A relaxed quasinormality condition and the boundedness of dual augmented Lagrangian sequences*

R. Andreani[†] G. Haeser[‡] M. L. Schuverdt[§] L. D. Secchin[¶]

January 5, 2024

Abstract

Global convergence of augmented Lagrangian methods to a first-order stationary point is well-known to hold under considerably weak constraint qualifications. In particular, several constant rank-type conditions have been introduced for this purpose which turned out to be relevant also beyond this scope. In this paper we show that in fact under these conditions the sequence of approximate Lagrange multipliers generated by the algorithm remains bounded. This important stability property is associated with both the practical effectiveness of the algorithm and also its computational complexity. In order to obtain this result we introduce a relaxed version of the quasinormality constraint qualification which adequately treats equality constraints by means of informative Lagrange multipliers, a topic that has been extensively studied.

1 Introduction

In this paper we are interested in the general smooth nonlinear programming problem with equality and inequality constraints. More specifically, we are interested in the properties of implementable algorithms for solving the problem. Most algorithms are primal-dual, in the sense that they build a sequence of primal iterates that hopefully converges to a solution, but they also build approximations of Lagrange multipliers (dual solutions) to help guiding the algorithm towards a solution. These sequences play different roles in the analysis as, for instance, boundedness of the primal iterates may be guaranteed by adding large enough box constraints to the problem, while the dual solutions may be unbounded.

The most well known approach for bounding the dual sequence generated by an algorithm is assuming the Mangasarian-Fromovitz constraint qualification (MFCQ) at the point of interest. This is equivalent to saying that the set of Lagrange multipliers at the point is bounded. However, this may be considered too stringent for practical purposes as, for instance, it does not allow redundancies in the problem formulation; MFCQ always fails when an equality constraint is replaced by two inequalities or when an equality constraint appears twice in the problem formulation. Of course, these situations may sometimes be prevented by pre-processing the problem, but this may not be possible or it may be very time consuming, especially when the optimization process appears in the middle of a more complicated application.

Thus, we are interested in bounding the dual sequence generated by the algorithm even when the primal sequence is converging to a point that fails to satisfy MFCQ. For instance, it has been

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

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

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

[§]CONICET, Department of Mathematics, FCE, University of La Plata, La Plata, Bs. As., Argentina. Email: schuverd@mate.unlp.edu.ar

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

shown in [19] that the popular interior point method IPOPT tends to find an unbounded dual sequence when MFCQ fails, hindering its practical performance, in contrast with other interior point methods [21].

In the context of augmented Lagrangian methods, several constraint qualifications have been used to show that the primal iterate converges to a stationary point [15]; that is, even though the dual sequence may be unbounded, an approximate KKT point is achieved. Boundedness of the dual sequence has been shown only recently in [3, 17] under the so-called quasinormality constraint qualification [20], which is weaker than MFCQ, the constant rank constraint qualification (CRCQ [22]), and the constant positive linear dependence condition (CPLD [28]), which were used in the original global convergence analysis of the popular augmented Lagrangian method ALGENCAN [1]. See [9]. However, global convergence to a stationary point is known to hold under considerably weaker conditions. In [26], it has been shown that equality constraints should be treated differently in the formulation of CRCQ, giving rise to a relaxed variant of CRCQ (RCRCQ) which has essentially the same properties of the original formulation. This approach has been exploited in the definition of a relaxed variant of CPLD (RCPLD [7]), which later gave rise to the so-called constant rank of the subspace component constraint qualification (CRSC [8]), the weakest of the constant rank-type constraint qualifications. Besides global convergence of algorithms, several applications and extensions of these conditions have been discussed in the literature, for instance, concerning second-order necessary optimality conditions and a facial reduction procedure for removing redundancies in the problem formulation. We refer to [4] and the references therein for a thorough discussion on this topic, specifically on the central role played by CRSC in this context.

Condition CRSC and the relaxed variants of CRCQ and CPLD are not related with the quasinormality constraint qualification, thus boundedness of the dual sequence is not known under these conditions. The purpose of this paper is to bridge the gap in terms of the global convergence of the augmented Lagrangian method to a stationary point and the boundedness of the dual sequence. In particular, we will define a relaxed variant of the quasinormality constraint qualification that is implied by CRSC and that still guarantees boundedness of the dual augmented Lagrangian sequences. The definition is inspired by the well-known notion of an informative Lagrange multiplier [13] and the relaxed variants of CRCQ and CPLD.

The quasinormality condition, introduced by Hestenes [20] and further studied and adapted to different contexts [11, 12, 18, 23, 25, 30], has found several other applications such as in exact penalty [13], computation of the value function [29], and error bound properties [27, 29]. Although these applications are out of the scope of this paper, we believe that our version of quasinormality adequately treats equality constraints in a similar fashion as the relaxed variants of CRCQ and CPLD, preserving its properties, so that we expect these applications to be extended under relaxed quasinormality.

This paper is organized as follows: Section 2 introduces the preliminary results and definitions. In Section 3 we present the two main proofs that relaxed quasinormality implies CRSC and boundedness of dual augmented Lagrangian sequences. In Section 4 we present some more refined results in terms of feasibility of the primal sequence and a more general algorithm with a scaled criterion for solving the augmented Lagrangian subproblems. Section 5 presents some more refined comparisons with other constraint qualifications and Section 6 presents some concluding remarks.

Notation We use $\mathbb{R}_+$ to denote the set of nonnegative real numbers. Given $z \in \mathbb{R}^r$, $z_+ \in \mathbb{R}^r_+$ is the vector whose i-th coordinate is $\max\{0, z_i\}, i = 1, \ldots, r$. Given a function $q\colon \mathbb{R}^n \to \mathbb{R}^r$, $\nabla q(x)$ is the $n \times r$ matrix whose columns are the gradients $\nabla q_i(x)$, $i = 1, \ldots, r$ at a point $x \in \mathbb{R}^n$ (transposed Jacobian). We use $\|\cdot\|_2$ and $\|\cdot\|_\infty$ to denote the Euclidean norm and the sup-norm, respectively. When $g\colon \mathbb{R}^n \to \mathbb{R}^p$ and $x \in \mathbb{R}^n$ is such that $g(x) \leq 0$, the set $\mathcal{A}(x) = \{j \in \{1, \ldots, p\} \mid g_j(x) = 0\}$ is the set of indices of active inequality constraints at x. Given real sequences $\{a_k\}$ and $\{b_k\}$, $a_k = o(b_k)$ means that there is a sequence $\{m_k\} \subset \mathbb{R}$, $m_k > 0$, converging to zero such that $|a_k| \leq m_k |b_k|$ for all k. Given a tuple (an ordered finite set) $\mathcal{J} = (j_1, \ldots, j_\ell) \subseteq \{1, \ldots, r\}$ and $z \in \mathbb{R}^r$, we define $z_{\mathcal{J}} = (z_{j_1}, \ldots, z_{j_\ell}) \in \mathbb{R}^\ell$. The number of elements in a tuple $\mathcal{J}$ is denoted by $|\mathcal{J}|$.

2 Preliminaries

We consider the nonlinear programming problem

$$\min f(x) \quad \text{s.t.} \quad h(x) = 0, \quad g(x) \leq 0, \tag{P}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$, $h\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. We denote by

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x)$$

the Lagrangian function associated with (P). A constraint qualification is any condition on the description of the feasible set of (P) such that whenever $\bar{x} \in \mathbb{R}^n$ is a local minimizer of (P), there exist so-called Lagrange multipliers $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that $\nabla L(x, \lambda, \mu) = 0$ with $\mu_i = 0$ for all $i \notin \mathcal{A}(\bar{x})$, where the derivative is taken with respect to x. In other words, it must be the case that $-\nabla f(\bar{x}) \in \mathcal{K}(\bar{x}; \bar{x})$ where

$$\mathcal{K}(x; \bar{x}) = \left\{ \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j \in \mathcal{A}(\bar{x})} \mu_j \nabla g_j(x) \;\middle|\; \mu_j \geq 0 \text{ for all } j \in \mathcal{A}(\bar{x}) \right\}.$$

When $x = \bar{x}$, we may write $\mathcal{K}(\bar{x}; \bar{x}) = \mathcal{K}(\bar{x})$, which is the polar of the linearized tangent cone at $\bar{x}$. It is easy to see that the set of Lagrange multipliers at $\bar{x}$ is bounded if, and only if, the gradients of equalities and active inequalities are *positively linearly independent* (that is, MFCQ holds):

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) + \sum_{j \in \mathcal{A}(\bar{x})} \mu_j \nabla g_j(\bar{x}) = 0, \; \mu \geq 0 \;\text{ implies }\; \lambda = 0, \; \mu_j = 0, \forall j \in \mathcal{A}(\bar{x}).$$

In this paper we consider the (safeguarded) augmented Lagrangian method described in [2, 15] whose implementation is known as Algencan[1]. Given the *penalty parameter* $\rho > 0$ and the *projected multipliers* $\bar{\lambda} \in \mathbb{R}^m$, $\bar{\mu} \in \mathbb{R}^p_+$, let us consider the Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian function

$$L_{\rho, \bar{\lambda}, \bar{\mu}}(x) = f(x) + \frac{\rho}{2} \left[\left\| h(x) + \frac{\bar{\lambda}}{\rho} \right\|_2^2 + \left\| \left(g(x) + \frac{\bar{\mu}}{\rho} \right)_+ \right\|_2^2 \right].$$

In Algencan, the iterate x^k is obtained by minimizing $L_{\rho_k, \bar{\lambda}^k, \bar{\mu}^k}(x)$ for fixed ρ_k, $\bar{\lambda}^k$ and $\bar{\mu}^k$. The projected multipliers sequences $\{\bar{\lambda}^k\}$ and $\{\bar{\mu}^k\}$ are computed within a pre-defined box (*safeguards*).

The dual (approximate Lagrange multipliers) sequences generated by Algorithm 1 are defined as

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k) \quad \text{and} \quad \mu^k = [\bar{\mu}^k + \rho_k g(x^k)]_+, k \geq 1. \tag{1}$$

The global convergence of Algorithm 1 was established and improved over several works. Let us recall here one of the main conditions used in this analysis.

Definition 1 (CRSC [8])**.** *A feasible point $\bar{x}$ for* (P) *satisfies the* constant rank of the subspace component condition (CRSC) *if the rank of the gradients*

$$\nabla h_i(x), \; i = 1, \ldots, m, \qquad \nabla g_j(x), \; j \in \mathcal{A}_-(\bar{x}),$$

remains constant for all x in a neighbourhood of $\bar{x}$, where

$$\mathcal{A}_-(\bar{x}) = \{ j \in \mathcal{A}(\bar{x}) \mid -\nabla g_j(\bar{x}) \in \mathcal{K}(\bar{x}) \}. \tag{2}$$

[1] freely available at `www.ime.usp.br/~egbirgin/tango`

Algorithm 1 Safeguarded augmented Lagrangian method

The parameters are $\tau \in [0,1)$, $\gamma > 1$, $-\infty < \lambda_{\min} \leq \lambda_{\max} < \infty$, $0 \leq \mu_{\max} < \infty$ and $\rho_1 > 0$. Let $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^1 \in [0, \mu_{\max}]^p$ and a sequence $\{\varepsilon_k\} \subset \mathbb{R}_+$ such that $\lim_{k\to\infty} \varepsilon_k = 0$. Initialize $k \leftarrow 1$.

Step 1 (Solving the subproblem): Compute an approximate stationary point x^k of $L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x)$, that is, x^k satisfying $\|\nabla L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x^k)\|_\infty \leq \varepsilon_k$.

Step 2 (Updating the penalty parameter): Compute

$$V^k = \min\left\{-g(x^k), \frac{\bar{\mu}^k}{\rho_k}\right\}.$$

If $k = 1$ or $\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\}$, set $\rho_{k+1} = \rho_k$. Otherwise, take $\rho_{k+1} \geq \gamma\rho_k$.

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

Step 4: Set $k \leftarrow k+1$ and go to Step 1.

This condition improves several others [7, 22, 26, 28] by considering a single set of constraints to have the constant rank property. Notice that when MFCQ holds, CRSC also holds with $\mathcal{A}_-(\bar{x}) = \emptyset$. Several applications of this condition have been found and we refer the interested reader to [4] and the references therein. In particular, under CRSC, the constraints indexed in the set $\mathcal{A}_-(\bar{x})$ behave locally as equality constraints in the description of the feasible set; this procedure is known as *facial reduction* [16] in the context of conic programming. Now let $\bar{x}$ be a limit point of a sequence $\{x^k\}$ generated by Algorithm 1. If $\bar{x}$ is feasible and satisfies CRSC, then it was shown in [8] that $\bar{x}$ is a stationary point of (P), however no information has been provided with respect to the dual sequences (1). In order to provide such information, one relies on the quasinormality constraint qualification, which is also weaker than MFCQ but it is independent of CRSC. The definition is as follows:

Definition 2 (Quasinormality [20])**.** *A feasible point $\bar{x}$ of* (P) *satisfies the* quasinormality (QN) *constraint qualification if there is no $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that*

1. $\nabla h(\bar{x})\lambda + \nabla g(\bar{x})\mu = 0$*;*
2. $(\lambda, \mu) \neq 0$*;*
3. *Defining the index sets*

$$I_{\neq} = \{i \mid \lambda_i \neq 0\} \quad \textit{and} \quad J_+ = \{j \mid \mu_j > 0\}, \tag{3}$$

there is a sequence $\{x^k\}$ converging to $\bar{x}$ such that, for all k, $\lambda_i h_i(x^k) > 0$, $\forall i \in I_{\neq}$, and $g_j(x^k) > 0$, $\forall j \in J_+$.

We start by recalling the fact that CRSC and QN are independent conditions.

Example 1. *Consider the constraint set defined by $h_1(x) = x$ and $h_2(x) = x^2$ at $\bar{x} = 0 \in \mathbb{R}$. QN does not hold since we can take $\lambda_1 = 0$ and $\lambda_2 = 1$ together with the sequence $x^k = 1/k$ where we have that $\lambda_1\nabla h_1(\bar{x}) + \lambda_2\nabla h_2(\bar{x}) = 0$ with $\lambda_2 h_2(x^k) > 0$ for all k. On the other hand, CRSC holds since the set $\{\nabla h_1(x), \nabla h_2(x)\}$ has full (constant) rank for all x nearby $\bar{x}$. The reverse situation occurs with the constraint set defined by $g_1(x) = -x_1$ and $g_2(x) = x_1 - x_2^2$ at $\bar{x} = (0,0) \in \mathbb{R}^2$. The set $\mathcal{K}(\bar{x})$ is equal to $\mathbb{R} \times \{0\}$ where $\mathcal{A}_-(\bar{x}) = \{1,2\}$ but the rank of $\{\nabla g_1(x), \nabla g_2(x)\}$ increases from 1 at $\bar{x}$ to 2 for x nearby $\bar{x}$ with $x_2 \neq 0$. Thus CRSC fails. In order the see that QN holds notice that $\mu_1\nabla g_1(\bar{x}) + \mu_2\nabla g_2(\bar{x}) = 0$ with $0 \neq \mu \geq 0$ implies that $\mu_1 = \mu_2 > 0$. However if $g_1(x) > 0$ for some x, it must be the case that $g_2(x) < 0$. Thus no sequence satisfying item 3. in Definition 2 exists and QN holds.*

Under quasinormality, it was proved in [3, 19] that if $\bar{x}$ is a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 1, that is, $\lim_{k\in K} x^k = \bar{x}$ for an infinite set of indexes K, then the sequences of approximate Lagrange multipliers $\{\lambda^k\}_{k\in K}$ and $\{\mu^k\}_{k\in K}$ as defined in (1) are bounded. In fact, by Step 1 of the algorithm, a simple computation gives $\nabla L(x^k, \lambda^k, \mu^k) \to 0$. Thus, assuming that the dual sequences are not both bounded one arrives at a pair $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that items 1. and 2. of Definition 2 are satisfied. Now, by (1) it is easy to see that when $\lambda_i^k \to +\infty$ then, since $\{\bar{\lambda}^k\}$ is bounded, it must be the case that $\rho_k \to +\infty$ and $h_i(x^k) > 0$ for sufficiently large k. A similar analysis holds when $\lambda_i^k \to -\infty$ or $\mu_j^k \to +\infty$ so that the sign condition given by item 3. is also satisfied. Therefore, under quasinormality, no such sequence exists and the dual sequences must be bounded.

An alternative motivation for the definition of quasinormality [20] comes from an enhanced Fritz-John theorem, where it is shown that around a local minimizer there exists a sequence that violates the constraints in a particular way. This inspired several different definitions of a Lagrange multiplier with additional requirements concerning constraint violation. The most general of these results is the following:

Theorem 1 ([13, Proposition 2.1])**.** *Let $\bar{x}$ be a local minimizer of* (P)*. Then there is $(\sigma, \lambda, \mu) \in \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}^p_+$ such that*

1. *$\sigma\nabla f(\bar{x}) + \nabla h(\bar{x})\lambda + \nabla g(\bar{x})\mu = 0$;*
2. *$(\sigma, \lambda, \mu) \neq 0$;*
3. *if $I_{\neq} \cup J_+ \neq \emptyset$, where $I_{\neq}$ and J_+ are defined in* (3)*, then there is a sequence $\{x^k\}$ converging to $\bar{x}$ such that, for all k,*
   (a) *$\lambda_i h_i(x^k) > 0$, $\forall i \in I_{\neq}$, and $g_j(x^k) > 0$, $\forall j \in J_+$;*
   (b) *$|h_i(x^k)| = o(w(x^k))$, $\forall i \notin I_{\neq}$ and $g_j(x^k)_+ = o(w(x^k))$, $\forall j \notin J_+$, where*

$$w(x^k) = \min\left\{\min_{i\in I_{\neq}} |h_i(x^k)|, \quad \min_{j\in J_+} g_j(x^k)_+\right\}. \tag{4}$$

It is easy to see that item 3(a) of the above theorem implies the usual complementary slackness $\mu_j g_j(\bar{x}) = 0$ for all $j = 1, \ldots, p$. Notice that Theorem 1 implies that QN is a constraint qualification, since Definition 2 prevents the existence of a sequence satisfying items 1, 2 and 3(a) of Theorem 1 when $\sigma = 0$. Thus, at a local minimizer, QN implies that there exists a Lagrange multiplier (λ, μ) with the additional constraint violation given by items 3(a-b) of Theorem 1. This has been called an *informative* Lagrange multiplier in [13]. It was shown in [10] that this additional dual information is not relevant for distinguishing a primal solution, that is, a feasible point for (P) admits an informative Lagrange multiplier if, and only if, it admits a standard Lagrange multiplier. However, this additional dual information will be crucial in our analysis. We start by noticing that it is clear that Theorem 1 suggests a weaker definition of QN by incorporating also item 3(b) as follows:

Definition 3. *A feasible point $\bar{x}$ for* (P) *satisfies the* relaxed quasinormality *(RQN) condition if there is no $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that*

1. *$\nabla h(\bar{x})\lambda + \nabla g(\bar{x})\mu = 0$;*
2. *$(\lambda, \mu) \neq 0$;*
3. *There is a sequence $\{x^k\}$ converging to $\bar{x}$ such that, for all k,*
   (a) *$\lambda_i h_i(x^k) > 0$, $\forall i \in I_{\neq}$, and $g_j(x^k) > 0$, $\forall j \in J_+$;*
   (b) *$|h_i(x^k)| = o(w(x^k))$, $\forall i \notin I_{\neq}$ and $g_j(x^k)_+ = o(w(x^k))$, $\forall j \notin J_+$, where $I_{\neq}$ and J_+ are defined as in* (3) *and $w(x^k)$ as in* (4)*.*

This definition has not been exploited yet in the literature and it will be the main focus in this paper. It is clearly a constraint qualification since it implies that it must be the case that $\sigma > 0$ in Theorem 1. Recalling Example 1 with the constraint set defined by $h_1(x) = x$ and $h_2(x) = x^2$

at $\bar{x} = 0 \in \mathbb{R}$, notice that the sequence $x^k = 1/k$ with $\lambda_1 = 0$ and $\lambda_2 = 1$ fails to comply with item 3(b) in Definition 3. Indeed, it is not the case that $|h_1(x^k)| = o(|h_2(x^k)|)$, that is, $\frac{|x^k|}{(x^k)^2} \not\to 0$. In fact, this is the case for any sequence $x^k \to \bar{x}, x^k \neq \bar{x}$ and since items 1-2 of Definition 3 implies that $\lambda_1 = 0$ and $\lambda_2 \neq 0$, this shows that RQN holds.

It turns out that RQN will provide an adequate way of dealing with equality constraints, in a similar way as it is done in the relaxed variants of CRCQ and CPLD. Namely, while QN implies CRCQ and CPLD, it does not imply the weaker variants RCRCQ, RCPLD and CRSC. We will show that CRSC (and all other constant rank-type constraint qualifications) strictly implies RQN. This will give rise to a new stability property under constant rank-type constraint qualifications as we will show that RQN will be enough for providing boundedness of the dual augmented Lagrangian sequences.

3 Main results

Our first main result concerning RQN is the fact that it subsumes all constant-rank type CQs. That is, we will show that CRSC implies RQN. Clearly, the implication is strict due to the second constraint set defined in Example 1 where CRSC fails and QN (thus RQN) holds. We will make use of the following lemma, which is a consequence of the inverse function theorem:

Lemma 1 ([9, Lemma 3.2])**.** *Let $\bar{x} \in \mathbb{R}^n$, $\mathcal{V} \subseteq \mathbb{R}^n$ an open neighbourhood of $\bar{x}$ and $c \colon \mathcal{V} \to \mathbb{R}^r$ a C^1 function. Suppose that the tuple $\mathcal{J} \subseteq \{1, \ldots, r\}$ is such that $\{\nabla c_j(\bar{x})\}_{j \in \mathcal{J}}$ is linearly independent and that $F \colon \mathcal{V} \to \mathbb{R}$ satisfies*

$$\nabla F(x) = \sum_{j \in \mathcal{J}} \alpha_j \nabla c_j(x)$$

for all $x \in \mathcal{V}$. Then there exists an open neighbourhood $\mathcal{U} \subseteq \mathbb{R}^{|\mathcal{J}|}$ of $c_{\mathcal{J}}(\bar{x})$ and a C^1 function $\varphi \colon \mathcal{U} \to \mathbb{R}$ such that $c_{\mathcal{J}}(x) \in \mathcal{U}$ and $F(x) = \varphi(c_{\mathcal{J}}(x))$ for all $x \in \mathcal{V}$, and

$$\alpha_j = \left[\nabla \varphi(c_{\mathcal{J}}(\bar{x}))\right]_j, \quad j = 1, \ldots, |\mathcal{J}|.$$

Theorem 2. *CRSC implies RQN.*

Proof. Suppose by contradiction that $\bar{x}$ satisfies CRSC but not RQN, and take $(\lambda, \mu) \neq 0$ and $\{x^k\}$ satisfying items 1 and 3 of Definition 3. First, we affirm that $\mu_j = 0$ for all $j \notin \mathcal{A}_-(\bar{x})$, or equivalently, $J_+ \subseteq \mathcal{A}_-(\bar{x})$. In fact, if $\mu_j > 0$ then $\nabla h(\bar{x})\lambda + \nabla g(\bar{x})\mu = 0$ implies

$$-\nabla g_j(\bar{x}) = \frac{1}{\mu_j}\left[\sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) + \sum_{\ell \neq j} \mu_\ell \nabla g_\ell(\bar{x})\right] \in \mathcal{K}(\bar{x}),$$

which in turn implies $j \in \mathcal{A}_-(\bar{x})$. Hence, item 1 of Definition 3 takes the form

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) + \sum_{j \in \mathcal{A}_-(\bar{x})} \mu_j \nabla g_j(\bar{x}) = 0.$$

Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ and $\mathcal{J} \subseteq \mathcal{A}_-(\bar{x})$ be such that $\{\nabla h_i(\bar{x})\}_{i \in \mathcal{I}} \cup \{\nabla g_j(\bar{x})\}_{j \in \mathcal{J}}$ is a basis for span $\{\nabla h(\bar{x}), \nabla g_{\mathcal{A}_-(\bar{x})}(\bar{x})\}$, the subspace generated by $\{\nabla h_i(\bar{x})\}_{i=1}^m \cup \{\nabla g_j(\bar{x})\}_{j \in \mathcal{A}_-(\bar{x})}$. The CRSC condition guarantees that

$$\operatorname{span}\{\nabla h_{\mathcal{I}}(x), \nabla g_{\mathcal{J}}(x)\} = \operatorname{span}\{\nabla h(x), \nabla g_{\mathcal{A}_-(\bar{x})}(x)\} \quad \text{for all } x \text{ close to } \bar{x}. \tag{5}$$

Now, let us define

$$F(x) = -\sum_{i \in \{1 \ldots, m\} \setminus \mathcal{I}} \lambda_i h_i(x) - \sum_{j \in \mathcal{A}_-(\bar{x}) \setminus \mathcal{J}} \mu_j g_j(x). \tag{6}$$

From (5), $\nabla F(x) \in \operatorname{span}\{\nabla h_{\mathcal{I}}(x), \nabla g_{\mathcal{J}}(x)\}$ for all x near $\bar{x}$. So, we can apply Lemma 1 to obtain

$$F(x) = \varphi(h_{\mathcal{I}}(x), g_{\mathcal{J}}(x)), \ \forall x \text{ near } \bar{x}, \quad \text{and} \quad (\lambda, \mu)_i = \nabla \varphi(h_{\mathcal{I}}(\bar{x}), g_{\mathcal{J}}(\bar{x}))_i, \ \forall i. \tag{7}$$

Note that $(h_{\mathcal{I}}(\bar{x}), g_{\mathcal{J}}(\bar{x})) = 0$, $\varphi(0) = F(\bar{x}) = 0$ and then the Taylor expansion of φ around the origin gives

$$\varphi(z) = \nabla\varphi(h_{\mathcal{I}}(\bar{x}), g_{\mathcal{J}}(\bar{x}))^T z + o(\|z\|_\infty).$$

Considering (6), (7), the fact that $\mu_j = 0$ for $j \notin \mathcal{A}_-(\bar{x})$, that $g_j(x^k) > 0$ when $j \in J_+$ by item 3(a) of Definition 3 and the above expression with $z = (h_{\mathcal{I}}(x^k), g_{\mathcal{J}}(x^k)_+)$, we have, for all k large enough,

$$\begin{aligned}\sum_{i=1}^m \lambda_i h_i(x^k) + \sum_{j\in\mathcal{J}} \mu_j g_j(x^k)_+ + \sum_{j\in\mathcal{A}_-(\bar{x})\setminus\mathcal{J}} \mu_j g_j(x^k) + w_k \\ = \sum_{i\in I_{\neq}} \lambda_i h_i(x^k) + \sum_{j\in J_+} \mu_j g_j(x^k)_+ + w_k = 0,\end{aligned} \tag{8}$$

where $I_{\neq}$, J_+ are as in (3) and $w_k = o(\|(h_{\mathcal{I}}(x^k), g_{\mathcal{J}}(x^k)_+)\|_\infty)$. Clearly,

$$w_k = o(\|(\ h(x^k)\ ,\ g_{\mathcal{J}}(x^k)_+\ ,\ g_{\mathcal{A}_-(\bar{x})\setminus\mathcal{J}}(x^k)\ )\|_\infty),$$

and by item 3 of Definition 3, we also have

$$w_k = o(\|(\ h_{I_{\neq}}(x^k)\ ,\ g_{J_+}(x^k)\ )\|_\infty).$$

Thus, dividing (8) by $\|(h_{I_{\neq}}(x^k), g_{J_+}(x^k))\|_\infty$ and passing to the limit on a subsequence if necessary, we obtain

$$(h^*, g^*) = \lim_k \frac{(\ h_{I_{\neq}}(x^k)\ ,\ g_{J_+}(x^k)_+)}{\|(h_{I_{\neq}}(x^k), g_{J_+}(x^k))\|_\infty} \neq 0$$

satisfying

$$\sum_{i\in I_{\neq}} \lambda_i h_i^* + \sum_{j\in J_+} \mu_j g_j^* = 0,$$

with either $h^* \neq 0$ or $g^* \neq 0$, and $\lambda_i h_i^* \geq 0$, $i \in I_{\neq}$, $\mu_i g_i^* \geq 0$, $j \in J_+$, which is impossible. We then conclude that RQN holds at $\bar{x}$. $\square$

We now show that, similar to what is known about QN, the dual sequences generated by Algorithm 1 are bounded under RQN.

Theorem 3. *Let $\{x^k\}$ be a sequence generated by Algorithm 1 and $\bar{x}$ a feasible limit point of it, let us say, $\lim_{k\in K} x^k = \bar{x}$. If $\bar{x}$ satisfies RQN, then the associated dual subsequences $\{\lambda^k\}_{k\in K}$ and $\{\mu^k\}_{k\in K}$ given in* (1) *are bounded. In particular, all limit points of these sequences are Lagrange multipliers associated with $\bar{x}$.*

Proof. Let $\{\rho_k\}$, $\{\bar{\lambda}^k\}$, and $\{\bar{\mu}^k\}$ be corresponding sequences produced by Algorithm 1 and suppose that $\{M_k := \|(1, \lambda^k, \mu^k)\|_\infty\}$ is unbounded. By (1), we have $\rho_k \to \infty$ and then $\mu_i^k = 0$ for all $i \notin \mathcal{A}(\bar{x})$. So, dividing the expression $\nabla L(x^k, \lambda^k, \mu^k) \to 0$ as provided by Step 1 by M_k and taking the limit in an appropriate subsequence, let us say with indices in K, we arrive at

$$\sum_{i=1}^m \lambda_i \nabla h_i(\bar{x}) + \sum_{i\in\mathcal{A}(\bar{x})} \mu_i \nabla g_i(\bar{x}) = 0,$$

with $(\lambda, \mu) \neq 0$ and

$$\lambda_i = \lim_{k\in K} \frac{\bar{\lambda}_i^k + \rho_k h_i(x^k)}{M_k}, \quad \mu_j = \lim_{k\in K} \frac{\left[\bar{\mu}_j^k + \rho_k g_j(x^k)\right]_+}{M_k} \tag{9}$$

for all i, j. If $\lambda_i \neq 0$, then we can extract a subsequence so that $h_i(x^k)$ always has the same sign of λ_i (the same is valid for μ_j). Thus, passing to a subsequence if necessary, we can suppose without loss of generality that,

$$\text{for all } k \in K, \quad \lambda_i h_i(x^k) > 0 \ \text{ if } \ \lambda_i \neq 0 \quad \text{and} \quad g_j(x^k) > 0 \ \text{ if } \ \mu_j > 0. \tag{10}$$

Therefore, if all entries of (λ, μ) are non-null, then item 3(b) of Definition 3 holds trivially with $\{x^k\}_{k\in K}$, contradicting the validity of RQN at $\bar{x}$.

Now, suppose that $\lambda_i \neq 0$ and $\lambda_\ell = 0$. Clearly $\lim_{k\in K} |\rho_k h_i(x^k)| = \infty$ and, by (10), $h_i(x^k) \neq 0$ for all $k \in K$. For each $k \in K$, let us define

$$A_k := \frac{|\bar{\lambda}_\ell^k + \rho_k h_\ell(x^k)|}{|\rho_k h_i(x^k)|} = \left| \frac{\bar{\lambda}_\ell^k}{\rho_k h_i(x^k)} + \frac{h_\ell(x^k)}{h_i(x^k)} \right|.$$

We affirm that $\liminf_{k\in K} A_k = 0$. In fact, if $A_k \geq \varepsilon > 0$ for all $k \in K$ large enough, we would have $|\bar{\lambda}_\ell^k + \rho_k h_\ell(x^k)| \geq \varepsilon |\rho_k h_i(x^k)|$ for all $k \in K$ large enough and therefore

$$0 < \varepsilon |\lambda_i| = \lim_{k\in K} \varepsilon \left| \frac{\bar{\lambda}_i^k + \rho_k h_i(x^k)}{M_k} \right| = \lim_{k\in K} \varepsilon \frac{|\rho_k h_i(x^k)|}{M_k} \leq \lim_{k\in K} \frac{|\bar{\lambda}_\ell^k + \rho_k h_\ell(x^k)|}{M_k} = |\lambda_\ell| = 0,$$

a contradiction. Hence, there is an infinite set of indices $K_1 \subseteq K$ such that

$$\lim_{k\in K_1} \frac{|h_\ell(x^k)|}{|h_i(x^k)|} = \lim_{k\in K_1} A_k = 0.$$

A similar argument is valid changing $\lambda_i \neq 0$ to $\mu_j > 0$, $\lambda_\ell = 0$ to $\mu_\ell = 0$ and/or $|h_\ell(x^k)|$ to $g_\ell(x^k)_+$. Thus, applying it successively we obtain an infinite set $K_* \subseteq \cdots \subseteq K_1 \subseteq K$ such that

$$\lim_{k\in K_*} \frac{|h_\ell(x^k)|}{w(x^k)} = 0 \text{ if } \lambda_\ell = 0 \quad \text{and} \quad \lim_{k\in K_*} \frac{g_\ell(x^k)_+}{w(x^k)} = 0 \text{ if } \mu_\ell = 0, \tag{11}$$

where $w(x^k)$ is as in (4). Finally, from (10) and (11) we conclude that RQN at $\bar{x}$ is violated using the sequence $\{x^k\}_{k\in K_*}$, and the proof is complete. $\square$

Theorem 3 was known only under QN [3, 19]. The following example shows that when RQN fails, the dual sequences of Algorithm 1 may in fact be unbounded.

Example 2. *Consider the problem*

$$\min f(x) = x_1^2 + x_2^2 \quad s.t. \quad x \in \Omega$$

where

$$\Omega = \{x \in \mathbb{R}^2 \mid g_1(x) = x_1^3 - x_2 \leq 0,\ g_2(x) = x_1^3 + x_2 \leq 0,\ g_3(x) = -x_1 \leq 0\},$$

and its feasible point $\bar{x} = (0, 0)$. We affirm that this point can be reached by Algorithm 1 with unbounded multiplier sequences (1). In fact, let us consider the sequence $x^k = (1/\rho_k^a, 0)$, $\rho_k > 0$, where $a \in (1/5, 1/3)$ is a constant. For each k, take any $\bar{\mu}_1^k = \bar{\mu}_2^k \geq 0$ and $\bar{\mu}_3^k = 0$. The multipliers estimates (1) with these sequences are

$$\mu_2^k = \mu_1^k = \bar{\mu}_1^k + \rho_k^{1-3a}, \quad \mu_3^k = 0.$$

We have

$$\nabla L_{\rho_k,\bar{\mu}^k}(x^k) =$$

$$\begin{bmatrix} \frac{2}{\rho_k^a} \\ 0 \end{bmatrix} + (\bar{\mu}_1^k + \rho_k^{1-3a}) \begin{bmatrix} \frac{3}{\rho_k^{2a}} \\ -1 \end{bmatrix} + (\bar{\mu}_1^k + \rho_k^{1-3a}) \begin{bmatrix} \frac{3}{\rho_k^{2a}} \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{\rho_k^a} + \frac{6\bar{\mu}_1^k}{\rho_k^{2a}} + \frac{6}{\rho_k^{5a-1}} \\ 0 \end{bmatrix}$$

and

$$V_2^k = V_1^k = \min\left\{ -\frac{1}{\rho_k^{3a}}, \frac{\bar{\mu}_1^k}{\rho_k} \right\} = -\frac{1}{\rho_k^{3a}}, \quad V_3^k = \min\left\{ \frac{1}{\rho_k^a}, \frac{\bar{\mu}_3^k}{\rho_k} \right\} = 0.$$

If $\{\rho_k\}$ remained constant for all $k \geq k_0$ then we would have $\|V^k\|_\infty = 1/\rho_k^{3a} > \tau/\rho_{k-1}^{3a} = \tau\|V^{k-1}\|_\infty$ for any $\tau \in [0, 1)$, $k \geq k_0$, contradicting Step 2. On the contrary, $\rho_k \to \infty$ implies $\nabla L_{\rho_k,\bar{\mu}^k}(x^k) \to 0$ since $a > 1/5$. Therefore the sequence $(x^k, \rho_k, \bar{\mu}^k)$ with $\rho_k \to \infty$ can be generated by Algorithm 1. In this case $\mu_2^k = \mu_1^k \to \infty$ since $a < 1/3$.

In order to see that RQN fails at $\bar{x}$, just note that $\mu = (1, 1, 0)$ and $x^k = (1/k, 0)$ fulfill items 1, 2 and 3 of Definition 3.

Notice that $\bar{x} = (0, 0)$ in Example 2 is a KKT point that satisfies a weak constraint qualification called *constant positive generators* (CPG), as we will see in Section 5. Actually, we will show that RQN and CPG are independent conditions.

4 Extensions

The result we presented related to the boundedness of the dual augmented Lagrangian sequences (Theorem 3) assumes that the limit point $\bar{x}$ is feasible. This is not a serious drawback since the algorithm tends to find feasible limit points, when they exist, as their limit points are stationary to the problem of minimizing $\|h(x)\|_2^2 + \|g(x)_+\|_2^2$ [1]. These points are feasible when the gradients of equality constraints and violated or active inequality constraints are positively linearly independent, what is known as Extended-MFCQ [24]. However no feasibility result is known under a condition weaker than Extended-MFCQ. Let us show that the boundedness of the dual sequences is enough for ensuring feasibility and this is obtained by an extension of RQN to infeasible points. The definition is exactly the same as Definition 3, however it is simply not assumed that the point is feasible. We opted to present a simpler version of this result in Theorem 3 for clarity of exposition, but in fact this theorem is a particular case of the result we prove next.

Definition 4. *A point $\bar{x} \in \mathbb{R}^n$, not necessarily feasible, satisfies the* Extended-RQN *condition if there is no $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ and sequence $x^k \to \bar{x}$ such that items 1, 2, 3a) of Definition 3 hold and*

b') $|h_i(x^k)| = o(w(x^k))$, $\forall i \notin I_{\neq}$ with $h_i(\bar{x}) = 0$ and $g_j(x^k)_+ = o(w(x^k))$, $\forall j \notin J_+$ with $g_j(\bar{x}) = 0$, where $I_{\neq}$ and J_+ are defined as in (3) *and $w(x^k)$ as in* (4).

Theorem 4. *Let $\bar{x}$ be a limit point generated by Algorithm 1. If $\bar{x}$ satisfies Extended-RQN, then $\bar{x}$ is feasible and the sequences* (1) *are bounded. In particular, all limit points of these sequences are Lagrange multipliers associated with $\bar{x}$.*

Proof. In view of Theorem 3, it is sufficient to show that $\bar{x}$ is feasible. Let $\{\rho_k\}, \{\bar{\lambda}^k\}$, and $\{\bar{\mu}^k\}$ be corresponding sequences produced by Algorithm 1. If $\{\rho_k\}$ is bounded, the test in Step 2 of Algorithm 1 ensures that $\bar{x}$ is feasible. As a consequence, $\{M_k = \|(1, \lambda^k, \mu^k)\|_\infty\}$ is trivially bounded. If $\rho_k \to \infty$ and $\{M_k\}$ is bounded then by (1) and the boundedness of $\{(\bar{\lambda}^k, \bar{\mu}^k)\}$ we conclude that $\bar{x}$ is feasible. From now on, we suppose that $\rho_k \to \infty$ and $\{M_k\}$ is unbounded.

By (1), we have $\mu_i^k = 0$ for all i with $g_i(\bar{x}) < 0$. So, dividing $\nabla L(x^k, \lambda^k, \mu^k) \to 0$ by M_k and taking the limit in a subsequence if necessary, we obtain $(\lambda, \mu) \neq 0$ as in (9) and satisfying item 1 of Definition 4.

If some $\lambda_i = 0$ and $h_i(\bar{x}) \neq 0$, then $\lim_k \rho_k / M_k = 0$. However, as $(\lambda, \mu) \neq 0$, there is an index ℓ such that $\lambda_\ell \neq 0$ or $\mu_\ell > 0$. In both cases, we would arrive at $\lambda_\ell = \lim_k \rho_k h_\ell(x^k)/M_k = 0$ or $\mu_\ell = \lim_k [\bar{\mu}^k + \rho_k g_\ell(x^k)]_+ / M_k = 0$, a contradiction. Therefore, $h_i(\bar{x}) = 0$ whenever $\lambda_i = 0$. Analogously, $g_j(\bar{x}) \leq 0$ whenever $\mu_j = 0$. The case where $\lambda_i \neq 0$ or $\mu_j > 0$ follows similarly to the proof of Theorem 3, as we can conclude that this contradicts the validity of Extended-RQN at $\bar{x}$, independently of the feasibility of $\bar{x}$. This completes the proof. □

Theorems 3 and 4 can also be extended in a different direction, by considering a more general variation of Algorithm 1. Namely, in Step 1, instead of computing x^k such that $\|\nabla L_{\rho_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k)\|_\infty \leq \varepsilon_k$ for a sequence $\varepsilon_k \to 0$, we do so for a sequence ε_k such that $\varepsilon_k = o(M_k)$, where $M_k = \|(1, \lambda^k, \mu^k)\|_\infty$ as defined in (1). A sequence of this type is computed when one aims at achieving a scaled stopping criterion, so that the subproblems may be solved to a less stringent accuracy, improving the efficiency without hindering its robustness. See the discussion and numerical experiments in [6] and [14]. In other words, [14, Theorem 2.1] may be proved under Extended-RQN, which we state below.

Theorem 5. *Let $\bar{x}$ be the limit of a subsequence $\{x^k\}_{k \in K}$ as generated by Algorithm 1 that satisfies Extended-RQN where the subproblem tolerance ε_k is such that $\varepsilon_k = o(M_k)$, where $M_k = \|(1, \lambda^k, \mu^k)\|_\infty$ as defined in* (1). *Then $\bar{x}$ is feasible and $\{M_k\}_{k \in K}$ is bounded. In particular, all limit points of these sequences are Lagrange multipliers associated with $\bar{x}$.*

Proof. The proof is essentially the same as the ones presented previously. In the proof of Theorem 3, notice that the first step is to divide $\|\nabla L(x^k, \lambda^k, \mu^k)\|_\infty \leq \varepsilon_k$ by M_k and use that the right-hand side goes to zero. This remains to be the case under our assumptions. □

5 Relationship between RQN and other known CQs

In this section we analyse the relationship between RQN and other known CQs from the literature besides CRSC and QN. Next, we recall the *AL-regular CQ* (or *AL-regularity* condition), which is associated with the sequences generated by Algorithm 1 [5]. To this end, we consider the function $\mathcal{K}^{\mathrm{AL}}\colon \mathbb{R}^{n+1} \to \mathbb{R}^n$ defined by

$$\mathcal{K}^{\mathrm{AL}}(x, \rho) = \nabla h(x)[\rho h(x)] + \nabla g(x)[\rho g(x)]_+.$$

Let $\bar{x}$ be a feasible point for (P). The upper limit of $\mathcal{K}^{\mathrm{AL}}(x, \rho)$ as $x \to \bar{x}$ and $\rho \to \infty$ is the set

$$\begin{aligned}&\limsup_{x \to \bar{x}, \rho \to \infty} \mathcal{K}^{\mathrm{AL}}(x, \rho) \\ &\qquad = \left\{ \bar{y} \in \mathbb{R}^n \mid \exists \{(x^k, y^k)\} \to (\bar{x}, \bar{y}),\ \exists \{\rho_k\} \to \infty \text{ s.t. } y^k = \mathcal{K}^{\mathrm{AL}}(x^k, \rho_k)\ \forall k \right\}.\end{aligned}$$

We have $\mathcal{K}(\bar{x}) \subseteq \limsup_{x \to \bar{x}, \rho \to \infty} \mathcal{K}^{\mathrm{AL}}(x, \rho)$ [5], but the contrary inclusion is not always true. AL-regularity is exactly that.

Definition 5. *A feasible $\bar{x}$ for* (P) *satisfies the* AL-regularity *condition (or it is an* AL-regular *point) if*

$$\limsup_{x \to \bar{x}, \rho \to \infty} \mathcal{K}^{AL}(x, \rho) \subseteq \mathcal{K}(\bar{x}).$$

The AL-regularity condition is in some sense the weakest possible property that guarantees that any feasible limit point of Algorithm 1 satisfies the KKT conditions.

Theorem 6. *Let $\bar{x}$ be a feasible limit point of a sequence generated by Algorithm 1. If $\bar{x}$ satisfies the AL-regularity condition then $\bar{x}$ satisfies the KKT conditions. Conversely, if for every objective function that attains a constrained local minimum at $\bar{x}$, the KKT conditions are satisfied, then $\bar{x}$ is an AL-regular point.*

Proof. The statement follows from [5, Theorems 1 and 6]. □

The above theorem implies that, indeed, AL-regularity is a constraint qualifications since every local minimizer is the limit point of a sequence generated by Algorithm 1 [5, Theorem 2] (actually, it implies Abadie's constraint qualification [5, Theorem 9]). This fact also gives an alternative proof that RQN is a constraint qualification.

Theorem 7. *RQN implies AL-regularity.*

Proof. The statement follows directly from Theorems 3 and 6, as the boundedness of the multiplier sequence of Algorithm 1 implies the validity of the KKT conditions. □

Another CQ of interest is the constant positive generators (CPG), which we recall next. Let us consider the set

$$\begin{aligned}&\mathcal{K}(x; \bar{x}, \mathcal{I}, \mathcal{J}) = \\ &\quad \left\{ \sum_{i \in \mathcal{I}} \lambda_i \nabla h_i(x) + \sum_{j \in \mathcal{J}} \mu_j \nabla g_j(x) + \sum_{j \in \mathcal{A}_+(\bar{x})} \mu_j \nabla g_j(x) \;\middle|\; \mu_j \geq 0 \text{ for all } j \in \mathcal{A}_+(\bar{x}) \right\},\end{aligned}$$

where $\mathcal{I} \subseteq \{1, \ldots, m\}$, $\mathcal{J} \subseteq \mathcal{A}_-(\bar{x})$, $\mathcal{A}_+(x) = \mathcal{A}(x) \backslash \mathcal{A}_-(x)$ and $\mathcal{A}_-(x)$ is given in (2). In this set, inequality constraints with indices in $\mathcal{J}$ are treated as equalities in the sense that their associated multipliers are free of sign.

Definition 6 (CPG [8])**.** *A feasible point $\bar{x}$ for* (P) *satisfies CPG if there are index sets $\mathcal{I} \subseteq \{1, \ldots, m\}$ and $\mathcal{J} \subseteq \mathcal{A}_-(\bar{x})$ such that the gradients within $\mathcal{K}(\bar{x}; \bar{x}, \mathcal{I}, \mathcal{J})$ are positively linearly independent (that is, the unique way to $0 \in \mathcal{K}(\bar{x}; \bar{x}, \mathcal{I}, \mathcal{J})$ is taking $(\lambda, \mu) = 0$) and*

$$\mathcal{K}(x; \bar{x}) \subseteq \mathcal{K}(x; \bar{x}, \mathcal{I}, \mathcal{J})$$

for all x in a neighbourhood of $\bar{x}$.

It is known that CRSC implies CPG [8], which in turn implies AL-regularity [5]. However, CPG and RQN are independent of each other, as the next example shows.

Example 3 (CPG does not imply RQN)**.** *Let*

$$\Omega = \{x \in \mathbb{R}^2 \mid g_1(x) = x_1 \leq 0,\ h_1(x) = x_1^3 + x_2 = 0,\ h_2(x) = e^{x_2} - 1 = 0\}$$

and $\bar{x} = (0,0) \in \Omega$*, for which* $\mathcal{A}(\bar{x}) = \{1\}$*. It is easy to see that* $\mathcal{A}_-(\bar{x}) = \emptyset$*, so* $\mathcal{A}_+(\bar{x}) = \{1\}$*. The gradients* $\nabla g_1(\bar{x}) = (1,0)$ *and* $\nabla h_2(\bar{x}) = (0,1)$ *are linearly independent, which lead us to take* $\mathcal{I} = \{2\}$ *and* $\mathcal{J} = \emptyset$ *in Definition 6. We have* $\mathcal{K}(x;\bar{x}) = \mathbb{R}_+ \times \mathbb{R} = \mathcal{K}(x;\bar{x},\mathcal{I},\mathcal{J})$ *for all* x*, and therefore CPG holds at* $\bar{x}$*.*

To show that RQN does not hold at $\bar{x}$*, it is enough to consider* $\mu_1 = 0$*,* $\lambda = (-1,1)$ *and* $x^k = (-1/k, 1/k^4)$ *for all* $k \geq 2$*. In fact, we have* $\lim_k x^k = \bar{x}$*,* $\mu_1 \nabla g_1(\bar{x}) + \lambda_1 \nabla h_1(\bar{x}) + \lambda_2 \nabla h_2(\bar{x}) = 0$ *and, for all* $k \geq 2$*,* $\lambda_1 h_1(x^k) = 1/k^3 - 1/k^4 > 0$*,* $\lambda_2 h_2(x^k) = e^{1/k^4} - 1 > 0$ *and* $(x_1^k)_+ = 0 = o(w(x^k))$*.*

Example 4 (RQN does not imply CPG)**.** *As in Example 1, consider the constraints* $g_1(x) = -x_1 \leq 0$*,* $g_2(x) = x_1 - x_2^2 \leq 0$ *and the point* $\bar{x} = (0,0)$*. It was shown previously that QN holds at* $\bar{x}$*, so RQN also holds. On the other hand, CPG is not valid at* $\bar{x}$*. In fact, we have* $\mathcal{A}_-(\bar{x}) = \{1,2\}$*. It can not be* $\mathcal{J} = \{1,2\}$ *since* $\nabla g_1(\bar{x})$ *and* $\nabla g_2(\bar{x})$ *are not positively linearly independent. Furthermore, for any* $\delta \neq 0$ *we have* $(1, -3\delta^2) \in \mathcal{K}((0,\delta);\bar{x}) \backslash \mathcal{K}((0,\delta);\bar{x},\emptyset,\{1\})$ *and* $(-1,0) \in \mathcal{K}((0,\delta);\bar{x}) \backslash \mathcal{K}((0,\delta);\bar{x},\emptyset,\{2\})$*, so CPG does not hold at* $\bar{x}$*.*

We summarize all the relations discussed in this section in Figure 1, which brings several known CQs from the literature not considered in this work. The reader is refereed to [5] and references therein for details.

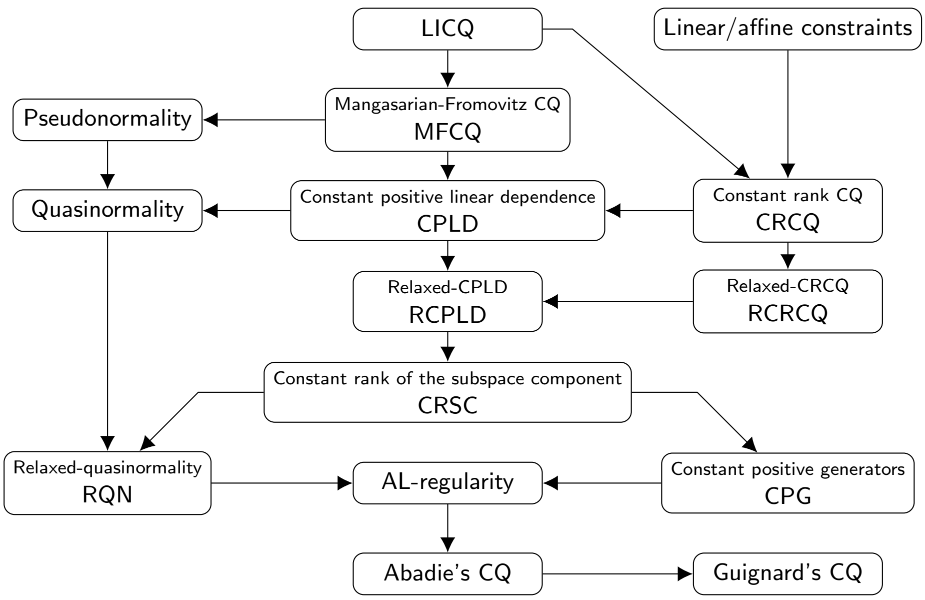

Figure 1: Landscape of constraint qualifications for nonlinear programming problems.

6 Conclusions

Weak constraint qualifications (in particular, those weaker than MFCQ and LICQ) have been largely used for several different purposes in nonlinear programming and more general optimization problems; namely for studying stability properties, error bound estimates, differentiability of the value function, global convergence of algorithms, among other applications. In particular, several studies have appeared related with constant rank-type constraint qualifications, which are the most well known of these conditions. On the other hand, the quasinormality constraint qualification has appeared in the context of enhanced Fritz-John conditions connected with the notion of a more precise (enhanced) class of Lagrange multipliers.

In some sense, constant rank constraint qualifications introduced in the recent years dictate that equality constraints should be treated differently than inequality constraints, with the exception of some inequalities that behave like equalities. In this paper we proposed a similar relaxation of the quasinormality condition, which turned out to be connected with the notion of informative Lagrange multipliers, where a Lagrange multiplier that vanishes must also somehow conform to a sign constraint with respect to how the constraints may be violated nearby the point of interest.

Concerning the global convergence properties of a safeguarded augmented Lagrangian method, several constraint qualifications have been used for this purpose but only the strongest ones were known to provide boundedness of the dual sequences. This property is particularly relevant for complexity analysis and for applications where a dual solution is actually sought (such as in energy pricing applications, among others). In this paper we showed that our relaxed quasinormality condition is enough for ensuring this result, which implies that all constant rank-type constraint qualifications also inherit this property. This is particularly relevant due to the pivotal role played by the so-called constant rank of the subspace component condition (CRSC).

Other applications and extensions have been discussed, in particular connected with attaining a feasible limit point and the use of a scaled stopping criterion. We expect future research to expand the applicability of relaxed quasinormality to other areas where quasinormality was previously used.

References

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

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

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

[4] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez. A minimal face constant rank constraint qualification for reducible conic programming, 2023.

[5] 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.

[6] R. Andreani, G. Haeser, M. L. Schuverdt, L. D. Secchin, and P. J. S. Silva. On scaled stopping criteria for a safeguarded augmented Lagrangian method with theoretical guarantees. *Mathematical Programming Computation*, 14(1):121–146, 2022.

[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, and M. L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125(2):473–483, 2005.

[10] R. Andreani, M. L. Schuverdt, and L. D. Secchin. On enhanced KKT optimality conditions for smooth nonlinear optimization. *Optimization online*, 2022.

[11] K. Bai, Y. Song, and J. Zhang. Second-order enhanced optimality conditions and constraint qualifications. *Journal of Optimization Theory and Applications*, 198(3):1264–1284, Aug. 2023.

[12] D. P. Bertsekas. *Nonlinear Programming: 2nd Edition*. Athena Scientific, 1999.

[13] D. P. Bertsekas and A. E. Ozdaglar. Pseudonormality and a Lagrange multiplier theory for constrained optimization. *Journal of Optimization Theory and Applications*, 114(2):287–343, 2002.

[14] E. G. Birgin, G. Haeser, and J. Martínez. Safeguarded augmented Lagrangian algorithms with scaled stopping criterion for the subproblems. *Optimization online*, 2023.

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

[16] J. M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *Journal of the Australian Mathematical Society*, 30(3):369–380, 1981.

[17] L. Bueno, G. Haeser, and F. Rojas. Optimality conditions and constraint qualifications for generalized nash equilibrium problems and their practical implications. *SIAM Journal on Optimization*, 29(1):31–54, 2019.

[18] L. Guo, J. J. Ye, and J. Zhang. Mathematical programs with geometric constraints in banach spaces: Enhanced optimality, exact penalty, and sensitivity. *SIAM Journal on Optimization*, 23(4):2295–2319, jan 2013.

[19] G. Haeser, H. Liu, and Y. Ye. Optimality condition and complexity analysis for linearly-constrained optimization without differentiability onthe boundary. *Mathematical Programming*, 178:263–299, 2019.

[20] M. R. Hestenes. *Otimization Theory: The Finite Dimensional Case*. John Wiley & Sons, New York, 1975.

[21] O. Hinder and Y. Ye. A one-phase interior point method for nonconvex optimization, 2018.

[22] R. Janin. Directional derivative of the marginal function in nonlinear programming. In A. V. Fiacco, editor, *Sensitivity, Stability and Parametric Analysis*, volume 21 of *Mathematical Programming Studies*, pages 110–126. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.

[23] C. Kanzow and A. Schwartz. Mathematical programs with equilibrium constraints: enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results. *SIAM Journal on Optimization*, 20(5):2730–2753, 2010.

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

[25] A. Khare and T. Nath. Enhanced fritz john stationarity, new constraint qualifications and local error bound for mathematical programs with vanishing constraints. *Journal of Mathematical Analysis and Applications*, 472(1):1042–1077, apr 2019.

[26] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011.

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

[28] L. Qi and Z. Wei. On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000.

[29] J. J. Ye and J. Zhang. Enhanced Karush-Kuhn-Tucker condition and weaker constraint qualifications. *Mathematical Programming*, 139(1):353–381, 2013.

[30] J. J. Ye and J. Zhang. Enhanced Karush-Kuhn-Tucker conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 163(3):777–794, 2014.