IMA Journal of Numerical Analysis (2017) **37**, 1902–1929
doi: 10.1093/imanum/drw064
Advance Access publication on January 20, 2017

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

ROBERTO ANDREANI
Department of Applied Mathematics, University of Campinas, Rua Sérgio Buarque de Holanda 651, Cidade Universitária, 13083-859, Campinas, SP, Brazil
andreani@ime.unicamp.br

GABRIEL HAESER* AND ALBERTO RAMOS‡
Department of Applied Mathematics, University of São Paulo, Rua do Matão 1010 Cidade Universitária, 05508-090, São Paulo, SP, Brazil
‡Present address: *Department of Mathematics, Federal University of Paraná, Curitiba 19.081, 81531-980, PR, Brazil*
*Corresponding author: ghaeser@ime.usp.br albertoramos@ufpr.brand

AND

PAULO J. S. SILVA
Department of Applied Mathematics, University of Campinas, Rua Sérgio Buarque de Holanda 651, Cidade Universitária, 13083-859, Campinas, SP, Brazil
pjssilva@ime.unicamp.br

[Received on 30 October 2015; revised on 10 October 2016]

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 article, 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. This means that we can prove second-order global convergence of well–established algorithms even when the set of Lagrange multipliers is unbounded, which was a limitation of previous results based on Mangasarian–Fromovitz constraint qualification. We prove global convergence of well-known variations of the augmented Lagrangian and regularized sequential quadratic programming methods to second-order stationary points under this new weak constraint qualification.

Keywords: 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 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.

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 on 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 that is 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 (Fletcher, 1981; Nocedal & Wright, 2006). 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 Petersen (1973), Solodov (2010), Andreani *et al.* (2012a,b) and references therein. We mention two of them. The Mangasarian–Fromovitz condition (MFCQ), defined in Mangasarian & Fromovitz (1967), 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 Janin (1984), 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 Andreani *et al.* (2011). See also Qi & Wei (2000), Martinez & Svaiter (2003), Birgin & Martinez (2014).

DEFINITION 1.1 The approximate Karush–Kuhn–Tucker (AKKT) optimality condition is said to hold at a point $x^* \in \mathbb{R}^n$ if there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$ and $\{\varepsilon_k\} \subset \mathbb{R}_+$, such that $x^k \to x^*$, $\varepsilon_k \to 0^+$,

$$\|h(x^k)\| \le \varepsilon_k, \qquad \|\max\{0, g(x^k)\}\| \le \varepsilon_k, \tag{1.3}$$

$$\left\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \right\| \le \varepsilon_k \tag{1.4}$$

and

$$\mu_j^k = 0 \text{ if } g_j(x^k) < -\varepsilon_k. \tag{1.5}$$

We note that the AKKT condition can be equivalently stated as $x^* \in \Omega$ and $\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j|g_j(x^*)=0} \mu_j^k \nabla g_j(x^k) \to 0$, with $x^k \to x^*$ and $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$.

The AKKT condition tries to capture a natural property of many optimization algorithms: they are designed to generate a primal sequence together with approximating multipliers for which the KKT residual goes to zero. In this case, one may expect that the cluster points of the primal sequence should conform to the KKT conditions. Such convergence to KKT points holds naturally if the approximating multipliers are bounded but it may fail otherwise. Therefore, the analysis of the global convergence of such algorithms usually revolves around stringent constraints qualifications, like MFCQ, and special algorithmic properties that allow us to bound, or indirectly control, the multiplier sequence. However, it has been shown recently that such *ad hoc* analysis is not necessary. The generation of an AKKT sequence together with a very mild constraint qualification is enough to ensure global convergence to KKT points even if the multiplier sequence is unbounded. See Andreani *et al.* (2012a,b, 2016) for details.

The attractiveness of sequential optimality conditions such as AKKT is associated to three properties. First, they are genuine necessary optimality conditions, independently of the fulfillment of constraint qualifications (Andreani *et al.*, 2011; Birgin & Martinez, 2014). Second, they are strong, in the sense that they imply the classical first-order optimality condition 'KKT or not CQ1' for weak constraint qualifications (Andreani *et al.*, 2012a,b, 2016). Third, there are many algorithms that generate sequences whose limit points satisfy them. Particularly in the case of AKKT, many optimization algorithms (but not all; see Andreani *et al.*, 2014b), 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 (1.4) and (1.5) are fulfilled (Andreani *et al.*, 2012a). 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 stating that whenever the point of interest is the limit of an AKKT sequence it is also a KKT point, was recently characterized in Andreani *et al.* (2016). 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 more desirable. Such conditions have been extensively studied in the literature (see Fiacco & McCormick, 1968; Fletcher, 1981; Bonnans *et al.*, 1999; Bonnans & Shapiro, 2000; Nocedal & Wright, 2006), with important applications to mathematical programming (Penot, 1998; Arutyunov & Pereira, 2006), composite optimization (Penot, 1994), optimal control (Casas & Troltzsch, 2002; Bonnans *et al.*, 2009), etc. In the development of algorithms, most of the second-order necessary optimality conditions used are of the form 'if a local minimizer satisfies some constraint qualification (CQ2), then the WSOC condition holds'. That is,

$$\text{WSOC or not CQ2,} \tag{1.6}$$

where WSOC stands for the weak second-order 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 2.2. Our focus on necessary optimality conditions of the type 'WSOC or not CQ2' comes from algorithmic considerations. To the best of our knowledge, there is no algorithm with global convergence to a point that satisfies a second-order stationarity measure stronger than WSOC. In particular, there is no 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 WSOC. There is also strong evidence that even simple second-order methods will fail to find points conforming to more stringent second-order conditions (Gould *et al.*, 1998).

Several algorithms for (1.1) that converge to second-order stationary points (i.e., points where WSOC holds) have been proposed in the literature over the years. Andreani *et al.* (2010a) (see also Andreani *et al.*, 2007), 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 *et al.* (1987) employ a sequential quadratic programming (SQP) approach, where the second-order stationarity is obtained due to the use of second-order correction steps. Coleman *et al.* (2002) also use an SQP approach with quadratic penalty functions for equality constrained minimization. Conn *et al.* (1998) employ the logarithmic barrier method for inequality constrained optimization with linear equality constraints. Dennis & Vicente (1997) use affine scaling directions and the SQP approach for optimization with equality constraints and simple bounds (see also Dennis *et al.*, 1997. DiPillo *et al.* (2005) 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 & Lucidi (1998) use negative-curvature directions in the context of inequality constrained problems. Recently, Gill, Kungurtsev and Robinson used a variant of the SQP method, specifically, the regularized SQP defined in Gill & Robinson (2013) and Gill *et al.* (2013). 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 & Prieto (2003) employ an interior-point algorithm for nonconvex problems and uses directions of negative curvature. The convergence to second-order critical points of trust-region algorithms for convex constraints is studied in detail in Conn *et al.* (2000).

Even with 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 conditions can play in the convergence theory of (first-order) algorithms. This is the main purpose of this article.

We will introduce a sequential second-order optimality condition that we call AKKT2. As with 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 'WSOC 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 3.7, which was introduced in Andreani *et al.* (2007) and is used in the analysis of convergence of the second-order augmented Lagrangian method proposed in Andreani *et al.* (2010a) and also the regularized SQP (Gill *et al.*, 2013). It is also strictly weaker than the CRCQ condition (or its relaxed version Minchenko & Stakhovski, 2011), which proves convergence to a second-order stationary point even when Lagrange multipliers 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 (Andreani *et al.*, 2010a), the regularized SQP (Gill *et al.*, 2013) and the trust-region method (El-Alem, 1996) generate AKKT2 sequences; see Section 5. These results indicate that AKKT2 can be used as a unifying tool for global convergence analysis of 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 WSOC, extending the convergence result of algorithms that assumed more stringent constraint qualifications.

We organize the rest of this article as follows. In Section 2, we survey some basic results and preliminary considerations that will be useful to understand the main results of the article. 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 'WSOC or not (MFCQ and WCR)'. Finally, we present an 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 such 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 $\mathbb{R}_- := -\mathbb{R}_+$. Here $\mathbb{I}$ denotes the identity matrix of appropriate dimension, e_i denotes the ith 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. Sym(n) denotes the set of symmetric matrices. Sym$_+(n)$ stands for the set of order n symmetric positive-semidefinite matrices. Given two symmetric matrices A, B in 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 $A(v,v) := \langle v, Av\rangle = v^{\mathrm{T}}Av$. Finally, we define the set of indexes $I := \{1,\ldots,m\}$ of equality constraints and $A(x^*) = \{i \in \{1,\ldots,p\} \mid g_i(x^*) = 0\}$ of active inequality constraints at $x^* \in \Omega$.

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

Lemma 2.1 (Debreu, 1952; Bertsekas, 1982) Let $P \in$ Sym(n) and vectors $a_1,\ldots,a_r \in \mathbb{R}^n$. Define the subspace $\mathscr{C} = \{d \in \mathbb{R}^n : \langle a_j, d\rangle = 0 \text{ for } j \in \{1,\ldots,r\}\}$. Suppose that $P(v,v) > 0$ for all $v \in \mathscr{C}$. Then, there exist positive scalars $\{c_j, j \in \{1,\ldots,r\}\}$ such that $P + \sum_{j=1}^r c_j a_j a_j^{\mathrm{T}} \succ 0$.

Given a set-valued mapping (multifunction) $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$, the *sequential Painlevéa–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 CQs (see Baccari & Trad, 2005; Bazaraa *et al.*, 2006; Andreani *et al.*, 2007, 2014a for details), one can prove that a local minimizer x^* of (1.1) fulfills the WSOC condition stated below.

DEFINITION 2.2 A feasible point $x^* \in \Omega$ satisfies the (WSOC) 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

$$\left(\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j^* \nabla^2 g_j(x^*) \right) (d, d) \geq 0 \text{ for all } d \in \mathscr{C}^W(x^*), \tag{2.5}$$

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

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

That is, WSOC 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 $\mathscr{C}^{\mathscr{W}}(x^*)$, for some Lagrange multiplier (λ^*, μ^*).

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

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

we say that the strong second-order necessary optimality condition (SSOC) holds. The SSOC is a well-studied condition (Fiacco & McCormick, 1968; Fletcher, 1981; Bertsekas, 1999; Bazaraa *et al.*, 2006) that holds under the classical LICQ. In fact, one can prove that a local minimizer of (1.1) fulfills SSOC imposing a variety of other conditions. See Baccari & Trad (2005), Bazaraa *et al.* (2006), Andreani *et al.* (2010b), Minchenko & Stakhovski (2011), Andreani *et al.* (2014a). We note that conditions in Andreani *et al.* (2010b), Andreani *et al.* (2014a), Minchenko & Stakhovski (2011) yield WSOC or SSOC for every Lagrange multiplier, which can be relevant in some applications.

However, it is well known that MFCQ by itself is not enough to ensure the validity of SSOC or WSOC (Arutyunov, 1991; Anitescu, 2000). There are other second-order conditions that hold under MFCQ (for instance, see Ben-Tal & Zowe, 1982, Remark 9.3 or Bonnans & Shapiro, 2000, Theorem 3.45), or without any CQ (see Arutyunov, 1998, Theorem 3.1; Bonnans & Shapiro, 2000, Theorem 3.50 or Ben-Tal & Zowe, 1982, Theorem 9.3). These conditions do not suit our framework since they require the knowledge of the whole set of Lagrange multipliers in order to be verified, whereas in practice, only an approximation to a single Lagrange multiplier is available.

Also from the computational point of view, even establishing if SSOC holds is, in general, an NP-hard problem (Murty & Kabadi, 1987) and to our knowledge, no algorithm has been shown to converge to a point at which SSOC holds. Gould *et al.* (1998) showed a simple box-constrained optimization problem where the barrier method generates a sequence where SSOC fails to be attained at the limit, while the sequence of barrier minimizers satisfies the second-order sufficient optimality condition.

From the above considerations, WSOC is the natural condition to be considered in the convergence analysis of second-order algorithm 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 WSOC 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. By a sequential optimality condition, we mean a condition with the following three properties: (i) it is a necessary optimality condition, independently of any constraint qualification, (ii) it should be as strong as possible, in our case, it must imply (1.6) for weak constraint qualifications and (iii) it must be possible to verify its validity in sequences generated by algorithms.

DEFINITION 3.1 We say that the feasible point $x^* \in \Omega$ is an approximate second-order stationary point for 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 $x^k \to x^*$, $\delta_k \to 0$,

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

and

$$\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^{\mathrm{T}} + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^{\mathrm{T}} + \delta_k \mathbb{I} \tag{3.2}$$

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

Since Definition 3.1 is a second-order version of the sequential optimality condition AKKT, we say that a point that satisfies it is an AKKT2 point. The rest of this section is devoted to showing 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 3.2 (Andreani *et al.*, 2010a, 2007) Let $\bar{f} : \mathbb{R}^n \to \mathbb{R}$, $\bar{g}_j : \mathbb{R}^n \to \mathbb{R}$ for $j \in \{1, \ldots, p\}$ be 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

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

is positive semidefinite at $\bar{x}$.

The following lemma is an adaptation of the exterior penalty method (Fiacco & McCormick, 1968). See also Bertsekas (1999), Andreani *et al.* (2011).

LEMMA 3.3 Let $\mathscr{C}$ 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

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

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

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

Now, we will show that AKKT2 is a necessary optimality condition. The idea is to get second-order information using penalization techniques. This result can be obtained from others in the literature (see Remark 3.1) and its proof is included here for the sake of completeness.

THEOREM 3.4 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 an $\varepsilon > 0$ such $f(x^*) \leq f(x)$ for all feasible x such that $\|x - x^*\| \leq \varepsilon$. So x^* is the unique solution of

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

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

$$\text{minimize } f(x) + \frac{1}{4}\|x - x^*\|^4 + \frac{1}{2}\rho_k \left\{ \sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p} \max\{0, g_j(x)\}^2 \right\} \text{ subject to } x \in \mathbb{B}(x^*, \varepsilon). \tag{3.5}$$

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 3.3, the sequence $\{x^k\}$ converges to x^* and $x^k \in \operatorname{Int} \mathbb{B}(x^*, \varepsilon)$ 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:

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

By Lemma 3.2 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

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

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 3.5 The notion of AKKT2 has been implicitly stated in the optimization literature; in particular, we do not claim that the result from Theorem 3.4 is new. The idea of obtaining second-order information from penalization techniques is well known. See, for instance, Fiacco & McCormick (1968), Auslender (1979), Arutyunov (1998), Arutyunov (2000), Bertsekas (1999). In particular, the result above can be derived from the Stage I of the proof of Theorem 4.1 in Arutyunov (2000). 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.

3.2 *Strength of the AKKT2 condition*

The AKKT2 condition is a strong second-order optimality condition in the sense that it implies (1.6) for weak constraint qualifications. In this subsection, we will prove that the joint condition MFCQ and WCR serves as corresponding to CQ2; see Proposition 3.9. 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 4.12) and as a consequence the RCRCQ can be used in the global convergence analysis of algorithms. In order to prove that AKKT2 implies 'WSOC or not (MFCQ and WCR)', we recall the definition of the WCR condition introduced by Andreani *et al.* (2007).

DEFINITION 3.6 Let $x^* \in \Omega$ be a feasible point. We say that the WCR condition 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 WCR condition is a weaker property, weaker than CRCQ defined as follows.

DEFINITION 3.7 Let $x^* \in \Omega$ be a feasible point. We say that the CRCQ holds if there is a neighborhood V of x^* such that for every $I_1 \subset I$ and $J_1 \subset A(x^*)$, the rank of $\{\nabla h_i(x), \nabla g_j(x) : i \in I_1, j \in J_1\}$ remains constant for all $x \in V$.

The key property of the WCR condition is the following.

LEMMA 3.8 (Andreani *et al.*, 2007) Assume that WCR holds at a feasible point $x^* \in \Omega$. Then, for every $d \in \mathscr{C}^{\mathscr{W}}(x^*)$ 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 $i \in \{1,\dots,m\}$ 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 3.9 Let $x^* \in \Omega$ be such that AKKT2 holds. If the joint condition MFCQ and WCR holds at x^*, then WSOC 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) $\nabla_x^2 L(x^k,\lambda^k,\mu^k) + \sum \eta_i^k \nabla h_i(x^k)\nabla h_i(x^k)^{\mathrm{T}} + \sum \theta_j^k \nabla g_j(x^k)\nabla g_j(x^k)^{\mathrm{T}} \succeq -\delta_k \mathbb{I}$.

By MFCQ, the sequence $\{(\lambda^k,\mu^k)\}$ is bounded; otherwise, dividing ε^k by $\|(\lambda^k,\mu^k)\|$ and taking the 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 the limit in item (a), we have that x^* satisfies the KKT condition with multipliers μ^* and λ^*.

Now we will prove that WSOC holds in x^* with these multipliers. Take any d in $\mathscr{C}^{\mathscr{W}}(x^*)$; by Lemma 3.8, there is a sequence d^k with $d^k \to d$ such that $\langle \nabla h_i(x^k), d^k\rangle = 0$, for $i \in \{1,\dots,m\}$ 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 the limit in (3.8), we get

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

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 3.10 (AKKT2 is stronger than AKKT) Consider $f(x_1,x_2) = -x_1 - x_2$, $g(x_1,x_2) = x_1^2x_2^2 - 1$ 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 $\nabla g(x_1,x_2) = 2x_1x_2(x_2,x_1)$, the KKT condition holds at x^*. Second, 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 *An algorithm that generates AKKT2 sequences*

In this subsection, we will show that the Augmented Lagrangian algorithm proposed by Andreani *et al.* (2010a) (see also Andreani *et al.*, 2007) generates an AKKT2 sequence. In Section 5, we will show that this is also the case for the regularized SQP method of Gill *et al.* (2013) and for the trust-region method of Dennis & Vicente (1997).

Let us recall the augmented Lagrangian method from Andreani *et al.* (2010a) for problem (1.1), which is equivalent to that proposed in Andreani *et al.* (2007), but without box constraints.

Consider the following augmented Lagrangian function:

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

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 Andreani *et al.* (2007) that coincides with the second derivative operator ∇^2 at twice-differentiable points, and

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

Now we will proceed to analyse Algorithm 1 below.

Algorithm 1 (Andreani *et al.*, 2007, 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

$$\|\nabla L_{\rho_k}(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k \quad \text{and} \quad \bar{\nabla}^2 L_{\rho_k}(x^k, \lambda^k, \mu^k) \succeq -\varepsilon_k \mathbb{I}. \tag{3.13}$$

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 $\lambda_i^{k+1} := \operatorname{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\lambda_i^k + \rho_k h_i(x^k)) \in [\lambda_{\min}, \lambda_{\max}]$ for all $i \in \{1, \ldots, m\}$ and $\mu_j^{k+1} := \operatorname{proj}_{[0, \mu_{\max}]} \max\{0, \mu_j^k + \rho_k g_j(x^k)\} \in [0, \mu_{\max}]$, $j \in \{1, \ldots, p\}$. Set $k \leftarrow k + 1$ and go to step 1.

From (3.13), we have that any limit point of $\{x^k\}$ fulfills the AKKT2 condition. To see this we will take a closer look at Andreani *et al.* (2007, 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 (3.13) is equivalent to

$$\|\nabla L(x^k, \hat{\lambda}^k, \hat{\mu}^k)\| \leq \varepsilon_k, \tag{3.14}$$

and

$$\nabla^2 L(x^k, \hat{\lambda}^k, \hat{\mu}^k) + \rho_k \sum_{i=1}^{m} \nabla h_i(x) \nabla h_i(x)^{\mathrm{T}} + \rho_k \sum_{j \in A(x^*)} \nabla g_j(x) \nabla g_j(x)^{\mathrm{T}} \succeq -\varepsilon_k \mathbb{I}, \tag{3.15}$$

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, Andreani *et al.* (2007, Theorem 4.1) shows that for k large enough, $\hat{\mu}_j^k = 0$ for every $j \notin A(x^*)$. We see from (3.14) and (3.15) that the AKKT2 condition holds at x^*. Thus, we have shown that the augmented Lagrangian method generates AKKT2 sequences.

We end this subsection showing that a result similar to Proposition 3.9, replacing WSOC by SSOC, does not hold, even under LICQ. More specifically, we will show that Algorithm 1 can generate a sequence that converges to a point that does not satisfy SSOC, even if the criterion (3.13) for solving each subproblem is strengthened.

Let us choose $x^0 = 0$, $\lambda^0 = 0$ and $\rho_0 > 1$. For the next iteration, the algorithm will accept the point $x^1 = 0$ as an approximate solution of the subproblem independently of the precision ε, whenever $|\nabla_x L_{\rho_0}(x^1, \lambda^0)| = |2x^1(\rho_0 - 1)| \leq \varepsilon$ and $\overline{\nabla}^2_{xx} L_{\rho_0}(x^1, \lambda^0) = 2(\rho_0 - 1) > 0 > -\varepsilon$. Furthermore, $x^1 = 0$ is a global minimizer of the quadratic model $q_1(x) := \nabla_x L_{\rho_0}(x^1, \lambda^0)(x - x^1)^{\mathrm{T}} + \frac{1}{2}\overline{\nabla}^2_{xx} L_{\rho_0}(x^1, \lambda^0)(x - x^1)^2$. Following the usual multiplier update rule, we have that $\lambda^1 := \lambda^0 + \rho_1 \min\{x^1, 0\} = 0$. Continuing the process, we obtain a sequence of iterates $x^k = 0$, $\lambda^k = 0$, $\rho_k > 1$, for all $k \geq 1$, where each x^k is a global minimizer of the corresponding quadratic model. We can see that the limit $x^* = 0$ does not satisfy SSOC for the original problem, while it still satisfies WSOC.

4. Second-order global convergence under weak constraint qualifications

The global convergence to second-order stationary points of the augmented Lagrangian method (Andreani *et al.*, 2007) and regularized SQP method (Gill *et al.*, 2013) 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 3.9 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

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

The set $\mathscr{C}^{\mathscr{W}}(x, x^*)$ can be considered as a perturbation of the weak critical cone $\mathscr{C}^{\mathscr{W}}(x^*)$ around the feasible point $x^* \in \Omega$. Clearly, $\mathscr{C}^{\mathscr{W}}(x, x^*)$ is a subspace and $\mathscr{C}^{\mathscr{W}}(x^*, x^*)$ coincides with the weak critical cone $\mathscr{C}^{\mathscr{W}}(x^*)$. Using a variational language, we can re-state Lemma 3.8 as a WCR condition implies the inner semicontinuity (isc) of the set-valued mapping $x \rightrightarrows \mathscr{C}^{\mathscr{W}}(x, x^*)$ at $x = x^*$, that is,

$\mathscr{C}^{\mathscr{W}}(x^*, x^*) \subset \liminf_{x\to x^*} \mathscr{C}^{\mathscr{W}}(x, x^*)$; in fact, the inner semicontinuity of $\mathscr{C}^{\mathscr{W}}(x, x^*)$ at x^* turns out be equivalent to the WCR condition (Ramos, 2016).

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

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

The set $K_2^{\mathscr{W}}(x)$ is a convex cone included in $\mathbb{R}^n \times \mathrm{Sym}(n)$ and it allows us to write the WSOC in a more compact form, namely,

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

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) $x \mapsto K_2^{\mathscr{W}}(x)$, defined in (4.2), is outer semicontinuous at x^*, that is,

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

The CCP2 condition is the weakest condition that can be used to generalize Proposition 3.9, as the next theorem shows.

THEOREM 4.2 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^*, WSOC holds at x^*.

Proof. First, suppose CCP2 holds at x^* and there is a function f such that the AKKT2 is satisfied. From the definition of AKKT2, there exist sequences $\{x^k\}$, $\{\lambda^k\}$, $\{\mu^k\}$, $\{\eta^k\}$, $\{\theta^k\}$, $\{\delta_k\}$, 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) $\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^{\mathrm{T}} + \sum \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^{\mathrm{T}} \succeq -\delta_k \mathbb{I}$.

From items (a) and (b), we see that $(-\nabla f(x^k) + \varepsilon^k, -\nabla^2 f(x^k) - \delta_k \mathbb{I}) \in K_2^{\mathscr{W}}(x^k)$. Now using the continuity of $\nabla f(x)$ and $\nabla^2 f(x)$ jointly with the outer semicontinuity of $K_2^{\mathscr{W}}(x)$ at x^*, we obtain that $(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^{\mathscr{W}}(x^*)$ and as a consequence WSOC holds. Now let us prove the other implication. Let (w, W) be an element of $\limsup K_2^{\mathscr{W}}(x)$ when $x \to x^*$. We will show that (w, W) is in $K_2^{\mathscr{W}}(x^*)$.

By definition of the 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 $x^k \to x^*$,

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

and

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

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

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

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

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

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, WSOC holds, and by (4.3), $(w, W) = (-\nabla f(x^*), -\nabla^2 f(x^*))$ belongs to $K_2^{\mathscr{W}}(x^*)$ as we wanted to prove. □

Since AKKT2 is a necessary optimality condition, by Theorem 4.2, we have the following corollary.

COROLLARY 4.3 If x^* is a local minimizer of (1.1) such that CCP2 holds, then WSOC holds.

REMARK 4.4 Some CQ2's are easily verifiable (for example, LICQ) and others are verifiable with different degrees of difficulty. CCP2 is not easily verifiable. This is not one of our objectives in the analysis of constraint qualifications. We are mainly interested in the weakness of constraint qualifications since, when a CQ2 is weak, the condition 'WSOC or not-CQ2' is strong and so the corresponding sequential optimality condition is strong. Clearly, stopping an algorithm with the fulfillment of a strong optimality condition increases our chances of obtaining minimizers.

By Proposition 3.9 and Theorem 4.2 we have the following corollary.

COROLLARY 4.5 The joint condition MFCQ and WCR implies CCP2.

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

EXAMPLE 4.6 (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 $K_2^{\mathscr{W}}(x)$ 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 $\mathscr{C}^{\mathscr{W}}(x, x^*) = \{0\}$. From this, every $H \in \mathrm{Sym}(1) = \mathbb{R}$ satisfies

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

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

Another constraint qualification that yields WSOC is the following one introduced by Baccari & Trad (2005): 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.[1] Although the Baccari–Trad condition, as CCP2, guarantees the fulfillment of WSOC at a local minimizer, these conditions are not equivalent.

[1] The original Baccari–Trad condition has an extra assumption, called GSCS; see Baccari & Trad (2005, Definition 7.4). However, in this work they were trying to assert the validity of SSOC, instead of its weaker version WSOC. To derive WSOC using Baccari and Trad's results one does not need this extra assumption. Just use Baccari & Trad (2005, Lemma 7.2 and Theorem 4.1) with the weak critical cone as the first-order cone.

EXAMPLE 4.7 (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, the Baccari–Trad condition holds at x^* but CCP2 fails.

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

To see that CCP2 does not imply the Baccari–Trad condition, it is enough to see that the MFCQ+WCR condition implies CCP2 while not implying the Baccari–Trad condition, see Andreani *et al.* (2007, counterexample 5.2).

The independence of CCP2 and the Baccari–Trad condition has interesting implications. Due to Theorem 4.2, the Baccari–Trad condition is not enough to guarantee that a limit point of an AKKT2 sequence satisfies WSOC. See the next example.

EXAMPLE 4.8 (AKKT2 under the Baccari–Trad condition does not imply WSOC) Consider the optimization problem.

$$\text{minimize}\, f(x_1,x_2) = -2x_1^2 \text{ subject to } g_1(x_1,x_2) = -x_2 \leq 0,\ \ g_2(x_1,x_2) = x_1^2 - x_2 \leq 0. \tag{4.9}$$

By Example 4.7, the Baccari–Trad condition 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 these 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

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

where the last matrix is positive semidefinite. Also, by direct calculation we have that WSOC fails and $x^* = (0,0)$ is not an optimal solution. So, in this example, we have a point $x^* = (0,0)$ that is not an optimal solution and neither satisfies WSOC, 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 happen if instead of the Baccari–Trad condition, we consider any other constraint qualification that implies CCP2.

Another constraint qualification weaker than LICQ is the constant-rank constraint qualification (CRCQ); cf. Janin (1984). Let us recall the definition of CRCQ. We say that a feasible point $x^* \in \Omega$ verifies CRCQ if there exists a neighborhood 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 Andreani *et al.* (2010b), it was proved that under CRCQ, a local minimizer conforms to SSOC for every Lagrange multiplier. A relaxed version of the CRCQ has been defined in Minchenko & Stakhovski (2011), called relaxed CRCQ (RCRCQ), which also enjoys the same second-order property (Andreani *et al.*, 2010b, 2012b). In RCRCQ, fewer subsets should conform to the constant rank property, namely, subsets that include gradients of every equality constraint. 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. Spivak, 1965, Theorem 2.13).

LEMMA 4.9 Assume that the gradients $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathscr{I}, j \in \mathscr{J}\}$ have locally constant rank for y in a neighborhood of some $x \in \mathbb{R}^n$. Then for each $d \in \mathbb{R}^n$ such that

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

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 $i \in \mathscr{I}$ and $j \in \mathscr{J}$ we have $h_i(\phi(t)) = h_i(x)$ and $g_j(\phi(t)) = g_j(x)$ for all $t \in (-T, T)$.

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

LEMMA 4.10 (Andreani *et al.*, 2012b, Lemma 1) Suppose that $v = \sum_{i\in\mathscr{I}} \alpha_i p_i + \sum_{j\in\mathscr{J}} \beta_j q_j$ with $p_i, q_j \in \mathbb{R}^n$ for every $i \in \mathscr{I}, j \in \mathscr{J}$, $\{p_i\}_{i\in\mathscr{I}}$ are linearly independent and α_i, β_j are nonzero for every $i \in \mathscr{I}, j \in \mathscr{J}$. Then there is a subset $\mathscr{J}' \subset \mathscr{J}$ and scalars $\hat{\alpha}_i, \hat{\beta}_j$ for all $i \in \mathscr{I}, j \in \mathscr{J}'$ such that

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

A useful characterization of RCRCQ is given below.

THEOREM 4.11 (Andreani *et al.*, 2012b, 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 $\operatorname{span}\{\nabla h_i(x) : i \in \{1, \ldots, m\}\}$. 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 4.12 RCRCQ implies CCP2.

Proof. Let (w, W) be an element of $\limsup K_2^{\mathscr{W}}(x)$ 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 $\mathscr{C}^{\mathscr{W}}(x^k, x^*)$.
Take an index subset $\mathscr{I} \subset \{1, \ldots, m\}$ such that the gradients $\{\nabla h_i(x^*) : i \in \mathscr{I}\}$ form a linear basis for the subspace generated by $\{\nabla h_i(x^*) : i \in \{1, \ldots, m\}\}$. From continuity $\{\nabla h_i(x^k) : i \in \mathscr{I}\}$ is linearly independent for k large enough. By Theorem 4.11 item (a), we have that $\{\nabla h_i(x^k) : i \in \mathscr{I}\}$ is a linear

basis for the subspace generated by $\{\nabla h_i(x^k) : i \in \{1,\dots,m\}\}$, for k sufficiently large. Then, there is a sequence $\{\bar\lambda_i^k : i \in \mathscr{I}\} \subset \mathbb{R}$ such that $\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) = \sum_{i\in\mathscr{I}} \bar\lambda_i^k \nabla h_i(x^k)$. So, we may write

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

Applying Lemma 4.10 to the expression above, we find a subset $\mathscr{J}_k \subset A(x^*)$ and multipliers $\hat\lambda_i^k \in \mathbb{R}, i \in \mathscr{I}$ and $\hat\mu_j^k \in \mathbb{R}_+, j \in \mathscr{J}_k$ for k large enough such that

$$w^k = \sum_{i\in\mathscr{I}} \hat\lambda_i^k \nabla h_i(x^k) + \sum_{j\in \mathscr{J}_k} \hat\mu_j^k \nabla g_j(x^k), \tag{4.13}$$

and $\{\nabla h_i(x^k), \nabla g_j(x^k) : i \in \mathscr{I}, j \in \mathscr{J}_k\}$ is a linearly independent set. Since $A(x^*)$ is a finite index set, we may take $\mathscr{J} := \mathscr{J}_k$ for an appropriate subsequence. By RCRCQ (Theorem 4.11, item (b), we have that $\{\nabla h_i(x^*), \nabla g_j(x^*) : i \in \mathscr{I}, j \in \mathscr{J}\}$ is a linearly independent set and as a consequence $\{\hat\lambda_i^k, \hat\mu_j^k : i \in \mathscr{I}, j \in \mathscr{J}\}_{k\in\mathbb{N}}$ form 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 the limit in (4.13) we get $w = \sum_{i\in\mathscr{I}} \lambda_i \nabla h_i(x^*) + \sum_{j\in\mathscr{J}} \mu_j \nabla g_j(x^*)$. Define $\hat\lambda_i^k = 0$ for $i \notin \mathscr{I}$ and $\hat\mu_j^k = 0$ for $j \notin \mathscr{J}$ for every $k \in \mathbb{N}$; also define $\lambda_i = 0$ for $i \notin \mathscr{I}$ and $\mu_j = 0$ for $j \notin \mathscr{J}$. Now we will prove that for every $d \in \mathscr{C}^{\mathscr{W}}(x^*)$ 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 $d \in \mathscr{C}^{\mathscr{W}}(x^*)$. Since RCRCQ implies the WCR condition, we have that there is a sequence $d^k \to d$ such that $d^k \in \mathscr{C}^{\mathscr{W}}(x^k, x^*)$, given by Lemma 3.8. Thus

$$H^k(d^k,d^k) \le \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k)(d^k,d^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla^2 g_j(x^k)(d^k,d^k) \tag{4.15}$$

$$\le \sum_{i=1}^m \hat\lambda_i^k \nabla^2 h_i(x^k)(d^k,d^k) + \sum_{j\in A(x^*)} \hat\mu_j^k \nabla^2 g_j(x^k)(d^k,d^k) + \Xi^k, \tag{4.16}$$

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 4.9, 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

$$\langle \nabla h_i(x^k), v^k \rangle + \nabla^2 h_i(x^k)(d^k, d^k) = 0 \text{ for } i \in \{1, \dots, m\}, \tag{4.18}$$

$$\langle \nabla g_j(x^k), v^k \rangle + \nabla^2 g_j(x^k)(d^k, d^k) = 0 \text{ for } j \in A(x^*). \tag{4.19}$$

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

$$\varXi^k = -\sum_{i=1}^{m} \varLambda_i^k \langle \nabla h_i(x^k), v^k \rangle - \sum_{j \in A(x^*)} \varUpsilon_j^k \langle \nabla g_j(x^k), v^k \rangle \tag{4.20}$$

$$= -\left\langle \sum_{i=1}^{m} \varLambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \varUpsilon_j^k \nabla g_j(x^k), v^k \right\rangle = 0, \tag{4.21}$$

where in the last equality we have used (4.14). Now, since $\varXi^k = 0$ 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 the limit in (4.22), the assertion is proved. □

EXAMPLE 4.13 (RCRCQ is strictly stronger than CCP2)

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

$$h_1(x_1, x_2) = x_1;$$
$$g_1(x_1, x_2) = -x_1^2 + x_2;$$
$$g_2(x_1, x_2) = -x_1^2 + x_2^3.$$

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 $\mathscr{C}^{\mathscr{W}}(x, x^*) = \{0\}$, we get $K_2^{\mathscr{W}}(x) = \mathbb{R} \times \mathbb{R}_+ \times \mathrm{Sym}(2)$. Clearly, $K_2^{\mathscr{W}}(x)$ is osc on $\mathbb{R}^2$.

REMARK 4.14 We have just proved that CCP2 is a constraint qualification that yields WSOC 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 WSOC, as proved in Theorem 4.2. 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 algorithm that generates AKKT2 sequences, even when the set of Lagrange multipliers at the limit point is unbounded. This comes from the fact that the MFCQ assumption can be dropped.

Now, suppose that we want a condition that guarantees that every limit point of any AKKT2 sequence fulfills not only WSOC but also the strong second-order condition, SSOC, even though, to the best of our knowledge, no algorithm has been shown to converge to a point where the SSOC holds. With this in mind we shall define the next constraint qualification in the spirit of Theorem 4.2, replacing WSOC with SSOC. Our goal is to understand why algorithms are not expected to converge to a point fulfilling SSOC.

DEFINITION 4.15 We say that the strong CCP2 (SCCP2) holds at $x^* \in \Omega$ if

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

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

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

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

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

We note that the critical cone $\mathscr{C}^S(x^*, \mu)$ is well defined for every $\mu \geq 0$, 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 $\mathscr{C}^S(x^*)$ instead of $\mathscr{C}^S(x^*, \mu)$. 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 $\mathscr{C}^S(x^*)$ and $\mathscr{C}^{\mathscr{W}}(x^*)$ coincide and SSOC is equivalent to WSOC.

We observe that SSOC holds at x^* for problem (1.1) if and only if the pair $(-\nabla f(x^*), -\nabla^2 f(x^*))$ belongs to $K_2^S(x^*)$. We also note that $K_2^S(x^*)$ is a subset of $K_2^{\mathscr{W}}(x^*)$, due to $\mathscr{C}^{\mathscr{W}}(x^*) \subset \mathscr{C}^S(x^*, \mu)$ for every $\mu \geq 0$ and as a consequence the SCCP2 condition is stronger than CCP2.
Following the same reasoning as Theorem 4.2 we obtain the following theorem.

THEOREM 4.16 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 SSOC holds at x^*.

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

EXAMPLE 4.17 (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.

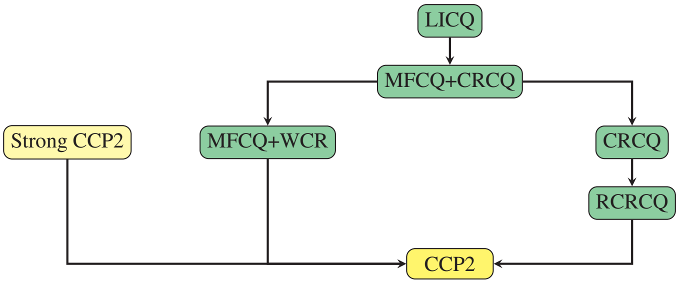

FIG. 1. Relationship of CQs associated with second-order global convergence of algorithms.

Clearly the set Ω is defined by the inequality constraints $g_j(x) = -x_j$ for $j = \{1, \dots, n\}$. Now we will calculate the weak cone $K_2^{\mathscr{W}}(x)$. Thus, we have to calculate the cone $\mathscr{C}^{\mathscr{W}}$ given by (4.1); in fact, $\mathscr{C}^{\mathscr{W}}(x, x^*) := \{d \in \mathbb{R}^n : \langle \nabla g_i(x), d \rangle = 0 \text{ for all } i \in A(x^*) = \{1, \dots, n\}\}$. Since $x^* = 0$, the set of active indexes $A(x^*)$ is $\{1, \dots, n\}$. 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 $\mathscr{C}^{\mathscr{W}}(x, x^*) = \{0\}$, and as a consequence the weak cone

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

is equal to $\mathbb{R}^n_- \times \text{Sym(n)}$ independently of x. Thus $K_2^{\mathscr{W}}$ is osc at x^* and CCP2 holds. Furthermore, since $\limsup_{x \to x^*} K_2^{\mathscr{W}}(x) = \mathbb{R}^m_- \times \text{Sym}(n)$, 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 $H(w, w) := w^{\mathrm{T}} H w > 0$ for some $w \in \mathscr{C}^S(x^*, \hat{\mu})$, because in this case, the pair $(-\hat{\mu}, H) \in K_2^{\mathscr{W}}(x) = \mathbb{R}^m_- \times \text{Sym}(n)$ but $(-\hat{\mu}, H)$ does not belong to $K_2^S(x^*)$. Choose $\hat{\mu} := e - e_1$ and $H := e_1 e_1^{\mathrm{T}}$. From the definition of the strong critical cone $\mathscr{C}^S(x^*, \hat{\mu})$, we have that $e_1 \in \mathscr{C}^S(x^*, \hat{\mu})$ 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 $\limsup_{x \to x^*} K_2^{\mathscr{W}}(x) = \mathbb{R}^n_- \times \text{Sym}(n)$ but it does not belong to the critical cone $K_2^S(x^*)$.

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

EXAMPLE 4.18 (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, $\mathscr{C}^{\mathscr{W}}(x, x^*) = \{0\}$ and $\mathscr{C}^S(x^*, \mu) = \{0\}$ for all $\mu \in \mathbb{R}^2_+$, so $K^{\mathscr{W}}(x) = \mathbb{R} \times \text{Sym}(1) = K^S(x^*)$ for all $x \in \mathbb{R}$, which implies the SCCP2 holds. On the other hand, clearly, LICQ fails.

In Fig. 1, we show the relationships among the constraint qualifications discussed in this article.

5. Algorithms that generate AKKT2 points

In this section, we show several algorithms in the literature that generate sequences whose limit points satisfy the sequential second-order optimality condition AKKT2. Besides the augmented Lagrangian algorithm of Andreani *et al.* (2010a), we will show that the regularized SQP of Gill *et al.* (2013) and the trust region method of Dennis & Vicente (1997) generate AKKT2 sequences.

For each of the aforementioned algorithms, we will use the same notation as the original paper, in order to facilitate the verification.

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 types of SQP methods were designed: regularized and stabilized SQP; see Izmailov & Solodov (2012); Gill & Robinson (2013). Gill *et al.* (2013) extended the regularized SQP method of Gill & Robinson (2013) to allow convergence to points satisfying the WSOC condition under the constraint qualification MFCQ+WCR. See also Kungurtsev (2013) and Gill *et al.* (2017).

Let us show that the method proposed by Gill *et al.* (2013) generates sequences that satisfy the sequential second-order optimality condition AKKT2. The problem analysed is

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

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 Gill *et al.* (2013). Let $H(x, \lambda) := \nabla^2 f(x) - \sum \lambda_i \nabla^2 c_i(x)$, $J(x)^{\mathrm{T}}$ be the matrix whose 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 $\widetilde{\mathscr{C}}(x) = \{d : J(x)^{\mathrm{T}} d = 0, d_j = 0 \text{ for } j \in A(x^*)\}$. Given positive scalar γ and $\varepsilon_a \in (0, 1)$, the ε-active set is defined as $\mathscr{A}_\varepsilon(x, \lambda, \mu) = \{i : x_i \leq \varepsilon, \text{ with } \varepsilon = \min(\varepsilon_a, \max(\mu, r(x, \lambda)^\gamma))\}$. The ε-free set is defined as $\mathscr{F}_\varepsilon(x, \lambda, \mu) := \{1, \ldots, n\} \setminus \mathscr{A}_\varepsilon(x, \lambda, \mu)$. The proposed algorithm in Gill *et al.* (2013) is based on the first-order primal–dual SQP method of Gill & Robinson (2013). 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

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

where ν is a non-negative 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 Gill *et al.* (2013), expression (2.1). The matrix $\hat{B}(x, \lambda; \mu)$ is a positive-definite matrix equal to $B(x, \lambda; \mu)$ when $B(x, \lambda; \mu)$ is sufficiently positive definite, otherwise it takes a specific form (see Gill *et al.*, 2013, expression (2.3)) that depends on a matrix $\hat{H}(x, \lambda)$ such that $\hat{H}(x, \lambda) + \mu^{-1} J(x) J(x)^{\mathrm{T}}$ is positive definite; cf. Gill & Robinson (2013), 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 kth estimate of a primal–dual solution of problem (5.1). Each iterate can be classified as V-, O-, M- or F-iterates (see Gill *et al.*, 2013, Algorithm 3), where the union of index sets of V-, O- and M-iterates is always infinite (Gill *et al.*, 2013, Theorem 3.2). Numerical experiments indicate that M-iterates occur infrequently relative to the total number of iterations. We give a summary of (Gill *et al.*, 2013, Algorithm 3) in Algorithm 2.

Algorithm 2 (Gill *et al.*, 2013, Algorithm summary)

The computation associated with the kth 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 $\mathscr{F}_\varepsilon(x^k, \lambda^k, \mu^R_{k-1})$ 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 non-negative 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 Gill *et al.*, 2013, 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 kth iteration (see Gill *et al.*, 2013, 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 Gill *et al.*, 2013, 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 Gill *et al.*, 2013, 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 Gill *et al.*, 2013, 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 bounded by a positive constant and the sequence of lowest eigenvalue of $\hat{H}(x^k, \lambda^k) + (1/\mu^R_k)J(x^k)J(x^k)^{\mathrm{T}}$ is bounded below by a positive constant 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 (Gill *et al.*, 2013, proof of Theorem 3.4). Let $\{v^k = (x^k, \lambda^k)\}$ be the sequence generated by Algorithm 3 of Gill *et al.* (2013) and suppose that the algorithm generates infinitely many V- or O-iterates. Let $\{(x^k, \lambda^k)\}$ be a sequence such that every iterate is a V- or O-iterate and x^* be a limit point of $\{x^k\}$. So we have, from Algorithm 3 of Gill *et al.* (2013), since the quantities $\phi^{\max}_V$ and $\phi^{\max}_O$ are positive bounds that are reduced by half during the solution process (see Gill *et al.* (2013), (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 expressions Gill *et al.* (2013), (3.25) and Gill *et al.* (2013), (2.6) we deduce that

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

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

$$H(x^k,\lambda^k) + \frac{1}{\mu_{k-1}^R} J_k J_k^{\mathrm{T}} + \left(\frac{1}{\theta}\epsilon_k^{(1)} + \frac{1}{k}\right)\mathbb{I} + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j \nabla g_j^{\mathrm{T}} \succ 0,$$

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

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

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 algorithm generates AKKT2 sequences. The algorithm is the one proposed by Dennis & Vicente (1997), which is an extension of the work of Dennis *et al.* (1997). They only consider equality constraints

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

We use the same notation as Dennis & Vicente (1997). Let $C : \mathbb{R}^n \to \mathbb{R}^m$ $(m < n)$, $C = (c_1, \ldots, c_m)^{\mathrm{T}}$ 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 $\mathrm{Ker}\nabla C(x^k)^{\mathrm{T}}$. Let H_k be an approximation to $\nabla^2 \ell(x^k,\mu^k)$, $\hat{H}_k = W_k^{\mathrm{T}} H_k W_k$ and $\hat{g}_k = W_k^{\mathrm{T}} \nabla q_k(s_k^n)$, q_k a 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 the s_k step of the method. See Dennis & Vicente (1997), Section 2. The general trust-region algorithm is given by Algorithm 3.

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*.

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

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

H.3 Functions $f, \nabla f, \nabla^2 f, C, \nabla C, \nabla^2 c_i$, $i = 1, \ldots, m$ are bounded in $\hat{\Omega}$. The matrix defined by $(\nabla C(x)^{\mathrm{T}} \nabla C(x))^{-1}$ is bounded below by a positive constant over $\hat{\Omega}$.

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

H.5 The Hessian approximation H_k is exact, that is, $H_k = \nabla_{xx}^2 \ell_k$, and $\nabla^2 f$ and $\nabla^2 c_i, i = 1, \ldots, m$ are Lipschitz continuous in $\hat{\Omega}$.

Now, we will prove that the method generates AKKT2 sequences when the Lagrange multipliers are updated in a consistent way (Dennis & Vicente, 1997, (4.7)). First, we will prove that (3.2) from the

Algorithm 3 (Dennis & Vicente, 1997, 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 $k = 0, 1, 2, \ldots$ do
   (a) If $\|W_k^{\mathrm{T}} \nabla \ell(x^k, \lambda^k)\| + \|C(x^k)\| + \gamma_k = 0$ where γ_k is given by (Dennis & Vicente, 1997, (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 (Dennis & Vicente, 1997, (2.1),(2.2),(2.3)) and $\|s_k^n\| \leq r\delta_k$.
   If $\|W_k^{\mathrm{T}} \nabla \ell(x^k, \lambda^k)\| + \gamma_k \neq 0$ then compute $\bar{s}_k^t$ satisfying (Dennis & Vicente, 1997, (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 (Dennis & Vicente, 1997, (2.8)).
   (d) Compute $pred(s_k, \rho_{k-1})$. See (Dennis & Vicente, 1997, 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 $\|\triangle\lambda^k\| \leq \kappa_3 \delta_k$.

definition of AKKT2 holds for $\{\lambda_k\}$ satisfying only Dennis & Vicente (1997), (2.8). From the Karush–Kuhn–Tucker conditions there exists a $\gamma_k \geq 0$ (Dennis & Vicente, 1997, (2.10)), such that

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

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

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

By Dennis & Vicente (1997), Theorem 3.10, $\liminf(\|W_k^{\mathrm{T}} \nabla \ell(x^k, \lambda^k)\| + \|C(x^k)\| + \gamma_k) = 0$. 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 $x^* \in \mathbb{R}^n$,

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

From $\gamma_k \to 0$ and (5.4), (3.2) holds. To prove that (3.1) is fulfilled we choose the Lagrange multipliers λ^k as (Dennis & Vicente, 1997, Lemma 4.2), that is , $\lambda^k = -(\nabla C(x^k)^{\mathrm{T}}\nabla C(x^k))^{-1}\nabla C_k^{\mathrm{T}}\nabla f(x^k)$. 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 $\mathrm{Ker}(\nabla C(x^k)^{\mathrm{T}})$ and $\nabla C(x^k)v^k$ belongs to $\mathrm{Ker}(\nabla C(x^k)^{\mathrm{T}})^{\perp} = \mathrm{Im}(\nabla C(x^k))$ for some u^k, v^k. Multipliying the expression (5.6) by $(u^k)^{\mathrm{T}} W_k^{\mathrm{T}}$ and using $\lim \|W_k^{\mathrm{T}}\nabla_x \ell(x^k, \lambda^k)\| = 0$, we have that $W_k u^k \to 0$. Now, we proceed to multiply (5.6) by $\nabla C(x^k)^{\mathrm{T}}$ and use the existence of the inverse $(\nabla C(x^k)^{\mathrm{T}}\nabla C(x^k))^{-1}$ to get $v^k = (\nabla C(x)^{\mathrm{T}}\nabla C(x))^{-1}\nabla C_k^{\mathrm{T}}\nabla f(x^k) + \lambda^k = 0$. 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 Dennis *et al.* (1997); El-Alem (1996), also generate AKKT2 sequences.

6. Final remarks

Over the years, several algorithms with convergence to second-order stationary points have been proposed in the literature. Their global convergence is guaranteed by using strong constraint qualification, as LICQ or MFCQ+WCR, which imply boundedness of the set of Lagrange multipliers. Guided by the necessity of explaining some observed aspects of these methods, we took a closer look into their stopping criteria and the associated sequential second-order optimality condition AKKT2. We were able to prove second-order global convergence of such algorithms under CCP2, a constraint qualification that is weaker than both MFCQ+WCR and RCRCQ. In particular, it does not imply boundedness of the set of Lagrange multipliers. This framework also gives a tool to prove second-order global convergence results of other second-order algorithms under CCP2. 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 that AKKT does for first-order methods.

Funding

This work was supported by PRONEX-Optimization (PRONEX-CNPq/FAPERJ E-26/111.449/2010-APQ1), CEPID-CeMEAI (FAPESP 2013/07375-0), National Council for Scientific and Technological Development (CNPq) (Grants 300907/2012-5, 481992/2013-8, 304618/2013-6, 482549/2013-0, and 303013/2013-3), São Paulo Research Foundation (FAPESP) (Grants 2010/19720-5, 2012/20339-0 and 2013/05475-7), Coordination for the Improvement of Higher Level Education Personnel (CAPES), and National Project on Industrial Mathematics.

References

Andreani, R., Behling, R., Haeser, G. & Silva, P. J. S. (2014a) On second-order optimality conditions for nonlinear optimization. *Optim. Meth. and Soft.*, **32**, 22–38.

Andreani, R., Birgin, E. G., Martinez, J. M. & Schuverdt, M. L. (2010a) Second-order negative-curvature methods for box-constrained and general constrained optimization. *Comput. Optim. Appl.*, **45**, 209–236.

Andreani, R., Echague, C. E. & Schuverdt, M. L. (2010b) Constant-rank condition and second-order constraint qualification. *J. Optim. Theory Appl.*, **146**, 255–266.

Andreani, R., Haeser, G. & Martinez, J. M. (2011) On sequential optimality conditions for smooth constrained optimization. *Optimization*, **60**, 627–641.

Andreani, R., Haeser, G., Schuverdt, M. L. & Silva, P. J. S. (2012a) Two new weak constraint qualifications and applications. *SIAM J. Optim.*, **22**, 1109–1135.

Andreani, R., Haeser, G., Schuverdt, M. L. & Silva, P. J. S. (2012b) A relaxed constant positive linear dependence constraint qualification and applications. *Math. Program.*, **135**, 255–273.

Andreani, R., Martinez, J., Ramos, A. & Silva, P. (2016) A cone-continuity constraint qualification and algorithmic consequences. *SIAM J. Optim.*, **26**, 96–110.

Andreani, R., Martinez, J. M., Santos, L. T. & Svaiter, B. F. (2014b) On the behaviour of constrained optimization methods when Lagrange multipliers do not exist. *Optim. Methods Softw.*, **29**, 646–657.

Andreani, R., Martinez, J. M. & Schuverdt, M. L. (2007) On second-order optimality conditions for nonlinear programming. *Optimization*, **56**, 529–542.

Anitescu, M. (2000) Degenerate nonlinear programming with a quadratic growth condition. *SIAM J. Optim.*, **10**, 1116–1135.

Arutyunov, A. V. (1991) Pertubations of extremum problems with constraints and necessary optimality conditions. *J. Soviet Math.*, **54**, 1342–1400.

Arutyunov, A. V. (1998) Second-order conditions in extremal problems. The abnormal points. *Trans. Am. Math. Soc.*, **350**, 4341–4365.

Arutyunov, A. V. (2000) *Optimality Conditions: Abnormal and Degenerate Problems*. Dordrecht: Kluwer Academic Publishers.

Arutyunov, A. V. & Pereira, F. L. (2006) Second-order necessary optimality conditions for problems without a priori normality assumptions. *Math. Oper. Res.*, **31**, 1–12.

Auslender, A. (1979) Penalty methods for computing points that satisfy second-order necessary conditions. *Math. Program.*, **17**, 229–238.

Baccari, A. & Trad, A. (2005) On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints. *SIAM J. Optim.*, **15**, 394–408.

Bazaraa, M. S., Sherali, H. D. & Shetty, C. M. (2006) *Practical Methods of Optimization: Theory and Algorithms*. NJ: John Wiley.

Ben-Tal, A. & Zowe, J. (1982) A unified theory of first and second-order conditions for extremum problems in topological vector spaces. *Math. Program. Study*, **19**, 39–76.

Bertsekas, D. P. (1982) *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic Press.

Bertsekas, D. P. (1999) *Nonlinear Programming*. Belmont, MA: Athenas Scientific.

Birgin, E. & Martinez, J. M. (2014) *Practical Augmented Lagrangian Methods for Constrained Optimization*. Philadelphia, PA: SIAM Publications.

Bonnans, J. F., Cominetti, R. & Shapiro, A. (1999) Second-order optimality conditions based on parabolic second-order tangent sets. *SIAM J. Optim.*, **9**, 466–492.

Bonnans, J. F., Hermant, A., Arutyunov, A. V. & Pereira, F. L. (2009) No-gap second-order optimality conditions for optimal control problems with a single state constraint and control. *Math. Program., Serie B.*, **117**, 21–50.

Bonnans, J. F. & Shapiro, A. (2000) *Pertubation Analysis of Optimization Problems*. Berlin: Springer.

Byrd, R. H., Schnabel, R. B. & Shultz, G. A. (1987) A trust region algorithm for nonlinearly constrained optimization. *SIAM J. Numer. Anal.*, **24**, 1152–1170.

Casas, E. & Troltzsch, F. (2002) Second-order necessary and sufficient optimality conditions for optimization problems and applications to control theory. *SIAM J. Optim.*, **13**, 406–431.

Coleman, T. F., Liu, J. & Yuan, W. (2002) A new trust-region algorithm for equality constrained optimization. *Comput. Optim. Appl.*, **21**, 177–199.

Conn, A. R., Gould, N. I. M. & Toint, Ph. L. (1999) A primal-dual algorithm for minimizing a nonconvex function subject to bound and linear equality constraints. In *Nonlinear Optimization and Related Topics* (G. Di Pillo and F. Giannnessi, eds), Kluwer Academic Publishers, pp. 15–50.

Conn, A. R., Gould, N. I. M. & Toint, P. L. (ed.) (2000) *Trust Region Methods*. MPS/SIAM Series on Optimization, Philadelphia, PA: SIAM.

Debreu, G. (1952) Definite and semidefinite quadratic forms. *Econometrica*, **20**, 295–300.

DENNIS, J. E., EL-ALEM, M. & MACIEL, M. C. (1997) A global convergence theory for general trust-region-based algorithms for equality constrained optimization. *SIAM J. Optim.*, **7**, 177–207.

DENNIS, J. E. & VICENTE, L. N. (1997) On the convergence theory of trust-region-based algorithms for equality-constrained optimization. *SIAM J. Optim.*, **7**, 927–950.

DIPILLO, G., LUCIDI, S. & PALAGI, L. (2005) Convergence to second-order stationary points of a primal–dual algorithm model for nonlinear programming. *Math. Oper. Res.*, **30**, 897–915.

EL-ALEM, M. M. (1996) Convergence to a second-order point of a trust-region algorithm with a nonmonotonic penalty parameter for constrained optimization. *J. Optim. Theory Appl.*, **91**, 61–79.

FACCHINEI, F. & LUCIDI, S. (1998) Convergence to second-order stationary points in inequality constrained optimization. *Math. Oper. Res.*, **23**, 746–766.

FIACCO, A. V. & MCCORMICK, G. P. (1968) *Nonlinear Programming Sequential Unconstrained Minimization Techniques*. New York: John Wiley.

FLETCHER, R. (1981) *Practical Methods of Optimization: Constrained Optimization*. New York: Wiley.

GILL, P. E., KUNGURTSEV, V. & ROBINSON, D. P. (2013) A regularized SQP method with convergence to second-order optimal points. *Technical Report CCoM-13-4. Center for Computational Mathematics, University of California, San Diego*. Hoboken NJ, USA: John Wiley and Sons.

GILL, P. E. & ROBINSON, D. P. (2013) A globally convergent stabilized SQP method. *SIAM J. Optim.*, **23**, 1983–2013.

GILL, P. E., KUNGURTSEV, V. & ROBINSON D. P. (2017) A stabilized SQP method: global convergence. *IMA J. Numer. Anal.*, **37**, 407–443.

GOULD, N. I. M. & TOINT, P. L. (1998) A note on the convergence of barrier algorithms for second-order necessary points. *Math. Program.*, **85**, 433–438.

IZMAILOV, A. F. & SOLODOV, M. V. (2012) Stabilized SQP revisited. *Math. Program.*, **133**, 93–120.

JANIN, R. (1984) Direction derivative of the marginal function in nonlinear programming. *Math. Program. Studies*, **21**, 127–138.

KUNGURTSEV, V. (2013) Second-derivative sequential quadratic programming methods for nonlinear optimization. *Ph.D. Thesis*, University of California, San Diego CA, USA.

MANGASARIAN, O. L. & FROMOVITZ, S. (1967) The Fritz-John necessary conditions in presence of equality and inequality constraints. *J. Math. Anal. Appl.*, **17**, 37–47.

MARTINEZ, J. M. & SVAITER, B. F. (2003) A practical optimality condition without constraint qualifications for nonlinear programming. *J. Optim. Theory Appl.*, **118**, 117–133.

MINCHENKO, L. & STAKHOVSKI, S. (2011) On relaxed constant rank regularity condition in mathematical programming. *Optimization*, **60**, 429–440.

MORGUERZA, J. M. & PRIETO, F. J. (2003) An augmented Lagrangian interior-point method using directions of negative curvature. *Math. Program.*, **95**, 573–616.

MURTY, K. G. & KABADI, S. N. (1987) Some NP-complete problems in quadratic and nonlinear programming. *Math. Program.*, **39**, 117–129.

NOCEDAL, J. & WRIGHT, S. J. (2006) *Numerical Optimization*. New York: Springer.

PENOT, J. P. (1994) Optimality conditions in mathematical programming and composite optimization. *Math. Program.*, **67**, 225–245.

PENOT, J. P. (1998) Second-order conditions for optimization problems with constraints. *SIAM J. Control Optim.*, **37**, 303–318.

PETERSEN, D. W. (1973) A review of constraint qualifications in finite-dimensional spaces. *SIAM Rev.*, **15**, 639–654.

QI, L. & WEI, Z. (2000) On the constant positive linear dependence conditions and its application to SQP methods. *SIAM J. Optim.*, **10**, 963–981.

RAMOS, A. (2016) *Tópicos em Condições de Otimalidade para Otimização não Linear*. *Ph.D. Thesis*, IME-USP, Departamento de Matemática Aplicada, São Paulo-SP, Brazil.

SOLODOV, M. V. (2010) Constraint qualifications. In *Wiley Encyclopedia of Operations Research and Management Science*. James J. Cochran, *et al*. (eds), John Wiley.

SPIVAK, M. (1965) *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. New York: Addison-Wesley.