

On optimality conditions for nonlinear conic programming

Roberto Andreani ^{*} Walter Gómez [†] Gabriel Haeser [‡] Leonardo M. Mito [§]
Alberto Ramos [¶]

Submitted on March 3, 2020. Last revised on April 20, 2021.

Abstract

Sequential optimality conditions have played a major role in proving stronger global convergence results of numerical algorithms for nonlinear programming. Several extensions have been described in conic contexts, where many open questions have arisen. In this paper, we present new sequential optimality conditions in the context of a general nonlinear conic framework, which explains and improves several known results for specific cases, such as semidefinite programming, second-order cone programming, and nonlinear programming. In particular, we show that feasible limit points of sequences generated by the Augmented Lagrangian method satisfy the so-called Approximate Gradient Projection optimality condition, and, under an additional smoothness assumption, the so-called Complementary Approximate Karush-Kuhn-Tucker condition. The first result was unknown even for nonlinear programming while the second one was unknown, for instance, for semidefinite programming.

Key words: Nonlinear conic optimization, Optimality conditions, Numerical methods, Constraint qualifications.

1 Introduction

We are interested in the general *nonlinear conic programming* (NCP) problem, which is usually presented in the form:

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\ & \text{subject to} && G(x) \in \mathcal{K}, \end{aligned} \tag{NCP}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $G : \mathbb{R}^n \rightarrow \mathbb{E}$ are continuously differentiable mappings, \mathbb{E} is a finite-dimensional vector space over \mathbb{R} equipped with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \sqrt{\langle x, x \rangle}$ induced by it, and $\mathcal{K} \subseteq \mathbb{E}$ is a nonempty closed convex cone. Let us denote its feasible set by Ω . This is a general class of optimization problems that encompasses, for instance, some well-known particular cases such as *nonlinear programming* (NLP), *nonlinear semidefinite programming* (NLSDP), and *nonlinear second-order cone programming* (NSOCP). It has applications in several areas which include, but are not restricted to, control theory [32], truss design problems and combinatorial optimization [77], portfolio optimization [54], structural optimization [50], and others. For more details, see [14, 77, 78] and references therein. It is worth mentioning that the content of this paper can be straightforwardly extended to a (NCP) with separate equality constraints in the form: $h(x) = 0$ and $G(x) \in \mathcal{K}$, but we stick to (NCP) as is for the sake of simplicity.

Algorithms for solving nonlinear optimization problems are mostly iterative and their convergence theories are usually built around the limit points of their output sequences. However, numerical methods must employ stopping criteria to properly truncate those sequences, which are often based on necessary optimality conditions. Under a *constraint qualification* (CQ), every local minimizer of (NCP) satisfies the classical *Karush-Kuhn-Tucker* (KKT) conditions, but even simple problems, such as minimizing x subject to $x^2 \in \{0\}$, may have minimizers that do not satisfy the KKT conditions. For NLP problems, a “sequential” alternative condition with high practical appeal was proposed in [3] under the name of *Approximate KKT* (AKKT) condition, that holds at local minimizers independently of CQs, similarly to the *Fritz-John* condition [56], but strictly stronger [3, Theorem 2.2].

This project was funded by FAPESP grants 2013/05475-7, 2013/07375-0, 2017/17840-2, 2017/18308-2, and 2018/24293-0; by CNPq grants 301888/2017-5, 404656/2018-8, 303427/2018-3, 438185/2018-8, and 307270/2019-0; and by FONDECYT grant 1170907.

^{*}Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil. Email: andreani@ime.unicamp.br

[†]Department of Mathematical Engineering, Universidad de la Frontera, Temuco, Chile. Email: walter.gomez@ufrontera.cl

[‡]Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil. Email: ghaeser@ime.usp.br

[§]Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil. Email: leokoto@ime.usp.br

[¶]Department of Mathematics, Federal University of Paraná, Curitiba, PR, Brazil. Email: albertoramos@ufpr.br

Roughly speaking, sequential optimality conditions such as AKKT are general characterizations of feasible limit points of an algorithm’s output sequence. It was proved in [12] that these sequential conditions imply KKT under very weak CQs; for instance, strictly weaker than the *Mangasarian-Fromovitz constraint qualification* (MFCQ). These results assemble a simple unified tool for proving global convergence of algorithms, without assuming boundedness of the Lagrange multiplier set at the limit point. Indeed, proving global convergence of an algorithm under such weak CQs reduces to proving that it generates limit points that satisfy a given sequential optimality condition. This was successfully done, for instance, in [1, 20] for *Augmented Lagrangian* methods, in [37] for a *Shifted Primal-Dual Penalty-Barrier* method, in [38, 67] for *Sequential Quadratic Programming* (SQP) methods, in [8, 30, 40, 41] for *Interior Point* methods, and also in [8, 21] for *Inexact Restoration* methods – see also [7, 11, 12] and references therein for more details. Conversely, sequential optimality conditions may suggest adaptations for practical algorithms that ensure a better theoretical performance, for instance, in [41] the authors analysed a sequential optimality condition satisfied by Interior Point methods, which characterized the effects of a certain control over the feasibility of the method on its convergence. In particular, they proved that a specific type of control guarantees that the sequence of approximate Lagrange multipliers of the method is bounded, even in the absence of a constraint qualification. Moreover, in [62], the authors develop a complexity analysis for an algorithm based on a log-barrier function, Newton’s method, and conjugate gradients, that converges to second-order stationary points via sequential optimality conditions. This kind of complexity analysis via sequential optimality conditions can be done even for some special problems such as in [42], where the KKT conditions are not defined at the limit point due to lack of differentiability; then, the AKKT notion serves as a natural optimality condition for it.

Naturally, such idea has been carried over for several other contexts, for example: *Nash equilibrium* problems [29], *mathematical programs with equilibrium constraints* (MPECs) [68, 69], *mathematical programs with complementarity constraints* (MPCCs) [9], *nonlinear vector optimization with conic constraints* [76], the *multiobjective* case [39], *variational problems in Banach spaces* [48], *quasi-equilibrium problems* [27], and several others. The first extension to a conic context is due to [10] followed by [2] for NLSDP and NSOCP, respectively, which gave rise to a more theoretical range of applications of sequential optimality conditions. For instance, there is a recent work that uses sequential optimality conditions (from [10]) to prove that every local minimizer of a general nonconvex NLSDP problem satisfies a second-order condition that depends on a single Lagrange multiplier over the lineality space of the critical cone, without assuming neither nondegeneracy nor strict complementarity (the common assumptions for this kind of analysis) [35]. The sequential framework has also allowed [4] to define and study weaker variants of the nondegeneracy condition, which are designed to aid in proving global convergence of methods that rely on spectral decompositions, or for problems that present some structural sparsity.

This paper aims at expanding the strongest known sequential optimality conditions from NLP to the general conic framework (NCP). The most difficult aspect of such generalizations is dealing with complementarity. For NLP constraints $g(x) \leq 0$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g = (g_1, \dots, g_m)$, and a Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^m, \bar{\lambda} \geq 0$, $\bar{\lambda} = (\bar{\lambda}_1, \dots, \bar{\lambda}_m)$, the complementarity constraint $\langle \bar{\lambda}, g(\bar{x}) \rangle = 0$ means precisely that, for every $i \in \{1, \dots, m\}$, the multiplier $\bar{\lambda}_i$ is complementary with respect to the constraint $g_i(x) \leq 0$, in the sense that $\bar{\lambda}_i g_i(\bar{x}) = 0$ at a feasible point of interest \bar{x} . It turns out that, when considering perturbations of \bar{x} , the latter gives a stronger optimality condition. In the conic case (NCP), it is not clear how to exploit a complementarity-like structure in a statement of the form $\langle \bar{\Lambda}, G(\bar{x}) \rangle = 0$ for a Lagrange multiplier $\bar{\Lambda} \in \mathcal{K}^\circ$ (the polar of the cone \mathcal{K}) where $G(\bar{x}) \in \mathcal{K}$. In the context of NLSDP [10], the eigenvalues are heavily employed to exploit a complementarity-like structure, where one must carefully consider how to order consistently the eigenvalues of $G(\bar{x})$ and $\bar{\Lambda}$. In [2], this approach is extended to so-called symmetric cones, where an eigenvalue structure is still available but a more elegant solution is given by making use of a so-called Jordan product \circ , which is inherent to the cone $\mathcal{K} = \{u \circ u : u \in \mathbb{E}\}$. Note that self-duality of \mathcal{K} plays an important role in defining these optimality conditions [2, 10].

In this paper we propose a much more general and unified approach for defining such conditions. Here, we propose splitting $\langle \bar{\Lambda}, G(\bar{x}) \rangle = 0$, by means of Moreau’s decomposition, into two complementarity-like statements of the form $\langle \bar{\Lambda}, \Pi_{\mathcal{K}} G(\bar{x}) \rangle = 0$ and $\langle \bar{\Lambda}, \Pi_{\mathcal{K}^\circ} G(\bar{x}) \rangle = 0$, where $\Pi_{\mathcal{K}}$ and $\Pi_{\mathcal{K}^\circ}$ denote orthogonal projections onto \mathcal{K} and its polar, respectively. Hence, no particular structure of the cone \mathcal{K} is needed. We then show that a primal dual sequence $\{(x^k, \Lambda^k)\} \subset \mathbb{R}^n \times \mathcal{K}^\circ$ generated by an Augmented Lagrangian method is such that $\langle \Lambda^k, \Pi_{\mathcal{K}} G(x^k) \rangle \rightarrow 0$. In the context of NLP, an optimality condition associated with this measure of complementarity turns out to be equivalent to the so-called Approximate Gradient Projection optimality condition (AGP, [13]), which is strictly stronger than the more common AKKT [3] optimality condition. The revelation of this property of the Augmented Lagrangian sequence is somewhat surprising, and it was achieved as a corollary of our more general approach. Also, under an additional smoothness assumption, the other complementarity-like statement $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ} G(x^k) \rangle \rightarrow 0$ is also satisfied. This answers an open question of [10] in the context of NLSDP by presenting a stronger complementarity-like structure, generated by the Augmented Lagrangian method, which was not achieved in [10]. In [2], although an optimality condition that reveals a strong complementarity-like structure was defined for general symmetric cones, the proof that the sequence generated by the Augmented Lagrangian fulfills this property was only done in the context of NSOCP.

Finally, we show that our global convergence results are strictly stronger than the ones usually employed for conic constraints, namely, where Robinson's CQ is employed. Note that our results do not require that \mathcal{K} has a nonempty interior nor self-duality, hence our results are relevant even when the constraints are linear. Also, since Robinson's CQ may fail, our results imply that even when the set of Lagrange multipliers is unbounded at a feasible limit point of a sequence generated by the algorithm, a global convergence result is available, that is, the dual sequence $\{\Lambda^k\}$ may diverge.

This paper is organized as follows: Section 2 presents some basic definitions and a short literature review on sequential optimality conditions for NLP and NLSDP. In Section 3, we define sequential conditions for NCP and some standard properties are proven. Section 4 presents an Augmented Lagrangian algorithm and its convergence theory in terms of sequential conditions. Section 5 is dedicated to a distinguished extension of AKKT and its relation to the other conditions. Section 6 is focused on contextualizing our conditions when NCP is reduced to NLP, NLSDP, and NSOCP. Section 7 introduces new constraint qualifications that can be useful for the convergence analysis of numerical methods. Lastly, Section 8 is dedicated to summarizing our main contributions, while presenting our prospective work.

2 Preliminaries

In this section, we recall some basic concepts and results of convex analysis and we make a more detailed review of sequential optimality conditions for NLP and NLSDP.

2.1 Notations and convex analysis background

Our notation is standard in optimization and variational analysis: \mathbb{N} denotes the set of natural numbers (with $0 \in \mathbb{N}$) and \mathbb{R}^n stands for the n -dimensional real Euclidean space. Let $x \in \mathbb{R}^n$, we use $B[x, \delta]$ to denote the closed ball with center at x and radius $\delta > 0$. For $a, b \in \mathbb{R}^n$ with components a_i and b_i respectively, we use $\max\{a, b\}$ to represent the vector with components $\max\{a_i, b_i\}$. The vector $\min\{a, b\}$ has a similar meaning. We denote the *interior* of a set A by $\text{int } A$. Moreover, we recall that \mathbb{E} is a finite-dimensional linear space equipped with an inner product, which we denote by $\langle \cdot, \cdot \rangle$.

Given a set-valued mapping $\Gamma : \mathbb{R}^s \rightrightarrows \mathbb{E}$, the sequential (Painlevé-Kuratowski) *outer limit* of $\Gamma(z)$ as $z \rightarrow \bar{z}$, is the set $\{\bar{w} \in \mathbb{E} : \exists (z^k, w^k) \rightarrow (\bar{z}, \bar{w}), w^k \in \Gamma(z^k)\}$, which is denoted by $\limsup_{z \rightarrow \bar{z}} \Gamma(z)$. Moreover, we say that Γ is outer semi continuous at \bar{z} when

$$\limsup_{z \rightarrow \bar{z}} \Gamma(z) \subseteq \Gamma(\bar{z}).$$

For a differentiable mapping $G : \mathbb{R}^n \rightarrow \mathbb{E}$, we use $DG(x)$ to denote the derivative of G at x , and $DG(x)^* : \mathbb{E} \rightarrow \mathbb{R}^n$ to denote the *adjoint* of $DG(x)$, which is characterized by the following property: $\langle DG(x)d, \Lambda \rangle = \langle d, DG(x)^*\Lambda \rangle$, for every $d \in \mathbb{R}^n$, $\Lambda \in \mathbb{E}$. For a differentiable real-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, we use $\nabla f(x)$ to denote the transpose of $Df(x)$, seen as a $1 \times n$ matrix.

Given a closed convex cone $K \subset \mathbb{E}$, the *polar* of K is the set $K^\circ := \{w \in \mathbb{E} : \langle w, k \rangle \leq 0, \forall k \in K\}$ and $(K^\circ)^\circ = K$. The distance of $w \in \mathbb{E}$ to K is defined as $\text{dist}_K(w) := \min\{\|w - v\| : v \in K\}$ and the *orthogonal projection* of w onto K , denoted by $\Pi_K(w)$, is the point where the minimum is attained. Moreover, it can be proved that $\Pi_K(w)$ is nonexpansive, that is,

$$\|\Pi_K(w) - \Pi_K(v)\| \leq \|w - v\|, \forall v, \forall w,$$

so it is a Lipschitz continuous function, and it can also be proved that $\text{dist}_K^2(w)$ is a continuously differentiable function whose derivative is given by (1). For a proof, see [34].

$$D(\text{dist}_K^2)(w) = 2(w - \Pi_K(w)), \forall w. \quad (1)$$

The following lemma (see, e.g., [47, Theorems 3.2.3 and 3.2.5]) encompasses other well-known properties of projections:

Lemma 2.1. *Let $K \subset \mathbb{E}$ be a closed convex cone and $w \in \mathbb{E}$. Then:*

1. $v = \Pi_K(w)$ if, and only if, $v \in K$, $w - v \in K^\circ$, and $\langle w - v, v \rangle = 0$;
2. $\Pi_K(\alpha w) = \alpha \Pi_K(w)$, for every $\alpha \geq 0$, and $\Pi_K(-w) = -\Pi_{-K}(w)$;
3. (Moreau's decomposition) For every $w \in \mathbb{E}$, we have $w = \Pi_K(w) + \Pi_{K^\circ}(w)$ and $\langle \Pi_K(w), \Pi_{K^\circ}(w) \rangle = 0$.

2.2 Sequential optimality conditions for NLP and NLSDP

In order to start a deeper discussion on sequential optimality conditions, we will make a brief exposition of the most important results around them in NLP, where it has been extensively studied, and a summary of some recent advances in NLSDP.

Consider the following NLP problem in standard form:

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\ & \text{subject to} && G(x) \leq 0, \end{aligned} \tag{NLP}$$

which is (NCP) with $\mathbb{E} = \mathbb{R}^m$ and $\mathcal{K} = \mathbb{R}_-^m := \{z \in \mathbb{R}^m : \forall i \in \{1, \dots, m\}, z_i \leq 0\}$. Following [3], we say that the *Approximate KKT* (AKKT) condition holds at a feasible point \bar{x} when there exist sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^m := -\mathbb{R}_-^m$ such that

$$\nabla f(x^k) + \sum_{i=1}^m \Lambda_i^k \nabla G_i(x^k) \rightarrow 0, \tag{2}$$

and $\Lambda_i^k = 0$, whenever $G_i(\bar{x}) < 0$, for sufficiently large k . Note that AKKT allows divergence of the multiplier sequences associated with active constraints. It has been proved that under some constraint qualifications weaker than the *linear independence constraint qualification* (LICQ) and MFCQ, for example, every AKKT point also satisfies the KKT conditions (see [11, 7, 8]). Then, as mentioned in the Introduction, since many algorithms generate AKKT sequences, this improves their convergence theory in a unified manner. Another practical advantage of sequential conditions is their relation to natural choices of stopping criteria for algorithms, for example, it is elementary to verify that AKKT holds at \bar{x} if, and only if, for every $\varepsilon > 0$ there is some $x_\varepsilon \in B[\bar{x}, \varepsilon]$ and some approximate multiplier $\Lambda_\varepsilon \geq 0$ such that

$$\|\max\{0, G(x_\varepsilon)\}\| \leq \varepsilon, \quad \|\nabla f(x_\varepsilon) + \sum_{i=1}^m (\Lambda_\varepsilon)_i \nabla G_i(x_\varepsilon)\| \leq \varepsilon, \quad \|\min\{\Lambda_\varepsilon, -G(x_\varepsilon)\}\| \leq \varepsilon. \tag{3}$$

The properties that made AKKT useful motivate the following general description of a “good” sequential optimality condition, that provides guidelines for defining new ones:

1. It must be a necessary optimality condition independently of the fulfillment of any CQ;
2. There must be meaningful numerical methods that generate sequences whose limit points satisfy it;
3. It must imply optimality conditions in the form “KKT or not-CQ” for very weak CQs.

The third property measures the strength of such sequential optimality condition in comparison with standard ones, while the first one guarantees that no local minimizer will be censured by it. In addition, the second property means that one must be able to employ it to formalize the convergence theory of at least one algorithm. It should be observed that, as long as they satisfy those three properties, the stronger the condition (in the logical implication sense), the better. The ability of strengthening global convergence results is of paramount importance, since otherwise stronger optimality conditions could be derived without resorting to the sequential approach.

For improving the AKKT condition for (NLP), it was proposed in [13] the so-called *Complementary AKKT* (CAKKT) condition, that holds at a feasible point \bar{x} when there are sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^m$ such that (2) holds and

$$\Lambda_i^k G_i(x^k) \rightarrow 0, \forall i \in \{1, \dots, m\}.$$

Indeed, the CAKKT condition is strictly stronger than the AKKT condition, but an additional property (the so-called *generalized Lojasiewicz inequality*) is needed in order to prove that the Augmented Lagrangian algorithm generates CAKKT sequences.

Another interesting sequential condition was introduced in [60] under the name *Approximate Gradient Projection* (AGP), that holds at a feasible point \bar{x} when there exists some sequence $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ such that

$$\|\Pi_{L(\Omega, x^k)}(-\nabla f(x^k))\| \rightarrow 0, \tag{4}$$

where $L(\Omega, x) := \{d \in \mathbb{R}^n : \min\{0, G_i(x)\} + \nabla G_i(x)^T d \leq 0, \text{ for all } i \text{ such that } G_i(\bar{x}) = 0\}$. One of the most highlighted features of AGP is its lack of Lagrange multiplier approximations, using projections instead. This makes it useful for supporting the global convergence of numerical optimization methods where multiplier approximations are not explicitly available; for example, algorithms based on *Inexact Restoration* (IR) procedures. See [58, 59, 60, 33, 28] and references therein for details.

Now, consider the following NLSDP problem:

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\ & \text{subject to} && G(x) \in \mathbb{S}_-^m, \end{aligned} \tag{NLSDP}$$

which is a particular case of (NCP) where $\mathbb{E} = \mathbb{S}^m$ is the linear space of $m \times m$ symmetric matrices, $\mathcal{K} = \mathbb{S}_-^m$ is the cone of $m \times m$ symmetric negative semidefinite matrices. The AKKT extension for (NLSDP) presented in [10, Definition 3.1] holds at a feasible point \bar{x} when there are sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{S}_+^m := -\mathbb{S}_-^m$ such that

$$\nabla f(x^k) + DG(x^k)^* \Lambda^k \rightarrow 0 \tag{5}$$

and $\lambda_i^{U^k}(\Lambda^k) = 0$ whenever $\lambda_i^U(G(\bar{x})) < 0$ and k is sufficiently large, for some sequence of orthogonal matrices $\{U^k\}_{k \in \mathbb{N}} \rightarrow U$, where U^k diagonalizes Λ^k , for each k , and U diagonalizes $G(\bar{x})$. The notation $\lambda_i^U(G(\bar{x}))$ stands for the i -th eigenvalue in the diagonal of $U^T G(\bar{x}) U$, and the same goes for $\lambda_i^{U^k}(\Lambda^k)$. This is done for imbuing the notion of ordering into the eigenvalues of the multipliers and for establishing a proper correspondence with the eigenvalues of $G(\bar{x})$, which makes this extension natural from the NLP context, but very dependent on the structure of \mathbb{S}^m . Still under the same analogy, the most natural extension of CAKKT, discussed in [10], would simply require (5) and

$$\lambda_i^{S^k}(G(x^k)) \lambda_i^{U^k}(\Lambda^k) \rightarrow 0, \tag{6}$$

where $\{S^k\}_{k \in \mathbb{N}} \rightarrow U$ is a sequence of orthogonal matrices that diagonalizes $G(x^k)$, for each k . However, although this is an actual optimality condition, it was not possible at the moment to provide an algorithm capable of generating sequences with these properties, even under generalized Łojasiewicz. Then, instead of using the eigenvalue product, the authors of [10] used the canonical inner product of \mathbb{S}^m (given by the trace of the matrix product) to define a new condition called *Trace AKKT* (TAKKT), that requires (5) and

$$\langle \Lambda^k, G(x^k) \rangle \rightarrow 0.$$

Surprisingly, TAKKT has been proven to be completely independent of AKKT (see [10, Example 5.2] and [2, Example 3.1]) and it also requires the generalized Łojasiewicz inequality to hold for it to be generated by the Augmented Lagrangian algorithm. However, observe that TAKKT can be equivalently stated in NLP using diagonal matrices, and in this context, it is strictly implied by CAKKT.

3 New optimality conditions for nonlinear conic programming

In this section, we propose new sequential optimality conditions for general optimization problems, we prove some of their properties, and we clarify the relations among them.

Before we begin, recall (from [25, 70], for example) that the *Karush-Kuhn-Tucker* (KKT) conditions hold at a feasible point \bar{x} of (NCP) when there exists some *Lagrange multiplier* $\bar{\Lambda} \in \mathcal{K}^\circ$ such that

$$\nabla f(\bar{x}) + DG(\bar{x})^* \bar{\Lambda} = 0, \tag{7}$$

$$\langle \bar{\Lambda}, G(\bar{x}) \rangle = 0. \tag{8}$$

By taking the natural relaxation of (7) and (8), we obtain a trivial extension of the TAKKT condition [10] from NLSDP to NCP, replacing the trace product with an arbitrary inner product:

Definition 3.1 (TAKKT). Let \bar{x} be a feasible point of (NCP). We say that \bar{x} satisfies the TAKKT condition if there exist sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that

$$\nabla f(x^k) + DG(x^k)^* \Lambda^k \rightarrow 0, \tag{9}$$

$$\langle \Lambda^k, G(x^k) \rangle \rightarrow 0. \tag{10}$$

Points that satisfy TAKKT are usually called “TAKKT points” and the sequences associated with them are called “TAKKT sequences”. Similar names hold for the other sequential conditions. At the end of this section, we prove that TAKKT is an optimality condition for (NCP) as well. Before that, note that at a KKT pair $(\bar{x}, \bar{\Lambda}) \in \mathbb{R}^n \times \mathcal{K}^\circ$, we also have

$$\langle \bar{\Lambda}, \Pi_{\mathcal{K}}(G(\bar{x})) \rangle = \langle \bar{\Lambda}, G(\bar{x}) \rangle - \langle \bar{\Lambda}, \Pi_{\mathcal{K}^\circ}(G(\bar{x})) \rangle = \langle \bar{\Lambda}, G(\bar{x}) \rangle = 0,$$

by Moreau’s decomposition, so (8) can be equivalently stated as $\langle \bar{\Lambda}, \Pi_{\mathcal{K}}(G(\bar{x})) \rangle = 0$. Relaxing this alternative expression leads to the following:

Definition 3.2. (AGP) Let \bar{x} be a feasible point of (NCP). We say that \bar{x} satisfies the Approximate Gradient Projection (AGP) condition if there exist sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds and

$$\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \rightarrow 0. \quad (11)$$

A similar optimality condition has appeared in [48, Definition 5.2], where the authors deal with a version of (NCP) over infinite-dimensional *Banach spaces* where \mathcal{K} is contained in a *Hilbert lattice*. However, the authors refer to it as ‘‘Asymptotic KKT (AKKT)’’, which does not make it clear how strong their results are. We point out that AGP might be a more appropriate name, for when Definition 3.2 is reduced to NLP, it is equivalent to the concept with the same name introduced in [60], which is given by (4). What follows is a proof of our claim:

Theorem 3.1. *Consider (NLP), which is (NCP) with $\mathbb{E} = \mathbb{R}^m$ and $\mathcal{K} = \mathbb{R}_+^m$. Let \bar{x} be a feasible point for it. Then, AGP as in Definition 3.2 holds at \bar{x} if, and only if, AGP as in (4) holds at \bar{x} .*

Proof. Let \bar{x} satisfy Definition 3.2. Then, there exist sequences $x^k \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^m$ such that (9) and (11) hold. Now, define $d^k := \Pi_{L(\Omega, x^k)}(-\nabla f(x^k))$, $\forall k \in \mathbb{N}$. By definition, d^k is a solution of

$$\begin{aligned} \text{Minimize}_{d \in \mathbb{R}^n} \quad & \frac{1}{2} \| -\nabla f(x^k) - d \|^2, \\ \text{subject to} \quad & \min\{0, G_i(x^k)\} + DG_i(x^k)d \leq 0, \forall i \in \mathcal{A}(\bar{x}), \end{aligned} \quad (12)$$

where $\mathcal{A}(\bar{x}) := \{i \in \{1, \dots, m\} : G_i(\bar{x}) = 0\}$. Since the constraints are linear, by the KKT conditions, there exists some $\hat{\Lambda}^k \in \mathbb{R}_+^{|\mathcal{A}(\bar{x})|}$ such that the first-order conditions hold for it. Define $\bar{\Lambda}_i^k \in \mathbb{R}_+^m$ such that

$$\bar{\Lambda}_i^k := \begin{cases} \hat{\Lambda}_i^k, & \text{if } i \in \mathcal{A}(\bar{x}), \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$\nabla f(x^k) + d^k + DG(x^k)^* \bar{\Lambda}^k = 0 \text{ and } \langle \bar{\Lambda}^k, \min\{0, G(x^k)\} + DG(x^k)d^k \rangle = 0. \quad (13)$$

Multiplying (13) by d^k , we obtain

$$\begin{aligned} \|d^k\|^2 &= -\langle \nabla f(x^k), d^k \rangle - \langle d^k, DG(x^k)^* \bar{\Lambda}^k \rangle \\ &= -\langle \nabla f(x^k), d^k \rangle - \langle DG(x^k)d^k, \bar{\Lambda}^k \rangle \\ &= -\langle \nabla f(x^k), d^k \rangle - \langle DG(x^k)d^k + \min\{0, G(x^k)\}, \bar{\Lambda}^k \rangle + \\ &\quad + \langle \min\{0, G(x^k)\}, \bar{\Lambda}^k \rangle \\ &\leq -\langle \nabla f(x^k), d^k \rangle \\ &= -\langle \nabla f(x^k) + DG(x^k)^* \Lambda^k, d^k \rangle + \langle DG(x^k)d^k + \min\{0, G(x^k)\}, \Lambda^k \rangle - \\ &\quad - \langle \min\{0, G(x^k)\}, \Lambda^k \rangle \\ &\leq \|\nabla f(x^k) + DG(x^k)^* \Lambda^k\| \|d^k\| + \\ &\quad + \sum_{i \notin \mathcal{A}(\bar{x})} (DG_i(x^k)d^k + \min\{0, G_i(x^k)\}) \Lambda_i^k - \langle \min\{0, G(x^k)\}, \Lambda^k \rangle, \end{aligned}$$

where in the last inequality we used the Cauchy-Schwarz inequality and that d^k is feasible for (12). Moreover, since $\|d^k\| \leq \|\nabla f(x^k)\|$ and (9) and (11) hold, we obtain $d^k \rightarrow 0$ and (4) holds, because $\Lambda_i^k \rightarrow 0$ for every $i \notin \mathcal{A}(\bar{x})$.

Conversely, assume that \bar{x} satisfies AGP as in (4) and set $d^k := \Pi_{L(\Omega, x^k)}(-\nabla f(x^k))$. Analogously, since d^k is a global minimizer of (12), there is an analogous choice of $\bar{\Lambda}^k \in \mathbb{R}_+^m$ such that (13) holds. Then,

$$\langle \bar{\Lambda}^k, \min\{0, G(x^k)\} \rangle = -\langle \bar{\Lambda}^k, DG(x^k)d^k \rangle = -\langle DG(x^k)^* \bar{\Lambda}^k, d^k \rangle = \langle \nabla f(x^k) + d^k, d^k \rangle. \quad (14)$$

By (4), (13) and (14), we obtain $\nabla f(x^k) + DG(x^k)^* \bar{\Lambda}^k = -d^k \rightarrow 0$ and $\langle \bar{\Lambda}^k, \min\{0, G(x^k)\} \rangle \rightarrow 0$. Consequently, Definition 3.2 holds at \bar{x} . \square

Surprisingly, although AGP and TAKKT look like twins, they are completely independent. The following counterexample shows that TAKKT does not imply AGP:

Example 3.1. (TAKKT does not imply AGP) In \mathbb{R}^2 , consider the nonlinear programming problem to minimize $-x_2$ subject to $G(x_1, x_2) \in \mathcal{K}$, where $G(x_1, x_2) := (-x_1, x_1 \exp(x_2))$, $\mathcal{K} := \mathbb{R}_-^2$, and the feasible point $\bar{x} = (0, 1)$. In this case, $\Lambda^k := (\lambda_1^k, \lambda_2^k) \in \mathbb{R}_+^2$ and $\nabla f(x^k) + DG(x^k)^* \Lambda^k \rightarrow 0$ reduces to $-\lambda_1^k + \lambda_2^k \exp(x_2^k) \rightarrow 0$ and $-1 + \lambda_2^k x_1^k \exp(x_2^k) \rightarrow 0$.

TAKKT holds at \bar{x} : Take $x_1^k := 1/k$, $x_2^k := 1$, $\lambda_2^k := (x_1^k \exp(x_2^k))^{-1}$, $\lambda_1^k := \exp(x_2^k) \lambda_2^k$. It is elementary to verify that $\{x^k = (x_1^k, x_2^k)\}$ is a TAKKT sequence.

AGP fails at \bar{x} : Assume that there is an AGP sequence $\{x^k\}$. We observe that the approximate complementarity condition $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \rightarrow 0$ implies that $\lambda_1^k \min\{-x_1^k, 0\} + \lambda_2^k \exp(x_2^k) \min\{x_1^k, 0\} \rightarrow 0$. If there is an AGP sequence with $x_1^k > 0$ for an infinite set of indices, from the complementarity condition we will have that $-x_1^k \lambda_1^k \rightarrow 0$, and thus $x_1^k \lambda_2^k \exp(x_2^k) \rightarrow 0$, which is a contradiction with $-1 + \lambda_2^k x_1^k \exp(x_2^k) \rightarrow 0$. Similar results are obtained if there is an AGP sequence with $x_1^k < 0$ (or $x_1^k = 0$), for an infinite set of indices.

Now, we show that AGP does not imply TAKKT either:

Example 3.2 (AGP does not imply TAKKT). Consider the nonlinear programming problem in \mathbb{R}^2 to minimize x_2 subject to $G(x_1, x_2) := x_2 h(x_1) \in \mathcal{K} = \{0\} \subset \mathbb{R}$, where $h : \mathbb{R} \rightarrow \mathbb{R}$ is the C^1 function introduced in [13], defined as

$$h(z) := \begin{cases} z^4 \sin(z^{-1}) & \text{if } z \neq 0; \\ 0 & \text{if } z = 0. \end{cases} \quad (15)$$

Consider the point $\bar{x} := (0, 1)$. Following [13], we see that there exists a sequence $\{z^k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ such that $z^k \rightarrow 0$, $h'(z^k) = -(z^k)^5$ and $\sin(1/z^k) \rightarrow 1$.

AGP holds at \bar{x} . First, choose a sequence $x^k := (z^k, 1)$ with $\Lambda := -(z^k)^{-4} \in \mathcal{K}^\circ = \mathbb{R}$. Now, observe that $\nabla f(x^k) + DG(x^k)\Lambda^k$ goes to zero, because

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Lambda^k \begin{pmatrix} h'(z^k) \\ h(z^k) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{-1}{(z^k)^4} \begin{pmatrix} -(z^k)^5 \\ (z^k)^4 \sin(1/z^k) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} z^k \\ -\sin(1/z^k) \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (16)$$

Finally, the approximate complementarity condition trivially holds, since $\Pi_{\mathcal{K}}(G(x^k)) = 0$.

TAKKT fails at \bar{x} . Suppose that there exists a sequence $\{x^k := (x_1^k, x_2^k)\}_{k \in \mathbb{N}}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ conforming to the definition of TAKKT. Then,

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Lambda^k \begin{pmatrix} x_2^k h'(x_1^k) \\ h(x_1^k) \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (17)$$

The approximate complementarity condition of TAKKT implies $\Lambda^k G(x_1^k, x_2^k) = \Lambda^k x_2^k h(x_1^k) \rightarrow 0$. Since $x_2^k \rightarrow 1$, we get $\Lambda^k h(x_1^k) \rightarrow 0$, which is a contradiction with (17).

Through Definition 3.2 and Theorem 3.1, it is possible to see AGP as an incomplete CAKKT condition in (NLP), which is a different interpretation from [60]. Indeed, note that AGP holds at \bar{x} with sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^m$ if, and only if, $\Lambda_i^k G_i(x^k) \rightarrow 0$, whenever $G_i(x^k) \leq 0$. Since both AGP and CAKKT push Λ_i^k to zero when $G_i(\bar{x}) < 0$, they only differ when $G_i(\bar{x}) = 0$. Even though the CAKKT condition allows divergence of Λ_i^k in this case, it demands it to go to infinity slower than $G_i(x^k)$ goes to zero, while AGP may allow a faster growth as long as $\{x^k\}_{k \in \mathbb{N}}$ violates $G_i(x^k) \leq 0$. From this point of view, CAKKT improves AGP by introducing some control in the behavior of the multiplier sequences associated with the infeasible part of the constraints, using a quantitative measure of such infeasibility. Generalizing this reasoning, we obtain:

Definition 3.3. (CAKKT) Let \bar{x} be a feasible point. We say that \bar{x} satisfies the Complementary Approximate Karush-Kuhn-Tucker (CAKKT) condition if there exist sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) and (11) hold, and

$$\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \rightarrow 0. \quad (18)$$

Observe that CAKKT as in Definition 3.3 is indeed a generalization of the condition with the same name from [13] for (NLP), since $\Lambda_i^k G_i(x^k) \rightarrow 0$ in this case, independently on the sign of $G_i(x^k)$. Moreover, note that CAKKT is essentially AGP upgraded with (18). Therefore, in view of Moreau's Decomposition, CAKKT clearly implies both, AGP and TAKKT. But since they are independent, the implications are strict. We proceed by showing that CAKKT is a genuine necessary optimality condition, that is, a property that must be satisfied by every local minimizer, even the ones that do not satisfy any constraint qualification.

Theorem 3.2. If \bar{x} is a local minimizer of (NCP), then \bar{x} satisfies the CAKKT condition.

Proof. Let \bar{x} be a local minimizer of (NCP) in $B[\bar{x}, \delta]$, for some $\delta > 0$. Then, \bar{x} is a global minimizer of

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x) + \frac{1}{2} \|x - \bar{x}\|^2, \\ & \text{subject to} && G(x) \in \mathcal{K} \\ & && \|x - \bar{x}\| \leq \delta. \end{aligned}$$

Let $\{\rho_k\}_{k \in \mathbb{N}} \rightarrow \infty$ and, for each $k \in \mathbb{N}$, consider the penalized optimization problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & f(x) + \frac{1}{2} \|x - \bar{x}\|^2 + \rho_k \frac{1}{2} \|\Pi_{\mathcal{K}^\circ}(G(x))\|^2, \\ \text{subject to} \quad & \|x - \bar{x}\| \leq \delta. \end{aligned} \tag{19}$$

Denote by x^k a global solution of (19). Using analogous arguments as in the standard external penalty algorithm convergence proof [61], we see that $x^k \rightarrow \bar{x}$, and thus $\|x^k - \bar{x}\| < \delta$ for k large enough. Using Fermat's rule applied to (19), we have

$$\nabla f(x^k) + (x^k - \bar{x}) + DG(x^k)^* \Lambda^k = 0, \text{ where } \Lambda^k := \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k)) \in \mathcal{K}^\circ,$$

and the expression for the derivative follows from (1), along with the definition of orthogonal projection, and Moreau's decomposition. Thus, (9) holds. Furthermore, by the definition of Λ^k and Lemma 2.1, we see that $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle = \langle \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k)), \Pi_{\mathcal{K}}(G(x^k)) \rangle = 0$, that is, (11) holds. We proceed to show that (18) is satisfied. First, from the optimality of x^k we have

$$f(x^k) + \frac{1}{2} \|x^k - \bar{x}\|^2 + \frac{1}{2} \langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle = f(x^k) + \frac{1}{2} \|x^k - \bar{x}\|^2 + \rho_k \frac{1}{2} \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \leq f(\bar{x}),$$

which leads to

$$0 \leq \langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle = \rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \leq 2(f(\bar{x}) - f(x^k)) - \|x^k - \bar{x}\|^2,$$

and since $x^k \rightarrow \bar{x}$ and $f(x^k) \rightarrow f(\bar{x})$, we see that $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \rightarrow 0$. Thus, CAKKT holds at \bar{x} . \square

Consequently, from our previous discussion:

Corollary 3.3. *If \bar{x} is a local minimizer of (NCP), then \bar{x} satisfies AGP and TAKKT.*

In the next section, we propose an Augmented Lagrangian algorithm for NCP based on projections onto \mathcal{K}° and we build its convergence theory using the new conditions.

4 An Augmented Lagrangian algorithm

Employing Augmented Lagrangian methods to find the solution of optimization problems is a very successful technique for solving finite-dimensional problems, and it is described in several textbooks on continuous optimization, for example, [18, 19, 20, 61, 74] to cite a few of them. In this section, we will show that a variant of the *Powell-Hestenes-Rockafellar* [65, 46, 72] Augmented Lagrangian algorithm generates AGP sequences without any additional condition, and also CAKKT (and TAKKT) sequences under the so-called *generalized Lojasiewicz inequality* (see (28) for the definition). The Augmented Lagrangian variant that we consider is a direct generalization of the one considered in [10, 13] and in the book [20]; the main contrasting feature of it is that it adopts a step control strategy (Step 3) and a *safeguarding* strategy (Step 2). The effects of these modifications over the "quality of convergence" of the method are discussed, at least in the particular case of NLP, in [5, 49].

Given $\rho > 0$, let $L_\rho : \mathbb{R}^n \times \mathcal{K}^\circ \rightarrow \mathbb{R}$ be the *Augmented Lagrangian function* of (NCP), defined as

$$L_\rho(x, \Lambda) := f(x) + \frac{\rho}{2} \left[\left\| \Pi_{\mathcal{K}^\circ} \left(G(x) + \frac{\Lambda}{\rho} \right) \right\|^2 - \left\| \frac{\Lambda}{\rho} \right\|^2 \right],$$

whose partial derivative with respect to x is given by

$$\nabla_x L_\rho(x, \Lambda) = \nabla f(x) + DG(x)^* \left(\rho \Pi_{\mathcal{K}^\circ} \left(G(x) + \frac{\Lambda}{\rho} \right) \right). \tag{20}$$

The expression of the derivative in (20) is what motivates the particular choice of Lagrange multiplier update in the following algorithm:

Algorithm 1 General framework: Augmented Lagrangian

Inputs: A sequence $\{\varepsilon_k\}_{k \in \mathbb{N}}$ of positive scalars such that $\varepsilon_k \rightarrow 0$; a nonempty convex compact set $\mathcal{B} \subset \mathcal{K}^\circ$; real parameters $\tau > 1$, $\sigma \in (0, 1)$, and $\rho_0 > 0$; and initial points $(x^{-1}, \widehat{\Lambda}^0) \in \mathbb{R}^n \times \mathcal{B}$. Also, define $\|V^{-1}\| = \infty$.

For every $k \in \mathbb{N}$:

1. Compute some point x^k such that

$$\|\nabla_x L_{\rho_k}(x^k, \widehat{\Lambda}^k)\| \leq \varepsilon_k; \quad (21)$$

2. Update the multiplier

$$\Lambda^k := \rho_k \Pi_{\mathcal{K}^\circ} \left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k} \right), \quad (22)$$

and compute some $\widehat{\Lambda}^{k+1} \in \mathcal{B}$ (typically, the projection of Λ^k onto \mathcal{B});

3. Define

$$V^k := \frac{\widehat{\Lambda}^k}{\rho_k} - \Pi_{\mathcal{K}^\circ} \left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k} \right); \quad (23)$$

4. If $\|V^k\| \leq \sigma \|V^{k-1}\|$, set $\rho_{k+1} := \rho_k$. Otherwise, choose some $\rho_{k+1} \geq \tau \rho_k$.
-

Note that Step 4 implies that either $\rho^k \rightarrow \infty$ or there is some $k_0 \in \mathbb{N}$ such that $\rho_k = \rho_{k_0}$, for every $k > k_0$. In the latter case, it holds also $\|V^k\| \rightarrow 0$. With this in mind, we proceed by showing that Algorithm 1 generates sequences whose limit points satisfy AGP.

Theorem 4.1. *Let \bar{x} be a feasible limit point of a sequence $\{x^k\}_{k \in \mathbb{N}}$ generated by the Augmented Lagrangian method. Then, \bar{x} satisfies AGP.*

Proof. First, from (21) together with (22) and (20), we get

$$\|\nabla f(x^k) + DG(x^k)^* \Lambda^k\| \leq \varepsilon_k, \quad \text{with } \Lambda^k = \rho_k \Pi_{\mathcal{K}^\circ} (G(x^k) + \rho_k^{-1} \widehat{\Lambda}^k), \quad (24)$$

which implies (9) and $\Lambda^k \in \mathcal{K}^\circ$. Taking a subsequence if necessary, we can suppose that $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$. We consider two cases depending on whether the sequence $\{\rho_k\}_{k \in \mathbb{N}}$ is bounded or not:

1. Suppose that $\rho_k \rightarrow \infty$. By (24) and Lemma 2.1 item 3, we have $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k) + \rho_k^{-1} \widehat{\Lambda}^k) \rangle = 0$ for every $k \in \mathbb{N}$, which yields

$$\begin{aligned} |\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle| &= |\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) - \Pi_{\mathcal{K}}(G(x^k) + \rho_k^{-1} \widehat{\Lambda}^k) \rangle| \\ &\leq \|\Lambda^k\| \|\Pi_{\mathcal{K}}(G(x^k)) - \Pi_{\mathcal{K}}(G(x^k) + \rho_k^{-1} \widehat{\Lambda}^k)\| \\ &\leq \|\Lambda^k\| \|\rho_k^{-1} \widehat{\Lambda}^k\|, \end{aligned}$$

where the inequalities follow from the Cauchy-Schwarz inequality the nonexpansiveness of the projection, respectively. But note that $\|\rho_k^{-1} \Lambda^k\| \|\widehat{\Lambda}^k\| = \|\Pi_{\mathcal{K}^\circ}(G(x^k) + \rho_k^{-1} \widehat{\Lambda}^k)\| \|\widehat{\Lambda}^k\|$ converges to zero by the continuity of the projection, so $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \rightarrow 0$ and AGP holds at \bar{x} .

2. If $\{\rho_k\}_{k \in \mathbb{N}}$ is a bounded sequence, it must be constant for sufficiently large k . Note that

$$\begin{aligned} \langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle &= \rho_k^{-1} \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k)) \rangle \\ &= \rho_k^{-1} \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \widehat{\Lambda}^k) \rangle, \end{aligned} \quad (25)$$

where in the second equality, we use $\langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k) + \widehat{\Lambda}^k) \rangle = 0$. It remains to show that the right-hand side of (25) goes to zero. In fact, since $\Lambda^k = \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k)$ is a bounded sequence, we only need to prove that $\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \widehat{\Lambda}^k)$ converges to zero. Indeed, the reasoning is as follows:

By Step 4 of Algorithm 1, we see that V^k converges to zero, and hence $\rho_k V^k \rightarrow 0$. Using the definition of V^k we get that

$$\Lambda^k - \widehat{\Lambda}^k = \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \widehat{\Lambda}^k = -\rho_k V^k \rightarrow 0,$$

and thus $\Lambda^k = \widehat{\Lambda}^k - \rho_k V^k$ characterizes a bounded sequence. To show that $\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \widehat{\Lambda}^k)$ converges to zero, consider the next expression

$$\begin{aligned} \rho_k G(x^k) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \widehat{\Lambda}^k) &= \rho_k G(x^k) + \widehat{\Lambda}^k - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \widehat{\Lambda}^k \\ &= \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \widehat{\Lambda}^k = \Lambda^k - \widehat{\Lambda}^k \rightarrow 0. \end{aligned} \quad (26)$$

Using the above expression and the nonexpansiveness of the projection onto \mathcal{K} , we get that

$$\|\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \widehat{\Lambda}^k)\| \leq \|\rho_k G(x^k) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \widehat{\Lambda}^k)\| \rightarrow 0. \quad (27)$$

Thus, we obtain that \bar{x} is an AGP point associated with $\{x^k\}_{k \in \mathbb{N}}$. \square

Remark 1. *Inexact Restoration* (IR) algorithms are well-known methods for solving NLP problems (see [23, 58, 59] for details). The philosophy behind them consists of dealing with feasibility and optimality in different stages. Hence, IR methods fit well in difficult problems whose structure allows the implementation of an efficient feasibility restoration procedure [28]. The AGP condition plays a pivotal role in obtaining global convergence results for IR methods [21], but its applicability beyond that class of algorithms was still unclear. Theorem 4.1 solves this issue by showing that the convergence theory of the Augmented Lagrangian is also supported by AGP, which is not an obvious result. From this point of view, Theorems 3.1 and 4.1 show that IR methods generate solution candidates at least as good as Augmented Lagrangian methods for NLP.

Next, we show that Algorithm 1 generates CAKKT sequences under an additional condition called *generalized Lojasiewicz inequality*, that is satisfied by a point \bar{x} and a function Ψ when there exist some $\delta > 0$ and a continuous function $\psi(x) : B(\bar{x}, \delta) \subset \mathbb{R}^n \rightarrow \mathbb{R}$ such that $\psi(x) \rightarrow 0$ when $x \rightarrow \bar{x}$, and

$$|\Psi(x) - \Psi(\bar{x})| \leq \psi(x) \|D\Psi(x)\| \text{ for every } x \in B(\bar{x}, \delta). \quad (28)$$

This property coincides with the inequality with the same name that was proposed in [13]. Such types of property have been extensively used in optimization methods, complexity theory, stability of gradient systems etc. See, for instance, [15, 16, 24, 31, 52, 53, 55] and references therein. For instance, all analytic functions satisfy it, and so does every function that satisfies the classical Lojasiewicz inequality [55]. Now, we may resume our results:

Theorem 4.2. *Let \bar{x} be a feasible limit point of a sequence $\{x^k\}_{k \in \mathbb{N}}$ generated by the Augmented Lagrangian Algorithm. If \bar{x} satisfies (28) for $\Psi(x) = (1/2) \|\Pi_{\mathcal{K}^\circ}(G(x))\|^2$, then \bar{x} satisfies CAKKT.*

Proof. For the sake of simplicity, we can suppose that $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$. By Theorem 4.1, $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle$ converges to zero. Thus, it suffices to show that $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle$ converges to zero as well. Similarly to the proof of the previous theorem, we split this proof into two cases, depending on whether $\{\rho_k\}_{k \in \mathbb{N}}$ is a bounded sequence or not:

1. Suppose that $\{\rho_k\}_{k \in \mathbb{N}}$ is unbounded. We start by showing that $\rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \rightarrow 0$.

From the generalized Lojasiewicz inequality, there exists some function ψ such that

$$\frac{1}{2} \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 = |\Psi(x^k)| \leq \psi(x^k) \|D\Psi(x^k)\| = \psi(x^k) \|DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\|,$$

and we obtain the following inequality

$$\rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \leq 2\psi(x^k) \|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\|.$$

Now, we proceed by finding an upper bound for the sequence $\|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\|$:

$$\begin{aligned} \|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\| &\leq \|DG(x^k)^* (\Lambda^k - \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k)))\| + \|DG(x^k)^* \Lambda^k\| \\ &\leq \|DG(x^k)^*\| \|\Lambda^k - \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k))\| + \|DG(x^k)^* \Lambda^k\| \\ &\leq \|DG(x^k)^*\| \|\Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k))\| + \|DG(x^k)^* \Lambda^k\|. \end{aligned}$$

Furthermore, from $\|\Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k))\| \leq \|\widehat{\Lambda}^k\|$, we see that

$$\begin{aligned} \|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\| &\leq \|DG(x^k)^*\| \|\widehat{\Lambda}^k\| + \|DG(x^k)^* \Lambda^k\| \\ &\leq \|DG(x^k)^*\| \|\widehat{\Lambda}^k\| + \|\nabla f(x^k)\| + \varepsilon_k, \end{aligned} \quad (29)$$

where in the second inequality, we use that $\|\nabla f(x^k) + DG(x^k)^* \Lambda^k\| \leq \varepsilon_k$, (Step 1 of Algorithm 1). Thus, (29) is bounded by some scalar $M > 0$, due to the continuity of DG and ∇f near \bar{x} . Thus,

$\rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \leq 2\psi(x^k)M$. Using the fact that $\psi(x^k) \rightarrow 0$, we get that $\rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \rightarrow 0$. We proceed by computing $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle$. Indeed,

$$\begin{aligned} \langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle &= \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k), \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \\ &= \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)), \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle + \\ &\quad + \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)), \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \\ &= \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)), \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle + \\ &\quad + \rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2. \end{aligned} \tag{30}$$

Since $\rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \rightarrow 0$, we only need to show that the first expression of (30) goes to zero. Now, since $\Pi_{\mathcal{K}^\circ}(G(x^k)) \rightarrow \Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$ and from the boundedness of $\{\widehat{\Lambda}^k\}$, we get that:

$$|\langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)), \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle| \leq \|\widehat{\Lambda}^k\| \|\Pi_{\mathcal{K}^\circ}(G(x^k))\| \rightarrow 0.$$

Thus $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \rightarrow 0$, and as a consequence CAKKT holds at \bar{x} .

2. Suppose that $\{\rho_k\}_{k \in \mathbb{N}}$ is a bounded sequence. By the proof of Theorem 4.1, we see that $\{\Lambda^k\}_{k \in \mathbb{N}}$ is a bounded sequence, and hence $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle$ goes to zero, since $\Pi_{\mathcal{K}^\circ}(G(x^k)) \rightarrow \Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$.

In both cases, we have shown that \bar{x} is a CAKKT point associated with $\{x^k\}_{k \in \mathbb{N}}$. \square

The Augmented Lagrangian method presented in [10, Algorithm 1] for NLSDP was proven to generate TAKKT sequences under generalized Lojasiewicz [10, Theorem 5.2]. Hence, Theorem 4.2 improves this result not only in terms of generality, but also in terms of refinement of the convergence theory. If (28) is not satisfied, then Algorithm 1 may generate sequences that do not satisfy CAKKT, as it is shown by the counterexample after [13, Theorem 5.1] (note that Algorithm 1 is a direct generalization of the augmented Lagrangian presented in [13]).

For completing the convergence theory of Algorithm 1, it is necessary to know how likely it is to reach feasible points. In fact, even though one can not guarantee that every limit point of the sequence $\{x^k\}_{k \in \mathbb{N}}$ generated by Algorithm 1 will be always feasible, at least it is possible to prove that it has the tendency of finding feasible points, in the following sense:

Proposition 4.3. *Every limit point \bar{x} of a sequence $\{x^k\}_{k \in \mathbb{N}}$ generated by Algorithm 1 is a stationary point of*

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad \|\Pi_{\mathcal{K}^\circ}(G(x))\|^2. \tag{31}$$

Proof. Taking a subsequence if necessary, suppose that $x^k \rightarrow \bar{x}$. If $\{\rho^k\}_{k \in \mathbb{N}}$ is bounded, it must converge to some $\bar{\rho}$ and then $\widehat{\Lambda}^k \rightarrow \widehat{\Lambda}$. Also, in this case $V^k \rightarrow 0$, which means $\widehat{\Lambda} = \Pi_{\mathcal{K}^\circ}(\widehat{\Lambda} + \bar{\rho}G(\bar{x}))$. Then, by Lemma 2.1, item 1, we get $\bar{\rho}G(\bar{x}) \in \mathcal{K}$, so $\Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$ and \bar{x} is a global solution of (31). On the other hand, if $\{\rho^k\}_{k \in \mathbb{N}}$ is unbounded, by Step 1 and (20), note that

$$\frac{1}{\rho^k} (\nabla f(x^k) + DG(x^k)^* \Lambda^k) = \frac{\nabla f(x^k)}{\rho^k} + DG(x^k)^* \Pi_{\mathcal{K}^\circ} \left(\frac{\widehat{\Lambda}^k}{\rho^k} + G(x^k) \right) \rightarrow 0.$$

Then, since $\nabla f(x^k) \rightarrow \nabla f(\bar{x})$ and $\{\widehat{\Lambda}^k\}_{k \in \mathbb{N}}$ is bounded, we obtain $DG(\bar{x})^* \Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$, which means \bar{x} is a stationary point of (31). \square

Remark 2. We highlight that our convergence theory for Algorithm 1 allows the set \mathcal{K} to have empty interior and it does not demand self-duality. For instance, it can be applied to optimization problems involving the classical set of *Euclidean Distance Matrices* (EDM) of dimension m , which is defined as

$$\mathcal{E}^m := \{M \in \mathbb{S}^m : \exists p_1, \dots, p_m \in \mathbb{R}^r, \forall i, j \in \{1, \dots, m\}, M_{ij} = \|p_i - p_j\|_2^2\}.$$

It is a closed convex cone because it can be seen as the image of \mathbb{S}_+^m through the linear operator

$$\mathcal{T}(Y) := \text{diag}(Y)e^T + e \text{diag}(Y)^T - 2Y,$$

where e is a vector of ones and $\text{diag}(M) := (M_{11}, \dots, M_{mm})$. Also, \mathcal{E}^m is not self dual. Since every $M \in \mathcal{E}^m$ is *hollow*, that is, $\text{diag}(M) = 0$, the EDM cone has empty interior. In this case it is possible to build a general convergence theory of algorithms over the EDM cone via sequential conditions, even when Robinson's CQ does not hold. In particular, algorithms based on projections onto \mathcal{E}^m or its polar, such as the Augmented Lagrangian method we propose, can benefit from the recent advances towards efficient numerical methods to compute projections while keeping a fixed *embedding dimension* [66]. More details about Euclidean Distance Matrices can be found in [51].

5 The AKKT condition

Until this point, we have presented generalizations of CAKKT and AGP via projections onto \mathcal{K} and \mathcal{K}° , but we have not addressed yet a generalization of AKKT, which is the most natural and simple condition in NLP. However, in Sections 3 and 4, everything was built starting from TAKKT instead of AKKT. Historically, AKKT was born in NLP as a natural way of representing limit points of sequences generated by algorithms and studying their properties. But, in NCP, we will present it arising from a much more theoretical field, which is the theory of perturbations in optimization problems.

Let us recall the KKT conditions at a point $\bar{x} \in \mathbb{R}^n$ with a multiplier $\bar{\Lambda} \in \mathcal{K}^\circ$ in the form of a generalized equation (in the sense of Robinson [71]):

$$\underbrace{\begin{pmatrix} \nabla f(\bar{x}) + DG(\bar{x})^* \bar{\Lambda} \\ G(\bar{x}) \end{pmatrix}}_{\mathcal{F}(\bar{x}, \bar{\Lambda})} \in \underbrace{\begin{bmatrix} \{0\} \\ \{Y \in \mathcal{K} : \langle Y, \bar{\Lambda} \rangle = 0\} \end{bmatrix}}_{\mathcal{N}(\bar{\Lambda})}.$$

Given some $\varepsilon > 0$, the standard perturbation theory (see, for example, [25, 44, 45]) can be used as inspiration to say that a point $x \in B(\bar{x}, \varepsilon)$ satisfies the KKT conditions with error ε when there is a multiplier $\Lambda \in \mathcal{K}^\circ$ and some perturbation vector $\xi \in \mathbb{R}^n \times \mathbb{E}$ such that $\mathcal{F}(x, \Lambda) + \xi \in \mathcal{N}(\Lambda)$ and $\|\xi\| \leq \varepsilon$. This strongly suggests a sequential optimality condition:

Definition 5.1. (AKKT) A feasible point \bar{x} satisfies the Approximate KKT (AKKT) condition when there exist sequences $\{y^k\}_{k \in \mathbb{N}} \rightarrow 0$, $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds, $G(x^k) + y^k \in \mathcal{K}$, and

$$\langle \Lambda^k, G(x^k) + y^k \rangle = 0, \forall k \in \mathbb{N}. \quad (32)$$

It turns out that Definition 5.1 coincides with [67, Definition 2.5], which to the best of our knowledge, employed for the first time perturbed KKT ideas to improve global convergence results of algorithms (for instance, *Sequential Quadratic Programming* (SQP) methods). At the time, the authors did not prove it was an optimality condition. Note that AKKT as in Definition 5.1 is distinguished for not directly relying on projections, eigenvalues or other similar objects, but giving some degree of freedom to the approximation instead, what makes it much more versatile and simple than the others. On the other hand, since the perturbation is inside the inner product, AKKT has a more solid structure to work with, when compared to the others. Also, when Definition 5.1 is specialized to the NLP, the NLSQP or the NSOCP contexts, it is consistent with the existing concepts with the same name from [3, Section 2], [10, Definition 3.1], and [2, Definition 3.3], respectively. This is clarified in Subsection 6.1.

For relating AKKT with the other conditions in NCP, we begin by proving that AGP implies AKKT:

Proposition 5.1. If \bar{x} satisfies the AGP condition, then it must also satisfy AKKT.

Proof. If \bar{x} satisfies AGP, then there are sequences $\{\tilde{\Lambda}^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ and $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ satisfying (9) and (11). Denote by y^k the global solution of

$$\begin{aligned} & \underset{y \in \mathbb{E}}{\text{Minimize}} && \frac{1}{2} \|y\|^2 - \langle \tilde{\Lambda}^k, G(x^k) + y \rangle, \\ & \text{subject to} && G(x^k) + y \in \mathcal{K} \end{aligned} \quad (33)$$

and consider the feasible point $y := -\Pi_{\mathcal{K}^\circ}(G(x^k))$ of (33). Then,

$$\begin{aligned} (1/2) \|y^k\|^2 - \langle \tilde{\Lambda}^k, G(x^k) + y^k \rangle & \leq (1/2) \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 - \langle \tilde{\Lambda}^k, G(x^k) - \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \\ & = (1/2) \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 - \langle \tilde{\Lambda}^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle. \end{aligned} \quad (34)$$

Taking $k \rightarrow \infty$ in (34), since $\Pi_{\mathcal{K}^\circ}(G(x^k)) \rightarrow 0$ and AGP holds, we see that $y^k \rightarrow 0$ and $\langle \tilde{\Lambda}^k, G(x^k) + y^k \rangle \rightarrow 0$, because $\tilde{\Lambda}^k \in \mathcal{K}^\circ$ and $G(x^k) + y^k \in \mathcal{K}$. Now, the necessary optimality condition for y^k in (33) implies

$$0 \in y^k - \tilde{\Lambda}^k + \{\Theta \in \mathcal{K}^\circ : \langle \Theta, G(x^k) + y^k \rangle = 0\}$$

Now, choosing $\Lambda^k := \tilde{\Lambda}^k - y^k$ for all $k \in \mathbb{N}$, we obtain that AKKT holds at \bar{x} with the sequences $\{x^k\}_{k \in \mathbb{N}}$, $\{\Lambda^k\}_{k \in \mathbb{N}}$, and $\{y^k\}_{k \in \mathbb{N}}$. \square

Though, the converse is not necessarily true, due to [3, Counterexample 3.1] that shows that AKKT does not imply AGP in NLP (and Theorem 3.1). Also, [10, Example 5.2] and [2, Example 3.1] show that TAKKT and AKKT do not imply each other in NLP, hence the same is valid for NCP.

Since AKKT is a weak condition, in comparison with AGP and CAKKT, we expect it to be more commonly generated by algorithms designed for solving (NCP). An interesting fact is that the vector y^k of Definition 5.1

formalizes the idea of seeing joint feasibility-complementarity measures, which are commonly used in algorithms, as certificates of approximate optimality along with the Lagrangian residue. For example, it is easy to verify that Algorithm 1 generates sequences $\{x^k\}_{k \in \mathbb{N}}$ and $\{\Lambda^k\}_{k \in \mathbb{N}}$ such that every feasible limit point \bar{x} of $\{x^k\}_{k \in \mathbb{N}}$ satisfy AKKT. In this case, without loss of generality, we can assume $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$. Then AKKT can be verified by taking the sequence $\{y^k\}_{k \in \mathbb{N}}$ such that $y^k = V^k$ for every k ; then $G(x^k) + y^k \in \mathcal{K}$ and

$$\langle \Lambda^k, G(x^k) + y^k \rangle = \rho_k \left\langle \Pi_{\mathcal{K}^\circ} \left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k} \right), \left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k} \right) - \Pi_{\mathcal{K}^\circ} \left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k} \right) \right\rangle = 0,$$

by Lemma 2.1. Condition (9) follows directly from Step 1, since $\nabla_x L_{\rho^k}(x^k, \Lambda^k) = \nabla f(x^k) + DG(x^k)^* \Lambda^k$ for the multiplier choice of Step 2. Also, if $\{\rho_k\}_{k \in \mathbb{N}}$ is unbounded, it follows that $y^k \rightarrow -\Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$; but if it is bounded, then $\|y^k\| \rightarrow 0$ due to Step 4. Thus, it can be said that AKKT is the most simple and natural sequential condition amongst the ones we presented when viewed from the algorithmic and the theoretical perspective.

When \mathcal{K} is a product of closed convex cones, say $\mathcal{K} = \mathcal{K}_1 \times \dots \times \mathcal{K}_r$ and $\mathbb{E} = \mathbb{E}_1 \times \dots \times \mathbb{E}_m$ where $\mathcal{K}_i \subset \mathbb{E}_i$ for all $i \in \{1, \dots, r\}$, it is possible to see that Definition 5.1 resembles the classical AKKT from NLP via the following lemma:

Lemma 5.2. *Let \bar{x} be a feasible point of (NCP). If \bar{x} satisfies AKKT, then there are sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds and $\Lambda_i^k = 0$ whenever $G_i(\bar{x}) \in \text{int } \mathcal{K}_i$, for sufficiently large k .*

Proof. Let \bar{x} be an AKKT point associated with the sequences $\{(y_1, \dots, y_r)^k\}_{k \in \mathbb{N}} \rightarrow 0$, $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{(\Lambda_1, \dots, \Lambda_r)^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$. Proving (9) is trivial. Then, for every index $i \in \{1, \dots, r\}$ such that $G_i(\bar{x}) \in \text{int } \mathcal{K}_i$, there is some $k_0 \in \mathbb{N}$ such that $z_i^k := G_i(x^k) + y_i^k \in \text{int } \mathcal{K}_i$ for every $k > k_0$. Hence, for every such k and i , there is some $\alpha_i^k > 0$ such that $z_i^k + \alpha_i^k \Lambda_i^k \in \mathcal{K}_i$, so we have $0 \geq \langle \Lambda_i^k, z_i^k + \alpha_i^k \Lambda_i^k \rangle = \alpha_i^k \|\Lambda_i^k\|^2$. Thus, $\Lambda_i^k = 0$ when $G_i(\bar{x}) \in \text{int } \mathcal{K}_i$, for all $k > k_0$. \square

Note that the converse holds in NLP when $\mathcal{K} = \mathbb{R}_+^m$ is seen as the cartesian product of m copies of \mathbb{R}_+ . Therefore, Definition 5.1 is consistent with the usual form of AKKT in NLP. In the next section, we contextualize the other conditions in other classical particular cases of (NCP) as well.

Remark 3. In Steck's PhD thesis [75], it was introduced a new sequential optimality condition, which was further developed in [26, Definitions 3.1 and 3.2]. In view of [26, Remark 4.1], since \mathbb{E} is finite-dimensional and \mathcal{K} is a closed convex cone, their condition can be defined for (NCP) similarly to Definition 3.1, but replacing (10) by

$$\liminf_{k \rightarrow \infty} \langle \Lambda^k, G(x^k) \rangle \geq 0. \quad (35)$$

Following [26], we will call this condition s-AKKT. It is immediate to see that s-AKKT is implied by TAKKT. Moreover, if AGP holds at a feasible point \bar{x} and \mathcal{K} is self-dual, then Moreau's decomposition tells us that

$$\liminf_{k \rightarrow \infty} \langle \Lambda^k, G(x^k) \rangle = \liminf_{k \rightarrow \infty} -\langle \Lambda^k, \Pi_{\mathcal{K}}(-G(x^k)) \rangle \geq 0.$$

Then, if \mathcal{K} is self-dual, AGP implies s-AKKT. However, since AGP and TAKKT are independent, then s-AKKT is strictly implied by both of them in this case. Since TAKKT does not imply AKKT, then s-AKKT does not imply AKKT either. Conversely, Example 3.2 shows that AKKT does not imply s-AKKT; in fact, any sequence that satisfies (17) must have $\liminf_{k \rightarrow \infty} \Lambda^k G(x_1^k, x_2^k) = \Lambda^k x_2^k h(x_1^k) \rightarrow -1 < 0$, since $x_2^k \rightarrow 1$. In summary, if \mathcal{K} is self-dual, s-AKKT is strictly weaker than TAKKT, strictly weaker than AGP, but independent of AKKT. If \mathcal{K} is not self-dual, Example 3.2 also shows that AGP is independent of s-AKKT, but the other relations still hold. It is worth mentioning that all notions of convergence considered in [26] are equivalent in a finite dimensional setting, which is why we only address s-AKKT.

The relations between the new sequential optimality conditions of this paper are presented in the figure below, where the arrows indicate (strict) implications.

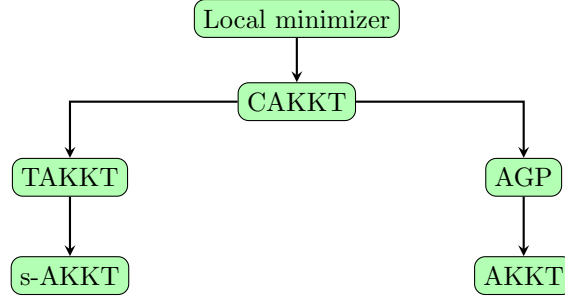


Figure 1: Relationship of the new sequential optimality conditions.

6 Contextualization in some particular cases of NCP

In this section, we specialize CAKKT, AGP, and AKKT, in the contexts of NLP, NLSDP, and NSOCP, to illustrate the stopping criteria associated with each of them in more practical terms. The TAKKT condition does not acquire any specific format when reduced to any context, so it is not included in this section.

6.1 Nonlinear programming

Consider the standard nonlinear programming problem with q inequality constraints and p equality constraints:

$$\begin{aligned}
 & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\
 & \text{subject to} && h(x) = 0, \\
 & && g(x) \leq 0.
 \end{aligned} \tag{NLP}$$

The most straightforward way of viewing (NLP) as a particular case of (NCP) is directly phrasing the constraints as $(h(x), g(x)) \in \{0\}^p \times \mathbb{R}_-^q$, but Moreau's Decomposition is meaningless when $\mathcal{K} = \{0\}^p$, and CAKKT reduces to TAKKT in this case. However, the CAKKT from [13] is strictly stronger than TAKKT, which indicates that this is not the best way of modelling (NLP) as a particular case of (NCP) concerning CAKKT. The AKKT and AGP conditions do not present the same behavior. It is noteworthy that if one considers (NCP) with separate equality constraints, i.e., with two blocks of constraints: $h(x) = 0$ and $G(x) \in \mathcal{K}$, then every definition from this paper can be straightforwardly extended to this new problem and, in this case, \mathcal{K} can be taken as \mathbb{R}_-^q . Then, we consider an auxiliary formulation of (NLP), where equality constraints are treated as two inequalities:

$$\begin{aligned}
 & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\
 & \text{subject to} && G(x) := (h(x), -h(x), g(x)) \in \mathbb{R}_-^p \times \mathbb{R}_-^p \times \mathbb{R}_-^q.
 \end{aligned} \tag{NLP-Auxiliary}$$

The Lagrange multipliers are then written as

$$\Lambda^k = (\omega_{1+}^k, \dots, \omega_{p+}^k, \omega_{1-}^k, \dots, \omega_{p-}^k, \mu_1^k, \dots, \mu_q^k) \in \mathbb{R}_+^p \times \mathbb{R}_+^p \times \mathbb{R}_+^q.$$

Our next result concerns about the specializations of CAKKT, AGP, and AKKT, from (NCP) to (NLP), but we use (NLP-Auxiliary) as an intermediary step for recovering the original (NLP) definitions of such conditions.

Proposition 6.1. *Let \bar{x} be a feasible point of (NLP). Then \bar{x} satisfies:*

- (a) *AKKT for (NLP-Auxiliary) if, and only if, $\mu_j^k = 0$ whenever $g_j(\bar{x}) < 0$, for every sufficiently large k and some sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$, $\{\omega^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^p$, and $\{\mu^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^q$ such that*

$$\nabla f(x^k) + \sum_{i=1}^p \omega_i^k \nabla h_i(x^k) + \sum_{j=1}^q \mu_j^k \nabla g_j(x^k) \rightarrow 0; \tag{36}$$

- (b) *AGP for (NLP-Auxiliary) if, and only if, $\mu_j^k \min\{0, g_j(x^k)\} \rightarrow 0$ and $\liminf_{k \rightarrow \infty} \omega_i^k h_i(x^k) \geq 0$, for every $i \in \{1, \dots, p\}$ and every $j \in \{1, \dots, q\}$; for some sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$, $\{\mu^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^q$, and $\{\omega^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^p$, such that (36) holds;*

(c) **CAKKT** for **(NLP-Auxiliary)** if, and only if, $\mu_j^k g_j(x^k) \rightarrow 0$ and $\omega_i^k h_i(x^k) \rightarrow 0$, for every $i \in \{1, \dots, p\}$ and every $j \in \{1, \dots, q\}$; for some sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$, $\{\mu^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+^q$, and $\{\omega^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^p$, such that (36) holds;

Proof. (a) **AKKT**: If AKKT holds for **(NLP-Auxiliary)**, there exist sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow x^k$, $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$, and $\{Y^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^q$, where $Y^k = (z_+^k, z_-^k, y^k) \rightarrow 0$, such that $h(x^k) + z_+^k \leq 0$, $-h(x^k) + z_-^k \leq 0$, and $g(x^k) + y^k \leq 0$ for every k , and

$$\langle G(x^k) + Y^k, \Lambda^k \rangle = \sum_{i=1}^p \omega_{i+}^k (h_i(x^k) + z_{i+}^k) + \sum_{i=1}^p \omega_{i-}^k (-h_i(x^k) + z_{i-}^k) + \sum_{j=1}^q \mu_j^k (g_j(x^k) + y_j^k) = 0. \quad (37)$$

Since all terms are nonnegative, they are all zero and, consequently, whenever $g_j(\bar{x}) < 0$ we have $g_j(x^k) + y_j^k < 0$ for sufficiently large k . That implies $\mu_j^k = 0$ for every such k and j . Take $\omega_i^k := \omega_{i+}^k - \omega_{i-}^k$ to obtain (36) from (9).

Conversely, consider the following natural choice: $Y^k := (-h(x^k), h(x^k), \widehat{g}_j(x^k)) \rightarrow 0$, where

$$\widehat{g}_j(x^k) := \begin{cases} 0, & \text{if } g_j(\bar{x}) < 0 \\ -g_j(x^k), & \text{otherwise,} \end{cases}$$

Then, for sufficiently large k , we have (37) and $G(x^k) + Y^k \in \mathcal{K}$, AKKT is satisfied with $\omega_{i+}^k := \max\{0, \omega_i^k\}$ and $\omega_{i-}^k := -\min\{0, \omega_i^k\}$.

(b) **AGP**: In this case, $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \rightarrow 0$ is equivalent to saying that

$$\sum_{i=1}^p \omega_{i+}^k \min\{0, h_i(x^k)\} + \sum_{i=1}^p \omega_{i-}^k \min\{0, -h_i(x^k)\} + \sum_{j=1}^q \mu_j^k \min\{0, g_j(x^k)\} \rightarrow 0.$$

Since each part of the sum has the same sign, we get that $\omega_{i-}^k \min\{0, -h_i(x^k)\} \rightarrow 0$, $\omega_{i+}^k \min\{0, h_i(x^k)\} \rightarrow 0$, and $\mu_j^k \min\{0, g_j(x^k)\} \rightarrow 0$. Hence, we have that $\mu_j^k \min\{0, g_j(x^k)\} \rightarrow 0$ for every $j \in \{1, \dots, q\}$ and, defining $\omega_i^k := \omega_{i+}^k - \omega_{i-}^k$, we observe that

$$\omega_i^k h_i(x^k) = \omega_{i+}^k \min\{0, h_i(x^k)\} - \omega_{i-}^k \min\{0, h_i(x^k)\} - \omega_{i+}^k \min\{0, -h_i(x^k)\} + \omega_{i-}^k \min\{0, -h_i(x^k)\}. \quad (38)$$

Then, note that the first and the last term of the right-hand side of (38) both vanish in the limit, and the middle terms are nonnegative. Hence, $\liminf_{k \rightarrow \infty} \omega_i^k h_i(x^k) \geq 0$, for every $i \in \{1, \dots, p\}$.

Conversely, set $\omega_{i+}^k := \max\{0, \omega_i^k\}$ and $\omega_{i-}^k := -\min\{0, \omega_i^k\}$. Then, for each k , only one term of the right-hand side of (38) can be nonzero, so its first and last terms must converge to zero, since they are nonpositive.

(c) **CAKKT**: Here, we get

$$\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle = \langle \omega_+^k, \min\{h(x^k), 0\} \rangle + \langle \omega_-^k, \min\{-h(x^k), 0\} \rangle + \langle \mu^k, \min\{g(x^k), 0\} \rangle$$

and

$$\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle = \langle \omega_+^k, \max\{h(x^k), 0\} \rangle + \langle \omega_-^k, \max\{-h(x^k), 0\} \rangle + \langle \mu^k, \max\{g(x^k), 0\} \rangle,$$

so if both tend to zero, define $\widehat{\omega}^k := \omega_+^k - \omega_-^k$ and we obtain $\widehat{\omega}_i^k h_i(x^k) \rightarrow 0$ and $\mu_j^k g_j(x^k) \rightarrow 0$.

The converse is analogous to the previous item, with the same choice of multipliers:

$$\omega_{i+}^k := \max\{0, \omega_i^k\} \quad \text{and} \quad \omega_{i-}^k := -\min\{0, \omega_i^k\}.$$

□

Note that from this point of view, our definitions are consistent with the original AKKT from [3], the AGP from [60] (which follows from Theorem 3.1), and the CAKKT from [13], respectively.

We summarize our results in the following table:

	Approximate complementarity condition
AKKT	$g_j(\bar{x}) < 0 \Rightarrow \mu_j^k = 0, \forall j$ and k sufficiently large.
AGP	$\mu_j^k \min\{g_j(x^k), 0\} \rightarrow 0$ and $\liminf_{k \rightarrow \infty} \omega_i^k h_i(x^k) \geq 0, \forall i, \forall j$.
CAKKT	$\mu_j^k g_j(x^k) \rightarrow 0$ and $\omega_i^k h_i(x^k) \rightarrow 0, \forall i, \forall j$.

Table 6.1: Sequential conditions when specialized to (NLP).

Since the sequential conditions mostly differ in how they deal with approximate complementarity, only this condition is made explicit in Table 6.1.

6.2 Nonlinear semidefinite programming

Here, we recall the classical form of an NLSDP problem:

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\ & \text{subject to} && G(x) \in \mathbb{S}_-^m, \end{aligned} \tag{NLSDP}$$

which is (NCP) with $\mathbb{E} = \mathbb{S}^m$, $\langle A, B \rangle := \text{tr}(AB)$ is the (Frobenius) inner product given by the trace of AB , and $\mathcal{K} = \mathbb{S}_-^m$. We recall that $\mathcal{K}^\circ = -\mathcal{K}$ and every symmetric matrix $A \in \mathbb{E}$ has a spectral decomposition, that is, there exists an orthogonal matrix U such that $A = UDU^T$, where $D = \text{Diag}(\lambda_1^U(A), \dots, \lambda_m^U(A))$ is a diagonal matrix and $\lambda_i^U(A)$ are eigenvalues of A ordered according to the eigenvectors in the columns of U . Moreover,

$$\Pi_{\mathcal{K}}(A) = U \text{Diag}(\min\{\lambda_1^U(A), 0\}, \dots, \min\{\lambda_m^U(A), 0\}) U^T$$

and a similar relation holds for $\Pi_{\mathcal{K}^\circ}(A)$ with max instead of min.

When no order is specified, we consider $\lambda_1(A) \leq \dots \leq \lambda_m(A)$. Note that for every $i \in \{1, \dots, m\}$ we have $\lambda_i(-A) = -\lambda_{m-i+1}(A)$. Also, the following inequality is important to our analyses: For every $A, B \in \mathbb{S}_-^m$, we have the inequality

$$\sum_{i=1}^m \lambda_i(A) \lambda_{m-i+1}(B) \leq \text{tr}(AB) \leq \sum_{i=1}^m \lambda_i(A) \lambda_i(B). \tag{39}$$

For its proof see [57].

Now, we specialize our conditions from (NCP) to (NLSDP):

Proposition 6.2. *Let \bar{x} be a feasible point of (NLSDP). Then \bar{x} satisfies:*

(a) *AKKT if, and only if, there exist sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$, $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$, and a sequence of orthogonal matrices $S^k \rightarrow U$, where U diagonalizes $G(\bar{x})$ and each S^k diagonalizes Λ^k , such that (9) holds and $\lambda_i^{S^k}(\Lambda^k) = 0$, if $\lambda_i^U(G(\bar{x})) < 0$, for sufficiently large k .*

(b) *AGP implies that, for every i ,*

$$\min\{0, \lambda_i(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0,$$

for some sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds;

(c) *CAKKT implies that, for every i ,*

$$\min\{0, \lambda_i(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0 \quad \text{and} \quad \min\{0, -\lambda_{m-i+1}(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0,$$

for some sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds;

Proof. (a) **AKKT:** If there is a sequence $\{y^k\}_{k \in \mathbb{N}} \rightarrow 0$ such that $G(x^k) + y^k \in \mathcal{K}$ and $\langle \Lambda^k, G(x^k) + y^k \rangle = 0$ for every k , the latter implies that $G(x^k) + y^k$ and Λ^k are simultaneously diagonalizable, that is, for every k , there is a matrix S^k such that $G(x^k) + y^k = S^k \Theta^k (S^k)^T$ and $\Lambda^k = S^k \Gamma^k (S^k)^T$, where $\Theta^k = \text{Diag}(\lambda_i^{S^k}(G(x^k) + y^k))$ and $\Gamma^k = \text{Diag}(\lambda_i^{S^k}(\Lambda^k))$. The continuity of G and the convergence of $\{x^k\}_{k \in \mathbb{N}}$ imply that $S^k \rightarrow U$, for some orthogonal matrix U . Then U diagonalizes $G(\bar{x})$ since $G(x^k) + y^k \rightarrow G(\bar{x})$, and if $\lambda_i^U(G(\bar{x})) < 0$, then for sufficiently large k we have $\lambda_i^{S^k}(G(x^k) + y^k) < 0$ as well. Then $\lambda_i^{S^k}(\Lambda^k) = 0$ for those k . Conversely, let \bar{x} be a feasible point associated with the sequences $\{\Lambda^k\}_{k \in \mathbb{N}}$, $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$, and $\{S^k\}_{k \in \mathbb{N}} \rightarrow U$, where each S^k diagonalize Λ^k and U diagonalizes $G(\bar{x})$, such that $\lambda_i^{S^k}(\Lambda^k) = 0$ whenever

$\lambda_i^U(G(\bar{x})) < 0$. Without loss of generality and to simplify the notation, we can suppose that U leaves the eigenvalues of $G(\bar{x})$ increasingly ordered. Also suppose that there are α negative eigenvalues and β zero eigenvalues in $G(\bar{x})$, with $\alpha + \beta = m$, and for some indexed matrix A^k , define $\tilde{A}^k := (S^k)^T A^k S^k$. Then, choose

$$y^k := S^k \begin{bmatrix} 0 & -\tilde{G}(x^k)_{\alpha\beta} \\ -\tilde{G}(x^k)_{\beta\alpha} & -\tilde{G}(x^k)_{\beta\beta} \end{bmatrix} (S^k)^T, \text{ where } \tilde{G}(x^k) = \begin{bmatrix} \tilde{G}(x^k)_{\alpha\alpha} & \tilde{G}(x^k)_{\alpha\beta} \\ \tilde{G}(x^k)_{\beta\alpha} & \tilde{G}(x^k)_{\beta\beta} \end{bmatrix}$$

and the partition $\alpha\alpha$ refers to the first α rows of the matrix, for example. Then, if $\Gamma^k := \tilde{\Lambda}^k$, we get

$$\langle \Lambda^k, G(x^k) + y^k \rangle = \langle \Gamma^k, \tilde{G}(x^k) + \tilde{y}^k \rangle = 0,$$

for sufficiently large k , because $\Gamma_{\alpha\alpha}^k = 0$ for large enough k . Since every block of $\tilde{G}(x^k)$ converges to zero, except for $\tilde{G}(x^k)_{\alpha\alpha}$, we know that $y^k \rightarrow 0$. Also, since $\tilde{G}(x^k) + \tilde{y}^k \in \mathbb{S}_-^m$, we get $G(x^k) + y^k \in \mathbb{S}_-^m$.

- (b) **AGP**: In this case, observe that $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle = \text{tr}(\Lambda^k \Pi_{\mathcal{K}}(G(x^k))) \rightarrow 0$ can be simplified using the left inequality of (39) for $A = \Pi_{\mathcal{K}}(G(x^k))$, $B = -\Lambda^k$. With this, we obtain $\lambda_i(\Pi_{\mathcal{K}}(G(x^k))) \lambda_{m-i+1}(-\Lambda^k) \rightarrow 0$, for all i . Since $\lambda_i(\Lambda^k) = -\lambda_{m-i+1}(-\Lambda^k)$, we conclude that AGP implies $\lambda_i(\Pi_{\mathcal{K}}(G(x^k))) \lambda_i(\Lambda^k) \rightarrow 0$, for every $i \in \{1, \dots, m\}$. This means

$$\min\{0, \lambda_i(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0.$$

- (c) **CAKKT**: Similarly to the previous item, from $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \rightarrow 0$ we get $\min\{0, \lambda_i(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0$. Also, from $\langle \Lambda^k, \Pi_{\mathcal{K}^o}(G(x^k)) \rangle \rightarrow 0$ and (39), we obtain $\lambda_i(-\Pi_{\mathcal{K}^o}(G(x^k))) \lambda_i(\Lambda^k) \rightarrow 0$, for every $i \in \{1, \dots, m\}$, which is equivalent to $\lambda_i(\Pi_{\mathcal{K}}(-G(x^k))) \lambda_i(\Lambda^k) \rightarrow 0$. Then,

$$\min\{0, -\lambda_{m-i+1}(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0.$$

□

Note that the characterizations for AGP and CAKKT are unilateral in this case, but since the purpose of specializing our conditions is to define stopping criteria, this is not an issue. For instance, if an algorithm employs the stopping criterion related to AGP (given by item (b) of Proposition 6.2), the proposition states that its feasible limit points are at least as good as AGP. Nevertheless, we point out that the converse statements hold when Λ^k and $G(x^k)$ commute, for every k .

We summarize our results for (NLSDP) in the following table:

	Approximate complementarity condition
AKKT	$\lambda_i^U(G(\bar{x})) < 0 \Rightarrow \lambda_i^{S^k}(\Lambda^k) = 0, \forall i, \forall k$ sufficiently large, where S^k diagonalizes Λ^k , for every k , and U diagonalizes $G(\bar{x})$.
AGP	$\min\{0, \lambda_i(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0, \forall i$.
CAKKT	$\min\{0, \lambda_i(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0$ and $\max\{0, \lambda_{m-i+1}(G(x^k))\} \lambda_i(\Lambda^k) \rightarrow 0, \forall i$.

Table 6.2: Sequential optimality conditions when specialized to (NLSDP). Recall that for $C \in \mathbb{S}^m$, the symbols $\lambda_1(C), \dots, \lambda_m(C)$ represent the eigenvalues of C , increasingly ordered.

Recall from Definition 3.3 that CAKKT incorporates the idea of controlling the behavior of the Lagrange multiplier through a vanishing measure of infeasibility. In NLP, this control can be understood in terms of growth, but in more general contexts such as NLSDP, it can have different meanings. As mentioned in the Introduction, the authors of [10] conjectured that the ideal definition of CAKKT should control the growth of all eigenvalues of the multiplier. In our case, only $\max\{0, m - 2r\}$ eigenvalues have their growths controlled, where r is the number of nonzero eigenvalues of $G(\bar{x})$. This suggests that even though our definition of CAKKT generalizes one of the multiple interpretations of the nonlinear programming CAKKT, it is still imperfect. We conjecture that our definition, the one presented in [2], and (6) are all independent. If this is the case, then there would be multiple correct ways of generalizing CAKKT. However, we were not able to find examples that support our claim, at this moment.

AGP was not yet defined for NLSDP and AKKT is consistent with the definition presented in [10, Definition 3.1]. Also, employing analogous reasoning, it is possible to recover AKKT from [2] in symmetric cones after imbuing a Jordan product into \mathbb{E} , by making use of the *Spectral Theorem* from [17].

6.3 Nonlinear second-order cone programming

Consider the following particular case of (NCP):

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\ & \text{subject to} && G_i(x) \in \mathcal{K}_i, i \in \{1, \dots, r\}, \end{aligned} \tag{NSOCP}$$

where $\mathbb{E} = \mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_r}$, with $m_1 + \dots + m_r = m$, and each $\mathcal{K}_i \subset \mathbb{R}^{m_i}$ is a *second-order cone* (or *Lorentz cone*), that is,

$$\mathcal{K}_i := \{(z_0, \hat{z}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} : \|\hat{z}\| \leq z_0\}, \quad i \in \{1, \dots, r\}.$$

Denote $G(x) := (G_1(x), \dots, G_r(x))$ and $\mathcal{K} := \mathcal{K}_1 \times \dots \times \mathcal{K}_r$. From [36], the interior and the boundary of \mathcal{K}_i are described by

$$\begin{aligned} \text{int } \mathcal{K}_i &:= \{(z_0, \hat{z}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} : \|\hat{z}\| < z_0\}, \\ \text{bd } \mathcal{K}_i &:= \{(z_0, \hat{z}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} : \|\hat{z}\| = z_0\}. \end{aligned}$$

Moreover, consider the following sets of indices:

$$\begin{aligned} \mathcal{I}_{\text{int}} &:= \{i \in \{1, \dots, r\} : G_i(\bar{x}) \in \text{int } \mathcal{K}_i\}, \\ \mathcal{I}_{\text{bd}_+} &:= \{i \in \{1, \dots, r\} : G_i(\bar{x}) \in \text{bd } \mathcal{K}_i \setminus \{0\}\}. \end{aligned} \tag{40}$$

It is well-known that $\mathcal{K}_i^\circ = -\mathcal{K}_i$, $\forall i$, and hence $\mathcal{K}^\circ = -\mathcal{K}$. In [36], there is a formula for the projection onto a single Lorentz cone \mathcal{K}_i . Following [36], every $v = (v_0, \hat{v}) \in \mathbb{R} \times \mathbb{R}^{m_i-1}$ can be decomposed as

$$v = \mu_1(v)c_1(v) + \mu_2(v)c_2(v), \tag{41}$$

where $\mu_\ell(v) \in \mathbb{R}$ and $c_\ell(v) \in \mathcal{K}$, for $\ell \in \{1, 2\}$, are given by the following expressions:

$$\mu_\ell(v) = v_0 + (-1)^\ell \|\hat{v}\| \quad \text{and} \quad c_\ell(v) = \begin{cases} (1/2)(1, (-1)^\ell \hat{v} \|\hat{v}\|^{-1}) & \text{if } \hat{v} \neq 0 \\ (1/2)(1, (-1)^\ell w) & \text{if } \hat{v} = 0, \end{cases} \tag{42}$$

where w is any unitary vector in \mathbb{R}^{m_i-1} . Clearly, we always have that $\mu_1(v) \leq \mu_2(v)$ and $0 \leq \langle c_i(v), c_j(w) \rangle \leq 1$ for every v, w and $i, j \in \{1, 2\}$. Now, for every $v = (v_0, \hat{v}) \in \mathbb{R} \times \mathbb{R}^{m_i-1}$, we have $\Pi_{\mathcal{K}_i}(v) := \max\{\mu_1(v), 0\}c_1(v) + \max\{\mu_2(v), 0\}c_2(v)$. Using $\mathcal{K}_i^\circ = -\mathcal{K}_i$, we obtain

$$\Pi_{\mathcal{K}_i^\circ}(v) := \min\{\mu_1(v), 0\}c_1(v) + \min\{\mu_2(v), 0\}c_2(v).$$

Finally, for $v = (v_1, \dots, v_r) \in \mathbb{E}$ with $v_i \in \mathbb{R} \times \mathbb{R}^{m_i-1}$, we have that $\Pi_{\mathcal{K}}(v) = (\Pi_{\mathcal{K}_1}(v_1), \dots, \Pi_{\mathcal{K}_r}(v_r))$ and $\Pi_{\mathcal{K}^\circ}(v) = (\Pi_{\mathcal{K}_1^\circ}(v_1), \dots, \Pi_{\mathcal{K}_r^\circ}(v_r))$.

We recall from [64, Lemma 6.2.3] that, for every $z_i, v_i \in \mathbb{R}^{m_i}$,

$$\mu_1(z_i)\mu_2(v_i) + \mu_2(z_i)\mu_1(v_i) \leq 2\langle z_i, v_i \rangle \leq \mu_1(z_i)\mu_1(v_i) + \mu_2(z_i)\mu_2(v_i). \tag{43}$$

Now, let \bar{x} be a feasible point of (NSOCP), and $\{x^k\}_{k \in \mathbb{N}}$ be a sequence with $x^k \rightarrow \bar{x}$ associated with some sequential optimality condition with multipliers $\{\Lambda^k\}_{k \in \mathbb{N}}$, then (9) can be stated as

$$\nabla f(x^k) + \sum_{i=1}^r DG_i(x^k)^T \lambda_i^k \rightarrow 0, \tag{44}$$

where $\Lambda^k = (\lambda_1^k, \dots, \lambda_r^k) \in \mathcal{K}_1^\circ \times \dots \times \mathcal{K}_r^\circ$.

Similarly to the previous subsections, the following result exhibits the formats of our sequential condition when specialized to NSOCP:

Proposition 6.3. *Let \bar{x} be a feasible point of (NSOCP). Then \bar{x} satisfies:*

- (a) *AKKT if, and only if, there exist sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$, $\{\lambda_i^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$, such that (44) holds, $\lambda_i^k = 0$ for every $i \in \mathcal{I}_{\text{int}}$ and sufficiently large k , and when $i \in \mathcal{I}_{\text{bd}_+}$, then $-\lambda_i^k \in \text{bd } \mathcal{K}_i$ and either $\lambda_i^k \rightarrow 0$ or*

$$\frac{\widehat{\lambda}_i^k}{\|\widehat{\lambda}_i^k\|} \rightarrow \frac{\widehat{G}_i(\bar{x})}{\|\widehat{G}_i(\bar{x})\|}; \tag{45}$$

- (b) *AGP implies that*

$$\mu_\ell(\lambda_i^k) \max\{\mu_\ell(G_i(x^k)), 0\} \rightarrow 0,$$

for every $\ell \in \{1, 2\}$ and every $i \in \{1, \dots, r\}$, for some sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\lambda_i^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that (44) holds;

(c) CAKKT implies that

$$\mu_\ell(\lambda_i^k) \max\{\mu_\ell(G_i(x^k)), 0\} \rightarrow 0, \quad \ell \in \{1, 2\} \quad \text{and} \quad \mu_\ell(\lambda_i^k) \min\{\mu_s(G_i(x^k)), 0\} \rightarrow 0,$$

for every $\ell, s \in \{1, 2\}$, with $\ell \neq s$, and every $i \in \{1, \dots, r\}$, for some sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that (44) holds.

Proof. (a) **AKKT:** If there are sequences $\{y_i^k\}_{k \in \mathbb{N}} \rightarrow 0$ such that $z_i^k := G_i(x^k) + y_i^k \in \mathcal{K}_i$ for every i and

$$\sum_{i=1}^r \langle \lambda_i^k, G_i(x^k) + y_i^k \rangle = 0,$$

then we have, for every i , that $\langle \lambda_i^k, G_i(x^k) + y_i^k \rangle = 0$. So if $G_i(\bar{x}) \in \text{int } \mathcal{K}_i$, we get $\lambda_i^k = 0$ for k sufficiently large, due to Lemma 5.2. But also, if $G_i(\bar{x}) \in \text{bd}_+ \mathcal{K}_i$, then $\|\widehat{G}_i(\bar{x})\| = (G_i(\bar{x}))_0 > 0$, so for large enough k , we must have $(z_i^k)_0 > 0$ as well. Now, note that

$$0 = \langle \lambda_i^k, z_i^k \rangle = \langle \widehat{\lambda}_i^k, \widehat{z}_i^k \rangle + (\lambda_i^k)_0 (z_i^k)_0 \leq \|\widehat{\lambda}_i^k\| \|\widehat{z}_i^k\| + (\lambda_i^k)_0 (z_i^k)_0 \leq (\|\widehat{\lambda}_i^k\| + (\lambda_i^k)_0) (z_i^k)_0,$$

but since $(z_i^k)_0 > 0$ we get $\|\widehat{\lambda}_i^k\| \geq -(\lambda_i^k)_0$. On the other hand, since $-\lambda_i^k \in \mathcal{K}_i$, we know that $\|\widehat{\lambda}_i^k\| \leq -(\lambda_i^k)_0$. Hence, $\|\widehat{\lambda}_i^k\| = -(\lambda_i^k)_0$, which means $-\lambda_i^k \in \text{bd } \mathcal{K}_i$. If $\lambda_i^k \neq 0$, then $(\lambda_i^k)_0 \neq 0$, but we still have $\langle \lambda_i^k, z_i^k \rangle = 0$, whence

$$1 = \lim_{k \rightarrow \infty} \left\langle \frac{\widehat{\lambda}_i^k}{-(\lambda_i^k)_0}, \frac{\widehat{z}_i^k}{(z_i^k)_0} \right\rangle = \lim_{k \rightarrow \infty} \left\langle \frac{\widehat{\lambda}_i^k}{\|\widehat{\lambda}_i^k\|}, \frac{\widehat{G}_i(x^k)}{(G_i(x^k))_0} \right\rangle = \lim_{k \rightarrow \infty} \left\langle \frac{\widehat{\lambda}_i^k}{\|\widehat{\lambda}_i^k\|}, \frac{\widehat{G}_i(\bar{x})}{\|\widehat{G}_i(\bar{x})\|} \right\rangle,$$

which means $\widehat{\lambda}_i^k / \|\widehat{\lambda}_i^k\| \rightarrow \widehat{G}_i(\bar{x}) / \|\widehat{G}_i(\bar{x})\|$. In order to check this, keep in mind that both vectors are unitary, so the cosine of the angle between them must tend to 1.

Conversely, without loss of generality, we can suppose that every λ_i^k such that $i \in \mathcal{I}_{\text{bd}_+}$ and $\lambda_i^k \rightarrow 0$, is equal to zero for k sufficiently large, and set

$$y_i^k := \begin{cases} \begin{cases} -G_i(x^k), & \text{if } G_i(\bar{x}) = 0, \\ -\Pi_{\mathcal{K}^\circ}(G_i(x^k)), & \text{if } i \in \mathcal{I}_{\text{int}} \text{ or } i \in \mathcal{I}_{\text{bd}_+} \text{ with } \lambda_i^k \rightarrow 0, \end{cases} \\ \frac{\|\widehat{G}_i(x^k)\|}{(\lambda_i^k)_0} \left((\lambda_i^k)_0, -\widehat{\lambda}_i^k \right) - G_i(x^k), & \text{if } i \in \mathcal{I}_{\text{bd}_+} \text{ with } \lambda_i^k \neq 0. \end{cases}$$

Then, we have $G_i(x^k) + y_i^k \in \mathcal{K}_i$, for every $i \in \{1, \dots, r\}$, because $((\lambda_i^k)_0, -\widehat{\lambda}_i^k) \in \mathcal{K}^\circ$ and $(\lambda_i^k)_0 \leq 0$. If $G_i(\bar{x}) \in \{0\} \cup \text{int } \mathcal{K}_i$, we clearly have $\langle G_i(x^k) + y_i^k, \lambda_i^k \rangle = 0$ for k sufficiently large. If $i \in \mathcal{I}_{\text{bd}_+}$, since $-\lambda_i^k \in \text{bd } \mathcal{K}_i$, we get

$$\langle G_i(x^k) + y_i^k, \lambda_i^k \rangle = \frac{\|\widehat{G}_i(x^k)\|}{(\lambda_i^k)_0} \left((\lambda_i^k)_0^2 - \|\widehat{\lambda}_i^k\|^2 \right) = 0.$$

Also, note that

$$(y_i^k)_0 = \|\widehat{G}_i(x^k)\| - (G_i(x^k))_0 \quad \text{and} \quad \widehat{y}_i^k = \frac{\|\widehat{G}_i(x^k)\|}{\|\widehat{\lambda}_i^k\|} \widehat{\lambda}_i^k - \widehat{G}_i(x^k),$$

so in case $\lambda_i^k \neq 0$ we get $y_i^k \rightarrow 0$ from $G_i(\bar{x}) \in \text{bd } \mathcal{K}_i \setminus \{0\}$ and (45).

(b) **AGP:** In this case, since

$$\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle = \sum_{i=1}^r \langle \lambda_i^k, \Pi_{\mathcal{K}_i}(G_i(x^k)) \rangle \rightarrow 0$$

and $\mathcal{K}^\circ = -\mathcal{K}$, we obtain $\langle \lambda_i^k, \Pi_{\mathcal{K}_i}(G_i(x^k)) \rangle \rightarrow 0$ for every $i \in \{1, \dots, r\}$. Now, using the spectral decomposition (41) and the right-hand side of (43), we obtain

$$2\langle \lambda_i^k, \Pi_{\mathcal{K}_i}(G_i(x^k)) \rangle \leq \mu_1(\lambda_i^k) \mu_1(\Pi_{\mathcal{K}_i}(G_i(x^k))) + \mu_2(\lambda_i^k) \mu_2(\Pi_{\mathcal{K}_i}(G_i(x^k))) \leq 0.$$

Hence, taking $k \rightarrow \infty$ in the above expression, we see that:

- $\mu_1(\lambda_i^k) \mu_1(\Pi_{\mathcal{K}_i}(G_i(x^k))) = \mu_1(\lambda_i^k) \max\{\mu_1(G_i(x^k)), 0\} \rightarrow 0;$
- $\mu_2(\lambda_i^k) \mu_2(\Pi_{\mathcal{K}_i}(G_i(x^k))) = \mu_2(\lambda_i^k) \max\{\mu_2(G_i(x^k)), 0\} \rightarrow 0.$

Furthermore, from above and since $\mu_1(\lambda^k) \leq \mu_2(\lambda^k) \leq 0$, we see that $\mu_2(\lambda_i^k) \max\{\mu_1(G_i(x^k)), 0\} \rightarrow 0$. Thus, $\langle \lambda_i^k, \Pi_{\mathcal{K}_i}(G_i(x^k)) \rangle \rightarrow 0$ implies $\mu_\ell(\lambda_i^k) \max\{\mu_\ell(G_i(x^k)), 0\} \rightarrow 0$ for $\ell \in \{1, 2\}$ and $\forall i \in \{1, \dots, r\}$.

- (c) **CAKKT**: Again, $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \rightarrow 0$ and $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \rightarrow 0$ are equivalent to $\langle \lambda_i^k, \Pi_{\mathcal{K}_i}(G_i(x^k)) \rangle \rightarrow 0$ and $\langle \lambda_i^k, \Pi_{\mathcal{K}_i^\circ}(G_i(x^k)) \rangle \rightarrow 0$, for every $i \in \{1, \dots, r\}$.

From $\langle \lambda_i^k, \Pi_{\mathcal{K}_i^\circ}(G_i(x^k)) \rangle \rightarrow 0$, using (41) and the left-hand side of (43), we have

$$0 \leq \mu_1(\lambda_i^k) \mu_2(\Pi_{\mathcal{K}_i^\circ}(G_i(x^k))) + \mu_2(\lambda_i^k) \mu_1(\Pi_{\mathcal{K}_i^\circ}(G_i(x^k))) \leq 2 \langle \lambda_i^k, \Pi_{\mathcal{K}_i^\circ}(G_i(x^k)) \rangle.$$

Then, from (41) and taking $k \rightarrow \infty$ in the above expression, we see that:

- $\mu_1(\lambda_i^k) \mu_2(\Pi_{\mathcal{K}_i^\circ}(G_i(x^k))) = \mu_1(\lambda_i^k) \min\{\mu_2(G_i(x^k)), 0\} \rightarrow 0$;
- $\mu_2(\lambda_i^k) \mu_1(\Pi_{\mathcal{K}_i^\circ}(G_i(x^k))) = \mu_2(\lambda_i^k) \min\{\mu_1(G_i(x^k)), 0\} \rightarrow 0$.

Furthermore, from above, we see that $\mu_2(\lambda_i^k) \min\{\mu_2(G_i(x^k)), 0\} \rightarrow 0$. Thus, $\langle \lambda_i^k, \Pi_{\mathcal{K}_i^\circ}(G_i(x^k)) \rangle \rightarrow 0$ implies $\mu_\ell(\lambda_i^k) \min\{\mu_s(G_i(x^k)), 0\} \rightarrow 0$ for $\ell, s \in \{1, 2\}$, $s \neq \ell$ and $\forall i \in \{1, \dots, r\}$. See Table 6.3. \square

Note that AKKT is consistent with [2, Definition 3.3] in view of [2, Theorem 4.1], which gives the exact same characterization as item (a) of Proposition 6.3, and AGP was not yet defined for NSOCP. Also, our version of CAKKT comprises eigenvalue products, which is similar to what we expected to obtain in the NLSDP case.

The following table summarizes our results:

	Approximate complementarity condition
AKKT	$i \in \mathcal{I}_{int}$, $\lambda_i^k \rightarrow 0$, and for $i \in \mathcal{I}_{bd+}$, $-\lambda_i^k \in \text{bd } \mathcal{K}_i$ and either $\lambda_i^k \rightarrow 0$ or $\widehat{\lambda}_i^k / \ \widehat{\lambda}_i^k\ \rightarrow \widehat{G}_i(\bar{x}) / \ \widehat{G}_i(\bar{x})\ $.
AGP	$\mu_\ell(\lambda_i^k) \max\{\mu_\ell(G_i(x^k)), 0\} \rightarrow 0$ for $\forall \ell \in \{1, 2\}$, $\forall i$.
CAKKT	$\mu_\ell(\lambda_i^k) \max\{\mu_\ell(G_i(x^k)), 0\} \rightarrow 0$ and $\mu_\ell(\lambda_i^k) \min\{\mu_s(G_i(x^k)), 0\} \rightarrow 0$, $\forall \ell, s \in \{1, 2\}$ ($\ell \neq s$), $\forall i$.

Table 6.3: Sequential optimality conditions when specialized to (NSOCP). We use $\{\mu_\ell(v), \ell \in \{1, 2\}\}$ to denote spectral values of $v \in \mathcal{K}_i$ (see (42)). For the definitions of \mathcal{I}_{bd+} and \mathcal{I}_{int} , see (40).

In NSOCP, the relation between Definition 3.3 and CAKKT as in [2, Definition 3.4] is not clear. We conjecture they are independent, which may endorse the possibility of existence of multiple independent extensions of CAKKT. Observe that the most important feature of our approach is its simplicity and its generality, since it only uses inner products and projections. On the other hand, [2, Definition 3.4] relies on the Jordan algebra structure, which is limited to symmetric cones, but it is more elegant than our approach in certain aspects.

7 Strength of the sequential optimality conditions

A sequential optimality condition carries the convergence properties of the algorithms supported by them and this is what gives them a practical meaning. However, even though we compared sequential conditions among themselves, we have not yet shown any improvement regarding the usual convergence theory of any algorithm. In other words, to complete our results, we still need to clarify the relation between our sequential conditions and other optimality conditions of the form “KKT or not-CQ” for some CQ. This section is dedicated to filling this gap.

Recall that the classical *Robinson’s CQ* [70] holds at some feasible point x when

$$0 \in \text{int}(G(x) + DG(x)\mathbb{R}^n - \mathcal{K}), \quad (46)$$

where $DG(x)\mathbb{R}^n := \{DG(x)d : d \in \mathbb{R}^n\}$. It is widely known that Robinson’s CQ generalizes the *Mangasarian-Fromovitz constraint qualification* (MFCQ) from NLP (see [56]). We proceed by re-proving the classical convergence results to KKT points under Robinson’s CQ, via sequential conditions:

Proposition 7.1. *If Robinson’s CQ holds at an AKKT point \bar{x} associated with the sequences $\{x^k\}_{k \in \mathbb{N}}$ and $\{\Lambda^k\}_{k \in \mathbb{N}}$, then \bar{x} satisfies the KKT conditions for (NCP).*

Proof. We begin by proving that $\{\Lambda^k\}_{k \in \mathbb{N}}$ is bounded. In order to do that, by contradiction, suppose not. Then, we can assume $\|\Lambda^k\| \rightarrow \infty$, but $\widehat{\Lambda}^k := \Lambda^k / \|\Lambda^k\| \rightarrow \bar{\Lambda} \in \mathcal{K}^\circ$. Then from (9) we get $\nabla f(x^k) / \|\Lambda^k\| + DG(x^k)^* \widehat{\Lambda}^k \rightarrow 0$ and, consequently, $DG(\bar{x})^* \bar{\Lambda} = 0$. Moreover,

$$0 = \langle \Lambda^k, G(x^k) + y^k \rangle = \langle \widehat{\Lambda}^k, G(x^k) + y^k \rangle \rightarrow \langle \bar{\Lambda}, G(\bar{x}) \rangle = 0$$

and, by Robinson's CQ, there exists some small $\alpha > 0$ such that $-\alpha \bar{\Lambda} \in (G(\bar{x}) + DG(\bar{x})\mathbb{R}^n - \mathcal{K})$. Let $d \in \mathbb{R}^n$ and $z \in \mathcal{K}$ be such that $-\alpha \bar{\Lambda} = G(\bar{x}) + DG(\bar{x})d - z$, so we have

$$-\alpha \langle \bar{\Lambda}, \bar{\Lambda} \rangle = \langle \bar{\Lambda}, G(\bar{x}) \rangle + \langle \bar{\Lambda}, DG(\bar{x})d \rangle - \langle \bar{\Lambda}, z \rangle = \langle DG(\bar{x})^* \bar{\Lambda}, d \rangle - \langle \bar{\Lambda}, z \rangle \geq 0,$$

which implies $\bar{\Lambda} = 0$. Since this contradicts the definition of $\bar{\Lambda}$, we conclude that $\{\Lambda^k\}_{k \in \mathbb{N}}$ must be bounded. Hence, without loss of generality we can assume it converges to $\bar{\Lambda} \in \mathcal{K}^\circ$. Trivially, $\nabla f(\bar{x}) + DG(\bar{x})^* \bar{\Lambda} = 0$ and

$$|\langle \bar{\Lambda}, G(\bar{x}) \rangle| = \lim_{k \rightarrow \infty} |\langle \Lambda^k, G(x^k) \rangle| = \lim_{k \rightarrow \infty} \langle \Lambda^k, y^k \rangle = 0.$$

Also, since \mathcal{K} is closed and $\mathcal{K} \ni G(x^k) + y^k \rightarrow G(\bar{x})$ we have $G(\bar{x}) \in \mathcal{K}$. Thus $(\bar{x}, \bar{\Lambda})$ is a KKT pair of (NCP). \square

Analogously, it can be proved that TAKKT also satisfies the KKT conditions under Robinson's CQ. Moreover, since AKKT is implied by CAKKT and AGP, the same holds for both. That means every algorithm that is supported by one of our sequential conditions converges to KKT points under Robinson's CQ, but sequential conditions tell us more than that. Following [12], for each sequential optimality condition (OC) it is possible to define conditions, so-called *strict constraint qualifications* (SCQ), such that " $OC + SCQ \Rightarrow KKT$ " and, among them, characterize the weakest one.

We use the same nomenclature style of [12]. For instance, the weakest SCQ associated with the AKKT condition is called *AKKT-regularity*, and similar names are given for the other sequential optimality conditions presented in Section 3.

Definition 7.1. Consider the following sets:

1. $\mathbb{K}_A(x, r) := \{DG(x)^* \Lambda : |y| \leq r, \Lambda \in \mathcal{K}^\circ, G(x) + y \in \mathcal{K} \cap \{\Lambda\}^\perp\}$;
2. $\mathbb{K}_T(x, r) := \{DG(x)^* \Lambda : |\langle \Lambda, G(x) \rangle| \leq r, \Lambda \in \mathcal{K}^\circ\}$;
3. $\mathbb{K}_{AGP}(x, r) := \{DG(x)^* \Lambda : |\langle \Lambda, \Pi_{\mathcal{K}}(G(x)) \rangle| \leq r, \Lambda \in \mathcal{K}^\circ\}$;
4. $\mathbb{K}_C(x, r) := \{DG(x)^* \Lambda : \max\{|\langle \Lambda, \Pi_{\mathcal{K}}(G(x)) \rangle|, |\langle \Lambda, \Pi_{\mathcal{K}^\circ}(G(x)) \rangle|\} \leq r, \Lambda \in \mathcal{K}^\circ\}$.

We say that the AKKT-regularity condition holds at \bar{x} , if the set-valued mapping $(x, r) \mapsto \mathbb{K}_A(x, r)$ is outer semi continuous at $(\bar{x}, 0)$. The constraint qualification conditions TAKKT-, AGP- and CAKKT-regularity have analogous definitions using the sets $\mathbb{K}_T(x, r)$, $\mathbb{K}_{AGP}(x, r)$, and $\mathbb{K}_C(x, r)$, respectively.

Remark 4. For a feasible point \bar{x} of (NCP), we see that, at $(\bar{x}, 0)$, all the sets from Definition 7.1 coincide with $\{DG(x)^* \Lambda : \langle \Lambda, G(x) \rangle = 0, \Lambda \in \mathcal{K}^\circ\}$. Thus, given an objective function f for (NCP), the KKT conditions hold at \bar{x} if, and only if, $-\nabla f(x) \in \mathbb{K}_A(\bar{x}, 0)$. Similar statements hold for the other sets.

The next theorem states that each SCQ is, indeed, the weakest SCQ associated with each optimality condition.

Theorem 7.2. *A feasible point \bar{x} for (NCP) satisfies CAKKT-regularity if, and only if, for every continuously differentiable objective function, the CAKKT condition at \bar{x} implies the KKT conditions. Similar conclusions are valid for AKKT-, AGP-, and TAKKT-regularity.*

Proof. We use the same techniques as [11, 12]. We just prove the statement for CAKKT-regularity, since the other ones are analogous.

Suppose that CAKKT-regularity holds at \bar{x} , and take any objective function f having \bar{x} as a CAKKT point. From definition, there exist sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^\circ$ such that

$$w^k := \nabla f(x^k) + DG(x^k)^* \Lambda^k \rightarrow 0, \quad \langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \rightarrow 0, \quad \text{and} \quad \langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \rightarrow 0.$$

Set $r_k := \max\{|\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle|, |\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle|\}, \forall k \in \mathbb{N}$. Thus, $-\nabla f(x^k) + w^k \in \mathbb{K}_C(x^k, r_k)$. Taking limits in the last expression, we get $-\nabla f(\bar{x}) \in \mathbb{K}_C(\bar{x}, 0)$ from the outer semi continuity of \mathbb{K}_C , and hence the KKT conditions holds at \bar{x} . Conversely, suppose that for every continuously differentiable objective function, the CAKKT condition at \bar{x} implies the KKT conditions. We will show that CAKKT-regularity hold at \bar{x} . Take $\omega \in \limsup_{(x,r) \rightarrow (\bar{x},0)} \mathbb{K}_C(x, r)$. Then, there exists an infinite subset $I \subseteq \mathbb{N}$ and sequences $\{x^k\}_{k \in I} \rightarrow \bar{x}$, $\{r_k\}_{k \in I} \rightarrow 0$,

and $\{w^k\}_{k \in I} \rightarrow w$, such that $w^k \in \mathbb{K}_C(x^k, r_k)$ for every $k \in I$, which means each w^k is associated with some $\Lambda^k \in \mathcal{K}^\circ$. Now, define the function $f(x) := -\langle \omega, x \rangle$, and let $i(k)$ denote the element of I that is closest to a given $k \in \mathbb{N}$; in particular, when $k \in I$, then $i(k) = k$. Defining $\tilde{x}^k := x^{i(k)}$ and $\tilde{\Lambda}^k := \Lambda^{i(k)}$ for every $k \in \mathbb{N}$, we see that \bar{x} is a CAKKT point for f associated with the sequences $\{\tilde{x}^k\}_{k \in \mathbb{N}}$ and $\{\tilde{\Lambda}^k\}_{k \in \mathbb{N}}$. Then, by assumption, the KKT conditions hold, which is equivalent to $-\nabla f(\bar{x}) = w \in \mathbb{K}_C(\bar{x}, 0)$, and thus CAKKT-regularity holds at \bar{x} . \square

From Theorem 7.1 and Definition 7.1, we observe that Robinson's CQ implies AKKT-regularity, which strictly implies AGP-regularity, and the latter strictly implies CAKKT-regularity due to Theorem 7.2 and the relations among the sequential conditions from Sections 3 and 5. Thus, our previous considerations show that the algorithms supported by the CAKKT condition are guaranteed to converge to KKT points under assumptions that are weaker than Robinson's CQ, for example.

Now, to conclude our analyses, we make explicit the relation between our SCQs and the very weak Abadie's CQ, that holds at a feasible point \bar{x} if, and only if, $T_\Omega(\bar{x}) = L_\Omega(\bar{x})$ and the set

$$\mathbb{K}_C(\bar{x}, 0) = \{DG(\bar{x})^* \Lambda : \langle \Lambda, G(\bar{x}) \rangle = 0, \Lambda \in \mathcal{K}^\circ\}$$

is closed, where

$$T_\Omega(\bar{x}) := \{d \in \mathbb{R}^n : \exists t_k \downarrow 0, d^k \rightarrow d \text{ with } \bar{x} + t_k d^k \in \Omega\}$$

is the *tangent cone* to Ω at \bar{x} and

$$L_\Omega(\bar{x}) := \{d \in \mathbb{R}^n : DG(\bar{x})d \in T_{\mathcal{K}}(G(\bar{x}))\}$$

is the so-called *linearized tangent cone* to Ω at \bar{x} .

The only affirmation that requires a proof is ‘‘CAKKT-regularity implies Abadie's CQ’’, but note that even in the finite-dimensional setting of (NCP) the set $\mathbb{K}_C(\bar{x}, 0)$ may not be closed [63]. Therefore, the first thing we prove is that this condition is guaranteed under CAKKT-regularity.

Lemma 7.3. *Let \bar{x} be a feasible point of (NCP) that satisfies CAKKT-regularity. Then, $\mathbb{K}_C(\bar{x}, 0)$ is closed, which in turn implies that it coincides with $L_\Omega(\bar{x})^\circ$.*

Proof. Recall that $\limsup_{(x,r) \rightarrow (\bar{x},0)} \mathbb{K}_C(x, r)$ is always a closed set, so if it coincides with $\mathbb{K}_C(\bar{x}, 0)$, then the latter is also closed. Moreover, it follows directly from [73, Corollary 11.25(d)] that if $\mathbb{K}_C(\bar{x}, 0)$ is closed, then it coincides with $L_\Omega(\bar{x})^\circ$. \square

In light of Lemma 7.3, the rest of the proof follows the same recipe as in NLP. To present it, we first recall the *regular normal cone* to Ω at $\bar{z} \in \Omega$, which is defined as

$$\widehat{N}_\Omega(\bar{z}) := \{w \in \mathbb{R}^n : \limsup_{z \rightarrow \bar{z}, z \in \Omega} \|z - \bar{z}\|^{-1} \langle w, z - \bar{z} \rangle \leq 0\},$$

and the *limiting normal cone* to Ω at $\bar{x} \in \Omega$, which is $N_\Omega(\bar{z}) := \limsup_{z \rightarrow \bar{z}, z \in \Omega} \widehat{N}_\Omega(z)$. Now, we present a technical lemma:

Lemma 7.4. *We always have that $N_\Omega(\bar{x}) \subset \limsup_{(x,r) \rightarrow (\bar{x},0)} \mathbb{K}_C(x, r)$.*

Proof. Analogous to the proof of [12, Lemma 4.3]. \square

And finally, the result:

Theorem 7.5. *CAKKT-regularity implies Abadie's CQ.*

Proof. Let \bar{x} be a feasible point such that CAKKT-regularity holds at \bar{x} . Using the definition of outer semi continuity of $\mathbb{K}_C(x, r)$ and Lemma 7.4, we get that $N_\Omega(\bar{x}) \subset \mathbb{K}_C(\bar{x}, 0)$, and by Lemma 7.3 we have that $\mathbb{K}_C(\bar{x}, 0)$ is closed and $\mathbb{K}_C(\bar{x}, 0) = L_\Omega(\bar{x})^\circ$. Therefore, $N_\Omega(\bar{x}) \subset L_\Omega(\bar{x})^\circ$. Now, since $T_\Omega(\bar{x}) \subset L_\Omega(\bar{x})$ always holds for every set, to show that Abadie's CQ holds at \bar{x} , it suffices to prove the inclusion $L_\Omega(\bar{x}) \subset T_\Omega(\bar{x})$. Now, from the inclusion $N_\Omega(\bar{x}) \subset L_\Omega(\bar{x})^\circ$, we get that $L_\Omega(\bar{x}) \subset (L_\Omega(\bar{x})^\circ)^\circ \subset N_\Omega(\bar{x})^\circ \subset T_\Omega(\bar{x})$, where the first inclusion follows from polarity theorem, since $L_\Omega(\bar{x})$ is a closed convex cone due to the fact that \mathcal{K} is also a closed convex cone, and the last inclusion comes from [73, Theorems 6.28(b) and 6.26]. \square

Remark 5. [26, Theorem 4.6] states that *s-AKKT-regularity* [26, Definition 4.4] is the weakest constraint qualification that guarantees equivalence between KKT and s-AKKT, similarly to Theorem 7.2. Therefore, s-AKKT-regularity is independent of AKKT-regularity, independent of AGP-regularity, and strictly stronger than TAKKT-regularity. When \mathcal{K} is self-dual, however, it implies AGP regularity strictly. Moreover, we obtain from [26, Theorem 5.2 and Example 5.4], that s-AKKT-regularity is strictly implied by Robinson's CQ.

For summarizing our results, we illustrate the position of the new SCQs among the existing CQs: s-AKKT-regularity, Abadie's CQ, and Robinson's CQ, in the diagram of Figure 2.

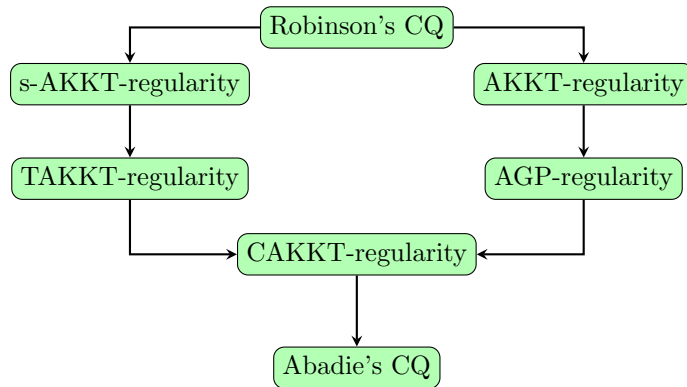


Figure 2: Relationship between the new (strict) constraint qualifications and existing ones.

8 Final remarks

Powerful modelling languages and other recent technological advances extended the possibilities for solving complex real-life problems. Such complexity is often translated in terms of (NCP), which is a large family of optimization problems, that generalizes NLP, NLSDP, and NSOCP, for example. In this paper, we extended to the NCP context some of the so-called sequential optimality conditions, which have been useful in particular cases of NCP for improving the global convergence analysis of several practical algorithms in a unified manner. Also, we presented a variant of the Augmented Lagrangian method for NCP, whose global convergence theory was built via sequential optimality conditions. We proved that every feasible limit point of this method satisfies AGP and, under an additional smoothness assumption, it also satisfies CAKKT, which is a strictly stronger condition. The meaning of such results lies in the fact that every CAKKT (respectively, AGP) point also satisfies the KKT conditions in the presence of a constraint qualification called CAKKT-regularity (respectively, AGP-regularity), which is strictly weaker than Robinson's condition. That means, for instance, that Algorithm 1 is at least as strong as the classical variants of the Augmented Lagrangian method, despite being much more general. To the best of our knowledge, the convergence of the Augmented Lagrangian to CAKKT points was only known in NLP and, more recently, in NSOCP, but its convergence to AGP points was not yet known even in NLP.

Intuitively, one may expect general environments to be more complicated or to be less likely to achieve strong results, in comparison with more structured ones. However, in this work we see the opposite since we were able to recover and improve most of the existing results from NLP, NLSDP, and NSOCP while employing simpler techniques in our analyses. In fact, we limited ourselves to using only somewhat simple structures, such as inner products and projections, to make our results as applicable as possible. Our efforts lead us to believe that NCP encompasses most of the fundamental aspects of the classical optimization theory in a natural way, which may encourage further research in this field. For instance, the relation between CAKKT and the concept with the same name, from [2], is still unclear. Another subject of further investigation is the role of sequential conditions in perturbation theory and error estimation, which may clarify their value as a theoretical local optimality analysis tool, as an alternative to the punctual KKT conditions. Second-order sequential conditions have recently appeared in [6, 22, 40, 43] for NLP and we intend to extend them to more general contexts as well. Moreover, as mentioned in the Introduction, there are several algorithms for NLP that have had their convergence theories (re)built via sequential optimality conditions of first- and second-order, such as [8, 21, 30, 37, 38, 40, 41, 67]; and we believe that this work can be useful for extending such algorithms to NCP as well.

Acknowledgments.

We are thankful to two anonymous referees for their remarkable suggestions. In particular, one referee pointed out [26] and suggested Remarks 3 and 5.

This project was funded by FAPESP grants 2013/05475-7, 2013/07375-0, 2017/17840-2, 2017/18308-2, and 2018/24293-0; by CNPq grants 301888/2017-5, 404656/2018-8, 303427/2018-3, 438185/2018-8, and 307270/2019-0; PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016); and by FONDECYT grant 1170907.

References

- [1] Andreani R, Birgin EG, Martínez JM, Schuverdt ML (2008) Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming* 111(1-2):5–32.
- [2] Andreani R, Fukuda EH, Haeser G, Santos DO, Secchin LD (2019) Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. *Optimization Online* URL http://www.optimization-online.org/DB_HTML/2019/10/7436.html.
- [3] Andreani R, Haeser G, Martínez JM (2011) On sequential optimality conditions for smooth constrained optimization. *Optimization* 60:627–641.
- [4] Andreani R, Haeser G, Mito LM, Ramírez H (2020) Weak notions of nondegeneracy in nonlinear semidefinite programming. *Optimization Online* URL http://www.optimization-online.org/DB_HTML/2020/12/8182.html.
- [5] Andreani R, Haeser G, Mito LM, Ramos A, Secchin LD (2020) On the best achievable quality of limit points of augmented Lagrangian schemes. *Optimization Online* URL http://www.optimization-online.org/DB_HTML/2020/07/7929.html.
- [6] Andreani R, Haeser G, Ramos A, Silva PJS (2017) A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA Journal of Numerical Analysis* 37(4):1902–1929.
- [7] Andreani R, Haeser G, Schuverdt ML, Silva PJS (2012) A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming, Ser. A* 135(1-2):255–273.
- [8] Andreani R, Haeser G, Schuverdt ML, Silva PJS (2012) Two new weak constraint qualifications and applications. *SIAM Journal on Optimization* 22:1109–1135.
- [9] Andreani R, Haeser G, Secchin LD, Silva PJS (2019) New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. *SIAM Journal on Optimization* 29(4):3201–3230.
- [10] Andreani R, Haeser G, Viana DS (2020) Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming* 180(1-2):203–235.
- [11] Andreani R, Martínez JM, Ramos A, Silva PJS (2016) A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization* 26(1):96–110.
- [12] Andreani R, Martínez JM, Ramos A, Silva PJS (2018) Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research* C2:693–1050.
- [13] Andreani R, Martínez JM, Svaiter BF (2010) A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization* 20:3533–3554.
- [14] Anjos MF, Lassere JB (2012) *Handbook on Semidefinite, Conic and Polynomial Optimization* (International Series in Operations Research and Management Science. Springer US. New York).
- [15] Attouch H, Bolte J, Redont P, Soubeyran A (2010) Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research* 35(2):438–457.
- [16] Attouch H, Bolte J, Svaiter BF (2013) Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting and regularized Gauss-Seidel methods. *Mathematical Programming* 137(1-2):91–129.
- [17] Baes M (2007) Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. *Linear Algebra and its Applications* 422(2):664–700.
- [18] Bertsekas DP (1982) *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press. New York).
- [19] Bertsekas DP (1999) *Nonlinear Programming* (Athenas Scientific. Belmont, Mass).
- [20] Birgin E, Martínez JM (2014) *Practical Augmented Lagrangian Methods for Constrained Optimization* (SIAM Publications. Philadelphia).
- [21] Birgin EG, Bueno LF, Martínez JM (2015) Assessing the reliability of general-purpose inexact restoration methods. *Journal of Computational and Applied Mathematics* 282:1–16.
- [22] Birgin EG, Haeser G, Ramos A (2018) Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. *Computational Optimization and Applications* 69(1):51–75.
- [23] Birgin EG, Krejić N, Martínez JM (2018) On the employment of inexact restoration for the minimization of functions whose evaluation is subject to errors. *Mathematics of Computation* 87:1307–1326.
- [24] Bolte J, Daniilidis A, Ley O, Mazet L (2010) Characterization of Łojasiewicz inequalities and applications: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society* 362:3319–3363.
- [25] Bonnans JF, Shapiro A (2000) *Perturbation Analysis of Optimization Problems* (Springer-Verlag. Berlin).
- [26] Börgens E, Kanzow C, Mehrlitz P, Wachsmuth G (2020) New Constraint Qualifications for Optimization Problems in Banach Spaces Based on Asymptotic KKT Conditions. *SIAM Journal on Optimization* 30(4):2956–2982.
- [27] Bueno LF, Haeser G, Lara F, Rojas FN (2020) An augmented Lagrangian method for quasi-equilibrium problems. *Computational Optimization and Applications* 76:737–766.
- [28] Bueno LF, Haeser G, Martínez JM (2015) A flexible inexact restoration method for constrained optimization. *Journal of Optimization Theory and Applications* 165(1):188–208.

- [29] Bueno LF, Haeser G, Rojas FN (2019) Optimality conditions and constraint qualifications for generalized nash equilibrium problems and their practical implications. *SIAM Journal of Optimization* 29(1):31–54.
- [30] Chen L, Goldfarb D (2006) Interior-point 2-penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming* 108:1026.
- [31] Chill R, Mildner S (2018) The Kurdyka-Lojasiewicz-Simon inequality and stabilisation in nonsmooth infinite-dimensional gradient systems. *Proceeding of the American Mathematical Society* 146:4307–4314.
- [32] Fares B, Apkarian P, Noll D (2001) An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. *International Journal of Control* 74(4):348–360.
- [33] Fischer A, Friedlander A (2010) A new line search inexact restoration approach for nonlinear programming. *Computational Optimization and Applications* 46(2):333–346.
- [34] Fitzpatrick S, Phelps RR (1982) Differentiability of the metric projection in Hilbert space. *Transactions of the American Mathematical Society* 270(2):483.
- [35] Fukuda EH, Haeser G, Mito LM (2020) Second-order analysis for semidefinite and second-order cone programming via sequential optimality conditions. *Optimization Online* URL http://www.optimization-online.org/DB_HTML/2020/08/7951.html.
- [36] Fukushima M, Luo ZQ, Tseng P (2006) Smoothing functions for second-order-cone complementarity problems. *SIAM Journal on Optimization* 12(2):436–460.
- [37] Gill PE, Kungurtsev V, Robinson DP (2020) A shifted primal-dual penalty-barrier method for nonlinear optimization. *SIAM Journal on Optimization* 30(2):1067–1093.
- [38] Gill PE, Kungurtsev V, Robinson DP (2017) A stabilized SQP method: global convergence. *IMA Journal of Numerical Analysis* 37(1):407–443.
- [39] Giorgi G, Jiménez B, Novo V (2016) Approximate Karush-Kuhn-Tucker condition in multiobjective optimization. *Journal of Optimization Theory and Applications* 171(1):70–89.
- [40] Haeser G (2018) A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. *Computational Optimization and Applications* 70(2):615–639.
- [41] Haeser G, Hinder O, Ye Y (2019) On the behavior of Lagrange multipliers in convex and nonconvex infeasible interior point methods. *To appear in Mathematical Programming* URL <https://doi.org/10.1007/s10107-019-01454-4>.
- [42] Haeser G, Liu H, Ye Y (2019) Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary. *Mathematical Programming* 178(1):263–299.
- [43] Haeser G, Ramos A (2020) New constraint qualifications with second-order properties in nonlinear optimization. *Journal of Optimization Theory and Applications* 182(2):494–506.
- [44] Hager WW, Gowda MS (1999) Stability in the presence of degeneracy and error estimation. *Mathematical Programming* 85:181–192.
- [45] Hager WW, Mico-Umutesi D (2014) Error estimation in nonlinear optimization. *Journal of Global Optimization* 59:327–341.
- [46] Hestenes MR (1969) Multiplier and gradient methods. *Journal of Optimization Theory and Applications* 4:303–320.
- [47] Hiriart-Urruty JB, Lemaréchal C (2001) *Fundamentals of Convex Analysis* (Springer-Verlag Berlin Heidelberg, Berlin).
- [48] Kanzow C, Steck D, Wachsmuth D (2018) An augmented Lagrangian method for optimization problems in Banach spaces. *SIAM Journal on Control and Optimization* 56(1):272–291.
- [49] Kanzow C, Steck D (2017) An example comparing the standard and safeguarded augmented Lagrangian methods. *Operations Research Letters* 45(6):598–603.
- [50] Kocvara M, Stingl M (2004) Solving nonconvex SDP problems of structural optimization with stability control. *Optimization Methods and Software* 19(5):595–609.
- [51] Krislock N, Wolkowicz H (2012) Euclidean distance matrices and applications. *International Series in Operations Research and Management Science* 166:879–914.
- [52] Kurdyka K (1998) On gradients of functions definable in o-minimal structures. *Annales de l’institut Fourier (Genève)* 48:769–783.
- [53] Li G, Pong TK (2018) Calculus of the exponent of Kurdyka-Lojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations of Computational Mathematics* 18(5):1199–1232.
- [54] Lobo MS, Vandenberghe L, Boyd S, Lebret H (1998) Applications of second-order cone programming. *Linear Algebra and Applications* 284:193–228.
- [55] Lojasiewicz S (1984) Sur les trajectoires du gradient d’une fonction analytique. *in Seminari di Geometria 1982-1983, Università di Bologna, Istituto di Geometria, Dipartimento di Matematica* 115–117.
- [56] Mangasarian OL, Fromovitz S (1967) The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications* 17(1):37–47.
- [57] Marcus M (1956) An eigenvalue inequality for the product of normal matrices. *The American Mathematical Monthly* 63:173–174.

- [58] Martínez JM, Pilotta EA (2000) Inexact restoration algorithms for constrained optimization. *Journal of Optimization Theory and Applications* 104:135–163.
- [59] Martínez JM, Pilotta EA (2005) Inexact restoration methods for nonlinear programming: advances and perspectives. Qi L, Teo K, Yang X, eds., *Optimization and Control with applications*, 271–292 (Springer).
- [60] Martínez JM, Svaiter BF (2003) A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications* 118:117–133.
- [61] Nocedal J, Wright S (New York. 2006) *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, ISBN 0387987932.
- [62] O’Neill M, Wright SJ (2020) A log-barrier Newton-CG method for bound constrained optimization with complexity guarantees. *IMA Journal of Numerical Analysis*, drz074 URL <https://doi.org/10.1093/imanum/drz074>.
- [63] Pataki G (2007) On the closedness of the linear image of a closed convex cone. *Mathematics of Operations Research* 32(2):395–412.
- [64] Peng J, Roos C, Terlaky T (2002) *Self-Regularity: A new paradigm for Primal-Dual Interior Point Algorithms*. Princeton series in Applied Mathematics (Princeton University Press).
- [65] Powell MJD (1982) *A method for nonlinear constraints in minimization problems* (Academic Press. New York).
- [66] Qi HD, Yuan X (2013) Computing the nearest Euclidean distance matrix with low embedding dimensions. *Mathematical Programming* 147(1-2):351–389.
- [67] Qi L, Wei Z (2000) On the constant positive linear dependence conditions and its application to SQP methods. *SIAM Journal on Optimization* 10:963–981.
- [68] Ramos A (2019) Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. *To appear in Optimization Methods and Software* URL <https://doi.org/10.1080/10556788.2019.1702661>.
- [69] Ramos A (2019) Two new weak constraint qualifications for mathematical programs with equilibrium constraints and applications. *Journal on Optimization Theory and Applications* 183:566–591.
- [70] Robinson SM (1976) Stability theory for systems of inequalities, part II: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis* 13:497–513.
- [71] Robinson SM (1982) Generalized equations and their solutions, part II: Applications to nonlinear programming. *Mathematical programming Study* 19:200–221.
- [72] Rockafellar RT (1974) Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization* 12:268–285, ISSN 0363-0129.
- [73] Rockafellar RT, Wets R (2009) *Variational Analysis* (Grundlehren der mathematischen Wissenschaften, v. 317. Springer-Verlag Berlin Heidelberg. Berlin).
- [74] Schnabel RB (1982) Determining feasibility of a set of nonlinear inequality constraints. *Mathematical Programming Studies* 16:137–148.
- [75] Steck D (2018) Lagrange multiplier methods for constrained optimization and variational problems in Banach spaces. *PhD thesis, Institute of Mathematics, Universität Würzburg* URL <https://arxiv.org/abs/urn:nbn:de:bvb:20-opus-174444>.
- [76] Tuyen NV, Xiao YB, Son TQ (2019) On AKKT optimality conditions for cone-constrained vector optimization problems. *to appear in Journal of Nonlinear Convex Analysis* URL <https://arxiv.org/abs/1902.07416>.
- [77] Wolkowicz H, Saigal R, Vandenberghe L (2000) *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications* (International Series in Operations Research and Management Science. Springer US. New York).
- [78] Yamashita H, Yabe H (2015) A survey of numerical methods for nonlinear semidefinite programming. *Journal of the Operations Research Society of Japan* 58(1):24–60.