

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

G. Haeser*

January 30, 2017. Revised on February 2, 2018

Abstract

In second-order algorithms, a second-order Approximate-KKT (AKKT2) point is expected to be found without additional assumptions. Under a constraint qualification, 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 is not controlled. Since no algorithm directly controls their growth, 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, Constraint qualification, 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 with 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 problems that are very degenerate or ill-conditioned 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

$$\begin{aligned} & \text{Minimize} && f(x), \\ & \text{subject to} && h(x) = 0, \\ & && g(x) \leq 0, \end{aligned}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}, h : \mathbb{R}^n \rightarrow \mathbb{R}^m, g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ 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

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

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 (e.g., the constant rank constraint qualification; see [8]), the Weak Second-order necessary Optimality Condition (WSOC) holds with some Lagrange multiplier $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$:

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

where

$$S(x^*) = \{d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0, i = 1, \dots, m, \nabla g_i(x^*)^T d = 0, i \in A(x^*)\}$$

is the critical subspace and $A(x^*) = \{i \in \{1, \dots, p\} \mid g_i(x^*) = 0\}$ 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

$$\nabla L_g(x^*, \lambda_0, \lambda, \mu) = 0 \text{ with } \mu^T g(x^*) = 0,$$

where $x \mapsto L_g(x, \lambda_0, \lambda, \mu) = \lambda_0 f(x) + \lambda^T h(x) + \mu^T g(x)$ 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 $\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla g_i(x^k) \rightarrow 0$ with $\mu_i^k \rightarrow 0$ for $i \in A(x^*)$, $x^k \rightarrow x^*$ and $\|(\lambda^k, \mu^k)\| \rightarrow +\infty$, dividing by $\|(1, \lambda^k, \mu^k)\|$ 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 the Mangasarian-Fromovitz constraint qualification fails. However, when dealing explicitly with the aforementioned sequences, one 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 the 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

$$\max_{(\lambda_0, \lambda, \mu) \in \Lambda_g(x^*)} d^T \nabla^2 L_g(x^*, \lambda_0, \lambda, \mu) d \geq 0 \text{ for all } d \in C(x^*) \setminus \{0\},$$

where $C(x^*)$ is the critical cone $C(x^*) = \{d \in \mathbb{R}^n \mid \nabla f(x^*)^T d = 0, \nabla h_i(x^*)^T d = 0, i = 1, \dots, m, \nabla g_i(x^*)^T d \leq 0, i \in A(x^*)\}$. This gives a necessary optimality condition for local optimality, and also a sufficient one if we replace the inequality sign “ ≥ 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); and 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 $\{(x^k, \lambda^k, \mu^k, \theta^k, \eta^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p \times \mathbb{R}_+^m \times \mathbb{R}_+^p$. Assume that $x^k \rightarrow x^*$, $\nabla L(x^k, \lambda^k, \mu^k) \rightarrow 0$,

$$\liminf_{k \rightarrow +\infty} \lambda_{\min}(\nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \theta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^\top + \sum_{i=1}^p \eta_i^k \nabla g_i(x^k) \nabla g_i(x^k)^\top) \geq 0, \quad (2.1)$$

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

$$\forall i = 1, \dots, p, \text{ if } g_i(x^*) < 0, \text{ then } \mu_i^k = 0 \text{ and } \eta_i^k = 0 \text{ for sufficiently large } k. \quad (2.2)$$

When such sequences exist, we call x^* an AKKT2 point, and the sequences $\{x^k\}$, $\{\lambda^k, \mu^k\}$, and $\{\theta^k, \eta^k\}$ primal, dual, and tangent (respectively) 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 these 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].

The notion of an AKKT2 point is 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 $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^s$, $x \mapsto F(x) \subseteq \mathbb{R}^s$, the *outer limit* of $F(x)$ as $x \rightarrow x^*$ is given by:

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

while its *inner limit* is given by:

$$\liminf_{x \rightarrow x^*} F(x) = \{w^* \in \mathbb{R}^s : \forall \{x^k\} \rightarrow x^*, \exists \{w^k\} \rightarrow w^* \text{ with } w^k \in F(x^k)\}.$$

The multifunction $F(x)$ is said to be *outer semicontinuous* (osc) at x^* if $\limsup_{x \rightarrow x^*} F(x) \subseteq F(x^*)$ and *inner semicontinuous* (isc) at x^* if $F(x^*) \subseteq \liminf_{x \rightarrow x^*} F(x)$.

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 $\{x^k\} \rightarrow x^* \in \mathbb{R}^n$ and $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ with $\mu_i^k = 0$ 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, $\nabla L(x^k, \lambda^k, \mu^k) \rightarrow 0$ and for some $\varepsilon_k \rightarrow 0^+$,*

$$d^T \nabla^2 L(x^k, \lambda^k, \mu^k) d \geq -\varepsilon_k \|d\|^2, \quad (3.1)$$

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

$$S(x, x^*) = \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \nabla h_i(x)^T d = 0, i = 1, \dots, m, \text{ and} \\ \nabla g_i(x)^T d = 0, i \in A(x^*) \end{array} \right\}. \quad (3.2)$$

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

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

is a positive definite matrix for sufficiently large k , where \mathcal{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 $d^T (\nabla^2 L(x^k, \lambda^k, \mu^k) + (\varepsilon_k + 1/k) \mathcal{I}) d > 0$ 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

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

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

Let us consider the following example:

$$\text{Minimize } x, \text{ s.t. } \frac{x^2}{2} = 0.$$

We note that AKKT2 gives no second-order information at the global solution $x^* = 0$. Namely, there exist $\{x^k\} \rightarrow x^*$ and $\{\lambda^k\} \subset \mathbb{R}$ such that $1 + \lambda^k x^k \rightarrow 0$, $d^T \lambda^k d \geq 0$ for all $d \in S(x^k, x^*) = \{d \in \mathbb{R} \mid d^T x^k = 0\}$. 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 [14, 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 $S(x, x^*)^\circ = \{w \in \mathbb{R}^n \mid w = \sum_{i=1}^m \alpha_i \nabla h_i(x) + \sum_{i \in A(x^*)} \beta_i \nabla g_i(x), \text{ for some } \alpha_i, \beta_i \in \mathbb{R}\}$, while the inner semicontinuity of $S(x, x^*)$ is equivalent to the outer semicontinuity of its polar (see [7, Theorem 1.1.8]). \square

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 $x^k \rightarrow x^*$ and $d^k \rightarrow d$ with $\nabla h_i(x^k)^\top d_k = 0, i = 1, \dots, m$ and $\nabla g_i(x^k)^\top d_k = 0, i \in A(x^*)$, then taking the limit we get $d \in S(x^*)$. Therefore, WCR is equivalent to $S(x^*) \subseteq \liminf_{x \rightarrow x^*} S(x, x^*) \subseteq \limsup_{x \rightarrow x^*} S(x, x^*) \subseteq S(x^*)$ 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 (λ, μ) . Then, every sequence $\{x^k\} \rightarrow x^*$ and $\{(\lambda^k, \mu^k)\} \rightarrow (\lambda, \mu)$, is a primal and, respectively, dual AKKT2 sequence associated with x^* .*

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

In [2], it was proved that under the 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 (λ, μ) a limit point of $\{(\lambda^k, \mu^k)\}$. Let $x^k \rightarrow x^*$ be a primal AKKT2 sequence. From Proposition 2, $S(x, x^*)$ is isc. Hence, there is some sequence $\{d_k\} \rightarrow d$ with $d_k \in S(x^k, x^*)$ and $\|d_k\| = 1$. Thus, $\nabla L(x^k, \lambda^k, \mu^k) \rightarrow 0$ and $d_k^\top \nabla^2 L(x^k, \lambda^k, \mu^k) d_k \geq -\varepsilon_k$ for some $\varepsilon_k \rightarrow 0^+$. Taking limit in an appropriate subsequence we have $\nabla L(x^*, \lambda, \mu) = 0$ and $d^\top \nabla^2 L(x^*, \lambda, \mu) d \geq 0$. \square

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 $Q = \text{diag}(-1, 1, \dots, 1)$ and let us define the objective function $f(x) = \frac{1}{2}x^T Q(x - 2x^*)$. 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, let $(\lambda^k, \mu^k) \rightarrow (0, 0)$ 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 (θ^k, η^k) such that

$$\liminf_{k \rightarrow +\infty} d^T \left(Q + \sum_{i=1}^m \theta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{i=1}^p \eta_i^k \nabla g_i(x^k) \nabla g_i(x^k)^T \right) d \geq 0.$$

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 $\theta_{i_0}^k$ or $\eta_{i_0}^k$ large enough, AKKT2 holds. \square

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 $d \in \liminf_{x \rightarrow x^*} S(x, x^*)$, 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 with explicitly within the algorithm, such as to bound their growth.

4 Discussion

Second-order algorithmic computations are based on the available information at the current iterate x^k , where one can only hopefully compute the perturbed critical subspace $S(x^k, x^*)$. More specifically, most second-order algorithms aim at finding at each iteration a negative curvature direction $d^k \in \mathbb{R}^n$ such that $(d^k)^\top \nabla^2 L(x^k, \lambda^k, \mu^k) d^k < 0$ with $\nabla h_i(x^k)^\top d^k = 0, i = 1, \dots, m; \nabla g_i(x^k)^\top d^k = 0$, if $g_i(x^k) > -\varepsilon$ for some tolerance $\varepsilon > 0$, and combining a step along this direction with a step along a descent direction. Besides first-order stationarity, a stopping criterion associated with the satisfaction of second-order stationarity is activated when such negative curvature directions do not exist. See, for instance, [1, 15]. In this way, most second-order algorithms are expected to generate AKKT2 sequences. This fact alone is enough to prove global convergence of algorithms to a WSOC point under a weak constraint qualification, which suggests the AKKT2 property is the most relevant optimality property shared by most algorithms. Hence, in respect to second-order global convergence, the particularities of the algorithm are not relevant other than that it generates an AKKT2 sequence.

According to Proposition 1, the definition of an AKKT2 sequence can be done without the notion of a tangent sequence, and in fact, algorithms do not include a specific computation of these parameters. That is, 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 $d^\top P d > 0$ for all $d \neq 0$ with $Qd = 0$ for some $m \times n$ matrix Q if, and only if, $P + \theta Q^\top Q$ 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 Augmented Lagrangians and Interior Point Methods (see [17]), with the use of a penalty or barrier parameter, each iterate is computed by approximately minimizing a modified objective function. This allows for second-order global convergence results without explicitly computing a negative curvature direction in the perturbed critical subspace, as long as the subproblem is solved up to second-order. A careful analysis of the sequences generated by these methods shows that an explicit tangent sequence is generated with an additional bound on its growth given by

$$\theta_i^k = o(1/h_i(x^k)^2), i = 1, \dots, m \text{ and } \eta_i^k = o(1/g_i(x^k)^2), i \in A(x^*). \quad (4.1)$$

An AKKT2 point where (4.1) holds is called a Complementary-AKKT2 (CAKKT2) point, and the fact that feasible limit points of Augmented Lagrangians and Interior Point Methods are CAKKT2 points is the main result of [17]. Our Proposition 5 suggests that the control of the growth of the tangent sequence is fundamental to avoid the theoretical limitation discussed in this paper, however, the following example shows that the difficulties associated with a discontinuous perturbed critical subspace do not disappear under the control given by (4.1):

Example 1. Consider the problem

$$\begin{aligned} \text{Minimize} \quad & -x_1^2 - x_2, \\ \text{subject to} \quad & x_2 \leq 0, -x_1^2 + x_2 \leq 0, \end{aligned}$$

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 $x^k = (x_1^k, 0) \rightarrow 0$, dual sequence $\mu^k \rightarrow (\mu_1, \mu_2)$ and tangent sequence $\eta^k = (0, 1/(x_1^k)^3)$. A simple calculation shows that x^* is an AKKT2 point with the additional control of the tangent sequence given by (4.1), while WSOC fails.

Considering the definition of an AKKT2 point x^* , it is easy to see by a direct computation that even when the critical subspace is discontinuous, x^* is guaranteed to be a WSOC point when the rate of growth of the tangent sequence is controlled in the following way:

$$\begin{aligned} \eta_i^k \max\{(\nabla h_i(x^k)^T d)^2, d \in S(x^*), \|d\| = 1\} &\rightarrow 0, i = 1, \dots, m, \\ \eta_i^k \max\{(\nabla g_i(x^k)^T d)^2, d \in S(x^*), \|d\| = 1\} &\rightarrow 0, i \in A(x^*), \end{aligned}$$

and $\eta_i^k \rightarrow 0, i \notin A(x^*)$. However, this does not give a computable condition, as it depends on the knowledge of the true critical subspace $S(x^*)$ at iteration k , and not its perturbed counterpart $S(x^k, x^*)$.

These results stress the importance of controlling the rate of growth of the tangent sequence in order to obtain second-order global convergence results beyond the constant rank case, but they do not suggest an algorithmic way of doing it beyond the Augmented Lagrangian and Interior Point Methods.

More importantly, to conclude this paper, we mention a related problem of defining an algorithm with a global convergence result to a second-order condition stronger than WSOC. Note that the continuity of the linearized cone $L(x, x^*) = \{d \mid \nabla h_i(x)^T d = 0, i = 1, \dots, m, \nabla g_i(x)^T d \leq 0, i \in A(x^*)\}$ is equivalent to the continuity of the KKT cone $L(x, x^*)^\circ = \{w \in \mathbb{R}^n \mid w = \sum_{i=1}^m \alpha_i \nabla h_i(x) + \sum_{i \in A(x^*)} \beta_i \nabla g_i(x^*), \alpha_i \in \mathbb{R}, \beta_i \in \mathbb{R}_+\}$, where this last property is the weakest condition to ensure that when $\nabla L(x^k, \lambda^k, \mu^k) \rightarrow 0$ with $x^k \rightarrow x^*$ and $\mu_i^k = 0$ for $i \notin A(x^*)$ and sufficiently large k , then x^* 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^*, λ, μ) :

$$\begin{aligned} \tilde{C}(x, x^*) = \{d \mid \nabla h_i(x)^T d = 0, i = 1, \dots, m; \nabla g_i(x)^T d = 0, i \text{ with } \mu_i > 0; \\ \nabla g_i(x)^T d \leq 0, i \in A(x^*), \mu_i = 0\}, \end{aligned}$$

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

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 in the way augmented Lagrangian and interior point methods control them, we show that although the second-order information is not always recovered, this control can be seen as an approximation to an adequate rate of growth of the tangent sequence.

Acknowledgement

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

References

- [1] Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications* **45**, 209–236 (2010)
- [2] Andreani, R., Haeser, G., Ramos, A., Silva, P.: A second-order sequential optimality condition associated to the convergence of algorithms. *IMA Journal of Numerical Analysis* **37**(4), 1902–1929 (2017)
- [3] Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization* **26**(1), 96–110 (2016)
- [4] Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research* (2018). DOI: 10.1287/moor.2017.0879
- [5] Andreani, R., Martínez, J.M., Schuverdt, M.L.: On second-order optimality conditions for nonlinear programming. *Optimization* **56**, 529–542 (2007)

- [6] Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization* **20**(6), 3533–3554 (2010)
- [7] Aubin, J.P., Frankowska, H.: *Set-Valued Analysis*. Birkhauser (1990)
- [8] Behling, R., Haeser, G., Ramos, A., Viana, D.S.: On a conjecture in second-order optimality conditions. *Journal of Optimization Theory and Applications* (2018). DOI: 10.1007/s10957-018-1229-1 (see also the extended technical report at arXiv:1706.07833)
- [9] Bian, W., Chen, X., Ye, Y.: Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization. *Mathematical Programming* **149**(1), 301–327 (2015)
- [10] Birgin, E.G., Haeser, G., Ramos, A.: Augmented lagrangians with constrained subproblems and convergence to second-order stationary points. *Computational Optimization and Applications* **69**(1), 51–75 (2018)
- [11] Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*. Springer-Verlag (2000)
- [12] Chen, L., Goldfarb, D.: Interior-point ℓ_2 -penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming* **108**(1), 1–36 (2006)
- [13] Dennis, J.E., Vicente, L.N.: On the convergence theory of trust-region-based algorithms for equality-constrained optimization. *SIAM Journal of Optimization* **7**(4) (1997)
- [14] Facchinei, F., Pang, J.S.: *Finite-dimensional variational inequalities and complementarity problems, volume I*. Springer (2003)
- [15] Gill, P.E., Kungurtsev, V., Robinson, D.P.: A stabilized SQP method: global convergence. *IMA Journal of Numerical Analysis* **37**(1), 407–443 (2016)
- [16] Gould, N.I.M., Conn, A.R., Toint, P.L.: A note on the convergence of barrier algorithms for second-order necessary points. *Mathematical programming* **85**, 433–438 (1998)
- [17] Haeser, G.: A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. *Optimization-online* (2016)
- [18] Haeser, G., Liu, H., Ye, Y.: Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary. *Optimization Online* (2017)

- [19] Liu, H., Yao, T., Li, R., Ye, Y.: Folded concave penalized sparse linear regression: sparsity, statistical performance, and algorithmic theory for local solutions. *Mathematical Programming* **166**(1-2), 207–240 (2017)
- [20] Ramos, A.: Tópicos em condições de otimalidade para otimização não linear. Ph.D. thesis, University of São Paulo, Department of Applied Mathematics (2016)
- [21] Ye, Y.: On affine scaling algorithms for nonconvex quadratic programming. *Mathematical Programming* **56**(1), 285–300 (1992)
- [22] Ye, Y.: On the complexity of approximating a KKT point of quadratic programming. *Mathematical Programming* **80**(2), 195–211 (1998)