Journal of Optimization Theory and Applications (2025) 206:54
https://doi.org/10.1007/s10957-025-02731-3

Global Convergence of a Second-order Augmented Lagrangian Method Under an Error Bound Condition

Roberto Andreani[1] · Gabriel Haeser[2] · Renan William Prado[2] · Maria L. Schuverdt[3] · Leonardo D. Secchin[4]

Received: 28 September 2024 / Accepted: 22 May 2025 / Published online: 10 June 2025

Abstract

This work deals with convergence to points satisfying the weak second-order necessary optimality conditions of a second-order safeguarded augmented Lagrangian method from the literature. To this end, we propose a new second-order sequential optimality condition that is, in a certain way, based on the iterates generated by the algorithm itself. This also allows us to establish the best possible global convergence result for the method studied, from which a companion constraint qualification is derived. The companion constraint qualification is independent of the Mangasarian-Fromovitz and constant-rank constraint qualifications and remains verifiable without them, as it can be certified by other known constraint qualifications. Furthermore, unlike similar results from previous works, the new constraint qualification cannot be weakened by another one with second-order global convergence guarantees for the method and assures second-order stationarity without the need for constant rank hypotheses. To guarantee the latter result, we established the convergence of the method under a

Communicated by Alexey F. Izmailov.

✉ Renan William Prado
renanwilliamprado@gmail.com

Roberto Andreani
andreani@unicamp.br

Gabriel Haeser
ghaeser@ime.usp.br

Maria L. Schuverdt
mlschuverdt@gmail.com

Leonardo D. Secchin
secchinleo@gmail.com

1 University of Campinas, Campinas, São Paulo, Brazil

2 University of São Paulo, São Paulo, São Paulo, Brazil

3 National University of La Plata, La Plata, Buenos Aires Province, Argentina

4 Federal University of Espírito Santo, São Mateus, Espírito Santo, Brazil

property slightly stronger than the error bound constraint qualification, which, until now, has not been known to be associated with nonlinear optimization methods.

Keywords Sequential optimality conditions · Second-order optimization methods · Constraint qualifications · Error bound · Constant rank

Mathematics Subject Classification 90C46 · 90C30 · 65K05

1 Introduction

We consider the general nonlinear programming problem (NLP)

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

where $f\colon \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{g}\colon \mathbb{R}^n \to \mathbb{R}^p$ and $\boldsymbol{h}\colon \mathbb{R}^n \to \mathbb{R}^q$ are all twice continuously differentiable. The feasible set of problem (1) is denoted by

$$\Omega \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0},\ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}.$$

Due to advances in computing hardware and algorithmic efficiency, there has been an increase in interest in optimization methods that utilize second-order information [1, 26, 29, 34]. It allows achieving high accuracy solutions [20], as it provides additional information about the curvature of the objective and constraints; see, for example, [11] and references therein. This work focuses on the second-order safeguarded augmented Lagrangian method (SALM) proposed in [4], whose first-order counterpart method ALGENCAN [3] is a celebrated algorithm with excellent convergence results [6, 10, 15] and robust numerical performance [21, 22].

Ideally, one would like methods to find a local minimizer. However, practical methods can only guarantee stationary points. In the first-order case, the strength of an algorithm can be measured by the strength of constraint qualifications that ensure the Karush-Kuhn-Tucker (KKT) conditions at its accumulation points. Likewise, the validity of second-order stationary conditions can be considered.

Preferably, as a second-order condition, one wants not only the KKT conditions to hold but also that, for some Lagrange multiplier vector, no direction within the linearized cone is simultaneously a non-ascending direction for the objective function ($\nabla f(\boldsymbol{x})^T\boldsymbol{d} \leq 0$) and of negative curvature for the Hessian of the Lagrangian function (w.r.t. $\boldsymbol{x}$); that is, one wants that the Hessian of the Lagrangian function is positive semi-definite in the strong critical cone

$$\mathscr{C}^S(\boldsymbol{x}) = \left\{ \boldsymbol{d} \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} \nabla f(\boldsymbol{x})^T\boldsymbol{d} & \leq 0, \\ \nabla h_j(\boldsymbol{x})^T\boldsymbol{d} & = 0 \text{ for all } j \in \{1,\ldots,q\}, \\ \nabla g_i(\boldsymbol{x})^T\boldsymbol{d} & \leq 0 \text{ for all } i \in \mathcal{A}(\boldsymbol{x}) \end{array} \right\},$$

where $\mathcal{A}(\boldsymbol{x}) = \{j \mid g_j(\boldsymbol{x}) = 0\}$ is the set of indexes of active inequality constraints at $\boldsymbol{x}$. Together with the KKT conditions, the resultant condition is called *strong second-order optimality condition* (SSOC). Nevertheless, even testing such a condition at a given and fixed point is hard [33]. For that reason, all commonly applied methods with second-order optimality guarantees deal with the weak critical cone

$$\mathscr{C}^W(\boldsymbol{x}) = \left\{ \boldsymbol{d} \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla h_j(\boldsymbol{x})^T \boldsymbol{d} & = 0 \text{ for all } j \in \{1, \ldots, q\}, \\ \nabla g_i(\boldsymbol{x})^T \boldsymbol{d} & = 0 \text{ for all } i \in \mathcal{A}(\boldsymbol{x}) \end{array} \right\}; \tag{2}$$

for instance, the methods in [11]. As easily seen, this cone is properly included in the strong critical cone when KKT conditions hold. The practical condition associated with the weak critical cone is called *weak second-order condition* (WSOC); see Definition 2.3.

Remarkably, even global minimizers fail to satisfy WSOC under assumptions that are less stringent than the *Mangasarian-Fromovitz Constraint Qualification* (MFCQ); see the counterexample in [18, Section 2.4]. Consequently, global convergence to second-order stationary points of any algorithm whose set of acceptable solutions includes points sufficiently close to local or global minimizers cannot be ensured even when MFCQ is satisfied. For this reason, until recently, second-order methods had the *linear independence constraint qualification* (LICQ) and strict complementarity as requirements for convergence, conditions that guarantee that $\mathscr{C}^W(\boldsymbol{x}) = \mathscr{C}^S(\boldsymbol{x})$ and uniqueness of the multipliers. To guarantee convergence of implementable second-order methods, one may require MFCQ together with some other constant rank-type condition. In [14], this was done assuming MFCQ and that the gradients of all active constraints in a neighborhood of $\boldsymbol{x}^*$ have constant rank – a condition known as *weak constant rank* (WCR), see Definition 2.4. In [5], it was proved that a constraint qualification (CQ) of constant rank-type also fulfills this purpose. Finally, these results were extended in [11] using a second-order sequential optimality condition called *second-order approximate KKT* (AKKT2).

Motivated by [11] and the recent results obtained for the first-order optimality in [10], we identify precisely the strongest global convergence result associated with SALM. This is done by showing that all possible accumulation points reached by the method are fully characterized by a new proposed second-order sequential optimality condition that we call *second-order augmented Lagrangian AKKT* (AL-AKKT2), following the terminology adopted in [10]. As a consequence, we prove that WSOC holds at these points, which include minimizers of (1), under new weak constraint qualifications. In particular, AL-AKKT2 is stronger than AKKT2 and allows us to obtain convergence of SALM in situations where WCR does not hold. It is worth noting that this was not possible before, as AKKT2 requires the validity of WCR to attest WSOC [25, Proposition 5]. This is done with a slightly more stringent condition than the error bound (EB) property, which we call *strong-EB*. Strong-EB is proven weaker than a recently proposed CQ for WSOC in the conic programming context, called strong-CRSC [8]. To the best of our knowledge, this is the first time that a second-order method has its global convergence established under an error bound-type condition.

It has been recently discovered that a relaxed version of the *quasinormality* (QN) [19, pp. 337] is sufficient for the boundedness of the multiplier sequence associated with the first-order SALM [16]. This *relaxed QN* (RQN) condition is a weak CQ, less stringent than QN itself and one of the weakest known CQs of constant rank-type, the *constant rank of the subspace component* (CRSC) [12]. However, the boundedness of dual sequences of SALM ensured by RQN is insufficient to assert WSOC. To address this concern, we show that WSOC holds under RQN as long as WCR or even strong-EB holds. This generalizes all the previous results from the literature regarding second-order algorithms.

The paper is organized as follows. In section 2, we present SALM and the fundamental theory involving WSOC. In section 3, we establish the new sequential optimality condition AL-AKKT2. In section 4, we prove that strong-EB guarantees WSOC under AL-AKKT2 and the boundedness of the AL-AKKT2 multipliers. Furthermore, we prove that WCR is independent of strong-EB and that WCR can also guarantee WSOC under boundedness of the AL-AKKT2 dual sequences. In section 5, we prove that the AL-AKKT2 conditions fully characterize the accumulation points of sequences generated by SALM. Additionally, we define the weakest CQ for WSOC under the AL-AKKT2 condition. Finally, section 6 brings our conclusions and future works.

Notation

- We typically use boldface letters (i.e., $\boldsymbol{v}$) for vectors and vector-valued functions;
- v_i is the $i-$th component of the vector $\boldsymbol{v}$. If $\boldsymbol{v}$ is a vector-valued function, v_i is the function given by its $i-$th component;
- For $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v}_+ = (\max\{0, v_1\}, \ldots, \max\{0, v_n\})$;
- $\|\cdot\|$ denotes the Euclidean norm;
- $\mathcal{B}(\boldsymbol{x}, \epsilon)$ is the open ball with center $\boldsymbol{x}$, radius ϵ, and the Euclidean distance;
- $\mathbb{R}_+$ stands for the set of non-negative real numbers;
- $\mathbb{R}_{++}$ denotes the set of strictly positive real numbers;
- Given a natural ℓ, $I_\ell = \{k \in \mathbb{N} \mid 1 \le k \le \ell\}$;
- $\mathbb{I}$ denotes the identity matrix of appropriate size;
- Given an infinite subset $\mathcal{K} \subset \mathbb{N}$, $\lim_{k\in\mathcal{K}} \boldsymbol{w}^k$ denotes the limit of the subsequence of $\{\boldsymbol{w}^k\}_{k\in\mathbb{N}}$ with indexes $k \in \mathcal{K}$;
- $\mathrm{Sym}(n)$ is the set of the symmetric matrices of order n;
- Given two symmetric matrices A and B of the same order, $A \succeq B$ indicates that $A - B$ is positive semi-definite.

2 Technical Background and the Augmented Lagrangian Method

The algorithm of interest for solving nonlinear programming problems is the (Powell-Hestenes-Rockafellar – PHR) safeguarded augmented Lagrangian method (SALM). This method successively minimizes the PHR-augmented Lagrangian function described below.

Definition 2.1 Fixed a scalar $\rho > 0$, the Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian function $\mathcal{L}_\rho : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q_+$ is given by

$$\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \stackrel{\text{def}}{=} f(\boldsymbol{x}) + \frac{\rho}{2}\left(\left\| \left(\boldsymbol{g}(\boldsymbol{x}) + \frac{\boldsymbol{\mu}}{\rho} \right)_+ \right\|^2 + \left\| \boldsymbol{h}(\boldsymbol{x}) + \frac{\boldsymbol{\lambda}}{\rho} \right\|^2 \right)$$

for all $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{\mu} \in \mathbb{R}^p_+$ and $\boldsymbol{\lambda} \in \mathbb{R}^q$.

Specifically, the idea behind SALM is to solve (1) through a sequence of problems

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \mathcal{L}_{\rho_k}(\boldsymbol{x}, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k) \tag{3}$$

where $\{\rho_k\}_{k \in \mathbb{N}}$ is a non-decreasing and positive sequence of real numbers 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, which are safeguarded estimates of the Lagrange multipliers for (1). The scalar ρ_k is called the penalty parameter. Under reasonable conditions, local minimizers of (3) are also local minimizers of (1), c.f. [19, Prop. 4.2.3]. Essentially, SALM is a method that controls the increasing behavior of the sequence of penalty parameters to avoid ill-conditioned subproblems.

Unless (3) is a convex problem, globally minimizing it in practice is tough. Despite the global problem (3) being hard, finding an approximate second-order stationary point is not. Consequently, a natural idea would be to apply to (3) globally convergent algorithms for unconstrained optimization that generate sequences whose accumulation points satisfy a second-order optimality condition. However, the PHR augmented Lagrangian function is not twice differentiable everywhere. To overcome this, we work with a substitute for its Hessian, as stated in the next definition.

Definition 2.2 Considering the functions that appear in problem (1), $\boldsymbol{x} \in \mathbb{R}^n$, $\bar{\boldsymbol{\mu}} \in \mathbb{R}^p_+$, $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^q$, $\eta \in \mathbb{R}_+$ and $\rho \in \mathbb{R}_+$, define the matrix

$$\begin{aligned} \nabla^2_{\boldsymbol{x}} \mathcal{L}^\eta_\rho(\boldsymbol{x}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \stackrel{\text{def}}{=} \nabla^2 f(\boldsymbol{x}) + \sum_{i=1}^p (\bar{\mu}_i + \rho g_i(\boldsymbol{x}))_+ \nabla^2 g_i(\boldsymbol{x}) + \rho \sum_{\bar{\mu}_i + \rho g_i(\boldsymbol{x}) \geq -\eta} \nabla g_i(\boldsymbol{x}) \nabla g_i(\boldsymbol{x})^T \\ + \sum_{j=1}^q \left(\bar{\lambda}_j + \rho h_j(\boldsymbol{x})\right) \nabla^2 h_j(\boldsymbol{x}) + \rho \sum_{j=1}^q \nabla h_j(\boldsymbol{x}) \nabla h_j(\boldsymbol{x})^T. \end{aligned}$$

By [4, Proposition 1], it is possible to show that $\nabla^2_{\boldsymbol{x}} \mathcal{L}^\eta_\rho(\cdot, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}})$ is positive semidefinite at unconstrained local minima of $\mathcal{L}_\rho(\cdot, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}})$, independently of η. Thus, it is a suitable substitute for the second-order derivative of $\mathcal{L}_\rho(\cdot, \boldsymbol{\lambda}, \boldsymbol{\mu})$ w.r.t. $\boldsymbol{x}$, and it can be employed in methods for unconstrained optimization with second-order convergence guarantees. We expect that such methods applied to (3) are able to find $\boldsymbol{x}^k$ fulfilling (4) below. With this discussion, we enunciate the safeguarded augmented Lagrangian method in Algorithm 1.

Remark 2.1 A small extension of [38, Corollary 1.4.3] can be used to guarantee the existence of points fulfilling (4) under the assumption that the PHR augmented

Algorithm 1 Second-order safeguarded augmented Lagrangian method (SALM)

step 1 Choose $\lambda_{\min} \leq \lambda_{\max}$, $0 \leq \mu_{\max}$, $\gamma > 1$, $0 < r < 1$, $\{\theta_k\}_{k\in\mathbb{N}}$ and $\{\eta_k\}_{k\in\mathbb{N}}$ with $\theta_k \downarrow 0$ and $\eta_k \downarrow 0$. Let $\bar{\boldsymbol{\mu}}^1 \in [0, \mu_{\max}]^p$, $\bar{\boldsymbol{\lambda}}^1 \in [\lambda_{\min}, \lambda_{\max}]^q$ and $\rho_1 > 0$. Initialize $k \leftarrow 1$.

step 2 Solve approximately (3) obtaining $\boldsymbol{x}^k$ satisfying

$$\|\nabla_{\boldsymbol{x}} \mathcal{L}_{\rho_k}(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\| \leq \theta_k \quad \text{and} \quad \nabla^2_{\boldsymbol{x}} \mathcal{L}^{\eta_k}_{\rho_k}(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k) \succeq -\theta_k \mathbb{I}. \tag{4}$$

step 3 Calculate $\boldsymbol{\mu}^k \stackrel{\text{def}}{=} \left(\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k)\right)_+$ and $\boldsymbol{\lambda}^k \stackrel{\text{def}}{=} \left(\bar{\boldsymbol{\lambda}}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k)\right)$.

step 4 For each $i \in I_p$, define

$$V_i^k \stackrel{\text{def}}{=} \min\left\{-g_i(\boldsymbol{x}^k), \frac{\bar{\mu}_i^k}{\rho_k}\right\}.$$

step 5 If $k = 1$ or

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

do $\rho_{k+1} = \rho_k$. Otherwise, choose $\rho_{k+1} \geq \gamma\rho_k$. Also, take $\bar{\boldsymbol{\mu}}^{k+1} \in [0, \mu_{\max}]^p$ and $\bar{\boldsymbol{\lambda}}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^q$. Do $k \leftarrow k+1$ and return to 1.

Lagrangian function is bounded below; a similar result for twice continuously differentiable functions can be seen in [17]. This can be ensured by assuming that the objective function of (1) has a lower bound. Under such an assumption, any method generating sequences with second-order properties for unconstrained optimization applied to (3) can be used to obtain (4). Andreani et al. proposed an implementation of a SALM meeting the requirements of this section, which they called ALGENCAN- SECOND, see [4, Algorithm 3.1].

From now on, we will refer to the second-order augmented Lagrangian method stated above as SALM. When we want to refer to the first-order method, that is, the one with only the first condition in (4), we explicitly write "first-order SALM". Additionally, any sequence satisfying the steps of SALM for a proper choice of parameters is said to be generated by SALM. Similarly, we might also say that a sequence is generated by the first-order SALM.

When a new method is proposed, it is desirable that the KKT conditions, i.e., for some $\boldsymbol{\mu} \in \mathbb{R}^p_+$, $\boldsymbol{\lambda} \in \mathbb{R}^q$ and $\boldsymbol{x}^* \in \Omega$,

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{x}^*) = 0,$$

where $\mathcal{L}$ is the usual Lagrangian function associated with (1), hold at its accumulation points under at least one constraint qualification. Moreover, the interest in CQs lies in pursuing the least stringent condition. This is because weaker CQs, following from their stronger counterparts, hold over a considerably broader set of feasible points. In other words, any point satisfying a stronger CQ automatically meets the requirements of the weaker one. Consequently, if an algorithm's global convergence is proven under a weaker condition, then any of its primal sequence limits that satisfy a stronger one are guaranteed to comply with the KKT criteria. Furthermore, the existence of feasible points that satisfy the weaker conditions but not the stronger ones significantly extends the applicability of the convergence result.

With regard to SALM, in [10], it was established that the weakest CQ for which it is possible to ensure the KKT conditions at the accumulation points of the first-order SALM, namely, *AL-regularity*. In particular, *AL-regularity* is implied by the *quasinormality CQ* and CRSC [12], two weak CQs used to guarantee convergence in first-order optimization methods [7, 12, 23].

The most accepted second-order optimality notion linked to the convergence of algorithms is the following:

Definition 2.3 We say that a feasible point $\boldsymbol{x}^*$ satisfies the *weak second-order condition (WSOC)* whenever there are vectors $\boldsymbol{\mu} \in \mathbb{R}^p_+$ and $\boldsymbol{\lambda} \in \mathbb{R}^q$ so that

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \boldsymbol{0},\ \boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{x}^*) = 0 \text{ and } \boldsymbol{d}^T \nabla^2_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) \boldsymbol{d} \geq 0\ \ \forall \boldsymbol{d} \in \mathscr{C}^W(\boldsymbol{x}^*),$$

where $\mathscr{C}^W(\boldsymbol{x}^*)$ is the weak critical cone (2). In this case, $\boldsymbol{x}^*$ is called a *WSOC point*.

In other words, WSOC holds whenever the KKT conditions hold with $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ and the Hessian of the Lagrangian $\mathcal{L}(\cdot, \boldsymbol{\mu}, \boldsymbol{\lambda})$ is positive semi-definite in the subspace orthogonal to the gradients of active constraints at $\boldsymbol{x}^*$.

Since WSOC is stronger than the KKT conditions, and local minima need not satisfy the first-order stationarity conditions, WSOC may also fail at certain local minimizers. A suitable CQ can be imposed for this. The convergence analysis of algorithms involving WSOC usually revolves around rigorous CQs. However, using second-order sequential optimality conditions (SOC2), it has recently been shown that this can be relaxed, allowing, for instance, the sequences of multipliers generated by the algorithm to be unbounded. Two such conditions were considered: the *second-order approximate-KKT* (AKKT2) [11] condition and the *second-order complementarity-approximate-KKT* (CAKKT2) condition [24]. Optimization methods that employ second-order techniques, like SALM, can ensure both conditions; see [11, 24] for more methods.

When an appropriate property on the constraints holds at a point that satisfies an SOC2, such a point also satisfies WSOC. These properties were usually composed of a CQ together with a constant rank-type property, which we recall below.

Definition 2.4 A feasible point $\boldsymbol{x}^*$ is said to satisfy the *weak constant rank* (WCR) property if there exists a neighborhood V of $\boldsymbol{x}^*$ such that the dimension of the space generated/spanned by $\{\nabla g_i(\boldsymbol{x})\}_{i \in \mathcal{A}(\boldsymbol{x}^*)} \cup \{\nabla h_j(\boldsymbol{x})\}_{j=1,\ldots,q}$ is constant for all $\boldsymbol{x} \in V$.

Given an arbitrary SOC2, the weakest associated CQ can be built by the techniques used in [11], which, in the case of AKKT2, is the *second-order cone continuity property* (CCP2) [11]. This CQ, together with WCR, guarantees WSOC. A similar discussion is valid for the weakest CQ associated with CAKKT2, namely *CAKKT2-regularity* [24].

The CQs required for convergence of methods like SALM are commonly related to the problem's stability. An instability happens when small changes in the constraints cause significant differences in the optimum values. A well-known property linked to the stability of a problem is the *error bound* [37, §3]. This work uses the validity of a variation of the error bound condition in order to guarantee WSOC. As said in the introduction, this is the first time a condition slightly stronger than the error bound has been used to reach such a strong result.

Definition 2.5 Let Ω be the feasible set of (1). We say that $\boldsymbol{x}^* \in \Omega$ meets the *error bound* (EB) CQ when there are constants $L > 0$ and $\delta > 0$ such that

$$\operatorname{dist}(\boldsymbol{x}, \Omega) \leq L\left(\|[\boldsymbol{g}(\boldsymbol{x})]_+\| + \|\boldsymbol{h}(\boldsymbol{x})\|\right) \text{ for all } \boldsymbol{x} \in \mathcal{B}(\boldsymbol{x}^*, \delta).$$

We call L a local error bound constant for Ω at $\boldsymbol{x}^*$.

SALM constructs a sequence of approximations of the primal and dual solutions. Even though the primal sequence has an accumulation point, the dual sequence may be unbounded if no extra assumptions about the problem hold. In [16], a weak extension of the quasinormality CQ called *relaxed quasinormality* has been proposed that assures boundedness of the dual sequences of the method. We will use this condition in the following sections to obtain WSOC.

Definition 2.6 ([16]) It is said that the *relaxed quasinormality* (RQN) CQ holds at the feasible point $\boldsymbol{x}^*$ of (1) whenever there is no sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ converging to $\boldsymbol{x}^*$ and vector $(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ satisfying the following requirements:

1. $\sum_{i=1}^p \mu_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^q \lambda_j \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}$;
2. $(\boldsymbol{\mu}, \boldsymbol{\lambda}) \neq (\mathbf{0}, \mathbf{0})$;
3. for all $i \in I_p$, $j \in I_q$ and $k \in \mathbb{N}$,

$$\begin{aligned} \text{if } \mu_i \neq 0, \text{ then } g_i(\boldsymbol{x}^k) > 0; \quad \text{if } \lambda_j \neq 0, \text{ then } \lambda_j h_j(\boldsymbol{x}^k) > 0; \\ \text{if } \mu_i = 0, \text{ then } g_i(\boldsymbol{x}^k)_+ = o(t_k); \quad \text{if } \lambda_j = 0, \text{ then } h_j(\boldsymbol{x}^k) = o(t_k), \end{aligned}$$

where $t_k = \min\{\min_{\mu_i > 0} g_i(\boldsymbol{x}^k)_+, \min_{\lambda_j \neq 0} |h_j(\boldsymbol{x}^k)|\}$.

Some CQs, necessarily not weaker than MFCQ, guarantee WSOC at local minimizers. An example is the *relaxed constant rank constraint qualification* (RCRCQ) [11, Proposition 4.12]. This CQ needs the local constant dimension of the subspace generated by many subsets of the gradients of active constraints. This was relaxed in the context of nonlinear conic programming with the *strong-constant rank of the subspace component* (strong-CRSC) [8, Definition 4.1], by reducing the number of subsets needed to obtain WSOC. In the next sections, we will prove, in addition, the global convergence of SALM under strong-CRSC.

Definition 2.7 Considering a feasible point $\boldsymbol{x}^*$ of (1), we define the set

$$\mathcal{J}_-(\boldsymbol{x}^*) = \left\{ i \in \mathcal{A}(\boldsymbol{x}^*) \,\middle|\, \begin{array}{l} \exists \boldsymbol{\mu} \in \mathbb{R}^p_+,\ \exists \boldsymbol{\lambda} \in \mathbb{R}^q, \\ -\nabla g_i(\boldsymbol{x}^*) = \sum_{i\in\mathcal{A}(\boldsymbol{x}^*)} \mu_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j\in I_q} \lambda_j \nabla h_j(\boldsymbol{x}^*) \end{array} \right\}.$$

Definition 2.8 The strong-CRSC condition is valid at a feasible point $\boldsymbol{x}^*$ of (1) when for every $\mathcal{J}_-(\boldsymbol{x}^*) \subset \mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$, the dimension of the subspace generated/spanned by the set $\{\nabla g_i(\boldsymbol{x})\}_{i\in\mathcal{A}} \cup \{\nabla h_j(\boldsymbol{x})\}_{j\in I_q}$ remains constant in a neighborhood of $\boldsymbol{x}^*$.

3 A New Second-Order Sequential Optimality Condition

In recent years, the global convergence of the first-order SALM has been improved by associating it with increasingly strengthened sequential optimality conditions [6, 7, 11, 15, 24, 35]. A recent proposal has provided a definitive answer to such an issue. In [10] (see also [9]), a sequential optimality condition inspired by the augmented Lagrangian method itself was proposed, namely Augmented Lagrangian AKKT (AL-AKKT). This condition provides the best possible global convergence result for the first-order SALM in a suitable sense. Motivated by these ideas, we define the second-order counterpart of AL-AKKT for SALM.

Definition 3.1 We say that a feasible point $\boldsymbol{x}^*$ for problem (1) fulfills the *second-order augmented Lagrangian AKKT* (AL-AKKT2) condition if there are sequences $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$, $\{\theta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and $\{\eta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ such that $\lim_{k\to\infty} \boldsymbol{x}^k = \boldsymbol{x}^*$, $\lim_{k\to\infty} \theta_k = \lim_{k\to\infty} \eta_k = 0$, the sequences $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}}$ are bounded,

$$\|\nabla_{\boldsymbol{x}} \mathcal{L}_{\rho_k}(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\| \le \theta_k, \text{ for all } k \in \mathbb{N}, \tag{5}$$

$$\nabla^2_{\boldsymbol{x}} \mathcal{L}^{\eta_k}_{\rho_k}(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k) \succeq -\theta_k\, \mathbb{I}, \text{ for all } k \in \mathbb{N}, \tag{6}$$

and

$$\lim_{k\to\infty} \min\left\{-g_i(\boldsymbol{x}^k), \frac{\bar{\mu}^k_i}{\rho_k}\right\} = 0, \text{ for all } i \in I_p. \tag{7}$$

As it is natural, we call $\{\rho_k\}_{k\in\mathbb{N}}$ the penalty parameter sequence and both $\{\left(\rho_k \boldsymbol{g}(\boldsymbol{x}^k) + \bar{\boldsymbol{\mu}}^k\right)_+\}_{k\in\mathbb{N}}$ and $\{(\rho_k \boldsymbol{h}(\boldsymbol{x}^k) + \bar{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$ the *AL-AKKT2 dual sequences* or *AL-AKKT2 multipliers*. Furthermore, the sequence of tuples $\{(\theta_k, \rho_k, \eta_k, \boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$ is said to be associated with the AL-AKKT2 conditions.

Remark 3.1 The AL-AKKT condition is defined precisely like the AL-AKKT2 condition but without requiring (6).

To begin the analysis, we first establish that AL-AKKT2 is a necessary optimality condition.

Theorem 3.1 *The AL-AKKT2 condition is valid at every local minimizer of (1).*

Proof Follow the proof of [11, Theorem 3.4], noting that the constructed sequence satisfies the AL-AKKT2 properties. □

4 Convergence of SALM Under an Error Bound Condition

The work [25] suggests that second-order algorithms generally try to compute a negative-curvature direction from the second-order derivative of the Lagrangian function. However, this procedure can fail to identify WSOC points correctly. As shown in [25], this problem is present in AKKT2 sequences, making WCR necessary; see [25, Definition 1 and Proposition 5]. On the other hand, closely investigating the proof of [25, Proposition 5], and comparing the terms in AL-AKKT2 sequences, the problematic part related to the necessity of WCR is closely linked with the uncontrolled growth of the penalty parameter sequence $\{\rho_k\}_{k\in\mathbb{N}}$, see Definition 5. Since ρ_k appears in the dual sequences of the AL-AKKT2 condition, its growth is somehow bounded by the growth of the dual sequences. This leads us to ask whether it is possible to guarantee convergence to WSOC points without requiring the WCR condition. This section shows it is possible.

Next, we will define the first two concepts required to derive the basic results of this section.

Definition 4.1 For each $\boldsymbol{x}^* \in \mathbb{R}^n$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$, define the vector space

$$\mathscr{C}^W_{\mathcal{A}}(\boldsymbol{x}, \boldsymbol{x}^*) = \left\{ \boldsymbol{d} \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} \nabla h_j(\boldsymbol{x})^T \boldsymbol{d} & = 0 \text{ for all } j \in I_q \\ \nabla g_i(\boldsymbol{x})^T \boldsymbol{d} & = 0 \text{ for all } i \in \mathcal{A} \end{array} \right\}$$

and the set

$$\Omega_{\mathcal{A}} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{g}_{\mathcal{A}}(\boldsymbol{x}) = \boldsymbol{0},\ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\},$$

where $\boldsymbol{g}_{\mathcal{A}}(\boldsymbol{x}) = (g_i(\boldsymbol{x}))_{i\in\mathcal{A}}$. Note that $\mathscr{C}^W_{\mathcal{A}(\boldsymbol{x}^*)}(\boldsymbol{x}^*, \boldsymbol{x}^*) = \mathscr{C}^W(\boldsymbol{x}^*)$ (see (2)). Additionally, define the set

$$\mathcal{A}_{=0} = \{i \in I_p \mid \exists \delta > 0, \forall \boldsymbol{y} \in \mathcal{B}(\boldsymbol{x}^*, \delta) \cap \Omega,\ g_i(\boldsymbol{y}) = 0\}.$$

Next, we define a slightly stronger condition than EB, which we use to guarantee WSOC at points fulfilling the AL-AKKT2 conditions.

Definition 4.2 We say that the *strong error bound* (strong-EB) condition holds at the point $\boldsymbol{x}^*$ for Ω whenever Ω and $\Omega_{\mathcal{A}}$ fulfill EB at $\boldsymbol{x}^*$ for each $\mathcal{A}$ with $\mathcal{A}_{=0} \subset \mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$, that is, there are $L > 0$ and $\delta > 0$ such that

$$\begin{aligned} \operatorname{dist}(\boldsymbol{x}, \Omega_{\mathcal{A}}) &\leq L \left(\|\boldsymbol{g}_{\mathcal{A}}(\boldsymbol{x})\| + \|\boldsymbol{h}(\boldsymbol{x})\|\right) \text{ and} \\ \operatorname{dist}(\boldsymbol{x}, \Omega) &\leq L \left(\|[\boldsymbol{g}(\boldsymbol{x})]_+\| + \|\boldsymbol{h}(\boldsymbol{x})\|\right) \end{aligned}$$

for all $\boldsymbol{x} \in \mathcal{B}(\boldsymbol{x}^*, \delta)$.

Abadie's CQ holds at $\boldsymbol{x}^*$ when the linearized cone of Ω at $\boldsymbol{x}^*$ and the tangent cone to Ω at $\boldsymbol{x}^*$,

$$\begin{aligned} T_\Omega(\boldsymbol{x}^*) \overset{\text{def}}{=} \limsup_{t\downarrow 0} \frac{\Omega - \boldsymbol{x}^*}{t} &= \{\boldsymbol{v} \in \mathbb{R}^n \mid \exists \{\boldsymbol{v}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n, \exists \{t_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++} \\ &\quad \text{such that } \lim_{k\to\infty} \boldsymbol{v}^k = \boldsymbol{v},\ \lim_{k\to\infty} t_k = 0 \text{ and } (t_k \boldsymbol{v}^k + \boldsymbol{x}^*) \in \Omega, \forall k \in \mathbb{N}\} \end{aligned}$$

[36, Definition 6.1], coincide. It is known that WSOC can be ensured at local minimizers if Abadie's CQ is valid regarding the set $\Omega_{\mathcal{A}(\boldsymbol{x}^*)}$ [2, Theorem 3.1], which has

$\mathscr{C}^W(\boldsymbol{x}^*)$ as the linearized cone at $\boldsymbol{x}^*$. The next technical lemma involves the validity of Abadie's CQ in an intermediate set $\Omega_{\mathcal{A}}$, see Remark 4.1.

Lemma 4.1 *For a set $\mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$, suppose that*

$$\liminf_{\boldsymbol{x} \xrightarrow[\Omega_{\mathcal{A}}]{} \boldsymbol{x}^*} T_{\Omega_{\mathcal{A}}}(\boldsymbol{x}) = \mathscr{C}^W_{\mathcal{A}}(\boldsymbol{x}^*, \boldsymbol{x}^*). \tag{8}$$

Then

$$\mathscr{C}^W_{\mathcal{A}(\boldsymbol{x}^*)}(\boldsymbol{x}^*, \boldsymbol{x}^*) \subset \liminf_{\boldsymbol{x} \xrightarrow[\Omega_{\mathcal{A}}]{} \boldsymbol{x}^*} \mathscr{C}^W_{\mathcal{A}}(\boldsymbol{x}, \boldsymbol{x}^*).$$

Proof For a set $\mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$, $\boldsymbol{v} \in \mathscr{C}^W_{\mathcal{A}(\boldsymbol{x}^*)}(\boldsymbol{x}^*, \boldsymbol{x}^*)$ and a sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \Omega_{\mathcal{A}}$, as $\mathscr{C}^W_{\mathcal{A}(\boldsymbol{x}^*)}(\boldsymbol{x}^*, \boldsymbol{x}^*) \subset \mathscr{C}^W_{\mathcal{A}}(\boldsymbol{x}^*, \boldsymbol{x}^*)$, by (8), there exists $\boldsymbol{v}^k \to \boldsymbol{v}$ with $\boldsymbol{v}^k \in T_{\Omega_{\mathcal{A}}}(\boldsymbol{x}^k)$ and, by the definition of $T_{\Omega_{\mathcal{A}}}(\boldsymbol{x}^k)$, we have $T_{\Omega_{\mathcal{A}}}(\boldsymbol{x}^k) \subset \mathscr{C}^W_{\mathcal{A}}(\boldsymbol{x}^k, \boldsymbol{x}^*)$, from which $\boldsymbol{v}^k \in \mathscr{C}^W_{\mathcal{A}}(\boldsymbol{x}^k, \boldsymbol{x}^*)$, and the statement follows. □

Remark 4.1 Clarke's regularity is a well-known type of regularity in variational analysis and optimization [27, 28, 36]. Considering a closed subset $\mathcal{X} \subset \mathbb{R}^n$, Clarke's regularity is defined in [36, Definition 6.4] by asking for the equality

$$N_{\mathcal{X}}(\boldsymbol{x}^*) = \limsup_{\boldsymbol{x} \xrightarrow[\mathcal{X}]{} \boldsymbol{x}^*} \hat{N}_{\mathcal{X}}(\boldsymbol{x}^*) = \hat{N}_{\mathcal{X}}(\boldsymbol{x}^*),$$

where $\hat{N}_{\mathcal{X}}(\boldsymbol{x}^*)$ is the regular normal cone and $N_{\mathcal{X}}(\boldsymbol{x}^*)$ is the limiting (Mordukhovich) normal cone. In view of [36, Corollary 6.29], Clarke's regularity can be stated as

$$T_{\mathcal{X}}(\boldsymbol{x}^*) = \limsup_{t\downarrow 0} \frac{\mathcal{X} - \boldsymbol{x}^*}{t} = \liminf_{\boldsymbol{x} \xrightarrow[\mathcal{X}]{} \boldsymbol{x}^*,\ t\downarrow 0} \frac{\mathcal{X} - \boldsymbol{x}}{t} \overset{\text{def}}{=} \hat{T}_{\mathcal{X}}(\boldsymbol{x}^*),$$

where $\hat{T}_{\mathcal{X}}(\boldsymbol{x}^*)$ is the regular tangent cone (see [36, Definition 6.25] or [32, Definition 1.8]). Additionally, by [36, Theorem 6.26],

$$\liminf_{\boldsymbol{x} \xrightarrow[\mathcal{X}]{} \boldsymbol{x}^*} T_{\mathcal{X}}(\boldsymbol{x}) = \hat{T}_{\mathcal{X}}(\boldsymbol{x}^*).$$

Therefore, since $\hat{T}_{\mathcal{X}}(\boldsymbol{x}^*) \subset T_{\mathcal{X}}(\boldsymbol{x}^*) \subset L_{\mathcal{X}}(\boldsymbol{x}^*)$, where $L_{\mathcal{X}}(\boldsymbol{x}^*)$ is the linearized cone of $\mathcal{X}$ at $\boldsymbol{x}^*$, the condition (8) is equivalent to asking for Abadie's CQ and Clarke's regularity at $\boldsymbol{x}^*$ w.r.t. the set $\Omega_{\mathcal{A}}$.

When applying a second-order method, the algorithm should obtain accumulation points fulfilling a second-order stationarity condition in at least one good context. In the case of methods that penalize infeasible constraints, such a context is when the sequence of penalty parameters ($\{\rho_k\}_{k\in\mathbb{N}}$) is bounded. The following result shows that it is possible to guarantee WSOC under the boundedness of this sequence.

Proposition 4.1 *Suppose that $\boldsymbol{x}^*$ is a feasible point fulfilling the AL-AKKT2 conditions, where the associated sequence of penalty parameters $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded. Then, WSOC holds at such a point.*

Proof Let us fix the point $\boldsymbol{x}^*$ for which AL-AKKT2 holds, and consider an associated sequence $\{(\theta_k, \rho_k, \eta_k, \boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$ as in Definition 3.1. By the hypothesis, the AL-AKKT2 multipliers $\boldsymbol{\lambda}^k = (\bar{\boldsymbol{\lambda}}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k))$ and $\boldsymbol{\mu}^k = (\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k))_+$ are bounded. Thus, there exists an infinite subset $\mathcal{K}_0 \subset \mathbb{N}$ such that

$$\lim_{k\in\mathcal{K}_0} (\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) = (\boldsymbol{\lambda}, \boldsymbol{\mu}) \quad \text{and} \quad \lim_{k\in\mathcal{K}_0} \rho_k = \rho_\infty,$$

with $\boldsymbol{\mu} \geq 0$ and $\rho_\infty > 0$. As $\lim_{k\to\infty} \min\{-g_i(\boldsymbol{x}^k), \bar{\mu}_i^k/\rho_k\} = 0$ for all i, we have $\lim_{k\to\infty} \bar{\mu}_i^k/\rho_k = 0$ whenever $g_i(\boldsymbol{x}^*) < 0$. Thus, $\mu_i = \lim_{k\to\infty} \rho_k((\bar{\mu}_i^k/\rho_k) + g_i(\boldsymbol{x}^k))_+ = 0$, and therefore the KKT complementarity is satisfied. Taking the limit in (5) over $\mathcal{K}_0$, we conclude that $(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda})$ is a KKT triple.

Second-order stationarity follows by taking an arbitrary direction $\boldsymbol{d} \in \mathscr{C}^W(\boldsymbol{x}^*)$, multiplying equation (6) on the right by $\boldsymbol{d}$ and on the left by $\boldsymbol{d}^T$, and then taking the limit with respect to $\mathcal{K}_0$ of the resulting expression. Since $\boldsymbol{d}$ does not depend on the multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, this ends the proof. □

The boundedness of the dual sequences generated by the method is an important point for the stability of algorithms. RQN condition [16] (Definition 2.6) guarantees the boundedness of multipliers generated by first-order SALM; it is weaker than CRSC [12] and QN, indicating that the boundedness of AL-AKKT multipliers can be obtained under weak and typical CQs. The compelling thing about this is that WSOC can be brought under the boundedness of the approximate Lagrange multipliers and strong-EB, as shown in the next result.

Proposition 4.2 *Suppose that $\boldsymbol{x}^* \in \Omega$ conforms to strong-EB and that there is an AL-AKKT2 sequence converging to $\boldsymbol{x}^*$ with bounded AL-AKKT2 multipliers. Then WSOC holds at $\boldsymbol{x}^*$.*

Proof Let us consider the AL-AKKT2 sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ associated with the bounded AL-AKKT2 multipliers $\{(\rho_k \boldsymbol{g}(\boldsymbol{x}^k) + \bar{\boldsymbol{\mu}}^k)_+\}_{k\in\mathbb{N}}$ and $\{(\rho_k \boldsymbol{h}(\boldsymbol{x}^k) + \bar{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$. By Proposition 4.1, the WSOC is valid when the sequence of penalty parameters is bounded. Thus, to prove this result, it is necessary to consider only the case in which $\{\rho_k\}_{k\in\mathbb{N}}$ is unbounded, which is assumed from now on. Observe that this allows us to relabel the sequences involved in the AL-AKKT2 conditions so that the entire sequence $\{\rho_k\}_{k\in\mathbb{N}}$ can be assumed to diverge to infinity, which is also assumed. Then there are vectors $\boldsymbol{\mu} \in \mathbb{R}^p_+$ and $\boldsymbol{\lambda} \in \mathbb{R}^q$ and an infinite set of indexes $\mathcal{M} \subset \mathbb{N}$ such that

$$\lim_{k\in\mathcal{M}} \left(\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k)\right)_+ = \boldsymbol{\mu} \qquad \text{and} \qquad \lim_{k\in\mathcal{M}} (\bar{\boldsymbol{\lambda}}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k)) = \boldsymbol{\lambda}.$$

Also, the sequences $\{\rho_k[\boldsymbol{g}(\boldsymbol{x}^k)]_+\}_{k\in\mathbb{N}}$ and $\{\rho_k \boldsymbol{h}(\boldsymbol{x}^k)\}_{k\in\mathbb{N}}$ are bounded, so there is $K > 0$ such that

$$\rho_k \left(\|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\| + \|\boldsymbol{h}(\boldsymbol{x}^k)\|\right) \le K, \text{ for all } k \in \mathbb{N}. \tag{9}$$

Now, by a recursive construction, it is possible to obtain an infinite subset $\mathcal{K}$ of $\mathcal{M}$ so that, for all $i \in I_p$,

$$\lim_{k\in\mathcal{K}} \rho_k g_i(\boldsymbol{x}^k) \in \{-\infty\} \cup \mathbb{R}.$$

Consider the set of indexes

$$\bar{\mathcal{A}} = \left\{ i \in I_p \;\middle|\; \lim_{k\in\mathcal{K}} \rho_k g_i(\boldsymbol{x}^k) \in \mathbb{R} \right\}. \tag{10}$$

We can, for each $k \in \mathbb{N}$, project $\boldsymbol{x}^k$ onto the feasible set Ω. Let us choose $\boldsymbol{y}^k \in \Omega$ to be one of these projections. We have $\boldsymbol{y}^k \in \Omega$ and, since the function that defines the distance from a point to a fixed set is Lipschitz with constant 1, $\|\boldsymbol{x}^k - \boldsymbol{y}^k\| = \operatorname{dist}(\boldsymbol{x}^k, \Omega) = |\operatorname{dist}(\boldsymbol{x}^k, \Omega) - \operatorname{dist}(\boldsymbol{x}^*, \Omega)| \le \|\boldsymbol{x}^k - \boldsymbol{x}^*\|$. Consequently, $\boldsymbol{y}^k \to \boldsymbol{x}^*$. Also, if $i \notin \bar{\mathcal{A}}$, necessarily,

$$\lim_{k\in\mathcal{K}} \rho_k g_i(\boldsymbol{x}^k) = -\infty.$$

Observe that, for each $k \in \mathbb{N}$ big enough, $\rho_k |g_i(\boldsymbol{y}^k) - g_i(\boldsymbol{x}^k)| \le L_{g_i} \rho_k \|\boldsymbol{y}^k - \boldsymbol{x}^k\| = L_{g_i} \rho_k \operatorname{dist}(\boldsymbol{x}^k, \Omega) \le L_{g_i} L_\Omega \rho_k \left(\|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\| + \|\boldsymbol{h}(\boldsymbol{x}^k)\|\right) \le L_{g_i} L_\Omega K$, where L_{g_i} is the local Lipschitz constant of the function g_i and L_Ω is a local error bound constant for Ω at $\boldsymbol{x}^*$ – strong-EB assumes EB, see Definition 4.2. Hence, for each $k \in \mathbb{N}$ big enough,

$$\rho_k g_i(\boldsymbol{y}^k) = \rho_k \left(g_i(\boldsymbol{y}^k) - g_i(\boldsymbol{x}^k)\right) + \rho_k g_i(\boldsymbol{x}^k) \le L_{g_i} L_\Omega K + \rho_k g_i(\boldsymbol{x}^k).$$

Thus, as $\lim_{k\in\mathcal{K}} \rho_k g_i(\boldsymbol{x}^k) = -\infty$, also $\lim_{k\in\mathcal{K}} \rho_k g_i(\boldsymbol{y}^k) = -\infty$. Consequently, since $\lim_{k\in\mathcal{K}} \boldsymbol{y}^k = \boldsymbol{x}^*$ and $\boldsymbol{y}^k \in \Omega$, it cannot be $i \in \mathcal{A}_{=0}$. This all means that $\mathcal{A}_{=0} \subset \bar{\mathcal{A}}$. Additionally, by the divergence to infinity of the sequence of penalty parameters, notice that it cannot be $i \notin \mathcal{A}(\boldsymbol{x}^*)$ and $i \in \bar{\mathcal{A}}$. Thus, $\bar{\mathcal{A}} \subset \mathcal{A}(\boldsymbol{x}^*)$.

Since $\mathcal{A}_{=0} \subset \bar{\mathcal{A}} \subset \mathcal{A}(\boldsymbol{x}^*)$, it is possible to take $\mathcal{A} = \bar{\mathcal{A}}$ and see that, by the definition of the strong-EB condition, see Definition 4.2, EB is valid for the set $\Omega_{\bar{\mathcal{A}}}$ at $\boldsymbol{x}^*$. Consequently, there exists $L > 0$ such that, for all $k \in \mathcal{K}$ big enough,

$$\rho_k \operatorname{dist}(\boldsymbol{x}^k, \Omega_{\bar{\mathcal{A}}}) \le L \rho_k \left(\sum_{i\in\bar{\mathcal{A}}} |g_i(\boldsymbol{x}^k)| + \sum_{j=1}^{q} |h_j(\boldsymbol{x}^k)| \right).$$

Now, considering that all the limits appearing in the definition (10) exist and, consequently, the associated sequences are bounded, the right-hand side of the last inequality must be bounded. Consequently, there exists $M > 0$ such that, for all $k \in \mathcal{K}$,

$$\rho_k \operatorname{dist}(\boldsymbol{x}^k, \Omega_{\bar{\mathcal{A}}}) \leq M.$$

Now, for each $k \in \mathcal{K}$, we can define the set $\mathcal{I}_k \stackrel{\text{def}}{=} \{i \in I_p \mid \rho_k g_i(\boldsymbol{x}^k) + \bar{\mu}_i^k \geq -\eta_k\}$. In this setting, for each $i \in \mathcal{I}_k$ and $k \in \mathcal{K}$, $\rho_k g_i(\boldsymbol{x}^k) \geq -\eta_k - \bar{\mu}_i^k$. Hence, due to (9), $\rho_k |g_i(\boldsymbol{x}^k)| \leq \max\{\eta_k + \bar{\mu}_i^k, K\}$. Consequently, $\mathcal{I}_k \subset \bar{\mathcal{A}}$, for $k \in \mathcal{K}$ big enough.

As $\Omega_{\bar{\mathcal{A}}}$ fulfills error bound, by [31, Proposition 1], it holds

$$\hat{T}_{\Omega_{\bar{\mathcal{A}}}}(\boldsymbol{x}^*) = L_{\Omega_{\bar{\mathcal{A}}}}(\boldsymbol{x}^*).$$

Consequently, considering Remark 4.1, not only $\Omega_{\bar{\mathcal{A}}}$ satisfies Clarke regularity but also Abadie's constraint qualification.

Now, we have all the necessary ingredients to prove the validity of WSOC at $\boldsymbol{x}^*$. Thus, take $\boldsymbol{d} \in \mathscr{C}^W(\boldsymbol{x}^*) = \mathscr{C}^W_{\mathcal{A}(\boldsymbol{x}^*)}(\boldsymbol{x}^*, \boldsymbol{x}^*)$. Hence, considering the sequence AL-AKKT2 $\{\boldsymbol{x}^k\}_{k \in \mathbb{N}}$, we can, for each $k \in \mathbb{N}$, project $\boldsymbol{x}^k$ onto the set $\Omega_{\bar{\mathcal{A}}}$. Let $\bar{\boldsymbol{y}}^k$ be a projection. Observe that $\bar{\boldsymbol{y}}^k \in \Omega_{\bar{\mathcal{A}}}$ and, since the function that defines the distance from a point to a fixed set is Lipschitz, we have also that $\bar{\boldsymbol{y}}^k \to \boldsymbol{x}^*$. Thus, by Lemma 4.1, we can find a sequence $\{\boldsymbol{d}^k\}_{k \in \mathbb{N}}$ converging to $\boldsymbol{d}$ such that $\boldsymbol{d}^k \in \mathscr{C}^W_{\bar{\mathcal{A}}}(\bar{\boldsymbol{y}}^k, \boldsymbol{x}^*)$ for all $k \in \mathbb{N}$. Consequently,

$$\nabla g_i(\bar{\boldsymbol{y}}^k)^T \boldsymbol{d}^k = 0 \text{ and } \nabla h_j(\bar{\boldsymbol{y}}^k)^T \boldsymbol{d}^k = 0, \text{ for all } i \in \bar{\mathcal{A}} \text{ and } j \in I_q.$$

Given $k \in \mathcal{K}$ big enough, as $\mathcal{I}_k \subset \bar{\mathcal{A}}$, it holds

$$\begin{aligned}
\rho_k \sum_{\rho_k g_i(\boldsymbol{x}^k) + \bar{\mu}_i^k \geq -\eta_k} \left(\nabla g_i(\boldsymbol{x}^k)^T \boldsymbol{d}^k\right)^2 &= \rho_k \sum_{i \in \mathcal{I}_k} \left(\left(\nabla g_i(\boldsymbol{x}^k) - \nabla g_i(\bar{\boldsymbol{y}}^k)\right)^T \boldsymbol{d}^k\right)^2 \\
&\leq \rho_k \sum_{i \in \bar{\mathcal{A}}} \left(\left(\nabla g_i(\boldsymbol{x}^k) - \nabla g_i(\bar{\boldsymbol{y}}^k)\right)^T \boldsymbol{d}^k\right)^2 \\
&\leq \rho_k \sum_{i \in \bar{\mathcal{A}}} L_{g_i}^2 \|\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k\|^2 \|\boldsymbol{d}^k\|^2 \\
&= \rho_k \operatorname{dist}(\boldsymbol{x}^k, \Omega_{\bar{\mathcal{A}}}) \sum_{i \in \bar{\mathcal{A}}} L_{g_i}^2 \|\boldsymbol{d}^k\|^2 \operatorname{dist}(\boldsymbol{x}^k, \Omega_{\bar{\mathcal{A}}}) \\
&\leq M \sum_{i \in \bar{\mathcal{A}}} L_{g_i}^2 \|\boldsymbol{d}^k\|^2 \operatorname{dist}(\boldsymbol{x}^k, \Omega_{\bar{\mathcal{A}}}) \stackrel{\mathcal{K}}{\to} 0
\end{aligned}$$

and, similarly,

$$\rho_k \sum_{j=1}^{q} \left(\nabla h_j(\boldsymbol{x}^k)^T \boldsymbol{d}^k\right)^2 \stackrel{\mathcal{K}}{\to} 0.$$

Finally, multiplying (6) by $\boldsymbol{d}^k$ on the right and $(\boldsymbol{d}^k)^T$ on the left, we obtain

$$
\begin{aligned}
&(\boldsymbol{d}^k)^T \left(\nabla^2_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^k, \left(\boldsymbol{\mu}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k)\right)_+, \boldsymbol{\lambda}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k)) \right) \boldsymbol{d}^k \\
&\quad + \rho_k \sum_{\rho_k g_i(\boldsymbol{x}^k) + \bar{\mu}_i^k \geq -\eta_k} \left(\nabla g_i(\boldsymbol{x}^k)^T \boldsymbol{d}^k \right)^2 + \rho_k \sum_{j=1}^{q} \left(\nabla h_j(\boldsymbol{x}^k)^T \boldsymbol{d}^k \right)^2 \geq -\theta_k \|\boldsymbol{d}^k\|^2
\end{aligned}
$$

for each $k \in \mathcal{K}$. Taking the limit w.r.t. $\mathcal{K}$, we get

$$
\begin{aligned}
&\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{p} \mu_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}, \\
&\boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{x}^*) = 0, \quad \boldsymbol{\mu} \geq \boldsymbol{0}, \\
&\boldsymbol{d}^T \left(\nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^{p} \mu_i \nabla^2 g_i(\boldsymbol{x}^*) + \sum_{j=1}^{q} \lambda_j \nabla^2 h_j(\boldsymbol{x}^*) \right) \boldsymbol{d} \geq 0,
\end{aligned}
$$

which implies WSOC, since $\boldsymbol{d}$ is a vector that does not depend on $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$. □

The following result gives a situation where WSOC is guaranteed at points satisfying AL-AKKT2 and strong-EB.

Corollary 4.1 *WSOC hold at AL-AKKT2 points under strong-EB and RQN.*

Proof The expression of the AL-AKKT2 multipliers is the same as the multipliers generated by the first-order SALM. Consequently, only assuming RQN, we can follow the proof of [16, Theorems 3] and prove that the dual sequences of the AL-AKKT2 conditions are bounded. Thus, strong-EB and RQN are enough to guarantee WSOC at AL-AKKT2 points. □

The following result directly applies the previous one and shows that strong-CRSC is enough to guarantee WSOC at points satisfying the AL-AKKT2 conditions.

Proposition 4.3 *Strong-CRSC guarantees strong-EB and RQN.*

Proof First, note that, by hypothesis, CRSC is valid. Thus, Ω not only satisfies EB at $\boldsymbol{x}^*$ (by [12, Theorem 5.5]) but also RQN (by [16, Theorem 2]). Additionally, $\mathcal{J}_-(\boldsymbol{x}^*) \subset \mathcal{A}_{=0}$ by [12, Lemma 5.3]. Hence, let us fix $\mathcal{A}$ with $\mathcal{A}_{=0} \subset \mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$ and consider the set $\Omega_{\mathcal{A}}$. It is only necessary to prove the validity of CRSC at $\boldsymbol{x}^*$ applied to $\Omega_{\mathcal{A}}$, since CRSC implies error bound [12, Theorem 5.5]. As $\Omega_{\mathcal{A}}$ has only equality constraints, it is only necessary to prove the constant rank of the gradients of the active constraints in a neighborhood of $\boldsymbol{x}^*$, i.e., the constant rank of $\{\nabla g_i(\boldsymbol{x})\}_{i \in \mathcal{A}} \cup \{\nabla h_j(\boldsymbol{x})\}_{j \in I_q}$ in a neighborhood of $\boldsymbol{x}^*$. Notice that $\mathcal{J}_-(\boldsymbol{x}^*) \subset \mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$. Consequently, by Definition 2.8, strong-CRSC says that the dimension of the subspace generated by $\{\nabla g_i(\boldsymbol{x})\}_{i \in \mathcal{A}} \cup \{\nabla h_j(\boldsymbol{x})\}_{j \in I_q}$ is constant for $\boldsymbol{x}$ in a neighborhood of $\boldsymbol{x}^*$, as we wanted to prove. □

The next result shows that the quasinormality CQ for $\Omega_{\mathcal{A}(\boldsymbol{x}^*)}$ is enough to guarantee WSOC at points satisfying the AL-AKKT2 conditions.

Proposition 4.4 *If $\Omega_{\mathcal{A}(\boldsymbol{x}^*)}$ satisfies QN, then Ω satisfies strong-EB and RQN.*

Proof Quasinormality applied to the set $\Omega_{\mathcal{A}(\boldsymbol{x}^*)}$ implies the quasinormality CQ applied to the set $\Omega_{\mathcal{A}}$ in which $\mathcal{A}$ satisfies $\mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$. Additionally, quasinormality w.r.t. the feasible set Ω can be derived from that in $\Omega_{\mathcal{A}(\boldsymbol{x}^*)}$. Consequently, not only quasinormality and relaxed quasinormality hold, but also EB for $\Omega_{\mathcal{A}}$ and Ω by [30, Theorem 2.1]. Thus, strong-EB and RQN hold. □

The following example shows that the hypothesis of Proposition 4.4 does not depend on the WCR condition. Thus, although the results previously known in the literature are not broad enough to guarantee WSOC without asking for WCR, the following result shows that this is no longer true.

Example 4.1 Consider the set $\Omega \subset \mathbb{R}^2$ formed by the constraints (example modified from [24, Example 1]): $h(x_1, x_2) = x_1 = 0$ and $g_1(x_1, x_2) = x_1 \mathrm{e}^{x_2} \leq 0$. Additionally, take the point $\boldsymbol{x}^* = (0, 0)^T$. The set Ω does not meet the WCR condition at the origin, but the set $\Omega_{\mathcal{A}(\boldsymbol{x}^*)}$ fulfills quasinormality. Furthermore, notice that the Mangasarian-Fromovitz and constant-rank constraint qualifications are both invalid at $\boldsymbol{x}^*$.

In the following example, WCR holds while strong-EB fails. Thus, both are independent conditions.

Example 4.2 Considering the functions $c_1, c_2, c_3 \colon \mathbb{R}^2 \to \mathbb{R}$ such that $c_1(x_1, x_2) = -x_1$, $c_2(x_1, x_2) = x_2 - x_1^2$ and $c_3(x_1, x_2) = -x_2 - x_1^2$ for each $(x_1, x_2)^T \in \mathbb{R}^2$, the set $\Omega = \{(x_1, x_2)^T \in \mathbb{R}^2 \mid c_1(x_1, x_2) \leq 0,\ c_2(x_1, x_2) \leq 0,\ c_3(x_1, x_2) \leq 0\}$ meets the WCR conditions at the origin. However, considering the subset $\Omega_{\mathcal{A}}$ with $\mathcal{A} = \{2, 3\}$, we have that $\Omega_{\mathcal{A}} = \{\mathbf{0}\}$, also, taking the sequence $\boldsymbol{x}^k = (1/k, 0)^T$, for all $k \in \mathbb{N}$, $\operatorname{dist}(\boldsymbol{x}^k, \Omega_{\mathcal{A}}) = \|\boldsymbol{x}^k\| = 1/k$, but $|c_2(\boldsymbol{x}^k)| = (1/k)^2$ and $|c_3(\boldsymbol{x}^k)| = (1/k)^2$. Hence,

$$\lim_{k \to \infty} \frac{\operatorname{dist}(\boldsymbol{x}^k, \Omega_{\mathcal{A}})}{|c_2(\boldsymbol{x}^k)| + |c_3(\boldsymbol{x}^k)|} = \infty.$$

Consequently, the set $\Omega_{\mathcal{A}}$ does not meet the error bound property at the origin, and strong-EB is violated.

The previous example shows that some problems are not naturally qualified concerning the local EB-type condition. Hence, presenting convergence results for WSOC points involving the WCR condition is still valid. Next, we use the boundedness of AL-AKKT2 multipliers and WCR conditions to obtain WSOC, as done in [11, 24].

Proposition 4.5 *WCR and boundedness of the AL-AKKT2 multipliers are sufficient for WSOC.*

Proof By the boundedness of the AL-AKKT2 multipliers, there are vectors $\boldsymbol{\mu} \in \mathbb{R}^p_+$ and $\boldsymbol{\lambda} \in \mathbb{R}^q$ and an infinite subset $\mathcal{K} \subset \mathbb{N}$ such that $\lim_{k \in \mathcal{K}} (\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k))_+ = \boldsymbol{\mu}$ and

$\lim_{k\in\mathcal{K}} \bar{\boldsymbol{\lambda}}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k) = \boldsymbol{\lambda}$. Now, given $\boldsymbol{d} \in \mathscr{C}^W(\boldsymbol{x}^*)$, we have that the WCR condition guarantees, by [14, Lemma 3.1],

$$\liminf_{\boldsymbol{x}\to\boldsymbol{x}^*} \mathscr{C}^W_{\mathcal{A}(\boldsymbol{x}^*)}(\boldsymbol{x}, \boldsymbol{x}^*) = \mathscr{C}^W(\boldsymbol{x}^*),$$

and thus, considering the AL-AKKT2 sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$, there is a sequence $\{\boldsymbol{d}^k\}_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} \boldsymbol{d}^k = \boldsymbol{d}$, $\nabla g_i(\boldsymbol{x}^k)^T \boldsymbol{d}^k = 0$ and $\nabla h_j(\boldsymbol{x}^k)^T \boldsymbol{d}^k = 0$, for all $i \in \mathcal{A}(\boldsymbol{x}^*)$ and $j \in I_q$, and for all $k \in \mathbb{N}$. Therefore, the first-order part of WSOC holds by taking limits w.r.t. $\mathcal{K}$ in (5). Additionally, the second-order part of WSOC holds by multiplying (6) by $(\boldsymbol{d}^k)^T$ on the left and $\boldsymbol{d}^k$ on the right and by taking limits w.r.t. $\mathcal{K}$ in the resulting expression, similarly to the end of the proof of Proposition 4.2. □

The following two results give practical situations where WSOC can be guaranteed at points satisfying AL-AKKT2.

Corollary 4.2 *WSOC holds at AL-AKKT2 points under WCR and RQN.*

Proof The boundedness of the AL-AKKT2 multipliers can be guaranteed under RQN, as discussed in the proof of the Corollary 4.1. Thus, we have, by Proposition 4.5, the validity of WSOC. □

Proposition 4.6 *Strong-CRSC implies WCR and RQN simultaneously.*

Proof By Definition 2.8, strong-CRSC ensures that the rank of the Jacobian matrix with only the active constraints is constant. Consequently, WCR holds. Additionally, strong-CRSC implies CRSC, which implies RQN [16, Theorem 2]. □

5 The Strongest Possible Sequential Optimality Condition for SALM

In this section, we show that AL-AKKT2 is the strongest second-order sequential optimality condition related to SALM. Part of the theory presented here could be seen as the counterpart of the theory for first-order SALM discussed in [10].

First, we prove that, as expected by the design of Definition 3.1, SALM's feasible accumulation points meet the AL-AKKT2 condition.

Theorem 5.1 *The AL-AKKT2 condition is valid at every feasible accumulation point of SALM.*

Proof Let $\boldsymbol{x}^*$ be a feasible accumulation point of SALM. In step k, an unconstrained optimization algorithm is applied to (3) to obtain $\boldsymbol{x}^k$ that fulfills (4). Hence, the first and second-order stationarity conditions related to AL-AKKT2 are valid.

It remains to prove (7), that is, $\lim_{k\to\infty} \|\boldsymbol{V}^k\|_\infty = 0$. If the sequence $\{\rho_k\}_{k\in\mathbb{N}}$ generated by the method is bounded, 1 and $r < 1$ guarantee $\lim_{k\to\infty} \|\boldsymbol{V}^k\|_\infty = 0$. If $\{\rho_k\}_{k\in\mathbb{N}}$ is unbounded, then by the definition of V_i^k we have $|V_i^k + g_i(\boldsymbol{x}^k)_+| \le \bar{\mu}_i^k/\rho_k$ for all $k \in \mathbb{N}$. Consequently, the boundedness of $\{\bar{\mu}_i^k\}_{k\in\mathbb{N}}$ and the feasibility of $\boldsymbol{x}^*$ imply $\lim_{k\to\infty} V_i^k = 0$ for all $i \in I_p$. This concludes the proof. □

Next, we show that if the sequence of penalty parameters associated with the second-order sequential optimality condition is bounded, a primal sequence generated by SALM exists. This key result is essential to prove that the AL-AKKT2 conditions are the strongest optimality conditions valid at SALM accumulation points.

Proposition 5.1 *Suppose that $\boldsymbol{x}^*$ is a feasible point fulfilling the AL-AKKT2 conditions, where the associated sequence of penalty parameters $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded. Then, there exists a convergent sequence approaching $\boldsymbol{x}^*$ that satisfies all SALM steps for a proper choice of input parameters.*

Proof Let us fix the point $\boldsymbol{x}^*$ for which AL-AKKT2 holds, and consider an associated sequence $\{(\theta_k, \rho_k, \eta_k, \boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$ as in Definition 3.1. By the hypothesis, the AL-AKKT2 multipliers $\boldsymbol{\lambda}^k = (\bar{\boldsymbol{\lambda}}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k))$ and $\boldsymbol{\mu}^k = (\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k))_+$ are bounded. Thus, there exists an infinite subset $\mathcal{K}_0 \subset \mathbb{N}$ such that

$$\lim_{k\in\mathcal{K}_0} (\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) = (\boldsymbol{\lambda}, \boldsymbol{\mu}) \quad \text{and} \quad \lim_{k\in\mathcal{K}_0} \rho_k = \rho_\infty,$$

with $\boldsymbol{\mu} \geq 0$ and $\rho_\infty > 0$. Furthermore, following the proof of Proposition 4.1, $(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is a KKT triple.

Now, it is necessary to construct a convergent sequence fulfilling all the steps of the SALM. In fact, for each $k \in \mathcal{K}_0$, let us define $\mathcal{A}^k = \{i \in I_p \mid \rho_k g_i(\boldsymbol{x}^k) + \bar{\mu}_i^k \geq -\eta_k\}$. As the number of subsets of I_p is finite, there are an infinite subset of natural numbers $\mathcal{M} \subset \mathcal{K}_0$ and a finite set $\mathcal{A}$ with $\mathcal{A} \subset I_p$, such that $\mathcal{A}^k = \mathcal{A}$ for all $k \in \mathcal{M}$. In such a way, by (6), given $k \in \mathcal{M}$,

$$\begin{aligned}
\nabla_{\boldsymbol{x}}^2 \mathcal{L}_{\rho_k}^{\eta_k}(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k) = {} & \nabla_{\boldsymbol{x}}^2 \mathcal{L}(\boldsymbol{x}^k, (\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k))_+, \bar{\boldsymbol{\lambda}}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k)) \\
& + \rho_k \sum_{i\in\mathcal{A}} \nabla g_i(\boldsymbol{x}^k) \nabla g_i(\boldsymbol{x}^k)^T \\
& + \rho_k \sum_{j=1}^{q} \nabla h_j(\boldsymbol{x}^k) \nabla h_j(\boldsymbol{x}^k)^T \succeq -\theta_k \mathbb{I}.
\end{aligned}$$

Consequently, taking the limit w.r.t. $\mathcal{M}$ in the above expression, $(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda})$ is a KKT triple such that, for some matrix H,

$$\begin{aligned}
H = {} & \nabla_{\boldsymbol{x}}^2 \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) \\
& + \rho_\infty \sum_{i\in\mathcal{A}} \nabla g_i(\boldsymbol{x}^*) \nabla g_i(\boldsymbol{x}^*)^T + \rho_\infty \sum_{j=1}^{q} \nabla h_j(\boldsymbol{x}^*) \nabla h_j(\boldsymbol{x}^*)^T \succeq 0,
\end{aligned}$$

Now, as $\lim_{k\to\infty} \min\{-g_i(\boldsymbol{x}^k), \bar{\mu}_i^k/\rho_k\} = 0$, if $g_i(\boldsymbol{x}^*) < 0$, necessarily, $\lim_{k\to\infty} \bar{\mu}_i^k = 0$. This means that there cannot exist an $i \in \mathcal{A}$ such that $i \notin \mathcal{A}(\boldsymbol{x}^*)$. Thus, $\mathcal{A} \subset \mathcal{A}(\boldsymbol{x}^*)$. Furthermore, notice that, for any $\eta \in \mathbb{R}_+$, $\mathcal{A}(\boldsymbol{x}^*) \subset \{i \in I_p \mid \rho_\infty g_i(\boldsymbol{x}^*) + \mu_i \geq -\eta\}$. All these observations implies that, for any $\eta \in \mathbb{R}_+$,

$$\nabla_{\boldsymbol{x}}^2 \mathcal{L}_{\rho_\infty}^{\eta}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) \succeq H, \tag{11}$$

since $\nabla^2_{\boldsymbol{x}} \mathcal{L}^{\eta}_{\rho_\infty}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda})$ is equal to H added to positive multiples of positive semi-definite matrices $\nabla g_i(\boldsymbol{x}^*)\nabla g_i(\boldsymbol{x}^*)^T$ with indexes in $\{i \in I_p \mid \rho_\infty g_i(\boldsymbol{x}^*) + \mu_i \geq -\eta\}$ but not in $\mathcal{A}$.

Now, choosing $\{\tilde{\eta}_k\}_{k\in\mathbb{N}}$ to be any positive sequence converging to zero, since $\nabla_{\boldsymbol{x}} \mathcal{L}_{\rho_\infty}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \boldsymbol{0}$ and (11) hold with the terms of the sequence $\{\tilde{\eta}_k\}_{k\in\mathbb{N}}$ in place of η, for all k we have

$$\nabla_{\boldsymbol{x}} \mathcal{L}_{\tilde{\rho}_k}(\tilde{\boldsymbol{x}}^k, \tilde{\boldsymbol{\mu}}^k, \tilde{\boldsymbol{\lambda}}^k) = \boldsymbol{0} \quad \text{and} \quad \nabla^2_{\boldsymbol{x}} \mathcal{L}^{\tilde{\eta}_k}_{\tilde{\rho}_k}(\tilde{\boldsymbol{x}}^k, \tilde{\boldsymbol{\mu}}^k, \tilde{\boldsymbol{\lambda}}^k) \succeq H, \tag{12}$$

with the constant sequences $\{\tilde{\boldsymbol{x}}^k = \boldsymbol{x}^*\}_{k\in\mathbb{N}}$, $\{\tilde{\boldsymbol{\mu}}^k = \boldsymbol{\mu}\}_{k\in\mathbb{N}}$, $\{\tilde{\boldsymbol{\lambda}}^k = \boldsymbol{\lambda}\}_{k\in\mathbb{N}}$ and $\{\tilde{\rho}_k = \rho_\infty\}_{k\in\mathbb{N}}$. Additionally, notice that, since, for all $k \in \mathbb{N}$,

$$\tilde{\boldsymbol{V}}^k \overset{\text{def}}{=} \min\left\{-\boldsymbol{g}(\tilde{\boldsymbol{x}}^k), \tilde{\boldsymbol{\mu}}^k/\tilde{\rho}_k\right\} = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{h}(\tilde{\boldsymbol{x}}^k) = \boldsymbol{0},$$

we have that the following measure always vanish:

$$\max\left\{\|\tilde{\boldsymbol{V}}^k\|, \|\boldsymbol{h}(\tilde{\boldsymbol{x}}^k)\|\right\} = 0, \quad \forall k \in \mathbb{N}. \tag{13}$$

Finally, it is necessary to define the input parameters of the SALM. In fact, we can choose $\{\tilde{\theta}_k\}_{k\in\mathbb{N}}$ to be a positive sequence converging to zero in which its terms are associated with the criterion in (4), $\mu_{\max}$, $\lambda_{\max}$, and $\lambda_{\min}$ to be large enough in modulus to satisfy $\tilde{\boldsymbol{\mu}}^k = \boldsymbol{\mu} \in [0, \mu_{\max}]^p$ and $\tilde{\boldsymbol{\lambda}}^k = \boldsymbol{\lambda} \in [\lambda_{\min}, \lambda_{\max}]^q$, for all k. In such a way, if we put $\tilde{\rho}_1 = \rho_\infty$ as the input penalty parameter, since (12) and (13) hold, for all $k \in \mathbb{N}$, $\tilde{\boldsymbol{x}}^k$ respects 1 (with the criterion (4)) and 1 is always done maintaining the penalty parameter unchanged from the previous iteration, which aligns with the constructed sequence of penalty parameters. Thus, the constant primal sequence $\{\tilde{\boldsymbol{x}}^k\}_{k\in\mathbb{N}}$ respects all the steps of SALM. □

The following lemma shows that the Hessian surrogate of the PHR-augmented Lagrangian function (see Definition 2.2) satisfies a lower semi-continuity property. This property is crucial for proving that the limit points of SALM are exactly the AL-AKKT2 points.

Lemma 5.1 *Assume that all functions describing problem (1) are twice continuously differentiable. For any $\rho \in \mathbb{R}_{++}$, $\eta, \tilde{\eta} \in \mathbb{R}_+$ with $\eta > \tilde{\eta}$, $(\boldsymbol{z}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}) \in \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^q$ and $\epsilon \in \mathbb{R}_{++}$, there is $\delta > 0$ such that for any $(\boldsymbol{x}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \in \mathcal{B}((\boldsymbol{z}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}), \delta)$,*

$$\nabla^2_{\boldsymbol{x}} \mathcal{L}^{\eta}_{\rho}(\boldsymbol{x}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \succeq \nabla^2_{\boldsymbol{x}} \mathcal{L}^{\tilde{\eta}}_{\rho}(\boldsymbol{z}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}) - \epsilon \mathbb{I}. \tag{14}$$

Proof First, notice that the function $(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \mapsto \nabla^2_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ is continuous. Now, the function $(\boldsymbol{x}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \mapsto (\boldsymbol{x}, (\bar{\boldsymbol{\mu}} + \rho \boldsymbol{g}(\boldsymbol{x}))_+, \bar{\boldsymbol{\lambda}} + \rho \boldsymbol{h}(\boldsymbol{x}))$ is also continuous. Thus, their composition is continuous. Furthermore, with ρ fixed, the functions $\boldsymbol{x} \mapsto \rho \nabla g_i(\boldsymbol{x}) \nabla g_i(\boldsymbol{x})^T$ and $\boldsymbol{x} \mapsto \rho \nabla h_j(\boldsymbol{x}) \nabla h_j(\boldsymbol{x})^T$ are all continuous, with $i \in I_p$

and $j \in I_q$. Thus, for the given $\epsilon > 0$, we can take a small $\delta_0 > 0$ so that, for all $(\boldsymbol{x}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \in \mathcal{B}((z, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}), \delta_0)$,

$$\nabla_{\boldsymbol{x}}^2 \mathcal{L}(\boldsymbol{x}, (\bar{\boldsymbol{\mu}} + \rho \boldsymbol{g}(\boldsymbol{x}))_+, \bar{\boldsymbol{\lambda}} + \rho \boldsymbol{h}(\boldsymbol{x})) - \nabla_{\boldsymbol{x}}^2 \mathcal{L}(z, (\tilde{\boldsymbol{\mu}} + \rho \boldsymbol{g}(z))_+, \tilde{\boldsymbol{\lambda}} + \rho \boldsymbol{h}(z)) \succeq -\epsilon/6\mathbb{I}, \tag{15}$$

$$\rho \left(\nabla g_i(\boldsymbol{x}) \nabla g_i(\boldsymbol{x})^T - \nabla g_i(z) \nabla g_i(z)^T \right) \succeq -\epsilon/(6p)\mathbb{I}, \quad \forall i \in I_p, \tag{16}$$

$$\text{and} \quad \rho \left(\nabla h_j(\boldsymbol{x}) \nabla h_j(\boldsymbol{x})^T - \nabla h_j(z) \nabla h_j(z)^T \right) \succeq -\epsilon/(6q)\mathbb{I}, \quad \forall j \in I_q. \tag{17}$$

In addition, take $\delta < \delta_0$ small enough so that if $(\boldsymbol{x}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \in \mathcal{B}((z, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}), \delta)$, then

$$\frac{\eta - \tilde{\eta}}{4\rho} > \|\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}(z)\|_\infty, \quad \frac{\eta - \tilde{\eta}}{4} > \|\bar{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}\|_\infty.$$

Under these considerations, for any $(\boldsymbol{x}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \in \mathcal{B}((z, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}), \delta)$, defining $\mathcal{I}_z \stackrel{\text{def}}{=} \{i \in I_p \mid \rho g_i(z) + \tilde{\mu}_i \geq -\tilde{\eta}\}$ and $\mathcal{I}_{\boldsymbol{x}} \stackrel{\text{def}}{=} \{i \in I_p \mid \rho g_i(\boldsymbol{x}) + \bar{\mu}_i \geq -\eta\}$, we have $\mathcal{I}_z \subset \mathcal{I}_{\boldsymbol{x}}$, since if $\rho g_i(z) + \tilde{\mu}_i \geq -\tilde{\eta}$, then $\rho g_i(\boldsymbol{x}) + \bar{\mu}_i = \rho g_i(z) + \tilde{\mu}_i + \rho(g_i(\boldsymbol{x}) - g_i(z)) + (\bar{\mu}_i - \tilde{\mu}_i) \geq -\tilde{\eta} - (\eta - \tilde{\eta})/2 \geq -\eta$. Consequently, we can sum the expressions (15), (16) with indexes in $i \in \mathcal{I}_z$ and (17) with indexes in $j \in I_q$, obtaining exactly

$$\nabla_{\boldsymbol{x}}^2 \mathcal{L}_\rho^\eta(\boldsymbol{x}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) - \nabla_{\boldsymbol{x}}^2 \mathcal{L}_\rho^{\tilde{\eta}}(z, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}) \succeq \rho \sum_{\ell \in \mathcal{I}_{\boldsymbol{x}} \setminus \mathcal{I}_z} \nabla g_\ell(\boldsymbol{x}) \nabla g_\ell(\boldsymbol{x})^T - (|\mathcal{I}_z|/p + 2)\epsilon/6\mathbb{I},$$

with $|\mathcal{I}_z|$ being the number of elements in $\mathcal{I}_z$. Thus, since $\nabla g_i(\boldsymbol{x}) \nabla g_i(\boldsymbol{x})^T$ are always positive semi-definite for all $i \in I_p$, we have (14). □

Next, we show that when the AL-AKKT2 condition is ensured, it is possible to find a sequence accepted by SALM for a particular choice of input parameters. This is consistent with previous results for the first-order SALM discussed in [10]. In other words, AL-AKKT2 provides the strongest global convergence result for SALM.

Theorem 5.2 *Let $\boldsymbol{x}^*$ be a feasible point of problem (1) satisfying the AL-AKKT2 conditions, then, for a proper choice of input parameters, there is a convergent sequence to $\boldsymbol{x}^*$ fulfilling the steps of SALM.*

Proof The proof is separated into two cases: $\boldsymbol{x}^*$ is an unconstrained second-order stationary point for the objective function f or not. In the first case, choosing $\{\tilde{\eta}_k\}_{k \in \mathbb{N}}$ to be any non-negative sequence converging to zero, notice that, for all k, (12) and (13) hold with $H = 0$, $\{\tilde{\boldsymbol{x}}^k = \boldsymbol{x}^*\}_{k \in \mathbb{N}}$, $\{\tilde{\boldsymbol{\mu}}^k = \boldsymbol{0}\}_{k \in \mathbb{N}}$, $\{\tilde{\boldsymbol{\lambda}}^k = \boldsymbol{0}\}_{k \in \mathbb{N}}$, $\{\tilde{\rho}_k = \rho_\infty\}_{k \in \mathbb{N}}$ and $\rho_\infty \in \mathbb{R}_{++}$. Thus, with similar reasoning, we might define the input parameters as the second part of the proof of the Proposition 5.1 and obtain that the constant sequence $\{\tilde{\boldsymbol{x}}^k\}_{k \in \mathbb{N}}$ satisfies all the steps of the SALM.

In the second case, suppose that $\boldsymbol{x}^*$ is not an unconstrained second-order stationary point for the function f. Additionally, consider the sequence $\{(\theta_k, \rho_k, \tilde{\eta}_k, \boldsymbol{z}^k, \tilde{\boldsymbol{\mu}}^k, \tilde{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$, which satisfies the AL-AKKT2 conditions and has $\{\boldsymbol{z}^k\}_{k\in\mathbb{N}}$ converging to $\boldsymbol{x}^*$ (see Definition 3.1, noting the ordering in the tuple). Next, we introduce a small perturbation in the sequence $\{(\theta_k, \rho_k, \tilde{\eta}_k, \boldsymbol{z}^k, \tilde{\boldsymbol{\mu}}^k, \tilde{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$. To do so, define the set

$$\Omega_{I_p} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\},$$

and observe that there cannot exist an infinite subset $\mathcal{K} \subset \mathbb{N}$ such that, for each $k \in \mathcal{K}$, there is at least one open neighborhood V of $\boldsymbol{z}^k$ satisfying $\boldsymbol{x} \in \Omega_{I_p}$ for all $\boldsymbol{x} \in V$. If, for the sake of contradiction, such a subset $\mathcal{K}$ did exist, then for all $k \in \mathcal{K}$, the expressions (5) and (6) would simplify to

$$\|\nabla f(\boldsymbol{z}^k)\| \leq \theta_k \quad \text{and} \quad \nabla^2 f(\boldsymbol{z}^k) \succeq -\theta_k \mathbb{I}.$$

Taking the limit w.r.t $\mathcal{K}$ would then imply that $\boldsymbol{x}^*$ is an unconstrained second-order stationary point of f, which leads to a contradiction. Therefore, for each sufficiently large k, there exists a sequence $\{\boldsymbol{y}_\ell^k\}_{\ell\in\mathbb{N}}$ converging to $\boldsymbol{z}^k$ such that $\boldsymbol{y}_\ell^k \notin \Omega_{I_p}$ for all $\ell \in \mathbb{N}$. Next, for simplicity, let us assume that, for every natural k, such a sequence exists.

Now, for each $k \in \mathbb{N}$, define uniformly bounded sequences $\{\bar{\boldsymbol{\mu}}_\ell^k\}_{\ell\in\mathbb{N}}$ and $\{\bar{\boldsymbol{\lambda}}_\ell^k\}_{\ell\in\mathbb{N}}$ that converge to $\tilde{\boldsymbol{\mu}}^k$ and $\tilde{\boldsymbol{\lambda}}^k$, respectively, with $\bar{\boldsymbol{\mu}}_\ell^k > \tilde{\boldsymbol{\mu}}^k$ for all $\ell \in \mathbb{N}$. Furthermore, take $\{\eta_k\}_{k\in\mathbb{N}}$, converging to zero, with $\eta_k > \tilde{\eta}_k$ for all $k \in \mathbb{N}$. In this setting, for each k, considering that the point of interest is $(\boldsymbol{z}^k, \tilde{\boldsymbol{\mu}}^k, \tilde{\boldsymbol{\lambda}}^k)$ and the fixed penalty parameter is ρ_k, we apply Lemma 5.1 and conclude that, for all sufficiently large ℓ,

$$\nabla_{\boldsymbol{x}}^2 \mathcal{L}_{\rho_k}^{\eta_k}(\boldsymbol{y}_\ell^k, \bar{\boldsymbol{\mu}}_\ell^k, \bar{\boldsymbol{\lambda}}_\ell^k) \succeq \nabla_{\boldsymbol{x}}^2 \mathcal{L}_{\rho_k}^{\tilde{\eta}_k}(\boldsymbol{z}^k, \tilde{\boldsymbol{\mu}}^k, \tilde{\boldsymbol{\lambda}}^k) - \theta_k \mathbb{I}.$$

Additionally, notice that

$$\lim_{\ell\to\infty} \nabla_{\boldsymbol{x}} \mathcal{L}_{\rho_k}(\boldsymbol{y}_\ell^k, \bar{\boldsymbol{\mu}}_\ell^k, \bar{\boldsymbol{\lambda}}_\ell^k) = \nabla_{\boldsymbol{x}} \mathcal{L}_{\rho_k}(\boldsymbol{z}^k, \tilde{\boldsymbol{\mu}}^k, \tilde{\boldsymbol{\lambda}}^k).$$

Consequently, for sufficiently large ℓ, we have

$$\|\nabla_{\boldsymbol{x}} \mathcal{L}_{\rho_k}(\boldsymbol{y}_\ell^k, \bar{\boldsymbol{\mu}}_\ell^k, \bar{\boldsymbol{\lambda}}_\ell^k)\| < 2\theta_k, \quad \text{and} \quad \nabla_{\boldsymbol{x}}^2 \mathcal{L}_{\rho_k}^{\eta_k}(\boldsymbol{y}_\ell^k, \bar{\boldsymbol{\mu}}_\ell^k, \bar{\boldsymbol{\lambda}}_\ell^k) \succeq -2\theta_k \mathbb{I}. \tag{18}$$

From these observations, for each $k \in \mathbb{N}$, we can select a term from the sequence $\{(\boldsymbol{y}_\ell^k, \bar{\boldsymbol{\mu}}_\ell^k, \bar{\boldsymbol{\lambda}}_\ell^k)\}_{\ell\in\mathbb{N}}$ satisfying (18) and define it as $(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)$, so that the sequence $\{(2\theta_k, \rho_k, \eta_k, \boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$ still satisfies the AL-AKKT2 conditions as in Definition 3.1, with $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ converging to $\boldsymbol{x}^*$, but now with $\bar{\boldsymbol{\mu}}^k > \boldsymbol{0}$ and $\boldsymbol{x}^k \notin \Omega_{I_p}$ for all $k \in \mathbb{N}$.

With the established modification, it is ensured that

$$\boldsymbol{V}^k \stackrel{\text{def}}{=} \min\left\{-\boldsymbol{g}(\boldsymbol{x}^k), \frac{\bar{\boldsymbol{\mu}}^k}{\rho_k}\right\} \neq \boldsymbol{0} \qquad \text{or} \qquad \boldsymbol{h}(\boldsymbol{x}^k) \neq \boldsymbol{0} \tag{19}$$

for all k. In fact, if this were not the case, for some k, since $\boldsymbol{x}^k \notin \Omega_{I_p}$, it must be $\boldsymbol{g}(\boldsymbol{x}^k) \neq \boldsymbol{0}$. Thus, there exists $i \in I_p$ with $g_i(\boldsymbol{x}^k) \neq 0$. In this case, since $V_i^k = 0$, it holds $\bar{\mu}_i^k = 0$. However, $\bar{\boldsymbol{\mu}}^k > \boldsymbol{0}$, which leads to an absurd. Therefore, (19) holds for all k.

Now, if $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, Proposition 5.1 allows us to conclude the result. On the other hand, if $\{\rho_k\}_{k\in\mathbb{N}}$ is unbounded, fixing any $r < 1$ and $\gamma > 1$, it is possible to recursively construct a strictly increasing sequence of natural numbers $\{\sigma_k\}_{k\in\mathbb{N}}$ such that, for all k,

$$\begin{aligned} &\rho_{\sigma_{k+1}} \geq \gamma \rho_{\sigma_k} \\ \text{and} \quad &\max\left\{\|\boldsymbol{V}^{\sigma_k}\|_\infty, \|\boldsymbol{h}(\boldsymbol{x}^{\sigma_k})\|_\infty\right\} \leq r \max\left\{\|\boldsymbol{V}^{\sigma_{k-1}}\|_\infty, \|\boldsymbol{h}(\boldsymbol{x}^{\sigma_{k-1}})\|_\infty\right\}. \end{aligned}$$

Define the new sequence $\{\bar{\sigma}_k\}_{k\in\mathbb{N}}$ by putting

$$\bar{\sigma}_k \stackrel{\text{def}}{=} \begin{cases} \sigma_{k/2} & \text{, if } k \text{ is even,} \\ \sigma_{(k+1)/2} & \text{, if } k \text{ is odd,} \end{cases}$$

for all $k \in \mathbb{N}$. As can be seen, the terms repeat for two iterations before changing their value:

$$\{\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3, \bar{\sigma}_4, \bar{\sigma}_5, \bar{\sigma}_6, \cdots\} = \{\sigma_1, \sigma_1, \sigma_2, \sigma_2, \sigma_3, \sigma_3, \cdots\}. \tag{20}$$

Let us take, for all $k \in \mathbb{N}$,

$$\begin{aligned} \bar{\rho}_k \stackrel{\text{def}}{=} \rho_{\bar{\sigma}_k}, \quad \boldsymbol{y}^k \stackrel{\text{def}}{=} \boldsymbol{x}^{\bar{\sigma}_k}, \quad \bar{\bar{\boldsymbol{\mu}}}^k \stackrel{\text{def}}{=} \bar{\boldsymbol{\mu}}^{\bar{\sigma}_k}, \quad \bar{\bar{\boldsymbol{\lambda}}}^k \stackrel{\text{def}}{=} \bar{\boldsymbol{\lambda}}^{\bar{\sigma}_k}, \\ \bar{\boldsymbol{V}}^k \stackrel{\text{def}}{=} \boldsymbol{V}^{\bar{\sigma}_k} = \min\left\{-\boldsymbol{g}(\boldsymbol{x}^{\bar{\sigma}_k}), \frac{\bar{\boldsymbol{\mu}}^{\bar{\sigma}_k}}{\rho_{\bar{\sigma}_k}}\right\}. \end{aligned} \tag{21}$$

We proceed by showing that these sequences respect the SALM steps. When $k = 1$, as it is easy to verify, $\bar{\sigma}_k = \bar{\sigma}_{k+1}$ and $\bar{\rho}_k = \bar{\rho}_{k+1}$ (see (20)), following 1 and all other steps of the method in the first iteration. Then, we analyze separately when k is even or odd for $k > 1$.

If k is even, we have $k = 2\ell$ for some ℓ and thus $\bar{\sigma}_k = \bar{\sigma}_{k-1}$ (see (20)). Furthermore,

$$\bar{\sigma}_{k+1} = \bar{\sigma}_{2\ell+1} = \sigma_{((2\ell+1)+1)/2} = \sigma_{\ell+1} \quad \text{and} \quad \bar{\sigma}_k = \bar{\sigma}_{2\ell} = \sigma_{2\ell/2} = \sigma_\ell.$$

Therefore, $\bar{\rho}_{k+1} = \rho_{\bar{\sigma}_{k+1}} = \rho_{\sigma_{\ell+1}} \geq \gamma\rho_{\sigma_\ell} = \gamma\rho_{\bar{\sigma}_k} = \gamma\bar{\rho}_k$ and

$$\begin{aligned}\max\left\{\|\bar{\boldsymbol{V}}^k\|_\infty, \|\boldsymbol{h}(\boldsymbol{y}^k)\|_\infty\right\} &= \max\left\{\|\boldsymbol{V}^{\bar{\sigma}_k}\|_\infty, \|\boldsymbol{h}(\boldsymbol{x}^{\bar{\sigma}_k})\|_\infty\right\}\\ &= \max\left\{\|\boldsymbol{V}^{\bar{\sigma}_{k-1}}\|_\infty, \|\boldsymbol{h}(\boldsymbol{x}^{\bar{\sigma}_{k-1}})\|_\infty\right\}\\ &= \max\left\{\|\bar{\boldsymbol{V}}^{k-1}\|_\infty, \|\boldsymbol{h}(\boldsymbol{y}^{k-1})\|_\infty\right\}.\end{aligned}$$

Thus, since (19) holds, $(\bar{\boldsymbol{V}}^k, \boldsymbol{h}(\boldsymbol{y}^k))$ does not vanish. Consequently,

$$\begin{aligned}\max\left\{\|\bar{\boldsymbol{V}}^k\|_\infty, \|\boldsymbol{h}(\boldsymbol{y}^k)\|_\infty\right\} &> r\max\left\{\|\bar{\boldsymbol{V}}^k\|_\infty, \|\boldsymbol{h}(\boldsymbol{y}^k)\|_\infty\right\}\\ &= r\max\left\{\|\bar{\boldsymbol{V}}^{k-1}\|_\infty, \|\boldsymbol{h}(\boldsymbol{y}^{k-1})\|_\infty\right\},\end{aligned}$$

agreeing with 1 from SALM when $\bar{\rho}_{k+1} \geq \gamma\bar{\rho}_k$.

Now, suppose that $k = 2\ell + 1$ for some $\ell \in \mathbb{N}$. Similarly,

$$\begin{aligned}&\bar{\sigma}_{k-1} = \bar{\sigma}_{(2\ell+1)-1} = \bar{\sigma}_{2\ell} = \sigma_{2\ell/2} = \sigma_\ell \quad \text{and}\\ &\bar{\sigma}_{k+1} = \bar{\sigma}_k = \bar{\sigma}_{(2\ell+1)+1} = \sigma_{((2\ell+1)+1)/2} = \sigma_{\ell+1}\end{aligned}$$

and so

$$\begin{aligned}\max\left\{\|\bar{\boldsymbol{V}}^k\|_\infty, \|\boldsymbol{h}(\boldsymbol{y}^k)\|_\infty\right\} &= \max\left\{\|\boldsymbol{V}^{\bar{\sigma}_k}\|_\infty, \|\boldsymbol{h}(\boldsymbol{x}^{\bar{\sigma}_k})\|_\infty\right\}\\ &= \max\left\{\|\boldsymbol{V}^{\sigma_{\ell+1}}\|_\infty, \|\boldsymbol{h}(\boldsymbol{x}^{\sigma_{\ell+1}})\|_\infty\right\}\\ &\leq r\max\left\{\|\boldsymbol{V}^{\sigma_\ell}\|_\infty, \|\boldsymbol{h}(\boldsymbol{x}^{\sigma_\ell})\|_\infty\right\}\\ &= r\max\left\{\|\boldsymbol{V}^{\bar{\sigma}_{k-1}}\|_\infty, \|\boldsymbol{h}(\boldsymbol{x}^{\bar{\sigma}_{k-1}})\|_\infty\right\}\\ &= r\max\left\{\|\bar{\boldsymbol{V}}^{k-1}\|_\infty, \|\boldsymbol{h}(\boldsymbol{y}^{k-1})\|_\infty\right\},\end{aligned}$$

which corroborates $\bar{\rho}_{k+1} = \bar{\rho}_k$. Thus, 1 is being respected.

The validity of the first- and second-order stationarity stated in equation (4) can be seen by noticing, considering expression (20), that the sequences $\{\boldsymbol{y}^k\}$, $\{\left(\bar{\bar{\boldsymbol{\mu}}}^k + \bar{\rho}_k\boldsymbol{g}(\boldsymbol{y}^k)\right)_+\}_{k\in\mathbb{N}}$, $\{(\bar{\bar{\boldsymbol{\lambda}}}^k + \bar{\rho}_k\boldsymbol{h}(\boldsymbol{y}^k))\}_{k\in\mathbb{N}}$ may be regarded as sequences containing some repeated terms of the AL-AKKT2 sequences $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$, $\{\left(\bar{\boldsymbol{\mu}}^k + \rho_k\boldsymbol{g}(\boldsymbol{x}^k)\right)_+\}_{k\in\mathbb{N}}$ and $\{(\bar{\boldsymbol{\lambda}}^k + \rho_k\boldsymbol{h}(\boldsymbol{x}^k))\}_{k\in\mathbb{N}}$. Hence, the defined sequences in (21) must also be AL-AKKT2 sequences.

The ideas regarding SALM's input parameters are straightforward. Here, it is only needed to choose $\mu_{\max}$, $\lambda_{\max}$ and $\lambda_{\min}$ to be big enough in modulus so that the bounded sequences satisfy $\bar{\bar{\boldsymbol{\mu}}}^k \in [0, \mu_{\max}]^p$ and $\bar{\bar{\boldsymbol{\lambda}}}^k \in [\lambda_{\min}, \lambda_{\max}]^q$ for all $k \in \mathbb{N}$. Notice that the sequences $\{\theta_k\}_{k\in\mathbb{N}}$ and $\{\eta_k\}_{k\in\mathbb{N}}$ in SALM can be taken as those associated with AL-AKKT2 (compare SALM's requirements with Definition 3.1). In such a way, $\{\boldsymbol{y}^k\}_{k\in\mathbb{N}}$ respects all the steps of SALM, concluding the proof. □

Remark 5.1 Our proof is built upon the original argument presented in [10, Theorem 1], and, as such, it shares certain similarities. However, the original proof may fail in extreme cases. Let $\{x^k\}_{k\in\mathbb{N}}$ be a primal AL-AKKT sequence and $\{(\bar{\mu}^k + \rho_k g(x^k))_+\}_{k\in\mathbb{N}}$ and $\{(\rho_k h(x^k) + \bar{\lambda}^k)\}_{k\in\mathbb{N}}$ be the associated dual sequences; cf. [10, Definition 1]. The proof fails when the given AL-AKKT sequences do not satisfy both conditions in (19) for infinitely many iterations, while the penalty parameter sequence remains unbounded. Specifically, within the notation of [10, Theorem 1], the case of consecutive elements where $\tilde{x}^{i_k-1/2} \rightsquigarrow \tilde{x}^{i_k}$ is not adequately addressed, because the term $\|\tilde{V}^{i_k-1/2}\|$ may vanish. This would invalidate the inequality $\|\tilde{V}^{i_k-1/2}\| > \tau\|\tilde{V}^{i_k-1/2}\|$, which is crucial for the proof to hold; see [10, Theorem 1].

Several approaches can be considered to address the issue in the original proof. One effective strategy is to divide the original proof into two cases: one where (19) fails for infinitely many iterations, and one where it does not. In the first case, when (19) fails for infinitely many iterations, it suffices to recognize that the KKT conditions are satisfied at the accumulation point of interest. Consequently, the first-order SALM with constant primal sequences converges to this KKT point, as rigorously analyzed in [10, Theorem 1] for the case where the penalty parameter sequence is bounded. In the second case, when (19) holds for sufficiently large indices k, we can proceed with the proof as originally presented, as the original technique remains valid for such values of k.

The next result is a direct consequence of Theorems 5.1 and 5.2.

Corollary 5.1 *AL-AKKT2 is the strongest sequential optimality condition valid at SALM accumulation points. In particular, it implies AKKT2.*

Remark 5.2 An analogous result to the above corollary is valid for CAKKT2 instead of AKKT2, provided that the quadratic-like penalty measure of infeasibility associated with problem (1) satisfies the generalized Kurdyka-Lojasiewicz (GKL) inequality [15], a mild hypothesis valid, for example, if the data functions of the problem are subanalytic. GKL is required for SALM to converge to CAKKT/CAKKT2 points, see [15, 24].

5.1 The Least Stringent CQ Associated With SALM

In this section, we present the least stringent CQ associated with AL-AKKT2. In other words, in view of Corollary 5.1, we state the weakest possible CQ for which SALM accumulation points fulfill WSOC. Evidently, the weaker the CQ used, the better the method's theoretical reliability, so ultimately, we are expanding the class of problems for which SALM converges.

As we already mentioned, whenever an SOC2 is proposed, it is possible to construct its weakest companion CQ following the techniques in [13] or [11]. Next, we recall some concepts necessary to build such a CQ for AL-AKKT2.

Consider a multivalued function $\mathcal{F}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $C \subset \mathbb{R}^n$. The *Painlevé–Kuratowski outer/upper limit* [36, p. 152] of $\mathcal{F}(x)$ as $x \to x^*$ is the set

$$\limsup_{x \xrightarrow[C]{} x^*} \mathcal{F}(x) \stackrel{\text{def}}{=} \{v \in \mathbb{R}^n \mid \exists\{x^k\}_{k\in\mathbb{N}}, \exists\{v^k\}_{k\in\mathbb{N}} \text{ such that}$$

$$\lim_{k\to\infty} \boldsymbol{v}^k = \boldsymbol{v}, \lim_{k\to\infty} \boldsymbol{x}^k = \boldsymbol{x}^*, \boldsymbol{x}^k \in C, \forall k \in \mathbb{N}, \text{ and } \boldsymbol{v}^k \in \mathcal{F}(\boldsymbol{x}^k), \forall k \in \mathbb{N}\}.$$

When it holds $\limsup_{\boldsymbol{x}\underset{C}{\longrightarrow}\boldsymbol{x}^*} \mathcal{F}(\boldsymbol{x}) \subset \mathcal{F}(\boldsymbol{x}^*)$, the multifunction $\mathcal{F}$ is said to be outer semicontinuous in $\boldsymbol{x}^*$ relative to C.

Below, we define the condition we call the weakest CQ for obtaining WSOC points associated with SALM's accumulation points.

Definition 5.1 Define the multivalued function $K^W : \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^q \times \mathbb{R}_{++} \times \mathbb{R}_{++} \rightrightarrows \mathbb{R}^n$ such that

$$K^W(\boldsymbol{x}, \bar{\bar{\boldsymbol{\mu}}}, \bar{\bar{\boldsymbol{\lambda}}}, \rho, \delta) \stackrel{\text{def}}{=} \left\{ \begin{array}{l} \left(\sum_{i=1}^{p}(\mu_i)_+\nabla g_i(\boldsymbol{x}) + \sum_{j=1}^{q}\lambda_j\nabla h_j(\boldsymbol{x}), H\right) \in \mathbb{R}^n \times \mathrm{Sym}(n) \mid \\ H \preceq \sum_{i=1}^{p}(\mu_i)_+\nabla^2 g_i(\boldsymbol{x}) + \sum_{j=1}^{q}\lambda_j\nabla^2 h_j(\boldsymbol{x}) \\ \quad + \rho\sum_{\mu_i\geq-\delta}\nabla g_i(\boldsymbol{x})\nabla g_i(\boldsymbol{x})^T + \rho\sum_{j=1}^{q}\nabla h_j(\boldsymbol{x})\nabla h_j(\boldsymbol{x})^T, \text{ where} \\ \boldsymbol{\mu} = (\bar{\bar{\boldsymbol{\mu}}} + \rho\, \boldsymbol{g}(\boldsymbol{x})) \quad \text{and} \quad \boldsymbol{\lambda} = (\bar{\bar{\boldsymbol{\lambda}}} + \rho\, \boldsymbol{h}(\boldsymbol{x})) \end{array} \right\},$$

for all $\boldsymbol{x} \in \mathbb{R}^n$, $\bar{\bar{\boldsymbol{\mu}}} \in \mathbb{R}^p_+$, $\bar{\bar{\boldsymbol{\lambda}}} \in \mathbb{R}^q$, $\rho \in \mathbb{R}_{++}$ and $\delta \in \mathbb{R}_{++}$. We say that a feasible point $\boldsymbol{x}^*$ fulfills the *AL2-regularity CQ* when, for all $\bar{\boldsymbol{\mu}} \in \mathbb{R}^p_+$ and $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^q$, the multivalued function fulfills

$$\limsup_{\left(\boldsymbol{x},\bar{\bar{\boldsymbol{\mu}}},\bar{\bar{\boldsymbol{\lambda}}},\rho,\delta\right)\to(\boldsymbol{x}^*,\bar{\boldsymbol{\mu}},\bar{\boldsymbol{\lambda}},+\infty,0^+)} K^W(\boldsymbol{x}, \bar{\bar{\boldsymbol{\mu}}}, \bar{\bar{\boldsymbol{\lambda}}}, \rho, \delta) \subset K^2(\boldsymbol{x}^*),$$

where

$$K^2(\boldsymbol{x}^*) \stackrel{\text{def}}{=} \left\{ \begin{array}{l} (\sum_{i=1}^{p}\mu_i\nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{q}\lambda_j\nabla h_j(\boldsymbol{x}^*), H) \in \mathbb{R}^n \times \mathrm{Sym}(n) | \\ \boldsymbol{\mu} \in \mathbb{R}^p_+, \quad \text{and} \quad \boldsymbol{\lambda} \in \mathbb{R}^q \\ H \preceq \sum_{i=1}^{p}\mu_i\nabla^2 g_i(\boldsymbol{x}^*) + \sum_{j=1}^{q}\lambda_j\nabla^2 h_j(\boldsymbol{x}^*) in \mathscr{C}^W(\boldsymbol{x}^*) \\ \text{and} \quad \boldsymbol{g}(\boldsymbol{x}^*)^T\boldsymbol{\mu} = 0 \end{array} \right\}.$$

Remark 5.3 Considering the previous definition, note that the point $\boldsymbol{x}^*$ satisfies WSOC if and only if $(-\nabla f(\boldsymbol{x}^*), -\nabla^2 f(\boldsymbol{x}^*)) \in K^2(\boldsymbol{x}^*)$.

Guaranteeing that AL2-regularity is sufficient to ensure WSOC under AL-AKKT2 shows that it is, in fact, a CQ for WSOC since AL-AKKT2 is valid at every minimizer. On the other hand, the reciprocal of this result guarantees that any CQ that guarantees WSOC under the validity of AL-AKKT2 must be stronger than (or equivalent to) AL2-regularity. These two results are fundamental to achieving our objective.

Theorem 5.3 *The point $\boldsymbol{x}^*$ fulfills the AL2-regularity CQ if and only if, for every twice continuously differentiable objective function f considered in (1), if $\boldsymbol{x}^*$ satisfies AL-AKKT2 with respect to this objective function, it also satisfies WSOC.*

Proof Suppose that $\boldsymbol{x}^*$ meets with AL2-regularity and that it fulfills AL-AKKT2. Let $\{(\theta_k, \rho_k, \eta_k, \boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\}_{k\in\mathbb{N}}$ be a sequence fulfilling Definition 3.1. We analyze the cases where $\{\rho_k\}_{k\in\mathbb{N}}$ has a bounded subsequence or not separately.

If $\{\rho_k\}_{k\in\mathbb{N}}$ has a bounded subsequence, by taking a subsequence, we may assume that $\{\rho_k\}_{k\in\mathbb{N}}$ is convergent since subsequences of AL-AKKT2 sequences are also AL-AKKT2 sequences. Then, by Proposition 4.1, WSOC is held, and the necessary result is proven. If no subsequence of $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, necessarily $\lim_{k\to\infty}\rho_k = \infty$. Taking $\delta_k = \eta_k$ for all $k \in \mathbb{N}$, we have that, for some sequence $\{\boldsymbol{v}^k\}_{k\in\mathbb{N}}$ converging to zero, $(-\nabla f(\boldsymbol{x}^k)+\boldsymbol{v}^k, -\nabla^2 f(\boldsymbol{x}^k)-\theta_k\mathbb{I}) \in K^W(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k, \rho_k, \delta_k)$. Thus, as $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}}$ are bounded, we can assume, without losing generality, that $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}}$ converges to $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\lambda}}$, respectively. Consequently, taking the limit, we derive that

$$(-\nabla f(\boldsymbol{x}^*), -\nabla^2 f(\boldsymbol{x}^*)) \in \limsup_{\left(\boldsymbol{x},\bar{\bar{\boldsymbol{\mu}}},\bar{\bar{\boldsymbol{\lambda}}},\rho,\delta\right)\to(\boldsymbol{x}^*,\bar{\boldsymbol{\mu}},\bar{\boldsymbol{\lambda}},+\infty,0^+)} K^W(\boldsymbol{x}, \bar{\bar{\boldsymbol{\mu}}}, \bar{\bar{\boldsymbol{\lambda}}}, \rho, \delta).$$

Hence, by the AL2-regularity at $\boldsymbol{x}^*$, $(-\nabla f(\boldsymbol{x}^*), -\nabla^2 f(\boldsymbol{x}^*)) \in K^2(\boldsymbol{x}^*)$. That is, $\boldsymbol{x}^*$ is a WSOC point.

Let us prove the reciprocal; that is, we now show that if the point $\boldsymbol{x}^*$ satisfies WSOC whenever it fulfills the AL-AKKT2 condition, then the function K^W satisfies, for a given $\bar{\boldsymbol{\mu}} \in \mathbb{R}^p_+$ and $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^q$,

$$\limsup_{\left(\boldsymbol{x},\bar{\bar{\boldsymbol{\mu}}},\bar{\bar{\boldsymbol{\lambda}}},\rho,\delta\right)\to(\boldsymbol{x}^*,\bar{\boldsymbol{\mu}},\bar{\boldsymbol{\lambda}},+\infty,0^+)} K^W(\boldsymbol{x}, \bar{\bar{\boldsymbol{\mu}}}, \bar{\bar{\boldsymbol{\lambda}}}, \rho, \delta) \subset K^2(\boldsymbol{x}^*).$$

Indeed, let

$$(\boldsymbol{d}, H) \in \limsup_{\left(\boldsymbol{x},\bar{\bar{\boldsymbol{\mu}}},\bar{\bar{\boldsymbol{\lambda}}},\rho,\delta\right)\to(\boldsymbol{x}^*,\bar{\boldsymbol{\mu}},\bar{\boldsymbol{\lambda}},+\infty,0^+)} K^W(\boldsymbol{x}, \bar{\bar{\boldsymbol{\mu}}}, \bar{\bar{\boldsymbol{\lambda}}}, \rho, \delta).$$

Then, there are sequences $\{\boldsymbol{d}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{H^k\}_{k\in\mathbb{N}} \subset \mathrm{Sym}(n)$, $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and $\{\delta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ such that $(\boldsymbol{d}^k, H^k) \in K^W(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k, \rho_k, \delta_k)$, $\{(\boldsymbol{d}^k, H^k)\}_{k\in\mathbb{N}}$ converges to $(\boldsymbol{d}, H)$, $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ converges to $\boldsymbol{x}^*$, $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}}$ converge to $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\lambda}}$, respectively, $\{\rho_k\}_{k\in\mathbb{N}}$ diverges to $+\infty$ and $\{\delta_k\}_{k\in\mathbb{N}}$ converges to zero. Therefore, considering the problem

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathbb{R}^n} & -\boldsymbol{d}^T(\boldsymbol{x}-\boldsymbol{x}^*) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^T H(\boldsymbol{x}-\boldsymbol{x}^*) \quad \text{s.t.} \quad \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0} \\ & \text{and} \quad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\end{aligned} \tag{22}$$

$\boldsymbol{x}^*$ is an AL-AKKT2 point with the primal sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$. Thus, by the hypothesis, the problem (22) fulfills WSOC at $\boldsymbol{x}^*$. And finally, $(\boldsymbol{d}, H) \in K^2(\boldsymbol{x}^*)$. This concludes the proof. □

The least stringent CQ associated with the AKKT2 conditions is called CCP2, while the one associated with CAKKT2 is called CAKKT2-regularity. Since AL-AKKT2 is stronger than AKKT2 and CAKKT2, it is natural to wonder whether the AL2-regularity is weaker than CCP2 and CAKKT2-regularity. Interestingly, it is, but not only that, AL2-regularity weakens all possible constraint qualifications that can be used to guarantee convergence to WSOC points for SALM's accumulation points. Understanding this matter is straightforward. A proof is given below.

Proposition 5.2 *The least stringent CQ necessary for WSOC points at the SALM accumulation points is AL2-regularity. In particular, CCP2 implies AL2-regularity.*

Proof Let $\boldsymbol{x}^*$ be a feasible point, and assume that a constraint qualification that guarantees WSOC at SALM accumulation points is valid at $\boldsymbol{x}^*$. Let

$$(\boldsymbol{d}, H) \in \limsup_{\left(\boldsymbol{x},\bar{\bar{\boldsymbol{\mu}}},\bar{\bar{\boldsymbol{\lambda}}},\rho,\delta\right)\to(\boldsymbol{x}^*,\bar{\boldsymbol{\mu}},\bar{\boldsymbol{\lambda}},+\infty,0^+)} K^W(\boldsymbol{x}, \bar{\bar{\boldsymbol{\mu}}}, \bar{\bar{\boldsymbol{\lambda}}}, \rho, \delta).$$

Then, there are sequences $\{\boldsymbol{d}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{H^k\}_{k\in\mathbb{N}} \subset \mathrm{Sym}(n)$, $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and $\{\delta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ such that $(\boldsymbol{d}^k, H^k) \in K^W(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k, \rho_k, \delta_k)$, with $\rho_k \to \infty$, $\{\delta_k\}_{k\in\mathbb{N}}$ converging to zero, $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ converging to $\boldsymbol{x}^*$, $\{(\boldsymbol{d}^k, H^k)\}_{k\in\mathbb{N}}$ converging to $(\boldsymbol{d}, H)$ and $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}}$ converging to $\bar{\boldsymbol{\mu}} \in \mathbb{R}^p_+$ and $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^q$, respectively. Therefore, considering the problem (22), $\boldsymbol{x}^*$ is an AL-AKKT2 point with the primal sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$. Since AL-AKKT2 guarantees a sequence convergent to $\boldsymbol{x}^*$ generated by SALM applied to (22), and a constraint qualification that guarantees WSOC at SALM accumulation points is valid at the same point, we have the validity of WSOC. Consequently, $(\boldsymbol{d}, H) \in K^2(\boldsymbol{x}^*)$ and the inclusion in Definition 5.1 follows.

Since AL-AKKT2 implies AKKT2, the second statement can be proved using the same reasoning applied to the first part of the proof. □

Remark 5.4 It is possible to prove straightforwardly that CAKKT2-regularity implies AL2-regularity under the GKL inequality, as discussed in Remark 5.2.

Considering the developments in this article, we provide the landscape of CQs associated with WSOC in Figure 1.

6 Final Remarks

Safeguarded augmented Lagrangian methods have been widely studied recently due to their broad applicability and general global convergence results. When one can solve the subproblems of an algorithm by obtaining second-order stationarity, a second-order stationary accumulation point for the original problem is expected to be found. The first-order variation of this idea and reasonable assumptions that guarantee a similar first-order property are well known. However, very few results are known for the second-order case. Previous results suggested that a constant rank property of the

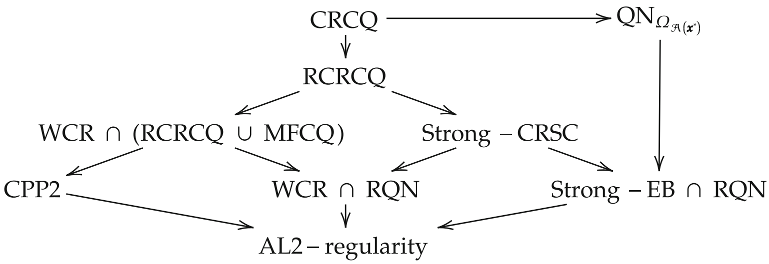

Fig. 1 Landscape of the new constraint qualifications for which accumulation points of the safeguarded augmented Lagrangian method are WSOC points. Here, $QN_{\Omega_{\mathcal{A}(x^*)}}$ is the quasinormality CQ applied to the set $\Omega_{\mathcal{A}(x^*)}$

gradients of active constraints was needed to guarantee the inner continuity of the linearized subspace, inherent to second-order stationarity.

In this paper, however, we propose a strong error bound property, which is enough to guarantee second-order stationarity. The condition corresponds to the traditional error bound constraint qualification for constraint sets formulated with selections of the constraint functions. Our result also requires boundedness of the dual sequences, which can be achieved with the so-called relaxed quasinormality condition. Connections with several other conditions known in the literature are studied; in particular, we show that the recently introduced *strong constant rank of the subspace component* is sufficient for guaranteeing second-order stationarity without any additional requirement.

To formulate our results, we present a new second-order sequential optimality condition, particularly tailored to the algorithm, which allows us to characterize the least restrictive condition that guarantees second-order stationarity, called AL2-regularity. This condition is the least stringent CQ associated with the algorithm in the sense that any CQ that guarantees second-order stationarity at feasible accumulation points reached by the algorithm is necessarily more stringent than AL2-regularity. This suggests a relevant research topic related to our results, which is the characterization of conditions that ensure the boundedness of the dual sequences generated by the algorithm. We conjecture that this property can be characterized in terms of error bound type conditions. This will be the subject of future research.

Acknowledgements This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (grants 2018/24293-0, 2017/18308-2, 2023/08706-1, 2024/12967-8 and 2023/08621-6), CNPq (grants 306988/2021-6, 302000/2022-4, 407147/2023-3 and 309136/2021-0), PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016), UNICAMP /PRP/FAEPEX (grants 3319/23 and 2997/24) and ANPCyT (grants PICT 2016-0921 and PICT 2019-02172)

Funding Fundaçäo de Amparo à Pesquisa do Estado de Sao Paulo

Data Availability Data sharing does not apply to this article as no datasets were generated or analyzed during the current theoretical study.

References

1. Anandkumar, A., Ge, R.: Efficient approaches for escaping higher order saddle points in non-convex optimization. In: V. Feldman, A. Rakhlin, O. Shamir (eds.) 29th Annual Conference on Learning Theory, Proceedings of Machine Learning Research, vol. 49, pp. 81–102. PMLR, Columbia University, New York, New York, USA (2016). https://proceedings.mlr.press/v49/anandkumar16.html
2. Andreani, R., Behling, R., Haeser, G., Silva, P.J.: On second-order optimality conditions in nonlinear optimization. Optimization Methods and Software **32**(1), 22–38 (2017). https://doi.org/10.1080/10556788.2016.1188926
3. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On Augmented Lagrangian Methods with General Lower-Level Constraints. SIAM J. Optim. **18**(4), 1286–1309 (2008). https://doi.org/10.1137/060654797
4. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Second-order negative-curvature methods for box-constrained and general constrained optimization. Comput. Optim. Appl. **45**(2), 209–236 (2010). https://doi.org/10.1007/s10589-009-9240-y
5. Andreani, R., Echagüe, C.E., Schuverdt, M.L.: Constant-rank condition and second-order constraint qualification. J. Optim. Theory Appl. **146**(2), 255–266 (2010). https://doi.org/10.1007/s10957-010-9671-8
6. 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 J. Optim. **29**(1), 743–766 (2019). https://doi.org/10.1137/17M1147330
7. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011). https://doi.org/10.1080/02331930903578700
8. Andreani, R., Haeser, G., Mito, L., Cabrera, H.: A minimal face constant rank constraint qualification for reducible conic programming (2023). https://optimization-online.org/?p=22785. Unpublished
9. 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 **90**(2), 879–880 (2022). https://doi.org/10.1007/s11075-021-01241-3
10. 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 **90**(2), 851–877 (2022). https://doi.org/10.1007/s11075-021-01212-8
11. Andreani, R., Haeser, G., Ramos, A., Silva, P.J.S.: A second-order sequential optimality condition associated to the convergence of optimization algorithms. IMA J. Numer. Anal. **37**(4), 1902–1929 (2017). https://doi.org/10.1093/imanum/drw064
12. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two New Weak Constraint Qualifications and Applications. SIAM J. Optim. **22**(3), 1109–1135 (2012). https://doi.org/10.1137/110843939
13. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: Strict Constraint Qualifications and Sequential Optimality Conditions for Constrained Optimization. Math. Oper. Res. **43**(3), 693–717 (2018). https://doi.org/10.1287/moor.2017.0879
14. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On second-order optimality conditions for nonlinear programming. Optimization **56**(5–6), 529–542 (2007). https://doi.org/10.1080/02331930701618617
15. Andreani, R., Martínez, J.M., Svaiter, B.F.: A New Sequential Optimality Condition for Constrained Optimization and Algorithmic Consequences. SIAM J. Optim. **20**(6), 3533–3554 (2010). https://doi.org/10.1137/090777189
16. Andreani R., Haeser G., Schuverdt M.L. and Secchin L. D.: A relaxed quasinormality condition and the boundedness of dual augmented Lagrangian sequences (2024). https://optimization-online.org/?p=25207. *Submitted*
17. Arutyunov, A., Bobylev, N., Korovin, S.: One remark to Ekeland's variational principle. Computers & Mathematics with Applications **34**(2), 267–271 (1997). https://doi.org/10.1016/S0898-1221(97)00128-4
18. Arutyunov, A.V.: Perturbations of extremal problems with constraints and necessary optimality conditions. J. Sov. Math. **54**(6), 1342–1400 (1991). https://doi.org/10.1007/BF01373649
19. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena scientific optimization and computation series. Athena Scientific, Belmont, Massachussets (1999)
20. Birgin, E.G., Gardenghi, J.L., Martínez, J.M., Santos, S.A., Toint, P.L.: Evaluation complexity for nonlinear constrained optimization using unscaled KKT conditions and high-order models. SIAM J. Optim. **26**(2), 951–967 (2016). https://doi.org/10.1137/15M1031631

21. Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA (2014). https://doi.org/10.1137/1.9781611973365
22. Birgin, E.G., Martínez, J.M.: Complexity and performance of an Augmented Lagrangian algorithm. Optimization Methods and Software **35**(5), 885–920 (2020). https://doi.org/10.1080/10556788.2020.1746962
23. Gomes-Ruggiero, M.A., Martínez, J.M., Santos, S.A.: Spectral Projected Gradient Method with Inexact Restoration for Minimization with Nonconvex Constraints. SIAM J. Sci. Comput. **31**(3), 1628–1652 (2009). https://doi.org/10.1137/070707828
24. Haeser, G.: A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. Comput. Optim. Appl. **70**(2), 615–639 (2018). https://doi.org/10.1007/s10589-018-0005-3
25. Haeser, G.: Some theoretical limitations of second-order algorithms for smooth constrained optimization. Oper. Res. Lett. **46**(3), 295–299 (2018). https://doi.org/10.1016/j.orl.2018.02.007 https://www.sciencedirect.com/science/article/pii/S0167637717306041
26. Jin, X.B., Zhang, X.Y., Huang, K., Geng, G.G.: Stochastic conjugate gradient algorithm with variance reduction. IEEE Transactions on Neural Networks and Learning Systems **30**(5), 1360–1369 (2019). https://doi.org/10.1109/TNNLS.2018.2868835
27. Jourani, A., Sene, M.: Characterization of the clarke regularity of subanalytic sets. Proceedings of the American Mathematical Society **146**, 1 (2017). https://doi.org/10.1090/proc/13847
28. Lewis, A.S., Luke, D.R., Malick, J.: Local Linear Convergence for Alternating and Averaged Nonconvex Projections. Found. Comput. Math. **9**(4), 485–513 (2009). https://doi.org/10.1007/s10208-008-9036-y
29. Martens, J.: Second-order optimization for neural networks. Ph.d. thesis, Department of Computer Science, University of Toronto, Toronto, Canada (2016). https://hdl.handle.net/1807/71732. Accessed: 16 Jul. 2024
30. Minchenko, L., Tarakanov, A.: On Error Bounds for Quasinormal Programs. J. Optim. Theory Appl. **148**(3), 571–579 (2011). https://doi.org/10.1007/s10957-010-9768-0
31. Minchenko, L.I., Sirotko, S.I.: On Local Error Bound in Nonlinear Programs. In: Olenev, N.N., Evtushenko, Y.G., Jaćimović, M., Khachay, M., Malkova, V. (eds.) Optimization and Applications, pp. 38–49. Springer International Publishing, Cham (2021)
32. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation I, Grundlehren der mathematischen Wissenschaften, vol. 330, 1st edn. Springer (2006). https://doi.org/10.1007/3-540-31247-1
33. Murty, K.G., Kabadi, S.N.: Some NP-complete problems in quadratic and nonlinear programming. Math. Program. **39**(2), 117–129 (1987). https://doi.org/10.1007/BF02592948
34. Nesterov, Y., Polyak, B.T.: Cubic regularization of Newton method and its global performance. Math. Program. **108**(1), 177–205 (2006). https://doi.org/10.1007/s10107-006-0706-8
35. Prado, R.W., Santos, S.A., Simões, L.E.A.: A novel sequential optimality condition for smooth constrained optimization and algorithmic consequences. Optimization , 1–30 (2023). https://doi.org/10.1080/02331934.2023.2168481
36. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, Grundlehren der mathematischen Wissenschaften, vol. 317, 1 edn. Springer Science & Business Media (2009). https://doi.org/10.1007/978-3-642-02431-3
37. Ye, J.J.: New uniform parametric error bounds. J. Optim. Theory Appl. **98**(1), 197–219 (1998). https://doi.org/10.1023/A:1022649217032
38. Zalinescu, C.: Convex Analysis in General Vector Spaces. World Scientific Publishing Company, Singapore (2002)