Some theoretical limitations of second-order algorithms for smooth constrained optimization

G. Haeser*


Abstract

In second-order algorithms, we investigate the relevance of the constant rank of the full set of active constraints in ensuring global convergence to a second-order stationary point. We show that second-order stationarity is not expected in the non-constant rank case if the growth of the so-called tangent sequence, associated with a second-order complementarity measure, is not controlled. We then investigate how these parameters should be controlled in order for the second-order information to remain present. Since no algorithm directly controls the growth of the tangent sequence, we argue that there is a theoretical limitation of present algorithms in finding second-order stationary points beyond the constant rank case.

Keywords: Global convergence, Second-order algorithms, Constant rank, Second-order optimality conditions

1 Introduction

In this paper we are interested in optimality properties of limit points of sequences generated by numerical algorithms, that is, global convergence results. Let us assume that $x^*$ is a feasible limit point of a sequence $\{x^k\}$ generated by an unspecified first-order numerical algorithm. It is well known that assuming that $x^*$ satisfies some constraint qualification is not enough to ensure that $x^*$ is a Karush-Kuhn-Tucker (KKT) point. Some specific constraint qualification, what has been called a strict one in [4], is needed. Similarly, if the algorithm is a second-order one, the Weak Second-order Optimality Condition (WSOC), the standard necessary optimality condition based on the critical subspace, can usually be proved at $x^*$ under a constraint qualification stronger than what is needed to prove that WSOC is necessary at a local minimizer (see [8]).

In this sense, there is an intrinsic theoretical gap between the types of problems in which local solutions are known to be KKT or WSOC points, and the

---

*Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil. Email: ghaeser@ime.usp.br. This research was conducted while holding a Visiting Scholar position at Department of Management Science and Engineering, Stanford University, Stanford CA, USA.
attainability of such points by a numerical algorithm. There is also a gap in
the sense that the second-order optimality condition can be theoretically refined
to consider a larger set of directions, namely, the critical cone, in which the
critical subspace is its lineality space. This cone encompasses the true second-
order information at \( x^* \), since this is the cone associated to sufficient optimality
conditions, but no algorithm is known to satisfy this stronger condition on its
limit points, so we only consider a gap with respect to fulfilling WSOC.

Of course, there are also numerical limitations of an algorithm, in the sense
that these borderline problems can be very degenerate and ill-conditioned, which
can cause an algorithm to fail in practice. These are not the types of deficiencies
that we are interested in this paper.

In particular, we discuss the role of the constant rank of the active constraints
gradients in ensuring WSOC, arguing that it is not expected to hold beyond
the non-constant rank case even when a Lagrange multiplier is known. Section
2 presents the basic framework regarding second-order optimality conditions
while Section 3 presents our main results. Section 4 presents a discussion while
Section 5 presents some concluding remarks.

## 2 Second-order optimality conditions

Let us consider the nonlinear optimization problem

\[
\begin{align*}
\text{Minimize} & \quad f(x), \\
\text{subject to} & \quad h(x) = 0, \\
& \quad g(x) \leq 0,
\end{align*}
\]

where \( f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p \) are twice continuously differentiable functions.

If the functions defining the constraint set satisfy a constraint qualification
at a local minimizer \( x^* \), there exist Lagrange multipliers \( (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \) such that the Karush-Kuhn-Tucker (KKT) conditions hold, that is

\[
\nabla L(x^*, \lambda, \mu) = 0 \quad \text{with} \quad \mu^T g(x^*) = 0,
\]

where \( x \mapsto L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x) \) is the Lagrangian function.

Under some more specific constraint qualifications (see [8]), the Weak Second-
order necessary Optimality Condition (WSOC) holds at some Lagrange multiplier \( (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \):

\[
d^T \nabla^2 L(x^*, \lambda, \mu) d \geq 0 \quad \text{for all} \quad d \in S(x^*),
\]

where

\[
S(x^*) = \{ d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m, \nabla g_i(x^*)^T d = 0, i \in A(x^*) \}
\]
is the critical subspace and \( A(x^*) = \{ i \in \{1, \ldots, p\} \mid g_i(x^*) = 0 \} \) is the index
set of active inequality constraints at \( x^* \).
When no constraint qualification is assumed, the usual notion of a first-order optimality condition is the so-called Fritz-John necessary optimality condition, which says that there exist Fritz-John multipliers \((\lambda_0, \lambda, \mu) \in \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}_+^p\) with \(\|(\lambda_0, \lambda, \mu)\| = 1\) such that

\[
\nabla L_g(x^*, \lambda_0, \lambda, \mu) = 0 \text{ with } \mu^\top g(x^*) = 0,
\]

where \(x \mapsto L_g(x, \lambda_0, \lambda, \mu) = \lambda_0 f(x) + \lambda^\top h(x) + \mu^\top g(x)\) is the generalized Lagrangian function. Note that this condition can cope with an unbounded Lagrange multiplier at a local minimizer \(x^*\) by means of a (bounded) Fritz-John multiplier with \(\lambda_0 = 0\), in the sense that if \(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla g_i(x^k) \to 0\) with \(\mu_i^k \to 0\) for \(i \in A(x^*), x^k \to x^*\) and \(\|(\lambda_i^k, \mu_i^k)\| \to +\infty\), dividing by \(\|(1, \lambda_i^k, \mu_i^k)\|\) and taking the limit of an appropriate subsequence, we get Fritz-John multipliers with \(\lambda_0 = 0\).

Note that when \(\lambda_0 = 0\), the optimality condition does not depend on the objective function, and it merely describes a specific type of degeneracy of the constraint functions, that is, that Mangasarian-Fromovitz constraint qualification fails. However, when dealing explicitly with the aforementioned sequences, we can show that \(x^*\) is a KKT point under the continuity of the KKT cone \([3]\), even when the approximate Lagrange multiplier sequence is unbounded, which is a strictly weaker assumption than Mangasarian-Fromovitz constraint qualification.

In the second-order case, the situation is not so straightforward. In optimization theory, a necessary (and sufficient, under a minor modification) second-order optimality condition, without constraint qualifications, is well studied. Let us denote by \(\Lambda_g(x^*)\) the set of all Fritz-John multipliers at a feasible point \(x^*\). The second-order Fritz-John necessary optimality condition \([11]\) is said to hold at \(x^*\) when \(\Lambda_g(x^*) \neq \emptyset\) and

\[
\max_{(\lambda_0, \lambda, \mu) \in \Lambda_g(x^*)} \langle d^\top \nabla^2 L_g(x^*, \lambda_0, \lambda, \mu) d \rangle \geq 0 \text{ for all } d \in C(x^*)\setminus\{0\},
\]

where \(C(x^*)\) is the critical cone \(C(x^*) = \{d \in \mathbb{R}^n \mid \nabla f(x^*)^\top d = 0, \nabla h_i(x^*)^\top d = 0, i = 1, \ldots, m, \nabla g_i(x^*)^\top d \leq 0, i \in A(x^*)\}\). This gives a necessary optimality condition for local optimality, and also a sufficient one if we replace the inequality sign \(\geq 0\) by the strict one \(> 0\), so it can be argued that it really captures all second-order information available, without constraint qualifications. Nonetheless, this optimality condition is not relevant to our considerations since: i) it relies on the true critical cone \(C(x^*)\), rather than the critical subspace \(S(x^*)\), which is intractable by numerical algorithms; ii) it includes a multiplier for the objective function, which is not usual in numerical algorithms. That is, algorithms usually treat objective function and constraints in different manners, in such a way that a multiplier for the objective function is not present; iii) it depends on the knowledge of the whole set of multipliers, and not on one particular approximation, as it is usually the case in an algorithm. See \([8]\) for more details.

Thus, due to its association with global convergence results, we consider...
the Second-order Approximate-KKT (AKKT2, [2]) optimality condition defined below:

**Definition 1.** Let \( x^* \) be a feasible point and a sequence \( \{(x^k, \lambda^k, \mu^k, \theta^k, \eta^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^p_+ \). Assume that \( x^k \to x^* \), \( \nabla L(x^k, \lambda^k, \mu^k) \to 0 \),

\[
\lim_{k \to +\infty} \lambda_{\text{min}}(\nabla^2 L(x^k, \lambda^k, \mu^k)) + \sum_{i=1}^m \theta^k_i \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{i=1}^p \eta^k_i \nabla g_i(x^k) \nabla g_i(x^k)^T \geq 0,
\]

(2.1)

where \( \lambda_{\text{min}}(Q) \) denotes the least eigenvalue of the symmetric matrix \( Q \), and

\[ \forall i = 1, \ldots, p, \text{ if } g_i(x^*) < 0, \text{ then } \mu^k_i = 0 \text{ and } \eta^k_i = 0 \text{ for sufficiently large } k. \]

(2.2)

When such sequences exist, we call \( x^* \) an AKKT2 point, \( \{x^k\} \) a primal, \( \{(\lambda^k, \mu^k)\} \) a dual, and \( \{ (\theta^k, \eta^k) \} \) a tangent AKKT2 sequences associated with \( x^* \).

Condition AKKT2 is an optimality condition that holds without constraint qualifications. Its relevance is associated with global convergence of second-order algorithms. Many second-order algorithms generate this type of sequences (trust region [13], augmented Lagrangian [1, 10], regularized sequential quadratic programming [15] and interior point [22, 9, 18] methods). See [2, 17]. A minimal companion constraint qualification is defined in [2] such that AKKT2 points are guaranteed to be standard WSOC points. This constraint qualification is weaker than previous constraint qualifications used to prove global convergence of a second-order algorithm to a WSOC point. See [2].

These tools are also relevant for the definition of a second-order stationary point within a tolerance \( \varepsilon > 0 \), which is relevant for the computation of worst-case complexity bounds, and also for the definition of optimality conditions for some degenerate problems. See [18].

Although one can argue that WSOC is too weak, in the sense that it is not based on the true critical cone \( C(x^*) \), a WSOC point has some important properties related to optimality in some statistical learning applications [19]. Also, it is believed that WSOC is the best optimality condition one can hope to achieve with a second-order algorithm, see [16].

### 3 Main results

In this paper we are interested in measuring how much of the true second-order information, meaning WSOC, the optimality condition AKKT2 can capture.

We start with the following basic definitions. Given a multifunction \( F : \mathbb{R}^n \rightharpoonup \mathbb{R}^s, x \mapsto F(x) \subseteq \mathbb{R}^s \), the **outer limit** of \( F(x) \) as \( x \to x^* \) is given by:

\[
\limsup_{x \to x^*} F(x) = \{ w^* \in \mathbb{R}^s : \exists \{ (x^k, w^k) \} \to (x^*, w^*) \text{ with } w^k \in F(x^k) \},
\]

while its **inner limit** is given by:

\[
\liminf_{x \to x^*} F(x) = \{ w^* \in \mathbb{R}^s : \forall \{ x^k \} \to x^*, \exists \{ w^k \} \to w^* \text{ with } w^k \in F(x^k) \}.
\]
The multifunction \( F(x) \) is said to be outer semicontinuous (osc) at \( x^* \) if \( \limsup_{x \to x^*} F(x) \subseteq F(x^*) \) and inner semicontinuous (isc) at \( x^* \) if \( F(x^*) \subseteq \liminf_{x \to x^*} F(x) \).

The next proposition is a characterization of AKKT2, which can be deduced from the proof of \[10\] Theorem 3.2, but we include a proof for completeness.

**Proposition 1 (\[10\]).** Given sequences \( \{x^k\} \to x^* \in \mathbb{R}^n \) and \( \{\lambda^k, \mu^k\} \in \mathbb{R}^m \times \mathbb{R}^+_n \) with \( \mu^k_i = 0 \) for sufficiently large \( k \) when \( g_i(x^*) < 0 \), the point \( x^* \) satisfies AKKT2 with primal-dual sequences \( \{x^k, \lambda^k, \mu^k\} \) if, and only if, \( \nabla L(x^k, \lambda^k, \mu^k) \to 0 \) and for some \( \varepsilon_k \to 0^+ \),

\[
d^2 \nabla^2 L(x^k, \lambda^k, \mu^k) d \geq -\varepsilon_k \|d\|^2,
\]

for all \( d \in S(x^k, x^*) \), where \( x \mapsto S(x, x^*) \) is the perturbed critical subspace multifunction given by

\[
S(x, x^*) = \left\{ d \in \mathbb{R}^n \left| \begin{array}{c}
\sum_{i=1}^m \theta_i^k \nabla h_i(x^k)^T d = 0, \text{ and } \\
\sum_{i \in A(x^*)} \eta_i^k \nabla g_i(x^k)^T d = 0
\end{array} \right. \right\}.
\]

**Proof:** From (2.1) and (2.2), there is a sequence \( \varepsilon_k \to 0^+ \) such that

\[
\nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \sum_{i=1}^m \theta_i^k \nabla h_i(x^k)^T + \sum_{i \in A(x^*)} \eta_i^k \nabla g_i(x^k) + \varepsilon_k I
\]

is a positive definite matrix for sufficiently large \( k \), where \( I \) is the identity matrix. Then (3.1) follows for all \( d \in S(x^k, x^*) \). Conversely, if (3.1) holds for all \( d \in S(x^k, x^*) \) it follows that \( d^T (\nabla^2 L(x^k, \lambda^k, \mu^k) + \varepsilon_k + 1/k) I d > 0 \) for all \( d \in S(x^k, x^*), d \neq 0 \). Then, it is well known that for each \( k \) there is a constant \( \theta_k > 0 \) such that

\[
\nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \theta_i \nabla h_i(x^k)^T + \sum_{i \in A(x^*)} \theta_i \nabla g_i(x^k)^T + (\varepsilon_k + 1/k) I
\]

is positive definite (see, for instance, \[17\] Proposition 2.1). Now (2.1) follows taking limit.

Let us consider the following example:

Minimize \( x \), s.t. \( \frac{x^2}{2} = 0 \).

We note that AKKT2 gives no second-order information at the global solution \( x^* = 0 \). Namely, there exist \( \{x^k\} \to x^* \) and \( \{\lambda^k\} \subset \mathbb{R} \) such that \( 1 + \lambda^k x^k \to 0 \), \( d^T \lambda^k d \geq 0 \) for all \( d \in S(x^k, x^*) = \{ d \in \mathbb{R} \ | \ d^T x^k = 0 \} \). Note that the perturbed critical subspace \( S(x^k, x^*) \) is the trivial set \{0\} when \( x^k \neq x^* \), thus not giving any second-order information. Note that at \( x^* \) the critical subspace is given by \( S(x^*) = \mathbb{R} \), hence the lack of continuity of the critical subspace seems to play a key role. The next result is a characterization of the continuity of the perturbed critical subspace by means of the constant rank of the gradients of active constraints. It is essentially done in \[13\] Theorem 3.2.9. See also \[20\] Appendix.\]
Proposition 2 ([14]). The multifunction \( x \mapsto S(x,x^*) \) is inner semicontinuous at \( x^* \) if, and only if, \( \{ \nabla h_i(x) \}_{i=1}^m \cup \{ \nabla g_i(x) \}_{i \in A(x^*)} \) has the same rank for all \( x \) in some neighborhood of \( x^* \).

Proof: The result follows from [14] Theorem 3.2.9, by noting that the polar of \( S(x,x^*) \) is given by \( S(x,x^*)^c = \{ w \in \mathbb{R}^n \mid w = \sum_{i=1}^m \alpha_i \nabla h_i(x) + \sum_{i \in A(x^*)} \beta_i \nabla g_i(x), \text{ for some } \alpha_i, \beta_i \in \mathbb{R} \} \), while the inner semicontinuity of \( S(x,x^*) \) is equivalent to the outer semicontinuity of its polar (see [7, Theorem 1.1.8]). □

The constant rank of the active constraints \( \{ \nabla h_i(x) \}_{i=1}^m \cup \{ \nabla g_i(x) \}_{i \in A(x^*)} \) is known as the Weak Constant Rank (WCR) property at \( x^* \), discussed in [5].

We note that the outer semicontinuity of \( S(x,x^*) \) always holds due to the continuity of the gradients of constraints, namely, if \( x^k \to x^* \) and \( d^k \to d \) with \( \nabla h_i(x^k)^T d_k = 0, i = 1, \ldots, m \) and \( \nabla g_i(x^k)^T d_k = 0, i \in A(x^*) \), then taking the limit we get \( d \in S(x^*) \). Therefore, WCR is equivalent to \( S(x^*) \subseteq \lim\inf_{x \to x^*} S(x,x^*) \subseteq \lim\sup_{x \to x^*} S(x,x^*) \subseteq S(x^*) \) and hence, to the continuity of \( S(x,x^*) \).

To understand the deficiencies associated with AKKT2, let us consider only cases where the dual sequence \( \{(\lambda^k, \mu^k)\} \) accurately approximates a true Lagrange multiplier.

An interesting result, due to the outer semicontinuity of \( S(x,x^*) \), is the following:

Proposition 3. Let \( x^* \) be a WSOC point with a Lagrange multiplier \( (\lambda, \mu) \). Then, every sequence \( \{x^k \} \to x^* \) and \( \{(\lambda^k, \mu^k)\} \to (\lambda, \mu) \), is a primal and, respectively, dual AKKT2 sequence associated with \( x^* \).

Proof: Clearly, \( \nabla L(x^k, \lambda^k, \mu^k) \to \nabla L(x^*, \lambda, \mu) = 0 \). Assume that there is some sequence \( d_k \in S(x^k, x^*), \|d^k\| = 1 \) and \( \varepsilon > 0 \) such that \( d^*_k \nabla^2 L(x^k, \lambda^k, \mu^k) d_k < -\varepsilon \). Take a subsequence such that \( d_k \to d \). From the outer semicontinuity of \( S(x,x^*) \), we have \( d \in S(x^*) \). Taking limit we get \( \varepsilon^T \nabla^2 L(x^*, \lambda, \mu) d \leq -\varepsilon \) which contradicts WSOC. □

In [2], it was proved that under Mangasarian-Fromovitz constraint qualification, WCR implies that an AKKT2 point is always a WSOC point. The next proposition is a simple re-statement of this fact.

Proposition 4. Assume WCR holds at \( x^* \). Let \( \{(\lambda^k, \mu^k)\} \) be a dual AKKT2 sequence associated with \( x^* \). Then WSOC holds at \( x^* \) for all limit points of \( \{(\lambda^k, \mu^k)\} \).

Proof: Let \( d \in S(x^*) \) and \( (\lambda, \mu) \) a limit point of \( \{(\lambda^k, \mu^k)\} \). Let \( x^k \to x^* \) be a primal AKKT2 sequence. From Proposition 2, \( S(x,x^*) \) is isc. Hence, there is some sequence \( \{d_k\} \to d \) with \( d_k \in S(x^k, x^*) \) and \( \|d_k\| = 1 \). Thus, \( \nabla L(x^k, \lambda^k, \mu^k) \to 0 \) and \( d^*_k \nabla^2 L(x^k, \lambda^k, \mu^k) d_k \geq -\varepsilon_k \) for some \( \varepsilon_k \to 0^+ \). Taking limit in an appropriate subsequence we have \( \nabla L(x^*, \lambda, \mu) = 0 \) and \( d^*_k \nabla^2 L(x^*, \lambda, \mu) d \geq 0 \). □

Note that, in Proposition 4 when \( \{(\lambda^k, \mu^k)\} \) has a limit point, [8] Theorem 3.3] implies that WSOC holds at all Lagrange multipliers.
The next proposition shows that, in fact, the converse of Proposition 4 also holds. That is, WCR is necessary for an AKKT2 point, with bounded dual sequence, to always guarantee primal-dual convergence to a WSOC point.

**Proposition 5.** Given a feasible point \( x^* \), assume that for every objective function \( f \) such that \( x^* \) is an AKKT2 point, \( x^* \) also satisfies WSOC at all limit points of the dual sequence. Then, WCR holds.

**Proof:** Assume WCR fails. From Proposition 2 there is some \( d \in S(x^*) \) such that for some \( \{x^k\} \rightarrow x^* \), \( d \) is not the limit of any direction \( d^k \in S(x^k, x^*) \). We may assume without loss of generality that \( d = e_1 \notin S(x^k, x^*) \) for all \( k \).

Let \( Q = diag(-1, 1, \ldots, 1) \) and let us define the objective function \( f(x) = \frac{1}{2}x^TQ(x - 2x^*) \). We have \( \nabla f(x^*) = 0 \) and \( \nabla^2 f(x) = Q \) for all \( x \). Note that the Lagrange multiplier \((\lambda, \mu) = (0, 0)\) does not satisfy WSOC. To prove that AKKT2 holds, \((\lambda^k, \mu^k) \rightarrow (0, 0)\) and note that \( \nabla^2 L(x^k, \lambda^k, \mu^k) \rightarrow Q \), with \( d^T Q d' > 0 \) for all \( d' \) orthogonal to \( d \), hence, it is enough to define tangent sequences \((\theta^k, \eta^k)\) such that

\[
\liminf_{k \to +\infty} d^T (Q + \sum_{i=1}^m \theta^k_i \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{i=1}^p \eta^k_i \nabla g_i(x^k) \nabla g_i(x^k)^T) d \geq 0.
\]

Since \( d \notin S(x^k, x^*) \) for all \( k \), and there is a finite number of constraints, there is some equality or inequality constraint \( h_{i_0}(x) = 0 \) or \( g_{i_0}(x) \leq 0 \) such that \( \nabla h_{i_0}(x^k)^T d \neq 0 \) or, respectively, \( \nabla g_{i_0}(x^k)^T d \neq 0 \) for all \( k \) in a subsequence. Defining \( \theta^k_{i_0} \) or \( \eta^k_{i_0} \) large enough, AKKT2 holds.

Proposition 5 gives our main result and points out the main deficiency of AKKT2 in the sense that it is not expected that an AKKT2 point satisfies WSOC beyond the constant rank case.

Note that when \( x^* \) is an AKKT2 point with a dual sequence \((\lambda^k, \mu^k) \rightarrow (\lambda, \mu)\), using Proposition 1 and taking the limit we get \( d^T \nabla L(x^*, \lambda, \mu) d \geq 0 \) for \( d \in \liminf_{x \to x^*} S(x, x^*) \), as it can be seen from the proof of Proposition 4. Hence, WSOC is not expected to hold when WCR fails, that is, when \( S(x, x^*) \) is not isc. Proposition 5 shows that this is indeed the case. This issue with AKKT2 is related to the fact that the associated tangent sequence \( \{(\theta^k, \eta^k)\} \) is allowed to be arbitrarily large in the definition of AKKT2. See the end of the proof of Proposition 5. That is, we can make the second-order information vanish for \( d \notin S(x^k, x^*) \) by choosing the tangent multiplier large enough, independent of the sign of \( d^T \nabla^2 L(x^k, \lambda^k, \mu^k) d \).

For this reason, since many algorithms are known to generate AKKT2 sequences, we claim that there is a theoretical limitation of such algorithms in guaranteeing WSOC when WCR fails. Therefore, to overcome such issue, tangent multipliers should be dealt explicitly within the algorithm, such as to bound their growth.
4 Discussion

Most second-order algorithm aim at finding a negative curvature direction $d^k \in \mathbb{R}^n$ such that $(d^k)^T \nabla^2 L(x^k, \lambda^k, \mu^k) d^k < 0$ with $\nabla h_i(x^k)^T d^k = 0$, $\nabla g_i(x^k)^T d^k = 0$, if $g_i(x^k) > -\varepsilon$ for some tolerance $\varepsilon > 0$, while combining a step along this direction with a step along a descent direction. See, for instance, [1, 15]. Besides first-order stationarity, a stopping criterion associated with the satisfaction of second-order stationarity is activated when such negative curvature direction does not exist. In this way, a tangent sequence $\{(\theta^k, \eta^k)\}$ is not explicitly available, but it must exist due to a contradiction argument given that a symmetric $n \times n$ matrix $P$ satisfies $d^T P d > 0$ for all $d \neq 0$ with $Q d = 0$ for some $m \times n$ matrix $Q$ if, and only if, $P + \theta Q^T Q$ is positive definite for some $\theta > 0$. See [17] [10]. Hence, the existence of the tangent sequence is guaranteed only by a contradiction argument, thus imposing no control on its rate of growth, and hence, convergence to a WSOC point is not expected beyond the constant rank case.

In the case of augmented Lagrangian and interior point methods the tangent sequence appears with a particular structure in the following sense. We present below the optimality condition introduced in [17]:

**Definition 2.** Let the feasible point $x^*$ be an AKKT2 point with associated primal $\{x^k\}$, dual $\{(\lambda^k, \mu^k)\}$ and tangent $\{(\theta^k, \eta^k)\}$ sequences. We say that Complementarity-AKKT2 (CAKKT2) holds when:

$$ \forall i = 1, \ldots, m, \lambda^k_i h_i(x^k) \to 0 \text{ and } \forall i = 1, \ldots, p, \mu^k_i g_i(x^k) \to 0, \quad (4.1) $$

$$ \forall i = 1, \ldots, m, \theta^k_i h_i(x^k)^2 \to 0 \text{ and } \forall i = 1, \ldots, p, \eta^k_i g_i(x^k)^2 \to 0. \quad (4.2) $$

Note that the complementarity condition (4.1) imply in particular that $\mu^k_i \to 0$ and $\eta^k_i \to 0$ when $g_i(x^*) < 0$, hence, by continuity, we may replace these parameters by zero for sufficiently large $k$, and we get that (2.2) is a consequence of (4.1) [14]. The stronger first-order complementarity (4.1) is imposed to limit the growth of the dual sequence in the sense that $|\lambda^k_i| = o(1/h_i(x^k))$, $i = 1, \ldots, m$ and $\mu^k_i = o(1/g_i(x^k))$, $i \in A(x^*)$, which holds in augmented Lagrangian [3], interior point [12] and regularized sequential quadratic programming [15] methods. See [17]. Since we are assuming that the dual sequence has a limit point, the stronger first-order complementarity measure (4.1) plays no role in our analysis. Condition (4.2) limits the growth of the tangent sequence in a similar way in order to avoid deficiencies of AKKT2 associated with a discontinuous perturbed critical subspace. This type of tangent sequence is generated by second-order augmented lagrangians [5] [10] and interior point [21] [22] [9] [15] methods. In this case, a slightly weaker constraint qualification can be employed to yield a WSOC point. See [17].

Note that although the bound $\theta^k_i = o(1/h_i(x^k)^2)$, $i = 1, \ldots, m$ and $\eta^k_i = o(1/g_i(x^k)^2)$, $i \in A(x^*)$ limits the growth of the tangent sequence in a way, they do not guarantee that the difficulties associated to a discontinuous perturbed critical subspace disappear, as the next example shows:
Example: Consider the problem

\[
\begin{align*}
\text{Minimize} & \quad -x_1^2 - x_2, \\
\text{subject to} & \quad x_2 \leq 0, -x_1^2 + x_2 \leq 0,
\end{align*}
\]

at \(x^* = (0, 0)\). For any Lagrange multiplier \((\mu_1, \mu_2) \in \mathbb{R}_+^2, \mu_1 + \mu_2 = 1\), take the primal sequence \(x^k = (x_1^k, 0) \to 0\), dual sequence \(\mu^k \to (\mu_1, \mu_2)\) and tangent sequence \(\eta^k = (0, 1/(x_1^k)^3)\). A simple calculation shows that \(x^*\) is a CACKET2 point, while WSOC fails.

The next proposition shows that with a particular choice of the tangent sequence, the true second-order information can be recovered without assuming WCR.

Proposition 6. Let \(x^*\) be an AKKT2 point with primal sequence \(\{x^k\}\), dual sequence \(\{\lambda^k, \mu^k\}\) and tangent sequence \(\{\theta^k, \eta^k\}\). Assume that

\[
\begin{align*}
\theta^k & \to \max\{|(\nabla h_i(x^k))^T d|^2, d \in S(x^*), \|d\| = 1\} = 0, i = 1, \ldots, m, \\
\eta^k & \to \max\{|(\nabla g_i(x^k)^T d|^2, d \in S(x^*), \|d\| = 1\} = 0, i \in A(x^*),
\end{align*}
\]

and \(\eta^k \to 0\) otherwise. Then, \(x^*\) satisfies WSOC at all limit points of \(\{(\lambda^k, \mu^k)\}\).

Proof: Assume \(x^*\) is an AKKT2 point with primal, dual and tangent sequences as defined. From the definition, taking \(d \in S(x^*)\), we have

\[
\liminf_{k \to +\infty} d^T (\nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \theta^k_i \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{i=1}^p \eta^k_i \nabla g_i(x^k) \nabla g_i(x^k)^T) d \geq 0.
\]

Taking the limit in the expression above, from the definition of the tangent sequence we have that both sums vanish, where we conclude that \(d^T \nabla^2 L(x^*, \lambda, \mu) d \geq 0\).

It is not known if around a local minimizer, the choice of tangent multipliers given in Proposition 6 can be done under reasonable assumptions. The choice given in Definition 2 can always be done and is also satisfied by sequences generated by augmented Lagrangian and interior point methods [17]. Hence, it would be interesting to investigate conditions under which the choice given in Proposition 6 can be done around a local solution, and how to build an algorithm fulfilling it. This would give an algorithm with global convergence to a WSOC point under conditions weaker than the ones in [2] [17].

To conclude this paper we mention that the continuity of the linearized cone \(L(x, x^+) = \{d \mid \nabla h_i(x)^T d = 0, i = 1, \ldots, m, \nabla g_i(x)^T d \leq 0, i \in A(x^*)\}\) is equivalent to the continuity of the KKT cone \(L(x, x^+)^o = \{w \in \mathbb{R}^n \mid w = \sum_{i=1}^m \alpha_i \nabla h_i(x) + \sum_{i \in A(x^*)} \beta_i \nabla g_i(x^*), \alpha_i, \beta_i \in \mathbb{R}_+, \text{ where this last property is the weakest condition to ensure that when } \nabla L(x^k, \lambda^k, \mu^k) \to 0 \text{ with } x^k \to x^* \text{ and } \mu^k_i = 0 \text{ for } i \notin A(x^*) \text{ and sufficiently large } k, \text{ then } x^* \text{ is a KKT point, and hence, it is associated with global convergence of many first-order algorithms (see [3])}. \) Since the critical cone \(C(x^*)\) can be written as \(C(x^*) = L(x^*, x^*) \cap \{d \mid \nabla f(x^*)^T d = 0\}\), it would be interesting to develop algorithms satisfying the
condition given by Proposition 1 where \( S(x, x^*) \) is replaced by \( L(x, x^*) \cap \{ d \mid \nabla f(x)^T d = 0 \} \), a perturbed critical cone, in order to arrive at global convergence results related to a more accurate second-order optimality condition.

Alternatively, we might consider the following perturbation of the true critical cone at a primal-dual pair \((x^*, \lambda, \mu)\):

\[
\tilde{C}(x, x^*) = \{ d \mid \nabla h_i(x)^T d = 0, i = 1, \ldots, m; \nabla g_i(x)^T d = 0, i \text{ with } \mu_i > 0; \nabla g_i(x)^T d \leq 0, i \in A(x^*), \mu_i = 0 \},
\]

and from the definition of AKKT2 with \( \{(\lambda^k, \mu^k)\} \to (\lambda, \mu) \), if the tangent sequence is such that \( \eta^k_i \to 0 \) when \( \mu^k_i \to 0 \), we would have \( d^T \nabla^2 L(x^*, \lambda, \mu) d \geq 0 \) for all \( d \in \liminf_{x \to x^*} \tilde{C}(x, x^*) \), which shows that the proper choice of the tangent sequence can yield a second-order property related to the strong second-order optimality condition.

### Acknowledgement

This work was supported by FAPESP (grants 2013/05475-7 and 2016/02092-8) and CNPq.

### 5 Concluding remarks

In this paper we investigated the role of the tangent sequence in global convergence analysis of second-order algorithms. We showed that without controlling its growth, second-order global convergence to a WSOC point can only be expected when the WCR property holds, since it is equivalent to the continuity of the perturbed critical subspace. When controlling the growth of these parameters as in CAKKT2, that is, in the way augmented Lagrangian and interior point methods control them, we show that although it does not always recover the second-order information, it can be seen as an approximation to an adequate rate of growth of the tangent sequence.

### References


10


