A second-order sequential optimality condition associated to the convergence of optimization algorithms *

Roberto Andreani[†] Gabriel Haeser[‡] Alberto Ramos[§] Paulo J. S. Silva[¶]

October 8, 2015

Abstract

Sequential optimality conditions have recently played an important role on the analysis of the global convergence of optimization algorithms towards first-order stationary points, justifying their stopping criteria. In this paper we introduce a sequential optimality condition that takes into account second-order information and that allows us to improve the global convergence assumptions of several second-order algorithms, which is our main goal. We also present a companion constraint qualification that is less stringent than previous assumptions associated to the convergence of second-order methods, like the joint condition Mangasarian-Fromovitz and Weak Constant Rank. Our condition is also weaker than the Constant Rank Constraint Qualification, which associated it to the convergence of second-order algorithms. This means that we can prove second-order global convergence of well stablished algorithms even when the set of Lagrange multipliers is unbounded, which overcomes a limitation of previous results based on MFCQ. We prove global convergence of well known variations of the augmented Lagrangian and Regularized SQP methods to second-order stationary points under this new weak constraint qualification.

Key words: Nonlinear Programming, Constraint Qualifications, Algorithmic Convergence

1 Introduction

We are concerned with the general nonlinear optimization problem with equality and inequality constraints:

$$\text{Minimize } f(x), \text{ subject to } x \in \Omega, \tag{1.1}$$

where $\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \leq 0\}$ and $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable functions.

Practical algorithms for solving (1.1) are iterative. Hence, their implementations include stopping criteria to decide whether the current point is close to a solution or, at least, whether it verifies approximately a necessary optimality condition. By a necessary optimality condition we mean a computable

*This work was supported by PRONEX-Optimization (PRONEX-CNPq/FAPERJ E-26/111.449/2010-APQ1), CEPID-CeMEAI (FAPESP 2013/07375-0), CNPq (Grants 300907/2012-5, 481992/2013-8, 304618/2013-6, 482549/2013-0, and 303013/2013-3), FAPESP (Grants 2010/19720-5, 2012/20339-0 and 2013/05475-7), CAPES, and National Project on Industrial Mathematics.

[†]Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas, SP, Brazil. Email: andreani@ime.unicamp.br.

[‡]Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil. Email: ghaeser@ime.usp.br.

[§]Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil. Email: aramos27@gmail.com.

[¶]Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas, SP, Brazil. Email: pjssilva@ime.unicamp.br.

condition that must be verified by the minimizer of (1.1) and whose fulfillment indicates that the point under consideration is an acceptable candidate for a solution of the problem.

Necessary optimality conditions are more useful whenever the more they are restrictive, ruling out as many non minimizers as possible. In this case, its fulfillment will be a serious indication that a local minimizer has been found.

The most usual algebraic optimality conditions for (1.1) are associated to the Karush-Kuhn-Tucker (KKT) condition. In fact, many necessary optimality conditions can be stated as "if the description of the constraints at a local minimizer conform to a Constraint Qualification (CQ1), then the KKT condition holds". In other words, many necessary optimality conditions are propositions of the form:

$$\text{KKT or not CQ1.} \tag{1.2}$$

Such conditions use only first-order information of the functions that describe the optimization problem and are then called *first-order* necessary optimality conditions. A condition of this form will be stronger the less stringent is the associated constraint qualification used.

The most used constraint qualification is the Linear Independence Constraint Qualification (LICQ). It states that the gradients of the equality and active inequality constraints are linearly independent at the point of interest. It is interesting due to its many good properties, like uniqueness of the multiplier [48, 36]. It is however very stringent and, hence, the associated optimality condition is weak. There is a vast literature on constraint qualifications weaker than LICQ, see [6, 5, 54, 51] and references therein. We mention two of them. The Mangasarian-Fromovitz condition (MFCQ), defined in [43], says that the gradients of the equality and active inequality constraints are positive linearly independent at the feasible point of interest. The Constant-Rank Constraint Qualification (CRCQ), defined in [41], states that there is a neighborhood around the point of interest where the rank of any subset of the gradients of the equality and active inequality constraints does not change.

In practice, it is usually impossible to find a point that conforms exactly to the KKT condition even if a strong CQ1 holds. Hence, an algorithm may stop when such conditions are satisfied approximately. A sequential optimality condition makes a precise definition based on this practice. Let us consider the most popular of these conditions, the *Approximate KKT* (AKKT) condition introduced in [4]. See also [20, 44, 52].

Definition 1.1. The Approximate-Karush-Kuhn-Tucker (AKKT) optimality condition is said to hold at a feasible point $x^* \in \Omega$ if there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$ and $\{\mu^k\} \subset \mathbb{R}^p_+$, $\{x^k\}$ not necessarily feasible, such that $x^k \to x^*$,

$$\lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0, \tag{1.3}$$

and

$$\mu_j^k = 0 \text{ for } j \notin A(x^*), \tag{1.4}$$

where $A(x^*) = \{i \in \{1, \ldots, p\} \mid g_i(x^*) = 0\}$ is the set of indexes of active inequality constraints at $x^* \in \Omega$.

The attractiveness of sequential optimality conditions is associated to three properties. First, they are genuine necessary optimality conditions, independently of the fulfillment of a constraint qualifications [4, 20]. Second, they are strong, in the sense that they imply the classical first-order optimality condition "KKT or not CQ1" for weak constraint qualifications [5, 6, 9]. Third, there are many practical algorithms that generate sequences whose limit points satisfy them. Particularly in the case of AKKT, many practical optimization algorithms (but not all, see [7]), such as augmented Lagrangian methods, some Sequential Quadratic Programming (SQP) algorithms, interior point methods and inexact restoration methods generate primal-dual sequences $\{x^k, \lambda^k, \mu^k\}$ for which (1.3) and (1.4) are fulfilled [6]. In this case $\{x^k\}$ is called an AKKT sequence and we can say that these methods generate AKKT sequences. We would like to emphasize that this discussion means that sequential optimality conditions, like AKKT, are powerful tools in the global convergence analysis to first-order stationary points, under weak constraint qualifications, of optimization methods. In particular, a new CQ1 that is equivalent to ensure that whenever the point of interest is the limit of an AKKT sequence it is also a KKT point, was recently characterized in [9]. This result relaxed the convergence assumptions of many important algorithms.

Using second order information, one can formulate second-order optimality conditions. Such conditions are usually much stronger than first-order conditions and hence are mostly desirable. Such conditions have been extensively studied in the literature, see [35, 48, 36, 23, 21], with important applications to mathematical programming [50, 13], composite optimization [49], optimal control [22, 25], etc. In practice, most of the second-order necessary optimality conditions used are of the form: "If a local minimizer satisfies some constraint qualification (CQ2), then the WSONC condition holds". That is,

$$\text{WSONC or not CQ2,} \tag{1.5}$$

where WSONC stands for the Weak Second Order Necessary Condition that states that the Hessian of the Lagrangian at a KKT point is positive semidefinite on the subspace orthogonal to the gradients of active constraints, see Definition 2.1. Our focus on necessary optimality conditions of the type "WSONC or not CQ2" comes from algorithmic considerations. To the best of our knowledge, there is not any practical algorithm with global convergence to a point that satisfies a second order stationarity measure stronger than WSONC. In particular, there are not any algorithm that is guaranteed to converge to points where the Hessian of the Lagrangian is positive-semidefinite on the so-called critical cone, instead of on the smaller subspace considered in WSONC. There is also strong evidence that even simple second-order methods will fail to find points conforming to more stringent second-order conditions [39]. Finally, dealing with positive-semidefiniteness over a subspace is much more tractable, from the computational point of view, than dealing with it over a cone.

Several practical algorithms for (1.1) that converge to second-order stationary points (i.e. points where WSONC holds) have been proposed in the literature over the years. Andreani, Birgin, Martínez and Schuverdt [2], see also [8], used a second-order negative-curvature method for box-constrained minimization applied to certain classes of functions that do not possess continuous second derivatives. Byrd, Schnabel and Schultz [24] employ a sequential quadratic programming (SQP) approach, where the second-order stationarity is obtained due to the use of second-order correction steps. Coleman, Liu and Yuan [26] also use a SQP approach with quadratic penalty functions for equality constrained minimization. Conn, Gould, Orban and Toint [27] employ the logarithmic barrier method for inequality constrained optimization with linear equality constraints. Dennis and Vicente [31] use affine scaling directions and the SQP approach for optimization with equality constraints and simple bounds, see also [30]. Di Pillo, Lucidi and Palagi [32] define a primal-dual model algorithm for inequality constrained optimization problems where they take advantage of the equivalence between the original constrained problem and the unconstrained minimization of an exact Augmented Lagrangian function. They use a curvilinear line search technique using information on the nonconvexity of the Augmented Lagrangian function. Facchinei and Lucidi [34] use negative-curvature directions in the context of inequality constrained problems. Recently, Gill, Kungurtsev and Robinson used a variant of the sequential quadratic programming (SQP) method, specifically, the regularized SQP defined in [38], in [37]. Their method is based on performing a flexible line search along a direction formed from the solution of a strictly convex regularized quadratic programming subproblem and, when one exists, a direction of negative curvature for the primal-dual augmented Lagrangian. Morguerza and Prieto [46] employ an interior-point algorithm for non-convex problems and uses directions of negative curvature. The convergence to second-order critical points of trust-region algorithms for convex constraints is studied in details in [28].

Even though all this activity around second order conditions and related algorithms the authors are not aware of any attempt to define a sequential second order optimality condition that can play the same unification role that AKKT and other sequential first order condition can play in the convergence theory of (first-order) algorithms. This is the main purpose of this paper.

We will introduce a sequential second-order optimality condition that we call AKKT2. As every sequential optimality condition it has the associated three main desirable properties. It is a genuine necessary optimality condition (its fulfillment is independent of any constraint qualification). It is also strong in the sense that it implies "WSONC or not CQ2" for a new weak constraint qualification. This new constraint qualification is strictly weaker than the typical condition associated to the convergence of second-order algorithms, namely the joint condition MFCQ and Weak Constant Rank (WCR), see definition 3.2, which was introduced in [8] and used in the analysis of convergence of the second-order augmented Lagrangian proposed in [2] and also the regularized SQP [37]. It is also strictly weaker than the CRCQ condition (or its relaxed version [45]), which proves convergence to a second-order stationary point even when the Lagrange multipliers approximations are unbounded. Finally, we will show that many optimization algorithms with convergence to second-order points generate sequences whose limit points satisfy AKKT2. For instance, we show that the second-order augmented Lagrangian [2], the regularized SQP [37], and the trust-region method of [33] generate AKKT2 sequences, see Section 5. These results indicate that AKKT2 can be used as an unifying tool for global convergence analysis of practical algorithms that converge to second-order stationary points. In particular, we also present the companion CQ2 that fully characterizes the property that a convergent AKKT2 sequence will converge to a point conforming to WSONC, extending the convergence result of algorithms that assumed more stringent constraint qualifications.

We organize the rest of this paper as follows: In Section 2 we survey some basic results and preliminary considerations that will be useful to understand the main results of the paper. In Section 3 we introduce the new sequential second-order optimality condition and we prove that it is a genuine sequential optimality condition, that is, we prove that local minimizers necessarily satisfy it. We also show that it is a strong optimality condition, in the sense that it implies "WSONC or not (MFCQ and WCR)". Finally, we present a practical algorithm that generates sequences whose limit points naturally satisfy this new second-order condition. In Section 4 we refine the results of Section 3 by introducing a new weak constraint qualification associated to AKKT2 and we establish its relationship with other known constraint qualifications as CRCQ and MFCQ+WCR. In Section 5 we present other well-known algorithms with convergence to second-order stationary points that produce sequences whose limit points satisfy our second-order sequential condition. Finally, in Section 6 we give some conclusions and remarks.

2 Basic definitions and preliminary considerations

We denote by $\mathbb{B}$ the closed unit ball in $\mathbb{R}^n$, and $\mathbb{B}(x,\eta) := x+\eta\mathbb{B}$ the closed ball centered at x with radius $\eta > 0$. $\mathbb{R}_+$ is the set of positive scalars, $a^+ := \max\{0, a\}$, the positive part of $a \in \mathbb{R}$. Set $\mathbb{R}_- := -\mathbb{R}_+$. $\mathbb{I}$ denotes the identity matrix of appropriate dimension, e_i denotes the i-th column of $\mathbb{I}$ and $e := \sum e_i$. We use $\langle \cdot, \cdot \rangle$ to denote the Euclidean inner product on $\mathbb{R}^n$, $\|\cdot\|$ the associated norm. $\mathrm{Sym}(n)$ denotes the set of symmetric matrices. $\mathrm{Sym}_+(n)$ stands for the set of order n symmetric positive semidefinite matrices. Given two symmetric matrices A, B in $\mathrm{Sym}(n)$, we write $A \succeq B$ ($A \succ B$) if $A(v, v) \geq B(v, v)$ ($A(v, v) > B(v, v)$) for all $v \in \mathbb{R}^n$, where $A(v, v) := \langle v, Av \rangle$. Finally, we posit $I := \{1, \dots, m\}$.

We state the following well-known lemma for latter reference.

Lemma 2.1. *[18, 29] Let $P \in \mathrm{Sym}(n)$ and vectors $a_1, \dots, a_r \in \mathbb{R}^n$. Define the subspace $\mathcal{C} = \{d \in \mathbb{R}^n : \langle a_j, d \rangle = 0$ for $j \in \{1, \dots, r\}\}$. Suppose that $P(v, v) > 0$ for all $v \in \mathcal{C}$. Then, there exist positive scalars $\{c_j,\ j \in \{1, \dots, r\}\}$ such that $P + \sum_{j=1}^{r} c_j a_j a_j^T \succ 0$.*

Given a set-valued mapping (multifunction) $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$, the *sequential Painleve-Kuratowski outer/upper* *limit* of $F(z)$ as $z \to z^*$ is denoted by

$$\limsup_{z \to z^*} F(z) := \{w^* \in \mathbb{R}^d : \exists \ (z^k, w^k) \to (z^*, w^*) \text{ with } w^k \in F(z^k)\}. \tag{2.1}$$

We say that F is *outer semicontinuous* (osc) at z^* if

$$\limsup_{z \to z^*} F(z) \subset F(z^*). \tag{2.2}$$

Let $L(x, \lambda, \mu)$ be the Lagrangian function associated to (1.1):

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x), \tag{2.3}$$

where $\mu_j \geq 0$ for all $j = 1, \ldots, p$. Under some MFCQ-type [8, 15, 1] or an Abadie-type [16, 1] constraint qualifications, one can prove that a local minimizer x^* of (1.1) fulfills the WSONC condition stated below.

Definition 2.1. A feasible point $x^* \in \Omega$ satisfies the Weak Second-Order Necessary optimality Condition (WSONC) if there exist Lagrange multipliers $\lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^p_+$, $\mu^*_j = 0$ for $j \notin A(x^*)$, such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda^*_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu^*_j \nabla g_j(x^*) = 0 \tag{2.4}$$

and

$$\left(\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda^*_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu^*_j \nabla^2 g_j(x^*) \right) (d, d) \geq 0, \ \forall d \in \mathcal{C}^W(x^*), \tag{2.5}$$

where the weak critical cone $\mathcal{C}^W(x^*)$ is defined as the subspace

$$\mathcal{C}^{\mathcal{W}}(x^*) := \{d \in \mathbb{R}^n : \langle \nabla h_i(x^*), d \rangle = 0, i \in I, \langle \nabla g_j(x^*), d \rangle = 0, j \in A(x^*)\}. \tag{2.6}$$

That is, WSONC holds when the KKT condition holds and the Hessian of the Lagrangian $L(\cdot, \lambda^*, \mu^*)$ is positive semidefinite at x^* over the weak critical cone $\mathcal{C}^{\mathcal{W}}(x^*)$, for some Lagrange multiplier (λ^*, μ^*).

When the weak critical cone is replaced by the usual (strong) critical cone

$$\mathcal{C}^S(x^*) := \left\{ d \in \mathbb{R}^n : \begin{array}{l} \langle \nabla h_i(x^*), d \rangle = 0, i \in I, \langle \nabla g_j(x^*), d \rangle \leq 0, j \in A(x^*) \\ \langle \nabla f(x^*), d \rangle \leq 0 \end{array} \right\}, \tag{2.7}$$

we say that the Strong Second-Order Necessary optimality Condition (SSONC) holds. The SSONC is a well-established condition [16, 19, 35, 36] that holds under the classical LICQ. In fact, one can prove that a local minimizer of (1.1) fulfills SSONC imposing the relaxed constant rank constraint qualification [45, 3], an Abadie-type condition [16, 1], or a MFCQ-type condition [15, 1]. We note that conditions in [1, 45, 3] yield WSONC or SSONC for every Lagrange multiplier, which can be relevant in practical considerations.

However, it is well known that MFCQ by itself is not enough to ensure the validity of SSONC or WSONC [10, 11]. There are other second-order conditions that hold under MFCQ (for instance, see [17, Theorem 3.3] and [23, Theorem 3.45]). These conditions do not suit our practical framework since they require the knowledge of the whole set of Lagrange multipliers in order to be verified, whereas in practice, tipically, only an approximation to a single Lagrange multiplier is available.

Also from the practical point of view, even establishing if SSONC holds is, in general, a NP-hard problem [47] and to our knowledge, no algorithm has been shown to converge to a point at which SSONC holds. In [39] Gould and Toint showed a simple box-constrained optimization problem where the barrier method generates a sequence where SSONC fails to be attained at the limit, while the sequence of barrier minimizers satisfy the second-order sufficient optimality condition.

From the above practical considerations, WSONC is the natural condition to be considered in algorithmic convergence analysis and we focus on optimality conditions that imply it under weak assumptions. The attentive reader may notice that we call an optimality condition *strong* when it implies WSONC under a *weak* constraint qualification, which is not usual in classical second-order analysis.

3 A sequential second-order optimality condition

In this section we will proceed to define a sequential second-order optimality condition, which will play a key role in the convergence analysis of algorithms. As every sequential optimality condition in the literature, it should satisfy three properties: (i) It is an optimality condition, independently of any constraint qualification, (ii) it should be as strong as possible, in our case, it must imply (1.5) for weak constraint qualifications and (iii) it must be possible to verify its validity in sequences generated by practical algorithms.

Definition 3.1. We say that the feasible point $x^* \in \Omega$ is an Approximate Stationary Second-Order point (AKKT2) for the problem (1.1) if there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\eta^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$, $\{\theta^k\} \subset \mathbb{R}^p_+$, $\{\delta_k\} \subset \mathbb{R}_+$ with $\mu^k_j = 0$ for $j \notin A(x^*)$, $\theta^k_j = 0$ for $j \notin A(x^*)$ such that $x^k \to x^*$, $\delta_k \to 0$,

$$\lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^m \lambda^k_i \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu^k_j \nabla g_j(x^k) = 0 \tag{3.1}$$

and

$$\nabla^2_x L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta^k_i \nabla h_i(x^k)\nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta^k_j \nabla g_j(x^k)\nabla g_j(x^k)^T + \delta_k \mathbb{I}, \tag{3.2}$$

is positive semidefinite for $k \in \mathbb{N}$ sufficiently large.

Any sequence $\{x^k\}$ such that every one of its limit points is AKKT2 is called an AKKT2 sequence. The rest of this section is devoted to show that AKKT2 meets the three main properties required by a sequential optimality condition.

3.1 AKKT2 is a necessary optimality condition

In order to prove that AKKT2 is a necessary optimality condition, we will use the next lemmas.

Lemma 3.1. *[2, 8] Let $\bar{f} : \mathbb{R}^n \to \mathbb{R}$, $\bar{g}_j : \mathbb{R}^n \to \mathbb{R}$ for $j \in \{1, \dots, p\}$ functions with continuous second order derivatives in a neighborhood of a point $\bar{x}$. Let us define*

$$\bar{F}(x) := \bar{f}(x) + \frac{1}{2}\sum_{j=1}^p \max\{0, \bar{g}_j(x)\}^2$$

for all x in an open neighborhood of $\bar{x}$. Suppose that $\bar{x}$ is a local minimizer of $\bar{F}$. Then the symmetric matrix defined as

$$H(x) := \nabla^2 \bar{f}(x) + \sum_{j=1}^p \max\{0, \bar{g}_j(x)\}\nabla^2 \bar{g}_j(x) + \sum_{j:\bar{g}(\bar{x})\geq 0} \nabla \bar{g}_j(x)\nabla \bar{g}_j(x)^T$$

is positive semidefinite at $\bar{x}$. Furthermore the quadratic form with gradient $\nabla \bar{F}(\bar{x})$ and Hessian $H(\bar{x})$ is an overestimation of the increment $\bar{F}(x) - \bar{F}(\bar{x})$ in a neighborhood of $\bar{x}$.

The following lemma is an adaptation of the exterior penalty method [35]. See also [4, 19].

Lemma 3.2. *Let $\mathcal{C}$ be a closed subset of $\mathbb{R}^n$, and $\{\rho_k\}$ a positive sequence that tends to infinity. Assume that for all $k \in \mathbb{N}$, x^k is a global minimizer of the mathematical programming problem*

$$\text{Minimize} \;\; f(x) + \rho_k(\sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p} \max\{0, g_j(x)\}^2) \text{ subject to } x \in \mathcal{C}.$$

Then every limit point of $\{x^k\}$ is a global solution of

$$\text{Minimize} \;\; f(x) \;\; \text{subject to } h(x) = 0, \quad g(x) \leq 0, \quad x \in \mathcal{C}. \tag{3.3}$$

Now, we will show that AKKT2 is a necessary optimality condition.

Theorem 3.3. *If x^* is a local minimizer of* (1.1)*, then x^* satisfies the AKKT2 condition.*

Proof. Since x^* is a local minimizer of (1.1) there is a $\varepsilon > 0$ such $f(x^*) \leq f(x)$ for all feasible x such that $\|x - x^*\| \leq \varepsilon$. So x^* is the unique solution of

$$\text{Minimize } f(x) + \frac{1}{4}\|x - x^*\|^4 \;\; \text{subject to } h(x) = 0, \quad g(x) \leq 0, \quad x \in \mathbb{B}(x^*, \varepsilon). \tag{3.4}$$

Let $\{\rho_k\}$ be a sequence of positive scalar with $\rho_k \to \infty$. Consider the penalty method for (3.4).

$$\text{Minimize } f(x) + \frac{1}{4}\|x - x^*\|^4 + \frac{1}{2}\rho_k\{\sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p}(g_j^+(x))^2\} \text{ s.t. } x \in \mathbb{B}(x^*, \varepsilon). \tag{3.5}$$

Let x^k be a global solution of this subproblem (3.5), which is well defined by the compactness of $\mathbb{B}(x^*, \varepsilon)$ and continuity of the functions. Furthermore, by Lemma 3.2, the sequence $\{x^k\}$ converges to x^* and $x^k \in \text{Int } \mathbb{B}(x^*, \varepsilon)$ for k large enough. Then, using Fermat's rule, the gradient of the objective function of (3.5) must vanish at x^k for sufficiently large k:

$$\nabla f(x^k) + \sum_{i=1}^{m} \rho_k h_i(x^k) \nabla h_j(x^k) + \sum_{j=1}^{p} \rho_k g_j^+(x^k) \nabla g_j(x^k) + \|x^k - x^*\|^2(x^k - x^*) = 0. \tag{3.6}$$

By Lemma 3.1 with $\bar{F}(x) = f(x) + \frac{1}{2}\rho_k \sum_{i=1}^{m} h_i(x)^2 + \frac{1}{4}\|x - x^*\|^4$, $\bar{g}_j = \sqrt{\rho_k} g_j(x)$ for $j \in \{1, \ldots, p\}$ and $\bar{x} = x^k$, we can state that:

$$\begin{gathered}
\nabla^2 f(x^k) + \sum_{i=1}^{m} \rho_k h_i(x^k) \nabla^2 h_j(x^k) + \sum_{j=1}^{p} \rho_k g_j^+(x^k) \nabla^2 g_j(x^k) + \\
\sum_{i=1}^{m} \rho_k \nabla h_i(x^k) \nabla h_j(x^k)^T + \sum_{j: g_j(x^k) \geq 0} \rho_k \nabla g_j(x^k) \nabla g_j(x^k)^T + \\
2(x^k - x^*)(x^k - x^*)^T + \|x^k - x^*\|^2 \mathbb{I} \succeq 0.
\end{gathered} \tag{3.7}$$

Define $\lambda_i^k := \rho_k h_i(x^k), \eta_i^k := \rho_k$ for $i \in \{1, \ldots, m\}$, $\mu_j^k := \rho_k \max\{0, g_j(x^k)\}$ for $j \in \{1, \ldots, p\}$, $\theta_j^k = \rho_k$ if $g_j(x^k) \geq 0$ and $\theta_j^k = 0$, otherwise. Finally define $\delta_k := 3\|x^k - x^*\|^2$. Clearly, $\delta_k \to 0$ as k goes to infinity, $\mu_j^k = 0$ for $j \notin A(x^*)$ and $\theta_j^k = 0$ for $j \notin A(x^*)$. Now with these choices and from (3.6) and (3.7), we have that (3.1) and (3.2) are satisfied, which proves that AKKT2 holds. □ □

Remark 1. The notion of AKKT2 has been implicity stated in the optimization literature, in particular, we do not claim that the result from Theorem 3.3 is new. The idea of obtaining second-order information from penalization techniques is very well-known. See, for instance, [35, 19, 14, 12]. The contribution of this paper is to introduce these ideas in the context of sequential optimality conditions, which allows us to improve the global convergence assumptions of several well-known algorithms.

Remark 2. There are typical questions whenever optimality conditions are introduced, specially second-order ones. An important question is whether the necessary condition introduced can be made into a sufficient optimality condition with a small variation, for instance, by replacing positive semidefiniteness by positive definiteness over the same set. Sufficient second-order optimality conditions are very relevant from the theoretical and practical point of view since they are used, for instance, in local convergence and stability analysis. We do not pursue these issues in this paper since they do not play a role in our analysis. Our focus is more on the global convergence of algorithms rather than on optimality conditions. A more profound study of the AKKT2 condition is out of the scope of this paper and will be subject of future research.

3.2 Strength of the AKKT2 condition

The AKKT2 condition is a strong second-order optimality condition in the sense that it implies (1.5) for weak constraint qualifications. In this subsection we will prove that the joint condition MFCQ and Weak Constant Rank (WCR) serves as corresponding CQ2, see Proposition 3.5. In the next section we will show that the *relaxed constant rank constraint qualification* (RCRCQ), a weaker version of the CRCQ, also serves as CQ2 (see Proposition 4.7) and as a consequence the RCRCQ can be used in the global convergence analysis of algorithms. In order to prove that AKKT2 implies "WSONC or not (MFCQ and WCR)", we recall the definition of the WCR condition introduced by Andreani, Martínez and Schuverdt in [8].

Definition 3.2. Let $x^* \in \Omega$ be a feasible point. We say that the Weak Constant Rank condition (WCR) holds if there is a neighborhood V of x^* such that the rank of $\{\nabla h_i(x), \nabla g_j(x) : i \in I, j \in A(x^*)\}$ remains constant for all $x \in V$.

The key property of the WCR condition is the following

Lemma 3.4. *[8] Assume that WCR holds at a feasible point $x^* \in \Omega$. Then, for every $d \in C^{\mathcal{W}}(x^*)$ and for every sequence $\{x^k\} \subset \mathbb{R}^n$ with $x^k \to x^*$, there exists a sequence $\{d^k\} \subset \mathbb{R}^n$ with $d^k \to d$ such that for k sufficiently large, $\langle \nabla h_i(x^k), d^k \rangle = 0$ for $i \in \{1, \ldots, m\}$ and $\langle \nabla g_j(x^k), d^k \rangle = 0$ for $j \in A(x^*)$.*

The next proposition shows that the AKKT2 condition is a strong necessary optimality condition.

Proposition 3.5. *Let $x^* \in \Omega$ be such that AKKT2 holds. If the joint condition MFCQ and WCR holds at x^*, then WSONC is satisfied at x^*.*

Proof. From the definition of AKKT2 there exist sequences $\{x^k\}$, $\{\mu^k\}$, $\{\delta_k\}$, $\{\theta^k\}$ with $\mu_j^k = 0$ and $\theta_j^k = 0$ for $j \notin A(x^*)$ such that $x^k \to x^*$, $\delta_k \to 0$ and

a) $\varepsilon^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \to 0$;

b) $\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\delta_k \mathbb{I}$.

By MFCQ, the sequence $\{(\lambda^k, \mu^k)\}$ is bounded, otherwise, dividing ε^k by $\|(\lambda^k, \mu^k)\|$ and taking limit in a suitable subsequence we get a contradiction. Now, since $\{(\lambda^k, \mu^k)\}$ is bounded, it admits a convergent subsequence, by simplicity we will assume that $\mu^k \to \mu^*$ and $\lambda^k \to \lambda^*$, so $\mu_j^* = 0$ for $j \notin A(x^*)$. Taking limit in item a), we have that x^* satisfies the KKT condition with multipliers μ^* and λ^*.

Now we will prove that WSONC holds in x^* with these multipliers. Take any d in $C^{\mathcal{W}}(x^*)$, by Lemma 3.4, there is a sequence d^k with $d^k \to d$ such that $\langle \nabla h_i(x^k), d^k \rangle = 0$, for $i \in \{1, \ldots, m\}$ and $\langle \nabla g_j(x^k), d^k \rangle = 0$ for $j \in A(x^*)$. Thus, evaluating the quadratic form of item b) at d^k we obtain that

$$\nabla_x^2 L(x^k, \lambda^k, \mu^k)(d^k, d^k) \geq -\delta_k \|d^k\|^2. \tag{3.8}$$

Taking limit in (3.8), we get

$$(\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j^* \nabla^2 g_j(x^*))(d^*, d^*) \geq 0, \tag{3.9}$$

as we wanted to prove. □ □

Clearly, from (3.1), the AKKT condition is implied by the AKKT2 condition, in fact, AKKT2 is actually stronger than the AKKT condition as the following example shows.

Example 3.1 (AKKT2 is stronger than AKKT)**.** Consider $f(x_1, x_2) = -x_1 - x_2$, $g(x_1, x_2) = x_1^2 x_2^2 - 1$ and $x^* = (1, 1)$.

First, let us show that the sequential optimality condition AKKT holds at $x^* = (1, 1)$. In fact, since $\nabla f(x_1, x_2) = (-1, -1)$ and $\nabla g(x_1, x_2) = 2x_1 x_2 (x_2, x_1)$, the KKT condition holds at x^*. Secondly, let us show that AKKT2 fails. Suppose that (3.1) and (3.2) hold. Now, choose d^k as $(x_1^k, -x_2^k)$ where $\{(x_1^k, x_2^k)\}$ is the sequence given by the definition of AKKT2. Since $\left\langle \nabla g(x_1^k, x_2^k), d^k \right\rangle = 0$ for all $k \in \mathbb{N}$ we get from (3.2) that:

$$\mu^k \nabla^2 g(x_1^k, x_2^k)(d^k, d^k) + \delta_k \|d^k\|^2 \geq 0, \tag{3.10}$$

for some $\mu^k \geq 0$ and some positive scalar $\delta_k \to 0$.
Substituting $\nabla^2 g(x_1^k, x_2^k)(d^k, d^k) = -4(x_1^k x_2^k)^2$ into (3.10) we have that for all $k \in \mathbb{N}$, $-4\mu^k (x_1^k x_2^k)^2 + \delta_k \|d^k\|^2 \geq 0$. But this is impossible because by (3.1): $\nabla f(x_1, x_2) + \mu^k \nabla g(x_1, x_2) \to 0$ implies $-1 + 2\mu^k x_1^k (x_2^k)^2 \to 0$ and as a consequence $2\mu^k (x_1^k x_2^k)^2 \to 1$ and $-4\mu^k (x_1^k x_2^k)^2 + \delta_k \|d^k\|^2 \to -2$.

3.3 A practical algorithm that generates AKKT2 sequences

In this subsection we will show that the Augmented Lagrangian algorithm proposed by [2] (see also [8]) generates an AKKT2 sequence. In Section 5 we will show that this is also the case for the regularized SQP method of Gill, Kungurtsev and Robinson [37] and for the trust-region method of Dennis and Vicente [31].

Let us recall the augmented lagrangian method from [2] for problem (1.1), which is equivalent to the proposed in [8], but without box constraints.

Consider the following augmented lagrangian function

$$L_\rho(x, \lambda, \mu) := f(x) + \frac{\rho}{2} \left(\sum_{i=1}^{m} \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{j=1}^{p} \left[\max\{0, g_j(x) + \frac{\mu_j}{\rho}\} \right]^2 \right), \tag{3.11}$$

for all $x \in \mathbb{R}^n, \rho > 0$ and $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}_+^p$. The function L_ρ has continuous first derivatives with respect to x, but second derivatives are not defined at points satisfying $g_j(x) + \mu_j/\rho = 0$. For this reason, an operator $\bar{\nabla}^2$ is defined in [8], that coincides with the second derivative operator ∇^2 at twice differentiable points, and

$$\bar{\nabla}^2 \left(\max\{0, g_i(x) + \frac{\mu_i}{\rho}\} \right)^2 \quad := \quad \nabla^2 \left(g_i(x) + \frac{\mu_i}{\rho} \right)^2 \quad \text{if } g_i(x) + \frac{\mu_i}{\rho} = 0. \tag{3.12}$$

Now we will proceed to analyze Algorithm 1. From (3.15) we have that any limit point of $\{x^k\}$ fulfills the AKKT2 condition. To see this we will take a closer look at [8, Theorem 4.1] that proves global convergence of the algorithm. Let x^* be any limit point of $\{x^k\}$. For k sufficiently large the expression (3.15) is equivalent to

$$\|\nabla L(x^k, \hat{\lambda}^k, \hat{\mu}^k)\| \leq \varepsilon_k, \tag{3.13}$$

and

$$\nabla^2 L(x^k, \hat{\lambda}^k, \hat{\mu}^k) + \rho_k \sum_{i=1}^{m} \nabla h_i(x) \nabla h_i(x)^T + \rho_k \sum_{j \in A(x^*)} \nabla g_j(x) \nabla g_j(x)^T \succeq -\varepsilon_k \mathbb{I}, \tag{3.14}$$

where $\hat{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k)$ for every $i \in \{1, \ldots, m\}$ and $\hat{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}$ for every $j \in \{1, \ldots, p\}$. Moreover, [8, Theorem 4.1] shows that for k large enough $\hat{\mu}_j^k = 0$ for every $j \notin A(x^*)$.

Algorithm 1 [8, Algorithm 4.1]

Let $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $\rho_1 > 0$, $\tau \in (0, 1)$. Let ε_k be a sequence of positive scalars such that $\lim \varepsilon_k = 0$. Let $\lambda_i^1 \in [\lambda_{\min}, \lambda_{\max}], i \in \{1, \ldots, m\}$ and $\mu_j^1 \in [0, \mu_{\max}], j \in \{1, \ldots, p\}$. Let $x^0 \in \mathbb{R}^n$ be an arbitrary initial point. Define $V^0 = \max\{0, g(x^0)\}$. Initialize with $k = 1$.

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

$$\|\nabla L_{\rho_k}(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k \quad \text{and} \quad \bar{\nabla}^2 L_{\rho_k}(x^k, \lambda^k, \mu^k) \succeq -\varepsilon_k \mathbb{I} \tag{3.15}$$

2. Define $V_j^k := \max\{g_j(x^k), -\mu_j^k/\rho_k\}$ for $j \in \{1, \ldots, p\}$.
   If we have $\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\}$ set $\rho_{k+1} = \rho_k$, otherwise, put $\rho_{k+1} = \gamma \rho_k$;
3. Compute $\lambda_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}]$ for all $i \in \{1, \ldots, m\}$ and $\mu_j^{k+1} \in [0, \mu_{\max}]$, $j \in \{1, \ldots, p\}$. Set $k \leftarrow k + 1$ and go to Step 1.

We see from (3.13) and (3.14) that the AKKT2 condition holds at x^*.

4 Second-order global convergence under weak constraint qualifications

The global convergence to second order stationary points of the augmented Lagrangian [8] and regularized SQP [37] is based on the joint assumption of MFCQ and WCR conditions. As we will see in the next section, both methods generate AKKT2 sequences. Hence, a natural question is whether Proposition 3.5 can be proved using weaker constraint qualifications, since this would improve the global convergence theory of every algorithm that generates AKKT2 sequences. In this section we will answer this question affirmatively.

Define for each $x \in \mathbb{R}^n$ the cone

$$C^{\mathcal{W}}(x, x^*) := \{d \in \mathbb{R}^n \colon \langle \nabla h_i(x), d \rangle = 0, i \in I; \langle \nabla g_j(x), d \rangle = 0, j \in A(x^*)\}. \tag{4.1}$$

The set $C^{\mathcal{W}}(x, x^*)$ can be consider as a perturbation of the weak critical cone $C^{\mathcal{W}}(x^*)$ around the feasible point $x^* \in \Omega$. Clearly, $C^{\mathcal{W}}(x, x^*)$ is a subspace and $C^{\mathcal{W}}(x^*, x^*)$ coincides with the weak critical cone $C^{\mathcal{W}}(x^*)$. Using a variational language, we can re-state Lemma 3.4 as WCR implies the inner semicontinuity (isc) of the set-valued mapping $x \rightrightarrows C^{\mathcal{W}}(x, x^*)$ at $x = x^*$, that is, $C^{\mathcal{W}}(x^*, x^*) \subset$

$\liminf_{x\to x^*} C^{\mathcal{W}}(x,x^*)$, in fact, the inner semicontinuity of $C^{\mathcal{W}}(x,x^*)$ at x^* turns out be equivalent to WCR, [53].

Now we will proceed to define the main object of this section. For $x \in \mathbb{R}^n$, denote by $K_2^{\mathcal{W}}(x)$ the following set

$$\bigcup_{\substack{(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^p_+,\\ \mu_j=0 \text{ for } j\notin A(x^*)}} \left\{ \begin{array}{l} (\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j\in A(x^*)} \mu_j \nabla g_j(x), H), \text{ such that} \\ H \preceq \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) + \sum_{j\in A(x^*)} \mu_j \nabla^2 g_j(x) \text{ on } C^W(x,x^*) \end{array} \right\}. \tag{4.2}$$

The set $K_2^{\mathcal{W}}(x)$ is a convex cone included in $\mathbb{R}^n \times \mathrm{Sym}(n)$ and it allows to write the weak second-order necessary condition WSONC in a more compact form, namely,

$$(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^{\mathcal{W}}(x^*). \tag{4.3}$$

The next definition is our new constraint qualification associated to the AKKT2 condition.

Definition 4.1. We say that $x^* \in \Omega$ satisfies the Second-order Cone-Continuity Property CCP2 if the set-valued mapping (multifunction) $x \mapsto K_2^{\mathcal{W}}(x)$, defined in (4.2), is outer semicontinuous at x^*, that is,

$$\limsup_{x\to x^*} K_2^{\mathcal{W}}(x) \subset K_2^{\mathcal{W}}(x^*). \tag{4.4}$$

The CCP2 condition is the weakest condition that can be used to generalize Proposition 3.5 as the next theorem shows.

Theorem 4.1. *Let $x^* \in \Omega$. The conditions below are equivalent:*

- *CCP2 holds at x^*;*
- *For every objective function $f : \mathbb{R}^n \to \mathbb{R}$ of problem* (1.1) *such that AKKT2 holds at x^*, the condition WSONC holds at x^*.*

Proof. First, suppose CCP2 holds at x^* and there is a function f such the AKKT2 is satisfied. From definition of AKKT2, there exist sequences $\{x^k\}$, $\{M^k\}$, $\{\lambda^k\}$, $\{\mu^k\}$, $\{\eta^k\}$, $\{\theta^k\}$, $\{\delta_k\}$, with $\mu^k \geq 0$, $\mu_j^k = 0$ for $j \notin A(x^*)$ and $\theta^k \geq 0$, $\theta_j^k = 0$ for $j \notin A(x^*)$ such that $x^k \to x^*$, $\delta_k \to 0$ and

a) $\varepsilon^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) \to 0$;

b) $\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\delta_k \mathbb{I}$.

From items a) and b), we see that $(-\nabla f(x^k) + \varepsilon^k, -\nabla^2 f(x^k) - \delta_k \mathbb{I}) \in K_2^{\mathcal{W}}(x^k)$. Now using the continuity of $\nabla f(x)$ and $\nabla^2 f(x)$ jointly with the outer semicontinuity of $K_2^{\mathcal{W}}(x)$ at x^* we obtain that $(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^{\mathcal{W}}(x^*)$ and as a consequence WSONC holds. Now let us prove the other implication. Let (w, W) be an element of $\limsup K_2^{\mathcal{W}}(x)$ when $x \to x^*$. We will show that (w, W) is in $K_2^{\mathcal{W}}(x^*)$. By definition of outer limit, we have that there are sequences $\{x^k\}$, $\{\lambda_i^k\}$, $\{\mu_j^k\}$ with $\mu_j^k = 0$ for $j \notin A(x^*)$ and $\{H^k\} \subset \mathrm{Sym(n)}$ such that $x^k \to x^*$,

$$(\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k g_j(x^k), H^k) \to (w, W)$$

and

$$H^k \preceq \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla^2 g_j(x^k) \text{ over the set } C^{\mathcal{W}}(x^k, x^*).$$

Define the following function:

$$f(x) := -\langle w, x - x^* \rangle - \frac{1}{2} W(x - x^*, x - x^*).$$

We will show that AKKT2 holds at x^* with $f(x)$ as the objective function. Clearly, we have that $\nabla f(x) = -w - W(x - x^*)$ and $\nabla^2 f(x) = -W$. To prove (3.1), it is enough to see that $\lim_{k\to\infty} \nabla_x L(x^k, \lambda^k, \mu^k) = 0$, but this is trivial, since that limit is equal to

$$\lim_{k\to\infty} \left(\sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k g_j(x^k) - w \right) - \lim_{k\to\infty} W(x^k - x^*) = 0.$$

To prove that (3.2) holds we will use Lemma 2.1 with

$$P^k := \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla^2 g_j(x^k) - H^k + \frac{1}{k}\mathbb{I}, \tag{4.5}$$

and a_i are the columns of the matrix $[\nabla h_i(x^k), i \in I; \nabla g_j(x^k), j \in A(x^*)]$. By Lemma 2.1 there are positive sequences $\{\theta^k\}$ and $\{\eta^k\}$ such that:

$$S^k := P^k + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succ 0. \tag{4.6}$$

Put $\theta_j^k = 0$ for $j \notin A(x^*)$. Using (4.5), (4.6) and $\nabla^2 f(x) = -W$, we get

$$\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \tag{4.7}$$

is equal to $-W + H^k + S^k - \frac{1}{k}\mathbb{I}$. Now, we will proceed to find a lower bound for this matrix. By Rayleigh's principle we have

$$-W + H^k \succeq -|\lambda_1(W - H^k)|\mathbb{I},$$

where $\lambda_1(W - H^k)$ denotes the smallest eigenvalue of $W - H^k$. By (4.6), $S^k \succ 0$, so we have that

$$-W + H^k + S^k - \frac{1}{k}\mathbb{I} \succeq -|\lambda_1(W - H^k)|\mathbb{I} + S^k - \frac{1}{k}\mathbb{I} \succeq -|\lambda_1(W - H^k)|\mathbb{I} - \frac{1}{k}\mathbb{I} = -\delta_k \mathbb{I}, \tag{4.8}$$

where $\delta_k := |\lambda_1(W - H^k)| + 1/k$. Since $H^k \to W$ as k tends to infinity, δ_k tends to zero. From (4.8) and (4.7), we see that condition (3.2) holds, therefore x^* is an AKKT2 point. Then by hypothesis, WSONC holds, and by (4.3), $(w, W) = (-\nabla f(x^*), -\nabla^2 f(x^*))$ belongs to $K_2^{\mathcal{W}}(x^*)$ as we wanted to prove. □ □

Since AKKT2 is a necessary optimality condition, by Theorem 4.1, we have:

Corollary 4.2. *If x^* is a local minimizer of* (1.1) *such that CCP2 holds, then WSONC holds.*

By Proposition 3.5 and Theorem 4.1 we have:

Corollary 4.3. *The joint condition MFCQ and WCR implies CCP2.*

CCP2 is strictly weaker than MFCQ and WCR as the next example shows.

Example 4.1 (CCP2 does not imply MFCQ and WCR)**.** Consider in $\mathbb{R}$ the vector $x^* = 0$ and the inequality constraints defined by the functions $g_1(x) = x$ and $g_2(x) = -x$. Then, CCP2 holds at x^* but MFCQ does not (as a consequence MFCQ+WCR fails).

Let us compute the cone $K_2^{\mathcal{W}}(x)$ for every $x \in \mathbb{R}$. From direct calculations, $\nabla g_1(x) = 1, \nabla^2 g_1(x) = 0, \nabla g_2(x) = -1$ and $\nabla^2 g_2(x) = 0$. Thus, we have $C^{\mathcal{W}}(x, x^*) = \{0\}$. From this, every $H \in \mathrm{Sym}(1) = \mathbb{R}$ satisfies

$$H \preceq \mu_1 \nabla^2 g_1(x) + \mu_2 \nabla^2 g_2(x) = 0 \text{ on } C^{\mathcal{W}}(x, x^*) = \{0\} \text{ for any } \mu_1, \mu_2 \geq 0.$$

Then, we get $K_2^{\mathcal{W}}(x) = \mathbb{R} \times \mathbb{R}$, $x \in \mathbb{R}$, and subsequently $K_2^{\mathcal{W}}$ is osc on $\mathbb{R}$.

Another constraint qualification of the MFCQ-type that yields WSONC is the following one introduced by Baccari and Trad, [15]: The Baccari-Trad condition holds at $x^* \in \Omega$ if MFCQ holds and the rank of the active constraints is at most one less than the number of active constraints. Although the Baccari-Trad condition, as CCP2, guarantees the fullfilment of WSONC at a local minimizer, these conditions are not related.

Example 4.2 (Baccari-Trad condition does not imply CCP2)**.** Consider in $\mathbb{R}^2$ the vector $x^* = (0, 0)$ and the inequality constraints defined by the functions $g_1(x_1, x_2) = -x_2$ and $g_2(x_1, x_2) = x_1^2 - x_2$. Then, Baccari-Trad holds at x^* but CCP2 fails.

Clearly, Baccari-Trad holds at x^*. To see that CCP2 fails, it is enough to compute the cones $C^{\mathcal{W}}(x, x^*)$ around x^*. By direct calculations, $C^{\mathcal{W}}(x^*, x^*) = \mathbb{R} \times \{0\}$ and $C^{\mathcal{W}}(x, x^*) = \{(0, 0)\}, x_1 \neq x_1^*$, so $K_2^{\mathcal{W}}(x) = \mathbb{R}_+ \times \mathbb{R}_- \times \mathrm{Sym}(2)$ for every x such that $x_1 \neq x_1^*$ but, $K_2^{\mathcal{W}}(x^*)$ is a proper subset of $\{0\} \times \mathbb{R}_- \times \mathrm{Sym}(2)$. Then, CCP2 fails.

To see that CCP2 does not imply the Baccari-Trad condition, it is enough to see that MFCQ+WCR implies CCP2 while not implying the Baccari-Trad condition, see [8, counterexample 5.2].

The independence of CCP2 and the Baccari-Trad condition has practical implications. Due to Theorem 4.1, the Baccari-Trad condition is not enough to guarantee that a limit point of an AKKT2 sequence satisfies WSONC. See the next example.

Example 4.3 (AKKT2 under the Baccari-Trad condition does not imply WSONC)**.** Consider the optimization problem:

$$\text{Minimize } f(x_1, x_2) = -2x_1^2 \text{ s.t. } g_1(x_1, x_2) = -x_2 \leq 0, g_2(x_1, x_2) = x_1^2 - x_2 \leq 0. \tag{4.9}$$

By Example 4.2, Baccari-Trad holds at $x^* = (0, 0)$. To show that x^* is an AKKT2 point, choose $x_1^k := 1/k$, $x_2^k := x_1^k$, $\mu_1^k := 0$, $\mu_2^k := 0$, $\theta_2^k := 2(x_1^k)^{-2}$, $\theta_1^k := 2\theta_2^k$ and $\delta_k := 0$. With theses multipliers, we have $\nabla f(x^k) + \mu_1^k \nabla g_1(x^k) + \mu_2^k \nabla g_2(x^k) \to (0, 0)$ and

$$\nabla^2 L(x^k, \mu^k) + \theta_1^k \nabla g_1(x^k) \nabla g_1(x^k)^T + \theta_2^k \nabla g_2(x^k) \nabla g_2(x^k)^T = \begin{pmatrix} 4 & -2\theta_2^k x_2^k \\ -2\theta_2^k x_2^k & 3\theta_2^k \end{pmatrix},$$

where the last matrix is positive semidefinite. Also, by direct calculation we have that WSONC fails and $x^* = (0, 0)$ is not an optimal solution. So, in this example, we have a point $x^* = (0, 0)$ which is not an optimal solution neither satisfies WSONC, but can be achieved by an AKKT2 sequence (perhaps, generated by an augmented lagrangian method or a regularized SQP method) and as a consequence accepted as a candidate solution. This cannot happens if instead of the Baccari-Trad condition, we consider any other constraint qualification which implies CCP2.

Another constraint qualification weaker than LICQ is the Constant-Rank Constraint Qualification (CRCQ), cf. [41]. Let us recall the definition of CRCQ. We say that a feasible point $x^* \in \Omega$ verifies CRCQ if there exists a neighbourhood of x^* in which the rank of any subset of the gradients of equality and active inequality constraints does not change in a neighborhood. In [3], it was proved that under CRCQ, a local minimizer conforms to SSONC for every Lagrange multiplier. A relaxed version of the CRCQ has been defined in [45], called Relaxed-CRCQ (RCRCQ) which also enjoys the same second-order property [3, 5]. In RCRCQ, fewer subsets should conform to the constant rank property, namely, subsets that include gradients of every equality constraints. We will prove that RCRCQ is strictly stronger than CCP2. Let us consider the following lemmas. The first is a result from the classical constant rank theorem from analysis (cf. [55, Theorem 2.13]).

Lemma 4.4. *Assume that the gradients* $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ *have locally constant rank in a neighborhood of some* $x \in \mathbb{R}^n$*. Then for each* $d \in \mathbb{R}^n$ *such that*

$$\langle \nabla h_i(x), d\rangle = 0 \quad \text{for } i \in \mathcal{I} \text{ and } \langle \nabla g_j(x), d\rangle = 0 \quad \text{for } j \in \mathcal{J}, \tag{4.10}$$

there exists some curve $t \to \phi(t)$*,* $t \in (-T, T)$*,* $T > 0$ *twice differentiable such that* $\phi(0) = x$*,* $\phi'(0) = d$ *and for every* $i \in \mathcal{I}$ *and* $j \in \mathcal{J}$ *we have* $h_i(\phi(t)) = h_i(x)$ *and* $g_j(\phi(t)) = g_j(x)$ $\forall t \in (-T, T)$*.*

The next lemma is a variation of Carathéodory's lemma.

Lemma 4.5. *[5, Lemma 1] Suppose that* $v = \sum_{i\in\mathcal{I}} \alpha_i p_i + \sum_{j\in\mathcal{J}} \beta_j q_j$ *with* $p_i, q_j \in \mathbb{R}^n$ *for every* $i \in \mathcal{I}, j \in \mathcal{J}$*,* $\{p_i\}_{i\in\mathcal{I}}$ *are linearly independent and* α_i, β_j *are nonzero for every* $i \in \mathcal{I}, j \in \mathcal{J}$*. Then there is a subset* $\mathcal{J}' \subset \mathcal{J}$ *and scalars* $\hat{\alpha}_i, \hat{\beta}_j$ *for all* $i \in \mathcal{I}$*,* $j \in \mathcal{J}'$ *such that*

- $v = \sum_{i\in\mathcal{I}} \hat{\alpha}_i p_i + \sum_{j\in\mathcal{J}'} \hat{\beta}_j q_j$*;*
- *For every* $j \in \mathcal{J}'$ *we have* $\beta_j \hat{\beta}_j > 0$*;*
- $\{p_i, q_j\}_{i\in\mathcal{I}, j\in\mathcal{J}'}$ *is a linearly independent set.*

A useful characterization of RCRCQ is given below:

Theorem 4.6. *[5, Theorem 1] Let* $I \subset \{1, \ldots m\}$ *be an index set such that* $\{\nabla h_i(x) : i \in I\}$ *is a linear basis for span*$\{\nabla h_i(x) : i \in \{1, \ldots, m\}\}$*. A feasible point* $x \in \Omega$ *satisfies RCRCQ if, and only if, there is a neighborhood* V *of* x *such that:*

a) $\{\nabla h_i(y) : i \in \{1, \ldots, m\}\}$ *has the same rank for every* $y \in V$*;*

b) For every $J \subset A(x)$*, if the set* $\{\nabla h_i(x), \nabla g_j(x) : i \in I, j \in J\}$ *is linearly dependent, then* $\{\nabla h_i(y), \nabla g_j(y) : i \in I, j \in J\}$ *is linearly dependent for every* $y \in V$*.*

We are ready to prove the following:

Proposition 4.7. *RCRCQ implies CCP2.*

Proof. Let (w, W) be an element of $\limsup K_2^{\mathcal{W}}(x)$ when $x \to x^*$. By definition of outer limit, we have that there are sequences $\{x^k\}$, $\{\lambda_i^k\}$, $\{\mu_j^k\}$ with $\mu_j^k = 0$ for $j \notin A(x^*)$ and $\{H^k\}$ such that $x^k \to x^*$,

$$w^k := \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) \to w \text{ and } H^k \to W, \tag{4.11}$$

where $H^k \preceq \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla^2 g_j(x^k)$ over $C^{\mathcal{W}}(x^k, x^*)$.
Take an index subset $\mathcal{I} \subset \{1, \ldots, m\}$ such that the gradients $\{\nabla h_i(x^*) : i \in \mathcal{I}\}$ form a linear basis for the subspace generated by $\{\nabla h_i(x^*) : i \in \{1, \ldots, m\}\}$. From continuity $\{\nabla h_i(x^k) : i \in \mathcal{I}\}$ is linearly independent for k large enough. By Theorem 4.6 item a), we have that $\{\nabla h_i(x^k) : i \in \mathcal{I}\}$ is a linear basis for the subspace generated by $\{\nabla h_i(x^k) : i \in \{1, \ldots, m\}\}$, for k sufficiently large. Then, there is a sequence $\{\bar{\lambda}_i^k : i \in \mathcal{I}\} \subset \mathbb{R}$ such that $\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) = \sum_{i\in\mathcal{I}} \bar{\lambda}_i^k \nabla h_i(x^k)$. So, we may write:

$$w^k = \sum_{i\in\mathcal{I}} \bar{\lambda}_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k). \tag{4.12}$$

Applying Lemma 4.5 to the expression above, we find a subset $\mathcal{J}_k \subset A(x^*)$ and multipliers $\hat{\lambda}_i^k \in \mathbb{R}, i \in \mathcal{I}$ and $\hat{\mu}_j^k \in \mathbb{R}_+, j \in \mathcal{J}_k$ for k large enough such that

$$w^k = \sum_{i\in\mathcal{I}} \hat{\lambda}_i^k \nabla h_i(x^k) + \sum_{j\in\mathcal{J}_k} \hat{\mu}_j^k \nabla g_j(x^k), \tag{4.13}$$

and $\{\nabla h_i(x^k), \nabla g_j(x^k) : i \in \mathcal{I}, j \in \mathcal{J}_k\}$ is a linearly independent set. Since $A(x^*)$ is a finite index set, we may take $\mathcal{J} := \mathcal{J}_k$ for an appropriate subsequence. By RCRCQ (Theorem 4.6, item b), we have that $\{\nabla h_i(x^*), \nabla g_j(x^*) : i \in \mathcal{I}, j \in \mathcal{J}\}$ is a linearly independent set and as a consequence $\{\hat{\lambda}_i^k, \hat{\mu}_j^k : i \in \mathcal{I}, j \in \mathcal{J}\}_{k\in\mathbb{N}}$ formd a bounded sequence, so we can assume, without loss of generality, that $\hat{\lambda}_i^k \to \lambda_i$ and $\hat{\mu}_j^k \to \mu_j$. Taking limit in (4.13) we get $w = \sum_{i\in\mathcal{I}} \lambda_i \nabla h_i(x^*) + \sum_{j\in\mathcal{J}} \mu_j \nabla g_j(x^*)$. Define $\hat{\lambda}_i^k = 0$ for $i \notin \mathcal{I}$ and $\hat{\mu}_j^k = 0$ for $j \notin \mathcal{J}$ for every $k \in \mathbb{N}$, also define $\lambda_i = 0$ for $i \notin \mathcal{I}$ and $\mu_j = 0$ for $j \notin \mathcal{J}$. Now we will prove that for every $d \in C^{\mathcal{W}}(x^*)$ the following inequality holds: $H(d,d) \le \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*)(d,d) + \sum_{j\in A(x^*)} \mu_j \nabla^2 g_j(x^*)(d,d)$.

Define for every $k \in \mathbb{N}$, $\Lambda_i^k := \lambda_i^k - \hat{\lambda}_i^k$, $\Upsilon_j^k := \mu_j^k - \hat{\mu}_j^k$. From (4.13) and (4.11) we obtain

$$\sum_{i=1}^m \Lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \Upsilon_j^k \nabla g_j(x^k) = 0 \quad \text{for } k \in \mathbb{N} \text{ large enough.} \tag{4.14}$$

Take $d \in C^{\mathcal{W}}(x^*)$. Since RCRCQ implies the WCR condition, we have that there is a sequence $d^k \to d$ such that $d^k \in C^{\mathcal{W}}(x^k, x^*)$, given by Lemma 3.4. Thus

$$\begin{aligned} H^k(d^k,d^k) &\le \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k)(d^k,d^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla^2 g_j(x^k)(d^k,d^k) & (4.15)\\ &\le \sum_{i=1}^m \hat{\lambda}_i^k \nabla^2 h_i(x^k)(d^k,d^k) + \sum_{j\in A(x^*)} \hat{\mu}_j^k \nabla^2 g_j(x^k)(d^k,d^k) + \Xi^k & (4.16)\end{aligned}$$

where

$$\Xi^k := \sum_{i=1}^m \Lambda_i^k \nabla^2 h_i(x^k)(d^k,d^k) + \sum_{j\in A(x^*)} \Upsilon_j^k \nabla^2 g_j(x^k)(d^k,d^k). \tag{4.17}$$

By RCRCQ, we have that for k sufficiently large, x^k has a neighborhood where the rank of $\{\nabla h_i(x^k), \nabla g_j(x^k) : i \in \{1,\dots,m\}, j \in A(x^*)\}$ is locally constant, so by Lemma 4.4, there is an arc $t \to \phi_k(t)$ for $t \in (-T_k, T_k)$, $T_k > 0$ with $\phi_k(0) = x^k$, $\phi_k'(0) = d^k$ such that $h_i(\phi_k(t)) = h_i(x^k)$ for every $i \in \{1,\dots,m\}$ and $g_j(\phi_k(t)) = g_j(x^k)$ for every $j \in A(x^*)$. Defining $v^k = \phi_k''(0)$ and differentiating $h_i(\phi_k(t)) = h_i(x^k)$, $i \in \{1,\dots,m\}$ and $g_j(\phi_k(t)) = g_j(x^k)$, $j \in A(x^*)$ twice at $t = 0$, we obtain:

$$\begin{aligned} \langle \nabla h_i(x^k), v^k\rangle + \nabla^2 h_i(x^k)(d^k,d^k) &= 0, \text{ for } i \in \{1,\dots,m\}, & (4.18)\\ \langle \nabla g_j(x^k), v^k\rangle + \nabla^2 g_j(x^k)(d^k,d^k) &= 0, \text{ for } j \in A(x^*). & (4.19)\end{aligned}$$

So, replacing the expressions (4.18) and (4.19) into (4.17) we get

$$\begin{aligned} \Xi^k &= -\sum_{i=1}^m \Lambda_i^k \langle \nabla h_i(x^k), v^k\rangle - \sum_{j\in A(x^*)} \Upsilon_j^k \langle \nabla g_j(x^k), v^k\rangle & (4.20)\\ &= -\langle \sum_{i=1}^m \Lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \Upsilon_j^k \nabla g_j(x^k), v^k\rangle = 0, & (4.21)\end{aligned}$$

where in the last equality we have used (4.14). Now, since $\Xi^k = 0$ for every k sufficiently large, we have that (4.16) becomes

$$H^k(d^k,d^k) \le \sum_{i=1}^m \hat{\lambda}_i^k \nabla^2 h_i(x^k)(d^k,d^k) + \sum_{j\in A(x^*)} \hat{\mu}_j^k \nabla^2 g_j(x^k)(d^k,d^k). \tag{4.22}$$

Taking limit in (4.22), the assertion is proved. □ □

Example 4.4 (RCRCQ is strictly stronger than CCP2)**.**

In $\mathbb{R}^2$, consider $x^* = (0,0)$ and the following equality and inequality constraints:

$$\begin{array}{rcl} h_1(x_1,x_2) & = & x_1; \\ g_1(x_1,x_2) & = & -x_1^2 + x_2; \\ g_2(x_1,x_2) & = & -x_1^2 + x_2^3. \end{array}$$

We have $\nabla h_1(x_1,x_2) = (1,0)$, $\nabla g_2(x_1,x_2) = (-2x_1,1)$ and $\nabla g_3(x_1,x_2) = (-2x_1, 3x_2^2)$. From this, RCRCQ fails at $x^* = (0,0)$. Now, since $C^{\mathcal{W}}(x,x^*) = \{0\}$, we get $K_2^{\mathcal{W}}(x) = \mathbb{R} \times \mathbb{R}_+ \times Sym(2)$. Clearly, $K_2^{\mathcal{W}}(x)$ is osc on $\mathbb{R}^2$.

Remark 3. We just proved that CCP2 is a constraint qualification that yields WSONC at a local minimizer, that is weaker than the joint condition MFCQ+WCR and weaker than RCRCQ, furthermore, this condition is the minimal one to guarantee that every AKKT2 point fulfills WSONC, as proved in Theorem 4.1. This improves second-order global convergence results of algorithms that generate AKKT2 sequences in the sense that only CCP2 could be assumed. Even the weaker result under RCRCQ was not previously known. From the results presented above, it is possible to guarantee the global convergence to second-order stationary points for every algortihm that generates AKKT2 sequences, even when the sequence of approximate Lagrange multipliers generated by it is unbounded.

Now, suppose that we want a condition which guarantees that every limit point of any AKKT2 sequence fulfills not only WSONC but also the strong second order condition, SSONC, even though, to the best of our knowledge, no algorithm has been shown to converge to a point where the SSONC holds. With this in mind we shall define the next constraint qualification in the spirit of Theorem 4.1, replacing WSONC with SSONC. Our goal is to understand why practical algorithms are not expected to converge to a point fulfilling SSONC.

Definition 4.2. We say that the Strong CCP2 (SCCP2) holds at $x^* \in \Omega$ if

$$\limsup_{x \to x^*} K_2^{\mathcal{W}}(x) \subset K_2^{\mathcal{S}}(x^*),$$

where $K_2^{\mathcal{S}}(x^*)$ is the cone associated to the critical cone $C^{\mathcal{S}}(x^*,\mu)$, that is, the cone:

$$\bigcup_{\substack{(\lambda,\mu) \in \mathbb{R}^m \times \mathbb{R}^p_+, \\ \mu_j = 0 \text{ for } j \notin A(x^*)}} \left\{ \begin{array}{l} (\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*), H), \text{ such that} \\ H \preceq \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \text{ on } C^{\mathcal{S}}(x^*,\mu) \end{array} \right\}, \tag{4.23}$$

where $C^{\mathcal{S}}(x^*,\mu)$ is the (strong) critical cone given by

$$\mathcal{C}^S(x^*,\mu) := \left\{ d \in \mathbb{R}^n : \quad \begin{array}{l} \langle \nabla h_i(x^*), d\rangle = 0, i \in I, \langle \nabla g_j(x^*), d\rangle = 0, \text{ if } \mu_j > 0 \\ \langle \nabla g_j(x^*), d\rangle \le 0, \text{ if } \mu_j^* = 0, \quad j \in A(x^*) \end{array} \right\}. \tag{4.24}$$

We note that the critical cone $\mathcal{C}^S(x^*,\mu)$ is well-defined for every $\mu \ge 0$, and when μ is a Lagrange multiplier (i.e (2.4) holds for some λ) the critical cone coincides with the one defined in (2.7).

So, in this case, the multiplier is redundant and we write $\mathcal{C}^S(x^*)$ instead of $\mathcal{C}^S(x^*,\mu)$. It is worth noting that under the strict complementarity slackness condition, (i.e. μ satisfies (2.4) and $\mu_j > 0$ for all $j \in A(x^*)$), both cones $\mathcal{C}^S(x^*)$ and $\mathcal{C}^{\mathcal{W}}(x^*)$ coincide and SSONC is equivalent to WSONC.

We observe that SSONC holds at x^* for the problem (1.1) if and only if the pair $(-\nabla f(x^*), -\nabla^2 f(x^*))$ belongs to $K_2^{\mathcal{S}}(x^*)$. We also note that $K_2^{\mathcal{S}}(x^*)$ is a subset of $K_2^{\mathcal{W}}(x^*)$, due to $C^{\mathcal{W}}(x^*) \subset C^{\mathcal{S}}(x^*,\mu)$ for every $\mu \ge 0$ and as a consequence the SCCP2 condition is stronger than CCP2.

Following the same reasoning of Theorem 4.1 we obtain:

Theorem 4.8. *Let $x^* \in \Omega$. Then, the conditions below are equivalent*

- *SCCP2 holds at x^*;*
- *for every objective function $f : \mathbb{R}^n \to \mathbb{R}$ of problem* (1.1) *such that AKKT2 holds at x^*, the condition SSONC holds at x^*.*

The next example shows that SCCP2 is too strong that even in well-behaved problems where LICQ holds, it may fail.

Example 4.5 (SCCP2 fails even for simple box constraints)**.** Consider in $\mathbb{R}^n$ $(n \geq 0)$ the simple box constraint $\Omega = \{x \in \mathbb{R}^n : x \geq 0\}$. Then, SCCP2 fails at $x^* = 0$ and CCP2 holds.

Clearly the set Ω is defined by the inequality constraints $g_j(x) = -x_j$ for $j = \{1, \ldots, n\}$. Now we will calculate the weak cone $K_2^{\mathcal{W}}(x)$. Since $x^* = 0$, the set of active index $A(x^*)$ is $\{1, \ldots, n\}$. Using the fact that for all $x \in \mathbb{R}^n$, $\nabla g_i(x) = -e_i$ and $\nabla^2 g_i(x) = 0$ independently of i, we have that

$$C^{\mathcal{W}}(x, x^*) = \{d \in \mathbb{R}^n : \ \ \langle \nabla g_i(x), d\rangle = 0 \text{ for all } i \in \{1, \ldots, n\}\} = \{0\},$$

and as a consequence the weak cone $K_2^{\mathcal{W}}(x)$:

$$K_2^{\mathcal{W}}(x) = \{(\sum_{j \in A(x^*)} -\mu_j e_j, H) : H \preceq 0 \text{ on } C^{\mathcal{W}}(x, x^*) = \{0\}, \, \mu_j \geq 0\}$$

is equal to $K_2^{\mathcal{W}}(x) = \mathbb{R}^n_- \times \text{Sym(n)}$ independently of x. Thus $K_2^{\mathcal{W}}$ is osc at x^* and CCP2 holds. Furthermore, since $\limsup_{x \to x^*} K_2^{\mathcal{W}}(x) = \mathbb{R}^m_- \times \text{Sym}(n)$, to prove that SCCP2 does not hold it is sufficient to find a vector $\hat{\mu} \in \mathbb{R}^n_+$ and a symmetric matrix H such that $H(w, w) > 0$ for some $w \in C^{\mathcal{S}}(x^*, \hat{\mu})$, because in this case, the pair $(-\hat{\mu}, H) \in K_2^{\mathcal{W}}(x) = \mathbb{R}^m_- \times \text{Sym}(n)$ but $(-\hat{\mu}, H)$ does not belong to $K_2^{\mathcal{S}}(x^*)$. Choose $\hat{\mu} := e - e_1$ and $H := e_1 e_1^T$. From the definition of the strong critical cone $C^{\mathcal{S}}(x^*, \hat{\mu})$ we have that $e_1 \in C^{\mathcal{S}}(x^*, \hat{\mu})$ and from the definition of the matrix H, $H(e_1, e_1) = \|\langle e_1, e_1\rangle\|^2 > 0$. Thus the pair $(-\hat{\mu}, H)$ belongs to $\limsup_{x \to x^*} K_2^{\mathcal{W}}(x) = \mathbb{R}^n_- \times \text{Sym}(n)$ but it does not belong to the critical cone $K_2^{\mathcal{S}}(x^*)$.

Despite the strongness of SCCP2, the next example shows that SCCP2 may hold for problems where LICQ fails.

Example 4.6 (SCCP2 does not imply LICQ)**.**
Consider in $\mathbb{R}$, the point $x^* = 0$ and the inequality constraints given by $g_1(x) = -\exp(x) + 1$ and $g_2(x) = x$. Then, SCCP2 holds at $x^* = 0$ and LICQ fails.

First, we note that $x^* = 0$ is a feasible point with $A(x^*) = \{1, 2\}$. From the definition of g_1 and g_2 we have $\nabla g_1(x) = -\exp(x)$, $\nabla^2 g_1(x) = -\exp(x)$, $\nabla g_2(x) = 1$ and $\nabla^2 g_2(x) = 0$. Thus, $C^{\mathcal{W}}(x, x^*) = \{0\}$ and $C^{\mathcal{S}}(x^*, \mu) = \{0\}$ for all $\mu \in \mathbb{R}^2_+$, so $K^{\mathcal{W}}(x) = \mathbb{R} \times \text{Sym}(1) = K^{\mathcal{S}}(x^*)$ for all $x \in \mathbb{R}$, which implies the SCCP2 holds. On the other hand, clearly, LICQ fails.

In Figure 1 we show the relationship among the constraint qualifications discussed in this paper.

5 Practical algorithms that generate AKKT2 points

In this section we show several practical algorithms in the literature that generates sequences whose limit points satisfy the sequential second-order optimality condition AKKT2. Besides the Augmented Lagrangian algorithm of [2], we will show that the regularized SQP of [37] and the Trust Region method of [31] generate AKKT2 sequences.

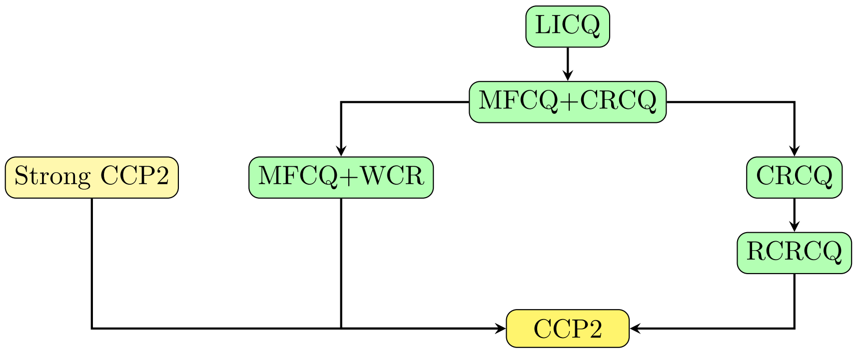

Figure 1: Relationship of CQs associated with second-order global convergence of algorithms.

5.1 Regularized Sequential Quadratic Programming with second-order global convergence

Sequential Quadratic Programming (SQP) methods are a popular class of methods for nonlinear constrained optimization, particularly effective for solving problems arising, for example, from mixed-integer nonlinear programming and PDE-constrained optimization. Due to some theoretical and numerical difficulties associated with ill-posed or degenerate nonlinear optimization problems, two type of SQP methods were designed, regularized and stabilized SQP, see [38, 40]. In [37] Gill, Kungurtsev and Robinson extended the regularized SQP method of [38] to allow convergence to points satisfying the WSONC condition under the constraint qualification MFCQ+WCR. See also [42].

Let us show that the method proposed by [37] generates sequences that satisfy the sequential second-order optimality condition AKKT2. The problem analyzed is

$$\text{Minimize } f(x) \text{ subject to } c(x) = 0, \quad x \geq 0, \tag{5.1}$$

where $c : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable functions. To simplify the analysis in this subsection we will use the same notation as [37]. Let $H(x,\lambda) := \nabla^2 f(x) - \sum \lambda_i \nabla^2 c_i(x)$, $J(x)^T$ is the matrix whoses rows are the gradients of $c_i(x)$ for all $i = 1, \ldots, m$. Note that if we define $h(x) = c(x)$ and $g(x) = -x$, the symmetric matrix $H(x,\lambda)$ coincides with the Hessian of the lagrangian $L(x,\lambda,\mu) = f(x)+\sum \lambda_i h_i(x)+\sum \mu_j g_j(x)$. Define the residual $r(x,\lambda) := \|(c(x), \min(x, \nabla f(x)-J(x)\lambda))\|$. For a feasible point x^*, the perturbed weak critical cone $\widetilde{\mathcal{C}}\ (x) = \{d : J(x)^T d = 0, d_j = 0 \text{ for } j \in A(x^*)\}$. Given positive scalar $\gamma, \varepsilon_a \in (0,1)$, ε-active set is defined as $\mathcal{A}_\varepsilon(x,\lambda,\mu) = \{i : x_i \leq \varepsilon, \text{ with } \varepsilon = \min(\varepsilon_a, \max(\mu, r(x,\lambda)^\gamma))\}$. The ε-free set is defined as $\mathcal{F}_\varepsilon(x,\lambda,\mu) := \{1,\ldots,n\} \setminus \mathcal{A}_\varepsilon(x,\lambda,\mu)$. The proposed algorithm in [37] is based on the first-order primal-dual SQP method of Gill and Robinson [38]. The line-search direction is augmented by a direction of negative curvature that facilitates convergence to points that satisfy the second-order necessary conditions for optimality and it is based on the properties of the primal-dual augmented Lagrangian function

$$M(x,\lambda;\lambda^E,\mu) = f(x) - c(x)^T\lambda^E + \frac{1}{2\mu}\|c(x)\|^2 + \frac{\nu}{2\mu}\|c(x) + \mu(\lambda - \lambda^E)\|^2,$$

where ν is a nonnegative scalar, μ is a positive penalty parameter and λ^E is an estimate of a Lagrangian multiplier. The matrix $B(x,\lambda;\mu)$ denotes the approximation of $\nabla^2 M$ given by [37, expression (2.1)]. $\hat{B}(x,\lambda;\mu)$ is a positive definite matrix equals to $B(x,\lambda;\mu)$ when $B(x,\lambda;\mu)$ is sufficiently positive definite, otherwise it takes an specific form, see [37, expression (2.3)], which depends on a matrix $\hat{H}(x,\lambda)$ such that $\hat{H}(x,\lambda) + \mu^{-1}J(x)J(x)^T$ is positive definite, cf. [38, Theorem 4.5].

For the remainder of the discussion, it is assumed that ν is a fixed positive scalar parameter. The algorithm generates a sequence $\{v_k\}$ where $v_k = (x_k, \lambda_k)$ is the k-th estimate of a primal-dual solution of problem (5.1). Each iterate can be classified as V-, O-, M- or F-iterate (see [37, Algorithm 3]), where the union of index sets of V-, O- and M-iterates is always infinite ([37, Theorem 3.2]). Numerical experiments indicate that M-iterates occur infrequently relative to the total number of iterations. We give a summary of Algorithm 3 from [37] in Algorithm 2.

Algorithm 2 [37, Algorithm summary]

The computation associated with the k-th iteration may be arranged into five main steps.

1. Given (x^k, λ^k) and the regularization parameter μ^R_{k-1} from the previous iteration, define $\mathcal{F}_\varepsilon(x^k, \lambda^k, \mu^R_{k-1})$ and $B(x^k, \lambda^k; \mu^R_{k-1})$. Compute the positive-definite matrix $\hat{B}(x^k, \lambda^k; \mu^R_{k-1})$ together with a nonnegative scalar $\epsilon^{(1)}_k$ and vector s_k such that if $\epsilon^{(1)}_k > 0$, then $(-\epsilon^{(1)}_k, s_k)$ approximates the most negative eigenpair of $B(x^k, \lambda^k; \mu^R_{k-1})$ (see [37, Section 2.1]).
2. Use $\epsilon^{(1)}_k$ and $r(x^k, \lambda^k)$ to define values of λ^E_k and μ^R_k for the k-th iteration (see [37, Section 2.2]).
3. Define a descent direction $d^k = (p^k, q^k)$ by solving a convex bound-constrained subproblem with Hessian $B(x^k, \lambda^k; \mu^R_{k-1})$ and gradient $\nabla M(x^k, \lambda^k; \mu^R_k)$. The primal part of d^k satisfies $x^k + p^k \geq 0$ (see [37, Section 2.3]).
4. Compute a direction of negative curvature $s^k = (u^k, w^k)$ by rescaling the direction s^k. The primal part of s^k satisfies $x^k + p^k + u^k \geq 0$ (see [37, Section 2.3]).
5. Perform a flexible line search along the vector $\triangle v^k = s^k + d^k = (p^k + u^k, q^k + w^k)$ (see [37, Section 2.4]). Update the line-search penalty parameter.

They used the following standard assumptions: (i) The sequence of matrices $\{\hat{H}(x^k, \lambda^k)\}_{k\in\mathbb{N}}$ is uniformly bounded and the sequence of lowest eigenvalue of $\hat{H}(x^k, \lambda^k) + (1/\mu^R_k)J(x^k)J(x^k)^T$ is uniformy bounded by below and (ii) the sequence $\{x^k\}$ is contained in a compact set.

To show that the method generates AKKT2 sequences let us take a closer look at the proof of Theorem 3.4 in [37]. Let $\{v^k = (x^k, \lambda^k)\}$ be the sequence generated by Algorithm 3 of [37] and suppose that the algorithm generates infinitely many V- or O-iterates. Let x^* be a limit point such that every iterate is a V- or O-iterate. So we have, from Algorithm 3 of [37], since the quantities $\phi^{\max}_V$ and $\phi^{\max}_O$ are positive bounds that are reduced by half during the solution process (see [37, (2.10)-(2.11)]), that:

$$\max\left(\|c(x^k)\|, \|\min(x^k, \nabla f(x^k) - J(x^k)\lambda^k)\|, \epsilon^{(1)}_k\right) \to 0. \tag{5.2}$$

From (5.2), we have that x^* is feasible and from $\|\min(x^k, \nabla f(x^k) - J(x^k)\lambda^k)\| \to 0$ we deduce that (3.1) from the definition of AKKT2 holds. Now, we will prove that (3.2) also holds. From the expression [37, (3.25)] and [37, (2.6)] we deduce that

$$\left(H(x^k, \lambda^k) + \frac{1}{\mu^R_{k-1}} J_k J_k^T\right)(v, v) \geq -\frac{1}{\theta}\epsilon^{(1)}_k \|v\|_2^2 \ \text{ for all } v \in \widetilde{\mathcal{C}}\,(x^k),$$

for some scalar θ independent of x^k and λ^k. By (5.2), $\epsilon^{(1)}_k \to 0$. Now using Lemma 2.1 with $P = H(x^k, \lambda^k) + \frac{1}{\mu^R_{k-1}} J_k J_k^T + (\frac{1}{\theta}\epsilon^{(1)}_k + \frac{1}{k})\mathbb{I}$ and $\mathcal{C}$ as $\widetilde{\mathcal{C}}\,(x^k)$, we can conclude that

$$H(x^k, \lambda^k) + \frac{1}{\mu^R_{k-1}} J_k J_k^T + (\frac{1}{\theta}\epsilon^{(1)}_k + \frac{1}{k})\mathbb{I} + \sum_{j\in A(x^*)} \theta^k_j \nabla g_j \nabla g_j^T \succ 0,$$

for some nonnegative scalars $\{\theta_j^k : j \in A(x^*)\}$, or equivalently

$$\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \frac{1}{\mu_{k-1}^R} \sum_{i=1}^m \nabla h_i(x^k) \nabla h_i^T(x^k) + \sum_{j \in A(x^*)} \theta_j^k \nabla g_j \nabla g_j^T \succ -\delta_k \mathbb{I}.$$

where $\delta_k := (\frac{1}{\theta}\epsilon_k^{(1)} + \frac{1}{k})$. Since $\delta_k \to 0$, we get that x^* is an AKKT2 point.

5.2 Trust-region methods with second-order global convergence

Now we will proceed to show that the following trust-region based algortihm generates AKKT2 sequences. The algorithm is the one proposed by Dennis and Vicente in [31] which is an extension of the work of [30]. They only consider equality constraints

$$\text{Minimize } f(x) \quad \text{subject to } C(x) = 0.$$

We use the same notation of [31]. Let $C : \mathbb{R}^n \to \mathbb{R}^m$ $(m < n)$, $C = (c_1, \dots, c_m)^T$ be a twice differentiable function. Each iterate of the method is denoted by x^k. Let W_k be a matrix such that its columns form a basis of $\text{Ker}\nabla C(x^k)^T$. Let H_k be an approximation to $\nabla^2 \ell(x^k, \mu^k)$, $\hat{H}_k = W_k^T H_k W_k$ and $\hat{g}_k = W_k^T \nabla q_k(s_k^n)$, q_k quadratic model of $\ell(x, \mu) = f(x) + \langle \lambda, h(x) \rangle$ at (x^k, λ^k) and s_k^n is called the quasi-normal component of s_k, step of the method. See [31, Section 2]. The general trust-region algorithm is given by Algorithm 3.

Algorithm 3 [31, ALGORITHM 2.1 general trust-region algorithm]

---

1. Choose $x^0, \delta^0, \lambda^0, H_0$ and W_0. Set $\rho_{-1} \geq 1$. Choose $\alpha_1, \eta_1, \delta_{min}, \delta_{max}, \bar{\rho}$ and r such that $0 < \alpha_1, \eta_1 < 1, 0 < \delta_{min} \leq \delta_{max}, \bar{\rho} > 0$ and $r \in (0, 1)$.

2. For $k = 0, 1, 2, \dots$ do

   (a) If $\|W_k^T \nabla \ell(x^k, \lambda^k)\| + \|C(x^k)\| + \gamma_k = 0$ where γ_k is given by [31, (2.10)], stop the algorithm and use x^k as solution.

   (b) Set $s_k^n = s_k^t = 0$.
   If $C(x^k) \neq 0$ then compute s_k^n satisfying [31, (2.1),(2.2),(2.3)] and $\|s_k^n\| \leq r\delta_k$.
   If $\|W_k^T \nabla \ell(x^k, \lambda^k)\| + \gamma_k \neq 0$ then compute $\bar{s}_k^t$ satisfying [31, (2.6)].
   Set $s_k = s_k^n + s_k^t = s_k^n + W_k \bar{s}_k^t$.

   (c) Compute λ^{k+1} satisfying [31, (2.8)].

   (d) Compute $pred(s_k, \rho_{k-1})$. See [31, Algorithm 2.1 (general trust-region algorithm). Item 2.4].

   (e) If $ared(s_k, \rho_k)/pred(s_k, \rho_k) < \eta_1$, set $\delta_{k+1} = \alpha_1 \|s_k\|$ and reject s_k. Otherwise accept s_k and choose δ_{k+1} such that $\max\{\delta_{min}, \delta_k\} \leq \delta_{k+1} \leq \delta_{max}$.

   (f) If s_k was rejected set $x^{k+1} = x^k$ and $\lambda^{k+1} = \lambda^k$. Otherwise $x^{k+1} = x^k + s_k$ and $\lambda^{k+1} = \lambda^k + \triangle\lambda^k$, with $\|\triangle\lambda^k\| \leq \kappa_3 \delta_k$.

Let $\hat{\Omega}$ be an open set of $\mathbb{R}^n$. Suppose that for all the iterations, x^k and $x^k + s_k$ are in $\hat{\Omega}$. Let us consider the following *general assumptions*:

A.1 Functions f, C are twice continuously differentiable in $\hat{\Omega}$.

A.2 The gradient matrix $\nabla C(x)$ has full column rank for all $x \in \hat{\Omega}$.

A.3 Functions $f, \nabla f, \nabla^2 f, C, \nabla C, \nabla^2 c_i, i = 1, \dots, m$ are bounded in $\hat{\Omega}$. The matrix defined by $(\nabla C(x)^T \nabla C(x))^{-1}$ is uniformly bounded in $\hat{\Omega}$.

A.4 Sequences $\{W_k\}$, $\{H_k\}$ and $\{\lambda_k\}$ are bounded.

A.5 The Hessian approximation H_k is exact, i.e., $H_k = \nabla^2_{xx}\ell_k$, and $\nabla^2 f$ and $\nabla^2 c_i, i = 1, \dots, m$ are Lipschitz continuous in $\hat{\Omega}$.

Now, we will prove that the method generates AKKT2 sequences when the Lagrangian multipliers are updated in a consistent way [31, (4.7)]. First, we will prove that (3.2) from the definition of AKKT2 holds for $\{\lambda_k\}$ satisfying only [31, (2.8)]. From the Karush-Kuhn-Tucker conditions there exists a $\gamma_k \geq 0$, [31, (2.10)], such that:

$$\begin{aligned} &\hat{H}_k + \gamma_k W_k^T W_k \text{ is positive semidefinite,} \\ &(\hat{H}_k + \gamma_k W_k^T W_k)\bar{s}_k = -\bar{g}_k, \\ &\gamma_k(\bar{\delta}_k - \|W_k \bar{s}_k\|) = 0. \end{aligned} \tag{5.3}$$

Furthemore, since $\hat{H}_k + \gamma_k W_k^T W_k = W_k^T(H_k + \gamma_k \mathbb{I})W_k$ is positive semidefinite and W_k is a matrix whose columns form a basis of $\text{Ker}\nabla C(x^k)^T$, we have by Lemma 2.1 that there are $\eta_i^k \geq 0$, $i = 1, \dots, m$ such that

$$H_k + \sum_{i=1}^m \eta_i^k \nabla c_i(x^k)\nabla c_i(x^k)^T + (\gamma_k + \frac{1}{k})\mathbb{I} \succ 0. \tag{5.4}$$

By [31, Theorem 3.10], $\liminf(\|W_k^T \nabla \ell(x^k, \lambda^k)\| + \|C(x^k)\| + \gamma_k) = 0$. Now, assume that x^* is a limit point of $\{x^k\}$. Taking an adequate subsequence we may assume that $x^k \to x^*$ for some $x^* \in \mathbb{R}^n$,

$$\gamma_k \to 0, \qquad \|C(x^k)\| \to 0 \qquad \text{and} \qquad \|W_k^T \nabla_x \ell(x^k, \lambda^k)\| \to 0. \tag{5.5}$$

Thus, since $\gamma_k \to 0$, we deduce from (5.4) that (3.2) holds. To prove that (3.1) is fulfilled we choose the Lagrange multipliers λ^k as [31, Lemma 4.2], that is, $\lambda^k = -(\nabla C(x^k)^T \nabla C(x^k))^{-1} \nabla C_k^T \nabla f(x^k)$. Now, for each k, we decompose $\nabla_x \ell(x^k, \lambda^k)$ as:

$$\nabla_x \ell(x^k, \lambda^k) = W_k u^k + \nabla C(x^k) v^k, \tag{5.6}$$

where $W_k u^k$ is in $Ker(\nabla C(x^k)^T)$ and $\nabla C(x^k) v^k$ belongs to $Ker(\nabla C(x^k)^T)^\perp = Im(\nabla C(x^k))$ for some u^k, v^k. Multipliying the expression (5.6) by $(u^k)^T W_k^T$ and using $\lim \|W_k^T \nabla_x \ell(x^k, \lambda^k)\| = 0$, we have that $W_k u^k \to 0$. Now, we proceed to multiply (5.6) by $\nabla C(x^k)^T$ and use the existence of the inverse $(\nabla C(x^k)^T \nabla C(x^k))^{-1}$ to get $v^k = (\nabla C(x)^T \nabla C(x))^{-1} \nabla C_k^T \nabla f(x^k) + \lambda^k = 0$. So, from (5.6), we get $\nabla_x \ell(x^k, \lambda^k) = W_k u^k \to 0$ and (3.1) holds. Finally, from $\|C(x^k)\| \to 0$ we get that x^* is feasible. Thus, x^* is an AKKT2 point as we wanted to show.

Other trust region based-algorithms, such as [30, 33], also generate AKKT2 sequences.

6 Final Remaks

Over the years, several algorithms with convergence to second-order stationary points have been proposed in the literature, whose global convergence is guaranteed by using some constraint qualification as LICQ or MFCQ+WCR. Guided by the necessity of explaining practical aspects of these methods, we took a closer look into their stopping criteria and hence into the sequential second-order optimality condition AKKT2, and we were able to prove second-order global convergence of such algorithms under a weaker assumption, namely, CCP2, which is weaker than MFCQ+WCR, which has been previously assumed, and is also weaker than RCRCQ. This framework also gives a tool to prove second-order global convergence results of second-order algorithms under a weak constraint qualification. In this sense, we believe that AKKT2 can play a unifying role in the global convergence analysis of second-order algorithms in the same way as AKKT plays a similar role for first-order methods.

References

[1] R. Andreani, R. Behling, G. Haeser, and P. J. S. Silva. On second order optimality conditions for nonlinear optimization. *Optimization Online*, Submmited:1–19, 2014.

[2] R. Andreani, E. G. Birgin, J. M. Martinez, and M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications*, 45:209–236, 2010.

[3] R. Andreani, C. E. Echague, and M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization theory and Applications*, 146(2):255–266, 2010.

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

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

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

[7] R. Andreani, J. M. Martinez, L. T. Santos, and B. F. Svaiter. On the behaviour of constrained optimization methods when lagrange multipliers do not exist. *Optimization Methods and Software*, 29:646–657, 2014.

[8] R. Andreani, J. M. Martinez, and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56:529–542, 2007.

[9] R. Andreani, J.M. Martínez, A. Ramos, and P.J.S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *Optimization-online*, 2015.

[10] M. Anitescu. Degenerate nonlinear programming with a quadratic growth condition. *SIAM Journal on Optimization*, 10(4):1116–1135, 2000.

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

[12] A. V. Arutyunov. Second-order conditions in extremal problems. the abnormal points. *Transactions of American Mathematical Society*, 350:4341–4365, 1998.

[13] A. V. Arutyunov and F. L. Pereira. Second-order necessary optimality conditions for problems without a priori normality assumptions. *Mathematics of Operations Research*, 31(1):1–12, 2006.

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

[15] 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(2):394–408, 2005.

[16] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Practical methods of Optimization: theory and algorithms*. John Wiley and Sons, New Jersey, USA, 2006.

[17] A. Ben-Tal. Second order and related extremality conditions in nonlinear programming. *Journal of Optimization Theory and Applications*, 31:143–165, 1980.

[18] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, NY, 1982.

[19] D. P. Bertsekas. *Nonlinear programming.* Athenas Scientific, Belmont, Mass, 1999.

[20] E. Birgin and J. M. Martinez. *Practical Augmented Lagrangian Methods for Constrained Optimization.* SIAM Publications, Philadelphia, PA, 2014.

[21] J. F. Bonnans, R. Cominetti, and A. Shapiro. Second-order optimality conditions based on parabolic second order tangent sets. *SIAM Journal of Optimization*, 9:466–492, 1999.

[22] J. F. Bonnans, A. Hermant, A. V. Arutyunov, and F. L. Pereira. No-gap second-order optimality conditions for optimal control problems with a single state constraint and control. *Mathematical Programming, Serie B.*, 117:21–50, 2009.

[23] J. F. Bonnans and A. Shapiro. *Pertubation Analysis of Optimization Problems.* Springer-Verlag, Berlin, 2000.

[24] R. H. Byrd, R. B. Schnabel, and G. A. Shultz. A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal of Numerical Analysis*, 24:1152–1170, 1987.

[25] E. Casas and F. Troltzsch. Second-order necessary and sufficient optimality conditions for optimization problems and applications to control theory. *SIAM Journal of Optimization*, 13 (2):406–431, 2002.

[26] T. F. Coleman, J. Liu, and W. Yuan. A new trust-region algorithm for equality constrained optimization. *Computacional Optimization and Applications*, 21:177–199, 2002.

[27] A. R. Conn, N. I. M. Gould, D. Orban, and P. L. Toint. A primal-dual trust-region algorithm for minimizing a non-convex fucntion subject to general inequality and linear equality constraints. In *Nonlinear optimization and related topics*, pages 15–49, Erice 1998.

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

[29] G. Debreu. Definite and semidefinite quadratic forms. *Econometrica*, 20:295–300, 1952.

[30] J. E. Dennis, M. El-Alem, and M. C. Maciel. A global convergence theory for general trust-region-based algorithms for equality constrained optimization. *SIAM Journal of Optimization*, 7(1):177–207, 1997.

[31] J. E. Dennis and L. N. Vicente. On the convergence theory of trust-region-based algorithms for equality-constrained optimization. *SIAM Journal of Optimization*, 7(4):927–950, 1997.

[32] G. DiPillo, S. Lucidi, and L. Palagi. Convergence to second-order stationary points of a primal-dual algorithm model for nonlinear programming. *Mathematics of Operations Research*, 30:897–915, 2005.

[33] M. M. El-Alem. Convergence to a second-order point of a trust-region algorithm with a nonmonotonic penalty parameter for constrained optimization. *Journal of Optimization theory and applications*, 91(1):61–79, 1996.

[34] F. Facchinei and S. Lucidi. Convergence to second-order stationary points in inequality constrained optimization. *Mathematics of Operations Research*, 23:746–766, 1998.

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

[36] R. Fletcher. *Practical methods of Optimization: Constrained Optimization.* Wiley, New York, USA, 1981.

[37] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A regularized SQP method with convergence to second-order optimal points. *Optimization Online*, 2013.

[38] P. E. Gill and D. P. Robinson. A globally convergent stabilized SQP method. *SIAM Journal of Optimization*, pages 1983–2013, 2013.

[39] N. I. M. Gould, Conn, and P. L. Toint. A note on the convergence of barrier algorithms for second-order necessary points. *Mathematical programming*, 85:433–438, 1998.

[40] A. F. Izmailov and M. V. Solodov. Stabilized SQP revisited. *Mathematical Programming*, pages 133(93–120), 2012.

[41] R. Janin. Direction derivative of the marginal function in nonlinear programming. *Mathematical Programming Studies*, 21:127–138, 198ch4.

[42] V. Kungurtsev. Second-derivative sequential quadratic programming methods for nonlinear optimization. *Ph. D. Thesis, UC San Diego*, 2013.

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

[44] J. M. Martinez and B. F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, 118 No 1:117–133, 2003.

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

[46] J. M. Morguerza and F. J. Prieto. An augmented lagrangian interior-point method using directions of negative curvature. *Mathematical Programming*, 95 (3):573–616, 2003.

[47] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.

[48] J. Nocedal and S. J. Wright. *Numerical optimization.* Springer, New York, 2006.

[49] J. P. Penot. Optimality conditions in mathematical programming and composite optimization. *Mathematical Programming*, 67:225–245, 1994.

[50] J. P. Penot. Second-order conditions for optimization problems with constraints. *SIAM Journal of Control and Optimization*, 37:303–318, 1998.

[51] D. W. Petersen. A review of constraint qualifications in finite-dimensional spaces. *SIAM Review*, 15:639–654, 1973.

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

[53] A. Ramos. *Tópicos em Condições de Otimalidade para Otimização não linear.* PhD thesis, IME-USP, Departamento de Matemática Aplicada, São Paulo-SP, Brazil, 2015.

[54] M. V. Solodov. Constraint qualifications. In *Wiley Encyclopedia of Operations Research and Management Science.* James J. Cochran, et al. (editors), John Wiley and Sons, 2010.

[55] M. Spivak. *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. Addison-Wesley Publishing Company, 1965.