## 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 [2, 17] package provided freely by the TANGO project, compiled with GNU Fortran 7.5.0 using the “O3” flag. The Algencan 3.1.1 code 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 pure augmented Lagrangian 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, which is based on Algorithm 1, and which a primal-dual iteration of Algorithm 2 is applied when it is needed to force the fulfillment of the complementarity conditions. 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, which are also used to update the multipliers estimates \( \hat{\lambda}_{k+1} \) and \( \hat{\mu}_{k+1} \) of Step 3 by projecting (19) onto the boxes \([\lambda_{\min}, \lambda_{\max}]^m\) and \([0, \mu_{\max}]^p\); thus, solving these subproblems efficiently may require 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

\[
\|\nabla x^P_{\rho_k, \mu_k} (x^k)\|_\infty \leq \varepsilon_{\text{opt}}, \quad V_k \leq \varepsilon_{\text{opt}}, \quad \max \{\|\lambda_k \ast h(x^k)\|_\infty, \|\mu_k \ast g(x^k)\|_\infty\} > \sqrt{\varepsilon_{\text{opt}}},
\]

where \( \varepsilon_{\text{opt}} \) is 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 \( \zeta_k \leq \min\{a/\nu_k, \zeta_{\text{max}}^k\} \), the next iteration is also primal-dual. Therefore, we maintain the minimization on the primal and dual variables whenever both penalty parameters \( \rho \) and \( \nu \) 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 \( \rho \) becomes large (\( \geq 10^5 \)). 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 during consecutive iterations \( k, k + 1, \ldots, k + p \) this remains true, 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 \( k \) was 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 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, Algorithm 2, only if $V_k \leq \sqrt{\varepsilon_{\text{opt}}}$. That is, the penalty parameter $\nu$ can only be increased after a “sufficient” fulfillment of the (AKKT-type) complementarity $\min\{-g(x^k), \tilde{\mu}^k / \rho_k\}$. 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 ensure PCAKKT points under a Lojasiewicz hypothesis. Thus, we can expect CAKKT complementarity frequently. Then, our strategy aims to give measure $V_k$ a chance to achieve CAKKT points before increment $\nu$. 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 $\mu^{a,k} \geq 0$ in order to (19) equals to their first terms $\rho_k h(x^k) + \tilde{\lambda}^k$ and $[\rho_k g(x^k) + \tilde{\lambda}^k]_+$. 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 $\lambda^{a,k}$ and $\mu^{a,k} \geq 0$, 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 $\zeta_{\text{max}} = a/\nu_1$. All other parameters are initialized as ALGENCAN’s default values. 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 to recover 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 brings 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 “$f$”, “$\|\nabla L\|$”, “$\|\nabla L_{\text{PHR}}\|$” and “$V_k$” contains the final value of each quantity.

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
Hybrid strategy (Algorithm 1+Algorithm 2)

Problem | st | it | $f$ | $\|\nabla L\|$ | feas | $V_k$ | cpl (29)
---|---|---|---|---|---|---|---
HS56 | 0 | 71 | -3.46e+00 | 8.07e-07 | 1.22e-07 | 1.16e-06 |
QE226 | 0 | 29 | 2.13e+02 | 2.31e-07 | 2.31e-07 | 2.04e-07 |
QSHARE1B | 0 | 42 | 7.20e+05 | 2.56e-09 | 9.09e-12 | 2.56e-11 |

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

Finally, we stress that our proposal, at least from the practical point of view, can be viewed as a strategy to improve the convergence 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 compared 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 re-
lated 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). We shown that, curiously, PCAKKT points is not the simple union of PAKKT and CAKKT ones, it is in fact strictly stronger than both simultaneously (see Example 1). Thus, the PCAKKT condition really leads to an improvement in the convergence of methods over all previous (first order) conditions. In particular, we shown that the safeguarded PHR augmented Lagrangian method also generates PCAKKT points under a Lojasiewicz-type assumption (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 shown that it achieves PCAKKT points under mild assumptions. In particular, we provide strong convergence results, showing that in some situations the Lojasiewicz 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’s 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, as we mentioned before, the subproblems of Algorithm 2 are more challenging than those of Algorithm 2. We believe that a specialized solver, which exploits the structure derived from our primal-dual augmented Lagrangian function (3), is necessary in order to implement a competitive algorithm. Thus, a detailed investigation on how we can efficiently minimize our primal-dual augmented Lagrangian (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 $\lambda^a$ and $\mu^a$ and complementarity terms to the minimization of the subproblems. In order 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 establish adequate solvers for inner problems, rules for hybridizing the two strategies, such as those used in our tests, should be investigated.

References


