# A novel sequential optimality condition for smooth constrained optimization and algorithmic consequences

Renan W. Prado[1], Sandra A. Santos[1], and Lucas E. A. Simões[1]

[1] *Institute of Mathematics, University of Campinas, Rua Sergio Buarque de Holanda, 651, 13083-859, Campinas, São Paulo, Brazil*
Emails: renanwillianprado@hotmail.com, sandra@ime.unicamp.br, simoes.lea@gmail.com

March 31, 2022

**Abstract**

In the smooth constrained optimization setting, this work introduces the Domain Complementary Approximate Karush-Kuhn-Tucker (DCAKKT) condition, inspired by a sequential optimality condition recently devised for nonsmooth constrained optimization problems. It is shown that the augmented Lagrangian method can generate limit points satisfying DCAKKT, and it is proved that such a condition is not related to previously established sequential optimality conditions. An essential characteristic of the DCAKKT is to capture the asymptotic potential increasing of the Lagrange multipliers using a single parameter. Besides that, DCAKKT points satisfy the Strong Approximate Gradient Projection condition. Due to the intrinsic features of DCAKKT, which combine strength and generality, this novel and genuine sequential optimality condition may shed some light upon the practical performance of algorithms that are yet to be devised.

**Keywords:** Sequential optimality conditions, augmented Lagrangian method, nonlinear programming.

**AMS Classification:** 90C46, 90C30, 65K05

# 1 Introduction

This study focuses on the properties of the limit points generated by optimization methods used for solving problems of the form

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \\ &\text{s.t. } \boldsymbol{c}(\boldsymbol{x}) \leq \boldsymbol{0}, \quad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \end{aligned} \tag{1}$$

where the functions $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{c} : \mathbb{R}^n \to \mathbb{R}^p$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^q$ are all continuously differentiable. For such a purpose, we will explore the notion of Sequential Optimality Conditions (SOCs).

Optimality conditions are sets of properties that a point $\boldsymbol{x}^* \in \mathbb{R}^n$ must satisfy to be considered a good candidate for optimality [2]. Two types of optimality conditions are possible: pointwise and sequential ones. The first type encompasses those that rely just on the point $\boldsymbol{x}^*$ – under a constraint qualification, a good example would be the Karush-Kuhn-Tucker (KKT) conditions. The second ones are those that do not depend directly on $\boldsymbol{x}^*$ but rest upon a sequence of points that converge to $\boldsymbol{x}^*$. Since practical methods for solving minimization problems are, in general, iterative and possess asymptotic convergence, SOCs are more suitable to be used as stopping criteria than the pointwise conditions.

Every optimality condition aims at considering only local minimizers as valid candidates for optimality. However, since practical optimality conditions are usually necessary but not sufficient for a point to be a local minimizer, they might end up counting points that are not local minimizers as good candidates for optimality. Hence, the strength of an optimality condition is characterized by how well the condition at hand can avoid non-minimizers.

Because methods that satisfy strong sequential optimality conditions at their limit points can be said to have a strong convergence result, sequential optimality conditions are also tools for one to prove the convergence of an optimization method, and, in this sense, the stronger, the better. Moreover, SOCs that have general conditions can be applied to a broader class of algorithms – for example, the Approximate KKT (AKKT) condition [2] is fulfilled by a wide set of optimization methods [5, 8, 9, 17]. However, it is not desirable to have a very general sequential optimality condition, since it often implies that it is not strong enough to ensure good candidates for optimal points. As a result, in many cases, it is ideal to have a compromise between these two extremes.

Inspired by a smooth version of a sequential optimality condition recently developed for nonsmooth constrained optimization – the $\epsilon$-ASOC, cf. [11], this work introduces the *Domain Complementary Approximate Karush-Kuhn-Tucker* (DCAKKT) condition. The novelty of DCAKKT rests upon the fact that it implies the so-called *strong Approximate Gradient Projection* (SAGP) condition [18], and it is not related to the Positive Approximate KKT (PAKKT) condition [1], sequential optimality conditions also proved to be independent of each other, and to hold at limit points generated by the augmented Lagrangian method. Moreover, even by including the complementarity condition in the image space, which generates the *Domain and Image Complementary Approximate*

*Karush-Kuhn-Tucker* (DICAKKT) condition, no equivalence is achieved with already known sequential optimality conditions. Additionally, it is proved that SAGP is actually weaker than DCAKKT.

The contribution of this study lies in the presentation and contextualization of two new sequential optimality conditions – DCAKKT and DICAKKT. Both of them are generated by augmented Lagrangian methods and, at least the first one, is also satisfied by the algorithm called PACNO [12] whenever used for solving smooth optimization problems. Therefore, we believe our proposed SOCs are general enough to be applied to a set of existing methods and, possibly, to help the achievement of convergence proofs of upcoming optimization algorithms. Moreover, since they are stronger than many other sequential optimality conditions, we are convinced that a compromise has been achieved between strength and generality.

The roadmap of this work is the following. In Section 2, the new sequential optimality condition DCAKKT is defined, together with a stronger version of it (DICAKKT), accompanied by the results that relate them with the augmented Lagrangian method. In addition, examples put our contribution into perspective with the recently established conditions PAKKT [1] and PCAKKT [7]. Section 4 addresses the fact that the so-called *strong Approximate Gradient Projection* (SAGP) condition [18] is actually weaker than DCAKKT. Section 5 contains our final remarks.

**Notation** The following notation will be used along the present text:

- for $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v}_+ = (\max\{0, v_1\}, \ldots, \max\{0, v_n\})$;
- for $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v}_- = (\min\{0, v_1\}, \ldots, \min\{0, v_n\})$;
- $\|\cdot\|$ denotes the Euclidean norm;
- $\mathcal{B}[\boldsymbol{x}, \epsilon]$ is the closed ball with center $\boldsymbol{x}$, radius $\epsilon$, and the Euclidean distance;
- $\mathcal{P}_{\mathcal{X}}(\boldsymbol{v})$ denotes the Euclidean projection of $\boldsymbol{v}$ over the set $\mathcal{X}$;
- $\mathbb{R}_+$ stands for the non-negative real set;
- $\mathbb{R}_+^*$ denotes the strictly positive real set;
- $\mathbb{R}_-$ stands for the non-positive real set;
- given an integer $\ell$, $I_\ell = \{k \in \mathbb{N} : 1 \leq k \leq \ell\}$;
- the cardinality of a finite set $\mathbb{I}$ is denoted by $|\mathbb{I}|$;
- given a vector-valued function $\boldsymbol{\varphi} : \mathbb{R}^n \to \mathbb{R}^m$, its Jacobian matrix $(m \times n)$ computed at $\boldsymbol{x} \in \mathbb{R}^n$ is denoted by $J_{\boldsymbol{\varphi}}(\boldsymbol{x})$.

## 2 The novel sequential optimality condition

The Approximate Gradient Projection (AGP), introduced by Martínez and Svaiter in 2003 [18], is a well-known sequential optimality condition that is strictly stronger than the Approximate KKT (AKKT) conditions [2]. Linked to the AGP and the AKKT conditions is the sequential optimality condition known as Strong AKKT (SAKKT) [18]. Such a condition is stronger than the aforementioned ones – since both are implied by it – and the SAKKT can assure the KKT conditions under a very weak constraint qualification (quasi-normality [14]). One of the main reasons for the strength of the SAKKT condition relies on the sign control of the approximate Lagrange multipliers associated with the inequality constraints: the sign of the inequality constraint at a primal point must be the same as its associated approximate Lagrange multiplier. The fact that practical algorithms do not necessarily generate primal/dual iterates satisfying this property, however, usually prevents one to adopt the SAKKT condition as a stopping criterion.

Our new condition DCAKKT presents a relaxed version of the control over the sign of the approximate Lagrange multipliers possessed by the SAKKT condition, but with the positive side of having practical algorithms associated with it. This relaxation is established by zeroing the multipliers of the constraints that are sufficiently satisfied by a sequence $\{\boldsymbol{x}^k\}$ of primal approximations to a stationary point $\boldsymbol{x}^*$ of problem (1). Therefore, the meaningful multipliers are those associated with active or violated constraints in a given neighborhood of $\boldsymbol{x}^k$ – as it will be seen later, such relaxation is strong enough to still guarantee the validity of the AGP condition.

The next definition is in charge of addressing the corresponding indices; the sequential optimality condition in its weakest format is stated below it.

**Definition 2.1.** *With problem* (1) *as the reference, and given* $\epsilon > 0$ *and* $\boldsymbol{x} \in \mathbb{R}^n$*, we define the sets of indices*

$$\mathcal{I}_\epsilon(\boldsymbol{x}) := \{i \in I_p : \exists\ \boldsymbol{y} \in \mathcal{B}[\boldsymbol{x}, \epsilon]\ \text{with}\ c_i(\boldsymbol{y}) \geq 0\} \tag{2}$$

*and*

$$\mathcal{J}_\epsilon(\boldsymbol{x}) := \{j \in I_q : \exists\ \boldsymbol{y} \in \mathcal{B}[\boldsymbol{x}, \epsilon]\ \text{with}\ h_j(\boldsymbol{y}) = 0\}. \tag{3}$$

**Definition 2.2** (weak DCAKKT)**.** *A feasible point* $\boldsymbol{x}^*$ *of problem* (1) *is said to satisfy the* weak Domain Complementary Approximate Karush-Kuhn-Tucker *(weak DCAKKT) condition if there exist sequences* $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$*,* $\{\epsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^*_+$*,* $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$ *and* $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$ *such that* $\boldsymbol{x}^k \to \boldsymbol{x}^*$*,* $\epsilon_k \downarrow 0$*,*

$$\lim_{k\in\mathbb{N}} \left(\nabla f(\boldsymbol{x}^k) + \textstyle\sum_{i=1}^p \mu_i^k \nabla c_i(\boldsymbol{x}^k) + \sum_{j=1}^q \lambda_j^k \nabla h_j(\boldsymbol{x}^k)\right) = \mathbf{0}, \tag{4}$$

*and*

$$\begin{aligned} &\text{for } i \in I_p,\ \text{ if } i \notin \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k),\ \text{ then } \mu_i^k = 0, \\ &\text{for } j \in I_q,\ \text{ if } j \notin \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k),\ \text{ then } \lambda_j^k h_j(\boldsymbol{x}^k) \geq 0. \end{aligned} \tag{5}$$

It is worth noticing that the freedom of the sequence $\{\epsilon_k\}_{k\in\mathbb{N}}$ in the previous definition might turn the conditions in (5) innocuous, so that the weak DCAKKT condition would not add upon the Approximate Karush-Kuhn-Tucker (AKKT) condition, established in [2]. With a convenient control over the decreasing of $\epsilon_k$ towards zero, the next definition provides stronger sequential optimality conditions, as compared with the weak DCAKKT.

**Definition 2.3** (DCAKKT and DICAKKT)**.** *Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a feasible point that satisfies the weak DCAKKT condition. If the limits*

$$\epsilon_k\|\boldsymbol{\mu}^k\|_\infty \to 0 \quad \textit{and} \quad \epsilon_k\|\boldsymbol{\lambda}^k\|_\infty \to 0 \tag{6}$$

*also hold, we say that $\boldsymbol{x}^*$ fulfils the* Domain Complementary Approximate Karush-Kuhn-Tucker *(DCAKKT) condition. In this case, $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ is called a DCAKKT sequence. Moreover, if for all $i \in I_p$ and $j \in I_q$, we additionally have*

$$\mu_i^k c_i(\boldsymbol{x}^k) \to 0 \quad \textit{and} \quad \lambda_j^k h_j(\boldsymbol{x}^k) \to 0 \tag{7}$$

*then the point $\boldsymbol{x}^*$ is said to satisfy the* Domain-and-Image Complementary Approximate Karush-Kuhn-Tucker *(DICAKKT) condition. Under the additional assumption* (7)*, $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ is called a DICAKKT sequence.*

To start the analysis, we show that the DICAKKT is indeed a necessary optimality condition. Before presenting such a result, a classic auxiliary function is defined.

**Definition 2.4.** *Given the optimization problem* (1)*, we define its* quadratic measure of infeasibility *by the function $\Phi : \mathbb{R}^n \to \mathbb{R}$ stated as*

$$\Phi(\boldsymbol{x}) := \tfrac{1}{2}\left(\|\boldsymbol{c}(\boldsymbol{x})_+\|^2 + \|\boldsymbol{h}(\boldsymbol{x})\|^2\right). \tag{8}$$

**Theorem 2.1.** *Let $\boldsymbol{x}^*$ be a feasible point of problem* (1)*. If, in addition, $\boldsymbol{x}^*$ is a local minimizer of such a problem, then $\boldsymbol{x}^*$ fulfils the DICAKKT conditions.*

*Proof.* From the assumption that $\boldsymbol{x}^*$ is a local minimizer of (1), there exists $\delta > 0$ such that, for every $\boldsymbol{x}$ in $\mathcal{B}[\boldsymbol{x}^*, \delta]$, we have $f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$. Now, let us consider a sequence $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+$ such that $\lim_{k\in\mathbb{N}} \rho_k = +\infty$, and the sequence of the following unconstrained problems

$$\min_{\boldsymbol{x}\in\mathcal{B}[\boldsymbol{x}^*,\delta]} f(\boldsymbol{x}) + \rho_k\Phi(\boldsymbol{x}) + \tfrac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|^2, \tag{9}$$

with $\Phi(\cdot)$ defined in (8). Since the feasible set of the above problem is compact, and its objective function is continuous, for every $k \in \mathbb{N}$, problem (9) has a global solution. Hence, let $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ be the sequence of solutions of (9). Then, for each $k \in \mathbb{N}$,

$$f(\boldsymbol{x}^k) + \frac{1}{2}\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 + \rho_k\Phi(\boldsymbol{x}^k) \le f(\boldsymbol{x}^*), \tag{10}$$

and thus

$$\Phi(\boldsymbol{x}^k) \le \frac{f(\boldsymbol{x}^*) - f(\boldsymbol{x}^k)}{\rho_k}, \quad \forall k \in \mathbb{N}.$$

Now, since the right-hand side of the last inequality goes to zero as $k \to \infty$, the sequence $\{\Phi(\boldsymbol{x}^k)\}_{k\in\mathbb{N}}$ also converges to zero, which, by (8), ensures that every limit point of $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ is feasible for (1).

Let $\bar{\boldsymbol{x}}$ be a limit point of the sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ – which exists because $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ is bounded. Then, there must exist an infinite index set $\mathcal{K} \subset \mathbb{N}$ such that $\lim_{k\in\mathcal{K}} \boldsymbol{x}^k = \bar{\boldsymbol{x}}$. By (10), it follows that

$$f(\boldsymbol{x}^k) + \frac{1}{2}\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 \le f(\boldsymbol{x}^*) \Rightarrow \frac{1}{2}\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 \le f(\boldsymbol{x}^*) - f(\boldsymbol{x}^k).$$

By considering the limit in the index set $\mathcal{K}$, and since $\boldsymbol{x}^*$ is a local minimizer, the last inequality above gives

$$\|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|^2 \le f(\boldsymbol{x}^*) - f(\bar{\boldsymbol{x}}) \le 0, \tag{11}$$

so that $\bar{\boldsymbol{x}} = \boldsymbol{x}^*$, and because $\bar{\boldsymbol{x}}$ was an arbitrary limit point of the bounded sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$, it follows that

$$\lim_{k\in\mathbb{N}} \boldsymbol{x}^k = \boldsymbol{x}^*. \tag{12}$$

Now, due to (10), notice that we have $\rho_k\Phi(\boldsymbol{x}^k) \le f(\boldsymbol{x}^*) - f(\boldsymbol{x}^k)$, and because of (12), it follows that $\rho_k\Phi(\boldsymbol{x}^k) \to 0$. This yields

$$\rho_k c_i(\boldsymbol{x}^k)_+^2 \to 0,\ \forall\ i \in I_p \quad \text{and} \quad \rho_k h_j(\boldsymbol{x}^k)^2 \to 0,\ \forall\ j \in I_q. \tag{13}$$

Additionally, from the limit (12), for large values of $k \in \mathbb{N}$, the optimization problem in (9) becomes unconstrained, and therefore, $\boldsymbol{x}^k$ is a stationary point for the objective function in (9) for large values of $k \in \mathbb{N}$. Identifying

$$\boldsymbol{\mu}^k := \rho_k \boldsymbol{c}(\boldsymbol{x}^k)_+, \quad \boldsymbol{\lambda}^k := \rho_k \boldsymbol{h}(\boldsymbol{x}^k) \quad \text{and} \quad \epsilon_k := \frac{1}{\rho_k}, \tag{14}$$

it follows that (4) holds, since

$$\nabla f(\boldsymbol{x}^k) + \rho_k J_{\boldsymbol{c}}(\boldsymbol{x}^k)^T \boldsymbol{c}(\boldsymbol{x}^k)_+ + \rho_k J_{\boldsymbol{h}}(\boldsymbol{x}^k)^T \boldsymbol{h}(\boldsymbol{x}^k) = -\left(\boldsymbol{x}^k - \boldsymbol{x}^*\right) \overset{(12)}{\to} \boldsymbol{0}.$$

To obtain the remaining conditions (5), (6) and (7), notice that

$$\text{if } i \notin \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k), \text{ then } c_i(\boldsymbol{x}^k) < 0 \ \Rightarrow\ \mu_i^k \overset{(14)}{=} \rho_k c_i(\boldsymbol{x}^k)_+ = 0,$$

$$\text{if } j \notin \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k), \text{ then } h_j(\boldsymbol{x}^k) \ne 0 \ \Rightarrow\ \lambda_j^k h_j(\boldsymbol{x}^k) \overset{(14)}{=} \rho_k h_j(\boldsymbol{x}^k)^2 \ge 0,$$

$$\epsilon_k \|\boldsymbol{\mu}^k\|_\infty \overset{(14)}{=} \|\boldsymbol{c}(\boldsymbol{x}^k)_+\|_\infty \overset{(12)}{\to} 0,$$

$$\epsilon_k \|\boldsymbol{\lambda}^k\|_\infty \overset{(14)}{=} \|\boldsymbol{h}(\boldsymbol{x}^k)\|_\infty \overset{(12)}{\to} 0,$$

and

$$\mu_i^k c_i(\boldsymbol{x}^k) \stackrel{(14)}{=} \rho_k c_i(\boldsymbol{x}^k)_+ c_i(\boldsymbol{x}^k) = \rho_k c_i(\boldsymbol{x}^k)_+^2 \stackrel{(13)}{\to} 0,\ \forall i \in I_p,$$

$$\lambda_j^k h_j(\boldsymbol{x}^k) \stackrel{(14)}{=} \rho_k h_j(\boldsymbol{x}^k)^2 \stackrel{(13)}{\to} 0,\ \forall j \in I_q,$$

completing the proof. □

The next result assures that, whenever KKT is valid, the proposed sequential optimality condition also holds. Although such a result is expected, since sequential optimality conditions can be seen as a generalization of the KKT conditions, it will help us develop more meaningful results.

**Proposition 2.1.** *Let $\boldsymbol{x}^*$ be a KKT point for* (1). *Then, $\boldsymbol{x}^*$ also satisfies the DICAKKT condition.*

*Proof.* Since $\boldsymbol{x}^*$ is a KKT point for (1), then there must exist $\boldsymbol{\mu} \in \mathbb{R}_+^p$ and $\boldsymbol{\lambda} \in \mathbb{R}^q$ such that

$$\nabla f(\boldsymbol{x}^*) + \textstyle\sum_{i=1}^p \mu_i \nabla c_i(\boldsymbol{x}^*) + \sum_{j=1}^q \lambda_j \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}, \tag{15}$$

$$\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}, \tag{16}$$

$$\boldsymbol{c}(\boldsymbol{x}^*) \leq \boldsymbol{0}, \tag{17}$$

$$\boldsymbol{\mu}^T \boldsymbol{c}(\boldsymbol{x}^*) = 0. \tag{18}$$

Now, one can choose the sequences $\{\epsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+^*$, $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+^p$, $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$ to satisfy

$$\epsilon_k := 1/k,\ \boldsymbol{x}^k := \boldsymbol{x}^*,\ \boldsymbol{\mu}^k = \boldsymbol{\mu},\ \boldsymbol{\lambda}^k := \boldsymbol{\lambda}. \tag{19}$$

Notice that, except for $\{\epsilon_k\}_{k\in\mathbb{N}}$, the remanining sequences were chosen to be constant to facilitate the proof, but it is also possible to add small perturbations to these constant sequences to have a dynamic behavior. Thus, due to (19), for all $k \in \mathbb{N}$ we must have

$$\nabla f(\boldsymbol{x}^k) + \textstyle\sum_{i=1}^p \mu_i^k \nabla c_i(\boldsymbol{x}^k) + \sum_{j=1}^q \lambda_j^k \nabla h_j(\boldsymbol{x}^k) \stackrel{(15)}{=} \boldsymbol{0},$$

$$\text{given } i \in I_p \text{ such that } i \notin \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k), \text{ then } c_i(\boldsymbol{x}^k) < 0 \stackrel{(18)}{\Rightarrow} \mu_i^k = 0,$$

$$\text{given } j \in I_q \text{ such that } j \notin \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k), \text{ then } \lambda_j^k h_j(\boldsymbol{x}^k) \stackrel{(16)}{=} 0,$$

$$\epsilon_k \|\boldsymbol{\mu}^k\|_\infty \stackrel{(19)}{\to} 0 \quad \text{and} \quad \epsilon_k \|\boldsymbol{\lambda}^k\|_\infty \stackrel{(19)}{\to} 0,$$

$$\mu_i^k c_i(\boldsymbol{x}^k) \stackrel{(18),(19)}{=} 0,\ \forall i \in I_p \quad \text{and} \quad \lambda_j^k h_j(\boldsymbol{x}^k) \stackrel{(16),(19)}{=} 0,\ \forall j \in I_q.$$

The above relationships guarantee that $\boldsymbol{x}^*$ satisfies DICAKKT, as desired. □

# 3 Relating DCAKKT to the safeguarded augmented Lagrangian method

To show that the DCAKKT is linked not only to the PACNO algorithm, but other optimization methods may generate limit points fulfilling its conditions, we reserve the present section to show that the DCAKKT conditions are also produced by augmented Lagrangian methods. Next, we define the Lagrangian function and the augmented Lagrangian function of Powell-Hestenes-Rockafellar (cf. [13, 19, 20]) associated with our problem of interest.

**Definition 3.1.** *The Lagrangian function for problem* (1) *is defined by*

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f(\boldsymbol{x}) + \boldsymbol{\mu}^T \boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x})$$

*with* $\boldsymbol{\mu} \in \mathbb{R}^p_+$ *and* $\boldsymbol{\lambda} \in \mathbb{R}^q$.

**Definition 3.2.** *The augmented Lagrangian of Powell-Hestenes-Rockafellar (PHR) for problem* (1) *is given by*

$$\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f(\boldsymbol{x}) + \frac{\rho}{2}\left(\left\|\left(\frac{\boldsymbol{\mu}}{\rho} + \boldsymbol{c}(\boldsymbol{x})\right)_+\right\|^2 + \left\|\frac{\boldsymbol{\lambda}}{\rho} + \boldsymbol{h}(\boldsymbol{x})\right\|^2\right), \tag{20}$$

*with* $\rho > 0$, $\boldsymbol{\mu} \in \mathbb{R}^p_+$ *and* $\boldsymbol{\lambda} \in \mathbb{R}^q$.

In numerical analysis and optimization, it is a commom practice to replace the addressing of a difficult problem by solving a sequence, possibly infinite, of easier instances. Along this line, instead of problem (1), one may consider

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \mathcal{L}_{\rho_k}(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}^k, \bar{\boldsymbol{\mu}}^k) \tag{21}$$

in which $\{\rho_k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+$ is nondecreasing, and $\{\bar{\boldsymbol{\lambda}}^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^q$ and $\{\bar{\boldsymbol{\mu}}^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^p_+$ are bounded sequences. The dynamics for approximately solving (21), and updating the fundamental sequences of penalty parameters and Lagrange multipliers are established in the next algorithm, which follows the main model algorithmic presentation of [8].

---

**The safeguarded augmented Lagrangian method (ALM)**

---

**Step 0** Choose real numbers satisfying $\lambda_{\min} < \lambda_{\max}$, $0 < \mu_{\max}$, $\gamma > 1$, $0 < r < 1$, and define a sequence $\{\theta_k\}_{k \in \mathbb{N}}$ such that $\theta_k \downarrow 0$. Let $\bar{\boldsymbol{\lambda}}^1 \in [\lambda_{\min}, \lambda_{\max}]^q$, $\bar{\boldsymbol{\mu}}^1 \in [0, \mu_{\max}]^p$, $\rho_1 > 0$ and $\theta_{\text{opt}} > 0$. Initialize $k \leftarrow 1$.

**Step 1** Find $\boldsymbol{x}^k \in \mathbb{R}^n$ such that

$$\left\|\nabla f(\boldsymbol{x}^k) + \sum_{i=1}^p \mu_i^k \nabla c_i(\boldsymbol{x}^k) + \sum_{j=1}^q \lambda_i^k \nabla h_j(\boldsymbol{x}^k)\right\| \leq \theta_k, \tag{22}$$

with $$\boldsymbol{\mu}^k := \left(\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{c}(\boldsymbol{x}^k)\right)_+ \quad \text{and} \quad \boldsymbol{\lambda}^k := \bar{\boldsymbol{\lambda}}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k). \tag{23}$$

**Step 2** For each $i \in I_p$, set

$$V_i^k := \max\left\{c_i(\boldsymbol{x}^k), -\frac{\bar{\mu}_i^k}{\rho_k}\right\}. \tag{24}$$

If $\max\left\{\left\|\boldsymbol{V}^k\right\|_\infty, \left\|\boldsymbol{h}(\boldsymbol{x}^k)\right\|_\infty\right\} \le \theta_{\text{opt}}$ and

$$\left\|\nabla f(\boldsymbol{x}^k) + \textstyle\sum_{i=1}^p \mu_i^k \nabla c_i(\boldsymbol{x}^k) + \sum_{j=1}^q \lambda_i^k \nabla h_j(\boldsymbol{x}^k)\right\| \le \theta_{\text{opt}},$$

then terminate.

**Step 3** If $k = 1$ or

$$\max\left\{\left\|\boldsymbol{V}^k\right\|_\infty, \left\|\boldsymbol{h}(\boldsymbol{x}^k)\right\|_\infty\right\} \le r \max\left\{\left\|\boldsymbol{V}^{k-1}\right\|_\infty, \left\|\boldsymbol{h}(\boldsymbol{x}^{k-1})\right\|_\infty\right\}, \tag{25}$$

set $\rho_{k+1} = \rho_k$; otherwise, set $\rho_{k+1} = \gamma\rho_k$. Compute

$$\bar{\boldsymbol{\mu}}^{k+1} := \mathcal{P}_{[0,\mu^{\max}]^p}(\boldsymbol{\mu}^k) \quad \text{and} \quad \bar{\boldsymbol{\lambda}}^{k+1} := \mathcal{P}_{[\lambda^{\min},\lambda^{\max}]^q}(\boldsymbol{\lambda}^k). \tag{26}$$

Set $k \leftarrow k+1$ and go to **Step 1**.

---

The auxiliary result presented next will be useful to simplify the proofs that limit points of the algorithm ALM satisfy DCAKKT and DICAKKT (cf. Theorems 3.1 and 3.2 in the sequel, respectively).

**Lemma 3.1.** *Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a limit point of the sequence generated by the algorithm ALM. If $\boldsymbol{x}^*$ is feasible for problem* (1), *and the sequence $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, then $\boldsymbol{x}^*$ is a KKT point for problem* (1).

*Proof.* Let $\{\boldsymbol{x}^k\}_{n\in\mathbb{N}}$ be the sequence generated by the algorithm ALM, and let $\mathcal{K} \subset \mathbb{N}$ be the infinite set for which $\lim_{k\in\mathcal{K}} \boldsymbol{x}^k = \boldsymbol{x}^*$.

Under the boundedness assumption of $\{\rho_k\}_{k\in\mathbb{N}}$, it follows that

$$\mu_i^k := \left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+, \ \forall i \in I_p \quad \text{and} \quad \lambda_j^k := \left(\rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k\right), \ \forall j \in I_q$$

define bounded subsequences for $k \in \mathcal{K}$ as well.

Let $\mathcal{K}_1$, $\rho > 0$, $\boldsymbol{\mu} \ge \boldsymbol{0}$, $\bar{\boldsymbol{\mu}} \ge \boldsymbol{0}$, $\boldsymbol{\lambda}$ and $\bar{\boldsymbol{\lambda}}$ be such that $\mathcal{K}_1 \subset \mathcal{K}$ is an infinite set, $\lim_{k\in\mathcal{K}_1} \rho_k = \rho$, $\lim_{k\in\mathcal{K}_1} \boldsymbol{\mu}^k = \boldsymbol{\mu}$, $\lim_{k\in\mathcal{K}_1} \bar{\boldsymbol{\mu}}^k = \bar{\boldsymbol{\mu}}$, $\lim_{k\in\mathcal{K}_1} \boldsymbol{\lambda}^k = \boldsymbol{\lambda}$ and $\lim_{k\in\mathcal{K}_1} \bar{\boldsymbol{\lambda}}^k = \bar{\boldsymbol{\lambda}}$. Given $i \in I_p$, assume that $c_i(\boldsymbol{x}^*) < 0$. In such a case,

$$\mu_i = \lim_{k\in\mathcal{K}_1} \left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ = \left(\rho\ c_i(\boldsymbol{x}^*) + \bar{\mu}_i\right)_+.$$

From the boundedness of $\{\rho_k\}_{k\in\mathbb{N}}$, Step 4 of the algorithm ALM assures that $V_i^k \to 0$, and thus $\bar{\mu}_i = 0$. Then, $\mu_i = 0$ and $\mu_i c_i(\boldsymbol{x}^*) = 0$. Now, if $c_i(\boldsymbol{x}^*) = 0$, we also have $\mu_i c_i(\boldsymbol{x}^*) = 0$. Therefore, $\boldsymbol{\mu}^T \boldsymbol{c}(\boldsymbol{x}^*) = 0$, with $\boldsymbol{\mu} \in \mathbb{R}_+^p$.

Moreover, taking the limit for $k \in \mathcal{K}_1$ on the condition (22) of Step 1 of the algorithm ALM we obtain

$$\nabla f(\boldsymbol{x}^*) + \textstyle\sum_{i=1}^p \mu_i \nabla c_i(\boldsymbol{x}^*) + \sum_{j=1}^q \lambda_j \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},$$

so that the **KKT** conditions are valid at the point $\boldsymbol{x}^*$. □

A notation for the sets of indices associated with vanishing gradients of the constraints is established below. Such a definition aims to simplify the proof of the subsequent result, in which feasible limit points of the algorithm ALM are proven to satisfy the DCAKKT condition.

**Definition 3.3.** *Let $\boldsymbol{x} \in \mathbb{R}^n$ be a feasible point for problem* (1)*. We define the sets*

$$\Delta_{\boldsymbol{c}}(\boldsymbol{x}) := \{i \in I_p \mid \nabla c_i(\boldsymbol{x}) = \boldsymbol{0}\} \tag{27}$$

*and*

$$\Delta_{\boldsymbol{h}}(\boldsymbol{x}) := \{j \in I_q \mid \nabla h_j(\boldsymbol{x}) = \boldsymbol{0}\}. \tag{28}$$

**Theorem 3.1.** *Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a limit point of the sequence generated by the algorithm ALM, which is feasible for problem* (1)*. Then, $\boldsymbol{x}^*$ fulfils the DCAKKT condition.*

*Proof.* Let $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$, $\{\rho_k\}_{k\in\mathbb{N}}$, $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}}$, and $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}}$ be the sequences generated by the algorithm ALM . From the assumptions, there exists an infinite set $\mathcal{K} \subset \mathbb{N}$ such that $\lim_{k\in\mathcal{K}} \boldsymbol{x}^k = \boldsymbol{x}^*$. Concerning the sequence $\{\rho_k\}_{k\in\mathbb{N}}$, there are two possibilities to be analyzed: such a sequence is either bounded or unbounded.

Under the boundedness of the sequence of penalty parameters, the validity of the DCAKKT condition at the point $\boldsymbol{x}^*$ follows from Lemma 3.1 and Proposition 2.1.

Assume now that $\{\rho_k\}_{k\in\mathbb{N}}$ is unbounded. From the maintenance of the signal property of limits, for $k$ large enough, the sequence

$$\delta^k := \min_{i\notin\Delta_{\boldsymbol{c}}(\boldsymbol{x}^*),\, j\notin\Delta_{\boldsymbol{h}}(\boldsymbol{x}^*)} \left\{\|\nabla c_i(\boldsymbol{x}^k)\|, \|\nabla h_j(\boldsymbol{x}^k)\|\right\} \tag{29}$$

is bounded away from zero, with the sets $\Delta_{\boldsymbol{c}}(\boldsymbol{x}^*)$ and $\Delta_{\boldsymbol{h}}(\boldsymbol{x}^*)$ as in (27)–(28). Thus, for sufficiently large $k$, consider the sequence

$$\epsilon_k := \frac{\mu_{\max} + |\lambda_{\max}| + |\lambda_{\min}|}{\delta_k\, \rho_k}. \tag{30}$$

Moreover, let the characteristic function of the interval $[0, +\infty)$ be stated by

$$\mathrm{I}_{[0,+\infty)}(x) := \begin{cases} 1, & \text{if } x \in [0, +\infty) \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

and let the sequences $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}}$, $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{N}}$ be defined componentwise as

$$\mu_i^k := \alpha_i^k \hat{\mu}_i^k,\ \forall i \in I_p \quad \text{and} \quad \lambda_j^k := \beta_j^k \hat{\lambda}_j^k,\ \forall j \in I_q, \tag{32}$$

with

$$\alpha_i^k := \sup_{\boldsymbol{y}\in\mathcal{B}[\boldsymbol{x}^k,\epsilon_k]} \mathrm{I}_{[0,+\infty)}(c_i(\boldsymbol{y})), \qquad \hat{\mu}_i^k := \left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+, \tag{33}$$

and

$$\beta_j^k := \sup_{\boldsymbol{y}\in\mathcal{B}[\boldsymbol{x}^k,\epsilon_k]} \mathrm{I}_{[0,+\infty)}(\hat{\lambda}_j^k h_j(\boldsymbol{y})), \qquad \hat{\lambda}_j^k := \rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k. \tag{34}$$

Observe that, for all $i \in I_p$ and $j \in I_q$, we have that both $\alpha_i^k$ and $\beta_j^k$ belong to the set $\{0, 1\}$. Moreover, the sequences $\alpha_i^k$ and $\beta_j^k$ aim at zeroing, respectively, the multipliers $\mu_i^k$ and $\lambda_j^k$ whose signals are in disagreement with the signal of the corresponding constraint, for points sufficiently close to $\boldsymbol{x}^k$. Consequently, given $i \notin \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$ then $\mu_i^k = 0$, and given $j \notin \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k)$, then $\lambda_j^k h_j(\boldsymbol{x}^k) \geq 0$, so that (5) is verified. One should notice that even if $\mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$ or $\mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k)$ happens to be empty sets, the corresponding implication vacuously holds.

To prove the domain complementarity conditions (6), observe that, for each $i \in I_p$,

$$|\epsilon_k \mu_i^k| \leq \frac{\mu_{\max} + |\lambda_{\max}| + |\lambda_{\min}|}{\delta_k} \left( c_i(\boldsymbol{x}^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right)_+ \xrightarrow{k \in \mathcal{K}} 0,$$

and, for each $j \in I_q$,

$$|\epsilon_k \lambda_j^k| \leq \frac{\mu_{\max} + |\lambda_{\max}| + |\lambda_{\min}|}{\delta_k} \left| h_j(\boldsymbol{x}^k) + \frac{\bar{\lambda}_j^k}{\rho_k} \right| \xrightarrow{k \in \mathcal{K}} 0,$$

since $\{\bar{\mu}_i^k\}_{k \in \mathcal{K}}$ and $\{\bar{\lambda}_j^k\}_{k \in \mathcal{K}}$ are bounded and $\boldsymbol{x}^*$ is feasible.

It remains to show the validity of the stationarity condition (4). To accomplish this task, notice that the issue is the behavior of the multipliers of the constraints associated with $\alpha_i^k = 0$ and $\beta_j^k = 0$, (cf. (33) and (34), resp.) as defined in Step 1 of the algorithm ALM, since those multipliers are the ones that might asymptotically damage the stationarity condition. Therefore, for each $i \in I_p$ such that $\alpha_i^k = 0$, two possibilities must be analyzed: $i \in \Delta_{\boldsymbol{c}}(\boldsymbol{x}^*)$ and $i \notin \Delta_{\boldsymbol{c}}(\boldsymbol{x}^*)$, whereas for $j \in I_q$ such that $\beta_j^k = 0$, the cases to be considered are $j \in \Delta_{\boldsymbol{h}}(\boldsymbol{x}^*)$ and $j \notin \Delta_{\boldsymbol{h}}(\boldsymbol{x}^*)$.

Let us first fix $k \in \mathbb{N}$, and then we start the analysis with $\alpha_i^k = 0$ and $i \in \Delta_{\boldsymbol{c}}(\boldsymbol{x}^*)$. Notice that (33) implies that $c_i(\boldsymbol{x}^k) < 0$, and hence

$$0 \leq \mu_i^k \leq \left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ \leq \bar{\mu}_i^k \leq \mu_{\max}, \tag{35}$$

which will be useful in a while.

In case $\beta_j^k = 0$ and $j \in \Delta_{\boldsymbol{h}}(\boldsymbol{x}^*)$, the first observation is that $j \notin \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k)$ – the index set defined in (3) – and from (34), necessarily $\hat{\lambda}_j^k \neq 0$, since, otherwise, $\beta_j^k = 1$ would take place. Assume first that $\hat{\lambda}_j^k > 0$. From (34), to occur $\beta_j^k = 0$, for an arbitrary $\boldsymbol{y} \in \mathcal{B}[\boldsymbol{x}^k, \epsilon_k]$ we must have

$$h_j(\boldsymbol{y}) < 0. \tag{36}$$

Thus,

$$|\lambda_{\max}| \geq \lambda_{\max} \overset{(36)}{\geq} \rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k \overset{(34)}{=} \hat{\lambda}_j^k > 0 \geq -|\lambda_{\min}|.$$

Now, if $\hat{\lambda}_j^k < 0$, again from (34), to occur $\beta_j^k = 0$, for an arbitrary $\boldsymbol{y} \in \mathcal{B}[\boldsymbol{x}^k, \epsilon_k]$ it must hold

$$h_j(\boldsymbol{y}) > 0. \tag{37}$$

Hence,

$$|\lambda_{\max}| \geq 0 \geq \rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k \overset{(34)}{=} \hat{\lambda}_j^k \overset{(37)}{\geq} \bar{\lambda}_i^k \geq -|\lambda_{\min}|.$$

Consequently, both cases yield

$$|\hat{\lambda}_j^k| \leq \max\{|\lambda_{\max}|, |\lambda_{\min}|\}, \tag{38}$$

to be used ahead as well.

Now, for $i \in I_p$ such that $\alpha_i^k = 0$ and $i \notin \Delta_{\boldsymbol{c}}(\boldsymbol{x}^*)$, and for $j \in I_q$ such that $\beta_j^k = 0$ and $j \notin \Delta_{\boldsymbol{h}}(\boldsymbol{x}^*)$, we will show that the corresponding multipliers $\hat{\mu}_i^k$ and $\hat{\lambda}_k^j$, stated in (33)–(34), and in accordance with the updating (23) of Step 1 of the algorithm ALM, besides being bounded, actually converge to zero. First notice that, in this case, for $k \in \mathbb{N}$ and large enough, we have

$$\|\nabla c_i(\boldsymbol{x}^k)\| \geq \delta_k \quad \text{and} \quad \|\nabla h_j(\boldsymbol{x}^k)\| \geq \delta_k,$$

with $\delta_k$ given in (29). Then, for the aforementioned indices $i$ and $j$, by recalling the definition (30) and setting the sequences of vectors

$$\boldsymbol{y}_i^k := \boldsymbol{x}^k + \frac{\bar{\mu}_i^k}{\rho_k \|\nabla c_i(\boldsymbol{x}^k)\|} \frac{\nabla c_i(\boldsymbol{x}^k)}{\|\nabla c_i(\boldsymbol{x}^k)\|} \in \mathcal{B}[\boldsymbol{x}^k, \epsilon_k] \tag{39}$$

and

$$\boldsymbol{y}_j^k := \boldsymbol{x}^k + \frac{\operatorname{sign}(\hat{\lambda}_j^k)|\bar{\lambda}_j^k|}{\rho_k \|\nabla h_j(\boldsymbol{x}^k)\|} \frac{\nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|} \in \mathcal{B}[\boldsymbol{x}^k, \epsilon_k], \tag{40}$$

Taylor's theorem with the Lagrange form of the remainder ensures the existence of $\bar{t}_k^i \in (0,1)$ and $\bar{t}_k^j \in (0,1)$ such that

$$\bar{\boldsymbol{x}}_i^k := \boldsymbol{x}^k + \frac{\bar{t}_k^i \bar{\mu}_i^k}{\rho_k \|\nabla c_i(\boldsymbol{x}^k)\|^2} \nabla c_i(\boldsymbol{x}^k), \tag{41}$$

$$\bar{\boldsymbol{x}}_j^k := \boldsymbol{x}^k + \frac{\bar{t}_k^j \operatorname{sign}(\hat{\lambda}_j^k)|\bar{\lambda}_j^k|}{\rho_k \|\nabla h_j(\boldsymbol{x}^k)\|^2} \nabla h_j(\boldsymbol{x}^k), \tag{42}$$

$$c_i(\boldsymbol{y}_i^k) = c_i(\boldsymbol{x}^k) + \frac{\bar{\mu}_i^k}{\rho_k} \frac{\nabla c_i(\bar{\boldsymbol{x}}_i^k)^T \nabla c_i(\boldsymbol{x}^k)}{\|\nabla c_i(\boldsymbol{x}^k)\|^2}, \tag{43}$$

and

$$h_j(\boldsymbol{y}_j^k) = h_j(\boldsymbol{x}^k) + \frac{\operatorname{sign}(\hat{\lambda}_j^k)|\bar{\lambda}_j^k|}{\rho_k \|\nabla h_j(\boldsymbol{x}^k)\|^2} \frac{\nabla h_j(\bar{\boldsymbol{x}}_j^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2}. \tag{44}$$

We remark that the subindices of the vectors $\boldsymbol{y}_i^k$ and $\bar{\boldsymbol{x}}_i^k$ refer to the $i$-th inequality constraint, and not to the $i$-th component of the vector. The same notation applies to the subindices of the vectors $\boldsymbol{y}_j^k$ and $\bar{\boldsymbol{x}}_j^k$, in reference to the $j$-th equality constraint. Now, notice that, from (39), and since $\alpha_i^k = 0$, it follows that $c_i(\boldsymbol{y}_i^k) < 0$. Multiplying both sides of equation (43) by $\rho_k$, we get

$$\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k \frac{\nabla c_i(\bar{\boldsymbol{x}}_i^k)^T \nabla c_i(\boldsymbol{x}^k)}{\|\nabla c_i(\boldsymbol{x}^k)\|^2} < 0.$$

Then, by adding $\bar{\mu}_i^k$ in both sides of the inequality above, we obtain

$$\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k < \bar{\mu}_i^k \left(1 - \frac{\nabla c_i(\bar{\boldsymbol{x}}_i^k)^T \nabla c_i(\boldsymbol{x}^k)}{\|\nabla c_i(\boldsymbol{x}^k)\|^2}\right) \leq \mu_{\max} \left|1 - \frac{\nabla c_i(\bar{\boldsymbol{x}}_i^k)^T \nabla c_i(\boldsymbol{x}^k)}{\|\nabla c_i(\boldsymbol{x}^k)\|^2}\right|,$$

so that

$$\left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ < \mu_{\max} \left|1 - \frac{\nabla c_i(\bar{\boldsymbol{x}}_i^k)^T \nabla c_i(\boldsymbol{x}^k)}{\|\nabla c_i(\boldsymbol{x}^k)\|^2}\right|.$$

Therefore,

$$\hat{\mu}_i^k < \mu_{\max} \left|1 - \frac{\nabla c_i(\bar{\boldsymbol{x}}_i^k)^T \nabla c_i(\boldsymbol{x}^k)}{\|\nabla c_i(\boldsymbol{x}^k)\|^2}\right|. \tag{45}$$

To see that the multiplier associated with the $j$-th equality constraint goes to zero, for $j \in I_q$ such that $\beta_j^k = 0$ and $j \notin \Delta_{\boldsymbol{h}}(\boldsymbol{x}^*)$, two cases must be analyzed: $\hat{\lambda}_j^k > 0$ and $\hat{\lambda}_j^k < 0$. Addressing first the possibility $\hat{\lambda}_j^k > 0$, from the definition of $\beta_j^k$ stated in (34), since $\beta_j^k = 0$ and $\boldsymbol{y}_j^k \in \mathcal{B}[\boldsymbol{x}^k, \epsilon_k]$, it follows that $h_j(\boldsymbol{y}_i^k) < 0$. Multiplying both sides of equation (44) by $\rho_k$, we obtain

$$\rho_k h_j(\boldsymbol{x}^k) + |\bar{\lambda}_j^k| \frac{\nabla h_j(\bar{\boldsymbol{x}}_i^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2} < 0.$$

Hence,

$$\begin{aligned} 0 < \rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k &< \left(\bar{\lambda}_j^k - |\bar{\lambda}_j^k|\right) + |\bar{\lambda}_j^k| \left(1 - \frac{\nabla h_j(\bar{\boldsymbol{x}}_i^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2}\right) \\ &\leq |\bar{\lambda}_j^k| \left(1 - \frac{\nabla h_j(\bar{\boldsymbol{x}}_i^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2}\right). \end{aligned}$$

Assuming now that $\hat{\lambda}_j^k < 0$, with the same previous reasoning we must have $h_j(\boldsymbol{y}_i^k) > 0$, and from (44) we get

$$\rho_k h_j(\boldsymbol{x}^k) - |\bar{\lambda}_j^k| \frac{\nabla h_j(\bar{\boldsymbol{x}}_i^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2} > 0.$$

Thus,

$$\begin{aligned} 0 > \rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k &> \left(|\bar{\lambda}_j^k| - \bar{\lambda}_j^k\right) - |\bar{\lambda}_j^k| \left(1 - \frac{\nabla h_j(\bar{\boldsymbol{x}}_i^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2}\right) \\ &\geq - |\bar{\lambda}_j^k| \left(1 - \frac{\nabla h_j(\bar{\boldsymbol{x}}_i^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2}\right). \end{aligned}$$

Consequently, in both cases we have

$$|\rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k| < |\bar{\lambda}_j^k| \left|1 - \frac{\nabla h_j(\bar{\boldsymbol{x}}_i^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2}\right|.$$

As a result, the following bound holds

$$|\hat{\lambda}_j^k| < \max\{|\lambda_{\max}|, |\lambda_{\min}|\} \left|1 - \frac{\nabla h_j(\bar{\boldsymbol{x}}_i^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2}\right|. \tag{46}$$

For sufficiently large $k \in \mathcal{K}$, and due to the inequalities (22), (35), (38), (45) and (46), we have

$$
\begin{aligned}
\left\|\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)\right\| &= \left\| \nabla f(\boldsymbol{x}^k) + \sum_{i=1}^{p} \mu_i^k \nabla c_i(\boldsymbol{x}^k) \ + \ \sum_{j=1}^{q} \lambda_j^k \nabla h_j(\boldsymbol{x}^k) \right\| \\
= \quad & \left\| \nabla f(\boldsymbol{x}^k) + \sum_{i|\alpha_i^k=1} \hat{\mu}_i^k \nabla c_i(\boldsymbol{x}^k) \ + \sum_{j|\beta_j^k=1} \hat{\lambda}_j^k \nabla h_j(\boldsymbol{x}^k) \right\| \\
\leq \quad & \left\| \nabla f(\boldsymbol{x}^k) + \sum_{i=1}^{p} \hat{\mu}_i^k \nabla c_i(\boldsymbol{x}^k) \ + \ \sum_{j=1}^{q} \hat{\lambda}_j^k \nabla h_j(\boldsymbol{x}^k) \right\| \\
& + \left\| \sum_{i|\alpha_i^k=0} \hat{\mu}_i^k \nabla c_i(\boldsymbol{x}^k) \right\| + \left\| \sum_{j|\beta_j^k=0} \hat{\lambda}_j^k \nabla h_j(\boldsymbol{x}^k) \right\| \\
\overset{(22)}{\leq} \quad & \theta_k + \left\| \sum_{i|\alpha_i^k=0} \hat{\mu}_i^k \nabla c_i(\boldsymbol{x}^k) \right\| + \left\| \sum_{j|\beta_j^k=0} \hat{\lambda}_j^k \nabla h_j(\boldsymbol{x}^k) \right\| \\
\leq \quad & \theta_k + \sum_{\substack{i \ | \ \alpha_i^k = 0 \\ i \in \Delta(\boldsymbol{x}^*)}} \hat{\mu}_i^k \|\nabla c_i(\boldsymbol{x}^k)\| + \sum_{\substack{i \ | \ \alpha_i^k = 0 \\ i \notin \Delta(\boldsymbol{x}^*)}} \hat{\mu}_i^k \|\nabla c_i(\boldsymbol{x}^k)\| \\
& + \sum_{\substack{j \ | \ \beta_j^k = 0 \\ j \in \Delta_h(\boldsymbol{x}^*)}} |\hat{\lambda}_j^k| \|\nabla h_j(\boldsymbol{x}^k)\| + \sum_{\substack{j \ | \ \beta_j^k = 0 \\ j \notin \Delta_h(\boldsymbol{x}^*)}} |\hat{\lambda}_j^k| \|\nabla h_j(\boldsymbol{x}^k)\| \\
\overset{\substack{(35),(38) \\ (45),(46)}}{\leq} \quad & \theta_k + \sum_{\substack{i \ | \ \alpha_i^k = 0 \\ i \in \Delta(\boldsymbol{x}^*)}} \mu_{\max} \|\nabla c_i(\boldsymbol{x}^k)\| \\
& + \sum_{\substack{i \ | \ \alpha_i^k = 0 \\ i \notin \Delta(\boldsymbol{x}^*)}} \mu_{\max} \left| 1 - \frac{\nabla c_i(\bar{\boldsymbol{x}}_i^k)^T \nabla c_i(\boldsymbol{x}^k)}{\|\nabla c_i(\boldsymbol{x}^k)\|^2} \right| \|\nabla c_i(\boldsymbol{x}^k)\| \\
& + \sum_{\substack{j \ | \ \beta_j^k = 0 \\ j \in \Delta_h(\boldsymbol{x}^*)}} \max\{|\lambda_{\max}|, |\lambda_{\min}|\} \|\nabla h_j(\boldsymbol{x}^k)\| \\
& + \sum_{\substack{j \ | \ \beta_j^k = 0 \\ j \notin \Delta_h(\boldsymbol{x}^*)}} \max\{|\lambda_{\max}|, |\lambda_{\min}|\} \left| 1 - \frac{\nabla h_j(\bar{\boldsymbol{x}}_j^k)^T \nabla h_j(\boldsymbol{x}^k)}{\|\nabla h_j(\boldsymbol{x}^k)\|^2} \right| \|\nabla h_j(\boldsymbol{x}^k)\|.
\end{aligned}
$$

As the right-hand side of the above inequality converges to zero for $k \in \mathcal{K}$, the stationarity condition (4) is attained, concluding the proof. □

It is worth noticing that the modified multipliers defined within the proof of Theorem 3.1 (cf. equation (32)) might not necessarily fulfill the image complementary condition (7). Nevertheless, in the next result, following [6], and under the assumption that the infeasibility measure (8) satisfies the Kurdyka-Lojasiewicz inequality [15, 16], limit points of the algorithm ALM are shown to verify the DICAKKT condition.

**Theorem 3.2.** *Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a limit point of the sequence generated by the algorithm ALM, which is feasible for problem* (1)*. Assume that, for the function $\Phi$ defined in* (8)*, there exist $\delta > 0$ and $\varphi : \mathcal{B}[\boldsymbol{x}^*, \delta] \to \mathbb{R}$ such that*

$$\|\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{x}^*)\| \leq \varphi(\boldsymbol{x}) \|\nabla \Phi(\boldsymbol{x})\|, \tag{47}$$

*for all $\boldsymbol{x} \in \mathcal{B}[\boldsymbol{x}^*, \delta]$, with $\lim_{\boldsymbol{x} \to \boldsymbol{x}^*} \varphi(\boldsymbol{x}) = 0$. Then, $\boldsymbol{x}^*$ fulfils the DICAKKT condition.*

*Proof.* We start with the sequences generated by the algorithm ALM, namely $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$, $\{\rho_k\}_{k\in\mathbb{N}}$, $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}}$, and $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}}$. From the hypothesis, there exists an infinite set $\mathcal{K} \subset \mathbb{N}$ such that $\lim_{k\in\mathcal{K}} \boldsymbol{x}^k = \boldsymbol{x}^*$. The sequence of penalty parameters $\{\rho_k\}_{k\in\mathbb{N}}$ is either bounded or unbounded. Under the former assumption, the validity of the DICAKKT condition at the point $\boldsymbol{x}^*$ follows from Lemma 3.1 and Proposition 2.1.

Assume now that $\{\rho_k\}_{k\in\mathbb{N}}$ is unbounded. Let the indicator function defined in (31), and the sequences $\{\epsilon_k\}_{k\in\mathbb{N}}$, $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}}$ and $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{N}}$, given by (30) and (32), respectively. The validity of the conditions (4), (5) and (6) may be achieved following the same steps of the proof of Theorem 3.1.

To see that the image complementary condition (7) holds, first notice that, for all $k \in \mathcal{K}$, we have

$$\begin{aligned}
\rho_k \|\nabla \Phi(\boldsymbol{x}^k)\| &= \left\| \sum_{i=1}^p \rho_k c_i(\boldsymbol{x}^k)_+ \nabla c_i(\boldsymbol{x}^k) + \sum_{j=1}^q \rho_k h_j(\boldsymbol{x}^k) \nabla h_j(\boldsymbol{x}^k) \right\| \\
&= \left\| \sum_{i=1}^p \left[ \rho_k c_i(\boldsymbol{x}^k)_+ - \left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ \right] \nabla c_i(\boldsymbol{x}^k) \right. \\
&\quad + \sum_{j=1}^q \left[ \rho_k h_j(\boldsymbol{x}^k) - \left(\rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k\right) \right] \nabla h_j(\boldsymbol{x}^k) \\
&\quad + \sum_{i=1}^p \left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ \nabla c_i(\boldsymbol{x}^k) \\
&\quad \left. + \sum_{j=1}^q \left(\rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k\right) \nabla h_j(\boldsymbol{x}^k) + \nabla f(\boldsymbol{x}^k) - \nabla f(\boldsymbol{x}^k) \right\| \\
&\leq \sum_{i=1}^p \left| \rho_k c_i(\boldsymbol{x}^k)_+ - \left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ \right| \|\nabla c_i(\boldsymbol{x}^k)\| \\
&\quad \sum_{j=1}^q \left| \rho_k h_j(\boldsymbol{x}^k) - \left(\rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k\right) \right| \|\nabla h_j(\boldsymbol{x}^k)\|
\end{aligned}$$

$$
\begin{aligned}
&+ \left\| \nabla f(\boldsymbol{x}^k) + \sum_{i=1}^{p} \left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ \nabla c_i(\boldsymbol{x}^k) \right. \\
&+ \left. \sum_{j=1}^{q} \left(\rho_k h_j(\boldsymbol{x}^k) + \bar{\lambda}_j^k\right) \nabla h_j(\boldsymbol{x}^k) \right\| + \|\nabla f(\boldsymbol{x}^k)\| \\
\overset{(22)}{\leq} & \sum_{i=1}^{p} \bar{\mu}_i^k \|\nabla c_i(\boldsymbol{x}^k)\| + \sum_{j=1}^{q} \left|\bar{\lambda}_j^k\right| \|\nabla h_j(\boldsymbol{x}^k)\| + \theta_k + \|\nabla f(\boldsymbol{x}^k)\|,
\end{aligned}
$$

with the non-expansiveness of the application $(\cdot)_+$ also being used in the last inequality. As a result, there exists $M > 0$ such that, for all $k \in \mathcal{K}$, it holds $\|\rho_k \nabla \Phi(\boldsymbol{x}^k)\| \leq M$. This bound, together with (47) and the feability of $\boldsymbol{x}^*$ implies

$$
\|\rho_k \Phi(\boldsymbol{x}^k)\| \leq \varphi(\boldsymbol{x}^k) \|\rho_k \nabla \Phi(\boldsymbol{x}^k)\| \leq M \varphi(\boldsymbol{x}^k) \underset{k \in \mathcal{K}}{\to} 0.
$$

Hence, since

$$
\rho_k \Phi(\boldsymbol{x}^k) = \frac{1}{2} \left(\|\sqrt{\rho_k} \boldsymbol{c}(\boldsymbol{x}^k)_+\|^2 + \|\sqrt{\rho_k} \boldsymbol{h}(\boldsymbol{x}^k)\|^2\right) \quad \text{and} \quad \rho_k \Phi(\boldsymbol{x}^k) \underset{k \in \mathcal{K}}{\to} 0,
$$

then each component of the sequences $\{\sqrt{\rho_k} \boldsymbol{c}(\boldsymbol{x}^k)_+\}_{k \in \mathcal{K}}$ and $\{\sqrt{\rho_k} \boldsymbol{h}(\boldsymbol{x}^k)\}_{k \in \mathcal{K}}$ converges to zero, so that

$$
\lim_{k \in \mathcal{K}} \rho_k c_i(\boldsymbol{x}^k)_+^2 = 0, \ \forall i \in I_p \quad \text{and} \quad \lim_{k \in \mathcal{K}} \rho_k h_j(\boldsymbol{x}^k)^2 = 0, \ \forall j \in I_q. \tag{48}
$$

For each $i \in I_p$, if $i \in \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$, from the definition (32),

$$
\begin{aligned}
\left|\mu_i^k c_i(\boldsymbol{x}^k)_+ - \rho_k c_i(\boldsymbol{x}^k)_+^2\right| &= \left|\left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ c_i(\boldsymbol{x}^k)_+ - \rho_k c_i(\boldsymbol{x}^k)_+^2\right| \\
&\leq \left|\left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right)_+ - \rho_k c_i(\boldsymbol{x}^k)_+\right| c_i(\boldsymbol{x}^k)_+ \\
&\leq \left|\left(\rho_k c_i(\boldsymbol{x}^k) + \bar{\mu}_i^k\right) - \rho_k c_i(\boldsymbol{x}^k)\right| c_i(\boldsymbol{x}^k)_+ \\
&\leq \bar{\mu}_i^k c_i(\boldsymbol{x}^k)_+.
\end{aligned}
$$

If $i \notin \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$, again from the definition (32), since $c_i(\boldsymbol{x}^k) < 0$ in this case, we have

$$
\left|\mu_i^k c_i(\boldsymbol{x}^k)_+ - \rho_k c_i(\boldsymbol{x}^k)_+^2\right| = 0 \leq \bar{\mu}_i^k c_i(\boldsymbol{x}^k)_+.
$$

Therefore, in both cases,

$$
\left|\mu_i^k c_i(\boldsymbol{x}^k)_+ - \rho_k c_i(\boldsymbol{x}^k)_+^2\right| \leq \bar{\mu}_i^k c_i(\boldsymbol{x}^k)_+, \quad \forall i \in I_p. \tag{49}
$$

Now, for each $j \in I_q$, we have

$$
\begin{aligned}
|\lambda_j^k h_j(\boldsymbol{x}^k)| &\overset{(34)}{=} \|\beta_j^k \left(\bar{\lambda}_j^k + \rho_k h_j(\boldsymbol{x}^k)\right) h_j(\boldsymbol{x}^k)| \\
&\overset{\beta_j^k \in \{0,1\}}{\leq} |\left(\bar{\lambda}_j^k + \rho_k h_j(\boldsymbol{x}^k)\right) h_j(\boldsymbol{x}^k)| \\
&\leq |\bar{\lambda}_j^k h_j(\boldsymbol{x}^k)| + \rho_k h_j(\boldsymbol{x}^k)^2.
\end{aligned}
$$

By the limit on the right of (48) and the feasibility of $\boldsymbol{x}^*$, the upper bound above converges to zero. Hence, by the squeeze theorem,

$$\lim_{k\in\mathcal{K}} \lambda_j^k h_j(\boldsymbol{x}^k) = 0, \quad \forall j \in I_q. \tag{50}$$

On the other hand, because of (48), (49) and from the feasibility of $\boldsymbol{x}^*$, we obtain

$$\lim_{k\in\mathcal{K}} \mu_i^k c_i(\boldsymbol{x}^k)_+ = 0. \tag{51}$$

Define $z_k := \max_{i\in I_p} \left\{\left|\mu_i^k c_i(\boldsymbol{x}^k)\right|\right\}$ for every $k \in \mathcal{K}$. There exist an index $s \in I_p$ and an infinite set $\mathcal{K}_1 \subset \mathcal{K}$ such that $z_k = \left|\mu_s^k c_s(\boldsymbol{x}^k)\right|$ for all $k \in \mathcal{K}_1$. Recalling that we are under the assumption that $\{\rho_k\}_{k\in\mathbb{N}}$ is unbounded, to conclude the proof we must address the following two cases:

(i) There exists an infinite set $\mathcal{K}_2 \subset \mathcal{K}_1$ such that $c_s(\boldsymbol{x}^k) \geq 0$ for all $k \in \mathcal{K}_2$. In this case, the limit (50) implies that, for a given $i \in I_p$ and $k \in \mathcal{K}_2$,

$$0 \leq \left|\mu_i^k c_i(\boldsymbol{x}^k)\right| \leq z_k = \max_{i\in I_p} \left\{\left|\mu_i^k c_i(\boldsymbol{x}^k)\right|\right\} = \mu_s^k c_s(\boldsymbol{x}^k) = \mu_s^k c_s(\boldsymbol{x}^k)_+ \underset{k\in\mathcal{K}_2}{\longrightarrow} 0.$$

Therefore, together with the limit (51), we obtain that $\boldsymbol{x}^*$ verifies (7) with the sequences associated with the set $\mathcal{K}_2$.

(ii) For every $k \in \mathcal{K}_1$ and large enough, $c_s(\boldsymbol{x}^k) < 0$. The analysis is subdivided in two possibilities: (a) $c_s(\boldsymbol{x}^*) < 0$ or (b) $c_s(\boldsymbol{x}^*) = 0$. If (a) holds, from the unboundedness of $\{\rho_k\}_{k\in\mathbb{N}}$, and since $\{\bar{\mu}_s^k\}_{k\in\mathcal{K}_1}$ is bounded, we have $\lim_{k\in\mathcal{K}_1} \rho_k c_s(\boldsymbol{x}^k) + \bar{\mu}_s^k = -\infty$. Hence, $\mu_s^k = 0$ and $c_s(\boldsymbol{x}^k)\mu_s^k = 0$ for $k$ sufficiently large. In case (b) holds, we have $\mu_s^k = \left(\rho_k c_s(\boldsymbol{x}^k) + \bar{\mu}_s^k\right)_+ \leq \bar{\mu}_s^k$ since $c_s(\boldsymbol{x}^k) < 0$ for $k \in \mathcal{K}_1$ and large enough. Notice that, for $k \in \mathcal{K}_1$,

$$0 \leq |\mu_s^k c_s(\boldsymbol{x}^k)| \leq \bar{\mu}_s^k |c_s(\boldsymbol{x}^k)| \to 0.$$

As a result, together with the limit (51), we also obtain that $\boldsymbol{x}^*$ satisfies (7), now with the sequences associated with the set $\mathcal{K}_1$.

Summing up, the limit point $\boldsymbol{x}^*$ fulfills the DICAKKT condition, and the proof is complete. □

**Remark 3.1.** *Despite not being our initial proposal, whenever the multipliers defined in* (32) *are unbounded, and following the lines of the proof of [1, Theorem 4.1.], it is possible to show that the sequence* $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ *is also a PAKKT sequence, i.e., the sequence* $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ *is a DICAKKT+PAKKT sequence. To obtain a similar result with bounded multipliers, one must observe that the DCAKKT condition holds with bounded multipliers. Therefore, taking a subsequence, if necessary, and using the relations* (4) *and* (5)*, it is not hard to see that the KKT conditions hold. Thus, it is possible to construct a DICAKKT+PAKKT sequence. Indeed, the sequence built up in [1, Lemma 2.6] is a DICAKKT+PAKKT sequence. Now, in view of Example 3.1 coming ahead, which also shows that DICAKKT + PAKKT is not implied by PAKKT, this Remark strengthens [1, Theorem 4.1].*

**Remark 3.2.** *It should be mentioned that, recently, the strongest sequential optimality condition that can be linked to the safeguarded augmented Lagrangian method – called AL-AKKT – has been established [4], see also [3]. Therefore, since every limit point of the algorithm ALM is a DCAKKT point, this yields that the DCAKKT condition is implied by the AL-AKKT. However, this does not exclude DCAKKT from having new implications for the convergence theory of optimization methods. By being the strongest SOC of safeguarded augmented Lagrangian methods, the AL-AKKT condition is intrinsically connected to the algorithm ALM and, therefore, it should not be expected to have convergence implications for methods that are not based on augmented Lagrangian techniques. On the other hand, the DCAKKT is flexible enough to be applied outside the context of augmented Lagrangian methods – for example, limit points of the PACNO algorithm are also DCAKKT points whenever PACNO is used for solving smooth optimization problems. Additionally, as we will see in the next section, DCAKKT is strong enough to imply the Strong AGP condition.*

Next, we provide an example ensuring that the *positive complementary approximate KKT* (PCAKKT) condition (cf. [7]), which is strictly stronger than combining the elements from PAKKT [1] and CAKKT [6], does not imply the DCAKKT condition. Since the PCAKKT strictly implies not only PAKKT and CAKKT, but also the *approximate gradient projection* (AGP) condition [18] – see [7] for the details –, and each of these conditions is stronger than the *approximate KKT* (AKKT, cf. [2]), the next example aims at contextualizing our contribution.

**Example 3.1** (PCAKKT does not imply DCAKKT)**.** *Consider the problem*

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}\in\mathbb{R}^2} & x_1 - x_2 \\ s.t. & -x_1 \leq 0,\ -x_2 \leq 0,\ x_1x_2 \leq 0. \end{array}$$

*Despite not being a minimizer, the origin verifies the PCAKKT conditions (see [7, Definition 2.3]). Indeed, by considering the sequences* $\boldsymbol{x}^k = (1/k, 1/k^2)^T$ *and* $\boldsymbol{\mu}^k = (2, k-2, k^2-k)^T$*, notice that*

$$\begin{aligned} \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}^k, \boldsymbol{\mu}^k) &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \mu_1^k \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \mu_2^k \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \mu_3^k \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + (k-2) \begin{bmatrix} 0 \\ -1 \end{bmatrix} + (k^2-k) \begin{bmatrix} 1/k^2 \\ 1/k \end{bmatrix} \\ &= \begin{bmatrix} -1/k \\ 0 \end{bmatrix} \overset{k\to\infty}{\longrightarrow} \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

*Moreover, the following relations hold*

$$\begin{aligned} &\frac{\mu_1^k}{\max\{1, \|\boldsymbol{\mu}^k\|_\infty\}} = \frac{2}{k^2-k} \to 0, \quad \frac{\mu_2^k}{\max\{1, \|\boldsymbol{\mu}^k\|_\infty\}} = \frac{k-2}{k^2-k} \to 0, \ \textit{and} \\ &\frac{\mu_3^k}{\max\{1, \|\boldsymbol{\mu}^k\|_\infty\}} = \frac{k^2-k}{k^2-k} = 1 \to 1, \ \textit{with } c_3(\boldsymbol{x}^k) = \frac{1}{k^3} > 0, \forall k \in \mathbb{N}, \end{aligned}$$

*and the complementarity is verified as follows*

$$\mu_1^k c_1(\boldsymbol{x}^k) = 2\left(\frac{-1}{k}\right) \to 0,$$
$$\mu_2^k c_2(\boldsymbol{x}^k) = (k-2)\left(\frac{-1}{k^2}\right) \to 0, \quad \textit{and}$$
$$\mu_3^k c_3(\boldsymbol{x}^k) = \left(k^2 - k\right)\left(\frac{1}{k^3}\right) \to 0.$$

*Thus, $x^* = (0,0)^T$ satisfies the PCAKKT condition.*

*The same reasoning of [12, Example 3.4] concerning the $\epsilon$-ASOC may be used to show that the origin does not fulfill the DCAKKT condition.*

As every DICAKKT point also satisfies the DCAKKT conditions, Example 3.1 additionally evinces that there exist PCAKKT points that are not DICAKKT. When it comes to the extra requirement that distinguishes DICAKKT from DCAKKT, namely the complementarity (7), a pertinent question arises: would such a requirement make DICAKKT strong enough to imply PCAKKT? The next example gives a negative answer to this question.

**Example 3.2** (DICAKKT does not imply PAKKT)**.** *Consider the problem*

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}\in\mathbb{R}^2} & x_1 - x_2 \\ s.t. & \frac{-x_1^2}{2} + x_2 \leq 0, \quad \frac{-x_1^2}{2} - x_2 \leq 0. \end{array}$$

*The origin, a nonregular point, satisfies DICAKKT. In fact, taking the sequences $\boldsymbol{x}^k = (1/k, 0)^T$ and $\boldsymbol{\mu}^k = (k/2, (k-2)/2)^T$, we have*

$$\begin{aligned} \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^k, \boldsymbol{\mu}^k) &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \mu_1^k \begin{bmatrix} -x_1^k \\ 1 \end{bmatrix} + \mu_2^k \begin{bmatrix} -x_1^k \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \frac{k}{2} \begin{bmatrix} -1/k \\ 1 \end{bmatrix} + \frac{k-2}{2} \begin{bmatrix} -1/k \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 1/k \\ 0 \end{bmatrix} \overset{k\to\infty}{\longrightarrow} \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \end{aligned}$$

*that is,* (4) *holds. Setting $\epsilon_k = 1/k^2$, we obtain*

$$c_i\left(\boldsymbol{x}^k + (-1)^{1+i}\epsilon_k \boldsymbol{e}_2\right) = \frac{-1}{2k^2} + (-1)^{1+i}(-1)^{1+i}\frac{1}{k^2} = \frac{1}{2k^2} > 0,$$

*for $i \in \{1,2\}$ and $\boldsymbol{e}_2 = (0\ 1)^T$. Hence, for all $k \in \mathbb{N}$, we have $i \in \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$ for $i \in \{1,2\}$, i.e.,* (5) *is verified. Moreover,*

$$\epsilon_k \mu_1^k = \frac{1}{k^2}\left(\frac{k}{2}\right) \to 0 \quad \textit{and} \quad \epsilon_k \mu_2^k = \frac{1}{k^2}\left(\frac{k-2}{2}\right) \to 0,$$

*so that* (6) *holds. Finally, from*

$$\mu_1^k c_1(\boldsymbol{x}^k) = \frac{k}{2}\left(\frac{-1}{2k^2} + 0\right) \to 0 \quad \text{and} \quad \mu_2^k c_2(\boldsymbol{x}^k) = \frac{k-2}{2}\left(\frac{-1}{2k^2} - 0\right) \to 0,$$

*the complementarity* (7) *is obtained.*

*Now, assume by contradiction, that the origin satisfies PAKKT. Therefore, there exist* $\boldsymbol{x}^k = [x_1^k, x_k^2]^T \to \mathbf{0}$ *and* $\boldsymbol{\mu}^k = \left[\mu_1^k, \mu_2^k\right]^T \in \mathbb{R}_+^2$ *such that*

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^k, \boldsymbol{\mu}^k) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \mu_1^k \begin{bmatrix} -x_1^k \\ 1 \end{bmatrix} + \mu_2^k \begin{bmatrix} -x_1^k \\ -1 \end{bmatrix} \to \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{52}$$

$$\lim_{k \in \mathbb{N}} \max\{c_i(\boldsymbol{x}^k), -\mu_i^k\} = 0, \text{ for } i \in \{1, 2\}, \tag{53}$$

*and*

$$\mu_i^k c_i(\boldsymbol{x}^k) > 0, \text{ if } \lim_{k \in \mathbb{N}} \frac{\mu_i^k}{\max\{1, \|\boldsymbol{\mu}^k\|_\infty\}} > 0, \text{ for } i \in \{1, 2\}. \tag{54}$$

*The first row of expression* (52) *implies that* $\lim_{k \in \mathbb{N}} \|\boldsymbol{\mu}^k\|_\infty = +\infty$*. Thus, dividing the second row of* (52) *by* $\max\{1, \|\boldsymbol{\mu}^k\|_\infty\}$ *and taking the limit for* $k \in \mathbb{N}$*, we obtain*

$$\lim_{k \in \mathbb{N}} \frac{\mu_1^k}{\max\{1, \|\boldsymbol{\mu}^k\|_\infty\}} = \lim_{k \in \mathbb{N}} \frac{\mu_2^k}{\max\{1, \|\boldsymbol{\mu}^k\|_\infty\}}.$$

*But, for* $k$ *large enough,* $\mu_1^k / \max\{1, \|\boldsymbol{\mu}^k\|_\infty\} = 1$ *or* $\mu_2^k / \max\{1, \|\boldsymbol{\mu}^k\|_\infty\} = 1$*, so that the above limits are equal to 1. Therefore, the relations* (54) *yield* $-(x_1^k)^2/2 + x_2^k > 0$ *and* $-(x_1^k)^2/2 - x_2^k > 0$ *for all* $k \in \mathbb{N}$ *and sufficiently large. As a result,* $x_2^k > (x_1^k)^2/2 \geq -(x_1^k)^2/2 > x_2^k$ *for all large enough* $k \in \mathbb{N}$*, leading to a contradiction. Consequently, the origin does not satisfy PAKKT.*

Notice that, as every PCAKKT point is a PAKKT point, Example 3.2 also indicates the existence of DICAKKT points that do not verify the PCAKKT conditions. Moreover, the above examples show that DCAKKT and PAKKT are independent conditions, because neither PCAKKT implies DCAKKT nor DICAKKT implies PAKKT.

As a result, the proposed DCAKKT condition is not implied by any already known sequential optimality condition, namely AKKT, AGP, CAKKT, and PAKKT, nor by possible combinations of their key elements. This happens because PCAKKT is strictly stronger than any possible combination of the already known conditions, see [7, Fig.2]. To properly place DCAKKT within the landscape of the sequential optimality conditions from the literature for problem (1), we analyze next its relationship with the *strong AGP* (SAGP), established in [18].

# 4 Strong AGP is strictly weaker than DCAKKT

As the approximate gradient projection is essentially a property of the sequence of *primal* approximations of the stationary point of problem (1), we start with an equivalent definition of the *strong AGP* (SAGP), to include multipliers in the analysis.

**Definition 4.1** (strong AGP)**.** *A feasible point* $\boldsymbol{x}^*$ *of problem* (1) *is said to satisfy the* strong approximate gradient projection *(SAGP) condition if there exist sequences* $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$*,* $\{\vartheta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^*_+$*,* $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$ *and* $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$ *such that* $\boldsymbol{x}^k \to \boldsymbol{x}^*$*,* $\vartheta_k \downarrow 0$

$$\lim_{k\in\mathbb{N}} \Big(\nabla f(\boldsymbol{x}^k) + \textstyle\sum_{i=1}^p \mu_i^k \nabla c_i(\boldsymbol{x}^k) + \sum_{j=1}^q \lambda_j^k \nabla h_j(\boldsymbol{x}^k)\Big) = \mathbf{0}. \tag{55}$$

*and, for all* $k \in \mathbb{N}$*,*

$$c_i(\boldsymbol{x}^k)\mu_i^k \geq -\vartheta_k,\ \forall i \in I_p \quad \textit{and} \quad h_j(\boldsymbol{x}^k)\lambda_j^k \geq -\vartheta_k,\ \forall j \in I_q. \tag{56}$$

The equivalence between the SAGP as stated in [18, §4] and the reformulated Definition 4.1 is proved below. In addition, the result provides a straightforward connection between the multipliers associated with the quadratic penalty method and the SAGP condition. As a corollary, one obtains that *strong AKKT* (SAKKT) [10] and CAKKT imply SAGP for inequality constrained problems.

**Proposition 4.1.** *A point* $\boldsymbol{x}^*$ *that is feasible for problem* (1) *satisfies the SAGP conditions as stated in [18] if, and only if, such a point fulfills the SAGP stated in Definition 4.1.*

*Proof.* Thanks to the discussion in [18, §4], we know that the SAGP conditions are the AGP conditions related to problem (1) whenever each equality constraint is transformed into two inequalities in the obvious way. By [7, Theorem 2.7], this means that the SAGP conditions are equivalent to the existence of multipliers $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, $\{\boldsymbol{\lambda}^k_{(-)}\}_{k\in\mathbb{N}} \subset \mathbb{R}^q_+$ and $\{\boldsymbol{\lambda}^k_{(+)}\}_{k\in\mathbb{N}} \subset \mathbb{R}^q_+$ and a sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ such that $\boldsymbol{x}^k \to \boldsymbol{x}^*$,

$$\lim_{k\in\mathbb{N}} \Big(\nabla f(\boldsymbol{x}^k) + \textstyle\sum_{i=1}^p \mu_i^k \nabla c_i(\boldsymbol{x}^k) + \sum_{j=1}^q \Big((\lambda^k_{(+)})_j - (\lambda^k_{(-)})_j\Big) \nabla h_j(\boldsymbol{x}^k)\Big) = \mathbf{0}, \tag{57}$$

and for all $j \in I_q$, $i \in I_p$,

$$\lim_{k\in\mathbb{N}} (\lambda^k_{(+)})_j \min\{h_j(\boldsymbol{x}^k), 0\} = 0 \text{ and } \lim_{k\in\mathbb{N}} (\lambda^k_{(-)})_j \min\{-h_j(\boldsymbol{x}^k), 0\} = 0, \tag{58}$$

$$\lim_{k\in\mathbb{N}} \mu_i^k \min\{c_i(\boldsymbol{x}^k), 0\} = 0. \tag{59}$$

Now, the equivalence will follow directly. Indeed, let $\boldsymbol{x}^*$ satisfy the SAGP conditions. Then, we have the existence of multipliers $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, $\{\boldsymbol{\lambda}^k_{(-)}\}_{k\in\mathbb{N}} \subset \mathbb{R}^q_-$ and $\{\boldsymbol{\lambda}^k_{(+)}\}_{k\in\mathbb{N}} \subset \mathbb{R}^q_+$, and a sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ satisfying the limits (57)-(59). Defining

$$\vartheta_k := -\sum_{i=1}^{p} \mu_i^k \min\{c_i(\boldsymbol{x}^k), 0\}$$
$$- \sum_{j=1}^{q} (\lambda_{(+)}^k)_j \min\{0, h_j(\boldsymbol{x}^k)\} - (\lambda_{(-)}^k)_j \min\{0, -h_j(\boldsymbol{x}^k)\},$$

we have $\vartheta_k \geq 0$,

$$\left((\lambda_{(+)}^k)_j - (\lambda_{(-)}^k)_j\right) h(\boldsymbol{x}^k) \geq -\vartheta_k \text{ and } \mu_i^k g_i(\boldsymbol{x}^k) \geq -\vartheta_k,$$

for all $k \in \mathbb{N}$, which imply (55) and (56).

Conversely, if there exist sequences $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\vartheta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+^*$, $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+^p$ and $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$ such that $\boldsymbol{x}^k \to \boldsymbol{x}^*$, $\vartheta_k \downarrow 0$ satisfying (55) and (56), by choosing $\boldsymbol{\lambda}_{(-)}^k := (-\boldsymbol{\lambda}^k)_+$ and $\boldsymbol{\lambda}_{(+)}^k := \boldsymbol{\lambda}_+^k$ we have that $\boldsymbol{\lambda}^k = \boldsymbol{\lambda}_{(+)}^k - \boldsymbol{\lambda}_{(-)}^k$, for all $k \in \mathbb{N}$. This means that the limit (57) holds.

Moreover, for all $j \in I_q$, $i \in I_p$ we have

$$-\vartheta_k \leq \min\{h_j(\boldsymbol{x}^k)\lambda_j^k, 0\} \leq \min\{h_j(\boldsymbol{x}^k), 0\}(\lambda_{(+)}^k)_j \leq 0,$$
$$-\vartheta_k \leq \min\{h_j(\boldsymbol{x}^k)\lambda_j^k, 0\} \leq \min\{-h_j(\boldsymbol{x}^k), 0\}(\lambda_{(-)}^k)_j \leq 0,$$
$$\text{and } -\vartheta_k \leq \min\{c_i(\boldsymbol{x}^k)\mu_i^k, 0\} = \min\{c_i(\boldsymbol{x}^k), 0\}\mu_i^k \leq 0,$$

for all $k \in \mathbb{N}$. Therefore, the squeeze theorem yields

$$\lim_{k\in\mathbb{N}} \min\{h_j(\boldsymbol{x}^k), 0\}(\lambda_{(+)}^k)_j = 0,$$
$$\lim_{k\in\mathbb{N}} \min\{-h_j(\boldsymbol{x}^k), 0\}(\lambda_{(-)}^k)_j = 0,$$
$$\text{and } \lim_{k\in\mathbb{N}} \min\{c_i(\boldsymbol{x}^k), 0\}\mu_i^k = 0,$$

which prove (58) and (59), and consequently, the validity of the original SAGP conditions. This concludes the proof. □

The reformulated SAGP condition of Definition 4.1 is shown to be weaker than the DCAKKT condition in the next result, which also provides an explicit relationship between the multipliers of both conditions. This is a novelty as, up to our knowledge, it had not been demonstrated yet that the augmented Lagrangian method can generate limit points satisfying SAGP without assuming that the infeasibility measure (8) fulfills the Kurdyka-Lojasiewickz inequality [6, 15, 16]. Up to now, the SAGP has only been obtained under the assumption of CAKKT. Moreover, as a corollary, the safeguarded augmented Lagrangian method is guaranteed to converge to minimizers of convex problems just supposing continuity of the gradients of the constraints [18, Theorem 3.1].

**Theorem 4.1.** *Let $\boldsymbol{x}^*$ be a feasible point of problem* (1)*, and assume that the objective function and all the functions that describe the feasible set of problem* (1) *are locally Lipschitz continuous and differentiable around $\boldsymbol{x}^*$. If $\boldsymbol{x}^*$ satisfies the DCAKKT condition, then the multipliers associated with such a condition ensure that the reformulated SAGP conditions are verified.*

*Proof.* Assuming $\boldsymbol{x}^*$ to be a feasible point of problem (1) satisfying the DCAKKT condition, it follows that there exist $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$ and $\{\epsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^*_+$ such that $\boldsymbol{x}^k \to \boldsymbol{x}^*$, $\epsilon_k \downarrow 0$, and the relationships (4)–(6) hold. We must show that there exists a sequence $\{\vartheta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^*_+$ such that, for each $k \in \mathbb{N}$,

$$c_i(\boldsymbol{x}^k)\mu_i^k \geq -\vartheta_k \text{ for all } i \in I_p \quad \text{ and } \quad h_j(\boldsymbol{x}^k)\lambda_j^k \geq -\vartheta_k \text{ for all } j \in I_q.$$

By considering the local Lipschitz constants for the constraints $c_i$ and $h_j$ around $\boldsymbol{x}^*$, respectively, given by $L_{c_i}$ and $L_{h_j}$, $\vartheta_k$ may be taken, for each $k \in \mathbb{N}$ as

$$\vartheta_k := \max_{i\in I_p, j\in I_q}\{L_{c_i}, L_{h_j}\}\left(\|\boldsymbol{\lambda}^k\| + \|\boldsymbol{\mu}^k\|\right)\epsilon_k,$$

which converges to zero by the limits (6).

Let us start by analyzing the inequality constraints. For a given $k \in \mathbb{N}$, two possibilities must be addressed for $i \in I_p$: either $i \notin \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$ or $i \in \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$. In case $i \notin \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$, from (5) we have $\mu_i^k = 0$, and thus $\mu_i^k c_i(\boldsymbol{x}^k) = 0 \geq -\vartheta_k$. For $i \in \mathcal{I}_{\epsilon_k}(\boldsymbol{x}^k)$, from the definition (2), there exists $\boldsymbol{x}_i^k$ such that $c_i(\boldsymbol{x}_i^k) \geq 0$ and $\|\boldsymbol{x}^k - \boldsymbol{x}_j^k\| \leq \epsilon_k$. Consequently,

$$\begin{aligned}
\mu_i^k c_i(\boldsymbol{x}^k) &= \mu_i^k(c_i(\boldsymbol{x}^k) - c_i(\boldsymbol{x}_i^k)) + \mu_i^k c_i(\boldsymbol{x}_i^k) \\
&\geq \mu_i^k(c_i(\boldsymbol{x}^k) - c_i(\boldsymbol{x}_i^k)) \\
&\geq -L_{c_i}\mu_i^k\|\boldsymbol{x}^k - \boldsymbol{x}_i^k\| \\
&\geq -\max_{i\in I_p, j\in I_q}\{L_{c_i}, L_{h_j}\}\mu_i^k\epsilon_k \\
&\geq -\max_{i\in I_p, j\in I_q}\{L_{c_i}, L_{h_j}\}\left(\|\boldsymbol{\lambda}^k\| + \|\boldsymbol{\mu}^k\|\right)\epsilon_k. \\
&= -\vartheta_k,
\end{aligned}$$

so in both cases the desired inequality holds.

Now, for the multipliers associated with the equality constraints, let an arbitrary $k \in \mathbb{N}$. Then, for $j \in I_q$, either $j \notin \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k)$ or $j \in \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k)$. If $j \notin \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k)$, from (5), we have $h_j(\boldsymbol{x}^k)\lambda_i^k \geq 0$, and thus $h_j(\boldsymbol{x}^k)\lambda_i^k \geq 0 \geq -\vartheta_k$. If $j \in \mathcal{J}_{\epsilon_k}(\boldsymbol{x}^k)$, from (3) there exists $\boldsymbol{x}_j^k$ such that $h_j(\boldsymbol{x}_j^k) = 0$ and $\|\boldsymbol{x}^k - \boldsymbol{x}_j^k\| \leq \epsilon_k$, so that, for each $j$,

$$\begin{aligned}
\lambda_j^k h_j(\boldsymbol{x}^k) &= \lambda_j^k(h_j(\boldsymbol{x}^k) - h_j(\boldsymbol{x}_j^k)) \\
&\geq -L_{h_j}\lambda_j^k\|\boldsymbol{x}^k - \boldsymbol{x}_j^k\| \\
&\geq -L_{h_j}\lambda_j^k\epsilon_k
\end{aligned}$$

$$\geq -\max_{i \in I_p, j \in I_q} \{L_{c_i}, L_{h_j}\} \left( \|\boldsymbol{\lambda}^k\| + \|\boldsymbol{\mu}^k\| \right) \epsilon_k$$
$$= -\vartheta_k.$$

In both cases the aimed inequality is valid, and the proof is complete. □

**Remark 4.1.** *Concerning the assumptions made in Theorem 4.1 about the functions that describe the feasible set of problem* (1) *being locally Lipschitz continuous and differentiable around* $\boldsymbol{x}^*$*, a few words are in order. First, these assumptions imply that the related gradients are bounded around* $\boldsymbol{x}^*$*. Conversely, the boundedness of the gradient around a point ensures local Lipschitz continuity of the function. Hence, under differentiability, local Lipschitz continuity and local boundedness of the gradient are equivalent. Nevertheless, let the function* $\phi : \mathbb{R} \to \mathbb{R}$ *be such that* $\phi(0) = 0$ *and* $\phi(x) = x^2 \sin(\exp(1/x^2))$ *for all* $x \neq 0$*, which is differentiable, but not locally Lipschitz continuous around the origin. Consequently, the assumptions of the theorem are not redundant, providing slightly more generality to the result.*

To enlighten our motivation to create the new DCAKKT taking into consideration the already defined $\epsilon$-ASOC (cf. [11, Def. 2.1]), two reasons are in order. First, in the equality constrained case, the $\epsilon$-ASOC is equivalent to the AKKT condition, which is the weakest and broadest sequential optimality condition known. Hence, $\epsilon$-ASOC does not achieve our aim of having a strong, yet general, optimality condition. The second and more important reason is that the $\epsilon$-ASOC does not imply the SAGP condition, as shown next.

**Example 4.1** ($\epsilon$-ASOC does not imply SAGP)**.** *Consider the problem given in [12, Example 3.5.]:*

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} \frac{(x_2 - 2)^2}{2} \quad s.t \quad x_1 = 0, \; x_1 x_2 = 0.$$

*The point* $\boldsymbol{x}^* = (0, 1)^T$ *satisfies the $\epsilon$-ASOC conditions with the sequences* $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$, $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{N}}$, $\{\epsilon_k\}_{k\in\mathbb{N}}$ *given by* $\boldsymbol{x}^k := (1/k, 1)^T$, $\boldsymbol{\lambda}^k := (-k, k)^T$ *and* $\epsilon_k := 1/k^2$.

*To prove that* $\boldsymbol{x}^*$ *is not an SAGP point, let us suppose, by contradiction, that there exist sequences* $\{(v_1^k, v_2^k)^T\}_{k\in\mathbb{N}} \subset \mathbb{R}^2$, $\{(x_1^k, x_2^k)^T\}_{k\in\mathbb{N}} \subset \mathbb{R}^2$, $\{(\lambda_1^k, \lambda_2^k)^T\}_{k\in\mathbb{N}} \subset \mathbb{R}^2$ *and* $\{\vartheta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+^*$ *such that* $(x_1^k, x_2^k)^T \to (0, 1)^T$, $(v_1^k, v_2^k)^T \to (0, 0)^T$, $\vartheta_k \to 0$ *and*

$$\begin{aligned} \begin{bmatrix} v_1^k \\ v_2^k \end{bmatrix} = \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^k, \boldsymbol{\lambda}^k) &= \begin{bmatrix} 0 \\ x_2^k - 2 \end{bmatrix} + \lambda_1^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda_2^k \begin{bmatrix} x_2^k \\ x_1^k \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1^k + \lambda_2^k x_2^k \\ x_2^k - 2 + \lambda_2^k x_1^k \end{bmatrix}, \end{aligned} \tag{60}$$

*with*

$$\lambda_1^k x_1^k \geq -\vartheta_k \quad and \quad \lambda_2^k x_1^k x_2^k \geq -\vartheta_k, \tag{61}$$

*for all* $k \in \mathbb{N}$*. Multiplying the second row of* (60) *by* $x_2^k$*, we have* $x_2^k v_2^k - x_2^k(x_2^k - 2) = \lambda_2^k x_1^k x_2^k$ *for all* $k \in \mathbb{N}$*. Hence,* $\lim_{k\in\mathbb{N}} \lambda_2^k x_1^k x_2^k = 1$*. On the other hand, multiplying the first row of* (60) *by* $x_1^k$*, we have* $x_1^k v_1^k - \lambda_2^k x_1^k x_2^k = \lambda_1^k x_1^k$*, for all* $k \in \mathbb{N}$*. Thus,* $\lim_{k\in\mathbb{N}} \lambda_1^k x_1^k = -1$*, which contradicts* (61) *by taking limits.*

Figure 1 puts into perspective the aforementioned sequential optimality conditions and their interrelationships.

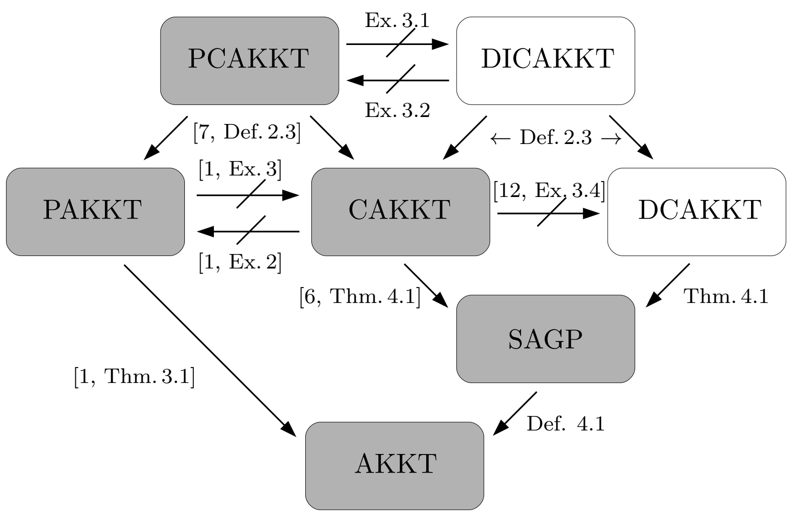


Figure 1: Contextualizing DCAKKT and DICAKKT among other well-established sequential optimality conditions.

# 5 Final remarks

In contrast with CAKKT, which handles feasibility in the image space, the devised sequential optimality condition DCAKKT controls feasibility in the domain, essentially using the distance of the analyzed primal sequence to the feasible set. DCAKKT is not related to the recently established PCAKKT, thus leading to a prospective improvement upon the convergence theory of a wide class of first-order augmented Lagrangian methods. This lack of interrelationship indicates that the class of augmented Lagrangian methods may be capable of generating limit points satisfying properties that are distinct from the already known sequential optimality conditions. Proved to be stronger than SAGP, the DCAKKT, together with its non-smooth counterpart ($\epsilon$-ASOC), turns out to be a stronger generalization of the AGP optimality condition.

**Acknowledgements** This work was partially supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) grants 131634/2018-6 and 305010/2020-4, and *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP) grants 2013/07375-0, 2016/22989-2, 2018/24293-0, 2019/18859-4.

# References

[1] Andreani, R., Fazzio, N.S., Schuverdt, M.L., Secchin, L.D.: A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. SIAM Journal on Optimization **29**(1), 743–766 (2019). URL `https://doi.org/10.1137/17M1147330`

[2] Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011). URL `https://doi.org/10.1080/02331930903578700`

[3] Andreani, R., Haeser, G., Mito, L.M., Ramos, A., Secchin, L.D.: Correction to: On the best achievable quality of limit points of augmented Lagrangian schemes. Numerical Algorithms (2021). URL `https://doi.org/10.1007/s11075-021-01241-3`

[4] Andreani, R., Haeser, G., Mito, L.M., Ramos, A., Secchin, L.D.: On the best achievable quality of limit points of augmented Lagrangian schemes. Numerical Algorithms (2021). URL `https://doi.org/10.1007/s11075-021-01212-8`

[5] Andreani, R., Martínez, J.M., Santos, L.T.: Newton's method may fail to recognize proximity to optimal points in constrained optimization. Mathematical Programming **160**(1), 547–555 (2016)

[6] Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM Journal on Optimization **20**(6), 3533–3554 (2010). URL `https://doi.org/10.1137/090777189`

[7] Andreani, R., Ramos, A., Ribeiro, A.A., Secchin, L.D., Velazco, A.R.: On the convergence of augmented Lagrangian strategies for nonlinear programming. IMA Journal of Numerical Analysis (2021). URL `https://doi.org/10.1093/imanum/drab021`

[8] Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA (2014). URL `https://locus.siam.org/doi/abs/10.1137/1.9781611973365`

[9] Gonzaga, C.C., Karas, E., Vanti, M.: A globally convergent filter method for nonlinear programming. SIAM Journal on Optimization **14**(3), 646–669 (2004)

[10] Haeser, G., Schuverdt, M.L.: On approximate KKT condition and its extension to continuous variational inequalities. Journal of Optimization Theory and Applications **149**(3), 528–539 (2011). URL https://doi.org/10.1007/s10957-011-9802-x

[11] Helou, E.S., Santos, S.A., Simões, L.E.A.: Analysis of a new sequential optimality condition applied to mathematical programs with equilibrium constraints. Journal of Optimization Theory and Applications **185**(2), 433–447 (2020). URL https://doi.org/10.1007/s10957-020-01658-1

[12] Helou, E.S., Santos, S.A., Simões, L.E.A.: A new sequential optimality condition for constrained nonsmooth optimization. SIAM Journal on Optimization **30**(2), 1610–1637 (2020). URL https://doi.org/10.1137/18M1228608

[13] Hestenes, M.R.: Multiplier and gradient methods. Journal of Optimization Theory and Applications **4**(5), 303–320 (1969)

[14] Hestenes, M.R.: Optimization Theory: the finite dimensional case. John Wiley, New York (1975)

[15] Kurdyka, K.: On gradients of functions definable in o-minimal structures. Annales de l'Institut Fourier **48**, 769–783 (1998)

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

[17] Martínez, J.M.: Inexact-restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. Journal of Optimization Theory and Applications **111**(1), 39–58 (2001)

[18] Martínez, J.M., Svaiter, B.F.: A practical optimality condition without constraint qualifications for nonlinear programming. Journal of Optimization Theory and Applications **118**(1), 117–133 (2003). URL https://doi.org/10.1023/A:1024791525441

[19] Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: R. Fletcher (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)

[20] Rockafellar, R.T.: The multiplier method of Hestenes and Powell applied to convex programming. Journal of Optimization Theory and Applications **12**(6), 555–562 (1973)