A relaxed constant positive linear dependence constraint qualification applied to an augmented Lagrangian method *

Roberto Andreani† Gabriel Haeser‡ María Laura Schuverdt§
Paulo J. S. Silva¶

June 16, 2010

Abstract

In this work we introduce a relaxed version of the constant positive linear dependence constraint qualification (CPLD) that we call RCPLD. This development is inspired by a recent generalization of the constant rank constraint qualification from Minchenko and Stakhovski that was called RCR. We show that RCPLD is enough to ensure the convergence of an augmented Lagrangian algorithm and asserts the validity of an error bound. We also provide proofs and counter-examples that show the relations of RCR and RCPLD with other known constraint qualifications, in particular, RCPLD is strictly weaker than CPLD and RCR, while still stronger than Abadie's constraint qualification. We also verify that RCR is a strong second order constraint qualification.

Key words: Nonlinear Programming, Constraint Qualifications, Practical Algorithms.
AMS Subject Classification: 90C30, 49K99, 65K05.

1 Introduction

In this paper, we consider the nonlinear programming problem

$$\text{Minimize } f(x), \quad \text{subject to} \quad x \in \Omega, \tag{1}$$

where $\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \le 0\}, f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. For each feasible point $x \in \Omega$, we define the set of active inequality constraints $A(x) = \{j \mid g_j(x) = 0, j = 1, \dots, p\}$.

We say that the constraints defining the feasible set Ω satisfy a *constraint qualification* if, independently of the objective function f, for every local solution x of (1), there exist Lagrange

*This work was supported by PRONEX-Optimization (PRONEX-CNPq/FAPERJ E-26/171.510/2006-APQ1), Fapesp (Grants 2006/53768-0 and 2009/09414-7) and CNPq (Grants 300900/2009-0, 303030/2007-0 and 474138/2008-9).

†Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas SP, Brazil. E-mail: andreani@ime.unicamp.br

‡Institute of Mathematics and Statistics, University of São Paulo, São Paulo SP, Brazil. On leave to Department of Science and Technology, Federal University of São Paulo, São José dos Campos SP, Brazil. E-mail: ghaeser@gmail.com

§CONICET, Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina. E-mail: schuverd@mate.unlp.edu.ar

¶Institute of Mathematics and Statistics, University of São Paulo, São Paulo SP, Brazil. E-mail: pjssilva@ime.usp.br

multipliers $\lambda \in \mathbb{R}^m$ and $\mu_i \geq 0$ for every $i \in A(x)$ such that the KKT condition holds:

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

Constraint qualifications are properties of the analytical description of the feasible set that ensure that the reconstruction of its geometrical structure from first order information is possible. The presence of a constraint qualification is then fundamental to derive (analytical) characterizations of the solutions to optimization and variational problems, as well as other theoretical properties related to duality and sensitivity. It is also essential in the development of computational methods and to study their convergence.

In this sense, we emphasize two desirable aspects of constraint qualifications. First, they should be associated to practical algorithms, unveiling weak conditions that ensure convergence. Second, they should assert that sensitivity information that may be used for practical purposes can be readily computed. One good example of such property is the presence of an error bound that can be used to analyze the convergence of computational methods. The error bound property, also known as R-regularity, was introduced by Ioffe [14] and explored by Robinson to study the Lipschitz properties of multifunctions [22] .

The most common constraint qualification is the linear independence constraint qualification (LICQ), which requires that the gradient vectors $\left(\{\nabla h_i(x)\}_{i=1}^m, \{\nabla g_i(x)\}_{i \in A(x)}\right)$ are linearly independent. A weaker condition is the Mangasarian-Fromovitz constraint qualification (MFCQ, [17, 23]), which requires only positive-linear independence of the gradient vectors[1].

Next, we define the constant rank constraint qualification of Janin (CRCQ, [15]), which is also weaker than LICQ.

Definition 1 (CRCQ) *We say that the* constant rank constraint qualification (CRCQ) *holds at a feasible point $x \in \Omega$ if there exists a neighborhood $N(x)$ of x such that for every $I \subset \{1, \ldots, m\}$ and every $J \subset A(x)$, the family of gradients $\{\nabla h_i(y)\}_{i \in I} \cup \{\nabla g_i(y)\}_{i \in J}$ has the same rank for every $y \in N(x)$.*

In [18], Minchenko and Stakhovski provide a relaxed form of the constant rank constraint qualification, which they called *relaxed constant rank* (RCR). Instead of requiring that the rank of every subset of equality and active inequality gradients to remain constant in a neighborhood of a feasible point, they only require constant rank of subsets consisting of *all* equality gradients and any subset of active inequality gradients. The authors proved that this is still a constraint qualification, and they used it to prove an error bound property.

It is well known that CRCQ can be equivalently stated as a constant linear dependence condition, that is: for every $I \subset \{1, \ldots, m\}$ and every $J \subset A(x)$, whenever $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_i(x)\}_{i \in J}$ is linearly dependent, we must have $\{\nabla h_i(y)\}_{i \in I} \cup \{\nabla g_i(y)\}_{i \in J}$ linearly dependent for every $y \in N(x)$, for some neighborhood $N(x)$ of x.

This motivates the definition of the constant positive linear dependence constraint qualification (CPLD, [21, 6]), which is weaker than MFCQ and CRCQ.

Definition 2 (CPLD) *We say that the* constant positive linear dependence condition (CPLD) *holds at a feasible point $x \in \Omega$ if there exists a neighborhood $N(x)$ of x such that for every $I \subset \{1, \ldots, m\}$ and every $J \subset A(x)$, whenever $(\{\nabla h_i(x)\}_{i \in I}, \{\nabla g_i(x)\}_{i \in J})$ is positive-linearly dependent, then $\{\nabla h_i(y)\}_{i \in I} \cup \{\nabla g_i(y)\}_{i \in J}$ is linearly dependent for every $y \in N(x)$.*

[1]The pair of families $\left(\{v_i\}_{i=1}^m, \{v_i\}_{i=m+1}^p\right)$ is said to be positive-linearly dependent if $\{v_i\}_{i=1}^p$ is linearly dependent with non-negative scalars associated to the second family of vectors. Otherwise we say that the pair of families is positive-linearly independent.

The CPLD is an interesting constraint qualification. Up to now, it is the weakest constraint qualification associated to the convergence of a practical augmented Lagrangian algorithm [2, 3], and it is also sufficient to ensure R-regularity [18]. In this last paper, the authors left open the question if the CPLD can be weakened in the same way RCR weakens CRCQ. In this work we will answer this question affirmatively introducing a relaxed version of the CPLD, that we call RCPLD. We will also extend the result on the convergence of the augmented Lagrangian algorithm and the error bound property to RCPLD. We provide proofs and counter-examples that give a complete picture of the relationship of RCR and RCPLD with other well known constraint qualifications.

We will use the following notation:

- $\|\cdot\| = \|\cdot\|_2$,
- $|J|$ denotes the number of elements of the finite set J,
- $\text{span}\{v_i\}_{i=1}^m$ denotes the subspace generated by the vectors $v_1, \dots, v_m$.

2 Relaxed constant rank constraint qualification

We study the relaxed constant rank constraint qualification of Minchenko and Stakhovski (RCR, [18]).

Definition 3 (RCR) *We say that the* relaxed constant rank condition (RCR) *holds at a feasible point $x \in \Omega$ if there exists a neighborhood $N(x)$ of x such that for every $J \subset A(x)$, the family of gradients $\{\nabla h_i(y)\}_{i=1}^m \cup \{\nabla g_i(y)\}_{i\in J}$ has the same rank for every $y \in N(x)$.*

Compared to original CRCQ this relaxation treats the set of equality constraints as a whole, without the need to impose restrictions on all their subsets. In [18], the authors proved that RCR is still a constraint qualification, by showing that it implies Abadie's constraint qualification [1]. They also showed that RCR is strictly weaker than CRCQ. In the case of only equality constraints, this condition was independently formulated in [4]. The RCR condition has also been studied in the context of parametric problems in [16].

Since CRCQ is equivalent to the fact that for every subset of equality and active inequality gradients, linearly dependent vectors remain linearly dependent on some neighborhood, one could conjecture that a similar equivalence holds true for RCR, considering only subsets that contain every equality gradient. But this is not the case. Consider the equality constraints $h_1(x_1, x_2) = x_1, h_2(x_1, x_2) = x_1$ and the inequality constraint $g_1(x_1, x_2) = x_2^2$ at the feasible point $x = (0, 0)$. RCR does not hold since $\{\nabla h_1(y), \nabla h_2(y), \nabla g_1(y)\}$ has rank one at $y = x$ and rank two for y arbitrarily close to x, but subsets that contain both equality gradients are linearly dependent on every neighborhood of x.

We will provide a reformulation of RCR in terms of constant linear dependence. We must keep the condition that the rank of the equality constraints gradients $\{\nabla h_i(y)\}_{i=1}^m$ is constant for every y in some neighborhood $N(x)$ of x. The key point is that in that situation, we may choose a subset $I \subset \{1, \dots, m\}$ such that $\{\nabla h_i(x)\}_{i\in I}$ is a basis for $\text{span}\{\nabla h_i(x)\}_{i=1}^m$, thus, since linearly independent vectors remain linearly independent in a neighborhood, and the rank is the same, we have that $\{\nabla h_i(y)\}_{i\in I}$ is a basis for $\text{span}\{\nabla h_i(y)\}_{i=1}^m$ for every y in some neighborhood $N(x)$ of x. The reformulation requires that for every $J \subset A(x)$, linear dependence is maintained in a neighborhood of x whenever $\{\nabla h_i(x)\}_{i\in I} \cup \{\nabla g_i(x)\}_{i\in J}$ is linearly dependent. Notice that when $\{\nabla h_i(x)\}_{i\in I} \cup \{\nabla g_i(x)\}_{i\in J}$ is linearly dependent, there must exist an index $j \in J$ such that $\nabla g_j(x)$ is a linear combination of the remaining gradients, otherwise this would contradict the linear independence of $\{\nabla h_i(x)\}_{i\in I}$. In order to prove our reformulation, we need:

Lemma 1 *If r is the rank of $\{\nabla h_i(x)\}_{i=1}^m$, then there exists a neighborhood $N(x)$ of x such that the rank of $\{\nabla h_i(y)\}_{i=1}^m$ is greater than or equal to r, for every $y \in N(x)$.*

Proof: This is a direct consequence of the fact that linear independence is preserved locally. $\square$

Theorem 1 *Let $I \subset \{1, \ldots, m\}$ be such that $\{\nabla h_i(x)\}_{i \in I}$ is a basis for span$\{\nabla h_i(x)\}_{i=1}^m$. A feasible point $x \in \Omega$ satisfies RCR if, and only if, there exists a neighborhood $N(x)$ of x such that*

- *$\{\nabla h_i(y)\}_{i=1}^m$ has the same rank for every $y \in N(x)$,*
- *For every $J \subset A(x)$, if $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_i(x)\}_{i \in J}$ is linearly dependent, then $\{\nabla h_i(y)\}_{i \in I} \cup \{\nabla g_i(y)\}_{i \in J}$ is linearly dependent for every $y \in N(x)$.*

Proof: Let $x \in \Omega$ satisfies RCR. The first claim follows by taking $J = \emptyset$ in the definition of RCR. Let $J \subset A(x)$ such that $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_i(x)\}_{i \in J}$ is linearly dependent. Since the gradients corresponding to the set I generate the remaining equality constraints gradients in a neighborhood, using RCR we have that the rank of $\{\nabla h_i(y)\}_{i \in I} \cup \{\nabla g_i(y)\}_{i \in J}$ is constant for every y in some neighborhood $N(x)$ of x, therefore, this set must be linearly dependent for $y \in N(x)$.

To prove the converse let $J \subset A(x)$. Choose $\hat{J} \subset J$ such that $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_i(x)\}_{i \in \hat{J}}$ is a basis for $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_i(x)\}_{i \in J}$. The case $\hat{J} = J$ is trivial. Now let $j \in J \backslash \hat{J}$. As $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_i(x)\}_{i \in \hat{J} \cup \{j\}}$ is linearly dependent, it must remain linearly dependent in $N(x)$. Hence the rank of $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_i(x)\}_{i \in J}$ is not greater than $|I| + |\hat{J}|$. The result now follows from Lemma 1. $\square$

Next we provide counter-examples to show where RCR fits among other well known constraint qualifications. The following counter-example shows that MFCQ does not imply RCR.

Counter-example 1: Consider the inequality constraints $g_1(x_1, x_2) = -x_2$ and $g_2(x_1, x_2) = x_1^2 - x_2$ at the feasible point $x = (0, 0)$. Clearly, MFCQ holds. RCR does not hold since $\{\nabla g_1(y), \nabla g_2(y)\}$ has rank one at $y = x$ and rank two for y arbitrarily close to x.

We say that quasinormality (see [13, 8]) holds at a feasible point $x \in \Omega$ if whenever $\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{i \in A(x)} \mu_i \nabla g_i(x) = 0$, there is not any sequence $y^k \to x$ such that $\lambda_i \neq 0 \Rightarrow \lambda_i h_i(y^k) > 0$ and $\mu_i > 0 \Rightarrow g_i(y^k) > 0$ for every k. The following counter-example shows that RCR does not imply the quasinormality constraint qualification.

Counter-example 2: Consider the equality constraint $h_1(x_1, x_2) = -(x_1 + 1)^2 - x_2^2 + 1$ and the inequality constraints $g_1(x_1, x_2) = x_1^2 + (x_2 + 1)^2 - 1, g_2(x_1, x_2) = -x_2$, at the feasible point $x = (0, 0)$. Quasinormality does not hold, since we can write $\nabla g_1(x) + 2\nabla g_2(x) = 0$ and by taking $y^k = \left(\sqrt{1 - (1 - \frac{1}{k})^2 + \frac{1}{k}}, -\frac{1}{k}\right)$ we have $g_1(y^k) > 0$ and $g_2(y^k) > 0$ for every k. RCR holds since there is a neighborhood $N(x)$ of x such that for every $y \in N(x), \{\nabla h_1(y)\}$ has rank one and $\{\nabla h_1(y), \nabla g_1(y)\}$, $\{\nabla h_1(y), \nabla g_2(y)\}$, $\{\nabla h_1(y), \nabla g_1(y), \nabla g_2(y)\}$ has rank two.

In Figure 1 we show relations of RCR with other well known constraint qualifications, where pseudonormality is the constraint qualification from [8]. The proof that RCR implies Abadie's constraint qualification [1] has been done in [18].

If the problem data is twice continuously differentiable there is the notion of second order constraint qualification. As it was mentioned in [4], a second order constraint qualification can be

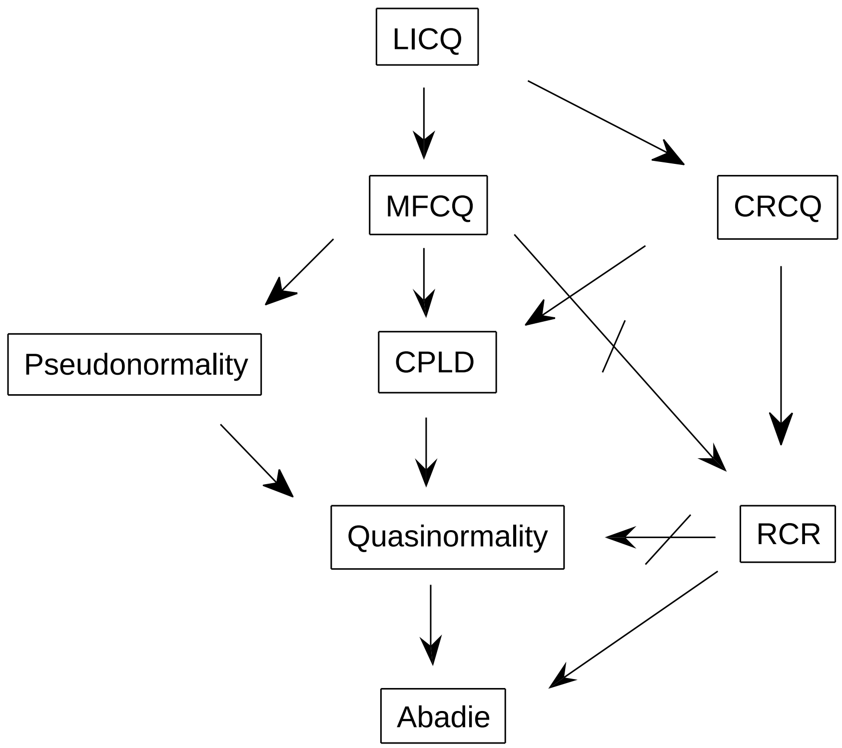

Figure 1: Complete diagram showing relations of RCR with other well known constraint qualifications, where an arrow between two constraint qualifications means that one is strictly stronger than the other.

weak or strong depending on the tangent subspace used to analyse the curvature of the Lagrangian function.

We say that a strong second order constraint qualification holds if whenever x is a local solution of (1) then there exist Lagrange multipliers $\lambda \in \mathbb{R}^n$, $\mu_i \geq 0 \quad \forall i \in A(x)$, satisfying the KKT condition for which

$$d^T \left(\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x) + \sum_{i=1}^{p} \mu_i \nabla^2 g_i(x) \right) d \geq 0, \tag{2}$$

for all directions $d \in \mathbb{R}^n$ in the following tangent subspace:

$$\begin{aligned} \tilde{V}_1(x) = \quad \{d \in \mathbb{R}^n : \quad & \nabla h_i(x)^T d = 0, i = 1, \ldots, m, \\ & \nabla g_j(x)^T d = 0, j \in A^+(x), \\ & \nabla g_j(x)^T d \leq 0, j \in A^0(x)\} \end{aligned}$$

where

$$A^+(x) = \{j \in A(x) : \mu_j > 0\}, A^0(x) = \{j \in A(x) : \mu_j = 0\}.$$

Analogously, we say that a weak second order constraint qualification holds if there exist at least one Lagrange multiplier vector such that (2) holds for all directions $d \in \mathbb{R}^n$ in the following smaller tangent subspace:

$$\tilde{V}_2(x) = \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0, i = 1, \ldots, m, \nabla g_j(x)^T d = 0, j \in A(x)\}.$$

In [4], the authors proved that CRCQ is a strong second order constraint qualification, and the counter-example defined in [7] shows that MFCQ is not even a weak second order constraint qualification. Thus, considering Figure 1, RCR can still be a second order constraint qualification. Actually, the proof that RCR is a strong second order constraint qualification follows from the Remark 3.2 in [4]. In fact, it was shown in [4] that under RCR, if x is a local solution of (1) then (2) holds for all $d \in \tilde{V}_1(x)$ and for every Lagrange multiplier vector.

3 Relaxed constant positive linear dependence constraint qualification

In [21], Qi and Wei proposed a relaxation of CRCQ, the constant positive linear dependence condition (CPLD), taking in consideration the positive sign of the multipliers associated to inequality constraints in the KKT condition. They used this condition to prove convergence of a sequential quadratic programming method.

In [6], it has been proved that CPLD is in fact a constraint qualification, and in [2, 3], the authors proved convergence of an augmented Lagrangian method under CPLD. We now propose a relaxation of CPLD in a way similar to RCR that we call *relaxed CPLD* (RCPLD). The definition is motivated by Theorem 1, considering only positive-linearly dependent gradients, as in the definition of CPLD.

Definition 4 (RCPLD) *Let $I \subset \{1, \ldots, m\}$ be such that $\{\nabla h_i(x)\}_{i \in I}$ is a basis for span$\{\nabla h_i(x)\}_{i=1}^m$. We say that a feasible point $x \in \Omega$ satisfies the* relaxed constant positive linear dependence constraint qualification (RCPLD) *if there exists a neighborhood $N(x)$ of x such that*

- *$\{\nabla h_i(y)\}_{i=1}^m$ has the same rank for every $y \in N(x)$,*
- *For every $J \subset A(x)$, if $(\{\nabla h_i(x)\}_{i \in I}, \{\nabla g_i(x)\}_{i \in J})$ is positive-linearly dependent, then $\{\nabla h_i(y)\}_{i \in I} \cup \{\nabla g_i(y)\}_{i \in J}$ is linearly dependent for every $y \in N(x)$.*

Clearly, Theorem 1 shows that RCR implies RCPLD. It is also clear from the definition that CPLD implies RCPLD. The constant rank of equality constraints gradients follows from the definition of CPLD with $J = \emptyset$ and the equivalence between CRCQ and its definition using constant linear dependence.

An important tool to deal with positive-linearly dependent vectors (in particular, to deal with CPLD or RCPLD) is Carathéodory's Lemma [8]. We will state here a similar result that will be suitable to study the RCPLD. This result can be seen as a corollary of Carathéodory's Lemma, but we include a full proof for completeness.

Lemma 2 *If $x = \sum_{i=1}^{m+p} \alpha_i v_i$ with $v_i \in \mathbb{R}^n$ for every i, $\{v_i\}_{i=1}^m$ linearly independent and $\alpha_i \neq 0$ for every $i = m+1, \ldots, m+p$, then there exist $J \subset \{m+1, \ldots, m+p\}$ and scalars $\bar{\alpha}_i$ for every $i \in \{1, \ldots, m\} \cup J$ such that*

- *$x = \sum_{i \in \{1,\ldots,m\} \cup J} \bar{\alpha}_i v_i$,*

- $\alpha_i \bar{\alpha}_i > 0$ *for every* $i \in J$,
- $\{v_i\}_{i \in \{1,\dots,m\} \cup J}$ *is linearly independent.*

Proof: We assume that $\{v_i\}_{i=1}^{m+p}$ is linearly dependent, otherwise the result follows trivially. Then, there exists $\beta \in \mathbb{R}^{m+p}$, such that $\sum_{i=m+1}^{m+p} |\beta_i| > 0$ and $\sum_{i=1}^{m+p} \beta_i v_i = 0$. Thus, we may write $x = \sum_{i=1}^{m+p} (\alpha_i - \gamma\beta_i) v_i$, for every $\gamma \in \mathbb{R}$. Choosing $\gamma \neq 0$ as the number of smallest modulus such that $\alpha_i - \gamma\beta_i = 0$ for at least one index $i \in \{m+1, \dots, m+p\}$, we are able to write the linear combination x with at least one vector v_i less, for some $i \in \{m+1, \dots, m+p\}$. We may repeat this procedure until the vectors are linearly independent. □

We point out that we can obtain bounds $|\bar{\alpha}_i| \leq 2^{p-1}|\alpha_i|, \forall i = m+1, \dots, m+p$ in the same way it is done in [11]. This may be useful, in particular, for applications to interior point methods.

We now prove that RCPLD is a constraint qualification. We will need a definition from [5]:

Definition 5 (AKKT) *We say that* $x \in \Omega$ *satisfies the* Approximate-KKT condition (AKKT) *if there exist sequences* $x^k \to x$, $\{\lambda^k\} \subset \mathbb{R}^m, \{\mu^k\} \subset \mathbb{R}^p, \mu^k \geq 0$ *such that*

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i \in A(x)} \mu_i^k \nabla g_i(x^k) \to 0.$$

Note that the definition of AKKT also depends on the objective function f, thus, it is a property of the optimization problem, rather than only of the constraint set. In Theorem 2.3 of [5] (with $I = \emptyset$), the authors proved that every local minimizer fulfills the AKKT condition (a simpler proof, specific for the case $I = \emptyset$, can be found in [12]). To prove that RCPLD is a constraint qualification, we need only to show that if RCPLD holds at a feasible point x such that AKKT also holds, then x is a KKT point. This property is also important because it ensures the convergence of an augmented Lagrangian algorithm as we discuss in the next section.

Theorem 2 *Let* $x \in \Omega$ *be such that RCPLD and AKKT hold, then* x *is a KKT point.*

Proof: From the definition of AKKT, there exist sequences $\varepsilon_k \to 0, x^k \to x$, $\lambda^k \in \mathbb{R}^m, \mu_j^k \geq 0, \forall j \in A(x)$, such that

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x)} \mu_j^k \nabla g_j(x^k) = \varepsilon_k, \quad \text{for every } k.$$

Consider a subset $I \subset \{1, \dots, m\}$ such that $\{\nabla h_i(x)\}_{i \in I}$ is a basis for $\text{span}\{\nabla h_i(x)\}_{i=1}^m$. From Lemma 2, we must have $\{\nabla h_i(x^k)\}_{i \in I}$ linearly independent for sufficiently large k, and since the rank of equality constraint gradients is constant, we have that $\{\nabla h_i(x^k)\}_{i \in I}$ is a basis for $\text{span}\{\nabla h_i(x^k)\}_{i=1}^m$ for sufficiently large k. Thus, there exist a sequence $\{\bar{\lambda}^k\} \subset \mathbb{R}^{|I|}$ such that $\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) = \sum_{i \in I} \bar{\lambda}_i^k \nabla h_i(x^k)$, and we may write

$$\nabla f(x^k) + \sum_{i \in I} \bar{\lambda}_i^k \nabla h_i(x^k) + \sum_{j \in A(x)} \mu_j^k \nabla g_j(x^k) = \varepsilon_k.$$

We apply Lemma 2 to obtain subsets $J_k \subset A(x)$ and multipliers $\tilde{\lambda}^k \in \mathbb{R}^{|I|}$ and $\tilde{\mu}_j^k \geq 0, \forall j \in J_k$ such that

$$\nabla f(x^k) + \sum_{i \in I} \tilde{\lambda}_i^k \nabla h_i(x^k) + \sum_{j \in J_k} \tilde{\mu}_j^k \nabla g_j(x^k) = \varepsilon_k,$$

and $\{\nabla h_i(x^k)\}_{i\in I} \cup \{\nabla g_i(x^k)\}_{i\in J_k}$ is linearly independent. We will consider a subsequence such that J_k is the same set J for every k (this can be done since there are finitely many possible sets J_k). Define $M_k = \max\{|\tilde{\lambda}_i^k|, \forall i \in I, \tilde{\mu}_j^k, \forall j \in J\}$. If there is a subsequence such that $M_k \to +\infty$, we may take a subsequence such that $\frac{(\tilde{\lambda}^k, \tilde{\mu}^k)}{M_k} \to (\lambda, \mu) \neq 0, \mu \geq 0$. Dividing by M_k and taking limits we have

$$\sum_{i\in I} \lambda_i \nabla h_i(x) + \sum_{j\in J} \mu_j \nabla g_j(x) = 0,$$

which contradicts RCPLD. Hence, we have that $\{M_k\}$ is a bounded sequence. Taking limits for a suitable subsequence such that $\lambda^k \to \lambda$ and $\mu^k \to \mu \geq 0$ we have

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j\in J} \mu_j \nabla g_j(x) = 0,$$

which proves that x is a KKT point. □

Corollary 1 *RCPLD is a constraint qualification.*

Given a new constraint qualification, it is important to know its relation with other well known constraint qualifications. In particular, we would like to know if RCPLD can still guarantee that the tangent cone is polyhedral. In the following theorem we prove that this is the case by showing that RCPLD implies Abadie's constraint qualification.

Let us consider the feasible set Ω and $x \in \Omega$. We define the (upper) tangent cone of Ω at x as (see for example [8, 10, 13, 25]):

$$T_\Omega(x) = \{0\} \cup \left\{ d \in \mathbb{R}^n : \text{there exists a sequence } \{x_k\} \subset \Omega, x_k \neq x,\ x_k \to x \text{ and } \frac{x_k - x}{\|x_k - x\|} \to \frac{d}{\|d\|} \right\}. \tag{3}$$

We define also the linearized tangent cone at x as:

$$V_\Omega(x) = \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0, i = 1, \ldots, m; \nabla g_j(x)^T d \leq 0, j \in A(x)\}. \tag{4}$$

We say that Abadie's constraint qualification [1] holds at a feasible point $x \in \Omega$ if $T_\Omega(x) = V_\Omega(x)$.

Theorem 3 *Let $x \in \Omega$ be such that RCPLD holds, then x satisfies Abadie's constraint qualification.*

Proof: The inclusion $T_\Omega(x) \subset V_\Omega(x)$ holds without any constraint qualification for every feasible point $x \in \Omega$.

The proof that $V_\Omega(x) \subset T_\Omega(x)$ relies on a simplification considered in [8, 13]. Let us define the set of indexes

$$\hat{J} = \{i \in A(x) : \nabla g_i(x)^T d = 0, \ \ \forall d \in V_\Omega(x)\}.$$

Define

$$\hat{X} = \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \ldots, m; g_j(x) \leq 0, j \in \hat{J}\}.$$

In the degenerate case, where there are no equalities and the set $\hat{J}$ is empty, we have $\hat{X} = \mathbb{R}^n$ by convention. In this case, every point of $\hat{X}$ verifies RCPLD and Abadie.

By the definition of RCPLD, if a feasible point $x \in \Omega$ verifies the RCPLD then it verifies the RCPLD as a point in $\hat{X}$. Using this simplification, let us prove that x verifies Abadie as a point in $\hat{X}$.

We have that $T_{\hat{X}}(x) \subset V_{\hat{X}}(x)$ always holds. Let us take a direction $d \in V_{\hat{X}}(x)$. Let $\varepsilon > 0$, $k > 0$ and let $y(t,k)$ be the minimizer of the function

$$H(y,t,k) = \|y - x - td\|^2 + tk\left(\sum_{i=1}^{m} h_i(y)^2 + \sum_{i\in\hat{J}} \max\{0, g_i(y)\}^2\right)$$

subject to $\|y - x\| \leq \varepsilon$.

We have that, for $t \geq 0$,

$$\|y(t,k) - x - td\|^2 \leq H(y(t,k),t,k) \leq H(x,t,k) = t^2\|d\|^2, \tag{5}$$

and analogously

$$0 \leq k\left(\sum_{i=1}^{m} h_i(y(t,k))^2 + \sum_{i\in\hat{J}} \max\{0, g_i(y(t,k))\}^2\right) \leq t\|d\|^2. \tag{6}$$

By (5) we have

$$\|y(t,k) - x\| \leq 2t\|d\|. \tag{7}$$

Thus, for each $t > 0$, we have that the sequence $\{y(t,k)\}_k$ is a bounded sequence and there exists $y(t), t > 0$ such that, taking a subsequence if necessary, we have

$$y(t,k) \to y(t).$$

Then, taking limits in (6) and by continuity:

$$0 \leq \left(\sum_{i=1}^{m} h_i(y(t))^2 + \sum_{i\in\hat{J}} \max\{0, g_i(y(t))\}^2\right) \leq \lim_{k\to\infty} \frac{t}{k}\|d\|^2 = 0.$$

This implies that $y(t) \in \hat{X}$ for all $t > 0$.

From (7) we have that $y(t) \to x$ as $t \to 0$. Moreover, we can select a sequence of positive numbers t_k with $t_k \to 0$ such that the limit

$$d_0 = \lim_{k\to\infty} \frac{y(t_k) - x}{t_k}$$

exists. Since $y(t_k) \in \hat{X}$ we obtain that $d_0 \in T_{\hat{X}}(x) \subset V_{\hat{X}}(x)$.

Let us define $y_k = y(t_k)$ and let us take k_0 such that $\|y_k - x\| < \varepsilon, \forall k \geq k_0$. By the definition of y_k we have that $\nabla_y H(y_k, t_k, k) = 0$, then

$$\frac{y_k - x - t_k d}{t_k} + k\left(\sum_{i=1}^{m} h_i(y_k)\nabla h_i(y_k) + \sum_{i\in\hat{J}} \max\{0, g_i(y_k)\}\nabla g_i(y_k)\right) = 0.$$

By the definition of RCPLD, we may take a subset $I \subset \{1,\ldots,m\}$ such that $\{\nabla h_i(y_k)\}_{i\in I}$ is a basis for $\operatorname{span}\{\nabla h_i(y_k)\}_{i=1}^m$ for sufficiently large k. Thus, there exist a sequence $\{\lambda^k\} \subset \mathbb{R}^m$ such that $\sum_{i=1}^m kh_i(y^k)\nabla h_i(y_k) = \sum_{i\in I} \lambda_i^k \nabla h_i(y_k)$. By applying Lemma 2, we have that there are subsets $J_k \subset \hat{J}$ and multipliers $\bar{\lambda}_i^k, \forall i \in I$ and $\bar{\mu}_i^k \geq 0, \forall i \in J_k$ such that

$$\frac{y_k - x - t_k d}{t_k} + \sum_{i\in I} \bar{\lambda}_i^k \nabla h_i(y_k) + \sum_{i\in J_k} \bar{\mu}_i^k \nabla g_i(y_k) = 0 \tag{8}$$

and

$$\{\nabla h_i(y_k)\}_{i\in I} \cup \{\nabla g_i(y_k)\}_{i\in J_k} \text{ is linearly independent.} \tag{9}$$

We will consider a subsequence such that J_k is the same set J (this can be done since there are finitely many possible sets J_k).

Denote,

$$M_k = \sqrt{1 + \sum_{i\in I}(\bar{\lambda}_i^k)^2 + \sum_{i\in J}(\bar{\mu}_i^k)^2}.$$

Then, dividing (8) by M_k and taking limit when $k \to \infty$ for k in an appropriate subsequence we have that there are scalars $\mu_0, \lambda_i, i \in I, \mu_j, j \in J, \mu_j \geq 0$ not all equal to zero such that

$$\mu_0(d_0 - d) + \sum_{i\in I}\lambda_i \nabla h_i(x) + \sum_{i\in J}\mu_i \nabla g_i(x) = 0. \tag{10}$$

If $\mu_0 = 0$, then we have that $(\{\nabla h_i(x)\}_{i\in I}, \{\nabla g_i(x)\}_{i\in J})$ is positive-linearly dependent, hence, since (9) holds, this contradicts RCPLD. Consequently, it must be $\mu_0 > 0$. Given $\hat{d} \in V_{\hat{X}}(x)$, let us prove that $\nabla g_j(x)^T\hat{d} = 0$ for every $j \in \hat{J}$. From the definition of $V_{\hat{X}}(x)$ we have that $\nabla g_j(x)^T\hat{d} \leq 0$ for every $j \in \hat{J}$, and from the definition of $\hat{J}$, for every $i \in A(x)\backslash\hat{J}$ there exists $d_i \in V_\Omega(x)$ such that $\nabla g_i(x)^T d_i < 0$. Defining $\bar{d} = \sum_{i\in A(x)\backslash\hat{J}} d_i$, we have that

$$\nabla g_i(x)^T\bar{d} < 0 \text{ for every } i \in A(x)\backslash\hat{J},$$

and

$$\nabla g_j(x)^T\bar{d} = 0 \text{ for every } j \in \hat{J}.$$

Thus, for sufficiently large $\alpha > 0$ we have

$$\nabla g_i(x)^T(\hat{d} + \alpha\bar{d}) = \nabla g_i(x)^T\hat{d} + \alpha\nabla g_i(x)^T\bar{d} < 0 \text{ for every } i \in A(x)\backslash\hat{J},$$

and

$$\nabla g_j(x)^T(\hat{d} + \alpha\bar{d}) = \nabla g_j(x)^T\hat{d} \leq 0 \text{ for every } j \in \hat{J}.$$

Hence $\hat{d} + \alpha\bar{d} \in V_\Omega(x)$, which implies that for all $j \in \hat{J}, \nabla g_j(x)^T(\hat{d} + \alpha\bar{d}) = \nabla g_j(x)^T\hat{d} = 0$ as we wanted to prove. Thus, multiplying (10) by $d, d_0 \in V_{\hat{X}}(x)$, we obtain that

$$d_0^T(d - d_0) = 0 = d^T(d - d_0),$$

which implies that $d = d_0 \in T_{\hat{X}}(x)$.

We have proved that a feasible point x that verifies RCPLD as a point in $\hat{X}$ verifies Abadie's as a point in $\hat{X}$.

Now we have to prove that this implies that x verifies Abadie as a point in Ω. Let us define the set $\tilde{V}_\Omega(x) = \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0, i = 1, \ldots, m; \nabla g_j(x)^T d = 0, j \in \hat{J}; \nabla g_j(x)^T d < 0, j \in A(x)\backslash\hat{J}\}$.

Since x verifies that $T_{\hat{X}}(x) = V_{\hat{X}}(x)$ and $\tilde{V}_\Omega(x) \subset V_{\hat{X}}(x)$ it is possible to prove that $\tilde{V}_\Omega(x) \subset T_\Omega(x)$. In general, $T_\Omega(x)$ is a closed cone, thus, $V_\Omega(x) = \text{cl}(\tilde{V}_\Omega(x)) \subset T_\Omega(x)$, where $\text{cl}(\cdot)$ denotes the closure operator. Thus, we have that x verifies Abadie as a point in Ω as we wanted to prove. □

We will now provide some counter-examples to completely state the relation of RCPLD with respect to other known constraint qualifications. We observe that since MFCQ implies RCPLD, Counter-example 1 shows that RCR is strictly stronger than RCPLD.

We say that pseudonormality (see [8]) holds at a feasible point $x \in \Omega$ if whenever $\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{i \in A(x)} \mu_i \nabla g_i(x) = 0$, there is not any sequence $y^k \to x$ such that $\sum_{i=1}^m \lambda_i h_i(y^k) + \sum_{i \in A(x)} \mu_i g_i(y^k) > 0$ for every k.

The following counter-example shows that pseudonormality does not imply RCPLD.

Counter-example 3: Consider the inequality constraints $g_1(x_1, x_2) = -x_1, g_2(x_1, x_2) = x_1 - x_1^2 x_2^2$, at the feasible point $x = (0, 0)$. RCPLD does not hold since $(\emptyset, \{\nabla g_1(y), \nabla g_2(y)\})$ is positive-linearly dependent at $y = x$ but linearly independent for y arbitrarily close to x. Pseudonormality holds, since we can write $\mu \nabla g_1(x) + \mu \nabla g_2(x) = 0$ for every $\mu > 0$, but $\mu g_1(y_1, y_2) + \mu g_2(y_1, y_2) = -\mu y_1^2 y_2^2 \leq 0$ for every y.

Since RCR does not imply quasinormality and RCR implies RCPLD, we have that RCPLD does not imply quasinormality. In Figure 2 we show a complete diagram picturing the relations of RCPLD with other constraint qualifications.

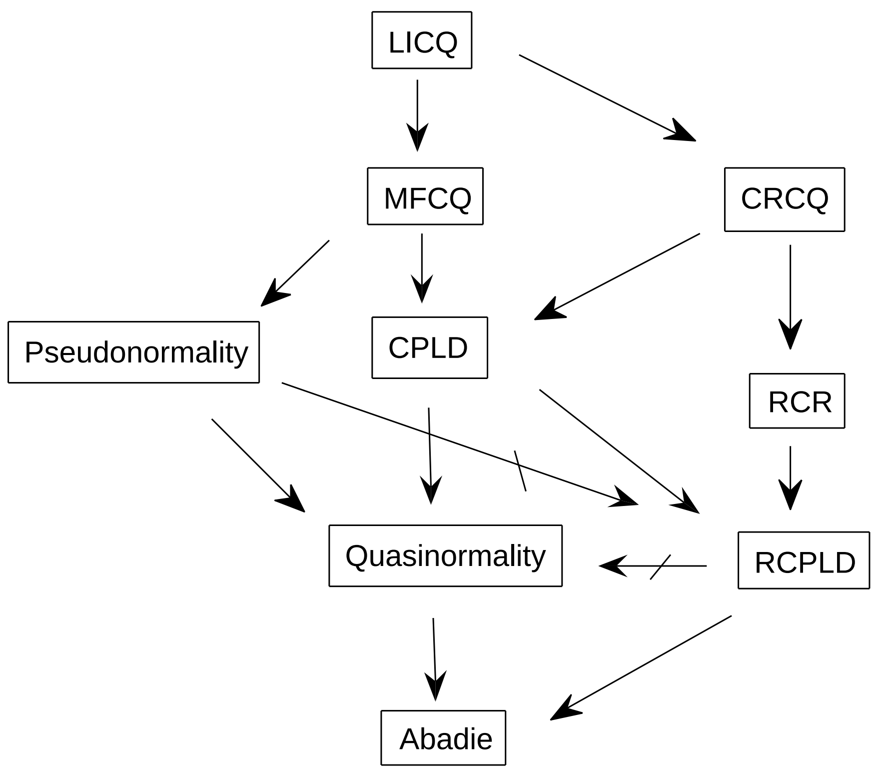

Figure 2: Complete diagram showing relations of RCPLD with other well known constraint qualifications, where an arrow between two constraint qualifications means that one is strictly stronger than the other.

Observe that, since MFCQ is not a second order constraint qualification and MFCQ implies RCPLD, this shows that RCPLD cannot be a second order constraint qualification.

It can be proved (see [24]) that the CPLD condition can be equivalently stated at a feasible point $x \in \Omega$ as: for every subset $I \subset \{1, \dots, m\}$ and $J \subset A(x)$, whenever $(\{\nabla h_i(x)\}_{i \in I}, \{\nabla g_i(x)\}_{i \in J})$ is positive-linearly dependent, we have that $(\{\nabla h_i(y)\}_{i \in I}, \{\nabla g_i(y)\}_{i \in J})$ is positive-linearly dependent for every y in some neighborhood of x. That is, requiring that the gradients are positive-linearly dependent in a neighborhood instead of the apparently weaker requirement of linear dependence, is in fact the same thing. This is an aesthetically pleasant result, that also guarantees that CPLD is stable in the sense that if a feasible point $x \in \Omega$ satisfies CPLD, then every feasible point in some neighborhood of x will also satisfy CPLD. We will prove an analogous equivalent definition for RCPLD that also guarantees stability of RCPLD.

Theorem 4 *Let $I \subset \{1, \dots, m\}$ be such that $\{\nabla h_i(x)\}_{i \in I}$ is a basis for span$\{\nabla h_i(x)\}_{i=1}^m$. A feasible point $x \in \Omega$ satisfies RCPLD if, and only if, there exists a neighborhood $N(x)$ of x such that*

- *$\{\nabla h_i(y)\}_{i=1}^m$ has the same rank for every $y \in N(x)$,*
- *For every $J \subset A(x)$, if $(\{\nabla h_i(x)\}_{i \in I}, \{\nabla g_i(x)\}_{i \in J})$ is positive-linearly dependent, then $(\{\nabla h_i(y)\}_{i \in I}, \{\nabla g_i(y)\}_{i \in J})$ is positive-linearly dependent for every $y \in N(x)$.*

Proof: Let us take $x \in \Omega$ that satisfies RCPLD and $J \subset A(x)$ such that $(\{\nabla h_i(x)\}_{i \in I}, \{\nabla g_i(x)\}_{i \in J})$ is positive-linearly dependent. Thus, there exist $\lambda_i \in \mathbb{R}, \forall i \in I, \mu_i \geq 0, \forall i \in J, \sum_{i \in J} \mu_i > 0$ such that $\sum_{i \in I} \lambda_i \nabla h_i(x) + \sum_{i \in J} \mu_i \nabla g_i(x) = 0$. We can assume that $\mu_i > 0$ for every $i \in J$. Since $J \neq \emptyset$, taking $j \in J$ we may write $\mu_j \nabla g_j(x) = \sum_{i \in I} -\lambda_i \nabla h_i(x) + \sum_{i \in J \backslash \{j\}} -\mu_i \nabla g_i(x)$. By Lemma 2, there exist $J' \subset J \backslash \{j\}$ and $\bar{\lambda}_i \in \mathbb{R}, \forall i \in I, \bar{\mu}_i > 0, \forall i \in J'$ such that

$$\mu_j \nabla g_j(x) = \sum_{i \in I} -\bar{\lambda}_i \nabla h_i(x) + \sum_{i \in J'} -\bar{\mu}_i \nabla g_i(x) \tag{11}$$

and

$$\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_i(x)\}_{i \in J'} \tag{12}$$

is linearly independent.

Now, RCPLD ensures that equation (11) has a solution in $\bar{\lambda}$ and $\bar{\mu}$ when we change x for y in a neighborhood of x. As all the functions involved are continuous and (12) holds, it follows from the pseudo-inverse formula that $\bar{\lambda}$ and $\bar{\mu}$ will change continuously in a neighborhood of x, in particular preserving $\bar{\mu}_i > 0$ for every $i \in J'$. $\square$

4 Applications of RCPLD

We now show how to apply the RCPLD constraint qualification to obtain a stronger convergence result to the general augmented Lagrangian method introduced in [2, 3]. We define the method with some small changes in the penalty parameter update suggested in [9].

We consider the problem

$$\text{Minimize } f(x), \text{ subject to } \quad h(x) = 0, g(x) \leq 0, \underline{h}(x) = 0, \underline{g}(x) \leq 0, \tag{13}$$

where $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p, \underline{h} : \mathbb{R}^n \to \mathbb{R}^{\underline{m}}$ and $\underline{g} : \mathbb{R}^n \to \mathbb{R}^{\underline{p}}$ are continuously differentiable functions. When the constraints $\underline{h}(x) = 0$ and $\underline{g}(x) \leq 0$ define a box in $\mathbb{R}^n$, this is the algorithm implemented in ALGENCAN[2]. Given $\rho > 0, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p, \mu \geq 0, x \in \mathbb{R}^n$ we define the augmented Lagrangian function

$$\mathcal{L}_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2}\left(\left\|h(x) + \frac{\lambda}{\rho}\right\|^2 + \left\|\max\left\{0, g(x) + \frac{\mu}{\rho}\right\}\right\|^2\right). \tag{14}$$

Algorithm Let $\varepsilon_k \geq 0, \varepsilon_k \to 0, \bar{\lambda}^k \in [\lambda_{min}, \lambda_{max}]^m, \bar{\mu}^k \in [0, \mu_{max}]^p$ for all k, $\rho_1 > 0, \tau \in (0,1), \eta > 1$.

For all k, compute $x^k \in \mathbb{R}^n$ such that there exist $v^k \in \mathbb{R}^{\underline{m}}, w^k \in \mathbb{R}^{\underline{p}}, w^k \geq 0$ satisfying:

$$\left\|\nabla_x \mathcal{L}_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) + \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^{\underline{p}} w_i^k \nabla \underline{g}_i(x^k)\right\| \leq \varepsilon_k, \tag{15}$$

$$\|\underline{h}(x^k)\| \leq \varepsilon_k, \|\max\{0, \underline{g}(x^k)\}\| \leq \varepsilon_k, \tag{16}$$

and

$$w_i^k = 0 \text{ whenever } \underline{g}_i(x^k) < -\varepsilon_k. \tag{17}$$

We define, for all $i = 1, \ldots, p$,

$$V_i^k = \max\left\{g_i(x^k), \frac{-\bar{\mu}_i^k}{\rho_k}\right\}. \tag{18}$$

If $k = 1$ or

$$\max\{\|h(x^k)\|, \|V^k\|\} \leq \tau \max\{\|h(x^{k-1})\|, \|V^{k-1}\|\} \tag{19}$$

we define $\rho_{k+1} \geq \rho_k$. Else, we define $\rho_{k+1} \geq \eta\rho_k$.

Remark: We can define the multiplier sequences $\{\bar{\lambda}^k\}$ and $\{\bar{\mu}^k\}$ using for example the first order update formula $\bar{\lambda}_i^{k+1} = P_{[\lambda_{min}, \lambda_{max}]^m}(\bar{\lambda}_i^k + \rho_k h_i(x^k)), i = 1, \ldots, m$ and $\bar{\mu}_i^{k+1} = P_{[0, \mu_{max}]^p}(\bar{\mu}_i^k + \rho_k g_i(x^k)), i = 1, \ldots, p$, where $P_X(\cdot)$ denotes the euclidean projection in X.

Theorem 5 *If x^* is a limit point of a sequence generated by the Algorithm, then x^* is an AKKT point of the problem*

$$\textit{Minimize } \|h(x)\|^2 + \|\max\{0, g(x)\}\|^2, \quad \textit{subject to} \quad \underline{h}(x) = 0, \underline{g}(x) \leq 0. \tag{20}$$

Proof: Consider a subsequence such that $x^k \to x^*$. Since $\varepsilon_k \to 0$, by (16) we have that $\underline{h}(x^*) = 0$ and $\underline{g}(x^*) \leq 0$. If $\{\rho_k\}$ is bounded we have that (19) is satisfied for every sufficiently large k, which implies $h(x^*) = 0$ and $g(x^*) \leq 0$. Hence, x^* is a global minimum for problem (20). Let us assume $\rho_k \to +\infty$. By (14) and (15) we have that

$$\begin{aligned}\nabla f(x^k) + \sum_{i=1}^{m}(\bar{\lambda}_i^k + \rho_k h_i(x^k))\nabla h_i(x^k) + \sum_{i=1}^{p}\max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\}\nabla g_i(x^k) + \\ + \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^{\underline{p}} w_i^k \nabla \underline{g}_i(x^k) = \delta^k,\end{aligned} \tag{21}$$

[2] freely available at *www.ime.usp.br/∼egbirgin/tango*

where $\delta_k \to 0$. If $\underline{g}_i(x^*) < 0$, then $\underline{g}_i(x^k) < -\varepsilon_k$ for sufficiently large k, which implies by (17) that $w_i^k = 0$ for sufficiently large k. Dividing (21) by ρ_k we may write:

$$\sum_{i=1}^{m} h_i(x^k)\nabla h_i(x^k) + \sum_{i=1}^{p} \max\{0, g_i(x^k)\}\nabla g_i(x^k) + \sum_{i=1}^{\underline{m}} \frac{v_i^k}{\rho_k}\nabla \underline{h}_i(x^k) + \sum_{\underline{g}_i(x^*)=0} \frac{w_i^k}{\rho_k}\nabla \underline{g}_i(x^k)$$
$$= \frac{\delta^k}{\rho_k} - \frac{\nabla f(x^k)}{\rho_k} - \sum_{i=1}^{m}\frac{\bar{\lambda}_i^k}{\rho_k}\nabla h_i(x^k) + \sum_{i=1}^{p}\left(\max\{0, g_i(x^k)\} - \max\left\{0, \frac{\bar{\mu}_i^k}{\rho_k} + g_i(x^k)\right\}\right)\nabla g_i(x^k). \tag{22}$$

Since $\{\bar{\lambda}^k\}$ and $\{\bar{\mu}^k\}$ are bounded sequences, the right hand side of (22) goes to zero, thus x^* satisfies the AKKT condition for problem (20). □

Theorem 6 *If x^* is a limit point of a sequence generated by the Algorithm such that x^* is feasible for (13), then x^* is an AKKT point of problem (13).*

Proof: By (14) and (15) we have that (21) holds, where $\delta_k \to 0$. As in the proof of Theorem 5 we have that if $\underline{g}_i(x^*) < 0$ then $w_i^k = 0$ for sufficiently large k. Define $\lambda_i^k = \bar{\lambda}_i^k + \rho_k h_i(x^k)$ and $\mu_i^k = \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \geq 0$. Now let us assume $g_i(x^*) < 0$. If $\{\rho_k\}$ is unbounded, then $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$, hence $\mu_i^k = 0$ for sufficiently large k. If $\{\rho_k\}$ is bounded we have that (19) is satisfied for every sufficiently large k, hence $V_i^k \to 0$, which implies by (18) that $\bar{\mu}_i^k \to 0$, thus $\mu_i^k = 0$ for sufficiently large k. Then, we may write (21) as:

$$\begin{aligned}\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{g_i(x^*)=0} \mu_i^k \nabla g_i(x^k) + \\ + \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{\underline{g}_i(x^*)=0} w_i^k \nabla \underline{g}_i(x^k) = \delta^k,\end{aligned} \tag{23}$$

hence x^* satisfies the AKKT condition for problem (13). □

Applying Theorem 2 we may sate the following:

Corollary 2 *If x^* is a limit point of a sequence generated by the Algorithm, then one of the following holds:*

- *x^* is a feasible point of (13).*
- *x^* is a KKT point of the problem*

$$\textit{Minimize } \|h(x)\|^2 + \|\max\{0, g(x)\}\|^2, \quad \textit{subject to} \quad \underline{h}(x) = 0, \underline{g}(x) \leq 0. \tag{24}$$

- *The constraint set defined by $\underline{h}(x) = 0, \underline{g}(x) \leq 0$ does not satisfy RCPLD at x^*.*

Corollary 3 *If x^* is a limit point of a sequence generated by the Algorithm such that x^* is feasible for (13), then one of the following holds:*

- *x^* is a KKT point of problem (13).*
- *The constraint set defined by $h(x) = 0, g(x) \leq 0, \underline{h}(x) = 0, \underline{g}(x) \leq 0$ does not satisfy RCPLD at x^*.*

In [3], the authors proved these same convergence theorems but employing a stronger constraint qualification (the CPLD). In particular, our result also shows convergence of this Augmented Lagrangian algorithm under RCR. We point out that we can also use RCPLD to prove convergence of a version of this augmented Lagrangian algorithm that do not require derivatives [20].

We now show that an error bound property holds under the RCPLD. This has been previously done for RCR and CPLD in [18], and alternatively for RCR in [16]. As mentioned in [19, 25], an initial motivation to study error bounds arose from a practical consideration in the computer implementation of iterative methods for solving optimization and equilibrium programs. An error bound is an estimate of the distance from a given feasible point in terms of computable quantities measuring the violation of the constraints. It is associated to a so-called residual function that plays a central role in the treatment of infeasible mathematical programming problems. We will prove the stronger result that the error bound property holds under RCPLD.

Definition 6 *We say that the point $x \in \Omega$ satisfies the* error bound property *with respect to the constraints $h(x) = 0$ and $g(x) \le 0$ if there exist $\alpha > 0$ and a neighborhood $N(x)$ of x such that for every $y \in N(x)$*

$$\min_{z\in\Omega} \|z - y\| \le \alpha r(y),$$

where r is a function that measures the infeasibility with respect to Ω that is readily computable.

Theorem 7 *If $x \in \Omega$ satisfies RCPLD and the functions h and g defining Ω admit second derivatives in a neighborhood of x, then x satisfies the error bound property with $r(y) = \max\{\|h(y)\|_\infty, \|\max\{0, g(y)\}\|_\infty\}$.*

Proof: If x is in the interior of Ω, then clearly the error bound property holds. We will assume that x lies in the frontier of Ω. For a fixed $y \in \mathbb{R}^n$, consider the problem

$$\text{Minimize } \|z - y\|, \quad \text{subject to} \quad h(z) = 0, g(z) \le 0. \tag{25}$$

In Theorem 2 of [18], the authors proved that if second derivatives are available, then the error bound property holds at $x \in \Omega$ if, and only if, there exists a neighborhood $N(x)$ of x such that there exist Lagrange multipliers to problem (25) that lie in a fixed compact set for all $y \in N(x)$, $y \notin \Omega$. Let us consider a sequence $y^k \to x$, $y^k \notin \Omega$ and let z^k be a solution to (25) for $y = y^k$. Since $\|z^k - y^k\| \le \|x - y^k\|$ we have also $z^k \to x$. It is a consequence of Theorem 4 that the RCPLD condition is preserved in a neighborhood, thus $z^k \in \Omega$ also satisfies RCPLD for sufficiently large k. Hence, there exist $\{\lambda^k\} \subset \mathbb{R}^m$ and $\{\mu^k\} \subset \mathbb{R}^p$, $\mu_i^k \ge 0$ such that

$$\frac{z^k - y^k}{\|z^k - y^k\|} + \sum_{i=1}^m \lambda_i^k \nabla h_i(z^k) + \sum_{i\in A(z^k)} \mu_i^k \nabla g_i(z^k) = 0,$$

for sufficiently large k, where $A(z^k) = \{i \in \{1, \dots, p\} \mid g_i(z^k) = 0\}$. From the definition of RCPLD we have that there exist $I \subset \{1, \dots, m\}$ and $\tilde{\lambda}_i^k$ for every $i \in I$ such that $\{\nabla h_i(z^k)\}_{i\in I}$ is linearly independent and $\sum_{i=1}^m \lambda_i^k \nabla h_i(z^k) = \sum_{i\in I} \tilde{\lambda}_i^k \nabla h_i(z^k)$ for sufficiently large k, hence, by Lemma 2, there exist $J_k \subset A(z^k)$, $\bar{\lambda}_i^k$ for every $i \in I$ and $\bar{\mu}_i^k \ge 0$ for every $i \in J_k$ such that:

$$\frac{z^k - y^k}{\|z^k - y^k\|} + \sum_{i\in I} \bar{\lambda}_i^k \nabla h_i(z^k) + \sum_{i\in J_k} \bar{\mu}_i^k \nabla g_i(z^k) = 0, \tag{26}$$

and

$$\{\nabla h_i(z^k)\}_{i\in I} \cup \{\nabla g_i(z^k)\}_{i\in J_k} \tag{27}$$

is linearly independent. Let us consider a subsequence such that J_k is the same set J for every k, where $J \subset A(z^k) \subset A(x)$. Define $M_k = \|(\bar{\lambda}^k, \bar{\mu}^k)\|_\infty$ and let us assume by contradiction that

$\{M_k\}$ is unbounded. Taking a subsequence such that $\frac{(\bar{\lambda}^k, \bar{\mu}^k)}{M_k} \to (\lambda, \mu) \neq 0, \mu \geq 0$, we may divide (26) by M_k and take limits for this subsequence to obtain:

$$\sum_{i \in I} \bar{\lambda}_i \nabla h_i(x) + \sum_{i \in J} \bar{\mu}_i \nabla g_i(x) = 0.$$

Since RCPLD holds at x we have that $\{\nabla h_i(z^k)\}_{i \in I} \cup \{\nabla g_i(z^k)\}_{i \in J}$ must be linearly dependent for sufficiently large k, which contradicts (27). This concludes the proof. $\square$

5 Final Remarks

We introduced a generalization of the RCR constraint qualification called RCPLD. We showed that this constraint qualification is strictly weaker than RCR and CPLD. The RCPLD shares with CPLD many of its important properties. In particular, it is enough to ensure the convergence of an Augmented Lagrangian algorithm and the presence of an error bound.

An interesting question that was not touched in this paper is whether it is possible to extend the RCR in a way that does not involve assumptions on the behavior of the gradients of *all subsets* of the active inequality constraints. Such extension would better fit the spirit of RCR when the description of the constraint does not have any inequalities. In this case, the assumption of constant rank has to be fulfilled only by the set of all gradients of the constraints.

Another question is whether RCPLD can still be weakened preserving the convergence of augmented Lagrangian algorithms. It may be also interesting to investigate its role in the convergence of other optimization methods, like sequential quadratic programming or inexact-restoration, as well as in the convergence of the extension of such methods to deal with variational inequalities. Finally, it may be valuable to search for an alternative proof that RCPLD implies the validity of an error bound that does not depend on the existence of second derivatives as required in Theorem 7.

Acknowledgement

The authors would like to thank Prof. Alexey Izmailov from the Moscow State University for pointing out the work of Minchenko and Stakhovski [18].

References

[1] J. Abadie. On the Kuhn-Tucker theorem. In J. Abadie, editor, *Nonlinear Programming*, pages 21–36. John Wiley, 1967.

[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*, 18:1286–1309, 2007.

[3] 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*, 112:5–32, 2008.

[4] R. Andreani, C.R. Echagüe, and M.L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization Theory and Applications*, 2010. DOI: 10.1007/s10957-010-9671-8.

[5] R. Andreani, G. Haeser, and J.M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 2010. (to appear). Available at Optimization Online http://www.optimization-online.org/DB_HTML/2009/06/2326.html.

[6] R. Andreani, J.M. Martínez, and M.L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125:473–485, 2005.

[7] A. V. Arutyunov. Perturbations of extremum problems with constraints and necessary optimality conditions. *Journal of Soviet Mathematics*, 54:1342–1400, 1991.

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

[9] E. Birgin, L.F. Bueno, N. Krejić, and J.M. Martınez. Low Order-Value approach for solving VaR-constrained optimization problems. 2010. submitted.

[10] G. Giorgi and A. Guerraggio. On the notion of tangent cone in mathematical programming. *Optimization*, 25:1:11–23, 1992.

[11] G. Haeser. On the global convergence of interior-point nonlinear programming algorithms. *Computational & Applied Mathematics*, 2010. (to appear). Available at Optimization Online http://www.optimization-online.org/DB_HTML/2009/06/2332.html.

[12] G. Haeser and M.L. Schuverdt. Approximate KKT condition for continuous variational inequality problems. 2010. submitted. Available at Optimization Online http://www.optimization-online.org/DB_HTML/2009/10/2415.html.

[13] M.R. Hestenes. *Optimization theory: the finite dimensional case.* John Wiley, 1975.

[14] A.D. Ioffe. Regular points of Lipschitz functions. *Transactions of American Mathematical Society*, 251:61–69, 1979.

[15] R. Janin. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*, 21:110–126, 1984.

[16] S. Lu. Relation between the constant rank and the relaxed constant rank constraint qualifications. 2010. submitted. Available at http://www.unc.edu/∼shulu/.

[17] O.L. Mangasarian and S. Fromovitz. The Fritz John optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17:37–47, 1967.

[18] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 2010. (to appear).

[19] J.-S. Pang. Error bound in mathematical programming. *Mathematical Programming*, 79:299–332, 1997.

[20] L.G. Pedroso. *Programação não linear sem derivadas.* PhD thesis, IMECC-UNICAMP, Departamento de Matemática Aplicada, Campinas-SP, Brazil, 2009. http://libdigi.unicamp.br/document/?code=000468366.

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

[22] S.M. Robinson. Stability theory for systems of inequalities. part II: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13(4):497–513, 1976.

[23] R.T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35(2):183–238, 1993.

[24] M.L. Schuverdt. *Métodos de Lagrangiano aumentado com convergência utilizando a condição de dependência linear positiva constante*. PhD thesis, IMECC-UNICAMP, Departamento de Matemática Aplicada, Campinas-SP, Brazil, 2006. http://libdigi.unicamp.br/document/?code=vtls000375801.

[25] M. Solodov. *Encyclopedia of Operations Research and Management Science*, chapter Constraints Qualifications. Kluwer Academic Publishers, 2010, to appear.