# Primal-dual global convergence of an augmented Lagrangian method under the error bound condition*

R. Andreani† G. Haeser‡ R. W. Prado§ L. D. Secchin¶

Revised July 10, 2026

**Abstract**

This work investigates global convergence properties of a safeguarded augmented Lagrangian method applied to nonlinear programming problems, with an emphasis on the role of constraint qualifications in ensuring boundedness of the computed multiplier sequences, also known as dual sequences. When functions with locally Lipschitz continuous derivatives define the constraint set, we prove that the Error Bound Constraint Qualification is the weakest constraint qualification that guarantees boundedness of the computed multiplier sequences generated by the method. The condition is also known as the local error bound property, metric subregularity constraint qualification, or calmness constraint qualification. We further show its equivalence to a Polyak-Łojasiewicz inequality for the quadratic infeasibility measure, which in turn is equivalent to the recently introduced relaxed quasinormality constraint qualification. Moreover, we prove the feasibility of accumulation points of primal sequences generated by the augmented Lagrangian method under a Polyak-Łojasiewicz inequality for the quadratic infeasibility measure. Our results provide a sound primal-dual global convergence result under a weak and well-known condition, reinforcing the effectiveness of safeguarded augmented Lagrangian methods over non-safeguarded approaches.

**Keywords:** Nonlinear programming, Augmented Lagrangian methods, Error bound, Constraint qualifications

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

## 1 Introduction

Augmented Lagrangian methods (ALMs) are one of the cornerstones of modern nonlinear programming methods, alongside interior point and sequential quadratic programming methods. At each augmented Lagrangian iteration, a penalized subproblem is solved where one uses estimates of the Lagrange multipliers as displacements of constraint violations. Using bounded Lagrange multiplier estimates is a fundamental idea of so-called safeguarded augmented Lagrangian methods, which allows obtaining primal global convergence for degenerate problems without hindering their overall performance. A natural question in the primal-dual convergence analysis of safeguarded augmented Lagrangian methods is whether the computed multipliers generated by this algorithm remain bounded or not. The primary objective of this paper is to propose a necessary and sufficient condition for the boundedness of the computed multiplier sequences. The condition is well known and essential for infeasible algorithms: one should be able to estimate distances to the feasible set by evaluating the constraints, avoiding the computational costs of measuring distances from the constraint set. We demonstrate that this attribute, commonly known as the *error bound property* (see Definition 2.2), is equivalent to the boundedness of the computed multipliers, and this connection sheds light on the practical efficacy of ALMs.

*This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (grants 2018/24293-0, 2017/18308-2, 2023/08706-1, 2023/08621-6 and 2024/12967-8), CNPq (grants 306988/2021-6, 302000/2022-4, 407147/2023-3 and 302520/2025-2), PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016), UNICAMP/PRP/FAEPEX (grants 3319/23 and 2997/24).

†Department of Applied Mathematics, Universidade Estadual de Campinas, Campinas, SP, Brazil. Email: andreani@unicamp.br

‡Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil. Email: ghaeser@ime.usp.br

§Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil. Email: pradorenanw@gmail.com

¶Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, ES, Brazil. Email: leonardo.secchin@ufes.br

Historically, achieving bounded computed multipliers demanded strong constraint qualifications (CQs). Sequential quadratic programming methods classically invoke *Mangasarian–Fromovitz Constraint Qualification* (MFCQ) or *Linear Independence Constraint Qualification* (LICQ); even regularized variants rely on a second-order condition to bound the computed multipliers, as discussed in [35, §2]. However, second-order conditions may fail at local minimizers when MFCQ breaks down [14, Sec. 2.4]. Interior point schemes ensure primal convergence under mild conditions [63, 8]. However, only recent enhancements achieve bounded duals without MFCQ, and only under still fairly stringent assumptions [39, Thm. 5]. Likewise, inexact restoration methods converge to KKT points under weak CQs [53, 37], yet their primal-dual versions still hinge on MFCQ for boundedness of the computed multipliers; we refer the reader to [12, §1] for background and context.

The augmented Lagrangian method introduced by Hestenes [40] demonstrated significant practical potential. Although Hestenes did not provide a formal convergence theory, it was evident that the algorithm converged to KKT points if the computed multipliers remained bounded (see Proposition 2.1). Rockafellar [65, 66] and Bertsekas [18] established the first rigorous convergence proofs for the convex case, incorporating Powell's treatment of inequality constraints [62]. For the non-convex case, Conn, Gould, and Toint [27, 29] later proved convergence, but under restrictive assumptions such as the linear independence of active or violated constraint gradients at infeasible points.

A breakthrough came with safeguarded augmented Lagrangian methods, epitomized by the ALGENCAN solver [21]: by projecting the computed multipliers onto a fixed compact set, they achieve KKT convergence under the much weaker *Constant Positive Linear Dependence* (CPLD) condition. The new method motivated the development of a unified theory and a hierarchy of ever-weaker CQs – *Constant Positive Generators* (CPG) [8, Theorem 3.3], *Approximate Karush-Kuhn-Tucker regularity* (AKKT-regularity) [9], and beyond; see [10] for a detailed and recent review. However, these frameworks aimed only at primal convergence to KKT points, leaving dual convergence unexplored.

Subsequent analyses revealed that *quasinormality* (QN) [19, pp. 337] is sufficient to ensure that the computed multipliers of the safeguarded ALM remain bounded. Recently, the result was generalized by utilizing a relaxed variant, *relaxed quasinormality* (RQN) [6, Thm. 2], which encompasses all constant rank-type constraint qualifications, such as *Constant Rank of the Subspace Component* (CRSC) [8, Definition 1.3]. Interestingly, except for RQN, all CQs known to guarantee boundedness of the computed multiplier sequence, including QN and CRSC, also imply the classical *error bound* (EB) constraint qualification; see Definition 2.2 ahead. In contrast, weaker CQs or those that do not imply EB fail to guarantee boundedness of the computed multipliers [8]. This observation has led us to hypothesize that the error bound condition may be necessary and sufficient for ensuring boundedness of the computed multipliers associated with the safeguarded ALM. The error bound condition is well-recognized across various contexts under several equivalent terms, including metric subregularity constraint qualification, calmness, or simply as the error bound property [16].

To investigate the conjecture, our primary objective is to identify the weakest possible CQ that ensures boundedness of computed multipliers generated by the method. The first of our contributions establishes that EB is sufficient for this (Theorem 3.1). Specifically, we prove that every primal convergent subsequence generated by the safeguarded ALM has the corresponding subsequence of computed multipliers bounded under the EB property. As a direct result, safeguarded ALM has KKT accumulation points under this condition.

Complementarily, we show that failure of the EB condition implies the existence of a nonlinear programming problem where the safeguarded ALM produces unbounded computed multipliers associated with a convergent primal sequence (Theorem 3.3). Taken together, these two results yield a tight characterization: the EB condition is the weakest possible CQ, ensuring that the safeguarded ALM generates bounded computed multipliers whenever the associated primal sequence converges. It is worth pointing out that the EB property ensures boundedness of the computed multipliers generated by the algorithm, not the entire set of Lagrange multipliers associated with the problem. This feature, which also applies to all CQs weaker than MFCQ, implies that the problem itself may admit an unbounded set of KKT multipliers, even though the algorithm yields bounded computed multiplier iterates.

To establish this fundamental characterization, we develop and connect auxiliary concepts. First, we introduce the *Polyak–Łojasiewicz Constraint Qualification* (PŁCQ, as in Definition 2.3). This condition arises from applying the PŁ inequality to the quadratic infeasibility measure associated with the problem. We prove that PŁCQ is equivalent to the EB condition (Theorem 3.2; see also Section 4.1), thus confirming that PŁCQ constitutes a genuine CQ. This equivalence provides a new and purely functional characterization of the EB property, expressed solely in terms of the function that defines the constraint set. The relevance of this reformulation is underscored by the central role of PŁ inequality across a broad spectrum of optimization settings, making PŁCQ a condition of independent interest [23, 26, 45, 48, 50, 64]. Second, we revisit the EB condition in light of another CQ, the RQN condition (Definition 2.1). RQN weakens the traditional QN condition, with additional assumptions reflecting sensitivity properties of the violated constraints around the target point, which is enough to ensure boundedness of computed multipliers generated by the safeguarded ALM [6]. Using PŁCQ, we prove that EB and RQN are equivalent (Theorem 4.1). Thus, considering that there are several studies in the literature concerning different quasinormality conditions as sufficient conditions for the error bound property in many contexts [15, 68, 17], our result shows that RQN is the ultimate quasinormality-type necessary and sufficient condition for EB, at least in the context of nonlinear programming.

Finally, we extend all previous results from this paper to include not only feasible but also infeasible points by introducing the Extended-PŁCQ condition, which generalizes recent advancements based on the Extended-RQN framework [6]. Collectively, the findings presented here complete the convergence theory of the safeguarded augmented Lagrangian method under the weakest known assumptions regarding both primal feasibility and dual boundedness.

**Notation** Vectors and vector-valued functions are denoted by boldface letters, where $v_i$ denotes the $i$-th component (or component function) of a vector (or vector-valued function) $\boldsymbol{v}$. The set of non-negative (positive) real numbers is denoted by $\mathbb{R}_+$ (respectively $\mathbb{R}_{++}$). For $\ell \in \mathbb{N}$, $I_\ell := \{k \in \mathbb{N} \,|\, 1 \leq k \leq \ell\}$ and for $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v}_+ := (\max\{0, v_1\}, \dots, \max\{0, v_n\})^T$. We use $\|\cdot\|$ to denote the Euclidean norm and $\mathcal{B}[\boldsymbol{x}, \epsilon]$ to denote the closed Euclidean ball centered at $\boldsymbol{x} \in \mathbb{R}^n$ with radius $\epsilon > 0$. For a set $S \subset \mathbb{R}^n$, co $S$ and $\bar{\text{co}}\, S$ denote its convex hull and closed convex hull, respectively. The Euclidean distance from $\boldsymbol{x} \in \mathbb{R}^n$ to $S \subset \mathbb{R}^n$ is denoted by $\operatorname{dist}(\boldsymbol{x}, S)$ and $\mathcal{P}_S(\boldsymbol{x}) := \{\boldsymbol{z} \in S : \|\boldsymbol{z} - \boldsymbol{x}\| \leq \operatorname{dist}(\boldsymbol{x}, S)\}$ is the set of points in $S$ closest to $\boldsymbol{x}$. For a given set $\mathcal{J} \subset \mathbb{R}^n$, we use $\mathcal{P}_S(\mathcal{J}) := \cup_{\boldsymbol{x} \in \mathcal{J}} \mathcal{P}_S(\boldsymbol{x})$.

## 2 Preliminaries

The nonlinear programming problems we are interested in are formulated as

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

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 functions with locally Lipschitz continuous first derivatives. For simplicity, we define the feasible set as $\Omega := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0},\ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$, which we assume to be nonempty.

To solve (1), we consider algorithms based on *penalty methods*, where the goal is to replace a complex constrained optimization problem with a sequence of simpler unconstrained problems. The *quadratic penalty method* addresses the original constrained problem by successively solving the unconstrained minimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) + \rho_k \Phi(\boldsymbol{x}),$$

where $\{\rho_k\}_{k \in \mathbb{N}}$ is a sequence of positive penalty parameters such that $\rho_k \to +\infty$ as $k \to \infty$ and $\Phi$ represents the quadratic measure of constraint violation (infeasibility),

$$\Phi(\boldsymbol{x}) := \frac{1}{2} \left( \|\boldsymbol{g}(\boldsymbol{x})_+\|^2 + \|\boldsymbol{h}(\boldsymbol{x})\|^2 \right) \quad \forall \boldsymbol{x} \in \mathbb{R}^n. \tag{2}$$

Safeguarded augmented Lagrangian methods, in turn, refine this approach by incorporating estimates of the *Lagrange multipliers* into the measure of constraint violation. This technique often mitigates the need for the penalty parameter to become excessively large, which can lead to numerical ill-conditioning of the subproblems. A central function utilized within this framework is the *Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian function*

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

for all $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{\mu} \in \mathbb{R}^p_+$ and $\boldsymbol{\lambda} \in \mathbb{R}^q$. The standard Lagrangian is given by $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f(\boldsymbol{x}) + \boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x})$.

Essentially, the strategy of the *safeguarded augmented Lagrangian method* is to replace the constrained optimization problem (1) with the task of solving (approximately) a sequence of unconstrained subproblems

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

Here, $\{\rho_k\}_{k\in\mathbb{N}}$ represents a non-decreasing sequence of positive penalty parameters which, unlike in pure penalty methods, does not necessarily need to satisfy $\rho_k \to \infty$. Concurrently, $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$ and $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$ are sequences of estimates for the Lagrange multipliers corresponding to equality and inequality constraints, respectively, of problem (1). In the classical safeguarded augmented Lagrangian framework considered here, the multiplier estimates that enter the subproblems are restricted to a fixed, bounded set.

To avoid the global minimization (3), for a given $k \in \mathbb{N}$, we seek a point $\boldsymbol{x}^k \in \mathbb{R}^n$ satisfying the approximate first-order stationarity condition

$$\|\nabla_{\boldsymbol{x}} \mathcal{L}_{\rho_k}(\boldsymbol{x}^k, \bar{\boldsymbol{\lambda}}^k, \bar{\boldsymbol{\mu}}^k)\| \le \theta_k, \tag{4}$$

where the gradient is taken with respect to $\boldsymbol{x}$ and $\{\theta_k\}_{k\in\mathbb{N}}$ is a sequence of positive tolerance parameters converging to zero. A point $\boldsymbol{x}^k$ satisfying (4) can be computed using standard iterative methods for unconstrained optimization applied to (3). This inexact minimization approach renders the overall method computationally feasible under mild assumptions. The framework of the method is presented next.

---

**Algorithm 1** Algorithmic scheme for the safeguarded augmented Lagrangian method (safeguarded ALM).

---

**Step 1** (*Choose input parameters*) $\lambda_{\min} \le \lambda_{\max}$, $0 \le \mu_{\max}$, $\gamma > 1$, $0 < r < 1$, and a sequence $\{\theta_k\}_{k\in\mathbb{N}}$ with $\theta_k \downarrow 0$. Initialize $\bar{\boldsymbol{\mu}}^1 \in [0, \mu^{\max}]^p$, $\bar{\boldsymbol{\lambda}}^1 \in [\lambda^{\min}, \lambda^{\max}]^q$, and $\rho_1 > 0$. Set $k \leftarrow 1$.

**Step 2** (*Solve subproblem*) Apply an unconstrained optimization algorithm to solve (3), obtaining a candidate solution $\boldsymbol{x}^k \in \mathbb{R}^n$ satisfying (4).

**Step 3** (*Compute multipliers*) Calculate:

$$\boldsymbol{\mu}^k := \left(\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k)\right)_+ \quad \text{and} \quad \boldsymbol{\lambda}^k := \bar{\boldsymbol{\lambda}}^k + \rho_k \boldsymbol{h}(\boldsymbol{x}^k). \tag{5}$$

**Step 4** (*Update penalty parameters*) Define

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

for each $i \in I_p$. If $k = 1$ or

$$\max\left\{\|\boldsymbol{V}^k\|, \|\boldsymbol{h}(\boldsymbol{x}^k)\|\right\} \le r \max\left\{\|\boldsymbol{V}^{k-1}\|, \|\boldsymbol{h}(\boldsymbol{x}^{k-1})\|\right\}, \tag{6}$$

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

**Step 5** (*Update safeguarded multipliers*) Choose $\bar{\boldsymbol{\mu}}^{k+1} \in [0, \mu^{\max}]^p$ and $\bar{\boldsymbol{\lambda}}^{k+1} \in [\lambda^{\min}, \lambda^{\max}]^q$. Set $k \leftarrow k + 1$ and return to **Step 2**.

---

Indeed, one of the most well-known properties proven at accumulation points of primal sequences generated by optimization methods for constrained optimization is the *KKT conditions*. It is well known that to ensure the validity of these properties at local minimizers, *constraint qualifications* are imposed. Clearly, optimization algorithms whose convergence is guaranteed under weaker CQs exhibit greater generality and can be applied to a broader class of problems.

The following result shows that the boundedness of the generated sequence of computed multipliers (the dual sequence) from the safeguarded ALM is sufficient for obtaining feasible KKT points.

**Proposition 2.1.** *Let $\mathbb{K} \subset \mathbb{N}$ be an infinite subset of natural numbers. Assume boundedness of the subsequences $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{K}}$ and $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{K}}$, as described in* (5)*. Then, every accumulation point of $\{\boldsymbol{x}^k\}_{k\in\mathbb{K}}$ is a KKT point.*

*Proof.* Let $\boldsymbol{x}^*$ be an accumulation point of $\{\boldsymbol{x}^k\}_{k\in\mathbb{K}}$, let us say, $\lim_{k\in\mathbb{L}} \boldsymbol{x}^k = \boldsymbol{x}^*$, $\mathbb{L} \subset \mathbb{K}$. Without loss of generality, we assume also $\lim_{k\in\mathbb{L}} \boldsymbol{\lambda}^k = \boldsymbol{\lambda} \in \mathbb{R}^q$ and $\lim_{k\in\mathbb{L}} \boldsymbol{\mu}^k = \boldsymbol{\mu} \in \mathbb{R}^p_+$. From the boundedness of $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{L}}$ and $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{L}}$, using (5), there exists $M > 0$ such that

$$\rho_k g_i(\boldsymbol{x}^k)_+ \leq M \quad \text{and} \quad \rho_k |h_j(\boldsymbol{x}^k)| \leq M \quad \forall i \in I_p, j \in I_q, \text{ and } k \in \mathbb{L}. \tag{7}$$

If $\lim_{k\in\mathbb{L}} \rho_k = \infty$, it immediately follows that $\boldsymbol{x}^*$ is feasible. Otherwise, since $\{\rho_k\}_{k\in\mathbb{N}}$ is nondecreasing and has a bounded subsequence, the entire sequence $\{\rho_k\}_{k\in\mathbb{N}}$ must be bounded. Thus, (6) cannot fail infinitely often, and so we have $\lim_{k\to\infty} V_i^k = 0$ and $\lim_{k\to\infty} \|\boldsymbol{h}(\boldsymbol{x}^k)\| = 0$. In particular, $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$. If $g_i(\boldsymbol{x}^*) > 0$ for some $i \in I_p$ then $V_i^k \leq -g_i(\boldsymbol{x}^k) \leq -g_i(\boldsymbol{x}^*)/2$ for sufficiently large $k \in \mathbb{L}$, which is impossible. Hence, $g_i(\boldsymbol{x}^*) \leq 0$, for all $i \in I_p$, proving $\boldsymbol{x}^*$ is feasible for (1). Let us show that $\boldsymbol{x}^*$ fulfills the KKT conditions. From **Step 2**, we have

$$\|\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^k, \boldsymbol{\mu}^k, \boldsymbol{\lambda}^k)\| = \|\nabla_{\boldsymbol{x}} \mathcal{L}_{\rho_k}(\boldsymbol{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\| \leq \theta_k.$$

Taking the limit over $\mathbb{L}$ yields $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \mathbf{0}$. Also, $\mu_i g_i(\boldsymbol{x}^*) = 0$ for all $i$. In fact, if $\lim_{k\in\mathbb{L}} \rho_k = \infty$ then when $g_i(\boldsymbol{x}^*) < 0$ we get from (5) that $\lim_{k\to\infty} \bar{\mu}_i^k + \rho_k g_i(\boldsymbol{x}^k) = -\infty$. Hence, $\mu_i^k = 0$ for sufficiently large $k \in \mathbb{L}$, with $\mu_i^k$ given in (5). Now, suppose that $\{\rho_k\}_{k\in\mathbb{L}}$ is bounded and $g_i(\boldsymbol{x}^*) < 0$. As before, $\lim_{k\to\infty} V_i^k = 0$. In this case, as $g_i(\boldsymbol{x}^*) < 0$ we have $V_i^k = \bar{\mu}_i^k/\rho_k$ for all large $k \in \mathbb{L}$, implying $\lim_{k\in\mathbb{L}} \bar{\mu}_i^k = 0$. Thus $\bar{\mu}_i^k + \rho_k g_i(\boldsymbol{x}^k) < 0$ for all $k \in \mathbb{L}$ large enough, implying $\mu_i^k = 0$ by (5). This concludes the proof. □

Classical conditions like LICQ and MFCQ are often used to control the growth of the computed multipliers $\{(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)\}_{k\in\mathbb{N}}$. However, these conditions can be too restrictive in real-world settings – for example, they may fail when the problem includes redundant constraints or when two inequalities replace an equality constraint with opposite signs. Recent developments have demonstrated that it is possible to guarantee convergence with bounded computed multipliers using the much weaker notion of relaxed quasinormality, which we describe next.

**Definition 2.1** ([6])**.** *It is said that relaxed quasinormality (RQN) CQ holds at the feasible point $\boldsymbol{x}^*$ of* (1) *whenever there is no sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ converging to $\boldsymbol{x}^*$ and vectors $\boldsymbol{\mu} \in \mathbb{R}^p_+$ and $\boldsymbol{\lambda} \in \mathbb{R}^q$ 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 $k \in \mathbb{N}$,*

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

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

The RQN condition is weaker than several CQs, notably all CQs from the constant-rank family. However, several other CQs are known to guarantee only primal stationarity without resorting to the boundedness of the dual sequence. These studies revealed a deeper property of safeguarded ALMs: one can guarantee convergence to KKT points even if the generated sequence of computed multipliers becomes unbounded. This finding suggests classifying CQs into two categories: those strong enough to ensure boundedness of the sequence of computed multipliers and those that still guarantee KKT convergence at accumulation points without ensuring boundedness of the computed multipliers. A detailed assessment reveals that CQs strictly stronger than RQN are known to imply the EB constraint qualification. However, RQN itself stands as a notable exception for which this implication was not previously known, a property we review in the following sections.

The EB condition was first established in the 1952 seminal work of A.J. Hoffman for measuring the constraint violation of systems of linear inequalities [41]. Since then, the error bound property has occupied a central position in modern optimization theory and practice, and within the framework of variational analysis and nonlinear optimization [58, 67]: serving as a link between the geometric regularity of feasible sets and the stability of solutions under perturbations [24, 67]; forming the basis for analyzing calmness and Aubin (Lipschitz) properties of the feasible/solution set mapping [58]; improving the understanding of the performance of numerical algorithms [32, 33, 59]; and analyzing the performance of methods in the presence of numerical errors [60]. Nowadays, the error bound property represents a fundamental measure of regularity that supports much of our understanding of the stability and solvability of constrained optimization problems. The condition is defined next.

**Definition 2.2.** *We say that a point* $\boldsymbol{x}^* \in \Omega$ *satisfies the* error bound *(EB) constraint qualification if there exist constants* $\kappa > 0$ *and* $\delta > 0$ *such that*

$$\operatorname{dist}(\boldsymbol{x}, \Omega) \leq \kappa \Phi(\boldsymbol{x})^{1/2} \quad \textit{for all} \quad \boldsymbol{x} \in \mathcal{B}[\boldsymbol{x}^*, \delta].$$

*The constant* $\kappa$ *is referred to as a local error bound constant for the set* $\Omega$ *at* $\boldsymbol{x}^*$*, relative to the function* $\Phi$ *defined in* (2).

**Remark 2.1.** *Definition 2.2 may be viewed as an instance of the metric subregularity constraint qualification (MSCQ) [36]. To see the connection, define* $\boldsymbol{F}(\boldsymbol{x}) := \big(\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{h}(\boldsymbol{x})\big), \mathcal{D} := \mathbb{R}^p_- \times \{\mathbf{0}\}^q$*, and let* $\mathcal{M}(\boldsymbol{x}) := \boldsymbol{F}(\boldsymbol{x}) - \mathcal{D}$*. Then* $\Omega = \mathcal{M}^{-1}(\mathbf{0})$ *and* $\operatorname{dist}\big(\mathbf{0}, \mathcal{M}(\boldsymbol{x})\big) = \sqrt{2\Phi(\boldsymbol{x})}$*. Hence, up to a rescaling of the local error bound constant, the error bound condition in Definition 2.2 is precisely the metric subregularity of the constraint mapping* $\mathcal{M}$ *at* $(\boldsymbol{x}^*, \mathbf{0})$*; see [15, 16].*

Here, we focus on distances measured with the Euclidean norm; however, any norm can be used to define equivalent conditions; see [60]. Interestingly, for general functions $f$, an inequality similar to the error bound property is known as *quadratic growth condition*, for some $\delta \in \mathbb{R}_{++}$:

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq \eta \operatorname{dist}(\boldsymbol{x}, \mathcal{S})^2 \quad \text{for all } \boldsymbol{x} \in \mathcal{B}[\boldsymbol{x}^*, \delta],$$

where $\mathcal{S} := \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) = f(\boldsymbol{x}^*)\}$ is the set of points attaining the minimum value $f(\boldsymbol{x}^*)$. Notably, this condition corresponds to the EB property when applied to the quadratic infeasibility measure $\Phi$ and at feasible points. Recently, a connection emerged for twice continuously differentiable functions: the quadratic growth condition at a local minimum is equivalent to a condition known as PŁ inequality [64], which states that for some $\delta \in \mathbb{R}_{++}$ and $\nu \in \mathbb{R}_{++}$ we have

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \leq \nu \|\nabla f(\boldsymbol{x})\|^2 \quad \text{for all } \boldsymbol{x} \in \mathcal{B}[\boldsymbol{x}^*, \delta].$$

This duality ensures that, under the quadratic growth condition, the function's value does not change "slowly" relative to how fast the gradient changes. One of its primary uses in optimization is to prove that the trajectory generated by gradient descent converges to a single critical point [64]. Our work focuses on the application of the PŁ inequality to the quadratic infeasibility measure to study boundedness of the computed multipliers generated by the safeguarded ALM. Therefore, we define the condition as follows.

**Definition 2.3.** *We say that* $\Phi$ *given in* (2) *meets the* Polyak-Łojasiewicz *(PŁ)* inequality *at* $\boldsymbol{x}^* \in \Omega$ *when there are constants* $\nu > 0$ *and* $\delta > 0$ *such that*

$$\Phi(\boldsymbol{x})^{1/2} \leq \nu \|\nabla \Phi(\boldsymbol{x})\| \quad \textit{for all } \boldsymbol{x} \in \mathcal{B}[\boldsymbol{x}^*, \delta]. \tag{8}$$

*We call* $\nu$ *a* Polyak-Łojasiewicz *constant at* $\boldsymbol{x}^*$*. In this context, we say that the* Polyak-Łojasiewicz Constraint Qualification/condition *(PŁCQ) is satisfied at* $\boldsymbol{x}^*$*. In this way, the reference to the function* $\Phi$ *is subsumed.*

Researchers employ a hierarchy of analytical techniques to establish the EB property for constrained optimization problems. A common and often practical approach involves verifying that the problem satisfies a standard CQ known to be stronger than EB. Numerous such CQs exist. More sophisticated analyses, which draw upon the powerful toolkit of variational analysis and nonsmooth geometry, establish the EB property via the *metric regularity* or *metric subregularity* of set-valued mappings associated with the constraints [46, 67]. Other techniques might address tilt stability associated with the derivative of the quadratic penalty function – typically, requiring the set of local minimizers to be isolated points, which might not be appropriate in some contexts; see, for instance, [31, Theorem 3.3] or, in a simpler form but in the convex context [32].

In contrast to these approaches, this paper investigates the least restrictive CQ sufficient to guarantee boundedness of computed multipliers generated by the safeguarded ALM. As detailed in the subsequent sections, this condition corresponds to EB, which provides a sharp characterization of when the computed multiplier sequence is guaranteed to be bounded for the safeguarded augmented Lagrangian schemes covered by our analysis.

# 3 Boundedness of computed multipliers under the Polyak-Łojasiewicz condition

In this section, we prove that safeguarded ALMs have bounded computed multipliers whenever the infeasibility measure satisfies the *Polyak-Łojasiewicz* (PŁ) *inequality*, that is, PŁCQ holds.

The following theorem presents a condition that ensures good algorithmic behavior for the safeguarded ALM. It demonstrates that the PŁ condition can indeed guarantee boundedness of computed multipliers associated with the method and, consequently, as stated in Proposition 2.1, KKT accumulation points.

**Theorem 3.1.** *Let $\boldsymbol{x}^* \in \Omega$ and assume that the functions defining the feasible set and the objective function are locally Lipschitz continuous, with Lipschitz constants strictly bounded by $L$ in a neighborhood of $\boldsymbol{x}^*$. Assume that the PŁ condition holds at $\boldsymbol{x}^*$ and let $\mathbb{K} \subset \mathbb{N}$ be an infinite subset such that the subsequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{K}}$ generated by the safeguarded ALM converges to $\boldsymbol{x}^*$. Then the associated subsequences of computed multipliers $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{K}}$ and $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{K}}$ as in* (5) *are bounded.*

*Proof.* Under the hypotheses, there exist sequences $\{\boldsymbol{v}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\bar{\boldsymbol{\mu}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, $\{\bar{\boldsymbol{\lambda}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ generated by the safeguarded ALM, and an infinite subset $\mathbb{K} \subset \mathbb{N}$ such that $\lim_{k\in\mathbb{K}} \boldsymbol{x}^k = \boldsymbol{x}^*$ and $\lim_{k\in\mathbb{K}} \boldsymbol{v}^k = \boldsymbol{0}$, where

$$\boldsymbol{v}^k := \nabla f(\boldsymbol{x}^k) + \sum_{i=1}^p \left(\bar{\mu}_i^k + \rho_k g_i(\boldsymbol{x}^k)\right)_+ \nabla g_i(\boldsymbol{x}^k) + \sum_{j=1}^q \left(\bar{\lambda}_j^k + \rho_k h_j(\boldsymbol{x}^k)\right) \nabla h_j(\boldsymbol{x}^k).$$

By the Lipschitz condition, we have that $\|\nabla f(\boldsymbol{x}^k) - \boldsymbol{v}^k\| \le L$ for all $k \in \mathbb{K}$ large enough. Let $M := \max\{\mu_{\max}, |\lambda_{\max}|, |\lambda_{\min}|\}$. Thus $\rho_k \left\|\nabla \Phi(\boldsymbol{x}^k)\right\|$ can be written equivalently as

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

By hypothesis, $\Phi$ satisfies the PŁ inequality at $\boldsymbol{x}^*$. Using (8), we have, for all sufficiently large $k \in \mathbb{K}$,

$$\rho_k \Phi(\boldsymbol{x}^k)^{1/2} \le \nu \rho_k \left\|\nabla \Phi(\boldsymbol{x}^k)\right\| \le \nu S,$$

for some $\nu > 0$. Thus, for all $k \in \mathbb{K}$ large enough, we have

$$\begin{aligned}
\|\boldsymbol{\mu}^k\| &= \|(\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k))_+\| \le \|(\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k))_+ - \rho_k \boldsymbol{g}(\boldsymbol{x}^k)_+\| + \rho_k \|\boldsymbol{g}(\boldsymbol{x}^k)_+\| \\
&\le \|\bar{\boldsymbol{\mu}}^k\| + \sqrt{2}\rho_k \Phi(\boldsymbol{x}^k)^{1/2} \le \sqrt{p}M + \sqrt{2}\nu S,
\end{aligned}$$

and similarly, $\|\boldsymbol{\lambda}^k\| \le \sqrt{q}M + \sqrt{2}\nu S$. Hence, $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{K}}$ and $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{K}}$ are bounded, as stated. □

**Remark 3.1.** *The fixed compact safeguarding set $[0, \mu^{\max}]^p \times [\lambda^{\min}, \lambda^{\max}]^q$ is not essential. The proof requires only that the safeguarded multiplier estimates remain uniformly bounded along the convergent primal subsequence under consideration. Hence, the result applies to any safeguarding strategy with this property, including schemes in which the safeguarding sets vary or are even allowed to expand without bound, as in [20, 30], provided that the corresponding safeguarded estimates remain uniformly bounded on the subsequence.*

We now focus on the reverse implication. To demonstrate that the Polyak-Łojasiewicz condition is a necessary condition for the boundedness of the Lagrange multiplier sequences in the safeguarded ALM, a preliminary technical lemma is required. This result can be understood as a specialized application of [9, Lemma 4.3].

**Lemma 3.1.** *Consider a point $\boldsymbol{x}^* \in \mathbb{R}^n$, not necessarily feasible for* (1)*, and $\delta > 0$ such that the derivatives of the constraint functions are Lipschitz continuous with constant $N$ on the compact set $\mathcal{K} := \overline{\text{co}}\left(\mathcal{J} \cup \mathcal{P}_\Omega(\mathcal{J}) + \mathcal{B}[\mathbf{0}, \delta]\right)$, where $\mathcal{J} := \mathcal{B}[\boldsymbol{x}^*, \delta]$. For any sequences $\{\boldsymbol{y}^k\}_{k\in\mathbb{N}} \subset \mathcal{J} \setminus \Omega$ and $\{\epsilon_k\}_{k\in\mathbb{N}} \subset (0, \delta)$, there exist sequences $\{\bar{\boldsymbol{y}}^k\}_{k\in\mathbb{N}}, \{\boldsymbol{w}^k\}_{k\in\mathbb{N}}, \{\boldsymbol{z}^k\}_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ and $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ satisfying, for all $k$, the following relations:*

$$\frac{\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k}{\|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\|} = \boldsymbol{w}^k + \rho_k \nabla\Phi(\boldsymbol{z}^k), \qquad \bar{\boldsymbol{y}}^k \in \mathcal{P}_\Omega(\boldsymbol{y}^k), \qquad \rho_k > 1/\min\{\text{dist}(\boldsymbol{y}^k, \Omega), \epsilon_k\},$$

$$\boldsymbol{g}(\bar{\boldsymbol{y}}^k)^T \boldsymbol{g}(\boldsymbol{z}^k)_+ = 0, \qquad \max\{\|\boldsymbol{z}^k - \bar{\boldsymbol{y}}^k\|, \|\boldsymbol{w}^k\|\} \le \min\{\text{dist}(\boldsymbol{y}^k, \Omega), \epsilon_k\},$$

$$\textit{and} \quad \text{dist}(\boldsymbol{y}^k, \Omega) \le \text{dist}(\boldsymbol{y}^k, \Omega)^2 + 2\rho_k \Phi(\boldsymbol{z}^k)^{1/2} \left( \frac{3\sqrt{2}}{4} \sqrt{p+q} N \,\text{dist}(\boldsymbol{y}^k, \Omega)^2 + \Phi(\boldsymbol{y}^k)^{1/2} \right).$$

*Proof.* Fix an arbitrary $k \in \mathbb{N}$ throughout the proof and choose an arbitrary $\bar{\boldsymbol{y}}^k \in \mathcal{P}_\Omega(\boldsymbol{y}^k)$. Then, $\text{dist}(\boldsymbol{y}^k, \Omega) = \left\|\bar{\boldsymbol{y}}^k - \boldsymbol{y}^k\right\|$. By continuity, for each $i \in I_p$,

$$\text{if } g_i(\bar{\boldsymbol{y}}^k) < 0, \text{ there exists } \delta_i^k > 0 \text{ such that } g_i(\boldsymbol{x}) < 0 \ \forall \boldsymbol{x} \in \mathcal{B}[\bar{\boldsymbol{y}}^k, \delta_i^k]. \tag{9}$$

Define

$$\bar{\epsilon}_k := \min\left\{ \min_{i \in I_p;\ g_i(\bar{\boldsymbol{y}}^k) < 0} \{\delta_i^k\},\ \text{dist}(\boldsymbol{y}^k, \Omega), \epsilon_k \right\}.$$

From [56, Lemma 2.1] we have $(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k)/\|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\| \in T_\Omega^\circ(\bar{\boldsymbol{y}}^k)$, where $T_\Omega^\circ(\bar{\boldsymbol{y}}^k)$ is the polar of the tangent cone to $\Omega$ at $\bar{\boldsymbol{y}}^k$.

By [9, Lemma 4.3], there exist $\rho_k > 0$, $\boldsymbol{z}^k \in \mathbb{R}^n$ and $\boldsymbol{w}^k \in \mathbb{R}^n$ such that

$$\frac{\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k}{\|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\|} = \boldsymbol{w}^k + \rho_k \nabla\Phi(\boldsymbol{z}^k), \quad \|\boldsymbol{z}^k - \bar{\boldsymbol{y}}^k\| \le \bar{\epsilon}_k,$$
$$\|\boldsymbol{w}^k\| \le \bar{\epsilon}_k, \quad \text{and} \quad \rho_k > 1/\bar{\epsilon}_k. \tag{10}$$

Now, if $g_i(\bar{\boldsymbol{y}}^k) < 0$ then, as $\|\boldsymbol{z}^k - \bar{\boldsymbol{y}}^k\| \le \bar{\epsilon}_k \le \delta_i^k$, expression (9) gives $g_i(\boldsymbol{z}^k) < 0$. So, $g_i(\boldsymbol{z}^k)_+ = 0$. Thus, we have proved the third of the five expressions in the statement. Consequently, except for the final inequality, all remaining expressions follow directly from observing the validity of (10).

Let us prove the last inequality in the statement. For each $i \in I_p$, since $\nabla g_i$ is Lipschitz continuous on $\mathcal{K}$, we have

$$\left|g_i(\boldsymbol{y}^k) - g_i(\bar{\boldsymbol{y}}^k) - \nabla g_i(\bar{\boldsymbol{y}}^k)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right)\right| \le \frac{N}{2} \left\|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right\|^2 = \frac{N}{2} \text{dist}(\boldsymbol{y}^k, \Omega)^2. \tag{11}$$

Since $\|\boldsymbol{z}^k - \bar{\boldsymbol{y}}^k\| \le \bar{\epsilon}_k$ and $\bar{\boldsymbol{y}}^k \in \mathcal{P}_\Omega(\boldsymbol{y}^k) \subset \mathcal{P}_\Omega(\mathcal{J})$, it follows that $\boldsymbol{z}^k \in \mathcal{P}_\Omega(\mathcal{J}) + \mathcal{B}[\mathbf{0}, \delta] \subset \mathcal{K}$. Using again the Lipschitz continuity of $\nabla g_i$ on $\mathcal{K}$, we have

$$\begin{aligned} &(\nabla g_i(\boldsymbol{z}^k) - \nabla g_i(\bar{\boldsymbol{y}}^k))^T (\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k) \\ \le\ & \left\|\nabla g_i(\boldsymbol{z}^k) - \nabla g_i(\bar{\boldsymbol{y}}^k)\right\| \left\|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right\| \\ \le\ & N \,\text{dist}(\boldsymbol{y}^k, \Omega)^2. \end{aligned} \tag{12}$$

Thus,

$$\begin{aligned} &\rho_k g_i(\boldsymbol{z}^k)_+ \nabla g_i(\boldsymbol{z}^k)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right) \\ =\ & \rho_k g_i(\boldsymbol{z}^k)_+ \left[(\nabla g_i(\boldsymbol{z}^k) - \nabla g_i(\bar{\boldsymbol{y}}^k))^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right) + \nabla g_i(\bar{\boldsymbol{y}}^k)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right)\right] \\ \le\ & \rho_k g_i(\boldsymbol{z}^k)_+ \left[N \,\text{dist}(\boldsymbol{y}^k, \Omega)^2 + g_i(\boldsymbol{y}^k) - g_i(\bar{\boldsymbol{y}}^k) + \frac{N}{2} \text{dist}(\boldsymbol{y}^k, \Omega)^2\right] \\ =\ & \frac{3N}{2} \rho_k g_i(\boldsymbol{z}^k)_+ \,\text{dist}(\boldsymbol{y}^k, \Omega)^2 + \rho_k g_i(\boldsymbol{z}^k)_+ g_i(\boldsymbol{y}^k) - \rho_k g_i(\boldsymbol{z}^k)_+ g_i(\bar{\boldsymbol{y}}^k) \\ \le\ & \frac{3N}{2} \rho_k g_i(\boldsymbol{z}^k)_+ \,\text{dist}(\boldsymbol{y}^k, \Omega)^2 + \rho_k g_i(\boldsymbol{z}^k)_+ g_i(\boldsymbol{y}^k)_+, \end{aligned} \tag{13}$$

where we used (11), (12) and the fact that $g_i(\boldsymbol{z}^k)_+ g_i(\bar{\boldsymbol{y}}^k) = 0$. Similarly, for $j \in I_q$,

$$\begin{aligned} &\rho_k h_j(\boldsymbol{z}^k)\nabla h_j(\boldsymbol{z}^k)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right) \\ &\leq \frac{3N}{2}\rho_k |h_j(\boldsymbol{z}^k)| \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 + \rho_k |h_j(\boldsymbol{z}^k)||h_j(\boldsymbol{y}^k)|. \end{aligned} \tag{14}$$

Additionally, as $\|\boldsymbol{w}^k\| \leq \bar{\epsilon}_k \leq \operatorname{dist}(\boldsymbol{y}^k, \Omega) = \|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\|$ we have

$$\left(\boldsymbol{w}^k\right)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right) \leq \|\boldsymbol{w}^k\|\|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\| \leq \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2. \tag{15}$$

By (2) and (10) we have

$$\begin{aligned} \operatorname{dist}(\boldsymbol{y}^k, \Omega) = \|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\| = \left(\boldsymbol{w}^k\right)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right) + \rho_k \nabla\Phi(\boldsymbol{z}^k)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right) \\ = \left(\boldsymbol{w}^k\right)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right) + \rho_k \left(\sum_{i=1}^{p} g_i(\boldsymbol{z}^k)_+ \nabla g_i(\boldsymbol{z}^k) + \sum_{i=1}^{q} h_j(\boldsymbol{z}^k)\nabla h_j(\boldsymbol{z}^k)\right)^T \left(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\right). \end{aligned}$$

Therefore, by (13), (14), and (15) we have

$$\begin{aligned} \operatorname{dist}(\boldsymbol{y}^k, \Omega) \leq\ & \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 + \frac{3N}{2}\rho_k \left(\sum_{i=1}^{p} g_i(\boldsymbol{z}^k)_+ + \sum_{j=1}^{q} |h_j(\boldsymbol{z}^k)|\right) \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 \\ & + \rho_k \left(\sum_{i=1}^{p} g_i(\boldsymbol{z}^k)_+ g_i(\boldsymbol{y}^k)_+ + \sum_{j=1}^{q} |h_j(\boldsymbol{z}^k)||h_j(\boldsymbol{y}^k)|\right) \\ \leq\ & \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 + \frac{3\sqrt{2}}{2}\sqrt{p+q} N \rho_k \Phi(\boldsymbol{z}^k)^{1/2} \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 \\ & + 2\rho_k \Phi(\boldsymbol{z}^k)^{1/2}\Phi(\boldsymbol{y}^k)^{1/2} \\ =\ & \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 + 2\rho_k \Phi(\boldsymbol{z}^k)^{1/2} \left(\frac{3\sqrt{2}}{4}\sqrt{p+q} N \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 + \Phi(\boldsymbol{y}^k)^{1/2}\right), \end{aligned}$$

where to obtain the last inequality, we used Cauchy-Schwarz's inequality and the equivalence of norms. This completes the proof. ☐

We will use the following theorem, which establishes the equivalence between EB and PŁCQ in the proof that follows. To keep the current discussion focused, its proof is deferred to Section 4.1.

**Theorem 3.2.** *A feasible point satisfies EB if and only if PŁCQ is satisfied.*

To place the preceding equivalence in context, recall that analogous equivalences are known for several notions of error bounds. In the convex setting, Hölderian error bounds are equivalent to Kurdyka–Łojasiewicz (KŁ) inequalities with the corresponding desingularizing functions; see [23]. The case of KŁ exponent one-half is especially relevant. For smooth functions, it corresponds to the Polyak–Łojasiewicz (PŁ) inequality, whereas, for convex functions, it is equivalent to several standard error-bound conditions, including quadratic growth, subdifferential error bounds, proximal error bounds, and Luo–Tseng-type error bounds; see [42, Chapter 8]. A subdifferential error bound estimates the distance to the solution set in terms of the distance from zero to the subdifferential; in the smooth case, this becomes a gradient, or first-order, error bound.

The gradient error bound case was studied by Karimi, Nutini, and Schmidt [45, Theorem 2] for smooth functions with Lipschitz continuous gradients. They show that it is equivalent to the PŁ inequality, and that either condition implies quadratic growth. In general, without convexity, the converse implication need not hold. Under convexity, however, the gradient error bound, the PŁ inequality, and quadratic growth are all equivalent.

The EB constraint qualification is equivalent to quadratic growth for the quadratic infeasibility measure; see the discussion after Remark 2.1. Theorem 3.2 shows that, for this particular infeasibility measure, quadratic growth is also equivalent to the PŁ inequality. Hence, the infeasibility measure considered here enjoys the same equivalence between quadratic growth and the PŁ inequality that holds in the smooth convex setting, even though the measure itself need not be convex.

A further relevant comparison is the theory of Rebjock and Boumal [64, Section 2], recalled in Section 2. This theory yields, among other equivalences, a local equivalence between the PŁ inequality and quadratic growth for twice continuously differentiable functions. It applies to the quadratic infeasibility measure only when this measure is twice continuously differentiable. Such regularity generally fails in the presence of inequality constraints, since the function need not be twice continuously differentiable at points with at least one active constraint. Our proof does not require this additional smoothness; instead, it uses the particular structure of the infeasibility measure and therefore also covers the inequality-constrained setting.

Finally, the following result shows that PŁCQ is the weakest condition ensuring that the computed multipliers of the safeguarded ALM remain bounded.

**Theorem 3.3.** *Let $\boldsymbol{x}^* \in \Omega$ be feasible for* (1)*. PŁCQ holds at $\boldsymbol{x}^*$ whenever, for every differentiable function $f$ and every application of the safeguarded ALM to solve problem* (1) *in which the primal sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ generated by it has $\boldsymbol{x}^*$ as an accumulation point (say, $\{\boldsymbol{x}^k\}_{k\in\mathbb{K}}$ converges to $\boldsymbol{x}^*$), the sequence of computed multipliers $\{\boldsymbol{\lambda}^k\}_{k\in\mathbb{K}}$ and $\{\boldsymbol{\mu}^k\}_{k\in\mathbb{K}}$, as given in* (5)*, possess a bounded subsequence.*

*Equivalently, if PŁCQ fails, then, for some objective function $f$, there exists an application of the safeguarded ALM that results in unbounded computed multipliers.*

*Proof.* The proof is by contradiction. Suppose that PŁCQ does not hold at $\boldsymbol{x}^* \in \Omega$. Then, by Theorem 3.2, the EB condition cannot hold at $\boldsymbol{x}^*$ either. Therefore, there must exist a sequence $\{\boldsymbol{y}^k\}_{k\in\mathbb{N}}$ of infeasible points converging to $\boldsymbol{x}^*$ such that

$$\operatorname{dist}(\boldsymbol{y}^k, \Omega) > k\Phi(\boldsymbol{y}^k)^{1/2} \quad \text{for all } k \in \mathbb{N}. \tag{16}$$

Thus, we can apply Lemma 3.1 to obtain sequences $\{\bar{\boldsymbol{y}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\boldsymbol{z}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, $\{\boldsymbol{w}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, and $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$, such that, for all $k \in \mathbb{N}$,

$$\frac{\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k}{\|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\|} = \boldsymbol{w}^k + \rho_k \nabla\Phi(\boldsymbol{z}^k), \tag{17}$$

$$\bar{\boldsymbol{y}}^k \in \mathcal{P}_\Omega(\boldsymbol{y}^k), \quad \|\boldsymbol{z}^k - \bar{\boldsymbol{y}}^k\| \le \operatorname{dist}(\boldsymbol{y}^k, \Omega), \tag{18}$$

$$\|\boldsymbol{w}^k\| \le \operatorname{dist}(\boldsymbol{y}^k, \Omega), \quad \rho_k > \frac{1}{\operatorname{dist}(\boldsymbol{y}^k, \Omega)} \tag{19}$$

$$\begin{aligned}\text{and} \quad \operatorname{dist}(\boldsymbol{y}^k, \Omega) &\le \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 \\ &+ 2\rho_k \Phi(\boldsymbol{z}^k)^{1/2} \left( \frac{3\sqrt{2}}{4} \sqrt{p+q}\, N \operatorname{dist}(\boldsymbol{y}^k, \Omega)^2 + \Phi(\boldsymbol{y}^k)^{1/2} \right).\end{aligned} \tag{20}$$

Since $\lim_{k\in\mathbb{N}} \operatorname{dist}(\boldsymbol{y}^k, \Omega) = 0$, the sequence $\{\rho_k\}_{k\in\mathbb{N}}$ must diverge to infinity and $\lim_{k\to\infty} \boldsymbol{w}^k = \boldsymbol{0}$. Also, for all $k \in \mathbb{N}$,

$$\|\boldsymbol{z}^k - \boldsymbol{x}^*\| \le \|\boldsymbol{z}^k - \bar{\boldsymbol{y}}^k\| + \|\bar{\boldsymbol{y}}^k - \boldsymbol{y}^k\| + \|\boldsymbol{y}^k - \boldsymbol{x}^*\| \le 2\operatorname{dist}(\boldsymbol{y}^k, \Omega) + \|\boldsymbol{y}^k - \boldsymbol{x}^*\|,$$

which implies that $\lim_{k\to\infty} \boldsymbol{z}^k = \boldsymbol{x}^*$ . We may assume without loss of generality that for some $\boldsymbol{v} \in \mathbb{R}^n$, $\lim_{k\to\infty}(\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k)/\|\boldsymbol{y}^k - \bar{\boldsymbol{y}}^k\| = \boldsymbol{v}$. Now, (17) gives $\lim_{k\to\infty} \|-\boldsymbol{v} + \rho_k \nabla\Phi(\boldsymbol{z}^k)\| = 0$, that is,

$$\lim_{k\to\infty} \left\| -\boldsymbol{v} + \left( \sum_{i=1}^{p} [\rho_k g_i(\boldsymbol{z}^k)_+] \nabla g_i(\boldsymbol{z}^k) + \sum_{j=1}^{q} [\rho_k h_j(\boldsymbol{z}^k)] \nabla h_j(\boldsymbol{z}^k) \right) \right\| = 0.$$

As shown in the proof of [4, Theorem 1], after possibly repeating some terms to comply with the penalty-update rule, the sequence $\{\boldsymbol{z}^k\}_{k\in\mathbb{N}}$ can be realized as a primal sequence generated by SALM applied to the auxiliary problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} -\boldsymbol{v}^T(\boldsymbol{x} - \boldsymbol{x}^*) \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}, \qquad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},$$

with $\lambda^{\min} = \lambda^{\max} = \mu^{\max} = 0$ as input data. Hence, the safeguarded estimates are identically zero, and the computed multipliers defined in (5) are

$$\boldsymbol{\mu}^k = \rho_k \boldsymbol{g}(\boldsymbol{z}^k)_+, \qquad \boldsymbol{\lambda}^k = \rho_k \boldsymbol{h}(\boldsymbol{z}^k). \tag{21}$$

Recall that the hypothesis of the theorem states that every primal sequence generated by the safeguarded ALM and converging to $\boldsymbol{x}^*$ admits an infinite set of indices along which the corresponding computed multiplier subsequence is bounded. After considering possible term repetitions described above, $\{\boldsymbol{z}^k\}_{k\in\mathbb{N}}$ is precisely such a sequence. Therefore, there exists an infinite set $\mathbb{L} \subset \mathbb{N}$ and a constant $L > 0$ such that

$$\rho_k \Phi(\boldsymbol{z}^k)^{1/2} \leq L \qquad \text{for every } k \in \mathbb{L}.$$

Notice that this conclusion remains valid even if $\rho_k \to +\infty$, since the hypothesis concerns the computed multipliers (21), rather than the penalty parameters themselves. Using the last bound in (20), we obtain

$$\begin{aligned} &\left[1 - \left(1 + \frac{3\sqrt{2}}{4}\sqrt{p+q}\, N\, 2L\right) \operatorname{dist}(\boldsymbol{y}^k, \Omega)\right] \operatorname{dist}(\boldsymbol{y}^k, \Omega) \\ &\qquad \leq 2L\Phi(\boldsymbol{y}^k)^{1/2} \qquad \text{for every } k \in \mathbb{L}. \end{aligned}$$

Hence, since $\lim_{k\to\infty} \operatorname{dist}(\boldsymbol{y}^k, \Omega) = 0$, for every sufficiently large $k \in \mathbb{L}$, we have

$$\frac{1}{2} \operatorname{dist}(\boldsymbol{y}^k, \Omega) \leq 2L\Phi(\boldsymbol{y}^k)^{1/2},$$

contradicting (16). Therefore, the proof is complete. □

To illustrate the theoretical results, we next consider two closely related nonlinear programs for which the iterates of the safeguarded augmented Lagrangian method can be computed explicitly. The example highlights the role played by the error bound condition in the behavior of the computed multipliers.

**Example 3.1.** *For $m \in \{1, 2\}$, consider*

$$(P_m) \qquad \min_{\boldsymbol{x}\in\mathbb{R}^2} -x_1 \quad \textit{subject to} \quad x_1 x_2 = 0, \qquad x_1^m = 0. \tag{22}$$

*Both problems have the same feasible set. Their infeasibility measures are*

$$\Phi_m(\boldsymbol{x}) = \frac{1}{2}\left(x_1^2 x_2^2 + x_1^{2m}\right).$$

*The error bound condition holds globally for $(P_1)$. By contrast, it fails at the origin for $(P_2)$.*

*We apply Algorithm 1 with exact minimization of the augmented Lagrangian subproblems. Fix $\lambda_{\max} > 1$ and $\lambda_{\min} = -\lambda_{\max}$, set $\bar{\boldsymbol{\lambda}}^1 = (0, 0)^T$ and $\rho_1 = 1$, and choose the following projection rule at Step 5:*

$$\bar{\boldsymbol{\lambda}}^{k+1} := \mathcal{P}_{[\lambda_{\min}, \lambda_{\max}]^2}(\boldsymbol{\lambda}^k), \qquad \boldsymbol{\lambda}^k := \bar{\boldsymbol{\lambda}}^k + \rho_k \begin{bmatrix} x_1^k x_2^k \\ (x_1^k)^m \end{bmatrix}.$$

*We first show by induction that the method may select $\boldsymbol{x}^k = (t_k, 0)^T$ with $\bar{\lambda}_1^k = \lambda_1^k = 0$ for every $k$. Indeed, suppose that $\bar{\lambda}_1^k = 0$. Since there are no inequality constraints, the PHR augmented Lagrangian is, for all $(x_1, x_2)^T \in \mathbb{R}^2$,*

$$\begin{aligned} \mathcal{L}_{\rho_k}^{(m)}\left((x_1, x_2)^T, \bar{\boldsymbol{\lambda}}^k\right) &= -x_1 + \frac{\rho_k}{2} x_1^2 x_2^2 + \frac{\rho_k}{2}\left(x_1^m + \frac{\bar{\lambda}_2^k}{\rho_k}\right)^2 \\ &\geq -x_1 + \frac{\rho_k}{2}\left(x_1^m + \frac{\bar{\lambda}_2^k}{\rho_k}\right)^2 = \mathcal{L}_{\rho_k}^{(m)}\left((x_1, 0)^T, \bar{\boldsymbol{\lambda}}^k\right). \end{aligned}$$

*The scalar function $t \mapsto \mathcal{L}_{\rho_k}^{(m)}((t, 0)^T, \bar{\boldsymbol{\lambda}}^k)$ is continuous and coercive. Let $t_k$ be one of its global minimizers. Consequently, $(t_k, 0)^T$ is a global minimizer of the augmented Lagrangian subproblem. Moreover, $\lambda_1^k = \bar{\lambda}_1^k + \rho_k \cdot t_k \cdot 0 = 0$, and the projection rule gives $\bar{\lambda}_1^{k+1} = 0$. The claim follows from $\bar{\lambda}_1^1 = 0$.*

*For $m = 1$, the scalar function $t \mapsto \mathcal{L}_{\rho_k}^{(1)}((t, 0)^T, \bar{\boldsymbol{\lambda}}^k)$ is differentiable and strictly convex. Since $t_k$ is its global minimizer, the first-order optimality condition yields*

$$0 = \frac{d}{dt}\mathcal{L}_{\rho_k}^{(1)}((t_k, 0)^T, \bar{\boldsymbol{\lambda}}^k) = -1 + \bar{\lambda}_2^k + \rho_k t_k.$$

*Therefore,*

$$t_k = \frac{1 - \bar{\lambda}_2^k}{\rho_k}, \qquad \lambda_2^k = \bar{\lambda}_2^k + \rho_k t_k = 1.$$

*Since* $\lambda_{\max} > 1$ *and* $\lambda_{\min} = -\lambda_{\max}$*, the projection is inactive. Thus,* $\boldsymbol{\lambda}^k = (0,1)^T$ *for every* $k$*, while* $\boldsymbol{x}^k = (0,0)^T$ *for every* $k \geq 2$*. In particular, the computed multiplier sequence is bounded.*

*For* $m = 2$*, we prove by induction that* $\bar{\lambda}_2^k \geq 0$ *for every* $k$*. The claim holds for* $k = 1$ *since* $\bar{\lambda}_2^1 = 0$*. Suppose that* $\bar{\lambda}_2^k \geq 0$*. Then*

$$\lambda_2^k = \bar{\lambda}_2^k + \rho_k t_k^2 \geq 0.$$

*The chosen projection rule therefore gives*

$$\bar{\lambda}_2^{k+1} = \min\{\lambda_{\max}, \lambda_2^k\} \geq 0.$$

*Thus,* $\bar{\lambda}_2^k \geq 0$ *for every* $k$*. Hence, one can now verify that the scalar function* $t \mapsto \mathcal{L}_{\rho_k}^{(2)}((t,0)^T, \bar{\boldsymbol{\lambda}}^k)$ *is differentiable and strictly convex. Since* $t_k$ *is its global minimizer, the first-order optimality condition yields*

$$0 = \frac{d}{dt}\mathcal{L}_{\rho_k}^{(2)}\big((t_k,0)^T, \bar{\boldsymbol{\lambda}}^k\big) = -1 + 2\bar{\lambda}_2^k t_k + 2\rho_k t_k^3.$$

*It follows that* $t_k > 0$ *and*

$$2\rho_k t_k^3 + 2\bar{\lambda}_2^k t_k = 1, \qquad \bar{\lambda}_2^{k+1} = \min\left\{\lambda_{\max}, \bar{\lambda}_2^k + \rho_k t_k^2\right\}. \tag{23}$$

*The second relation follows directly from the chosen projection rule.*

*We claim that* $\{\rho_k\}_{k\in\mathbb{N}}$ *diverges to infinity. Otherwise, the update rule in* ***Step 4*** *would imply that the nondecreasing penalty sequence is eventually constant, say* $\rho_k = \widehat{\rho}$*. Let* $\widehat{t} > 0$ *be the unique solution of* $2\widehat{\rho}\widehat{t}^3 + 2\lambda_{\max}\widehat{t} = 1$*. Since* $\bar{\lambda}_2^k \leq \lambda_{\max}$*, the first relation in* (23) *gives* $1 = 2\widehat{\rho}t_k^3 + 2\bar{\lambda}_2^k t_k \leq 2\widehat{\rho}t_k^3 + 2\lambda_{\max}t_k$*. As the map* $t \mapsto 2\widehat{\rho}t^3 + 2\lambda_{\max}t$ *is strictly increasing on* $[0, +\infty)$*, we obtain* $t_k \geq \widehat{t}$*. The second relation in* (23) *then implies that* $\bar{\lambda}_2^k = \lambda_{\max}$ *after finitely many iterations. From that point onward,* $t_k = \widehat{t}$ *and the constraint violation remains constant and positive. The sufficient-reduction test must therefore fail, forcing a penalty increase, which is a contradiction.*

*Finally, set* $s_k := \rho_k^{1/3} t_k$*. The first relation in* (23) *becomes* $2s_k^3 + 2\bar{\lambda}_2^k \rho_k^{-1/3} s_k = 1$*. This relation implies that* $s_k > 0$*. Hence,* $0 < s_k \leq 2^{-1/3}$*. Moreover,* $s_k = 2^{-1/3}(1 - 2\bar{\lambda}_2^k \rho_k^{-1/3} s_k)^{1/3}$*. Since* $\bar{\lambda}_2^k \leq \lambda_{\max}$*,* $\{\rho_k\}_{k\in\mathbb{N}}$ *diverges to infinity, and* $\{s_k\}_{k\in\mathbb{N}}$ *is bounded, it follows that*

$$\lim_{k\to\infty} s_k = 2^{-1/3}.$$

*Therefore,*

$$t_k = 2^{-1/3}\rho_k^{-1/3}\left(1 + o(1)\right),$$

*and hence*

$$\lambda_2^k = \bar{\lambda}_2^k + \rho_k t_k^2 \geq 2^{-2/3}\rho_k^{1/3}\left(1 + o(1)\right) \longrightarrow +\infty.$$

*Thus, two representations of the same feasible set lead to opposite behaviors: the computed multiplier sequence is bounded for* $(P_1)$ *and unbounded for* $(P_2)$*.*

Under sufficient second-order conditions (SOSC) and adequate control of subproblem errors, augmented Lagrangian methods are known to exhibit local primal-dual convergence. In particular, the local theory of Fernández–Solodov establishes $Q$-linear convergence of the primal-dual sequence for sufficiently large penalty parameters; see [34, Theorem 3.4]. Sufficiently large values of the penalty parameter are required in such local convergence results, even when strong constraint qualifications are met (see, e.g., [44, Theorem 6.4]). In practice, however, large penalty parameters may lead to numerical ill-conditioning. For this reason, the penalty parameter is often increased only moderately during the global phase, with the aim of improving primal feasibility before the method enters a local regime. From this perspective, boundedness of the computed multipliers is the missing link between the global convergence theory of safeguarded ALMs and the local primal-dual convergence theory of classical ALMs, as we now explain.

Indeed, the EB property guarantees that the computed multipliers remain bounded for the safeguarded ALM method covered by our assumptions. If, in addition, these bounded computed multipliers eventually lie inside the safeguarding box, then one may choose the safeguarded multiplier estimates to coincide with the computed multipliers. With this choice, the safeguarding mechanism becomes inactive from some iteration on, and the safeguarded ALM coincides with the classical ALM along the tail of the sequence. Therefore, under the additional assumptions required by Fernández–Solodov, their local $Q$-linear convergence result for the primal-dual sequence can be transferred to the safeguarded method.

Conversely, deficiencies in the algebraic description of the constraints, particularly the failure of the EB constraint qualification, may allow the safeguarded ALM to generate unbounded computed multipliers. In such a situation, the safeguarded multiplier update cannot, in general, be identified with the classical, non-safeguarded update. Thus, unbounded computed multipliers may prevent the safeguarded method from coinciding, along the tail of the sequence, with the classical ALM covered by local primal-dual convergence theory. This observation helps identify situations in which the algebraic description of the feasible set should be reconsidered. Along these lines, reduction techniques that provide alternative algebraic descriptions of the same feasible set for which the EB constraint qualification holds, such as those developed in [11, 55], may be worth further investigation.

Finally, boundedness of the computed multiplier sequence explains an important feature of safeguarded ALMs. We show that, in this situation, constraint violations are bounded by a term inversely proportional to the penalty parameter; see (7). This relation provides an explicit bound on the penalty growth needed to attain a prescribed level of feasibility and limits the method's exposure to penalty-induced ill-conditioning.

The following section investigates the validity of the EB constraint qualification without necessarily relying on a particular execution of the method.

# 4 Characterizing the Polyak-Łojasiewicz condition

The error bound concept is fundamental in variational analysis and optimization. Researchers have thoroughly explored its connections to metric regularity, stability analysis, and convergence of optimization algorithms [47, 52, 60]. A key related idea is the PŁ inequality [51, 61]. While the rich interplay between these concepts is well understood, using a PL-type condition explicitly as a CQ for nonlinear programming appears to be a novel approach. Our main contribution is to systematically propose and validate such a condition, derived directly from the PŁ inequality framework (as presented in Definition 2.3), as a legitimate CQ. Equivalently, we prove Theorem 3.2. After this is established, we prove the equivalence between the mentioned condition and RQN.

## 4.1 The equivalence between error bound and the Polyak-Łojasiewicz condition

In this section, we prove Theorem 3.2, that is, the equivalence between the error bound CQ (Definition 2.2) and the PŁ condition (Definition 2.3) in the context of constrained optimization. The proof is given in two lemmas. First, we establish that the PŁ inequality for the infeasibility measure, when satisfied at a feasible point, implies the error bound condition.

**Lemma 4.1.** *If PŁCQ holds at a feasible point $\boldsymbol{x}^*$, then the EB condition holds at $\boldsymbol{x}^*$.*

*Proof.* Assume that EB does not hold at the feasible point $\boldsymbol{x}^* \in \Omega$. Noting that Theorem 3.2 is used solely within the proof of Theorem 3.3 to establish the failure of the EB condition, we can instead assume the failure of EB directly and follow the same proof strategy from Theorem 3.3 to construct the sequences $\{\boldsymbol{y}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ of infeasible points converging to $\boldsymbol{x}^*$, $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ diverging to infinity, $\{\boldsymbol{w}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ converging to zero, $\{\bar{\boldsymbol{y}}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ converging to $\boldsymbol{x}^*$, and $\{\boldsymbol{z}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ converging to $\boldsymbol{x}^*$ such that for all $k \in \mathbb{N}$, equations (16), (17), (18), (19), and (20) hold, where $N$ is an upper bound on the local Lipschitz constant of all the gradients $\nabla g_i, \nabla h_j$ around $\boldsymbol{x}^*$. From (17), the sequence $\{\rho_k \nabla\Phi(\boldsymbol{z}^k)\}_{k\in\mathbb{N}}$ is bounded. Thus, there exists a constant $C > 0$ such that

$$\rho_k \|\nabla\Phi(\boldsymbol{z}^k)\| \leq C \tag{24}$$

for all $k \in \mathbb{N}$. By hypothesis, the infeasibility measure $\Phi$ satisfies the PŁ inequality at $\boldsymbol{x}^*$. This means there exist $\nu > 0$ and a neighborhood of $\boldsymbol{x}^*$ such that $\Phi(\boldsymbol{z})^{1/2} \leq \nu\|\nabla\Phi(\boldsymbol{z})\|$ for any $\boldsymbol{z}$ in this neighborhood. Thus, for sufficiently large $k$:

$$\rho_k \Phi(\boldsymbol{z}^k)^{1/2} \leq \nu\, \rho_k \|\nabla\Phi(\boldsymbol{z}^k)\| \leq \nu C. \tag{25}$$

Using this bound in (20) gives

$$\left[1 - \left(1 + \frac{3\sqrt{2}}{4}\sqrt{p+q}N2\nu C\right)\operatorname{dist}(\boldsymbol{y}^k, \Omega)\right]\operatorname{dist}(\boldsymbol{y}^k, \Omega) \leq 2\nu C\Phi(\boldsymbol{y}^k)^{1/2}$$

for sufficiently large $k$. Since $\lim_{k\in\mathbb{N}} \operatorname{dist}(\boldsymbol{y}^k, \Omega) = 0$, we arrive at

$$\frac{1}{2}\operatorname{dist}(\boldsymbol{y}^k, \Omega) \leq 2\nu C\Phi(\boldsymbol{y}^k)^{1/2}$$

for sufficiently large $k$, which contradicts (16). This proves that the EB condition must hold at $\boldsymbol{x}^*$. □

We need the following technical lemma to establish the reverse implication:

**Lemma 4.2.** *Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a point, not necessarily feasible, and let $\delta > 0$. Assume the gradients of the constraint functions $g_i, i \in I_p$, and $h_j, j \in I_q$, have a common Lipschitz constant $N$ within the compact set $\mathcal{K} := \overline{\operatorname{co}}\left(\mathcal{J} \cup \mathcal{P}_\Omega(\mathcal{J}) + \mathcal{B}[\mathbf{0}, \delta]\right)$, where $\mathcal{J} := \mathcal{B}[\boldsymbol{x}^*, \delta]$. Let $\mathbb{K} \subset \mathbb{N}$ be an infinite index set and let $\{\boldsymbol{x}^k\}_{k\in\mathbb{K}}$ be a sequence such that $\boldsymbol{x}^k \in \mathcal{J} \setminus \Omega$ for every $k \in \mathbb{K}$. Then, for each $k \in \mathbb{K}$, the following inequality holds:*

$$\left(2 - \sqrt{2(p+q)}N\frac{\operatorname{dist}(\boldsymbol{x}^k, \Omega)^2}{\Phi(\boldsymbol{x}^k)^{1/2}}\right)\frac{\Phi(\boldsymbol{x}^k)}{\operatorname{dist}(\boldsymbol{x}^k, \Omega)} \leq \left\|\nabla\Phi(\boldsymbol{x}^k)\right\|. \tag{26}$$

*Proof.* Fix an arbitrary index $k \in \mathbb{K}$. Let $\boldsymbol{y}^k \in \mathcal{P}_\Omega(\boldsymbol{x}^k)$. By definition, $\|\boldsymbol{x}^k - \boldsymbol{y}^k\| = \operatorname{dist}(\boldsymbol{x}^k, \Omega)$. For each $i \in I_p$ and $j \in I_q$, there exist $\boldsymbol{u}_i^k$ and $\boldsymbol{v}_j^k$ on the line segment connecting $\boldsymbol{x}^k$ and $\boldsymbol{y}^k$ such that:

$$\begin{aligned} g_i(\boldsymbol{x}^k) &= g_i(\boldsymbol{y}^k) + \nabla g_i(\boldsymbol{u}_i^k)^T(\boldsymbol{x}^k - \boldsymbol{y}^k), \\ h_j(\boldsymbol{x}^k) &= h_j(\boldsymbol{y}^k) + \nabla h_j(\boldsymbol{v}_j^k)^T(\boldsymbol{x}^k - \boldsymbol{y}^k). \end{aligned} \tag{27}$$

Since $\boldsymbol{x}^k \in \mathcal{J} \subset \mathcal{K}$ and $\boldsymbol{y}^k \in \mathcal{P}_\Omega(\mathcal{J}) \subset \mathcal{K}$, we have, by the convexity of $\mathcal{K}$, that $\boldsymbol{u}_i^k \in \mathcal{K}$ and $\boldsymbol{v}_j^k \in \mathcal{K}$. Now, using that $g_i(\boldsymbol{x}^k)_+ g_i(\boldsymbol{x}^k) = g_i(\boldsymbol{x}^k)_+^2$ and $\boldsymbol{y}^k \in \Omega$, by (27) we get

$$\begin{aligned} &\frac{2\Phi(\boldsymbol{x}^k)}{\operatorname{dist}(\boldsymbol{x}^k, \Omega)} \\ &= \frac{1}{\operatorname{dist}(\boldsymbol{x}^k, \Omega)}\left(\sum_{i=1}^p g_i(\boldsymbol{x}^k)_+ g_i(\boldsymbol{x}^k) + \sum_{j=1}^q h_j(\boldsymbol{x}^k)h_j(\boldsymbol{x}^k)\right) \\ &\leq \left(\sum_{i=1}^p g_i(\boldsymbol{x}^k)_+\nabla g_i(\boldsymbol{u}_i^k) + \sum_{j=1}^q h_j(\boldsymbol{x}^k)\nabla h_j(\boldsymbol{v}_j^k)\right)^T \frac{\boldsymbol{x}^k - \boldsymbol{y}^k}{\operatorname{dist}(\boldsymbol{x}^k, \Omega)} \\ &= \left(\sum_{i=1}^p g_i(\boldsymbol{x}^k)_+\left(\nabla g_i(\boldsymbol{u}_i^k) - \nabla g_i(\boldsymbol{x}^k)\right)\right. \\ &\quad \left. + \sum_{j=1}^q h_j(\boldsymbol{x}^k)\left(\nabla h_j(\boldsymbol{v}_j^k) - \nabla h_j(\boldsymbol{x}^k)\right)\right)^T \frac{\boldsymbol{x}^k - \boldsymbol{y}^k}{\operatorname{dist}(\boldsymbol{x}^k, \Omega)} \\ &\quad + \left(\sum_{i=1}^p g_i(\boldsymbol{x}^k)_+\nabla g_i(\boldsymbol{x}^k) + \sum_{j=1}^q h_j(\boldsymbol{x}^k)\nabla h_j(\boldsymbol{x}^k)\right)^T \frac{\boldsymbol{x}^k - \boldsymbol{y}^k}{\operatorname{dist}(\boldsymbol{x}^k, \Omega)}. \end{aligned}$$

Using the Cauchy-Schwarz inequality, (2) and the fact that $\|\boldsymbol{x}^k - \boldsymbol{y}^k\| = \operatorname{dist}(\boldsymbol{x}^k, \Omega)$, we arrive at

$$\begin{aligned} &\frac{2\Phi(\boldsymbol{x}^k)}{\operatorname{dist}(\boldsymbol{x}^k, \Omega)} \\ &\leq \sum_{i=1}^p g_i(\boldsymbol{x}^k)_+\left\|\nabla g_i(\boldsymbol{u}_i^k) - \nabla g_i(\boldsymbol{x}^k)\right\| \end{aligned}$$

$$+\sum_{j=1}^{q}|h_j(\boldsymbol{x}^k)|\left\|\nabla h_j(\boldsymbol{v}_j^k)-\nabla h_j(\boldsymbol{x}^k)\right\|+\left\|\nabla\Phi(\boldsymbol{x}^k)\right\|$$

$$\leq N\operatorname{dist}(\boldsymbol{x}^k,\Omega)\left(\sum_{i=1}^{p}g_i(\boldsymbol{x}^k)_+ +\sum_{j=1}^{q}|h_j(\boldsymbol{x}^k)|\right)+\left\|\nabla\Phi(\boldsymbol{x}^k)\right\|,$$

where in the last inequality we used Lipschitz continuity and the fact that $\|\boldsymbol{u}_i^k-\boldsymbol{x}^k\|\leq\|\boldsymbol{x}^k-\boldsymbol{y}^k\|$ and $\|\boldsymbol{v}_j^k-\boldsymbol{x}^k\|\leq\|\boldsymbol{x}^k-\boldsymbol{y}^k\|$. Since

$$\sum_{i=1}^{p}g_i(\boldsymbol{x}^k)_+ +\sum_{j=1}^{q}|h_j(\boldsymbol{x}^k)|\leq\sqrt{p+q}\sqrt{\|\boldsymbol{g}(\boldsymbol{x}^k)_+\|^2+\|\boldsymbol{h}(\boldsymbol{x}^k)\|^2}=\sqrt{2(p+q)}\Phi(\boldsymbol{x}^k)^{1/2},$$

we conclude that

$$\frac{2\Phi(\boldsymbol{x}^k)}{\operatorname{dist}(\boldsymbol{x}^k,\Omega)}\leq\sqrt{2(p+q)}N\frac{\operatorname{dist}(\boldsymbol{x}^k,\Omega)^2}{\Phi(\boldsymbol{x}^k)^{1/2}}\frac{\Phi(\boldsymbol{x}^k)}{\operatorname{dist}(\boldsymbol{x}^k,\Omega)}+\left\|\nabla\Phi(\boldsymbol{x}^k)\right\|.$$

Rearranging and factoring out the common terms yields the inequality in the statement. □

We can now demonstrate that the error bound condition implies the Polyak-Łojasiewicz condition.

**Lemma 4.3.** *If the EB condition holds at a feasible point $\boldsymbol{x}^*$, then PŁCQ holds at $\boldsymbol{x}^*$.*

*Proof.* Assume, for the sake of contradiction, that the error bound condition with constant $\nu>0$ holds at $\boldsymbol{x}^*$, but PŁCQ does not hold. Define $\rho_k:=k$ for all $k\in\mathbb{N}$. Then, there exists a sequence $\{\boldsymbol{y}^k\}_{k\in\mathbb{N}}$ of infeasible points converging to $\boldsymbol{x}^*$ such that

$$\Phi(\boldsymbol{y}^k)^{1/2}>\rho_k\|\nabla\Phi(\boldsymbol{y}^k)\|,\text{ for all }k\in\mathbb{N}.$$

The error bound condition implies that for all sufficiently large $k\in\mathbb{N}$,

$$\frac{\operatorname{dist}(\boldsymbol{y}^k,\Omega)^2}{\Phi(\boldsymbol{y}^k)^{1/2}}\leq\nu\operatorname{dist}(\boldsymbol{y}^k,\Omega)\to 0.$$

By Lemma 4.2, as the term inside parenthesis in (26) converges to 2, we conclude that

$$\frac{\rho_k\Phi(\boldsymbol{y}^k)}{\operatorname{dist}(\boldsymbol{y}^k,\Omega)}\leq\rho_k\left\|\nabla\Phi(\boldsymbol{y}^k)\right\|$$

for sufficiently large $k\in\mathbb{N}$. Consequently, for large $k$,

$$\rho_k\Phi(\boldsymbol{y}^k)^{1/2}=\frac{\rho_k\Phi(\boldsymbol{y}^k)}{\operatorname{dist}(\boldsymbol{y}^k,\Omega)}\frac{\operatorname{dist}(\boldsymbol{y}^k,\Omega)}{\Phi(\boldsymbol{y}^k)^{1/2}}\leq\nu\rho_k\|\nabla\Phi(\boldsymbol{y}^k)\|<\nu\Phi(\boldsymbol{y}^k)^{1/2},$$

which is a contradiction. □

The preceding analysis establishes an equivalence between two fundamental concepts in optimization. Connections between the error bound condition and other constraint qualifications are well-known in the literature. For instance, for convex optimization problems defined by inequality constraints, a global error bound holds across the entire feasible set if and only if *Abadie's CQ* (ACQ) is satisfied at the same points [49]. In this context, ACQ is the weakest CQ possible for convex problems, as it is equivalent to *Guignard's CQ* (GCQ) [38]. For non-convex optimization problems defined by equality constraints, ensuring an error bound requires not only ACQ but also Clarke's regularity. This necessity is established in [57, Theorem 1] and further discussed in [5, Remark 3], where the joint condition is characterized in terms of a linearized tangent subspace. This highlights the fact that non-convexity introduces geometric complexities that necessitate stronger assumptions than simply ACQ. For general optimization problems involving nonlinear equality and inequality constraints, stronger constraint qualifications, such as CRSC [8] and QN, are typically required to guarantee the error bound property. However, to the best of our knowledge, no direct characterization of the error bound property based solely on geometric properties of the feasible set is currently available. In contrast, this section provides a purely algebraic characterization of the error bound condition by establishing its equivalence with PŁCQ.

We proceed to characterize it in terms of a quasinormality-type CQ. We will show that condition RQN presented in Definition 2.1 is necessary and sufficient for the error bound property to hold. This is particularly relevant to PŁCQ since being equivalent to RQN implies that it is strictly weaker than the family of constant rank conditions [6].

## 4.2 The equivalence between the Polyak-Łojasiewicz condition and relaxed quasinormality

Relaxed quasinormality [6] has been conceived as a weakening of MFCQ-like conditions specifically for the safeguarded ALM, guaranteeing boundedness of the computed multipliers generated by the method. While this established the value of RQN as a generalization of previous CQs, such as quasinormality and constant rank-type CQs, its connection to other fundamental concepts remained an open question. For instance, whether RQN is sufficient to guarantee a local error bound remained unclear until now. In this section, we provide a definitive answer to this question and reveal the central role played by RQN in quasinormality-type conditions. We show that RQN not only guarantees the existence of an error bound, but it is, remarkably, equivalent to it. This result tightly binds the geometry of the feasible set to the behavior of infeasible sequences that reach feasible points. This is a direct consequence of the next theorem and Theorem 3.2, as summarized in Corollary 4.1.

**Theorem 4.1.** *A feasible point satisfies RQN if and only if PŁCQ holds at that point.*

*Proof.* By [6, Theorem 3.1], RQN implies that any primal sequence that has $\boldsymbol{x}^*$ as an accumulation point is such that the corresponding subsequence of computed multipliers is bounded. By Theorem 3.3, PŁCQ holds.

Let us prove the converse. Assume that RQN fails at $\boldsymbol{x}^*$, and let us show that PŁCQ also fails. By Definition 2.1, there exists a sequence of infeasible points $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \to \boldsymbol{x}^*$ and $(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ satisfying items 1, 2, and 3 of that definition. By Lemma 3.1, there exist sequences $\{\bar{\boldsymbol{y}}^k\}_{k\in\mathbb{N}}, \{\boldsymbol{w}^k\}_{k\in\mathbb{N}}, \{\boldsymbol{z}^k\}_{k\in\mathbb{N}}$ and $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ satisfying the following relations for all sufficiently large $k \in \mathbb{N}$:

$$\frac{\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k}{\|\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k\|} = \boldsymbol{w}^k + \rho_k \nabla\Phi(\boldsymbol{z}^k), \tag{28}$$

$$\bar{\boldsymbol{y}}^k \in \mathcal{P}_\Omega(\boldsymbol{x}^k), \quad \|\boldsymbol{z}^k - \bar{\boldsymbol{y}}^k\| \le \operatorname{dist}(\boldsymbol{x}^k, \Omega), \quad \text{and} \quad \|\boldsymbol{w}^k\| \le \operatorname{dist}(\boldsymbol{x}^k, \Omega), \tag{29}$$

with $\{\rho_k\}_{k\in\mathbb{N}}$ diverging to infinity. Since $\|\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k\| = \operatorname{dist}(\boldsymbol{x}^k, \Omega)$ for all $k \in \mathbb{N}$, we have $\lim_{k\to\infty} \bar{\boldsymbol{y}}^k = \lim_{k\to\infty} \boldsymbol{z}^k = \boldsymbol{x}^*$. Now, let us define the sequences $\bar{\boldsymbol{\mu}}^k := \rho_k \boldsymbol{g}(\boldsymbol{z}^k)_+$ and $\bar{\boldsymbol{\lambda}}^k := \rho_k \boldsymbol{h}(\boldsymbol{z}^k)$ for all $k \in \mathbb{N}$. By PŁCQ, for sufficiently large $k \in \mathbb{N}$, we have $\|(\bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\| = \sqrt{2}\rho_k \Phi(\boldsymbol{z}^k)^{1/2} \le \sqrt{2}\nu\rho_k \left\|\nabla\Phi(\boldsymbol{z}^k)\right\|$. Therefore $\{\|(\bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\|\}_{k\in\mathbb{N}}$ is bounded since by (28), $\{\rho_k\|\nabla\Phi(\boldsymbol{z}^k)\|\}_{k\in\mathbb{N}}$ is bounded. Using (2), equation (28) can be written as

$$\begin{aligned} \frac{\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k}{\|\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k\|} &= \boldsymbol{w}^k + \sum_{i=1}^p \bar{\mu}_i^k \nabla g_i(\boldsymbol{z}^k) + \sum_{j=1}^q \bar{\lambda}_j^k \nabla h_j(\boldsymbol{z}^k) \\ &= \boldsymbol{v}^k + \sum_{i=1}^p \bar{\mu}_i^k \nabla g_i(\bar{\boldsymbol{y}}^k) + \sum_{j=1}^q \bar{\lambda}_j^k \nabla h_j(\bar{\boldsymbol{y}}^k), \end{aligned} \tag{30}$$

where, for all $k \in \mathbb{N}$,

$$\boldsymbol{v}^k := \boldsymbol{w}^k + \sum_{i=1}^p \bar{\mu}_i^k \left(\nabla g_i(\boldsymbol{z}^k) - \nabla g_i(\bar{\boldsymbol{y}}^k)\right) + \sum_{j=1}^q \bar{\lambda}_j^k \left(\nabla h_j(\boldsymbol{z}^k) - \nabla h_j(\bar{\boldsymbol{y}}^k)\right),$$

which converges to zero. Let us also define, for all $k \in \mathbb{N}$, $\boldsymbol{u}^k := \sum_{i=1}^p \mu_i \nabla g_i(\bar{\boldsymbol{y}}^k) + \sum_{j=1}^q \lambda_j \nabla h_j(\bar{\boldsymbol{y}}^k)$, which converges to zero by item 1 of Definition 2.1. Let us choose a sequence of positive real numbers $\{s_k\}_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} s_k = \infty$, $\lim_{k\to\infty} s_k\|\boldsymbol{u}^k\| = 0$, and $\lim_{k\to\infty} s_k \operatorname{dist}(\boldsymbol{x}^k, \Omega) = 0$, which clearly exists. The inner product of $\boldsymbol{u}^k$ with $s_k(\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k)$ yields

$$s_k \left( \sum_{i=1}^{p} \mu_i \nabla g_i(\bar{\boldsymbol{y}}^k) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\bar{\boldsymbol{y}}^k) \right)^T (\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k) = s_k(\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k)^T \boldsymbol{u}^k,$$

while taking the inner product of (30) with $(\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k)$ gives:

$$(\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k)^T \boldsymbol{v}^k + \sum_{i=1}^{p} \bar{\mu}_i^k \nabla g_i(\bar{\boldsymbol{y}}^k)^T (\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k) + \sum_{j=1}^{q} \bar{\lambda}_j^k \nabla h_j(\bar{\boldsymbol{y}}^k)^T (\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k) = \operatorname{dist}(\boldsymbol{x}^k, \Omega).$$

Adding the above expressions gives

$$\begin{aligned} \operatorname{dist}(\boldsymbol{x}^k, \Omega) + (\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k)^T (s_k \boldsymbol{u}^k - \boldsymbol{v}^k) = &\sum_{i=1}^{p} \tilde{\mu}_i^k \nabla g_i(\bar{\boldsymbol{y}}^k)^T (\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k) \\ &+ \sum_{j=1}^{q} \tilde{\lambda}_j^k \nabla h_j(\bar{\boldsymbol{y}}^k)^T (\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k), \end{aligned} \tag{31}$$

where $\tilde{\boldsymbol{\mu}}^k := (s_k \boldsymbol{\mu} + \bar{\boldsymbol{\mu}}^k)$ and $\tilde{\boldsymbol{\lambda}}^k := (s_k \boldsymbol{\lambda} + \bar{\boldsymbol{\lambda}}^k)$ for all $k \in \mathbb{N}$. Let $N$ be a common local Lipschitz constant for all functions $\nabla g_i$ and $\nabla h_j$ around $\boldsymbol{x}^*$. Then we can write for all $k$ large enough

$$\begin{aligned} &\tilde{\mu}_i^k \nabla g_i(\bar{\boldsymbol{y}}^k)^T \left(\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k\right) \geq \tilde{\mu}_i^k (g_i(\boldsymbol{x}^k) - g_i(\bar{\boldsymbol{y}}^k)) - \frac{N}{2} \tilde{\mu}_i^k \operatorname{dist}(\boldsymbol{x}^k, \Omega)^2 \quad \text{and} \\ &\tilde{\lambda}_j^k \nabla h_j(\bar{\boldsymbol{y}}^k)^T \left(\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k\right) \geq \tilde{\lambda}_j^k (h_j(\boldsymbol{x}^k) - h_j(\bar{\boldsymbol{y}}^k)) - \frac{N}{2} |\tilde{\lambda}_j^k| \operatorname{dist}(\boldsymbol{x}^k, \Omega)^2. \end{aligned}$$

For the equality constraints, the multipliers $\tilde{\lambda}_j^k$ are free in sign. Hence, when Taylor's estimate is multiplied by $\tilde{\lambda}_j^k$, the remainder term must be controlled using $|\tilde{\lambda}_j^k|$. Now, using this in (31) and, rearranging, we arrive at

$$\begin{aligned} &\left[ 1 + \frac{(\boldsymbol{x}^k - \bar{\boldsymbol{y}}^k)^T}{\operatorname{dist}(\boldsymbol{x}^k, \Omega)} \left(s_k \boldsymbol{u}^k - \boldsymbol{v}^k\right) + \frac{N}{2} \operatorname{dist}(\boldsymbol{x}^k, \Omega) \left( \sum_{i=1}^{p} \tilde{\mu}_i^k + \sum_{j=1}^{q} |\tilde{\lambda}_j^k| \right) \right] \operatorname{dist}(\boldsymbol{x}^k, \Omega) \\ &\qquad \geq \sum_{i=1}^{p} \tilde{\mu}_i^k (g_i(\boldsymbol{x}^k) - g_i(\bar{\boldsymbol{y}}^k)) + \sum_{j=1}^{q} \tilde{\lambda}_j^k (h_j(\boldsymbol{x}^k) - h_j(\bar{\boldsymbol{y}}^k)). \end{aligned}$$

By the definitions of $s^k$, $\tilde{\boldsymbol{\mu}}^k$, and $\tilde{\boldsymbol{\lambda}}^k$, the term in square brackets on the left-hand side converges to 1. Therefore, for sufficiently large $k$,

$$\begin{aligned} 2 \operatorname{dist}(\boldsymbol{x}^k, \Omega) &\geq \sum_{i=1}^{p} \tilde{\mu}_i^k (g_i(\boldsymbol{x}^k) - g_i(\bar{\boldsymbol{y}}^k)) + \sum_{j=1}^{q} \tilde{\lambda}_j^k (h_j(\boldsymbol{x}^k) - h_j(\bar{\boldsymbol{y}}^k)) \\ &\geq \sum_{i=1}^{p} \tilde{\mu}_i^k g_i(\boldsymbol{x}^k) + \sum_{j=1}^{q} \tilde{\lambda}_j^k h_j(\boldsymbol{x}^k), \end{aligned}$$

where we used the fact that $\bar{\boldsymbol{y}}^k \in \Omega$. Thus, dividing both sides by $\Phi(\boldsymbol{x}^k)^{1/2} > 0$ and using the definitions of $\tilde{\boldsymbol{\mu}}^k$ and $\tilde{\boldsymbol{\lambda}}^k$ we obtain

$$\begin{aligned} \frac{2 \operatorname{dist}(\boldsymbol{x}^k, \Omega)}{\Phi(\boldsymbol{x}^k)^{1/2}} \geq s_k &\left( \sum_{i=1}^{p} \mu_i \frac{g_i(\boldsymbol{x}^k)_+}{\Phi(\boldsymbol{x}^k)^{1/2}} + \sum_{j=1}^{q} \lambda_j \frac{h_j(\boldsymbol{x}^k)}{\Phi(\boldsymbol{x}^k)^{1/2}} \right) \\ &+ \left( \sum_{i=1}^{p} \bar{\mu}_i^k \frac{g_i(\boldsymbol{x}^k)}{\Phi(\boldsymbol{x}^k)^{1/2}} + \sum_{j=1}^{q} \bar{\lambda}_j^k \frac{h_j(\boldsymbol{x}^k)}{\Phi(\boldsymbol{x}^k)^{1/2}} \right). \end{aligned} \tag{32}$$

Now, since the sequence $\{\Phi(\boldsymbol{x}^k)^{-1/2}(\boldsymbol{g}(\boldsymbol{x}^k), \boldsymbol{h}(\boldsymbol{x}^k))\}_{k\in\mathbb{N}}$ is bounded, we might pass to a subsequence, if necessary, to assume the existence of its limit.

Notice that the second term inside the parentheses on the right-hand side of (32) is bounded. Hence, as $\{s^k\}_{k\in\mathbb{N}}$ diverges to infinity, to show that the error bound condition fails and consequently, PŁCQ by Theorem 3.2, it remains to show that

$$\lim_{k\to\infty}\left(\sum_{i=1}^{p}\mu_i\frac{g_i(\boldsymbol{x}^k)_+}{\Phi(\boldsymbol{x}^k)^{1/2}}+\sum_{j=1}^{q}\lambda_j\frac{h_j(\boldsymbol{x}^k)}{\Phi(\boldsymbol{x}^k)^{1/2}}\right)>0. \tag{33}$$

Now, by condition 3 in the definition of RQN, we have $\mu_i g_i(\boldsymbol{x}^k)\geq 0$ and $\lambda_j h_j(\boldsymbol{x}^k)\geq 0$ for every $i\in I_p$, $j\in I_q$, and $k\in\mathbb{N}$. Hence, each term in the expression above is nonnegative. Therefore, if the limit in (33) is not positive, it must be zero. Suppose that the limit in (33) is zero. Therefore

$$\lim_{k\to\infty}\mu_i\frac{g_i(\boldsymbol{x}^k)_+}{\Phi(\boldsymbol{x}^k)^{1/2}}=0,\quad\forall i\in I_p\quad\text{ and }\quad\lim_{k\to\infty}\lambda_j\frac{h_j(\boldsymbol{x}^k)}{\Phi(\boldsymbol{x}^k)^{1/2}}=0,\quad\forall j\in I_q.$$

If $\lambda_j\neq 0$, we must have $\lim_{k\to\infty}h_j(\boldsymbol{x}^k)/\Phi(\boldsymbol{x}^k)^{1/2}=0$. If $\lambda_j=0$, since $\Phi(\boldsymbol{x}^k)\geq t_k$, where $t_k$ is given in item 3 of Definition 2.1, we get from this item that $\lim_{k\to\infty}h_j(\boldsymbol{x}^k)/\Phi(\boldsymbol{x}^k)^{1/2}=0$. Similarly, $\lim_{k\to\infty}g_i(\boldsymbol{x}^k)_+/\Phi(\boldsymbol{x}^k)^{1/2}=0$ for all $i\in I_p$. This is a contradiction, since $\|(\boldsymbol{g}(\boldsymbol{x}^k)_+,\boldsymbol{h}(\boldsymbol{x}^k))\|/\Phi(\boldsymbol{x}^k)^{1/2}=\sqrt{2}$ for all $k\in\mathbb{N}$. □

For completeness, we state the aforementioned equivalence between RQN and EB.

**Corollary 4.1.** *A feasible point satisfies RQN if and only if EB holds at that point.*

**Remark 4.1.** *It was shown in [5] that the RQN condition, combined with a constraint qualification termed strong error bound (Strong-EB), is sufficient to ensure that the accumulation points of a primal sequence generated by a second-order augmented Lagrangian algorithm necessarily satisfy the so-called* weak second-order necessary condition *(WSOC). However, Strong-EB includes the standard EB property in its definition. As EB is equivalent to RQN, the hypothesis required to guarantee convergence to a WSOC point can be simplified to requiring only the Strong-EB property. This highlights that the assumptions made in [5] are reasonably weak. The Strong-EB condition itself is implied by several established CQs from the literature, as summarized in [5, Figure 1].*

A powerful tool for analyzing the convergence of algorithms is the so-called sequential optimality conditions (AKKT [3], AGP [54], PAKKT [2], among others). AKKT is the simplest of these conditions and it holds at a feasible point $\boldsymbol{x}^*$ whenever one may find a sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}\to\boldsymbol{x}^*$ and (possibly unbounded) sequences $\{(\boldsymbol{\mu}^k,\boldsymbol{\lambda}^k)\}_{k\in\mathbb{N}}\subset\mathbb{R}^p_+\times\mathbb{R}^q$ such that $\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}^k,\boldsymbol{\mu}^k,\boldsymbol{\lambda}^k)\to\boldsymbol{0}$ and $\min\{\boldsymbol{\mu}^k,-\boldsymbol{g}(\boldsymbol{x}^k)\}\to\boldsymbol{0}$. Different conditions enforce different (stronger) complementarity measures. One can also define a constraint qualification (for example, associated with AKKT sequences) by imposing that, independently of the objective function, a point satisfying the AKKT condition is necessarily a KKT point. This is termed AKKT-regularity, see [9, 10]. Let us demonstrate that EB is insufficient to guarantee these CQs. The exception is AL-regularity [4], which is implied by EB due to [4, Proposition 1] and Theorem 3.2. Figure 1 shows the relations discussed here.

**Example 4.1.** *Consider the problem*

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}x_2\quad s.t.\quad x_1x_2=0,\quad x_1=0.$$

*As already pointed out in Example 3.1, this problem satisfies EB at the feasible point* $\boldsymbol{x}^*:=(0,0)^T$*. However,* $\boldsymbol{x}^*$ *is a non-KKT point that satisfies the AKKT condition. To see this, it is enough to consider the sequences* $(x_1^k,x_2^k):=(1/k,-1/k)$ *and* $(\lambda_1^k,\lambda_2^k):=(-k,-1)$*. Consequently,* $\boldsymbol{x}^*$ *cannot conform to AKKT-regularity. Similarly, one can check that AGP-regularity [10] and PAKKT-regularity [2] are also not satisfied. This example also shows that an AKKT sequence may require an unbounded dual sequence even when the error bound condition holds.*

The following figure synthesizes the known relationships from the literature with our findings. Detailed justifications and counterexamples for the strictness of implications between different CQs can be found in [7, 8, 10]. Notably, Theorem 3.3 shows that any condition in Figure 1 not implying the error bound leads, for at least one problem, to unbounded computed multipliers in the safeguarded augmented Lagrangian method.

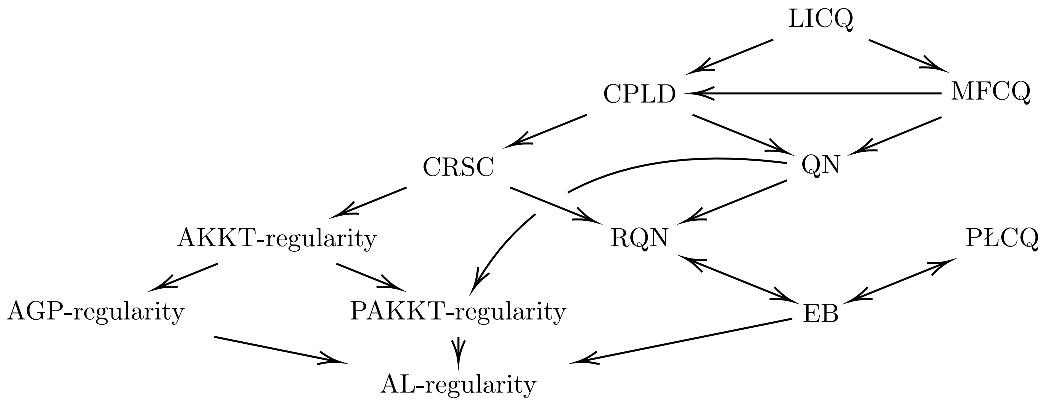


Figure 1: Landscape of constraint qualifications for which the safeguarded ALM accumulation points satisfy the KKT conditions. Conditions that imply EB also guarantee dual convergence. Arrows indicate strict implications.

# 5 Global convergence without assuming feasibility of the accumulation point

In our previous results, we always assumed that an accumulation point $\boldsymbol{x}^*$ of a sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ generated by the safeguarded ALM is feasible. The idea behind this assumption is that the algorithm is known to be naturally driven toward feasibility: any accumulation point $\boldsymbol{x}^*$ of the generated sequence is stationary for the infeasibility measure $\Phi$, that is, $\nabla\Phi(\boldsymbol{x}^*) = 0$ [1, Theorem 4.1]. However, Proposition 2.1 demonstrates that boundedness of the computed multipliers implies feasibility (and stationarity) of the accumulation point. This result suggests investigating a reasonable assumption that can be verified at potentially infeasible accumulation points and still guarantees boundedness of the computed multipliers. Then, the boundedness of the computed multipliers would yield feasibility in addition to the standard result concerning stationarity. To this end, we extend the definition of a constraint qualification to possibly infeasible points.

In [28], a condition fulfilling this requirement called Extended-LICQ was defined. However, this condition makes little sense as it may fail for elementary (linear) problems. An Extended-MFCQ condition was defined in [43, Definition 2.1], which does not suffer from this drawback; however, like MFCQ, this condition fails when the feasible region has an empty interior. On the other hand, in [6], a weaker condition called Extended-RQN was defined. It was proven that when this condition holds at a potentially infeasible accumulation point, the sequence of computed multipliers associated with safeguarded ALM is bounded.

This section aims to generalize the analysis presented earlier to accommodate infeasible points, thereby encompassing the results obtained under the Extended-RQN condition as a particular case. The definition of Extended-RQN is as follows:

**Definition 5.1** ([6]). *It is said that* Extended relaxed quasinormality *holds at an arbitrary point $\boldsymbol{x}^* \in \mathbb{R}^n$, not necessarily feasible, whenever there is no sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ converging to $\boldsymbol{x}^*$ and vectors $\boldsymbol{\mu} \in \mathbb{R}^p_+$ and $\boldsymbol{\lambda} \in \mathbb{R}^q$ satisfying requirements 1, 2 of Definition 2.1 and the following one*

*3'. for all $i \in I_p$, $j \in I_q$ and $k \in \mathbb{N}$,*

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

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

We define next an extended version of PŁCQ to infeasible points.

**Definition 5.2.** *We say that $\boldsymbol{x}^* \in \mathbb{R}^n$, not necessarily feasible, satisfies Extended-PŁCQ when there exist constants $\nu > 0$ and $\delta > 0$ such that*

$$\Phi(\boldsymbol{x})^{1/2} \leq \nu \|\nabla \Phi(\boldsymbol{x})\| \text{ for all } \boldsymbol{x} \in \mathcal{B}[\boldsymbol{x}^*, \delta].$$

Notice that the only difference with respect to Definition 2.3 of PŁCQ is the lack of the requirement that $\boldsymbol{x}^*$ is feasible. The following result shows that the computed multipliers generated by the safeguarded ALM remain bounded under Extended-PŁCQ; that is, feasibility is not required as an assumption to obtain, along a subsequence, primal-dual convergence to a feasible KKT pair.

**Proposition 5.1.** *Let $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ be a primal sequence generated by the safeguarded ALM. If $\boldsymbol{x}^*$ is the limit of a subsequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{K}}$ and satisfies the Extended-PŁCQ, then the sequence of computed multipliers $\{(\boldsymbol{\mu}^k, \boldsymbol{\lambda}^k)\}_{k\in\mathbb{K}}$ is bounded, with expressions given in* (5)*, making $\boldsymbol{x}^*$ a feasible point that satisfies the KKT conditions.*

*Proof.* The proof is essentially the same as Theorem 3.1. In view of Proposition 2.1, it is enough to prove boundedness of $\{(\boldsymbol{\mu}^k, \boldsymbol{\lambda}^k)\}_{k\in\mathbb{K}}$. Following the proof of Theorem 3.1, there exists $S > 0$ such that $\rho_k \left\|\nabla \Phi(\boldsymbol{x}^k)\right\| \leq S$ for all $k \in \mathbb{K}$. Hence,

$$\rho_k \Phi(\boldsymbol{x}^k)^{1/2} \leq \nu \rho_k \left\|\nabla \Phi(\boldsymbol{x}^k)\right\| \leq \nu S, \text{ for all } k \in \mathbb{K}.$$

Defining $M := \max\{\mu_{\max}, |\lambda_{\min}|, |\lambda_{\max}|\}$, for sufficiently large $k \in \mathbb{K}$ it holds that

$$\begin{aligned} \|\boldsymbol{\mu}^k\| &= \|(\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k))_+\| \leq \|(\bar{\boldsymbol{\mu}}^k + \rho_k \boldsymbol{g}(\boldsymbol{x}^k))_+ - \rho_k \boldsymbol{g}(\boldsymbol{x}^k)_+\| + \rho_k \|\boldsymbol{g}(\boldsymbol{x}^k)_+\| \\ &\leq \|\bar{\boldsymbol{\mu}}^k\| + \sqrt{2}\rho_k \Phi(\boldsymbol{x}^k)^{1/2} \leq \sqrt{p}M + \sqrt{2}\nu S \end{aligned}$$

and, similarly, $\|\boldsymbol{\lambda}^k\| \leq \sqrt{q}M + \sqrt{2}\nu S$. Thus, the sequence of computed multipliers is bounded, as we wanted to prove. □

Clearly, at feasible points, Extended-PŁCQ and Extended-RQN are equivalent to the error bound property. However, the situation is different at infeasible points. Extended-PŁCQ connects with the property that the accumulation points of the safeguarded ALM are stationary to the infeasibility measure $\Phi$ in the following way:

**Proposition 5.2.** *Extended-PŁCQ is valid at an infeasible point if and only if the gradient of the squared infeasibility measure $\Phi$ is non-zero at this point.*

*Proof.* If Extended-PŁCQ is valid at an infeasible point, it is clear from the definition that the gradient of the squared infeasibility measure is non-zero at that point. To prove the converse, it suffices to fix $\delta > 0$ such that $\min\{\|\nabla \Phi(\boldsymbol{x})\|, \Phi(\boldsymbol{x})^{1/2}\} \geq \delta$ for all $\boldsymbol{x} \in \mathcal{B}[\boldsymbol{x}^*, \delta]$ and define

$$\nu := \frac{\sup_{\boldsymbol{x}\in\mathcal{B}(\boldsymbol{x}^*,\delta)} \Phi(\boldsymbol{x})^{1/2}}{\inf_{\boldsymbol{x}\in\mathcal{B}(\boldsymbol{x}^*,\delta)} \|\nabla \Phi(\boldsymbol{x})\|} > 0.$$

Thus, considering such $\nu$, Definition 5.2 holds. □

Therefore, using the classical result that an accumulation point $\boldsymbol{x}^*$ is such that $\nabla \Phi(\boldsymbol{x}^*) = \boldsymbol{0}$ (see [21]), we have by Proposition 5.2 that under Extended-PŁCQ this accumulation point is necessarily feasible. Next, we establish that Extended-PŁCQ generalizes Extended-RQN.

**Proposition 5.3.** *The Extended-RQN condition at a point guarantees Extended-PŁCQ at the same point.*

*Proof.* The result is true for feasible points. Now, consider $\boldsymbol{x}^*$ an infeasible point that does not satisfy Extended-PŁCQ. Then $\nabla \Phi(\boldsymbol{x}^*) = \boldsymbol{0}$. Therefore, we may define $\boldsymbol{\mu} := \boldsymbol{g}(\boldsymbol{x}^*)_+$, $\boldsymbol{\lambda} := \boldsymbol{h}(\boldsymbol{x}^*)$, and the constant sequence $\boldsymbol{x}^k := \boldsymbol{x}^*$ for all $k \in \mathbb{N}$. With these definitions, items 1 and 2 of Definition 2.1 and item 3' of Definition 5.1 are satisfied. This shows that Extended-RQN fails at $\boldsymbol{x}^*$. □

**Remark 5.1.** *Note that Extended-PŁCQ at an infeasible point $\boldsymbol{x}^* \in \mathbb{R}^n$ can also be interpreted as a "linear independence" type condition. Specifically, from Proposition 5.2, it can be stated via the implication*

$$\sum_{i=1}^{p} g_i(\boldsymbol{x}^*)_+ \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{q} h_j(\boldsymbol{x}^*) \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0} \quad \Longrightarrow \quad (\boldsymbol{g}(\boldsymbol{x}^*)_+, \boldsymbol{h}(\boldsymbol{x}^*)) = (\boldsymbol{0}, \boldsymbol{0}),$$

*which is clearly weaker than both Extended-MFCQ and Extended-RQN.*

Let us provide an example in which Extended-PŁCQ holds, but Extended-RQN fails. More specifically, Extended-PŁCQ is well-posed in the sense that if it holds at a certain point, it also holds in an open neighborhood around that point. This observation follows directly from its definition. Well-posedness is a desirable property, as it is closely related to the stability of numerical algorithms (see [57]).

Surprisingly, the following example demonstrates that Extended-RQN is not well-posed. The existence of such an example is unexpected, since the classical RQN condition—being equivalent to the EB condition—always holds in a sufficiently small feasible neighborhood of a specified feasible point fulfilling it; in other words, the set of feasible points satisfying RQN is open relative to the constraint set. In contrast, Extended-RQN may not be valid in arbitrarily small neighborhoods of a point that meets the condition. That is, the set of points satisfying Extended-RQN is not always open.

**Example 5.1.** *Consider the set*

$$\Omega := \left\{ (x_1,x_2) \in \mathbb{R}^2 \;\middle|\; \begin{array}{l} g_1(x_1,x_2) := -(x_1-1)^2 - x_2^2 + 1 \le 0, \\ g_2(x_1,x_2) := -(x_1+1)^2 - x_2^2 + 1 \le 0, \\ g_3(x_1,x_2) := (x_1-x_2)^2 \le 0, \\ h_1(x_1,x_2) := x_2 = 0 \end{array} \right\} = \{(0,0)\}.$$

*We have*

$$\nabla g_1(\boldsymbol{x}) = \begin{bmatrix} -2(x_1-1) \\ -2x_2 \end{bmatrix}, \quad \nabla g_2(\boldsymbol{x}) = \begin{bmatrix} -2(x_1+1) \\ -2x_2 \end{bmatrix},$$
$$\nabla g_3(\boldsymbol{x}) = \begin{bmatrix} 2(x_1-x_2) \\ -2(x_1-x_2) \end{bmatrix}, \quad \nabla h_1(\boldsymbol{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

*The point $\boldsymbol{x}^* := (0,0)^T$ satisfies CRSC [8, Definition 1.3] given that $\nabla g_1(\boldsymbol{x}^*) = -\nabla g_2(\boldsymbol{x}^*)$ and $\nabla g_3(\boldsymbol{x}^*) = \boldsymbol{0}$ with the dimension of the space spanned by the set $\{\nabla g_1(\boldsymbol{x}), \nabla g_2(\boldsymbol{x}), \nabla g_3(\boldsymbol{x}), \nabla h_1(\boldsymbol{x})\}$ being 2 for every $\boldsymbol{x} \in \mathbb{R}^2$. Thus, [8, Theorem 5.5] attests that the error bound condition holds. Consequently, Extended-RQN also holds at $\boldsymbol{x}^*$. Let us now construct a sequence converging to the origin such that Extended-RQN is not satisfied at any of its terms. Take $\boldsymbol{y}^\ell := (1/\ell, 1/\ell)$ for every $\ell > 2$ and set $\boldsymbol{\mu}^\ell := (0,0,1)^T$ and $\boldsymbol{\lambda}^\ell := 0$ for every $\ell > 2$. Thus $\mu_1^\ell \nabla g_1(\boldsymbol{y}^\ell) + \mu_2^\ell \nabla g_2(\boldsymbol{y}^\ell) + \mu_3^\ell \nabla g_3(\boldsymbol{y}^\ell) + \lambda_1^\ell \nabla h_1(\boldsymbol{y}^\ell) = \boldsymbol{0}$ with $(\boldsymbol{\mu}^\ell, \boldsymbol{\lambda}^\ell) \neq \boldsymbol{0}$. Fixing $\ell > 2$, it remains to find a sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ converging to $\boldsymbol{y}^\ell$ satisfying the condition imposed by item 3' of Definition 5.1. Note that $g_1(\boldsymbol{y}^\ell) > 0$, $g_2(\boldsymbol{y}^\ell) < 0$, $g_3(\boldsymbol{y}^\ell) = 0$ and $h_1(\boldsymbol{y}^\ell) > 0$. Thus, item 3' simply requires that $g_3(\boldsymbol{x}^k) > 0$ for all $k \in \mathbb{N}$ for some sequence $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ converging to $\boldsymbol{y}^\ell$. It is enough to define $\boldsymbol{x}^k := \boldsymbol{y}^\ell + (1/k, 0)^T$ for all $k \in \mathbb{N}$. Thus, Extended-RQN fails at $\boldsymbol{y}^\ell$ for all $\ell > 2$. Since $\boldsymbol{y}^\ell$ is infeasible and $\nabla \Phi(\boldsymbol{y}^\ell) \neq \boldsymbol{0}$, Extended-PŁCQ holds at $\boldsymbol{y}^\ell$.*

We conclude this work by briefly stating how Extended-PŁCQ generalizes Theorem 3.3 to include infeasible points, establishing it as the weakest condition that guarantees a bounded sequence of computed multipliers. The reasoning is by contradiction. Assume that Extended-PŁCQ is false at an accumulation point. By definition, this means the point is stationary for the quadratic infeasibility measure. Consequently, a safeguarded ALM applied to problem (1) with null objective configured with the input parameters $\mu^{\max} = \lambda^{\min} = \lambda^{\max} = 0$ would accept such a stationary point of the infeasibility measure as a valid output of **Step 2** at each iteration. If we assume the validity of a condition guaranteeing boundedness of the computed multipliers associated with the method, then the point of interest must be feasible, as per Proposition 2.1. This, however, leads to a contradiction since we apply Theorem 3.3 to guarantee the standard PŁCQ, which is equivalent to Extended-PŁCQ at such a feasible point. This conclusion contradicts our initial assumption that Extended-PŁCQ was false. Thus, Extended-PŁCQ must be the minimal condition required to ensure that any converging primal sequence to the point of interest has a bounded associated multiplier sequence. It is also possible to see that the Extended-PŁCQ condition is the weakest condition that, when satisfied at an infeasible point, prevents the safeguarded ALM from having a subsequence converging to that specific point.

# 6 Final remarks

In optimization, many algorithms for constrained problems produce sequences of both primal and dual iterates. This work focuses on the boundedness of the computed multiplier sequence, also known as the dual sequence.

We identify the weakest constraint qualification that guarantees this boundedness property for the safeguarded augmented Lagrangian method considered here. The Mangasarian-Fromovitz Constraint Qualification is frequently used to ensure boundedness of computed multipliers, as it is equivalent to the set of Lagrange multipliers being bounded. However, strong constraint conditions depend on how the feasible set is described, and standard optimization techniques, such as presenting redundant constraints, may render these conditions false. On the other hand, computational experience has demonstrated that robust approaches, such as safeguarded augmented Lagrangian methods, can exhibit good numerical behavior even when standard constraint qualifications are not met. This work directly addresses the gap between theory and practice.

We investigated conditions under which the *safeguarded augmented Lagrangian method* ensures boundedness of its associated computed multipliers. Specifically, we investigated the least stringent constraint qualification required for this objective. In particular, we demonstrate that the sharp condition for boundedness of the computed multiplier sequence within our framework is a condition we term the *Polyak-Łojasiewicz Constraint Qualification* (Theorems 3.1 and 3.3). This constraint qualification refers to the Polyak-Łojasiewicz inequality applied to the quadratic infeasibility measure. We further show its equivalence to the *Error Bound Constraint Qualification* (Theorem 3.2; see also Section 4.1), also known as *metric subregularity constraint qualification* or *calmness* condition [16].

A novel characterization of the Polyak-Łojasiewicz constraint qualification is derived via gradients, which consequently also applies to the error bound condition. We prove that a recently proposed condition, weaker than several constraint qualifications typically used in algorithmic convergence—the *relaxed quasinormality* (Definition 2.1)—is equivalent to the Polyak-Łojasiewicz constraint qualification (Theorem 4.1). Relaxed quasinormality can be seen as a natural extension of the classic quasinormality condition, adding properties associated with penalty methods [6]. It is weaker than the weakest constant rank-type constraint qualification, known as *Constant Rank of the Subspace Component* [8, Definition 1.3]. Relaxed quasinormality is implied by a number of different constraint qualifications. This demonstrates the generality of the Polyak-Łojasiewicz condition and provides insight into the behavior of gradients near feasible points under the proposed constraint qualification.

Clearly, finding a solution to a problem is more relevant than merely identifying a feasible point. It is natural to question why one should focus on the feasibility problem rather than directly seeking a solution. Multiple justifications exist; see discussions in [25]. Additionally, the feasibility problem class is of great interest in a variety of applied mathematics fields [13, 22]. Interestingly, boundedness of the computed multipliers is sufficient to establish feasibility and stationarity of an accumulation point of the safeguarded augmented Lagrangian method (Proposition 2.1). In this work, we investigated the attainment of feasible points, along with the boundedness of the computed multipliers, using the safeguarded augmented Lagrangian method. We propose a condition, which we call Extended-PŁCQ. First, we prove that Extended-PŁCQ ensures that every accumulation point of the primal sequences generated by the safeguarded augmented Lagrangian method is a feasible solution (Proposition 5.1). We further establish that Extended-PŁCQ is equivalent to PŁCQ at feasible points. Additionally, we prove that, at infeasible points, Extended-PŁCQ is equivalent to the non-stationarity of the quadratic feasibility measure (Proposition 5.2), reducing to a classical result for the safeguarded augmented Lagrangian method concerning the stationarity of the quadratic infeasibility measure. Lastly, we prove that Extended-RQN is strictly stronger than Extended-PL (Proposition 5.3), improving a previously known result from the literature.


## Acknowledgments

The authors are grateful to the anonymous reviewers for their careful reading and constructive comments, which helped improve the presentation of this paper.


## References


[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309, 2008.

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

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

[4] R. Andreani, G. Haeser, L. M. Mito, A. Ramos, and L. D. Secchin. On the best achievable quality of limit points of augmented Lagrangian schemes. *Numerical Algorithms*, 90(2):851–877, 2022.

[5] R. Andreani, G. Haeser, R. W. Prado, M. L. Schuverdt, and L. D. Secchin. Global convergence of a second-order augmented Lagrangian method under an error bound condition. *Journal of Optimization Theory and Applications*, 206(2):54, Jun 2025.

[6] R. Andreani, G. Haeser, M. L. Schuverdt, and L. D. Secchin. A relaxed quasinormality condition and the boundedness of dual augmented Lagrangian sequences. *SIAM Journal on Optimization*, 35(4):2474–2489, 2025.

[7] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1):255–273, 2012.

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

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

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

[11] R. Andreani, M. da Rosa, and L. D. Secchin. A new constant-rank-type condition related to MFCQ and local error bounds. *Journal of Optimization Theory and Applications*, 209(1), 2026.

[12] R. Andreani, A. Ramos, and L. D. Secchin. Improving the global convergence of inexact restoration methods for constrained optimization problems. *SIAM Journal on Optimization*, 34(4):3429–3455, 2024.

[13] F. J. Aragón Artacho, R. Campoy, and M. K. Tam. The Douglas–Rachford algorithm for convex and nonconvex feasibility problems. *Mathematical Methods of Operations Research*, 91(2):201–240, Apr 2020.

[14] A. V. Arutyunov. Perturbations of extremal problems with constraints and necessary optimality conditions. *Journal of Soviet Mathematics*, 54(6):1342–1400, May 1991.

[15] K. Bai, J. J. Ye, and J. Zhang. Directional quasi-/pseudo-normality as sufficient conditions for metric subregularity. *SIAM Journal on Optimization*, 29(4):2625–2649, Jan. 2019.

[16] M. Benko, M. Červinka, and T. Hoheisel. Sufficient conditions for metric subregularity of constraint systems with applications to disjunctive and ortho-disjunctive programs. *Set-Valued and Variational Analysis*, 30(1):143–177, Mar 2022.

[17] M. Benko and P. Mehlitz. On the directional asymptotic approach in optimization theory. *Mathematical Programming*, 209(1):859–937, Jan. 2025.

[18] D. P. Bertsekas. Multiplier methods: A survey. *IFAC Proceedings Volumes*, 8(1):351–363, 1975. 6th IFAC World Congress (IFAC 1975), Part 1.

[19] D. P. Bertsekas. *Nonlinear Programming*. Athena scientific optimization and computation series. Athena Scientific, Belmont, Massachussets, 2 edition, 1999.

[20] E. G. Birgin, A. De Marchi, and P. Mehlitz. Elastically safeguarded augmented Lagrangian methods. Submitted. Available at `https://www.ime.usp.br/~egbirgin/publications/bdm2026-ealm.pdf`; accessed July 5, 2026, 2026.

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

[22] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[23] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165(2):471–507, Oct 2017.

[24] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, New York, NY, 1st edition, 2000.

[25] J. W. Chinneck. *Feasibility and Infeasibility in Optimization: Algorithms and Computational Methods*. Springer Publishing Company, Incorporated, 1st edition, 2007.

[26] W. Choi, C. Chun, Y. M. Jung, and S. Yun. On the linear convergence rate of Riemannian proximal gradient method. *Optimization Letters*, 19(3):667–687, Apr 2025.

[27] A. R. Conn, N. I. M. Gould, and P. L. Toint. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM Journal on Numerical Analysis*, 25(2):433–460, 1988.

[28] A. R. Conn, N. I. M. Gould, and P. L. Toint. *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*, volume 17 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, Heidelberg, 1 edition, 1992.

[29] A. R. Conn, N. I. M. Gould, and P. L. Toint. Numerical experiments with the LANCELOT package (release A) for large-scale nonlinear optimization. *Mathematical Programming*, 73(1):73–110, 1996.

[30] A. De Marchi, T. Hoheisel, and P. Mehlitz. Augmented Lagrangian methods for fully convex composite optimization, 2026. arXiv:2511.07117v2 [math.OC]. Available at `https://arxiv.org/abs/2511.07117`; accessed July 5, 2026.

[31] D. Drusvyatskiy and A. S. Lewis. Tilt stability, uniform quadratic growth, and strong metric regularity of the subdifferential. *SIAM Journal on Optimization*, 23(1):256–267, 2013.

[32] D. Drusvyatskiy and A. S. Lewis. Error bounds, quadratic growth, and linear convergence of proximal methods. *Mathematics of Operations Research*, 43(3):919–948, 2018.

[33] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Series in Operations Research and Financial Engineering. Springer, New York, NY, 1 edition, 2003.

[34] D. Fernández and M. V. Solodov. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. *SIAM Journal on Optimization*, 22(2):384–407, 2012.

[35] D. Fernández and M. Solodov. Stabilized sequential quadratic programming: a survey. *Pesquisa Operacional*, 34(3):463–479, 2014.

[36] H. Gfrerer and B. S. Mordukhovich. Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions. *SIAM Journal on Optimization*, 25(4):2081–2119, 2015.

[37] M. A. Gomes-Ruggiero, J. M. Martínez, and S. A. Santos. Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints. *SIAM Journal on Scientific Computing*, 31(3):1628–1652, 2009.

[38] M. Guignard. Generalized Kuhn–Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 7(2):232–241, 1969.

[39] G. Haeser, O. Hinder, and Y. Ye. On the behavior of Lagrange multipliers in convex and nonconvex infeasible interior point methods. *Mathematical Programming*, 186(1):257–288, Mar. 2021.

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

[41] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49(3):263–265, 1952.

[42] Q. Jin. Error bound conditions and linear convergence. In *Lectures on Nonsmooth Optimization*, volume 82 of *Texts in Applied Mathematics*, pages 289–353. Springer, Cham, 2025.

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

[44] C. Kanzow and D. Steck. On error bounds and multiplier methods for variational problems in Banach spaces. *SIAM Journal on Control and Optimization*, 56(3):1716–1738, 2018.

[45] H. Karimi, J. Nutini, and M. Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition. In P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken, editors, *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016*, volume 9851 of *Lecture Notes in Computer Science*, pages 795–811, Cham, 2016. Springer International Publishing.

[46] A. Y. Kruger. Error bounds and metric subregularity. *Optimization*, 64(1):49–79, 2015.

[47] A. S. Lewis and J.-S. Pang. Error bounds for convex inequality systems. In J.-P. Crouzeix, J.-E. Martinez-Legaz, and M. Volle, editors, *Generalized Convexity, Generalized Monotonicity: Recent Results*, pages 75–110. Springer US, Boston, MA, 1998.

[48] G. Li and T. K. Pong. Calculus of the exponent of Kurdyka–Łojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations of Computational Mathematics*, 18(5):1199–1232, 2018.

[49] W. Li. Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. *SIAM Journal on Optimization*, 7(4):966–978, 1997.

[50] L. Liu, M. B. Majka, and Ł. Szpruch. Polyak–Łojasiewicz inequality on the space of measures and convergence of mean-field birth-death processes. *Applied Mathematics & Optimization*, 87(3):48, 2023.

[51] S. Lojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. In *Les Équations aux Dérivées Partielles*, pages 87–89. Éditions du Centre National de la Recherche Scientifique, Paris, 1963.

[52] Z.-Q. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals of Operations Research*, 46(1):157–178, 1993.

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

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

[55] L. Minchenko. Note on Mangasarian–Fromovitz-like constraint qualifications. *Journal of Optimization Theory and Applications*, 182(3):1199–1204, 2019.

[56] L. Minchenko and A. Tarakanov. On error bounds for quasinormal programs. *Journal of Optimization Theory and Applications*, 148(3):571–579, 2011.

[57] L. I. Minchenko and S. I. Sirotko. On local error bound in nonlinear programs. In N. N. Olenev, Y. G. Evtushenko, M. Jaćimović, M. Khachay, and V. Malkova, editors, *Optimization and Applications*, pages 38–49, Cham, 2021. Springer International Publishing.

[58] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation I*, volume 330 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1st edition, 2006.

[59] I. Necoara, Y. Nesterov, and F. Glineur. Linear convergence of first order methods for non-strongly convex optimization. *Mathematical Programming*, 175(1):69–107, May 2019.

[60] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79(1):299–332, Oct 1997.

[61] B. T. Polyak. Gradient methods for minimizing functionals. *Soviet Mathematics Doklady*, 4:1436–1439, 1963.

[62] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic Press, New York, 1969.

[63] R. W. Prado, S. A. Santos, and L. E. A. Simões. On the fulfillment of the complementary approximate Karush–Kuhn–Tucker conditions and algorithmic applications. *Journal of Optimization Theory and Applications*, Mar 2023.

[64] Q. Rebjock and N. Boumal. Fast convergence to non-isolated minima: Four equivalent conditions for $C^2$ functions. *Mathematical Programming*, 213:151–199, 2025.

[65] R. T. Rockafellar. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical Programming*, 5(1):354–373, Dec 1973.

[66] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2):97–116, 1976.

[67] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer Science & Business Media, 1 edition, 2009.

[68] Z. Wu and J. J. Ye. Sufficient conditions for error bounds. *SIAM Journal on Optimization*, 12(2):421–435, Jan. 2002.