Optimization
Vol. 00, No. 00, Month-Month 200x, 1–18

On Second-Order Optimality Conditions for Nonlinear Programming

R. Andreani† & J. M. Martínez‡ & M. L. Schuverdt§

(Received 00 Month 200x; In final form 00 Month 200x)

Necessary Optimality Conditions for Nonlinear Programming are discussed in the present research. A new Second-Order condition is given, which depends on a weak constant rank constraint requirement. We show that practical and publicly available algorithms (www.ime.usp.br/∼egbirgin/tango) of Augmented Lagrangian type converge, after slight modifications, to stationary points defined by the new condition.

Key words: Nonlinear programming, Necessary Optimality Conditions, Constraint Qualifications, Practical Algorithms.

1 Introduction

The study of Necessary Optimality Conditions (NOC) is quite relevant for practical Nonlinear Programming. In constrained and unconstrained optimization one is generally interested in finding global minimizers. Since this is very hard, especially in large-scale problems, most practical algorithms are guaranteed to find only stationary points, that is, points that satisfy some NOC. Satisfactory NOC's must be strong. Obviously, the strongest possible NOC is Global Optimality. Global Optimality implies Local Optimality and Local Optimality implies most popular Necessary Optimality Conditions. Very weak NOC's (for example, mere feasibility) are almost unuseful.

December 11, 2006

† Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grants 06/53768-0 and 05-56773-1) and CNPq. e-mail: andreani@ime.unicamp.br
‡Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grant 06/53768-0) and CNPq. e-mail: martinez@ime.unicamp.br
§Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1) and FAPESP (Grants 06/53768-0 and 05-57684-2). e-mail: schuvertd@ime.unicamp.br

The predicted quality of an algorithm is linked to the strength of the Necessary Optimality Conditions fulfilled by its limit points. Algorithms that generate sequences that converge to points that satisfy strong NOC's are expected to be better than algorithms that converge only to weak-NOC points. A theoretical challenge for every nonlinear-programming algorithm is to prove that it is able to find points that satisfy strong NOC's. Very likely, this property is linked to practical efficiency. As a consequence, researchers concerned with algorithmic reliability are generally interested in the discovery of new strong Necessary Optimality Conditions and in the connection of new and old algorithms to NOC's.

Necessary Optimality Conditions may be first-order or second-order, according to the use of derivatives in its formulation. Higher order optimality conditions are seldom used in practical optimization. First-order NOC's are usually formulated in the following way: "If a feasible point satisfies some First-Order Constraint Qualification (CQ1), then the KKT (Karush-Kuhn-Tucker) conditions hold". In other words, first-order NOC's are propositions of the form:

$$\text{KKT or not-CQ1.}$$

Accordingly, strong Necessary Optimality Conditions correspond to weak Constraint Qualifications. The most popular CQ1 in Nonlinear Programming is the linear independence of the gradients of active constraints (LICQ), usually called *regularity*. A weaker constraint qualification (MFCQ) was given by Mangasarian and Fromovitz [38, 45]. Recently, we proved [4] that the Constant Positive Linear Dependence (CPLD) condition introduced by Qi and Wei [43] is an even weaker constraint qualification. Moreover, the safeguarded augmented Lagrangian algorithms introduced in [1, 2] find points that satisfy the strong first-order NOC that can be expressed as:

$$\text{KKT or not-CPLD.}$$

Second-Order Necessary Optimality Conditions are usually associated to the positive semidefiniteness of the Hessian of the Lagrangian function on a reduced (tangent) subspace. Therefore, points that satisfy Second-Order NOC's are KKT points which, if a Second-Order Constraint Qualification (CQ2) holds, satisfy a reduced positive semidefiniteness property (RSDP). So, a KKT point satisfies a Second-Order NOC if the following proposition is true:

$$\text{RSDP or not-CQ2.}$$

As in the first-order case, one is interested in strong Second-Order NOC's.

Therefore, we aim to find weak Second-Order Constraint Qualifications. Finally, we wish to associate practical algorithms to the fulfillment of the second-order strong optimality conditions discovered. This is the subject of this paper.

In Section 2 we introduce definitions and basic concepts. In Section 3 we prove the main theoretical result: the junction of MFCQ with a weak constant-rank condition is a Second-Order Constraint Qualification. In Section 4 we prove that a second-order variation of the algorithms introduced in [1,2] finds points that satisfy the new second-order optimality condition. In Section 5 we state relationships between Second-Order Constraint Qualifications. Conclusions are given in Section 6.

Notation

If $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)^{\mathrm{T}}$, we denote $\boldsymbol{v}_+ = (\max\{0, \boldsymbol{v}_1\}, \ldots, \max\{0, \boldsymbol{v}_n\})^{\mathrm{T}}$.

We denote $\mathbb{R}_+ = \{t \in \mathbb{R} \mid t \geq 0\}$.

If $K = (k_1, k_2, \ldots) \subset \mathbb{N}$ (with $k_j < k_{j+1}$ for all j), we denote $K \underset{\infty}{\subset} \mathbb{N}$.

If $\boldsymbol{h}(\boldsymbol{x}) = (\boldsymbol{h}_1(\boldsymbol{x}), \ldots, \boldsymbol{h}_m(\boldsymbol{x}))^{\mathrm{T}}$ we denote $\nabla \boldsymbol{h}(\boldsymbol{x}) = (\nabla \boldsymbol{h}_1(\boldsymbol{x}), \ldots, \nabla \boldsymbol{h}_m(\boldsymbol{x}))$.

The symbol $\|\cdot\|$ will denote the Euclidian norm, although many times it may be replaced by an arbitrary vector norm.

2 Definitions

We are concerned with the Nonlinear Programming (NLP) problem

$$\text{Minimize } f(\boldsymbol{x}) \text{ subject to } \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}, \boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m, \boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$ are twice continuosly differentiable on $\mathbb{R}^n$. A point $\boldsymbol{x} \in \mathbb{R}^n$ is said to be *feasible* if it satisfies the constraints $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$. If $\boldsymbol{x}$ is feasible, the equality constraints $\boldsymbol{h}_i(\boldsymbol{x}) = 0$ and the inequality constraints such that $\boldsymbol{g}_i(\boldsymbol{x}) = 0$ are said to be *active* at $\boldsymbol{x}$.

We say that a feasible point $\boldsymbol{x}$ is a KKT point if there exist (Lagrange multipliers) $\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^p_+$ such that

$$\nabla f(\boldsymbol{x}) + \nabla \boldsymbol{h}(\boldsymbol{x})\boldsymbol{\lambda} + \nabla \boldsymbol{g}(\boldsymbol{x})\boldsymbol{\mu} = \boldsymbol{0}, \tag{2}$$

where $\boldsymbol{\mu}_i = 0$ whenever $\boldsymbol{g}_i(\boldsymbol{x}) < 0$.

A First-Order Constraint Qualification is a property of feasible points of a Nonlinear Programming problem such that, when verified at a local minimizer, implies that the local minimizer is a KKT point. The Linear-Independence Constraint Qualification (LICQ), also called *regularity*, says that the gradients of the active constraints at the feasible point $\boldsymbol{x}$ are linearly independent.

Let $I(\boldsymbol{x}) \subset \{1, \ldots, p\}$ be the set of indices of the active inequality constraints at the feasible point $\boldsymbol{x}$. Let $I_1 \subset \{1, \ldots, m\}$, $I_2 \subset I(\boldsymbol{x})$. The subset of gradients of active constraints that correspond to the indices $I_1 \cup I_2$ is said to be *positively linearly dependent* if there exist multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ such that

$$\sum_{i \in I_1} \boldsymbol{\lambda}_i \nabla \boldsymbol{h}_i(\boldsymbol{x}) + \sum_{i \in I_2} \boldsymbol{\mu}_i \nabla \boldsymbol{g}_i(\boldsymbol{x}) = 0, \tag{3}$$

with $\boldsymbol{\mu}_i \geq 0$ for all $i \in I_2$ and $\sum_{i \in I_1} |\boldsymbol{\lambda}_i| + \sum_{i \in I_2} \boldsymbol{\mu}_i > 0$.

Otherwise, we say that these gradients are *positively linearly independent.*

The Mangasarian-Fromovitz Constraint Qualification MFCQ says that, at the feasible point $\boldsymbol{x}$, the gradients of the active constraints are positively linearly independent.

The CPLD Constraint Qualification says that, if a subset of gradients of active constraints is positively linearly dependent at the feasible point $\boldsymbol{x}$ (i.e. (3) holds), then there exists $\varepsilon > 0$ such that the vectors

$$\{\nabla \boldsymbol{h}_i(\boldsymbol{y})\}_{i \in I_1}, \{\nabla \boldsymbol{g}_j(\boldsymbol{y})\}_{j \in I_2}$$

are linearly dependent for all $\boldsymbol{y} \in \mathbb{R}^n$ such that $\|\boldsymbol{y} - \boldsymbol{x}\| \leq \varepsilon$.

Assume that $\boldsymbol{x}$ is a KKT point. Let S be the subspace generated by the gradients of active constraints at $\boldsymbol{x}$ and let T be its orthogonal complement. (If there are no active constraints at $\boldsymbol{x}$, then $S = \{\boldsymbol{0}\}, T = \mathbb{R}^n$.) We say that $\boldsymbol{x}$ satisfies the Weak Reduced Semidefiniteness Property (WRSP) if there exist KKT-multipliers $\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^p$ such that

$$\boldsymbol{d}^{\mathrm{T}}[\nabla^2 f(\boldsymbol{x}) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i \nabla^2 \boldsymbol{h}_i(\boldsymbol{x}) + \sum_{i=1}^{p} \boldsymbol{\mu}_i \nabla^2 \boldsymbol{g}_i(\boldsymbol{x})]\boldsymbol{d} \geq 0 \tag{4}$$

for all $\boldsymbol{d} \in T$.

A stronger second-order optimality condition (Strong-SOC), which is seldom associated to the convergence of practical algorithms [30], is defined in classical textbooks [12, 18, 22, 41]. Let $\bar{T}$ be the set of $\boldsymbol{d} \in \mathbb{R}^n$ such that satisfy the following conditions:

(i) $\boldsymbol{d}^{\mathrm{T}} \nabla \boldsymbol{h}_i(\boldsymbol{x}) = 0$ for all $i = 1, \ldots, m$.
(ii) $\boldsymbol{d}^{\mathrm{T}} \nabla \boldsymbol{g}_i(\boldsymbol{x}) = 0$ for all i such that $\boldsymbol{g}_i(\boldsymbol{x}) = 0$ and $\boldsymbol{\mu}_i > 0$.
(iii) $\boldsymbol{d}^{\mathrm{T}} \nabla \boldsymbol{g}_i(\boldsymbol{x}) \leq 0$ for all i such that $\boldsymbol{g}_i(\boldsymbol{x}) = 0$ and $\boldsymbol{\mu}_i = 0$.

It is said that the feasible point $\boldsymbol{x}$ satisfies Strong-SOC if (4) holds for all $\boldsymbol{d} \in \bar{T}$. Obviously, if the strict complementarity condition holds at $\boldsymbol{x}$, Strong-SOC coincides with WRSP.

In this paper we will deal with the Weak Reduced Semidefiniteness Property. The general question will be: under which conditions on the constraints, local minimizers satisfy WRSP? In other words, which properties of feasible points deserve to be called Second-Order Constraint Qualifications? We already said that LICQ is one of these properties, since Strong-SOC implies WRSP. However, we wish to obtain weaker Second-Order Constraint Qualifications (SOCQ's) (which mean stronger Second-Order NOC's).

A counterexample due to Arutyunov [7] (page 1350) shows that MFCQ is not a Second-Order Constraint Qualification. His counterexample is based on a construction provided in [36](pp. 131-133) and was rediscovered by Anitescu [5]. A simplification of the example was given by Baccari and Trad [10]. Since MFCQ implies CPLD, it turns out that CPLD is not a SOCQ either. In [9,10] it was proved that WRSP holds at a local minimizer $\boldsymbol{x}$ if MFCQ holds and, in addition, the number of linearly independent gradients of active constraints is, at least, $m+q-1$, where q is the number of all active inequality constraints at $\boldsymbol{x}$. A condition for the fulfillment of Strong-SOC is also given in [10].

In the following section, we prove that WRSP also holds under a different condition. To state the new condition we need to define the Weak Constant-Rank (WCR) property.

Assume that $i_1, \dots, i_q$ are the indices of all the active inequality constraints at the feasible point $\boldsymbol{x}$. We say that $\boldsymbol{x}$ satisfies WCR if the matrix $(\nabla \boldsymbol{h}_1(\boldsymbol{y}), \dots, \nabla \boldsymbol{h}_m(\boldsymbol{y}), \nabla \boldsymbol{g}_{i_1}(\boldsymbol{y}), \dots, \nabla \boldsymbol{g}_{i_q}(\boldsymbol{y}))$ has the same rank for all $\boldsymbol{y}$ in a neighborhood of $\boldsymbol{x}$. The WCR condition is weaker than the Constant-Rank condition introduced by Janin [32]. In particular, Janin's condition is a First-Order Constraint Qualification, but WCR is not.

The WCR condition is trivially satisfied at the feasible point $\boldsymbol{x}$ when this point is regular, when all the constraints are linear and when the gradients of active constraints at $\boldsymbol{x}$ span $\mathbb{R}^n$. In addition, if $\boldsymbol{x}$ satisfies the WCR condition for a set of constraints and one adds the constraint $\Phi(\boldsymbol{h}_1(\boldsymbol{x}), \dots, \boldsymbol{h}_m(\boldsymbol{x})) = 0$, for a smooth function $\Phi : \mathbb{R}^m \to \mathbb{R}$ such that $\Phi(\boldsymbol{0}) = 0$, the expanded set of constraints also fulfills WCR. This comes from the fact that the gradient of the new constraint is a linear combination of the gradients of the old constraints. Similar expansions can be made with respect to inequality constraints.

3 New necessary condition

In this section we prove that, if a local minimizer satisfies MFCQ and WCR, then it satisfies the Weak Second-Order Necessary Condition.

All along this section $\boldsymbol{x}^*$ will be a local minimizer of (1). To simplify the notation, we will assume, without loss of generality, that the active inequality constraints at $\boldsymbol{x}^*$ are the first q. Therefore, $\boldsymbol{g}_i(\boldsymbol{x}^*) = 0$ for all $i \le q$ and

$\boldsymbol{g}_i(\boldsymbol{x}^*) < 0$ for all $i > q$. For all $\boldsymbol{x} \in \mathbb{R}^n$ we define

$$T(\boldsymbol{x}) = \{\boldsymbol{d} \in \mathbb{R}^n \mid \nabla \boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{d} = \boldsymbol{0}, \nabla \boldsymbol{g}_i(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{0} = 0, i = 1, \dots, q\}.$$

We also assume that the Weak Constant Rank condition holds at $\boldsymbol{x}^*$. Therefore, for all $\boldsymbol{x}$ in a neighborhood of $\boldsymbol{x}^*$,

$$\text{rank } (\nabla \boldsymbol{h}(\boldsymbol{x}), \nabla \boldsymbol{g}_1(\boldsymbol{x}), \dots, \nabla \boldsymbol{g}_q(\boldsymbol{x})) = r \leq m + q.$$

Lemma 3.1. *Assume that the sequence $\{\boldsymbol{x}^k\} \subset \mathbb{R}^n$ converges to $\boldsymbol{x}^*$ and $\boldsymbol{d} \in T(\boldsymbol{x}^*)$. Then, there exists a sequence $\{\boldsymbol{d}^k\} \subset \mathbb{R}^n$ such that $\boldsymbol{d}^k \in T(\boldsymbol{x}^k)$ and $\lim_{k \to \infty} \boldsymbol{d}^k = \boldsymbol{d}$.*

Proof. Assume, without loss of generality, that the gradients $\{\nabla \boldsymbol{h}_1(\boldsymbol{x}^*), \dots, \nabla \boldsymbol{h}_{r_1}(\boldsymbol{x}^*), \nabla \boldsymbol{g}_1(\boldsymbol{x}^*), \dots, \nabla \boldsymbol{g}_{r_2}(\boldsymbol{x}^*)\}$ are linearly independent and $r = r_1 + r_2$. If $r = 0$ the result is trivial. From now on we assume that $r > 0$. For all $\boldsymbol{x} \in \mathbb{R}^n$ define $\mathbf{A}(\boldsymbol{x}) \in \mathbb{R}^{n \times r}$ by

$$\mathbf{A}(\boldsymbol{x}) = (\nabla \boldsymbol{h}_1(\boldsymbol{x}), \dots, \nabla \boldsymbol{h}_{r_1}(\boldsymbol{x}), \nabla \boldsymbol{g}_1(\boldsymbol{x}), \dots, \nabla \boldsymbol{g}_{r_2}(\boldsymbol{x})).$$

Since the columns of $\mathbf{A}(\boldsymbol{x}^*)$ are linearly independent, by the continuity of the gradients, the columns of $\mathbf{A}(\boldsymbol{x})$ are linearly independent for all $\boldsymbol{x}$ in a neighborhood of $\boldsymbol{x}^*$.

Therefore, by the WCR condition, for all $i = 1, \dots, m,\ j = 1, \dots, q$, $\nabla \boldsymbol{h}_i(\boldsymbol{x})$ and $\nabla \boldsymbol{g}_j(\boldsymbol{x})$ are linear combinations of the columns of $\mathbf{A}(\boldsymbol{x})$ for all $\boldsymbol{x}$ in a neighborhood of $\boldsymbol{x}^*$.

Define $\boldsymbol{d}^k$ as the orthogonal projection of $\boldsymbol{d}$ on $T(\boldsymbol{x}^k)$. Then:

$$\boldsymbol{d}^k = [\mathbf{I} - \mathbf{A}(\boldsymbol{x}^k)(\mathbf{A}(\boldsymbol{x}^k)^{\mathrm{T}} \mathbf{A}(\boldsymbol{x}^k))^{-1} \mathbf{A}(\boldsymbol{x}^k)^{\mathrm{T}}] \boldsymbol{d}.$$

Clearly, $\boldsymbol{d}^k$ is well defined for k large enough and $\mathbf{A}(\boldsymbol{x}^k)^{\mathrm{T}} \boldsymbol{d}^k = \boldsymbol{0}$ for all k. Moreover, taking limits, we obtain:

$$\lim_{k \to \infty} \boldsymbol{d}^k = [\mathbf{I} - \mathbf{A}(\boldsymbol{x}^*)(\mathbf{A}(\boldsymbol{x}^*)^{\mathrm{T}} \mathbf{A}(\boldsymbol{x}^*))^{-1} \mathbf{A}(\boldsymbol{x}^*)^{\mathrm{T}}] \boldsymbol{d}.$$

Since $\boldsymbol{d} \in T(\boldsymbol{x}^*)$, the projection of $\boldsymbol{d}$ on $T(\boldsymbol{x}^*)$ is $\boldsymbol{d}$. Therefore, $\lim_{k \to \infty} \boldsymbol{d}^k = \boldsymbol{d}$ as we wanted to prove. $\square$

For proving Theorem 3.1 below, we will use the following Lemma 3.2.

Lemma 3.2. *Assume that* $F : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{c}_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, p$ *admit continuous second derivatives in a neighborhood of* $\boldsymbol{x}^*$. *Define, for all* $\boldsymbol{x} \in \mathbb{R}^n$, $\Phi(\boldsymbol{x}) = F(\boldsymbol{x}) + \frac{1}{2}\sum_{i=1}^p \boldsymbol{c}_i(\boldsymbol{x})_+^2$. *Suppose that* $\boldsymbol{x}^*$ *is a local minimizer of* Φ. *Define, for all* $\boldsymbol{x} \in \mathbb{R}^n$,

$$\mathbf{H}(\boldsymbol{x}) = \nabla^2 F(\boldsymbol{x}) + \sum_{\boldsymbol{c}_i(\boldsymbol{x}^*) \geq 0} \nabla \boldsymbol{c}_i(\boldsymbol{x}) \nabla \boldsymbol{c}_i(\boldsymbol{x})^{\mathrm{T}} + \sum_{i=1}^p \boldsymbol{c}_i(\boldsymbol{x})_+ \nabla^2 \boldsymbol{c}_i(\boldsymbol{x}).$$

Then, $\mathbf{H}(\boldsymbol{x}^*)$ *is positive semidefinite.*

Proof. Define for all $\boldsymbol{x} \in \mathbb{R}^n$,

$$\Gamma(\boldsymbol{x}) = F(\boldsymbol{x}) + \frac{1}{2} \sum_{\boldsymbol{c}_i(\boldsymbol{x}^*) \geq 0} \boldsymbol{c}_i(\boldsymbol{x})^2.$$

Clearly, $\Gamma(\boldsymbol{x}^*) = \Phi(\boldsymbol{x}^*)$ and $\nabla\Gamma(\boldsymbol{x}^*) = \nabla\Phi(\boldsymbol{x}^*)$. Moreover, Γ has continuous second derivatives and $\nabla^2\Gamma(\boldsymbol{x}^*) = \mathbf{H}(\boldsymbol{x}^*)$.

Suppose, by contradiction, that $\mathbf{H}(\boldsymbol{x}^*)$ is not positive semidefinite. Therefore, $\boldsymbol{x}^*$ is not a local minimizer of Γ. Therefore, for all $\varepsilon > 0$ there exists $\boldsymbol{x} \in \mathbb{R}^n$ such that $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \varepsilon$ and

$$\Gamma(\boldsymbol{x}) < \Gamma(\boldsymbol{x}^*) = \Phi(\boldsymbol{x}^*).$$

But, by the continuity of $\boldsymbol{c}$, in a neighborhood of $\boldsymbol{x}^*$ one has that $\Gamma(\boldsymbol{x}) \geq \Phi(\boldsymbol{x})$. Thus, $\boldsymbol{x}^*$ is not a local minimizer of Φ, contradicting the hypotheses of the lemma. Therefore, $\mathbf{H}(\boldsymbol{x}^*)$ is positive semidefinite, as we wanted to prove. □

Theorem 3.1. *In addition to the general assumptions of this section (including the WCR condition), assume that* $\boldsymbol{x}^*$ *satisfies the Mangasarian-Fromovitz constraint qualification. Then,* $\boldsymbol{x}^*$ *satisfies the Weak Reduced Semidefinite property (4).*

Proof. Since $\boldsymbol{x}^*$ is a local minimizer, there exists $\delta > 0$ such that $\boldsymbol{x}^*$ is a global solution of

$$\text{Minimize } f(\boldsymbol{x}) \text{ subject to } \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \delta.$$

Therefore, $\boldsymbol{x}^*$ is the unique solution of

$$\text{Minimize } f(\boldsymbol{x}) + \frac{1}{4}\|\boldsymbol{x} - \boldsymbol{x}^*\|^4 \text{ subject to } \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \delta.$$

Consider the application of the External Penalty method [20] to this problem. Each subproblem is:

$$\text{Minimize } f(\boldsymbol{x})+\frac{1}{4}\|\boldsymbol{x}-\boldsymbol{x}^*\|^4+\frac{\rho_k}{2}[\|\boldsymbol{h}(\boldsymbol{x})\|^2+\|\boldsymbol{g}(\boldsymbol{x})_+\|^2] \text{ subject to } \|\boldsymbol{x}-\boldsymbol{x}^*\| \leq \delta. \tag{5}$$

Therefore, each subproblem is the minimization of a continuous function in a compact set. This implies that the solution $\boldsymbol{x}^k$ exists and, by the External Penalty theory, when $\rho_k \to \infty$,

$$\lim_{k\to\infty} \boldsymbol{x}^k = \boldsymbol{x}^*.$$

Clearly, $\|\boldsymbol{x}^k - \boldsymbol{x}^*\| < \delta$ for k large enough, so, the gradient of the objective function of (5) vanishes at $\boldsymbol{x}^k$.

Thus, there exists $k_1 \in \mathbb{N}$ such that for all $k \geq k_1$,

$$\nabla f(\boldsymbol{x}^k)+\nabla \boldsymbol{h}(\boldsymbol{x}^k)[\rho_k \boldsymbol{h}(\boldsymbol{x}^k)]+\nabla \boldsymbol{g}(\boldsymbol{x}^k)[\rho_k \boldsymbol{g}(\boldsymbol{x}^k)_+]+\|(\boldsymbol{x}^k-\boldsymbol{x}^*)\|^2(\boldsymbol{x}^k-\boldsymbol{x}^*) = \mathbf{0}. \tag{6}$$

Moreover, by Lemma 3.2, defining $F(\boldsymbol{x}) = f(\boldsymbol{x}) + \frac{1}{4}\|\boldsymbol{x} - \boldsymbol{x}^*\|^4 + \frac{\rho_k}{2}\|\boldsymbol{h}(\boldsymbol{x})\|^2$ and $\boldsymbol{c}(\boldsymbol{x}) = \sqrt{\rho_k}\boldsymbol{g}(\boldsymbol{x})$, we have:

$$\boldsymbol{v}^{\mathrm{T}}\left(\nabla^2 f(\boldsymbol{x}^k) + \rho_k \sum_{i=1}^{m} \boldsymbol{h}_i(\boldsymbol{x}^k)\nabla^2 \boldsymbol{h}_i(\boldsymbol{x}^k) + \rho_k \sum_{i=1}^{p} \boldsymbol{g}_i(\boldsymbol{x}^k)_+ \nabla^2 \boldsymbol{g}_i(\boldsymbol{x}^k)\right)\boldsymbol{v} \tag{7}$$

$$+\rho_k \sum_{i=1}^{m}(\nabla \boldsymbol{h}_i(\boldsymbol{x}^k)^{\mathrm{T}}\boldsymbol{v})^2 + \rho_k \sum_{\boldsymbol{g}_i(\boldsymbol{x}^k)\geq 0} (\nabla \boldsymbol{g}_i(\boldsymbol{x}^k)^{\mathrm{T}}\boldsymbol{v})^2$$

$$+\boldsymbol{v}^{\mathrm{T}}\left[2(\boldsymbol{x}^k - \boldsymbol{x}^*)(\boldsymbol{x}^k - \boldsymbol{x}^*)^{\mathrm{T}} + \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2\mathbf{I}\right]\boldsymbol{v} \geq 0,$$

for all $\boldsymbol{v} \in \mathbb{R}^n$ and k large enough.

Let us define $\boldsymbol{\lambda}^k = \rho_k \boldsymbol{h}(\boldsymbol{x}^k)$, $\boldsymbol{\mu}^k = \rho_k \boldsymbol{g}(\boldsymbol{x}^k)_+$.

It is standard to show that $\boldsymbol{x}^*$ satisfies KKT. We state the argument for the sake of completeness. If the sequence $\{(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)\}$ is unbounded, dividing by $\|(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)\|$ in (6) and taking limits for an appropriate subsequence, we obtain that

$$\nabla \boldsymbol{h}(\boldsymbol{x}^*)\widehat{\boldsymbol{\lambda}}^* + \nabla \boldsymbol{g}(\boldsymbol{x}^*)\widehat{\boldsymbol{\mu}}^* = \mathbf{0}.$$

But, by construction, $\widehat{\boldsymbol{\mu}}^* \geq \mathbf{0}$ and $\widehat{\boldsymbol{\mu}}_i^* = 0$ if $\boldsymbol{g}_i(\boldsymbol{x}^*) < 0$. This implies that $\boldsymbol{x}^*$ does not satisfy MFCQ, which contradicts the hypotheses of the theorem. Therefore, the sequence $\{(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)\}$ is bounded and admits a convergent subsequence. Thus, there exists an infinite subset of indices $K \underset{\infty}{\subset} \mathbb{N}$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ such that $\lim_{k \in K}(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) = (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. Therefore, $\boldsymbol{x}^*$ satisfy the KKT conditions with multipliers $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$.

Let $\boldsymbol{d} \in T(\boldsymbol{x}^*)$. By Lemma 3.1, there exists a sequence $\{\boldsymbol{d}^k\}$ such that $\boldsymbol{d}^k \in \mathbf{T}(\boldsymbol{x}^k)$ and $\lim_{k\to\infty} \boldsymbol{d}^k = \boldsymbol{d}$.

Replacing $\boldsymbol{v}$ by $\boldsymbol{d}^k$ in (7) we obtain:

$$(\boldsymbol{d}^k)^{\mathrm{T}}\left(\nabla^2 f(\boldsymbol{x}^k) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i^k \nabla^2 \boldsymbol{h}_i(\boldsymbol{x}^k) + \sum_{i=1}^{p} \boldsymbol{\mu}_i^k \nabla^2 \boldsymbol{g}_i(\boldsymbol{x}^k)\right)\boldsymbol{d}^k$$

$$+\rho_k \sum_{i=1}^{m} (\nabla \boldsymbol{h}_i(\boldsymbol{x}^k)^{\mathrm{T}}\boldsymbol{d})^2 + \rho_k \sum_{\boldsymbol{g}_i(\boldsymbol{x}^k)\geq 0} (\nabla \boldsymbol{g}_i(\boldsymbol{x}^k)^{\mathrm{T}}\boldsymbol{d}^k)^2$$

$$+(\boldsymbol{d}^k)^{\mathrm{T}}\left[2(\boldsymbol{x}^k - \boldsymbol{x}^*)(\boldsymbol{x}^k - \boldsymbol{x}^*)^{\mathrm{T}} + \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 \mathbf{I}\right]\boldsymbol{d}^k \geq 0$$

for all $k \in K$.

Now, if $\boldsymbol{g}_i(\boldsymbol{x}^*) < 0$, it turns out that $\boldsymbol{g}_i(\boldsymbol{x}^k) < 0$ for all $k \in K$ large enough. Therefore, for $k \in K$ large enough, $\boldsymbol{g}_i(\boldsymbol{x}^k) \geq 0$ implies that $\boldsymbol{g}_i(\boldsymbol{x}^*) = 0$. So, for those k, by the choice of $\boldsymbol{d}^k$, $\boldsymbol{g}_i(\boldsymbol{x}^k) \geq 0$ implies that $\nabla \boldsymbol{g}_i(\boldsymbol{x}^k)^{\mathrm{T}}\boldsymbol{d}^k = 0$. This means that the second term of the left-hand side of the previous inequality vanishes. Thus, for $k \in K$ large enough:

$$(\boldsymbol{d}^k)^{\mathrm{T}}\left(\nabla^2 f(\boldsymbol{x}^k) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i^k \nabla^2 \boldsymbol{h}_i(\boldsymbol{x}^k) + \sum_{i=1}^{p} \boldsymbol{\mu}_i^k \nabla^2 \boldsymbol{g}_i(\boldsymbol{x}^k)\right)\boldsymbol{d}^k$$

$$+(\boldsymbol{d}^k)^{\mathrm{T}}\left[2(\boldsymbol{x}^k - \boldsymbol{x}^*)(\boldsymbol{x}^k - \boldsymbol{x}^*)^{\mathrm{T}} + \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 \mathbf{I}\right]\boldsymbol{d}^k \geq 0.$$

So, taking limits on the last inequality, we get:

$$\boldsymbol{d}^{\mathrm{T}}\left(\nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i^* \nabla^2 \boldsymbol{h}_i(\boldsymbol{x}^*) + \sum_{i=1}^{p} \boldsymbol{\mu}_i^* \nabla^2 \boldsymbol{g}_i(\boldsymbol{x}^*)\right)\boldsymbol{d} \geq 0,$$

as we wanted to prove. □

4 Application to an Augmented Lagrangian Method

In [1, 2] safeguarded Augmented Lagrangian methods were defined that converge to KKT points under the CPLD constraint qualification and exhibit good properties in terms of penalty parameter boundedness. The methods are publicly available in `www.ime.usp.br/~egbirgin/tango`. Let us describe here the application of the main algorithm in [2] to problem (1) with the modification that makes it convergent to second-order stationary points.

We define the Augmented Lagrangian function, [31, 42, 44]:

$$L_\rho(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = f(\boldsymbol{x}) + \frac{\rho}{2}\left\{\sum_{i=1}^{m}\left[\boldsymbol{h}_i(\boldsymbol{x}) + \frac{\boldsymbol{\lambda}_i}{\rho}\right]^2 + \sum_{i=1}^{p}\left[\max\left(0, \boldsymbol{g}_i(\boldsymbol{x}) + \frac{\boldsymbol{\mu}_i}{\rho}\right)\right]^2\right\} \tag{8}$$

for all $\boldsymbol{x} \in \mathbb{R}^n, \rho > 0, \boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^p_+$.

The function L_ρ has continuous first derivatives with respect to $\boldsymbol{x}$, but second derivatives are not defined at the points defined by $\boldsymbol{g}_i(\boldsymbol{x}) + \frac{\boldsymbol{\mu}_i}{\rho} = 0$. We define:

$$\bar{\nabla}^2\left[\max\left(0, \boldsymbol{g}_i(\boldsymbol{x}) + \frac{\boldsymbol{\mu}_i}{\rho}\right)\right]^2 = \nabla^2\left(\boldsymbol{g}_i(\boldsymbol{x}) + \frac{\boldsymbol{\mu}_i}{\rho}\right)^2 \text{ if } \boldsymbol{g}_i(\boldsymbol{x}) + \frac{\boldsymbol{\mu}_i}{\rho} = 0,$$

$$\bar{\nabla}^2\left[\max\left(0, \boldsymbol{g}_i(\boldsymbol{x}) + \frac{\boldsymbol{\mu}_i}{\rho}\right)\right]^2 = \nabla^2\left[\max\left(0, \boldsymbol{g}_i(\boldsymbol{x}) + \frac{\boldsymbol{\mu}_i}{\rho}\right)\right]^2 \text{ otherwise.}$$

Algorithm 4.1

Let $\boldsymbol{\lambda}_{\min} < \boldsymbol{\lambda}_{\max}$, $\boldsymbol{\mu}_{\max} > 0$, $\gamma > 1$, $0 < \tau < 1$. Let $\{\varepsilon_k\}$ be a sequence of nonnegative numbers such that $\lim_{k\to\infty}\varepsilon_k = 0$. Let $\boldsymbol{\lambda}_i^1 \in [\boldsymbol{\lambda}_{\min}, \boldsymbol{\lambda}_{\max}], i = 1,\dots,m$, $\boldsymbol{\mu}_i^1 \in [0, \boldsymbol{\mu}_{\max}], i = 1,\dots,p$, and $\rho_1 > 0$. Let $\boldsymbol{x}^0 \in \Omega$ be an arbitrary initial point. Define $\boldsymbol{V}^0 = g(\boldsymbol{x}^0)_+$. Initialize $k \leftarrow 1$.

Step 1. Find an approximate minimizer $\boldsymbol{x}^k$ of $L_{\rho_k}(\boldsymbol{x}, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$. The conditions for $\boldsymbol{x}^k$ are:

$$\|\nabla L_{\rho_k}(\boldsymbol{x}^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)\| \le \varepsilon_k \tag{9}$$

and

$$\boldsymbol{v}^{\mathrm{T}}\bar{\nabla}^2 L_{\rho_k}(\boldsymbol{x}^k,\boldsymbol{\lambda}^k,\boldsymbol{\mu}^k)\boldsymbol{v} \geq -\varepsilon_k\|\boldsymbol{v}\|^2 \tag{10}$$

for all $\boldsymbol{v}\in\mathbb{R}^n$.
Step 2. Define

$$\boldsymbol{V}_i^k = \max\left\{\boldsymbol{g}_i(\boldsymbol{x}^k), -\frac{\boldsymbol{\mu}_i^k}{\rho_k}\right\}, i=1,\dots,p.$$

If

$$\max\{\|\boldsymbol{h}(\boldsymbol{x}^k)\|_\infty, \|\boldsymbol{V}^k\|_\infty\} \leq \tau\ \max\{\|\boldsymbol{h}(\boldsymbol{x}^{k-1})\|_\infty, \|\boldsymbol{V}^{k-1}\|_\infty\}, \tag{11}$$

define $\rho_{k+1}=\rho_k$. Otherwise, define $\rho_{k+1}=\gamma\rho_k$.
Step 3. Compute $\boldsymbol{\lambda}_i^{k+1}\in[\boldsymbol{\lambda}_{\min},\boldsymbol{\lambda}_{\max}], i=1,\dots,m$ and $\boldsymbol{\mu}_i^{k+1}\in[0,\boldsymbol{\mu}_{\max}], i=1,\dots,p$. Set $k\leftarrow k+1$ and go to Step 1.

Remark. Condition (10) states that $\bar{\nabla}^2 L_{\rho_k}$ must be approximately positive semidefinite at the subproblem solution $\boldsymbol{x}^k$. In other words, the smallest eigenvalue must be greater than ε_k. Very likely, variations of the Trust-Region Newton method [18] for solving the unconstrained subproblem converge to points that satisfy this requirement. However, the precise definition of such an algorithm must be subject of research.

Theorem 4.1. *Assume that the sequence $\{\boldsymbol{x}^k\}$ is generated by Algorithm 4.1 and admits a feasible limit point $\boldsymbol{x}^*$, that satisfies MFCQ and WCR. Then, $\boldsymbol{x}^*$ satisfies KKT and WRSP.*

Proof. Since MFCQ implies CPLD, the fact that $\boldsymbol{x}^*$ satisfies KKT is a consequence of Theorem 4.2 of [2].

By (9) and the definition of L_ρ we have that

$$\begin{aligned}&\|\nabla f(\boldsymbol{x}^k)+\textstyle\sum_{i=1}^m(\boldsymbol{\lambda}_i^k+\rho_k\boldsymbol{h}_i(\boldsymbol{x}^k))\nabla\boldsymbol{h}_i(\boldsymbol{x}^k)+\\ &+\textstyle\sum_{i=1}^p\max\{0,\boldsymbol{\mu}_i^k+\rho_k\boldsymbol{g}_i(\boldsymbol{x}^k)\}\nabla\boldsymbol{g}_i(\boldsymbol{x}^k)\|\leq\varepsilon_k.\end{aligned} \tag{12}$$

To simplify the notation, let us define $\widehat{\boldsymbol{\lambda}}^k = \boldsymbol{\lambda}^k+\rho_k\boldsymbol{h}(\boldsymbol{x}^k)$ and $\widehat{\boldsymbol{\mu}}^k = \max\{0,\boldsymbol{\mu}^k+\rho_k\boldsymbol{g}(\boldsymbol{x}^k)\}$ for all $k\in\mathbb{N}$.

If the sequence $\{(\widehat{\boldsymbol{\lambda}}^k,\widehat{\boldsymbol{\mu}}^k)\}$ is unbounded, dividing by $\|(\widehat{\boldsymbol{\lambda}}^k,\widehat{\boldsymbol{\mu}}^k)\|$ in (12) and taking limits for an appropriate subsequence, we have that there exist scalars

$\widehat{\boldsymbol{\lambda}} \in \mathbb{R}^m, \widehat{\boldsymbol{\mu}} \in \mathbb{R}^p_+$ (not all zero) such that $\widehat{\boldsymbol{\mu}}_i = 0$ if $\boldsymbol{g}_i(\boldsymbol{x}^*) < 0$ and

$$\nabla \boldsymbol{h}(\boldsymbol{x}^*)\widehat{\boldsymbol{\lambda}} + \nabla \boldsymbol{g}(\boldsymbol{x}^*)\widehat{\boldsymbol{\mu}} = 0.$$

This implies that $\boldsymbol{x}^*$ does not verify the MFCQ, contradicting the hypotheses of the theorem. Thus, the sequence $\{(\widehat{\boldsymbol{\lambda}}^k, \widehat{\boldsymbol{\mu}}^k)\}$ is bounded and there exists $K \underset{\infty}{\subset} \mathbb{N}$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ such that $\lim_{k\in K}(\widehat{\boldsymbol{\lambda}}^k, \widehat{\boldsymbol{\mu}}^k) = (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. So $\boldsymbol{x}^*$ satisfies the KKT conditions with multipliers $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$.

Let us prove now that WRSP also takes place.

Let $\boldsymbol{d} \in T(\boldsymbol{x}^*)$. By Lemma 3.1, there exists a sequence $\{\boldsymbol{d}^k\}$ such that $\boldsymbol{d}^k \in T(\boldsymbol{x}^k)$ and $\lim_{k\to\infty} \boldsymbol{d}^k = \boldsymbol{d}$. Replacing $\boldsymbol{h}$ by $\boldsymbol{d}^k$ in (10) and observing that $\boldsymbol{g}_i(\boldsymbol{x}^k) < -\boldsymbol{\mu}_i^k/\rho_k$ implies that $\widehat{\boldsymbol{\mu}}_i^k = 0$, we obtain that

$$-\varepsilon_k \|\boldsymbol{d}^k\|^2 \le (\boldsymbol{d}^k)^{\mathrm{T}} \bar{\nabla}^2 L_{\rho_k}(\boldsymbol{x}^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)\boldsymbol{d}^k$$

$$= (\boldsymbol{d}^k)^{\mathrm{T}} \left(\nabla^2 f(\boldsymbol{x}^k) + \sum_{i=1}^m \widehat{\boldsymbol{\lambda}}_i^k \nabla^2 \boldsymbol{h}_i(\boldsymbol{x}^k) + \sum_{i=1}^p \widehat{\boldsymbol{\mu}}_i^k \nabla^2 \boldsymbol{g}_i(\boldsymbol{x}^k) \right) \boldsymbol{d}^k$$

$$+\rho_k \left[\sum_{i=1}^m (\nabla \boldsymbol{h}_i(\boldsymbol{x}^k)^{\mathrm{T}} \boldsymbol{d}^k)^2 + \sum_{\boldsymbol{g}_i(\boldsymbol{x}^k) \ge -\boldsymbol{\mu}_i^k/\rho_k} (\nabla \boldsymbol{g}_i(\boldsymbol{x}^k)^{\mathrm{T}} \boldsymbol{d}^k)^2 \right]. \tag{13}$$

Let us analyze the case in which $\boldsymbol{g}_i(\boldsymbol{x}^*) < 0$. Then, $\boldsymbol{g}_i(\boldsymbol{x}^k) < 0$ for k large enough. Consider two cases: (a) the sequence $\{\rho_k\}$ is bounded; and, (b) the sequence $\{\rho_k\}$ is unbounded. In the first case, by (11), we have that $\boldsymbol{V}^k \to \boldsymbol{0}$ and, since $\boldsymbol{g}_i(\boldsymbol{x}^*) < 0$, we obtain that $\boldsymbol{g}_i(\boldsymbol{x}^k) < -\boldsymbol{\mu}_i^k/\rho_k$ for k large enough.

Now, consider the case (b). Since $\boldsymbol{\mu}_i^k$ is bounded, we have that $-\boldsymbol{\mu}_i^k/\rho_k \to 0$. Therefore, for k large enough, $\boldsymbol{g}_i(\boldsymbol{x}^*) < 0$ implies that $\boldsymbol{g}_i(\boldsymbol{x}^k) < -\boldsymbol{\mu}_i^k/\rho_k$.

Thus, for k large enough, $\boldsymbol{g}_i(\boldsymbol{x}^k) \ge -\boldsymbol{\mu}_i^k/\rho_k$ implies that $\boldsymbol{g}_i(\boldsymbol{x}^*) = 0$. Therefore, by (13),

$$-\varepsilon_k \|\boldsymbol{d}^k\|^2 \le (\boldsymbol{d}^k)^{\mathrm{T}} \left(\nabla^2 f(\boldsymbol{x}^k) + \sum_{i=1}^m \widehat{\boldsymbol{\lambda}}_i^k \nabla^2 \boldsymbol{h}_i(\boldsymbol{x}^k) + \sum_{i=1}^p \widehat{\boldsymbol{\mu}}_i^k \nabla^2 \boldsymbol{g}_i(\boldsymbol{x}^k) \right) \boldsymbol{d}^k$$

$$+\rho_k \left[\sum_{i=1}^m (\nabla \boldsymbol{h}_i(\boldsymbol{x}^k)^{\mathrm{T}} \boldsymbol{d}^k)^2 + \sum_{\boldsymbol{g}_i(\boldsymbol{x}^*) = 0} (\nabla \boldsymbol{g}_i(\boldsymbol{x}^k)^{\mathrm{T}} \boldsymbol{d}^k)^2 \right]$$

for k large enough. So, by the definition of $\boldsymbol{d}^k$,

$$-\varepsilon_k\|\boldsymbol{d}^k\|^2 \leq (\boldsymbol{d}^k)^{\mathrm{T}}\left(\nabla^2 f(\boldsymbol{x}^k) + \sum_{i=1}^{m} \widehat{\boldsymbol{\lambda}}_i^k \nabla^2 \boldsymbol{h}_i(\boldsymbol{x}^k) + \sum_{i=1}^{p} \widehat{\boldsymbol{\mu}}_i^k \nabla^2 \boldsymbol{g}_i(\boldsymbol{x}^k)\right)\boldsymbol{d}^k.$$

Taking limits for $k \in K$ in the last inequality, we obtain the WRSP condition. □

5 Counterexamples and Conjecture

In this section, we will show that the WCR condition is not a First-Order Constraint Qualification and we will analize the relation between the WCR condition and the Second-Order Constraint Qualification introduced by Baccari and Trad [10]. Recall that the Baccari-Trad (BT) condition holds at a feasible point x if MFCQ holds and the number of linearly independent gradients of active constraints at x is, at least, $m + q - 1$, where q is the number of active inequality constraints. Under an additional complementarity condition, Baccari and Trad [10] proved that Strong-SOC holds at local minimizers. Finally, we will formulate a conjecture in the context of Second-Order Constraint Qualifications.

Counterexample 5.1. *WCR is not a First-Order Constraint Qualification.*
Take $n = 2, m = 1, p = 2,$

$$f(\boldsymbol{x}_1, \boldsymbol{x}_2) = -\boldsymbol{x}_1,$$

$$\boldsymbol{h}_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_2 - \boldsymbol{x}_1^2,$$

$$\boldsymbol{g}_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = -\boldsymbol{x}_1,$$

$$\boldsymbol{g}_2(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_2,$$

$$\boldsymbol{x}^* = (0, 0).$$

The solution of the problem satisfies the WCR condition but it is not a KKT point.

Observe that the number of linearly independent gradients of active constraints at $\boldsymbol{x}^*$ is $2 = m + q - 1$. So, this example can be used to prove that the rank condition introduced by Baccari and Trad in [10] is not a First-Order

Constraint Qualification either.

Counterexample 5.2. *(MFCQ and WCR) do not imply BT.*
Take $n = 2, m = 0, p = 3$,

$$\boldsymbol{g}_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1 - \boldsymbol{x}_2,$$

$$\boldsymbol{g}_2(\boldsymbol{x}_1, \boldsymbol{x}_2) = 2\boldsymbol{x}_1 - 2\boldsymbol{x}_2,$$

$$\boldsymbol{g}_3(\boldsymbol{x}_1, \boldsymbol{x}_2) = 3\boldsymbol{x}_1 - 3\boldsymbol{x}_2,$$

$$\boldsymbol{x}^* = (0, 0).$$

The gradients satisfy the MFCQ condition and the constraints are linear functions so the rank of the active gradients is the same at every feasible point. Thus, $\boldsymbol{x}^*$ satisfies MFCQ and WCR. However, the number of linearly independent gradients of active constraints at $\boldsymbol{x}^*$ is 1, whereas $m + q - 1 = 2$.

Counterexample 5.3. *BT does not imply (MFCQ and WCR).*
Take $n = 2, m = 0, p = 2$,

$$\boldsymbol{g}_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1^2 - \boldsymbol{x}_2, \boldsymbol{g}_2(\boldsymbol{x}_1, \boldsymbol{x}_2) = -\boldsymbol{x}_2, \boldsymbol{x}^* = (0, 0).$$

The gradients satisfy the MFCQ condition and the rank at $\boldsymbol{x}^*$ is 1 but in every neighborhood there are points where the rank is 2. So, $\boldsymbol{x}^*$ does not verify the WCR condition. The number of linearly independent gradients of active constraints at $\boldsymbol{x}^*$ is 1 and $m+q-1 = 1$. So, (BT) is verified.

Counterexamples 5.2 and 5.3 show that the BT Second-Order Constraint Qualification and the constraint qualification that we introduced in this paper are independent. We finish this section formulating a conjecture under the general assumptions of Section 3.

Conjecture. *Assume that MFCQ holds,*

$$\mathit{rank}\ (\nabla \boldsymbol{h}(\boldsymbol{x}^*), \nabla \boldsymbol{g}_1(\boldsymbol{x}^*), \ldots, \nabla \boldsymbol{g}_q(\boldsymbol{x}^*)) = r,$$

and for all x in some neighborhood of $\boldsymbol{x}^*$,

$$\mathit{rank}\ (\nabla \boldsymbol{h}(\boldsymbol{x}), \nabla \boldsymbol{g}_1(\boldsymbol{x}), \ldots, \nabla \boldsymbol{g}_q(\boldsymbol{x})) \leq r + 1.$$

Then, $\boldsymbol{x}^*$ satisfies WRSP.

6 Final Remarks

Necessary Optimality Conditions for Nonlinear Programming are algorithmically relevant because, many times, it is possible to prove that practical algorithms converge to stationary limit points defined by those conditions. It is important to detect new strong optimality conditions and to analyze algorithms with respect to them. In this paper we introduced a new second-order condition that, in formal terms, can be formulated as:

$$\text{WRSP or not-(MFCQ and WCR).} \tag{14}$$

In other words, we proved that, if a KKT point satisfies the Mangasarian-Fromovitz constraint qualification and the Weak Constant Rank condition, then it satisfies the Weak Reduced Semidefiniteness of the Hessian of the Lagrangian. Moreover, we showed that a practical algorithm converges to stationary points in the sense defined by the new condition.

The main result in [10] is that the following proposition is a second-order optimality condition:

$$\text{Strong-SOC or not-BT+,} \tag{15}$$

where BT+ is a condition that involves BT and a complementarity-like assumption.

Although Strong-SOC is obviously stronger than WRSP, we showed that the second-order constraint qualifications *not-(MFCQ and WCR)* and *not-BT* are independent. Therefore, the optimality conditions (14) and (15) are independent as well. For example, take $f(\boldsymbol{x}_1, \boldsymbol{x}_2) = -\boldsymbol{x}_1^2 - \boldsymbol{x}_2^2$ subject to the constraints of Counterexample 5.2. Since BT does not hold at $\boldsymbol{x}^* \equiv (0, 0)$, it turns out that (15) is true. However, this point satisfies MFCQ and WCR and does not satisfy WRSP, so (14) does not hold. Obviously, $\boldsymbol{x}^*$ is not a local minimizer. Let us discuss the practical consequence of this. Assume that we have two nonlinear programming algorithms A and B and that, for Algorithm A, it can be proved that every KKT limit point that satisfies BT+, also fulfills Strong-SOC. Clearly, such an algorithm could converge to the maximizer $\boldsymbol{x}^*$ in the displayed example. On the other hand, assume that for Algorithm B, it can be proved that every KKT limit point that satisfies (MFCQ and WCR) also fulfills WRSP. This algorithm cannot converge to the wrong point $\boldsymbol{x}^*$. (Obviously, examples in the opposite sense can be exhibited, using the constraints of Counterexample 5.3.) In other words, an optimality result associated to

Strong-SOC is stronger than an optimality result associated to WRSP only if its associated Second Order Constraint Qualification is weaker.

It is interesting to comment that Fischer [21] introduced a restricted constant rank condition in the context of local convergence of Sequential Quadratic Programming methods. Fischer's condition involves constant rank of a submatrix of the matrix of gradients of active constraints, characterized by the existence of non-null associated multipliers. This condition is obviously independent of WCR and its relationship with second-order optimality conditions deserves to be investigated.

Much research is expected on the behavior of recent nonlinear programming methods [6, 11, 13–17, 23–29, 37, 39, 40, 46–52] with respect to new (First-Order and Second-Order) optimality conditions. Moreover, related problems that can be transformed in (1) after reformulations (Bilevel [19], Control [35], MPEC [3] and many others) should be revisited in order to verify which structural characteristics help to satisfy optimality conditions.

Finally, recent works by Jeyakumar, Rubinov and Wu [33, 34] point to a promising different direction in the study of necessary optimality conditions. Whereas the classical approach is to study NOC's that are satisfied by *local* minimizers, their approach is to look at optimality conditions that are satisfied by global minimizers but not necessarily by local ones. The algorithmic consequences of these studies should be the development of efficient algorithms with good escaping properties from local-nonglobal optima.

Acknowledgments

We are indebted to three anonymous referees, whose sharp comments helped us to improve the first version of this paper. In particular, we are grateful to the referee that called up our attention to Arutyunov's paper [7]. The other two referees motivated the idea of introducing Lemma 3.2 as a previous stage to the proof of Theorem 3.1. Our first proof of formula (7) used Theorem 3.2 of [8].

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*, Technical Report MCDO-040806 (see `www.ime.usp.br/~egbirgin`), Department of Applied Mathematics, UNICAMP, Brazil, 2004. To appear in Mathematical Programming.

[2] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, *On Augmented Lagrangian Methods with general lower-level constraints*, Technical Report MCDO-101505 (see `www.ime.unicamp.br/~martinez`), Department of Applied Mathematics, UNICAMP, Brazil, 2004. Available in Optimization On-Line.

[3] R. Andreani and J. M. Martínez, *On the solution of mathematical programming problems with equilibrium constraints*, Mathematical Methods of Operations Research 54 (2001) pp. 345–358.

[4] R. Andreani, J. M. Martínez and M. L. Schuverdt, *On the relation between the Constant*

Positive Linear Dependence condition and quasinormality constraint qualification, Journal of Optimization Theory and Applications 125 (2005), pp. 473–485.

[5] M. Anitescu, *Degenerate nonlinear programming with a quadratic growth condition*, SIAM Journal on Optimization 10 (2000) pp. 1116–1135.

[6] M. Argáez and R. A. Tapia, *On the global convergence of a modified augmented Lagrangian linesearch interior-point method for Nonlinear Programming*, Journal of Optimization Theory and Applications 114 (2002), pp. 1–25.

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

[8] A. Auslender, *Penalty methods for computing points that satisfy second order necessary conditions*, Mathematical Programming 17 (1979), pp. 229–238.

[9] A. Baccari, *On the classical necessary second-order optimality conditions*, Journal of Optimization Theory and Applications 123 (2004), pp. 213–221.

[10] A. Baccari and A. Trad, *On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints*, SIAM Journal on Optimization 15 (2004), pp. 394–408.

[11] S. Bakhtiari and A. L. Tits, *A simple primal-dual feasible interior-point method for nonlinear programming with monotone descent*, Computational Optimization and Applications 25 (2003), pp. 17–38.

[12] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts, 1999.

[13] R. H. Byrd, J. Ch. Gilbert and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Mathematical Programming 89 (2000), pp. 149–185.

[14] R. H. Byrd, N. I. M. Gould, J. Nocedal and R. A. Waltz, *An algorithm for nonlinear optimization using linear programming and equality constrained subproblems*, Mathematical Programming 100 (2004), pp. 27–48.

[15] R. H. Byrd, J. Nocedal and A. Waltz, *Feasible interior methods using slacks for nonlinear optimization*, Computational Optimization and Applications 26 (2003), pp. 35–61.

[16] L. Chen and D. Goldfarb, *Interior-Point ℓ_2 penalty methods for nonlinear programming with strong global convergence properties*, CORC Technical Report TR 2004-08, IEOR Department, Columbia University, 2005.

[17] A. R. Conn, N. I. M. Gould, D. Orban and Ph. L. Toint, *A primal-dual trust-region algorithm for nonconvex nonlinear programming*, Mathematical Programming 87 (2000), pp. 215–249.

[18] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *Trust Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[19] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002.

[20] A. V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons, New York, 1968.

[21] A. Fischer, *Modified Wilson's method for nonlinear problems with nonunique multipliers*, Mathematics of Operations Research 24 (1999), pp. 699–727.

[22] R. Fletcher, *Practical Methods of Optimization*, Academic Press, London, 1987.

[23] R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint and A. Wächter, *Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming*, SIAM Journal on Optimization 13 (2002), pp. 635–659.

[24] A. Forsgren, P. E. Gill and M. H. Wright, *Interior point methods for nonlinear optimization*, SIAM Review 44 (2002), pp. 525–597.

[25] E. M. Gertz and P. E. Gill, *A primal-dual trust region algorithm for nonlinear optimization*, Mathematical Programming 100 (2004), pp. 49–94.

[26] P. E. Gill, W. Murray and M. A. Saunders, *SNOPT: An SQP algorithm for large-scale constrained optimization*, SIAM Review 47 (2005), pp. 99–131.

[27] C. C. Gonzaga, E. Karas and M. Vanti, *A globally convergent filter method for Nonlinear Programming*, SIAM Journal on Optimization 14 (2003), pp. 646–669.

[28] N. I. M. Gould, D. Orban, A. Sartenaer and Ph. L. Toint, *Superlinear Convergence of Primal-Dual Interior Point Algorithms for Nonlinear Programming*, SIAM Journal on Optimization 11 (2000), pp. 974–1002.

[29] N. I. M. Gould, D. Orban and Ph. L. Toint, *GALAHAD: a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization*, ACM Transactions on Mathematical Software 29 (2003), pp. 353–372.

[30] N. I. M. Gould and Ph. L. Toint, *A note on the convergence of barrier algorithms for second-order necessary points*, Mathematical Programming 85 (1999), pp. 433–438.

[31] M. R. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications 4 (1969), pp. 303–320.
[32] R. Janin, *Direction derivative of the marginal function in nonlinear programming*, Mathematical Programming Study 21 (1984) pp. 127–138.
[33] V. Jeyakumar, A. M. Rubinov and Z. Y. Wu, *Sufficient global optimality conditions for nonconvex quadratic minimization problems with box constraints*, Journal of Global Optimization 36 (2006) pp. 471–481.
[34] V. Jeyakumar, A. M. Rubinov and Z. Y. Wu, *Nonconvex quadratic minimization problems with quadratic constraints: Global optimality conditions*, Mathematical Programming (2006). To appear.
[35] C. Y. Kaya and J. M. Martínez, *Euler discretization and inexact restoration for Optimal Control*, Technical Report MCDO050502, Department of Applied Mathematics, State University of Campinas, Brazil, 2005. To appear in *Journal of Optimization Theory and Applications.*
[36] E. S. Levitin, A. A. Milyutin and N. P. Osmolovskii, *Higher order conditions for a local minimum in problems with constraints*, Russian Math. Surveys 33 (1978) pp. 97–168.
[37] X. Liu and J. Sun, *A robust primal-dual interior point algorithm for nonlinear programs*, SIAM Journal on Optimization 14 (2004), pp. 1163–1186.
[38] O. L. Mangasarian and S. Fromovitz, *The Fritz-John necessary optimality conditions in presence of equality and inequality constraints*, Journal of Mathematical Analysis and Applications 17 (1967) pp. 37–47.
[39] J. M. Martínez, *Inexact Restoration Method with Lagrangian tangent decrease and new merit function for Nonlinear Programming*, Journal of Optimization Theory and Applications 111 (2001), pp. 39–58.
[40] J. M. Moguerza and F. J. Prieto, *An augmented Lagrangian interior-point method using directions of negative curvature*, Mathematical Programming 95 (2003), pp. 573–616.
[41] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.
[42] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in *Optimization*, R. Fletcher (ed.), Academic Press, New York, NY, pp. 283–298, 1969.
[43] L. Qi and Z. Wei, *On the constant positive linear dependence condition and its application to SQP methods*, SIAM Journal on Optimization 10 (2000), pp. 963–981.
[44] R. T. Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM Journal on Control and Optimization 12 (1974), pp. 268–285.
[45] R. T. Rockafellar, *Lagrange multipliers and optimality*, SIAM Review 35 (1993), pp. 183–238.
[46] D. F. Shanno and R. J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: orderings and high-order methods*, Mathematical Programming 87 (2000), pp. 303–316.
[47] A. L. Tits, A. Wächter, S. Bakhtiari, T. J. Urban and C. T. Lawrence, *A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties*, SIAM Journal on Optimization 14 (2003), pp. 173–199.
[48] P. Tseng, *A convergent infeasible interior-point trust-region method for constrained minimization*, SIAM Journal on Optimization 13 (2002), pp. 432–469.
[49] M. Ulbrich, S. Ulbrich and L. N. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming*, Mathematical Programming 100 (2004), pp. 379–410.
[50] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Technical Report, Optimization Technology Center, Northwestern University, Evanston, IL, 2003, to appear in Mathematical Programming.
[51] R. A. Waltz, J. L. Morales, J. Nocedal and D. Orban, *An interior algorithm for nonlinear optimization that combines line search and trust region steps*, to appear in Mathematical Programming.
[52] H. Yamashita and H. Yabe, *An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization*, SIAM Journal on Optimization 14 (2003), pp. 479–499.