A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms

G. Haeser*

October 28, 2016. Last revised on November 25, 2017.

Abstract

We develop a new notion of second-order complementarity with respect to the tangent subspace related to second-order necessary optimality conditions by the introduction of so-called tangent multipliers. We prove that around a local minimizer, a second-order stationarity residual can be driven to zero while controlling the growth of Lagrange multipliers and tangent multipliers, which gives a new second-order optimality condition without constraint qualifications stronger than previous ones associated with global convergence of algorithms. We prove that second-order variants of augmented Lagrangian and interior point methods generate sequences satisfying our optimality condition. We present also a companion minimal constraint qualification, weaker than the ones known for second-order methods, that ensures usual global convergence results to a classical second-order stationary point. Finally, our optimality condition naturally suggests definition of second-order stationarity suitable for the computation of iteration complexity bounds and for the definition of stopping criteria.

Keywords: second-order optimality conditions, complementarity, global convergence, constraint qualifications.

1 Introduction

This paper deals with the general smooth nonlinear optimization problem with constraints, where first- and second-order information of objective functions and constraints are available. With the increase of computational power, a renewed interest in second-order algorithms has appeared. It has been shown in [15] that the complexity of finding a first-order stationary point is reduced when allowing more derivatives to be used. In this case, one can expect to converge to a higher order stationary point. Recently, an efficient method for finding a third-order stationary point in unconstrained problems has appeared in [1]. Also, complexity analysis for finding higher-order stationary points in convexly-constrained nonlinear optimization is presented in [19].

When using second-order information, one could expect finding a second-order stationary point. This gives a more robust method than a first-order one since it has more chances of finding a true minimizer. Even when a true minimizer is not found, a second-order stationary point (even the weak notion considered here) can be of interest in practical situations, in particular, in machine and statistical learning, where meaningful statistical properties of a true solution can be recovered, see [34].

Different second-order optimality conditions can be found in the literature with distinct characteristics. We deal with a perturbation of the so-called Weak Second-order Optimality Condition (WSOC), since this is the optimality condition guaranteed by second-order algorithms, which requires positive semidefiniteness of the Hessian of the Lagrangian over a perturbed critical subspace. One difficulty that arises is that the critical subspace may have a discontinuous behavior at a solution (see [30]), which can trick second-order optimality conditions based on perturbations, which are suitable for global convergence analysis. We remove the dependency of the perturbed critical subspace by introducing so-called tangent multipliers. This concept is

*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.

based on the well known fact that positive definiteness of a matrix P over the kernel $\{d|Ad=0\}$ of a matrix A is equivalent to the positive definiteness of $P+\theta A^{\mathsf{T}}A$ for some $\theta>0$. In the semidefinite case, the equivalence is achieved with the limit inferior of the least eigenvalue of $P+\theta A^{\mathsf{T}}A$ being non-negative, as θ goes to infinity. However, if we allow θ to go to infinity too fast, the second-order information may be lost. Our optimality condition controls the growth of the tangent multipliers in order for the second-order information to remain present, whereas previous results allow for arbitrarilly fast growth ([5]). Also, our condition is satisfied at any local minimizer without constraint qualifications.

We show that the sequence required to check our optimality condition is in fact generated by augmented Lagrangian and interior point methods. By developing a minimal companion constraint qualification, we present a global convergence result of these methods to a WSOC point under a weaker assumption.

To motivate our optimality condition, let us consider the one-dimensional problem of minimizing x such that $x^2/2=0$. The Karush-Kuhn-Tucker (KKT) conditions for this problem says that $1+\lambda x=0$ should hold for the solution $x^*:=0$ with some Lagrange multiplier $\lambda\in\mathbb{R}$, which is impossible. Nonetheless, one may still rely on the KKT conditions to stop the execution of a primal-dual algorithm. For instance, the sequence $(x_k,\lambda_k):=(-\frac{1}{k},k)$ is such that the KKT residual $|1+\lambda_k x_k|$ is zero, and an algorithm may safely stop once x_k is feasible within a tolerance. In this case, it is easy to see that in order for the KKT residual to be driven to zero, the sequences $\{|\lambda_k|\}$ and $\{1/|x_k|\}$ must go to infinity with the same speed, however, in general, the KKT residual alone is not enough to conclude how to choose these sequences. In the Introduction of [10], an example is shown where the KKT residual is driven to zero far from the solution. It then proceeds to consider a redundant (for equality constraints) first-order complementarity condition, which would read as $\lambda x^2/2=0$ in our example, which suggests that one must build the sequences in such a way that $\lambda_k x_k^2/2\to 0$. In general, this introduces additional information on the problem. In our example, this information would read as $\{|\lambda_k|\}$ should go to infinity slower than $\{1/x_k^2\}$.

If we now consider the second-order information, the classical second-order optimality condition at the solution x^* would state that for an (inexisting) Lagrange multiplier λ, the Hessian matrix λ should be positive semidefinite over the subspace $S_x:=\{d\in\mathbb{R}\mid d^{\mathsf{T}}x=0\}$. At $x:=x^*$, this states that $\lambda\geq 0$, since $S_{x^*}=\mathbb{R}$, but at $x\neq x^*$, the second order information vanishes, given that $S_x=\{0\}$. We will show that, in this case, the second-order information can be recast by the introduction of a so-called tangent multiplier $\theta\geq 0$ such that $\lambda+\theta x^2\geq 0$. In particular, the sequential second-order optimality condition from [5] states the existence of a primal-dual sequence $\{(x_k,\lambda_k)\}$ such that the KKT residual converges to zero and $\liminf(\lambda^k+\theta_k x_k^2)\geq 0$ with some $\theta_k\geq 0$. However, without a control on how fast the sequence $\{\theta_k\}$ is allowed to grow to infinity, this second-order information is meaningless when $x_k\neq 0$, namely, it is sufficient to take $\{\theta_k x_k^2\}$ going to infinity as fast as $\{|\lambda_k|\}$ in order to satisfy it, independently of the primal-dual sequence $\{(x_k,\lambda_k)\}$. For the second-order information to be meaningful, we introduce a second-order complementarity with respect to the tangent multipliers that in this case says that $\theta_k x_k^4/4\to 0$, which limits the growth of the tangent multiplier $\{\theta_k\}$ to be slower than $\{1/x_k^4\}$. In other words, $\{\theta_k x_k^2\}$ must be slower than $\{1/x_k^2\}$. Since first-order complementarity gives that $\{|\lambda_k|\}$ is also slower than $\{1/x_k^2\}$, we have more chances of avoiding the critical case where the second-order information is meaningless.

This paper is organized as follows: In Section 2 we present our new optimality condition, while also generalizing previous results by considering strictly feasible sequences. In Section 3 we prove that an augmented Lagrangian and an interior point method generate sequences that fulfill our optimality condition. In Section 4 we measure the strength of our optimality condition by introducing a companion minimal constraint qualification that makes our optimality condition equivalent to the classical one. This gives global convergence results under a weaker assumption. Finally, Section 5 presents some conclusions.

Notation: Given a finite set $\mathcal{E}$ with cardinality $|\mathcal{E}|$, we denote $\theta_i, i\in\mathcal{E}$ the coordinates of a vector $\theta\in\mathbb{R}^{|\mathcal{E}|}$. Given functions $h_i:\mathbb{R}^n\to\mathbb{R}, i\in\mathcal{E}$, we denote $\nabla h_i(x)$ and $\nabla^2 h_i(x)$, for $x\in\mathbb{R}^n$, the gradient and Hessian matrix of $h_i(x)$, respectively, for $i\in\mathcal{E}$. Given a subset $E\subseteq\mathcal{E}$, we denote by h_E the function $h_E:\mathbb{R}^n\to\mathbb{R}^{|E|}$ with coordinate functions $h_i, i\in E$. By $\nabla h_E(x)$ we denote the $n\times|E|$ gradient matrix with columns $\nabla h_i(x), i\in E$. When $E:=\mathcal{E}$, we denote $h:=h_E$ and $\nabla h:=\nabla h_E$. We denote by $\mathrm{Sym}(n)$ the set of real $n\times n$ matrices and $\lambda_{min}(P)$ the smallest eigenvalue of $P\in\mathrm{Sym}(n)$. We also use the notation $P\succeq Q$ to denote that the symmetric matrix $P-Q$ is positive semidefinite. Finally, we denote I the $n\times n$ identity matrix and $\|\cdot\|$ the euclidean norm.

2 Optimality condition

Let us consider the smooth nonlinear optimization problem:

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{subject to} & h_i(x) = 0, i \in \mathcal{E}, \\ & g_i(x) \leq 0, i \in \mathcal{I}, \\ & c_i(x) \geq 0, i \in \mathcal{S}, \end{array} \tag{1}$$

where $\mathcal{E}, \mathcal{I}$ and $\mathcal{S}$ are distinct finite sets and all functions are twice continuously differentiable from $\mathbb{R}^n$ to $\mathbb{R}$. The reason for having two sets of inequalities is that we will assume the existence of a sufficient interior for the constraints $c_i(x) \geq 0, i \in \mathcal{S}$, which is relevant only for implications on the global convergence of interior point methods.

We denote by Ω the feasible set. For any fixed $x^* \in \Omega$, we denote by $\mathcal{I}_{x^*}$ the subset of indexes $i \in \mathcal{I}$ such that $g_i(x^*) = 0$ and by $\mathcal{S}_{x^*}$ the subset of indexes $i \in \mathcal{S}$ such that $c_i(x^*) = 0$. The index set of active constraints at $x^* \in \Omega$ is $\mathcal{E} \cup \mathcal{I}_{x^*} \cup \mathcal{S}_{x^*}$.

For fixed $(\lambda, \mu, s) \in \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{I}|}_+ \times \mathbb{R}^{|\mathcal{S}|}_+$, we denote the Lagrangian function by $x \mapsto L(x, \lambda, \mu, s) := f(x) + \langle h(x), \lambda \rangle + \langle g(x), \mu \rangle - \langle c(x), s \rangle$, with $\nabla L(x, \lambda, \mu, s)$ and $\nabla^2 L(x, \lambda, \mu, s)$ being its gradient and Hessian, respectively. When $\mathcal{E}$, $\mathcal{I}$ or $\mathcal{S}$ is $\emptyset$, we omit the corresponding argument of L and its derivatives.

For the moment, let us consider for the ease of exposition that $\mathcal{S} = \emptyset$. Under any constraint qualification a local minimizer x^* satisfies the Karush-Kuhn-Tucker (KKT) conditions, that is, there exist Lagrange multipliers $(\lambda^*, \mu^*) \in \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{I}|}_+$ such that

$$h(x^*) = 0, g(x^*) \leq 0, \tag{2}$$

$$\nabla L(x^*, \lambda^*, \mu^*) = 0, \tag{3}$$

$$\mu_i^* g_i(x^*) = 0 \text{ for all } i \in \mathcal{I}. \tag{4}$$

In a first-order algorithm, in order to try to find a local solution, one may aim at driving the KKT residual defined below to zero, by controlling the primal-dual variables $(x, \lambda, \mu), \mu \geq 0$:

$$R(x, \lambda, \mu) := \max\{\|h(x)\| + \|\max\{0, g(x)\}\|, \|\nabla L(x, \lambda, \mu)\|, \sum_{i \in \mathcal{I}} |\mu_i g_i(x)|\}.$$

In [10], it is suggested considering the redundant condition $\lambda_i^* h_i(x^*) = 0, i \in \mathcal{E}$ in the definition of a KKT point and hence, also driving the quantity $\sum_{i \in \mathcal{E}} |\lambda_i h_i(x)|$ to zero together with $R(x, \lambda, \mu)$. This can be understood in the following way: Since one does not know, a priori, if a solution of the problem under consideration is a KKT point, or, if the problem is regular enough in order for Lagrange multipliers to be bounded at the considered solution, driving the KKT residual to zero may drive the primal variable x to a local solution x^*, but some dual variable $\lambda_i, i \in \mathcal{E}$ or $\mu_i, i \in \mathcal{I}$ may be driven to $\pm\infty$. Ensuring also that $\lambda_i h_i(x) \to 0$ introduces a bound on the growth of the Lagrange multiplier approximation, in the sense that $|\lambda_i|$ should grow slower than $\frac{1}{|h_i(x)|}$. The same reasoning apply in favor of measuring inequality constraints complementarity (4) by driving $\sum_{i \in \mathcal{I}} |\mu_i g_i(x)|$ to zero, rather than by driving $\sum_{i \in \mathcal{I}} |\min\{\mu_i, -g_i(x)\}|$ to zero (this is associated with the optimality condition AKKT, see [4] and [10]). Although this previous form accurately detects that Lagrange multipliers associated with an inactive constraint $i \in \mathcal{I} \backslash \mathcal{I}_{x^*}$ must be driven to zero, it does not impose any limitation on the growth of the Lagrange multiplier approximation associated with an active constraint $i \in \mathcal{I}_{x^*}$. The interesting property is that this controlled reduction of the KKT residual by means of driving $R(x, \lambda, \mu)$ or $C(x, \lambda, \mu) := \max\{R(x, \lambda, \mu), \sum_{i \in \mathcal{E}} |\lambda_i h_i(x)|\}$ to zero can always be done around a local solution x^*, even if it has an unbounded set of Lagrange multipliers, or if x^* is not a KKT point at all. In particular, many algorithms (see [6]) are such that the primal-dual sequence $\{(x^k, \lambda^k, \mu^k)\}$ generated is such that $\max\{\|\nabla L(x^k, \lambda^k, \mu^k)\|, \sum_{i \in \mathcal{I}} |\min\{\mu_i^k, -g_i(x^k)\}|\} \to 0^+$ for any subsequence of $\{x^k\}$ converging to a feasible point. In [10], it is proved that an Augmented Lagrangian, under a mild additional smoothness assumption, is such that $C(x^k, \lambda^k, \mu^k) \to 0^+$ in a subsequence where x^k converges to a feasible point.

The main purpose of this paper is to propose a generalization of these ideas to the second-order case. Under some more specific constraint qualifications (see [11] for a discussion), it can be proved that a local solution x^*

is such that there exist Lagrange multipliers $(\lambda^*, \mu^*) \in \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|}$ such that (2)-(4) hold and

$$d^{\mathtt{T}}\nabla^2 L(x^*, \lambda^*, \mu^*)d \geq 0, \text{ for all } d \in S(x^*), \tag{5}$$

where $S(x^*) := \{d \in \mathbb{R}^n \mid \nabla h(x^*)^{\mathtt{T}} d = 0, \nabla g_{\mathcal{I}_{x^*}}(x^*)^{\mathtt{T}} d = 0\}$ is the so-called critical subspace (or weak critical cone). A point x^* is said to satisfy the Weak Second-order Optimality Condition (WSOC) if (2)-(5) hold for some Lagrange multiplier.

Under a different set of constraint qualifications, condition (5) can be strengthened to a larger set $C(x^*)$ (the critical cone) containing the subspace $S(x^*)$. This is relevant given that condition (5) replacing $S(x^*)$ by $C(x^*)\backslash\{0\}$ and the inequality ≥ 0 by the strict inequality > 0 is sufficient for x^* to be a strict local minimizer. Even stronger sufficient and necessary results (under MFCQ alone) can be derived by considering the possibility of verifying the inequality in (5) with different Lagrange multipliers for different directions $d \in C(x^*)$, or also, without constraint qualifications at all by considering Fritz-John multipliers. See, for instance, [18] for details. We focus our attention on the weaker second-order notion (5) given that this is the optimality condition associated with global convergence of second-order algorithms. In fact, we are not aware of any algorithm with a global convergence theory based on a second-order optimality condition defined in terms of the critical cone $C(x^*)$. The discussion in [28] suggests that probably such algorithms do not exist. Also, the stopping criterion of a second-order algorithm is, most of the times, based on checking the approximate fulfillment of (5) (see [5, 16]), which can be done in the case of a subspace, but turns out to be NP-hard for a pointed cone. See [11] and the references therein for a more detailed discussion on different second-order optimality conditions and their consequences to global convergence analysis, which is our main goal. We note also that a point that merely satisfies WSOC can be the goal in some important machine and statistical learning problems ([34]).

Let us start by motivating our results with the following characterization of WSOC:

Proposition 2.1. *Let $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|}$. Define $Q(\theta) := \nabla^2 L(x^*, \lambda^*, \mu^*) + \theta(\nabla h(x^*)\nabla h(x^*)^{\mathtt{T}} + \nabla g_{\mathcal{I}_{x^*}}(x^*)\nabla g_{\mathcal{I}_{x^*}}(x^*)^{\mathtt{T}})$.*

a) Condition (5) holds if, and only if $\liminf_{\theta\to+\infty} \lambda_{min}(Q(\theta)) \geq 0$.

b) There is a finite $\theta > 0$ such that $\lambda_{min}(Q(\theta)) > 0$ if, and only if, condition (5) holds and there is no $0 \neq d \in S(x^*)$ such that $d^{\mathtt{T}}\nabla^2 L(x^*, \lambda^*, \mu^*)d = 0$.

Proof: Let us prove a). Assume that $\liminf_{\theta\to+\infty} \lambda_{min}(Q(\theta)) \geq 0$ and let $d \in S(x^*)$. Then, there exists a sequence $\{\theta_k\} \to +\infty$ such that $\lambda_{min}(Q(\theta_k)) > -\frac{1}{k}$. For all k, $d = d_1^k + \cdots + d_n^k$ where $\{d_i^k\}_{i=1}^n$ are orthogonal eigenvectors of $Q(\theta_k)$ or zero. Hence, for all k,

$$d^{\mathtt{T}}\nabla^2 L(x^*, \lambda^*, \mu^*)d = d^{\mathtt{T}}Q(\theta_k)d \geq \lambda_{min}(Q(\theta_k)) \sum_{i=1}^{n} (d_i^k)^{\mathtt{T}} d_i^k \geq -\frac{1}{k}\|d\|^2.$$

Taking the limit $k \to +\infty$ gives $d^{\mathtt{T}}\nabla^2 L(x^*, \lambda^*, \mu^*)d \geq 0$.

Conversely, let (5) hold and assume $\liminf_{\theta\to+\infty} \lambda_{min}(Q(\theta)) < 0$. Then, there is some $\varepsilon > 0$ and a sequence $\{\theta_k\} \to +\infty$ such that $\lambda_{min}(Q(\theta_k)) < -\varepsilon$. Let $d_k, \|d_k\| = 1$, be an eigenvector of $Q(\theta_k)$ associated with the smallest eigenvalue. Hence, $d_k^{\mathtt{T}} Q(\theta_k) d_k = \lambda_{min}(Q(\theta_k)) < -\varepsilon$, that is

$$d_k^{\mathtt{T}}\nabla^2 L(x^*, \lambda^*, \mu^*)d_k^{\mathtt{T}} + \theta_k(\|\nabla h(x^*)^{\mathtt{T}} d_k\|^2 + \|\nabla g_{\mathcal{I}_{x^*}}(x^*)^{\mathtt{T}} d_k\|^2) < -\varepsilon. \tag{6}$$

Let us consider a subsequence such that $d_k \to d, \|d\| = 1$. From (6), taking $k \to +\infty$ and thus $\theta_k \to +\infty$, we must have $d \in S(x^*)$, which from the assumption implies $d^{\mathtt{T}}\nabla^2 L(x^*, \lambda^*, \mu^*)d \geq 0$. This gives a contradiction with (6). The proof of *b)* is similar and well known. □

Note that the additional condition imposed in *b)* is implied by the so-called noncriticality ([32]). Loosely speaking, we can say that condition (5) can be equivalently stated as the matrix

$$Q(\theta, \eta) := \nabla^2 L(x^*, \lambda^*, \mu^*) + \sum_{i\in\mathcal{E}} \theta_i \nabla h_i(x^*)\nabla h_i(x^*)^{\mathtt{T}} + \sum_{i\in\mathcal{I}} \eta_i \nabla g_i(x^*)\nabla g_i(x^*)^{\mathtt{T}}$$

being positive semidefinite possibly with "$\theta_i = +\infty$" for $i \in \mathcal{E}$, "$\eta_i = +\infty$" for $i \in \mathcal{I}_{x^*}$ and $\eta_i = 0$ for $i \in \mathcal{I} \backslash \mathcal{I}_{x^*}$. We call (θ, η) the *tangent multipliers* and we will develop a second-order complementarity measure for controlling the growth of these parameters in a similar way as CAKKT controls Lagrange multipliers.

Let us now consider problem (1) including the additional set of inequality constraint $c_i(x) \geq 0, i \in \mathcal{S}$ and let us specify our additional assumption with respect to these constraints.

Assumption 2.1. *(Sufficient interior property) We will assume that for all local solutions $x^* \in \Omega$, there exists a feasible sequence $\{z^\ell\} \subset \Omega$ with $z^\ell \to x^*$ and $c(z^\ell) > 0$, $\forall \ell$.*

Assumption 2.1 is a typical necessary assumption associated with the application of a feasible interior point method. Let us now present our general second-order optimality condition for problem (1).

Theorem 2.1. *Under Assumption 2.1, let $x^* \in \Omega$ be a local solution of (1). Then, there exist a sequence of approximate (primal) solutions $\{x^k\} \subset \mathbb{R}^n$ with $x^k \to x^*$, a sequence of approximate Lagrange multipliers $\{(\lambda^k, \mu^k, s^k)\} \subset \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}_+^{|\mathcal{S}|}$ and a sequence of approximate tangent multipliers $\{(\theta^k, \eta^k, \nu^k)\} \subset \mathbb{R}_+^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}_+^{|\mathcal{S}|}$ such that:*

i) $\nabla L(x^k, \lambda^k, \mu^k, s^k) \to 0$ *and* $\|h(x^k)\| + \|\max\{0, g(x^k)\}\| \to 0^+$,

ii) $\liminf_{k \to +\infty} \lambda_{min}(Q_k) \geq 0$, *where* $Q_k := \nabla^2 L(x^k, \lambda^k, \mu^k, s^k) + \sum_{i \in \mathcal{E}} \theta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^{\mathsf{T}} + \sum_{i \in \mathcal{I}} \eta_i^k \nabla g_i(x^k) \nabla g_i(x^k)^{\mathsf{T}} + \sum_{i \in \mathcal{S}} \nu_i^k \nabla c_i(x^k) \nabla c_i(x^k)^{\mathsf{T}}$,

iii) $h_i(x^k)\lambda_i^k \to 0, i \in \mathcal{E}; g_i(x^k)\mu_i^k \to 0, i \in \mathcal{I}$ *and* $c_i(x^k)s_i^k \to 0, i \in \mathcal{S}$,

iv) $h_i(x^k)^2\theta_i^k \to 0, i \in \mathcal{E}; g_i(x^k)^2\eta_i^k \to 0, i \in \mathcal{I}$ *and* $c_i(x^k)^2\nu_i^k \to 0, i \in \mathcal{S}$,

v) *for all k, $c(x^k) > 0$.*

Proof: Let x^* be a local solution of (1) and $\delta > 0$ small enough. Consider the regularized problem

$$\begin{array}{ll} \text{Minimize} & f(x) + \frac{1}{4}\|x - x^*\|^4, \\ \text{subject to} & h_i(x) = 0, i \in \mathcal{E}, \\ & g_i(x) \leq 0, i \in \mathcal{I}, \\ & c_i(x) \geq 0, i \in \mathcal{S}, \\ & \|x - x^*\|^2 \leq \delta. \end{array} \tag{7}$$

Clearly, x^* is the unique global minimizer of (7). Let $\{\rho_k\} \to +\infty$ and $\{r_k\} \to 0^+$ be arbitrary sequences. For all k, consider the problem

$$\begin{array}{ll} \text{Minimize} & \varphi_k(x), \\ \text{subject to} & c_i(x) > 0, i \in \mathcal{S}, \\ & \|x - x^*\|^2 \leq \delta, \end{array} \tag{8}$$

where $\varphi_k(x) := f(x) + \frac{1}{4}\|x - x^*\|^4 + \frac{\rho_k}{2}(\|h(x)\|^2 + \|\max\{0, g(x)\}\|^2) - r_k \sum_{i \in \mathcal{S}} \log(c_i(x))$. It is well known that under Assumption 2.1, a global solution x^k of (8) is well defined for all k and $x^k \to x^*$ ([4] and [26]). Let z^ℓ be the sequence given by Assumption 2.1 around x^*. Hence, for all k and ℓ large enough, we have $\varphi_k(x^k) \leq \varphi_k(z^\ell) = f(z^\ell) + \frac{1}{4}\|z^\ell - x^*\|^4 - r_k \sum_{i \in \mathcal{S}} \log(c_i(z^\ell))$, hence,

$$\frac{\rho_k}{2}(\|h(x^k)\|^2 + \|\max\{0, g(x^k)\}\|^2) - r_k \sum_{i \in \mathcal{S}} \log(c_i(x^k)) \leq$$
$$f(z^\ell) - f(x^k) + \frac{1}{4}\|z^\ell - x^*\|^4 - \frac{1}{4}\|x^k - x^*\|^4 - r_k \sum_{i \in \mathcal{S}} \log(c_i(z^\ell)).$$

Since ℓ is independent from k, for each k large enough, let us fix $\ell := \ell_k$ with $\ell_k \to +\infty$ as $k \to +\infty$ such that $-\sum_{i \in \mathcal{S}} \log(c_i(z^{\ell_k})) \leq \frac{1}{\sqrt{r_k}}$. This can be done by possibly taking non-distinct ℓ_k until the right hand side is large enough. Taking the limit $k \to +\infty$ we have

$$\begin{array}{l} \rho_k h_i(x^k)^2 \to 0^+ \text{ for all } i \in \mathcal{I}, \rho_k \max\{0, g_i(x^k)\}^2 \to 0^+ \text{ for all } i \in \mathcal{I}, \\ r_k \log(c_i(x^k)) \to 0 \text{ for all } i \in \mathcal{S}. \end{array} \tag{9}$$

Now, for k large enough, x^k is an unconstrained local minimizer of φ_k, hence, $\nabla\varphi_k(x^k)=0$, that is,

$$\begin{aligned}\nabla f(x^k)+(x^k-x^*)\|x^k-x^*\|^2+\textstyle\sum_{i\in\mathcal{E}}\rho_k h_i(x^k)\nabla h_i(x^k)+\\ \textstyle\sum_{i\in\mathcal{I}}\rho_k\max\{0,g_i(x^k)\}\nabla g_i(x^k)-\sum_{i\in\mathcal{S}}\frac{r_k}{c_i(x^k)}\nabla c_i(x^k)=0.\end{aligned}\tag{10}$$

Also, by [9, Lemma 3.2], the following matrix is positive semidefinite:

$$\begin{aligned}\nabla^2 f(x^k)+\sum_{i\in\mathcal{E}}\rho_k h_i(x^k)\nabla^2 h_i(x^k)+\sum_{i\in\mathcal{I}}\rho_k\max\{0,g_i(x^k)\}\nabla^2 g_i(x^k)\\ -\sum_{i\in\mathcal{S}}\frac{r_k}{c_i(x^k)}\nabla^2 c_i(x^k)+\sum_{i\in\mathcal{E}}\rho_k\nabla h_i(x^k)\nabla h_i(x^k)^{\mathtt{T}}\\ +\sum_{i:g_i(x^k)\geq 0}\rho_k\nabla g_i(x^k)\nabla g_i(x^k)^{\mathtt{T}}+\sum_{i\in\mathcal{S}}\frac{r_k}{c_i(x^k)^2}\nabla c_i(x^k)\nabla c_i(x^k)^{\mathtt{T}}\\ +2(x^k-x^*)(x^k-x^*)^{\mathtt{T}}+\|x^k-x^*\|^2 I.\end{aligned}\tag{11}$$

Let us define for all k, $\lambda_i^k:=\rho_k h_i(x^k)$ and $\theta_i^k:=\rho_k$ for all $i\in\mathcal{E}$, $\mu_i^k:=\rho_k g_i(x^k)$ and $\eta_i^k:=\rho_k$ for all $i\in\mathcal{I}$ such that $g_i(x^k)\geq 0$ and $\mu_i^k:=0,\eta_i^k:=0$ if $g_i(x^k)<0$. Finally, define $s_i^k:=\frac{r_k}{c_i(x^k)}$ and $\nu_i^k:=\frac{r_k}{c_i(x^k)^2}$ for all $i\in\mathcal{S}$. Thus, (9), (10) and the positive semidefiniteness of (11) together with $x^k\to x^*\in\Omega, c(x^k)>0$ gives *i)*-*v)*. □

Remark 1. Note that we have in fact proved that $\log(c_i(x^k))c_i(x^k)s_i^k\to 0$ and $\log(c_i(x^k))c_i(x^k)^2\nu_i^k\to 0, i\in\mathcal{S}$ but we have omitted the log term in the definition of the optimality condition since this does not seem to hold at sequences generated by the algorithms we analyse.

Condition *iii)* implies that if $g_i(x^*)<0$, then the corresponding Lagrange multiplier approximation $\mu_i^k\to 0$, and hence, by the continuity of the gradient $\nabla g_i(x^*)$ and the Hessian $\nabla^2 g_i(x^*)$, one can replace μ_i^k by zero, for k large enough, without altering the other properties. Similarly with s^k. Hence, *iii)* implies

iii)' for all $i\in\mathcal{I},\mu_i^k=0$ if $g_i(x^*)<0$ and for all $i\in\mathcal{S},s_i^k=0$ if $c_i(x^*)>0$.

Similarly, due to the continuity of the gradients at x^*, *iv)* implies

iv)' for all $i\in\mathcal{I},\eta_i^k=0$ if $g_i(x^*)<0$ and for all $i\in\mathcal{S},\nu_i^k=0$ if $c_i(x^*)>0$.

Definition 2.1. Let $x^*\in\Omega$ and a sequence $\{(x^k,\lambda^k,\mu^k,s^k,\theta^k,\eta^k,\nu^k)\}\subset\mathbb{R}^n\times\mathbb{R}^{|\mathcal{E}|}\times\mathbb{R}_+^{|\mathcal{I}|}\times\mathbb{R}_+^{|\mathcal{S}|}\times\mathbb{R}_+^{|\mathcal{E}|}\times\mathbb{R}_+^{|\mathcal{I}|}\times\mathbb{R}_+^{|\mathcal{S}|}$ with $x^k\to x^*$.

- If *i), iii)', v)* hold, then x^* is called an Approximate-KKT (AKKT) point ([4]).
- If *i), ii), iii)', iv)', v)* hold, then x^* is called a Second-Order Approximate-KKT (AKKT2) point ([5]).
- If *i), iii), v)* hold, then x^* is called a Complementarity-Approximate-KKT (CAKKT) point ([10]).
- If *i), ii), iii), iv), v)* hold, then x^* is called a Second-Order Complementarity-Approximate-KKT (CAKKT2) point.

We will also name the sequence according to the property of x^*. For instance, if x^* satisfies CAKKT2, then the sequence will be called a CAKKT2 sequence. Also, $\{x^k\}$ will be referred as the primal sequence, $\{(\lambda^k,\mu^k,s^k)\}$ as the dual sequence/Lagrange multipliers and $\{(\theta^k,\eta^k,\nu^k)\}$ as the tangent sequence/multipliers.

Note that if x^* satisfies any one of the definitions in Definition 2.1, then Assumption 2.1 holds (around x^*). When $\mathcal{S}=\emptyset$ this assumption holds vacuously, and we recover the classic definitions from [4, 5, 10]. The proof that *i)* and *iii)'* hold (an AKKT point with $\mathcal{S}:=\emptyset$) was implicitly done in [12]. See also [4], where *i), iii)'* and *v)* was proved. Conditions *i), ii), iii)'* and *iv)'* (an AKKT2 point with $\mathcal{S}:=\emptyset$) were proven in [5]. Conditions *i)* together with the stronger complementarity measure *iii)* (a CAKKT point with $\mathcal{S}:=\emptyset$) were proven in [10]. Theorem 2.1 proves that *i)*-*v)* (what we call a CAKKT2 point) can be satisfied simultaneously by the same set of sequences. In particular, it gives a second-order version of the optimality condition from [10], including a new strong complementarity measure *iv)* also for the approximate tangent multipliers, while also proving condition *v)* (interiority) for the constraints fulfilling the sufficient interior property (Assumption 2.1), in a similar fashion as [4]. Note that Theorem 2.1 also extends the definitions of AKKT2 and CAKKT to consider *v)* under Assumption 2.1.

The main contribution of Theorem 2.1 is the inclusion of the complementarity measure *iv)* for the approximate tangent multipliers, stronger than *iv)'*. Note that *iv)* and *iv)'* impose the same requirement for tangent multipliers associated with inactive constraints $(\mathcal{I}\backslash\mathcal{I}_{x^*}) \cup (\mathcal{S}\backslash\mathcal{S}_{x^*})$, that is, for k large enough, they must be zero, or replaceable by zero, but for the active constraints $\mathcal{E} \cup \mathcal{I}_{x^*} \cup \mathcal{S}_{x^*}$, where tangent multipliers are expected to go to $+\infty$, *iv)'* allows for arbitrarily fast divergence of tangent multipliers with respect to the other parameters, while *iv)* requires the growth of tangent multipliers to be slower than the square of the reciprocal of the constraint value, that is, $\theta_i^k = o\left(\frac{1}{h_i(x^k)^2}\right)$ for $i \in \mathcal{E}$, $\eta_i^k = o\left(\frac{1}{g_i(x^k)^2}\right)$ for $i \in \mathcal{I}_{x^*}$ and $\nu_i^k = o\left(\frac{1}{c_i(x^k)^2}\right)$, $i \in \mathcal{S}_{x^*}$.

Theorem 2.1 immediately suggests the following stopping criterion for algorithms that generate CAKKT2 sequences:

- $\|h(x^k)\| + \|\max\{0, g(x^k)\}\| \leq \varepsilon_{feas}$ (feasibility),
- $\|\nabla L(x^k, \lambda^k, \mu^k, s^k)\| \leq \varepsilon_{opt1}$ (first-order optimality),
- $\lambda_{min}(Q_k(x^k, \lambda^k, \mu^k, s^k, \theta^k, \eta^k, \nu^k)) \geq -\varepsilon_{opt2}$ (second-order optimality),
- $\sum_{i\in\mathcal{E}} |h_i(x^k)\lambda_i^k| + \sum_{i\in\mathcal{I}} |g_i(x^k)\mu_i^k| + \sum_{i\in\mathcal{S}} c_i(x^k)s_i^k \leq \varepsilon_{comp1}$ (first-order complementarity),
- $\sum_{i\in\mathcal{E}} h_i(x^k)^2\theta_i^k + \sum_{i\in\mathcal{I}} g_i(x^k)^2\eta_i^k + \sum_{i\in\mathcal{S}} c_i(x^k)^2\nu_i^k \leq \varepsilon_{comp2}$ (second-order complementarity),

for tolerances $\varepsilon_{feas} > 0, \varepsilon_{opt1} > 0, \varepsilon_{opt2} > 0, \varepsilon_{comp1} > 0$ and $\varepsilon_{comp2} > 0$. In the case of an AKKT2 sequence, a stopping criterion based on a perturbation of the critical subspace can be developed. See [16].

The following example shows that CAKKT2 is strictly stronger than AKKT2 and CAKKT together.

Example 2.1. Consider the problem

$$\begin{array}{ll} \text{Minimize} & f(x_1, x_2) := x_2 - e^{x_2}, \\ \text{subject to} & h(x_1, x_2) := x_1 = 0, \\ & g_1(x_1, x_2) := x_1^3 \leq 0, \\ & g_2(x_1, x_2) := x_1 e^{x_2} \leq 0, \end{array}$$

at the feasible point $x^* := (0, 0)$ and let us take the sequences $x_1^k := \frac{1}{k}, x_2^k := 0, \lambda^k := 0, \mu_1^k := 0, \mu_2^k := 0$ for $k \geq 1$. Clearly, this sequence attests that x^* satisfies CAKKT. To prove that this sequence also attests that AKKT2 holds, let us look at the matrix of second derivatives $Q_k(x^k, \lambda^k, \mu^k, \theta^k, \eta^k)$, given by

$$Q_k(\cdot) := \left(\begin{array}{cc} 6x_1^k\mu_1^k + \theta_1^k + 9\eta_1^k(x_1^k)^4 + \eta_2^k e^{2x_2} & \mu_2^k e^{x_2^k} + \eta_2^k x_1^k e^{2x_2^k} \\ \mu_2^k e^{x_2^k} + \eta_2^k x_1^k e^{2x_2^k} & -e^{x_2^k} + \mu_2^k x_1^k e^{x_2^k} + \eta_2^k (x_1^k)^2 e^{2x_2^k} \end{array} \right).$$

For $\theta_1^k := k^5, \eta_1^k := 0, \eta_2^k := k^3$ we have

$$Q_k(x^k, \lambda^k, \mu^k, \theta^k, \eta^k) = \left(\begin{array}{cc} k^5 + k^3 & k^2 \\ k^2 & -1 + k \end{array} \right),$$

which is positive definite for $k > 1$, hence, AKKT2 holds. Note that the stringent criterion *iv)* does not hold for this sequence. In fact, let us prove that x^* is not a CAKKT2 point. From *iv)*, any CAKKT2 sequence must be such that $\eta_2^k(x_1^k)^2 e^{2x_2^k} \to 0$, while *iii)* gives $\mu_2^k x_1^k e^{x_2^k} \to 0$. Hence the bottom-right element of the matrix converges to -1. This shows that $\liminf\limits_{k\to+\infty} \lambda_{min}(Q_k) < 0$, hence CAKKT2 fails.

This example shows that an algorithm may generate a sequence converging to a non-minimizer where the AKKT2 and CAKKT stopping criteria are satisfied, but CAKKT2 fails. This suggests that a second-order algorithm should include the more stringent stopping criterion associated with CAKKT2 to rule out as much as possible the possibility of declaring convergence to a non-solution.

Theorem 2.1 states that, in theory, around a local minimizer, one is always able to drive the second-order stationarity residual to zero while controlling the size of Lagrange multipliers *iii)* and tangent multipliers *iv)*. In the next section we prove that this can be done in practice by Augmented Lagrangian and Interior Point Methods.

We end this section by commenting that CAKKT2 does not give rise to a sufficient optimality condition. The reason is that it does not completely avoid the choice of a tangent multiplier that ignores the second-order information, in the sense that it can still go to infinity too fast. A counter-example would be the problem of minimizing x_1+x_2 subject to $x_1^2x_2=0$ at $x^*:=(0,0)$ with sequences $x^k:=(1/k,1/(2k)), \lambda^k:=-k^2, \theta^k:=2k^5$, where CAKKT2 holds with the matrix Q_k being positive definite for all k. In fact, it was not expected that CAKKT2 would give rise to a sufficient optimality condition, given that it is based on the critical subspace $S(x^*)$, and not the critical cone $C(x^*)$, but the example shows that this is not the case even when these sets coincide.

3 Algorithms satisfying the optimality condition

In the recent years, many algorithms with second-order global convergence properties have been developed for several classes of constrained optimization. See, for instance, [25, 38, 37, 36, 13, 3, 27, 35, 24, 21, 23], and many others. Recently, [5] has shown that it is expected that many different second-order algorithms generate AKKT2 sequences. In particular, it has been shown that trust region ([24]), augmented Lagrangian ([3]) and regularized sequential quadratic programming ([27]) generate AKKT2 sequences. For first-order algorithms, [10] shows that the general lower-level augmented Lagrangian ([2]) generate CAKKT sequences. The same holds true for the regularized sequential quadratic programming (see the Discussion at the end of Section 3 in [27]). Our proof of Theorem 2.1 suggests that first-order interior point methods would also generate CAKKT sequences. Indeed, to show this, we consider the first-order quasi-feasible interior point method from [20]. We consider a general nonlinear objective function f and inequality constraints $c(x)\geq 0$ conforming to Assumption 2.1. It is sufficient to note that [20, Equations (2.7) and (3.13)] ensures that CAKKT holds. See also [6, Equation (5.17)].

In this section we will show that, at least second-order variants of augmented Lagrangian and interior point methods are expected to fulfill the stronger second-order optimality condition CAKKT2. This is essentially due to the way constraints are penalized in those methods.

We start by showing that augmented Lagrangian methods satisfy our optimality condition under an additional smoothness assumption.

Let us consider $\mathcal{S}:=\emptyset$ and let us consider an augmented Lagrangian method for solving problem (1). For fixed Lagrange multipliers approximations $(\lambda,\mu)\in\mathbb{R}^{|\mathcal{E}|}\times\mathbb{R}^{|\mathcal{I}|}_+$ and penalty parameter $\rho>0$, let us define the augmented Lagrangian function $\mathbb{R}^n\ni x\mapsto L_\rho(x,\lambda,\mu)$ as

$$L_\rho(x,\lambda,\mu):=f(x)+\frac{\rho}{2}\left(\left\|h(x)+\frac{\lambda}{\rho}\right\|^2+\left\|\max\left\{0,g(x)+\frac{\mu}{\rho}\right\}\right\|^2\right).$$

The algorithm consists of approximately minimizing the augmented Lagrangian function at each iteration and updating the Lagrange multipliers and penalty parameter accordingly. See [14] for details. The gradient of L_ρ with respect to x is given by $\nabla L_\rho(x,\lambda,\mu)=\nabla f(x)+\sum_{i\in\mathcal{E}}(\lambda_i+\rho h_i(x))\nabla h_i(x)+\sum_{i\in\mathcal{I}}\max\{0,\mu_i+\rho g_i(x)\}\nabla g_i(x)$, and the matrix below plays a role of its Hessian matrix:

$$\begin{aligned}\nabla^2 L_\rho(x,\lambda,\mu) \quad := \quad & \nabla^2 f(x)+\sum_{i\in\mathcal{E}}(\lambda_i+\rho h_i(x))\nabla^2 h_i(x)+\sum_{i\in\mathcal{I}}\max\{0,\mu+\rho g_i(x)\}\nabla^2 g_i(x)\\ & +\sum_{i\in\mathcal{E}}\rho\nabla h_i(x)\nabla h_i(x)^{\mathtt{T}}+\sum_{i\in\mathcal{I}:\mu_i+\rho g_i(x)\geq 0}\rho\nabla g_i(x)\nabla g_i(x)^{\mathtt{T}}.\end{aligned}$$

Algorithm 3.1. *(Second-order Augmented Lagrangian ([9]))*
Step 0 (Initialization): *Set parameters* $\lambda_{min}<\lambda_{max}$, $\mu_{max}>0$, $\gamma>1$, $\tau\in[0,1)$ *and an iteration counter* $k:=1$. *Choose initial approximations* $x^0\in\mathbb{R}^n$, $\bar\lambda^1\in\mathbb{R}^{|\mathcal{E}|}$, $\bar\mu^1\in\mathbb{R}^{|\mathcal{I}|}_+$, *and an initial penalty parameter* $\rho_1>0$.

Set $V^0 := \max\{0, g(x^0)\}$ *and a sequence of subproblem tolerances* $\{\varepsilon_k\} \to 0^+$.
Step 1 (Solve subproblem): *Find* $x^k \in \mathbb{R}^n$ *such that*

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)\| \leq \varepsilon_k,$$
$$\lambda_{min}(\nabla^2 L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)) \geq -\varepsilon_k.$$

Step 2 (Estimate multipliers): *Define* $\lambda^k := \bar{\lambda}^k + \rho_k h(x^k)$ *and* $\mu^k := \max\{0, \bar{\mu}^k + \rho_k g(x^k)\}$. *Compute* $\bar{\lambda}^{k+1} \in \mathbb{R}^{|\mathcal{E}|}$ *and* $\bar{\mu}^{k+1} \in \mathbb{R}^{|\mathcal{I}|}_+$ *on the safeguarded boxes* $\bar{\lambda}_i^{k+1} \in [\lambda_{min}, \lambda_{max}], i \in \mathcal{E}$ *and* $\bar{\mu}_i^{k+1} \in [0, \mu_{max}], i \in \mathcal{I}$. *(Typically, these are computed as the projections of* λ^k *and* μ^k *on the corresponding safeguarded boxes.)*
Step 3 (Update penalty parameter): *Set* $V^k := \max\left\{g(x^k), -\frac{\bar{\mu}^k}{\rho_k}\right\}$.
If $\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\}$, *then define* $\rho_{k+1} \geq \rho_k$. *Otherwise, define* $\rho_{k+1} \geq \gamma\rho_k$.
Step 4 *Set* $k := k+1$ *and go to Step 1.*

To prove that Algorithm 3.1 generates CAKKT2 sequences, we will need the following smoothness assumption:

Assumption 3.1. *Any feasible limit point* $x^* \in \Omega$ *of Algorithm 3.1 satisfies the Generalized Lojasiewicz inequality below:*
There exist $\delta > 0$ *and a function* $\varphi : B(x^*, \delta) \to \mathbb{R}$ *with* $\varphi(x) \to 0$ *as* $x \to x^*$ *such that for all* $x \in B(x^*, \delta)$, $|\Phi(x) - \Phi(x^*)| \leq \varphi(x)\|\nabla\Phi(x)\|$, *where* $\Phi(x) := \|h(x)\|^2 + \|\max\{0, g(x)\}\|^2$.

The Generalized Lojasiewicz inequality holds, for instance, when all functions $h_i, i \in \mathcal{E}$ and $g_i, i \in \mathcal{I}$ are analytic, see [10]. For example, since $\Phi(x^*) = 0$ at a feasible x^*, it is easy to see that for polynomial functions h_i and g_i, we have $\frac{\Phi(x)}{\|\nabla\Phi(x)\|} \to 0$ as $x \to x^*$, implying that the condition holds. Much weaker smoothness conditions also imply Assumption 3.1, see [17] and [33].

Theorem 3.1. *Let* $x^* \in \Omega$ *be a feasible limit point of a sequence* $\{x^k\}$ *generated by Algorithm 3.1 such that Assumption 3.1 holds. Then* x^* *satisfies CAKKT2.*

Proof: By the definition of λ^k and μ^k from Step 2, the definition of the augmented Lagrangian function, and the resolution of the subproblems, we have $\|\nabla L(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k$ and $\lambda_{min}(Q_k) \geq -\varepsilon_k$, with

$$\nabla L(x^k, \lambda^k, \mu^k) = \nabla f(x^k) + \sum_{i \in \mathcal{E}} (\bar{\lambda}_i^k + \rho_k h_i(x^k))\nabla h_i(x^k) + \sum_{i \in \mathcal{I}} \max\{0, \bar{\mu}_i^k + \rho g_i(x^k)\}\nabla g_i(x^k)$$

and

$$\begin{aligned} Q_k \quad := \quad & \nabla^2 f(x^k) + \sum_{i \in \mathcal{E}} (\bar{\lambda}_i^k + \rho h_i(x^k))\nabla^2 h_i(x^k) + \sum_{i \in \mathcal{I}} \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\}\nabla^2 g_i(x^k) \\ & + \sum_{i \in \mathcal{E}} \rho_k \nabla h_i(x^k)\nabla h_i(x^k)^\mathtt{T} + \sum_{i \in \mathcal{I} : \bar{\mu}_i^k + \rho_k g_i(x^k) \geq 0} \rho_k \nabla g_i(x^k)\nabla g_i(x^k)^\mathtt{T}. \end{aligned}$$

Defining $\theta_i^k := \rho_k, i \in \mathcal{E}$ and $\eta_i^k := \rho_k$ if $\bar{\mu}_i^k + \rho_k g_i(x^k) \geq 0$ and $\eta_i^k := 0$ otherwise, we have that *i)* and *ii)* in the definition of CAKKT2 holds. If $g_i(x^*) < 0$ we have two cases: If $\rho_k \to +\infty$, since $\{\bar{\mu}^k\}$ is bounded, we have $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$ for all k sufficiently large. If $\{\rho_k\}$ is bounded, Step 3 of the Algorithm implies that $\rho_k := \rho_{k_0}$ for k large enough and $V^k \to 0$. In particular, $\frac{\bar{\mu}_i^k}{\rho_{k_0}} \to 0^+$, thus, $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$ for all k sufficiently large. Hence, from both cases, the definition of μ^k and η^k gives $\mu_i^k = 0$ and $\eta_i^k = 0$ for k large enough, which implies $\mu_i^k g_i(x^k) = 0$ and $\eta_i^k g_i(x^k)^2 = 0$ for k large enough (this gives, in particular, *iii)'* and *iv)'* of the AKKT2 definition). To prove the stronger conditions *iii)* and *iv)*, let us look at the set of active constraints $\mathcal{E} \cup \mathcal{I}_{x^*}$. When $\{\rho_k\}$ is bounded we have by definition that $\{\lambda^k\}, \{\mu^k\}, \{\theta^k\}$ and $\{\nu^k\}$ are all bounded sequences, hence *iii)* and *iv)* hold. Assume $\rho_k \to +\infty$.

Noting that $\{\nabla L(x^k, \lambda^k, \mu^k)\}$ is bounded, we can follow the proof of [10, Theorem 5.1] to see that under the Generalized Lojasiewicz inequality, we have $\rho_k h_i(x^k)^2 \to 0, i \in \mathcal{E}$ and $\rho_k \max\{0, g_i(x^k)\}^2 \to 0, i \in \mathcal{I}$.

Hence, for $i \in \mathcal{E}$, by the boundedness of $\{\bar{\lambda}^k\}$ we have $\lambda_i^k h_i(x^k) = (\bar{\lambda}_i^k + \rho_k h_i(x^k))h_i(x^k) \to 0$. Similarly, by the boundedness of $\{\bar{\mu}^k\}$, for $i \in \mathcal{I}_{x^*}$,

$$
\begin{aligned}
0 \leq \mu_i^k \max\{0, g_i(x^k)\} &= \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \max\{0, g_i(x^k)\} \\
&\leq (\bar{\mu}_i^k + \rho_k \max\{0, g_i(x^k)\}) \max\{0, g_i(x^k)\} \to 0.
\end{aligned}
$$

Now, if $g_i(x^k) \geq 0$ for all k sufficiently large, we have $\mu_i^k g_i(x^k) \to 0$, otherwise, we can take a subsequence such that $\mu_i^k := \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \leq \bar{\mu}_i^k$ and $\{\mu_i^k\}$ is bounded, which implies $\mu_i^k g_i(x^k) \to 0$ and *iii)* holds.

To prove *iv)*, for $i \in \mathcal{E}$, we have $\theta_i^k h_i(x^k)^2 = \rho_k h_i(x^k)^2 \to 0$. Now let $i \in \mathcal{I}_{x^*}$. By the definition of η_i^k, if there is an infinite subsequence such that $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$, we have $\eta_i^k = 0$ and we can take the limit for this subsequence to get $\eta_i^k g_i(x^k)^2 \to 0$. Otherwise, $\bar{\mu}_i^k + \rho_k g_i(x^k) \geq 0$ and $\eta_i^k = \rho_k$ for all sufficiently large k. Hence, $\eta_i^k \max\{0, g_i(x^k)\}^2 = \rho_k \max\{0, g_i(x^k)\}^2 \to 0$ and the result holds if $g_i(x^k) \geq 0$ for all k sufficiently large.

Finally, assume that there is a subsequence such that $g_i(x^k) < 0$ for all k in this subsequence. Since $\bar{\mu}_i^k + \rho_k g_i(x^k) \geq 0$, we have $|\rho_k g_i(x^k)| \leq \bar{\mu}_i^k$ and $\{\rho_k g_i(x^k)\}$ is bounded, which implies $\eta_i^k g_i(x^k)^2 \to 0$. ☐

Without Assumption 3.1, condition *iii)* may fail. See [10].

We mention that even though an approximate second-order stationary point is not guaranteed to be found at Step 3 in most well-known implementations of augmented Lagrangians such as the ones in [22] and [2], computational heuristics are included in the subproblem solver in such a way that second-order stationary points are usually found. A rigorous implementation of Algorithm 3.1 can be found in [3]. See also the discussion in [16], where more general augmented Lagrangian methods are shown to generate CAKKT2 sequences.

The natural case where our optimality condition is satisfied is in the relm of interior point methods due to the way first-order complementarity is treated in the definition of the central path. We will show that second-order complementarity is also fulfilled.

Let us consider the problem

$$
\begin{array}{ll}
\text{Minimize} & f(x), \\
\text{subject to} & x \geq 0,
\end{array}
$$

where f is smooth, and let us show that the interior-point trust-region algorithm from [13] is such that tangent multipliers can be defined such that second-order complementarity holds (Theorem 2.1, item *iv)*). The same holds true for the classical affine scaling method from [39, 40] when f is a possibly non-convex quadratic function under linear constraints.

Given a sequence of tolerances $\{\varepsilon_k\} \to 0^+$ and $x^k > 0$, the next iterate $x^{k+1} := x^k + X_k d > 0$ is built by computing the global minimizer d of the following ball-constrained quadratic programming:

$$
\text{Minimize } \frac{1}{2} d^{\mathrm{T}} X^k \nabla^2 f(x^k) X^k d + (X^k \nabla f(x^k))^{\mathrm{T}} d, \text{ s.t. } \|d\| \leq \beta \varepsilon_k,
$$

for some fixed $0 < \beta \leq \frac{1}{4}$, where X^k is the diagonal matrix defined by x^k. An efficient procedure for computing d is given in [39, 40]. In [13, Theorem 2] it was proved that for k sufficiently large, it holds that

$$
M_k := X^k \nabla^2 f(x^k) X^k + \varepsilon_k I \succeq 0.
$$

Since X^k is positive definite by construction, we have that $(X^k)^{-2} M_k \succeq 0$, where $(X^k)^{-2} M_k = \nabla^2 f(x^k) + \varepsilon_k (X^k)^{-2} = \nabla^2 f(x^k) + \sum_{i=1}^n \frac{\varepsilon_k}{(x_i^k)^2} e_i e_i^{\mathrm{T}}$, with e_i being the i-th canonical vector, $i = 1, \ldots, n$. Defining the tangent multipliers to be $\nu_i^k := \varepsilon_k / (x_i^k)^2, i = 1, \ldots, n$, we have that second-order complementarity holds given that $c(x) := x$. For more general interior point methods generating CAKKT2 sequences, see [31].

We note that the validity of CAKKT2, in contrast with AKKT2, is essentially due to the way constraints are penalized in augmented Lagrangian and interior point methods. To complete this discussion, we will show that when an active-set strategy is employed, CAKKT2 is not expected to hold. Consider the problem

$$
\begin{array}{ll}
\text{Minimize} & f(x), \\
\text{subject to} & h_i(x) = 0, i \in \mathcal{E}, \\
 & x \geq 0,
\end{array}
$$

that is, $c(x) := x$ and $\mathcal{I} := \emptyset$ in (1). We will show that the regularized sequential quadratic programming from [27] and the active-set method for box-constraints ($\mathcal{E} := \emptyset$) from [3] are not expected to generate CAKKT2 sequences, even though both methods generate AKKT2 sequences. To simpifly the exposition we consider only the case $\mathcal{E} := \emptyset$, but the argument generalizes easily to the more general case. In both methods, let $\{\varepsilon_k\} \to 0^+$ be a given sequence of tolerances. A primal sequence $\{x^k\}$ is generated in such a way that the second-order condition is imposed by ensuring that the least eigenvalue of the matrix H_k is at least $-\varepsilon_k$, where H_k is a matrix obtained from $\nabla^2 f(x^k)$ associated with the ε_k-free variables, that is, H_k is obtained from $\nabla^2 f(x^k)$ by removing its i-th row and i-th column whenever $x_i^k \leq \varepsilon_k$. In other words, it holds that $d^{\mathrm{T}} \nabla^2 f(x^k) d \geq -\varepsilon_k \|d\|^2$ for all d such that $e_i^{\mathrm{T}} d = 0, i : x_i^k \leq \varepsilon_k$, where $e_i = \nabla c_i(x^k)$ is the i-th canonical vector. In this case, it can be seen that the critical subspace is dealt explicitly by the algorithm, in such a way that tangent multipliers can be obtained only by means of Proposition 2.1, where there is no indication that second-order complementarity (Theorem 2.1, item *iv)*) would hold.

Finally, we note that checking that an algorithm satisfies our optimality condition, or its weaker variants, is usually easier than checking directly that it satisfies KKT or WSOC on its limit points under a suitable constraint qualification, since we can deal explicitly with the actual sequences generated by the algorithm without the need to bound the dual variables in order to take limits. Depending on which optimality condition is verified, a companion minimal constraint qualification is available (see the next section) such that KKT or WSOC is guaranteed to hold.

We end this section by noting the connection of our optimality condition with results on the iteration complexity for achieving a second-order stationary point perturbed by $\varepsilon > 0$, even when only linear constraints are considered. See, for instance, [13]. Usually, the introduction of such perturbations are motivated only by the fact that usual second-order stationarity is recovered when $\varepsilon := 0$. Our result suggests a way to formulate such perturbed conditions in such a way that complexity results are meaningful with respect to optimality.

For the connection of CAKKT2 with complexity analysis and more general interior point methods, we refer the reader to [31].

4 Strength of the optimality condition

It is clear that CAKKT2 implies the second-order optimality condition AKKT2 ([5]) and the first-order optimality condition CAKKT ([10]). In order to measure the strength of CAKKT2, let us review previous results about the strength of CAKKT and AKKT2.

We begin this discussion by considering $\mathcal{S} = \emptyset$ in (1). Under a constraint qualification, KKT conditions hold at a local minimizer. In other words, different constraint qualifications (CQ) gives rise to different optimality conditions of the form:

$$\text{KKT or not-CQ.} \tag{12}$$

The weaker the CQ, the stronger the optimality condition (12) is. It has been proved that AKKT is strictly stronger than (12), with CQ replaced by MFCQ, (R)CRCQ, (R)CPLD, CRSC or CPG. See [6] and references therein. Or the even weaker one CCP (AKKT-regularity) from [7]. Since CAKKT implies AKKT, we also have that CAKKT implies (12) for these CQs. Recently, an even weaker constraint qualification that precisely characterizes the strength of CAKKT has been defined in [8] under the name of CAKKT-regularity. In order to define it, we will consider the notion of outer semicontinuity of set-valued mappings.

Given a set-valued mapping $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$ with $\mathbb{R}^s \ni x \mapsto F(x) \subseteq \mathbb{R}^d$ and a set $D \subseteq \mathbb{R}^s$, the *sequential Painlevé-Kuratowski outer limit* over D of $F(z)$ as $z \to z^*$ is given by:

$$\limsup_{D \ni z \to z^*} F(z) := \left\{ \begin{array}{l} w^* \in \mathbb{R}^d : z^k \to z^*, w^k \to w^* \text{ for some } \{(z^k, w^k)\} \\ \text{with } w^k \in F(z^k), w^k \in D \text{ for all } k \end{array} \right\},$$

and $F(z)$ is said to be outer semicontinuous at z^* over D if $\limsup_{D \ni z \to z^*} F(z) \subseteq F(z^*)$. When $D := \mathbb{R}^s$, the set D is omitted from the notation.

Definition 4.1 ([8])**.** We say that $x^* \in \Omega$ satisfies CAKKT-regularity if the set-valued mapping $\mathbb{R}^n \times \mathbb{R}_+ \ni$

$(x,r) \mapsto K^C(x,r)$ is outer semicontinuous at $(x^*,0)$, where

$$K^C(x,r) := \left\{ \sum_{i\in\mathcal{E}} \lambda_i \nabla h_i(x) + \sum_{i\in\mathcal{I}_{x^*}} \mu_i \nabla g_i(x) : \begin{array}{r} \sum_{i\in\mathcal{E}} |\lambda_i h_i(x)| + \sum_{i\in\mathcal{I}_{x^*}} |\mu_i g_i(x)| \le r, \\ (\lambda,\mu) \in \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{I}|}_+ \end{array} \right\}.$$

The following theorem states that the constraint qualification CAKKT-regularity is the weakest constraint qualification such that the optimality condition (12) is implied by CAKKT.

Theorem 4.1 ([8]). *A feasible point $x^* \in \Omega$ satisfies CAKKT-regularity if, and only if, for every continuously differentiable objective function f such that x^* is CAKKT, x^* is also a KKT point.*

Similarly, under some constraint qualifications (CQ2), the following second-order optimality condition holds:

$$\text{WSOC or not-CQ2.} \tag{13}$$

It has been proved in [5] that AKKT2 is strictly stronger than (13) with CQ2 replaced by MFCQ+WCR or (R)CRCQ. See [5] and references therein for details. In particular, global convergence to a second-order stationary point of several algorithms has been proved even when the solution has an unbounded set of Lagrange multipliers (that is, without MFCQ). Also, since linear constraints imply (R)CRCQ, the results from [5] imply that second-order nonlinear solvers can find second-order stationary points in the case of linear constraints without removing redundancies introduced by linear dependent constraints. An even weaker constraint qualification is introduced in [5] that exactly measures the strength of AKKT2, that we define below:

Definition 4.2 ([5]). We say that $x^* \in \Omega$ satisfies the second-order cone continuity property (CCP2, or AKKT2-regularity) if the set-valued mapping $\mathbb{R}^n \ni x \mapsto K_2(x)$ is outer semicontinuous at x^*, where

$$K_2(x) := \bigcup_{\substack{(\lambda,\mu)\in\mathbb{R}^{|\mathcal{E}|}\times\mathbb{R}^{|\mathcal{I}|}_+, \\ \mu_i=0 \text{ for } i\notin\mathcal{I}_{x^*}}} \left\{ \begin{array}{l} (w,W) \in \mathbb{R}^n \times \mathrm{Sym}(n) : \\ w + \sum_{i\in\mathcal{E}} \lambda_i \nabla h_i(x) + \sum_{i\in\mathcal{I}_{x^*}} \mu_i \nabla g_i(x) = 0, \\ d^{\mathrm{T}} \left(W + \sum_{i\in\mathcal{E}} \lambda_i \nabla^2 h_i(x) + \sum_{i\in\mathcal{I}_{x^*}} \mu_i \nabla^2 g_i(x) \right) d \ge 0, \\ \hfill \forall d \in S(x,x^*) \end{array} \right\}, \tag{14}$$

where $S(x,x^*) := \{d \in \mathbb{R}^n \mid \nabla h(x)^{\mathrm{T}} d = 0, \nabla g_{\mathcal{I}_{x^*}}(x)^{\mathrm{T}} d = 0\}$ is the perturbed critical subspace.

Theorem 4.2 ([8]). *A feasible point $x^* \in \Omega$ satisfies AKKT2-regularity if, and only if, for every twice continuously differentiable objective function f such that x^* is AKKT2, x^* also satisfies WSOC.*

Clearly, since CAKKT2 implies AKKT2, CAKKT2 implies (13) with CQ2 replaced by MFCQ+WCR, (R)CRCQ or AKKT2-regularity. Let us now define a weaker constraint qualification that exactly measures the strength of CAKKT2. Let us consider again the additional constraints $c_i(x) \ge 0, i \in \mathcal{S}$.

Definition 4.3. We say that $x^* \in \Omega$ satisfies CAKKT2-regularity if the set-valued mapping $\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+ \ni (x,r_1,r_2) \mapsto K_2^C(x,r_1,r_2)$ is outer semicontinuous at x^* over $D := \{(x,r_1,r_2) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+ : c(x) > 0\}$, where

$$K_2^C(x,r_1,r_2) := \bigcup_{(\lambda,\mu,s)\in\mathbb{R}^{|\mathcal{E}|}\times\mathbb{R}^{|\mathcal{I}|}_+\times\mathbb{R}^{|\mathcal{S}|}_+} \limsup_{\bar{\theta}\to+\infty,\varepsilon\to 0^+} A(x,\lambda,\mu,s,r_1,r_2,\bar{\theta},\varepsilon), \tag{15}$$

with $A(x, \lambda, \mu, s, r_1, r_2, \bar{\theta}, \varepsilon)$ given by

$$A(\cdot) := \left\{ \begin{array}{l} (w, W) \in \mathbb{R}^n \times \mathrm{Sym}(n) : \\ w + \sum_{i\in\mathcal{E}} \lambda_i \nabla h_i(x) + \sum_{i\in\mathcal{I}} \mu_i \nabla g_i(x) - \sum_{i\in\mathcal{S}} s_i \nabla c_i(x) = 0 \\ \sum_{i\in\mathcal{E}} |\lambda_i h_i(x)| + \sum_{i\in\mathcal{I}} |\mu_i g_i(x)| + \sum_{i\in\mathcal{S}} s_i c_i(x) \le r_1, \\ \sum_{i\in\mathcal{E}} \theta_i h_i(x)^2 + \sum_{i\in\mathcal{I}} \eta_i g_i(x)^2 + \sum_{i\in\mathcal{S}} \nu_i c_i(x)^2 \le r_2, \text{ and} \\ \lambda_{min} \big(W + \sum_{i\in\mathcal{E}} \lambda_i \nabla^2 h_i(x) + \sum_{i\in\mathcal{I}} \mu_i \nabla^2 g_i(x) \\ + \sum_{i\in\mathcal{S}} s_i \nabla^2 c_i(x) + \sum_{i\in\mathcal{E}} \theta_i \nabla h_i(x) \nabla h_i(x)^\mathtt{T} \\ + \sum_{i\in\mathcal{I}} \eta_i \nabla g_i(x) \nabla g_i(x)^\mathtt{T} + \sum_{i\in\mathcal{S}} \nu_i \nabla c_i(x) \nabla c_i(x)^\mathtt{T} \big) \ge -\varepsilon, \\ \bar{\theta} \ge \sum_{i\in\mathcal{E}} \theta_i + \sum_{i\in\mathcal{I}} \eta_i + \sum_{i\in\mathcal{S}} \nu_i, \\ \text{for some } (\theta, \eta, \nu) \in \mathbb{R}^{|\mathcal{E}|}_+ \times \mathbb{R}^{|\mathcal{I}|}_+ \times \mathbb{R}^{|\mathcal{S}|}_+ \end{array} \right\}. \tag{16}$$

The following lemma shows that we can recast WSOC in terms of the cone $K_2^C(x^*, 0, 0)$.

Lemma 4.1. *$x^* \in \Omega$ satisfies WSOC if, and only if, $(\nabla f(x^*), \nabla^2 f(x^*)) \in K_2^C(x^*, 0, 0)$.*

Proof: Let x^* satisfy WSOC. Let (λ^*, μ^*, s^*) be correspondent Lagrange multipliers associated with x^*. From Proposition 2.1, $\liminf_{\bar{\theta}\to+\infty} \lambda_{min}(Q(\bar{\theta})) \ge 0$, where

$$Q(\bar{\theta}) := \nabla^2 f(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* \nabla^2 h_i(x^*) + \sum_{i\in\mathcal{I}_{x^*}} \mu_i^* \nabla^2 g_i(x^*) + \sum_{i\in\mathcal{S}_{x^*}} s_i^* \nabla^2 c_i(x^*) +$$
$$\sum_{i\in\mathcal{E}} \bar{\theta} \nabla h_i(x^*) \nabla h_i(x^*)^\mathtt{T} + \sum_{i\in\mathcal{I}_{x^*}} \bar{\theta} \nabla g_i(x^*) \nabla g_i(x^*)^\mathtt{T} + \sum_{i\in\mathcal{S}_{x^*}} \bar{\theta} \nabla c_i(x^*) \nabla c_i(x^*)^\mathtt{T}.$$

Thus, for $\varepsilon_k := \frac{1}{k}$, there exists a sequence $\{\bar{\theta}_k\} \to +\infty$ with $\lambda_{min}(Q(\bar{\theta}_k)) \ge -\varepsilon_k$, hence, for all k, $(\nabla f(x^*), \nabla^2 f(x^*)) \in A(x^*, \lambda^*, \mu^*, s^*, 0, 0, \bar{\theta}_k, \varepsilon_k)$, which implies $(\nabla f(x^*), \nabla^2 f(x^*)) \in K_2^C(x^*, 0, 0)$. Conversely, assume $(\nabla f(x^*), \nabla^2 f(x^*)) \in K_2^C(x^*, 0, 0)$. Then, there is $(\lambda^*, \mu^*, s^*) \in \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{I}|}_+ \times \mathbb{R}^{|\mathcal{S}|}_+$ and sequences $\{\bar{\theta}_k\} \to +\infty, \{\varepsilon_k\} \to 0^+, \{w^k\} \to \nabla f(x^*), \{W^k\} \to \nabla^2 f(x^*)$ such that $(w^k, W^k) \in A(x^*, \lambda^*, \mu^*, s^*, 0, 0, \bar{\theta}_k, \varepsilon_k)$ for all k, corresponding to some sequences $\{(\theta^k, \eta^k, \nu^k)\} \subset \mathbb{R}^{|\mathcal{E}|}_+ \times \mathbb{R}^{|\mathcal{I}|}_+ \times \mathbb{R}^{|\mathcal{S}|}_+$. From the definition of the set $A(\cdot)$ (16), we have that x^* is a KKT point and

$$W^k + \sum_{i\in\mathcal{E}} \lambda_i^* \nabla^2 h_i(x^*) + \sum_{i\in\mathcal{I}_{x^*}} \mu_i^* \nabla^2 g_i(x^*) + \sum_{i\in\mathcal{S}_{x^*}} s_i^* \nabla^2 c_i(x^*) +$$
$$\sum_{i\in\mathcal{E}} \theta_i^k \nabla h_i(x^*) \nabla h_i(x^*)^\mathtt{T} + \sum_{i\in\mathcal{I}_{x^*}} \eta_i^k \nabla g_i(x^*) \nabla g_i(x^*)^\mathtt{T} +$$
$$\sum_{i\in\mathcal{S}_{x^*}} \nu_i^k \nabla c_i(x^*) \nabla c_i(x^*)^\mathtt{T} + \varepsilon_k I \succeq 0.$$

Thus, for $d \in S(x^*)$, since $W^k \to \nabla^2 f(x^*)$ and $\varepsilon_k \to 0^+$ we have

$$d^\mathtt{T} \nabla L(x^*, \lambda^*, \mu^*, s^*) d \ge 0$$

and WSOC holds. □

We now present our characterization of the strength of CAKKT2.

Theorem 4.3. *A feasible point $x^* \in \Omega$ satisfies CAKKT2-regularity if, and only if, for every twice continuously differentiable objective function f such that x^* is CAKKT2, x^* also satisfies WSOC.*

Proof: Let us assume that $x^* \in \Omega$ satisfies CAKKT2-regularity. Let f be a twice continuously differentiable objective function such that x^* is a CAKKT2 point with sequences $\{(x^k, \lambda^k, \mu^k, s^k, \theta^k, \eta^k, \nu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^{|\mathcal{E}|} \times$

$\mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}_+^{|\mathcal{S}|} \times \mathbb{R}_+^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}_+^{|\mathcal{S}|}$ such that $x^k \to x^*$ and conditions *i)-v)* from Theorem 2.1 hold, that is, $c(x^k) > 0$, $\nabla L(x^k, \lambda^k, \mu^k, s^k) \to 0$, $Q_k := Q_k(x^k, \lambda^k, \mu^k, s^k, \theta^k, \eta^k, \nu^k)$ with $\liminf_{k \to +\infty} \lambda_{min}(Q_k) \geq 0$,

$$r_1^k := \sum_{i \in \mathcal{E}} |\lambda_i^k h_i(x^k)| + \sum_{i \in \mathcal{I}} |\mu_i^k g_i(x^k)| + \sum_{i \in \mathcal{S}} s_i^k c_i(x^k) \to 0^+$$

and

$$r_2^k := \sum_{i \in \mathcal{E}} \theta_i^k h_i(x^k)^2 + \sum_{i \in \mathcal{I}} \eta_i^k g_i(x^k)^2 + \sum_{i \in \mathcal{S}} \nu_i^k c_i(x^k)^2 \to 0^+.$$

For each k, let us define $w^k := \nabla f(x^k) - \nabla L(x^k, \lambda^k, \mu^k, s^k)$ and $W^k := \nabla^2 f(x^k) + \frac{1}{k} I$.

Let us take a subsequence such that $\lambda_{min}(Q_k) \geq -\frac{1}{k}$ for all k. Hence, for $\bar{\theta}_k := \sum_{i \in \mathcal{E}} \theta_i^k + \sum_{i \in \mathcal{I}} \eta_i^k + \sum_{i \in \mathcal{S}} \nu_i^k$, we have $(w^k, W^k) \in A(x^k, \lambda^k, \mu^k, s^k, r_1^k, r_2^k, \bar{\theta}_k, 0)$ and in particular $(w^k, W^k) \in K_2^C(x^k, r_1^k, r_2^k)$. From the continuity of ∇f and $\nabla^2 f$, we have $(w^k, W^k) \to (\nabla f(x^*), \nabla^2 f(x^*))$, hence $(\nabla f(x^*), \nabla^2 f(x^*)) \in \limsup_{D \ni (x, r_1, r_2) \to (x^*, 0, 0)} K_2^C(x, r_1, r_2)$, which, by CAKKT2-regularity implies that $(\nabla f(x^*), \nabla^2 f(x^*)) \in K_2^C(x^*, 0, 0)$. By Lemma 4.1, x^* satisfies WSOC.

To prove the reciprocal inclusion, let us take

$$(w, W) \in \limsup_{D \ni (x, r_1, r_2) \to (x^*, 0, 0)} K_2^C(x, r_1, r_2)$$

and let us prove that $(w, W) \in K_2^C(x^*, 0, 0)$. By definition, there are sequences $x^k \to x^*, c(x^k) > 0, r_1^k \to 0^+, r_2^k \to 0^+$ and $w^k \to w$, $W^k \to W$ where W^k is a symmetric $n \times n$ matrix with $(w^k, W^k) \in K_2^C(x^k, r_1^k, r_2^k)$ for all k. Hence, for all k, there are $(\lambda^k, \mu^k, s^k) \in \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}_+^{|\mathcal{S}|}$ and sequences $\{\bar{\theta}_{k,\ell}\} \to +\infty, \{\varepsilon_{k,\ell}\} \to 0^+, \{w^{k,\ell}, W^{k,\ell}\} \to (w^k, W^k)$ as $\ell \to +\infty$ such that $(w^{k,\ell}, W^{k,\ell}) \in A(x^k, \lambda^k, \mu^k, s^k, r_1^k, r_2^k, \bar{\theta}_{k,\ell}, \varepsilon_{k,\ell})$ for all k and all ℓ, associated with some $(\theta^{k,\ell}, \eta^{k,\ell}, \nu^{k,\ell}) \in \mathbb{R}_+^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}_+^{|\mathcal{S}|}$ given in the definition of the set $A(\cdot)$ (16).

Let us fix a subsequence $\ell := \ell_k$ such that $\ell_k \to +\infty$ when $k \to +\infty$. Let us define the objective function $f(x) := w^{\mathtt{T}}(x - x^*) + \frac{1}{2}(x - x^*)^{\mathtt{T}} W (x - x^*)$ and let us show that x^* satisfies CAKKT2 for this function. Note that $\nabla f(x^k) = w + W(x^k - x^*)$ and $\nabla^2 f(x^k) = W$. From the definition of the set $A(\cdot)$ (16) we have:

$$w^{k,\ell_k} + \sum_{i \in \mathcal{E}} \lambda_i^k \nabla h_i(x^k) + \sum_{i \in \mathcal{I}} \mu_i^k \nabla g_i(x^k) - \sum_{i \in \mathcal{S}} s_i^k \nabla c_i(x^k) = 0, \tag{17}$$

$$\begin{aligned} \lambda_{min} \Bigg(W^{k,\ell_k} &+ \sum_{i \in \mathcal{E}} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{i \in \mathcal{I}} \mu_i^k \nabla^2 g_i(x^k) + \sum_{i \in \mathcal{S}} s_i^k \nabla^2 c_i(x^k) + \\ &\sum_{i \in \mathcal{E}} \theta_i^{k,\ell} \nabla h_i(x^k) \nabla h_i(x^k)^{\mathtt{T}} + \sum_{i \in \mathcal{I}} \eta_i^{k,\ell} \nabla g_i(x^k) \nabla g_i(x^k)^{\mathtt{T}} + \\ &\sum_{i \in \mathcal{S}} \nu_i^{k,\ell} \nabla c_i(x^k) \nabla c_i(x^k)^{\mathtt{T}} \Bigg) \geq -\varepsilon_{k,\ell_k}, \end{aligned} \tag{18}$$

and

$$\sum_{i \in \mathcal{E}} |\lambda_i^k h_i(x^k)| + \sum_{i \in \mathcal{I}} |\mu_i^k g_i(x^k)| + \sum_{i \in \mathcal{S}} s_i^k c_i(x^k) \leq r_1^k,$$

$$\sum_{i \in \mathcal{E}} |\lambda_i^k h_i(x^k)^2| + \sum_{i \in \mathcal{I}} \mu_i^k g_i(x^k)^2 + \sum_{i \in \mathcal{S}} s_i^k c_i(x^k)^k \leq r_2^k$$

with $\sum_{i \in \mathcal{E}} \theta_i^{k,\ell} + \sum_{i \in \mathcal{I}} \eta_i^{k,\ell} + \sum_{i \in \mathcal{S}} \nu_i^{k,\ell} \leq \bar{\theta}_{k,\ell}$.

Note that $\nabla L(x^k, \lambda^{k,\ell_k}, \mu^{k,\ell_k}, s^{k,\ell_k}) = \nabla f(x^k) - w^{k,\ell_k} \to 0$ and

$$\begin{aligned} Q_k := \nabla^2 L(x^k, \lambda^k, \mu^k, s^k) &+ \sum_{i \in \mathcal{E}} \theta_i^{k,\ell_k} \nabla h_i(x^k) \nabla h_i(x^k)^{\mathtt{T}} + \\ &\sum_{i \in \mathcal{I}} \eta_i^{k,\ell_k} \nabla g_i(x^k) \nabla g_i(x^k)^{\mathtt{T}} + \sum_{i \in \mathcal{S}} \nu_i^{k,\ell_k} \nabla c_i(x^k) \nabla c_i(x^k)^{\mathtt{T}} \end{aligned}$$

is such that $Q_k - W + W^{k,\ell_k} + \varepsilon_{k,\ell_k} I \succeq 0$. Since $W^{k,\ell_k} \to W$ and $\varepsilon_{k,\ell_k} \to 0^+$ we have $\liminf_{k\to+\infty} \lambda_{min}(Q_k) \geq 0$. The fact that r_1^k and r_2^k go to zero gives the remaining properties in the definition of CAKKT2. The hypothesis implies that x^* satisfies WSOC, which by Lemma 4.1 implies that $(w, W) \in K_2^C(x^*, 0, 0)$ and CAKKT-regularity holds. □

Note that Theorems 2.1 and 4.3 imply that CAKKT2-regularity is a constraint qualification, that is, it can take the place of CQ in the first-order optimality condition (12). It can also take the place of CQ2 in the second-order optimality condition (13). Theorems 4.2 and 4.3 together with the fact that CAKKT2 implies AKKT2 shows that AKKT2-regularity implies CAKKT2-regularity. We note also that Definition 4.3 gives an alternative definition of AKKT2-regularity when $\mathcal{S} := \emptyset$ by simply removing the third condition from the definition of the set $A(\cdot)$ (16).

In comparison with the definition of AKKT-regularity ([7]), CAKKT-regularity ([8]) and AKKT2-regularity ([5]), our definition of CAKKT2-regularity also includes additional constraints $c_i(x) \geq 0, i \in \mathcal{S}$ (satisfying Assumption 2.1, in the non-trivial case). Our Theorem 2.1, in particular, proves that AKKT, CAKKT and AKKT2 sequences can always be build around a local minimizer that satisfies Assumption 2.1 with the additional property $c(x^k) > 0$ for all k. It is straightforward to generalize the corresponding minimal constraint qualifications associated with these optimality conditions by simply replacing, in the definition, the outer semicontinuity of the corresponding cone by the outer semicontinuity over the interior of the constraints in $i \in \mathcal{S}$. Note that constraint qualifications can become stronger by the inclusion of Assumption 2.1, for instance, when $\mathcal{E} := \emptyset$ and $\mathcal{I} := \emptyset$, Assumption 2.1 together with CPLD is equivalent to MFCQ ([29]).

Let us revisit Example 2.1. In this example, the critical subspace at x^* is $S(x^*) = \{(0)\} \times \mathbb{R}$, while the perturbed critical subspace $S(x, x^*) = \{(0, 0)\}$ for $x \neq x^*$. That is, the perturbed critical subspace does not recover the information of the true critical subspace, no matter how close x is from x^*. When the perturbed critical subspace is a trivial set, the second-order information of the problem is neglected by AKKT2. This can be seen, for instance, in the cone given in Definition 4.2. The definition of CAKKT2 tries to avoid this situation, controlling the size of tangent multipliers in order to give a chance for second order information to be considered. In the following example we show that CAKKT2-regularity is strictly weaker than AKKT2-regularity. Let us prove that the constraints from Example 2.1 conforms to CAKKT2-regularity, while AKKT2-regularity fails.

Example 4.1. Let us consider the constraints of Example 2.1.

$$
\begin{aligned}
&h(x_1, x_2) := x_1 = 0,\\
&g_1(x_1, x_2) := x_1^3 \leq 0,\\
&g_2(x_1, x_2) := x_1 e^{x_2} \leq 0,
\end{aligned}
$$

at the feasible point $x^* := (0, 0)$. Let us first compute the cone $K_2^C(x^*, 0, 0)$. This cone is formed by elements $(w^*, W^*) \in \mathbb{R}^2 \times \mathrm{Sym}(2)$ such that there are some $(\lambda^*, \mu_1^*, \mu_2^*) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$ and sequences $\varepsilon^\ell \to 0^+$, $\{(\bar{\theta}_\ell, \theta^\ell, \eta_1^\ell, \eta_2^\ell)\} \subset \mathbb{R}_+^4$, $\{(w^\ell, W^\ell)\} \subset \mathbb{R}^2 \times \mathrm{Sym}(2)$ such that $w^\ell \to w^*$, $W^\ell \to W^*$, $\bar{\theta}_\ell \geq \theta^\ell + \eta_1^\ell + \eta_2^\ell$,

$$
w^\ell + \lambda^* \nabla h(x^*) + \mu_1^* \nabla g_1(x^*) + \mu_2^* \nabla g_2(x^*) = 0, \tag{19}
$$

$$
\begin{aligned}
W^\ell + \lambda^* \nabla^2 h(x^*) + \mu_1^* \nabla^2 g_1(x^*) + \mu_2^* \nabla^2 g_2(x^*) + \theta^\ell \nabla h(x^*) \nabla h(x^*)^\mathtt{T} &+ \\
\nu_1^\ell \nabla g_1(x^*) \nabla g_1(x^*)^\mathtt{T} &+ \\
\nu_2^\ell \nabla g_2(x^*) \nabla g_2(x^*)^\mathtt{T} + \varepsilon_\ell I &\succeq 0.
\end{aligned} \tag{20}
$$

Computing derivatives we have that (19) and (20) can be written as:

$$
w_1^\ell + \lambda^* + \mu_2^* = 0, w_2^\ell = 0, \tag{21}
$$

$$
W^\ell + \mu_2^* \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \theta^\ell \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \nu_2^\ell \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \varepsilon_\ell I \succeq 0. \tag{22}
$$

These sequences can be built if, and only if, $w_2^* = 0$ and $W_{22}^* \geq 0$, which defines the cone $K_2^C(x^*, 0, 0)$. To prove that CAKKT2-regularity holds, let us take $(w^*, W^*) \in \limsup_{(x^k, r_1^k, r_2^k) \to (x^*, 0, 0)} K_2^C(x^k, r_1^k, r_2^k)$ and let us prove that

$(w^*, W^*) \in K_2^C(x^*, 0, 0)$, that is, $w_2^* = 0$ and $W_{22}^* \geq 0$. From the definition, we have that there exist sequences $(x^k, r_1^k, r_2^k) \to (x^*, 0, 0)$, $\{(\lambda^k, \mu_1^k, \mu_2^k)\} \subset \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$ and $(w^k, W^k) \to (w^*, W^*)$ such that

$$(w^k, W^k) \in \limsup_{\bar{\theta} \to +\infty, \varepsilon \to 0^+} A(x^k, \lambda^k, \mu^k, \emptyset, r_1^k, r_2^k, \bar{\theta}, \varepsilon),$$

that is, there exist sequences $\{(w^{k,\ell}, W^{k,\ell})\} \to (w^k, W^k)$ as $\ell \to +\infty$, $\bar{\theta}_{k,\ell} \to +\infty$ as $\ell \to +\infty$ and $\varepsilon_{k,\ell} \to 0^+$ as $\ell \to +\infty$ such that $(w^{k,\ell}, W^{k,\ell}) \in A(x^k, \lambda^k, \mu^k, \emptyset, r_1^k, r_2^k, \bar{\theta}_{k,\ell}, \varepsilon_{k,\ell})$ for all k and ℓ. Computing derivatives, from the definition (16) of the set $A(\cdot)$, for a fixed sequence $\ell := \ell_k$ such that $\ell \to +\infty$ when $k \to +\infty$ we have:

$$W^{k,\ell_k} + \lambda^k(1,0) + \mu_1^k(2(x_1^k)^2, 0) + \mu_2^k(e^{x_2^k}, x_1^k e^{x_2^k}) = (0,0), \tag{23}$$

$$|\lambda^k x_1^k| + |\mu_1^k x_1^3| + |\mu_2^k x_1^k e^{x_2^k}| \leq r_1^k, \tag{24}$$

$$\theta^k (x_1^k)^2 + \eta_1^k (x_1^k)^6 + \eta_2^k (x_1^k)^2 e^{2x_2^k} \leq r_2^k, \tag{25}$$

$$W^{k,\ell_k} + \begin{pmatrix} 6x_1^k\mu_1^k + \theta^k + 9\eta_1^k(x_1^k)^4 + \eta_2^k e^{2x_2^k} & \mu_2^k e^{x_2^k} + \eta_2^k x_1^k e^{2x_2^k} \\ \mu_2^k e^{x_2^k} + \eta_2^k x_1^k e^{2x_2^k} & \mu_2^k x_1^k e^{x_2^k} + \eta_2^k (x_1^k)^2 e^{2x_2^k} \end{pmatrix} + \varepsilon_{k,\ell_k} I \succeq 0, \tag{26}$$

$$\bar{\theta}_{k,\ell_k} \geq \theta^{k,\ell_k} + \eta_1^{k,\ell_k} + \eta_2^{k,\ell_k} \text{ for some } (\theta^{k,\ell_k}, \eta_1^{k,\ell_k}, \eta_2^{k,\ell_k}) \in \mathbb{R}_+^3. \tag{27}$$

From (23) and (24) and from the fact that $w^{k,\ell_k} \to w^* = (w_1^*, w_2^*)$ we have $w_2^* = 0$. From (24) and (25), since $W^{k,\ell_k} \to W^* = \begin{pmatrix} W_{11}^* & W_{12}^* \\ W_{12}^* & W_{22}^* \end{pmatrix}$ and $\varepsilon_{k,\ell_k} \to 0^+$, the element in the second column and second row of the matrix in (26) converges to W_{22}^*. Hence (26) gives $W_{22}^* \geq 0$, which proves CAKKT2-regularity. By a simple calculation, WSOC does not hold at x^* for any Lagrange multiplier, thus, since Example 2.1 exhibited an objective function such that x^* satisfies AKKT2, Theorem 4.2 implies that AKKT2-regularity fails. Note that even if we somewhat include first-order complementarity in the definition of AKKT2 and AKKT2-regularity, we would have equation (24) but not (25), which would not be enough to conclude continuity of the corresponding cone. This shows that the inclusion of second-order complementarity plays a key role in this example.

This example shows that for some objective function $f(\cdot)$, an algorithm may generate an AKKT2 and CAKKT sequence that converges to a point that does not satisfy WSOC as in Example 2.1, hence, not a solution, while whenever it generates a CAKKT2 sequence, it necessarily satisfies WSOC at a limit point.

5 Conclusions

We have developed a new second-order optimality condition CAKKT2 without constraint qualifications by limiting the rate of growth of Lagrange and tangent multipliers. This suggests that stopping criterion associated with the new optimality condition should be included in an algorithm in order to avoid, as much as possible, declaring convergence to a non-optimal solution. Usually, algorithms do not compute tangent multipliers explicitly, but at least for augmented Lagrangian and some interior point methods, the "hidden" tangent multipliers associated is such that CAKKT2 holds. We also developed a new constraint qualification that captures the rate in which Lagrange and tangent multipliers grow in these algorithms. This gives a new global convergence result to a second-order stationary point with a weaker assumption. We believe that the formulation of our optimality condition will help to guide the development of second-order algorithms by providing an adequate way to handle the critical subspace. Also, our optimality condition naturally suggests a way to define a perturbed version of the classical second-order necessary condition, even with simple constraints, which is relevant to the definition of a stationarity concept suitable to the development of iteration complexity bounds and stopping criteria.

Acknowledgments

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

References

[1] A. Anandkumar and R. Ge. Efficient approaches for escaping higher order saddle points in non-convex optimization. *arXiv:1602.05908v1*, 2016.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18:1286–1309, 2008.

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

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

[5] R. Andreani, G. Haeser, A. Ramos, and P.J.S. Silva. A second-order sequential optimality condition associated to the convergence of algorithms. *to appear in IMA Journal of Numerical Analysis*, 2017.

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

[7] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016.

[8] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *to appear in Mathematics of Operations Research*, 2017.

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

[10] R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554, 2010.

[11] R. Behling, G. Haeser, A. Ramos, and D. S. Viana. On a conjecture in second-order optimality conditions. *Optimization-online*, 2016.

[12] D. P. Bertsekas. *Nonlinear programming.* Athenas Scientific, 1999.

[13] W. Bian, X. Chen, and Y. Ye. Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization. *Mathematical Programming*, 149(1):301–327, 2015.

[14] E. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization.* SIAM Publications, 2014.

[15] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and Ph. L. Toint. Evaluation complexity for nonlinear constrained optimization using unscaled kkt conditions and high-order models. *SIAM Journal on Optimization*, 26(2):951–967, 2016.

[16] E. G. Birgin, G. Haeser, and A. Ramos. Augmented lagrangians with constrained subproblems and convergence to second-order stationary points. *Optimization-online*, 2016.

[17] J. Bolte, A. Daniilidis, and A. S. Lewis. The lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17:1205–1223, 2007.

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

[19] C. Cartis, N. I. M. Gould, and Ph. L. Toint. Second-order optimality and beyond: characterization and evaluation complexity in convexly-constrained nonlinear optimization. *Optimization Online*, 2016.

[20] L. Chen and D. Goldfarb. Interior-point $\ell 2$-penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming*, 108(1):1–36, 2006.

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

[22] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Lancelot: A Fortran Package for Large-Scale Nonlinear Optimization (Release A).* Springer, Berlin, 1992.

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

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

[25] F. Facchinei and S. Lucidi. Convergence to second order stationary points in inequality constrained optimization. *Mathematics of Operations Research*, 23(3):746–766, 1998.

[26] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, 1968.

[27] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A stabilized SQP method: global convergence. *IMA Journal of Numerical Analysis*, 37(1):407–443, 2016.

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

[29] G. Haeser. On the global convergence of interior-point nonlinear programming algorithms. *Computational & Applied Mathematics*, 29:125–138, 2010.

[30] G. Haeser. Some theoretical limitations of second-order algorithms for smooth constrained optimization. *Optimization-online*, 2017.

[31] G. Haeser, H. Liu, and Y. Ye. Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary. *Optimization Online*, 2017.

[32] A. F. Izmailov, A. S. Kurennoy, and M. V. Solodov. A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. *Mathematical Programming*, 142(1):591–604, 2013.

[33] S. G. Krantz and H. G. Parks. *A Primer on Real Analytic Functions.* Birkhäuser, 2002.

[34] H. Liu, T. Yao, R. Li, and Y. Ye. Folded concave penalized sparse linear regression: sparsity, statistical performance, and algorithmic theory for local solutions. *available at: https://web.stanford.edu/∼yyye/FoldedPenalty.pdf*, 2016.

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

[36] G. Di Pillo, G. Liuzzi, and S. Lucidi. A primal-dual algorithm for nonlinear programming exploiting negative curvature directions. *Numerical Algebra, Control and Optimization*, 1(3):509–528, 2011.

[37] G. Di Pillo, S. Lucidi, and L. Palagi. Convergence to second-order stationary points of a primal-dual algorithm model for nonlinear programming. *Mathematics of Operations Research*, 30(4):897–915, 2005.

[38] P. Tseng. Convergent infeasible interior-point trust-region methods for constrained minimization. *SIAM Journal on Optimization*, 13(2):432–469, 2002.

[39] Y. Ye. On affine scaling algorithms for nonconvex quadratic programming. *Mathematical Programming*, 56(1):285–300, 1992.

[40] Y. Ye. On the complexity of approximating a KKT point of quadratic programming. *Mathematical Programming*, 80(2):195–211, 1998.