On optimality conditions for nonlinear conic programming

Roberto Andreani ∗ Walter Gómez † Gabriel Haeser ‡ Leonardo M. Mito § Alberto Ramos ¶
March 3, 2020

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:

$$\begin{align*}
\text{Minimize} & \quad f(x), \\
\text{subject to} & \quad G(x) \in K,
\end{align*}$$

(NCP)

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{E}$ are continuously differentiable mappings, $\mathbb{E}$ is an $m$-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 $K \subseteq \mathbb{E}$ is a nonempty closed convex cone. Let us denote its feasible set by $\Omega$. 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 [29], truss design problems and combinatorial optimization [66], portfolio optimization [45], structural optimization [41], and others. For more details, see [67, 66, 12] and references therein.

Algorithms for solving optimization problems are usually iterative and their convergence theories are mostly 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 [47], but strictly stronger [3, Theorem 2.2]. Indeed, it has been specially useful for improving and unifying global convergence results of several numerical methods, such as Augmented Lagrangian methods [11, 13], some Sequential Quadratic Programming

---

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, 458185/2018-8, 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 S˜ ao Paulo, S˜ ao Paulo, SP, Brazil. Email: ghaeser@ime.usp.br
§Department of Applied Mathematics, University of S˜ ao Paulo, S˜ ao Paulo, SP, Brazil. Email: leokoto@ime.usp.br
¶Department of Mathematics, Federal University of Paraná, Curitiba, PR, Brazil. Email: albertoramos@ufpr.br
(SQP) algorithms [53], and interior point methods [27, 34] – see [6, 5, 9, 10] and references therein. It is also important to mention that such idea has been carried over for several other contexts, for example: Nash equilibrium problems [29], mathematical programs with equilibrium constraints (MPECs) [57, 58], mathematical programs with complementarity constraints (MPCCs) [7], nonlinear vector optimization with conic constraints [55], the multiobjective case [33], variational problems in Banach spaces [40], quasi-equilibrium problems [25], and several others.

This paper aims at expanding this well known class of stronger optimality conditions from NLP to the general conic framework (NCP). As far as we know, this problem have been addressed only in [8], where NLSDP is considered and in [42], where NSOCP and, more generally, optimization over symmetric cones are considered. The most difficult aspect of such generalizations is dealing with complementarity. For NLP constraints \( g(x) \leq 0, g : \mathbb{R}^n \to \mathbb{R}^m \), and a Lagrange multiplier \( \lambda \in \mathbb{R}^m, \lambda \geq 0, \lambda = (\lambda_1, \ldots, \lambda_m) \), the complementarity constraint \( (\lambda, g(x)) = 0 \) means precisely that, for every \( i \in \{1, \ldots, m\} \), the multiplier \( \lambda_i \) is complementary with respect to the constraint \( g_i(x) \leq 0 \), in the sense that \( \lambda_i g_i(x) = 0 \) at a feasible point of interest \( x \). It turns out that, when considering perturbations of \( x \), the latter gives a stronger optimality condition. In the conic case \( (\text{NCP}) \), it is not clear how to exploit a complementarity-like structure in a statement of the form \( \langle \Lambda, G(x) \rangle = 0 \) for a Lagrange multiplier \( \Lambda \in \mathbb{K}^n \) (the polar of the cone \( \mathbb{K} \)) where \( G(x) \in \mathbb{K} \). In the context of NLSDP [8], the eigenvalues are heavily employed to exploit a complementarity-like structure, where one must carefully consider how to order consistently the eigenvalues of \( G(x) \) and \( \Lambda \). 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 \( \mathbb{K} = \{ u \circ v : u \in \mathbb{E} \} \). Note that self-duality of \( \mathbb{K} \) plays an important role in defining these optimality conditions [8, 2].

In this paper we propose a much more general and unified approach for defining such conditions. Here, we propose splitting \( \langle \Lambda, G(x) \rangle = 0 \), by means of Moreau’s decomposition, into two complementarity-like statements of the form \( \langle \Lambda_k, G(x) \rangle = 0 \) and \( \langle \Lambda_k, G(x) \rangle = 0 \), where \( \Pi_K \) and \( \Pi_K^* \) denote orthogonal projections onto \( \mathbb{K} \) and its polar, respectively. Hence, no particular structure of the cone \( \mathbb{K} \) is needed. We then show that a primal dual sequence \( \{(x^k, \Lambda^k)\} \subset \mathbb{R}^n \times \mathbb{K}^n \) generated by an Augmented Lagrangian method is such that \( \langle \Lambda^k, \Pi_K G(x^k) \rangle \to 0 \). 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, [11]), 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 \( \langle \Lambda^k, \Pi_K G(x^k) \rangle \to 0 \) is also satisfied. This answers an open question of [8] in the context of NLSDP by presenting a stronger complementarity-like structure, generated by the Augmented Lagrangian method, which was not achieved in [8]. 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 \( \mathbb{K} \) has a nonempty interior, 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 int $A$.

Given a set-valued mapping $\Gamma : \mathbb{R}^k \rightrightarrows \mathbb{E}$, the sequential (Painlevé-Kuratowski) outer limit of $\Gamma(z)$ as $z \to \bar{z}$, is the set $\{\bar{w} \in \mathbb{E} : \exists (\bar{z}, \bar{w}) \in \Gamma(\bar{z})\}$, which is denoted by $\limsup \Gamma(z)$.

For a differentiable mapping $G : \mathbb{R}^n \to \mathbb{E}$, we use $DG(x)$ to denote the derivative of $G$ at $x$, and $DG(x)^* : \mathbb{E} \to \mathbb{R}^n$ 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 $f : \mathbb{R}^n \to \mathbb{R}$, 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 $K^0 := \{w \in \mathbb{E} : \langle w, k \rangle \leq 0, \forall k \in K\}$ and $(K^0)^0 = K$. 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 [31].

The following lemma encompasses other useful and well-known properties about projections:

**Lemma 2.1** (see, e.g., [39]). 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, $v \in K$, $w - v \in K^0$, 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 $w = \Pi_K(w) + \Pi_{K^0}(w)$ and $\langle \Pi_K(w), \Pi_{K^0}(w) \rangle = 0$.

### 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:

$$\begin{align*}
\text{Minimize} & \quad f(x), \\
\text{subject to} & \quad G(x) \leq 0,
\end{align*}$$

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

$$\nabla f(x^k) + \sum_{i=1}^m \Lambda^k_i \nabla g_i(x^k) \to 0,$$  

and $\Lambda^k = 0$, 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 the Mangasarian-Fromovitz constraint qualification (MFCQ), for example, every AKKT point also satisfies the KKT conditions (see [9, 5, 6]). 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

$$\|\max\{0, G(x_\varepsilon)\}\| \leq \varepsilon, \|\nabla f(x_\varepsilon) + \sum_{i=1}^m (\Lambda_\varepsilon_i) \nabla g_i(x_\varepsilon)\| \leq \varepsilon, \|\min\{\Lambda_\varepsilon, -G(x_\varepsilon)\}\| \leq \varepsilon.$$

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 [2] the so-called Complementary AKKT (CAKKT) condition, that holds at a feasible point $\bar{x}$ when there are sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n_+$ such that $[\Lambda^k G_i(x^k) \to 0, \forall i \in \{1, \ldots, m\}$. 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 [51] under the name Approximate Gradient Projection (AGP), that holds at a feasible point $\bar{x}$ when there exists some sequence $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ such that

$$||L_{(\Omega, x^k)}(-\nabla f(x^k))|| \to 0,$$

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 [49, 50, 51, 50, 23] and references therein for details.

Now, consider the following NLSDP problem:

Minimize \( f(x), \)
subject to \( G(x) \in S^m_\geq \)

which is a particular case of (NCP) where $E = S^m_\geq$ is the linear space of $m \times m$ symmetric matrices, $K = S^m_\leq$ is the cone of $m \times m$ symmetric negative semidefinite matrices. The AKKT extension for (NLP) presented in [8, Definition 3.1] holds at a feasible point $\bar{x}$ when there are sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset S^m_\geq := -S^m_\leq$ such that

$$\nabla f(x^k) + DG(x^k)^* \Lambda^k \to 0$$

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 $\{U^k\}_{k\in\mathbb{N}} \to U$, where $U^k$ diagonalizes $\Lambda^k$, 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 $S^m_\geq$. Still under the same analogy, the most natural extension of CAKKT, discussed in [8], would simply require [5] and

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

where $\{S^k\}_{k\in\mathbb{N}} \to U$ 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 Lojasiewicz. Then, instead of using the eigenvalue product, the authors of [8] used the canonical inner product of $S^m_\geq$ (given by the trace of the matrix product) to define a new condition called Trace AKKT (TAKKT), that requires [5] and

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

Surprisingly, TAKKT has been proven to be completely independent of AKKT (see [8, Example 5.2] and [2, Example 3.1]) and it also requires the generalized Lojasiewicz 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 \cite{22,50}, for example) that the Karush-Kuhn-Tucker (KKT) conditions hold at a feasible point \( \bar{x} \) of \( \text{(NCP)} \) when there exists some \textit{Lagrange multiplier} \( \lambda \in \mathcal{K}^o \) such that

\[
\nabla f(\bar{x}) + DG(\bar{x})^* \lambda = 0, \\
\langle \lambda, G(\bar{x}) \rangle = 0.
\]

By taking the natural relaxation of (7) and (8), we obtain a trivial extension of the TAKKT condition \cite{8} 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 \( \text{(NCP)} \). We say that \( \bar{x} \) satisfies the TAKKT condition if there exist sequences \( \{x^k\}_{k \in \mathbb{N}} \to \bar{x} \) and \( \{\lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^o \) such that

\[
\nabla f(x^k) + DG(x^k)^* \lambda^k \to 0, \\
\langle \lambda^k, G(x^k) \rangle \to 0.
\]

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 \( \text{(NCP)} \) as well. Before that, note that at a KKT point, \( \bar{x} \), we prove that TAKKT is an optimality condition for \( \text{(NCP)} \) as well. Before that, note that at a KKT point, \( \bar{x} \),

\[
\nabla f(\bar{x}) + DG(\bar{x})^* \lambda = 0, \\
\langle \lambda, G(\bar{x}) \rangle = 0.
\]

\[
\text{Let } \bar{\lambda} \text{ be a feasible point of } \text{(NCP)}. \text{ We say that } \bar{x} \text{ satisfies the Approximate Gradient Projection (AGP) condition if there exist sequences } \{x^k\}_{k \in \mathbb{N}} \to \bar{x} \text{ and } \{\lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^o \text{ such that (9) holds and}
\]

\[
\langle \lambda^k, \Pi_k(G(x^k)) \rangle \to 0.
\]

A similar optimality condition has appeared in \cite{10} Definition 5.2, where the authors deal with a version of \( \text{(NCP)} \) 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 \cite{51}, which is given by (11). What follows is a proof of our claim:

**Theorem 3.1.** Consider \( \text{(NLP)} \), which is \( \text{(NCP)} \) over \( \mathbb{E} = \mathbb{R}^m \) 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 (11) holds at \( \bar{x} \).

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

\[
\text{Minimize } \frac{1}{2} \| \nabla f(x^k) - d^k \|^2, \\
\text{subject to } \min \{0, G_i(x^k)\} + DG_i(x^k)d \leq 0, \forall i \in \mathcal{A}(\bar{x}),
\]

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

\[
\hat{\lambda}^k := \begin{cases} 
\tilde{\lambda}^k, & \text{if } i \in \mathcal{A}(\bar{x}), \\
0, & \text{otherwise}.
\end{cases}
\]

Hence,

\[
\nabla f(x^k) + d^k + DG(x^k)^* \lambda^k = 0 \text{ and } \langle \lambda^k, \min \{0, G(x^k)\} + DG(x^k)d^k \rangle = 0.
\]

Multiplying (13) by \( d^k \), we obtain

\[
\|d^k\|^2 = -\langle \nabla f(x^k), d^k \rangle - \langle d^k, DG(x^k)^* \lambda^k \rangle = -\langle \nabla f(x^k), d^k \rangle - \langle DG(x^k)d^k, \lambda^k \rangle \\
\leq -\langle \nabla f(x^k) + DG(x^k)^* \lambda^k, d^k \rangle + \langle DG(x^k)d^k + \min \{0, G(x^k)\}, \lambda^k \rangle \\
\leq \|\nabla f(x^k) + DG(x^k)^* \lambda^k\|\|d^k\| + \sum_{i \in \mathcal{A}(\bar{x})} (DG_i(x^k)d^k + \min \{0, G_i(x^k)\})\lambda^k_i - \langle \min \{0, G(x^k)\}, \lambda^k \rangle,
\]
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 $d^k \to 0$ and (4) holds, because $\Lambda^*_k \to 0$ for every $i \notin A(x)$.

Conversely, assume that $\bar{x}$ satisfies AGP as in (4) and set $d^k := \Pi_L(\alpha^k)\nabla f(x^k)$. 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 $\nabla f(x^k) + DG(x^k)^*\bar{\Lambda}^k \to -d^k \to 0$ and $\langle \bar{\Lambda}^k, \min\{0, G(x^k)\} \rangle \to 0$. Consequently, Definition 3.2 holds at $\bar{x}$.

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 $G(x_1, x_2) \in K$, where $G(x_1, x_2) := (-x_1, x_1 \exp(x_2))$, $K := \mathbb{R}^2_+$, and the feasible point $\bar{x} = (0, 1)$. In this case, $\Lambda^k := (\lambda^*_1, 0) \in \mathbb{R}^2_+$ and $-\nabla f(x^k) + DG(x^k)^*\bar{\Lambda}^k \to 0$ reduces to $-\lambda^*_1 + \lambda^*_2 \exp(x^k_2) \to 0$ and $\Lambda^k \to \bar{\Lambda}^k \exp(x^k_2) \to 0$.

**TAKKT holds at $\bar{x}$:** Take $x^k_1 := 1/k$, $x^k_2 := 1$, $\lambda^*_1 := (x^k_1 \exp(x^k_2))^{-1}$, $\lambda^*_2 := (x^k_2) \exp(x^k_2)$. It is elementary to verify that $\{x^k := (x^k_1, x^k_2)\}$ 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 $(\Lambda^k, \Pi_K(G(x^k))) \to 0$ implies that $\bar{\lambda}^k \min(\{-x^k_1\}, 0) + \lambda^*_2 \exp(x^k_2) \min(x^k_1, 0) \to 0$. If there is an AGP sequence with $x^k_1 > 0$ for an infinite set of indices, from the complementarity condition we have that $-x^k_1 \lambda^*_1 \to 0$, and thus $x^k_1 \lambda^*_2 \exp(x^k_2) \to 0$, which is a contradiction with $-1 + \lambda^*_2 \exp(x^k_2) \to 0$. Similar results are obtained if there is an AGP sequence with $x^k_1 < 0$ (or $x^k_1 = 0$), 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 $G(x_1, x_2) := x_2 h(x_1) \in K = \{0\} \subset \mathbb{R}$, where $h : \mathbb{R} \to \mathbb{R}$ is the $C^1$ function introduced in [11], 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 [11], we see that there exists a sequence $\{x^k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ such that $x^k \to 0$, $h'(z^k) = -z^k \sin(1/z^k)$ and $\sin(1/z^k) \to 1$.

**AGP holds at $\bar{x}$:** First, choose a sequence $x^k := (z^k, 1)$ with $\Lambda := -(z^k)^{-4} \in K^0 = \mathbb{R}$. Now, observe that $\nabla f(x^k) + DG(x^k)\Lambda^k$ goes to zero, because

$$\left(0, 1 + \Lambda^k \frac{h'(z^k)}{h(z^k)} \right) \to \left(0, 1 + \frac{-1}{(z^k)^4} \frac{(-z^k)^5}{(z^k)^4 \sin(1/z^k)} \right) \to \left(0, 1 + \frac{z^k}{-\sin(1/z^k)} \right) \to \left(0, 0 \right). \quad (16)$$

Finally, the approximate complementarity condition trivially holds, since $\Pi_K(G(x^k)) = 0$.

**TAKKT fails at $\bar{x}$:** Suppose that there exists a sequence $\{x^k := (x^k_1, x^k_2)\}_{k \in \mathbb{N}}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ conforming to the definition of TAKKT. Then,

$$\left(0, 1 + \Lambda^k \frac{x^k_2 h'(x^k_1)}{h(x^k_1)} \right) \to \left(0, 0 \right). \quad (17)$$

The approximate complementarity condition of TAKKT implies $\Lambda^k G(x^k_1, x^k_2) = \Lambda^k x^k_2 h(x^k_1) \to 0$. Since $x^k_2 \to 1$, we get $\Lambda^k h(x^k_1) \to 0$, 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 [51]. Indeed, note that AGP holds at $\bar{x}$ with sequences $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^m_+$ if, and only if, $\Lambda^*_k G_i(x^k) \to 0$, whenever $G_i(\bar{x}) \leq 0$. Since both AGP and CAKKT push $\Lambda^*_k$ 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 $\Lambda^*_k$ in this case, it demands it to go to infinity slower than $G_i(\bar{x})$. While AGP may allow a faster growth as long as $\{x^k\}_{k \in \mathbb{N}}$ violates $G_i(\bar{x}) \leq 0$. 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 \( \{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^o \) such that
\[
\langle \Lambda^k, \Pi_{\mathcal{K}^o}(G(x^k)) \rangle \rightarrow 0.
\] (18)

Observe that CAKKT as in Definition 3.3 is indeed a generalization of the condition with the same name from [11] for [NLP], since \( \Lambda^k_i G_i(x^k) \rightarrow 0 \) in this case, independently on the sign of \( G_i(x^k) \). Moreover, in view of Moreau’s Decomposition, (18) is a natural completion to (11) in order to reach (10) even in the general context. Hence, CAKKT clearly implies both, TAKKT and AGP. Though, 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
\[
\begin{align*}
\text{Minimize} \\
_{x \in \mathbb{R}^n} f(x) + \frac{1}{2} \|x - \bar{x}\|^2,
\text{subject to} \\
G(x) \in \mathcal{K}, \\
\|x - \bar{x}\| \leq \delta.
\end{align*}
\]

Let \( \{\rho_k\}_{k \in \mathbb{N}} \rightarrow \infty \) and, for each \( k \in \mathbb{N} \), consider the penalized optimization problem
\[
\begin{align*}
\text{Minimize} \\
_{x \in \mathbb{R}^n} f(x) + \frac{1}{2} \|x - \bar{x}\|^2 + \rho_k \frac{1}{2} \|\Pi_{\mathcal{K}^o}(G(x))\|^2, \\
\text{subject to} \\
\|x - \bar{x}\| \leq \delta.
\end{align*}
\] (19)

Denote by \( x^k \) a global solution of (19). Using analogous arguments as in the standard external penalty algorithm convergence proof [62], 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
\[
\nabla f(x^k) + (x^k - \bar{x}) + DG(x^k)^* \Lambda^k = 0, \quad \text{where} \quad \Lambda^k : = \rho_k \Pi_{\mathcal{K}^o}(G(x^k)) \in \mathcal{K}^o,
\]
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 \( \Lambda^k \) and Lemma 2.1, we see that
\[
\langle \Lambda^k, \Pi_{\mathcal{K}^o}(G(x^k)) \rangle = \langle \rho_k \Pi_{\mathcal{K}^o}(G(x^k)), \Pi_{\mathcal{K}^o}(G(x^k)) \rangle = 0, \quad \text{that is, (11) holds.}
\]

We proceed to show that (13) is satisfied. First, from the optimality of \( x^k \) we have
\[
f(x^k) + \frac{1}{2} \|x^k - \bar{x}\|^2 + \frac{1}{2} \langle \Lambda^k, \Pi_{\mathcal{K}^o}(G(x^k)) \rangle = f(x^k) + \frac{1}{2} \|x^k - \bar{x}\|^2 + \rho_k \frac{1}{2} \|\Pi_{\mathcal{K}^o}(G(x^k))\|^2 \leq f(\bar{x}),
\]
which leads to
\[
0 \leq \langle \Lambda^k, \Pi_{\mathcal{K}^o}(G(x^k)) \rangle = \rho_k \|\Pi_{\mathcal{K}^o}(G(x^k))\|^2 \leq 2(f(\bar{x}) - f(x^k)) - \|x^k - \bar{x}\|^2,
\]
and since \( x^k \rightarrow \bar{x} \) and \( f(x^k) \rightarrow f(\bar{x}) \), we see that \( \langle \Lambda^k, \Pi_{\mathcal{K}^o}(G(x^k)) \rangle \rightarrow 0 \). Thus, CAKKT holds at \( \bar{x} \).

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 \( \mathcal{K}^o \) 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, [16] [17] [18] [52] [64] to cite a few of them. In this section, we will show that the Powell-Hestenes-Rockafellar [53] [55] [61] Augmented Lagrangian algorithm generates AGP sequences without any additional condition, and also CAKKT (and TAKKT) sequences under the so-called generalized Lojasiewicz inequality (see [29] for the definition).
Given $\rho > 0$, let $L_\rho : \mathbb{R}^n \times K^0 \to \mathbb{R}$ be the \textit{Augmented Lagrangian function} of \textsc{NCP}, defined as

$$L_\rho(x, \Lambda) := f(x) + \frac{\rho}{2} \left\| \Pi_{K^0} \left( G(x) + \frac{\Lambda}{\rho} \right) \right\|^2 - \frac{1}{2} \|\Lambda\|^2,$$

whose partial derivative with respect to $x$ is given by

$$\nabla_x L_\rho(x, \Lambda) = \nabla f(x) + DG(x)^* \left( \rho \Pi_{K^0} \left( G(x) + \frac{\Lambda}{\rho} \right) \right). \quad (20)$$

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

\textbf{Algorithm 1} General framework: Augmented Lagrangian

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

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

1. Compute some point $x^k$ such that
   $$\|\nabla_x L_{\rho_k}(x^k, \hat{\Lambda}^k)\| \leq \epsilon_k;$$
   \quad \text{(21)}

2. Update the multiplier
   $$\Lambda^k := \rho_k \Pi_{K^0} \left( G(x^k) + \hat{\Lambda}^k \right),$$
   \quad \text{(22)}
   and compute some $\hat{\Lambda}^{k+1} \in B$ (typically, the projection of $\Lambda^k$ onto $B$);

3. Define
   $$V^k := \frac{\hat{\Lambda}^k}{\rho_k} - \Pi_{K^0} \left( G(x^k) + \frac{\hat{\Lambda}^k}{\rho_k} \right);$$
   \quad \text{(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 $\rho_k \to \infty$ 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 $\|V^k\| \to 0$. With this in mind, we proceed by showing that Algorithm 1 generates sequences whose limit points satisfy AGP.

\textbf{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.

\textbf{Proof.} First, from \textbf{(21)} together with \textbf{(22)} and \textbf{(20)}, we get

$$\|\nabla f(x^k) + DG(x^k)^* \Lambda^k\| \leq \epsilon_k, \quad \text{with } \Lambda^k = \rho_k \Pi_{K^0}(G(x^k) + \rho_k^{-1} \hat{\Lambda}^k),$$

which implies \textbf{(1)} and $\Lambda^k \in K^0$. Taking a subsequence if necessary, we can suppose that $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$. We consider two cases depending on whether the sequence $\{\rho_k\}_{k \in \mathbb{N}}$ is bounded or not:

1. Suppose that $\rho_k \to \infty$. Our aim is to show that $\langle \Lambda^k, \Pi_K(G(x^k)) \rangle$ converges to zero. In fact, we observe from Moreau’s decomposition that
   $$\rho_k G(x^k) + \hat{\Lambda}^k = \Pi_{K^0}(\rho_k G(x^k) + \hat{\Lambda}^k) + \Pi_K(\rho_k G(x^k) + \hat{\Lambda}^k) = \Lambda^k + \Pi_K(\rho_k G(x^k) + \hat{\Lambda}^k).$$
   Thus, $\rho_k G(x^k) = \Lambda^k - \hat{\Lambda}^k + \Pi_K(\rho_k G(x^k) + \hat{\Lambda}^k)$. From this, we see that
   $$\langle \Lambda^k, \Pi_K(G(x^k)) \rangle = \rho_k^{-1} \langle \Lambda^k, \rho_k G(x^k) - \Pi_{K^0}(\rho_k G(x^k)) \rangle = \rho_k^{-1} \langle \Lambda^k, \Lambda^k - \hat{\Lambda}^k - \Pi_{K^0}(\rho_k G(x^k)) \rangle$$
   $$= \rho_k^{-1} \langle \Lambda^k, \Pi_K(\rho_k G(x^k) + \hat{\Lambda}^k) - \Pi_{K^0}(\rho_k G(x^k)) \rangle - \rho_k^{-1} \langle \Lambda^k, \hat{\Lambda}^k \rangle,$$
   \quad \text{(24)}

where in the second line we use Moreau’s decomposition for $\rho_k G(x^k) + \hat{\Lambda}^k$ and also that $\langle \Lambda^k, \Pi_K(\rho_k G(x^k) + \hat{\Lambda}^k) \rangle = 0$, which follows from Lemma 2.1. Now, let us show that the right-hand side of \textbf{(24)} converges to
zero. Indeed, using Cauchy-Schwarz inequality for the first expression of (24) and the nonexpansiveness of the projection we obtain
\[ |\rho_k^{-1} \langle \bar{\Lambda}^k, \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k) \rangle - \Pi_{\mathcal{K}}(\rho_k G(x^k))| + |\rho_k^{-1} \langle \bar{\Lambda}^k, \hat{\Lambda}^k \rangle | \leq 2\|\rho_k^{-1} \bar{\Lambda}^k\| \|\hat{\Lambda}^k\|. \] (25)
but \(\|\rho_k^{-1} \bar{\Lambda}^k\| \|\hat{\Lambda}^k\| = \|\Pi_{\mathcal{K}}(G(x^k) + \rho_k \hat{\Lambda}^k)\| \) converges to zero by the continuity of the projection, and hence (25) converges to zero as well. Thus, \(\langle \bar{\Lambda}^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \rightarrow 0\).

2. If \(\{\rho_k\}_{k \in \mathbb{N}}\) is a bounded sequence, it must be constant for sufficiently large \(k\). Note that
\[ \langle \bar{\Lambda}^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle = \rho_k^{-1} \langle \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k)) \rangle \]
\[ = \rho_k^{-1} \langle \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k) \rangle \]
where in the second equality, we use \(\langle \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k) \rangle = 0\). It remains to show that the right-hand side of (26) goes to zero. In fact, \(\hat{\Lambda}^k = \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k)\) is a bounded sequence and \(\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k)\) converges to zero.

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
\(\Lambda^k - \hat{\Lambda}^k = \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k) - \hat{\Lambda}^k = -\rho_k V^k \rightarrow 0\),
and thus \(\Lambda^k - \rho_k V^k\) characterizes a bounded sequence. To show that \(\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k)\) converges to zero, consider the next expression
\[ \rho_k G(x^k) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k) = \rho_k G(x^k) + \hat{\Lambda}^k - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k) - \hat{\Lambda}^k \]
\[ = \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k) - \hat{\Lambda}^k = \Lambda^k - \hat{\Lambda}^k \rightarrow 0. \] (27)

Using the above expression and the nonexpansivity of the projection onto \(\mathcal{K}\), we get that
\[ \|\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k)\| \leq \|\rho_k G(x^k) - \Pi_{\mathcal{K}}(\rho_k G(x^k) + \hat{\Lambda}^k)\| \rightarrow 0. \] (28)
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 [19], [50], [21] 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. [23]. The AGP condition plays a pivotal role in obtaining global convergence results for IR methods [19], 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 \(\Psi\) 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
\[ |\Psi(x) - \Psi(\bar{x})| \leq \psi(x)\|D\Psi(x)\| \text{ for every } x \in B(\bar{x}, \delta). \] (29)
This property coincides with the inequality with the same name that was proposed in [11]. Such types of property have been extensively used in optimization methods, complexity theory, stability of gradient systems etc. See, for instance, [22], [43], [16], [14], [13], [28], [44] and references therein. 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 (29) for \(\Psi(x) = (1/2)\|\Pi_{\mathcal{K}}(G(x))\|^2\), 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, \(\langle \bar{\Lambda}^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle\) converges to zero. Thus, it suffices to show that \(\langle \bar{\Lambda}^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle\) 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 \(\rho_k\|\Pi_{\mathcal{K}}(G(x^k))\|^2 \rightarrow 0\).

From the generalized Lojasiewicz inequality, there exists some function \(\psi\) such that
\[ \frac{1}{2}\|\Pi_{\mathcal{K}}(G(x^k))\|^2 = |\Psi(x^k)| \leq \psi(x^k)\|D\Psi(x^k)\| = \psi(x^k)\|DG(x^k)\|\Pi_{\mathcal{K}}(G(x^k))\|^2. \]
and we obtain the following inequality
\[ \rho_k \| \Pi_{K^*}(G(x^k)) \|^2 \leq 2\psi(x^k) \| \rho_k DG(x^k) * \Pi_{K^*}(G(x^k)) \|. \]

Now, we proceed by finding an upper bound for the sequence \[ \| \rho_k DG(x^k) * \Pi_{K^*}(G(x^k)) \|: \]
\[ \| \rho_k DG(x^k) * \Pi_{K^*}(G(x^k)) \| \leq \| DG(x^k) * (\Lambda^k - \rho_k \Pi_{K^*}(G(x^k))) \| + \| DG(x^k) * \Lambda^k \| \]
\[ \leq \| DG(x^k) * |\Lambda^k - \rho_k \Pi_{K^*}(G(x^k))\| + \| DG(x^k) * \Lambda^k \| \]
\[ \leq \| DG(x^k) * (\Pi_{K^*}(\rho_k G(x^k) + \Lambda^k) - \Pi_{K^*}(\rho_k G(x^k))) \| + \| DG(x^k) * \Lambda^k \|. \]

Furthermore, from \[ \| \Pi_{K^*}(\rho_k G(x^k) + \Lambda^k) - \Pi_{K^*}(\rho_k G(x^k)) \| \leq \| \hat{\Lambda}^k \|, \]
we see that
\[ \| \rho_k DG(x^k) * \Pi_{K^*}(G(x^k)) \| \leq \| DG(x^k) * \| \hat{\Lambda}^k \| + \| DG(x^k) * \Lambda^k \| \]
\[ \leq \| DG(x^k) * \| \hat{\Lambda}^k \| + \| \nabla f(x^k) \| + \varepsilon_k, \quad (30) \]
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, \[ (30) \] is bounded by some scalar \( M > 0 \), due to the continuity of \( DG \) and \( \nabla f \) near \( \hat{x} \). Thus, \[ \rho_k \| \Pi_{K^*}(G(x^k)) \|^2 \leq 2\psi(x^k)M. \] Using the fact that \( \psi(x) \to 0 \), we get that \( \rho_k \| \Pi_{K^*}(G(x^k)) \|^2 \to 0 \). We proceed by computing \( \langle \Lambda^k, \Pi_{K^*}(G(x^k)) \rangle \). Indeed,
\[ \langle \Lambda^k, \Pi_{K^*}(G(x^k)) \rangle = \langle \Pi_{K^*}(\rho_k G(x^k) + \hat{\Lambda}^k), \Pi_{K^*}(G(x^k)) \rangle \]
\[ = \langle \Pi_{K^*}(\rho_k G(x^k) + \hat{\Lambda}^k) - \Pi_{K^*}(\rho_k G(x^k)), \Pi_{K^*}(G(x^k)) \rangle + \langle \Pi_{K^*}(\rho_k G(x^k)), \Pi_{K^*}(G(x^k)) \rangle \]
\[ = \langle \Pi_{K^*}(\rho_k G(x^k) + \hat{\Lambda}^k) - \Pi_{K^*}(\rho_k G(x^k)), \Pi_{K^*}(G(x^k)) \rangle + \rho_k \| \Pi_{K^*}(G(x^k)) \|^2 \quad \text{ (31)} \]

Since \( \rho_k \| \Pi_{K^*}(G(x^k)) \|^2 \to 0 \), we only need to show that the first expression of \( (31) \) goes to zero. Now, since \( \Pi_{K^*}(G(x^k)) \to \Pi_{K^*}(G(\hat{x})) = 0 \) and from the boundedness of \{\( \hat{\Lambda}^k \)\}, we get that:
\[ \langle \Pi_{K^*}(\rho_k G(x^k) + \hat{\Lambda}^k) - \Pi_{K^*}(\rho_k G(x^k)), \Pi_{K^*}(G(x^k)) \rangle \leq \| \hat{\Lambda}^k \| \| \Pi_{K^*}(G(x^k)) \| \to 0. \]
Thus \( \langle \hat{\Lambda}^k, \Pi_{K^*}(G(x^k)) \rangle \to 0 \), and as a consequence CAKKT holds at \( \hat{x} \).

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 \( \langle \Lambda^k, \Pi_{K^*}(G(x^k)) \rangle \) goes to zero, since \( \Pi_{K^*}(G(x^k)) \to \Pi_{K^*}(G(\hat{x})) = 0 \).

In both cases, we have shown that \( \hat{x} \) is a CAKKT point associated with \{\( x^k \)\}_{k \in \mathbb{N}}. \]

The Augmented Lagrangian method presented in Algorithm 1 for NLSDP was proven to generate TAKKT sequences under generalized Lojasiewicz [3] 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.

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 \( \hat{x} \) of a sequence \{\( x^k \)\}_{k \in \mathbb{N}} \] generated by Algorithm 1 is a stationary point of
\[ \min_{x \in \mathbb{R}^n} \| \Pi_{K^*}(G(x)) \|^2. \quad \text{ (32)} \]

**Proof.** Taking a subsequence if necessary, suppose that \( x^k \to \hat{x} \). If \{\( \rho_k \)\}_{k \in \mathbb{N}} \] is bounded, it must converge to some \( \bar{\rho} \) and then \( \hat{\Lambda}^k \to \hat{\Lambda} \). Also, in this case \( \nabla f \to 0 \), which means \( \hat{\Lambda} = \Pi_{K^*}(\hat{\Lambda} + \bar{\rho} G(\hat{x})) \). Then, by Lemma 2.1 item 1, we get \( \bar{\rho} G(\hat{x}) \in K \), so \( \Pi_{K^*}(G(\hat{x})) = 0 \) and \( \hat{x} \) is a global solution of \( (32) \). On the other hand, if \{\( \rho_k \)\}_{k \in \mathbb{N}} \] is unbounded, by Step 1 and \( (20) \), note that
\[ \frac{1}{\rho^k} (\nabla f(x^k) + DG(x^k) \Lambda^k) = \frac{\nabla f(x^k)}{\rho^k} + DG(x^k) \Pi_{K^*} \left( \frac{\hat{\Lambda}^k}{\rho^k} + G(x^k) \right) \to 0. \]
Then, since \( \nabla f(x^k) \to \nabla f(\hat{x}) \) and \{\( \hat{\Lambda}^k \)\}_{k \in \mathbb{N}} \] is bounded, we obtain \( DG(\hat{x}) * \Pi_{K^*}(G(\hat{x})) = 0 \), which means \( \hat{x} \) is a stationary point of \( (32) \). \]

**Remark 2.** We highlight that our convergence theory for Algorithm 1 allows the set \( K \) 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
\[ \mathcal{E}^m := \{ M \in \mathbb{S}^m \mid \exists p_1, \ldots, p_m \in \mathbb{R}^r, \forall i, j \in \{1, \ldots, m\}, M_{ij} = \| p_i - p_j \|^2 \}. \]
5 The AKKT condition

Until this point, we have presented generalizations of CAKKT and AGP via projections onto $K$ and $\mathcal{K}^o$, but we have not addressed yet a generalization of AKKT, which is the most natural and simple condition in NLP. 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 $\bar{\Lambda} \in \mathcal{K}^o$ in the form of a generalized equation (in the sense of Robinson [60]):

$$
\left( \nabla f(\bar{x}) + DG(\bar{x})^* \bar{\Lambda} \right) \in \left[ Y \in K \mid \langle Y, \bar{\Lambda} \rangle = 0 \right].
$$

Given some $\varepsilon > 0$, the standard perturbation theory (see, for example, [37, 38, 23]) can be used as inspiration to say that a point $x \in B(\bar{x}, \varepsilon)$ satisfies the KKT conditions with error $\varepsilon$ when there is a multiplier $\Lambda \in \mathcal{K}^o$ and some perturbation vector $\xi \in \mathbb{R}^n \times E$ such that $F(x, \Lambda) + \xi \in \mathcal{N}(\Lambda)$ 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 $\{y^k\}_{k \in \mathbb{N}} \to 0$, $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^o$ such that (29) holds, $G(x^k) + y^k \in \mathcal{K}$, and

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

It turns out that Definition 5.1 coincides with [59, 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 NLSDP or the NSOCP contexts, it is consistent with the existing concepts with the same name from [3, Section 2], [8, Definition 3.1], and [2, Definition 3.3], respectively. This is clarified in Subsection 6.4.

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 $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^o$ and $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$ satisfying (29) and (11). Denote by $y^k$ the global solution of

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

(34)

and consider the feasible point $y := -\Pi_{\mathcal{K}}(G(x^k))$ of (34). Then,

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

(35)
Taking \( k \to \infty \) in \([35]\), since \( \Pi_{\mathcal{K}}(G(x^k)) \to 0 \) and AGP holds, we see that \( y^k \to 0 \) and \( \langle \Lambda^k, G(x^k) + y^k \rangle \to 0 \), because \( \Lambda^k \in \mathcal{K}^0 \) and \( G(x^k) + y^k \in \mathcal{K} \). Now, the necessary optimality condition for \( y^k \) in \([34]\) implies
\[
0 \in y^k - \Lambda^k + \{ \Theta \in \mathcal{K}^0 \mid \langle \Theta, G(x^k) + y^k \rangle = 0 \}.
\]
Now, taking \( \{ y^k \}_{k \in \mathbb{N}} \) and choosing \( \hat{\Lambda}^k := \Lambda^k - y^k \) for all \( k \in \mathbb{N} \), AKKT holds.

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, \([8, \text{Example 5.2}] \) and \([2, \text{Example 3.1}] \) show that TAKKT satisfy AKKT. In this case, without loss of generality, we can assume that Algorithm 1 generates sequences as certificates of approximate optimality along with the Lagrangian residue. For example, it is easy to verify 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
\[
\langle \Lambda^k, G(x^k) + y^k \rangle = \rho_k \left( \Pi_{\mathcal{K}^0} \left( G(x^k) + \frac{\hat{\Lambda}^k}{\rho_k} \right) , \left( G(x^k) + \frac{\hat{\Lambda}^k}{\rho_k} \right) - \Pi_{\mathcal{K}^0} \left( G(x^k) + \frac{\hat{\Lambda}^k}{\rho_k} \right) \right) = 0,
\]
by Lemma 2.1. Condition \([9]\) follows directly from Step 1, since \( \nabla_x L_{\rho^k}(x^k, \Lambda^k) = \nabla f(x^k) + D G(x^k)^* \Lambda^k \) for the multiplier choice of Step 2. Also, if \( \{ \rho_k \}_{k \in \mathbb{N}} \) is unbounded, it follows that \( y^k \to -\Pi_{\mathcal{K}^0}(G(\bar{x})) = 0 \); but if it is bounded, then \( \| y^k \| \to 0 \) 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 \( \mathcal{K} = \mathcal{K}_1 \times \cdots \times \mathcal{K}_r \) and \( \mathbb{E} = \mathbb{E}_1 \times \cdots \times \mathbb{E}_m \) where \( \mathcal{K}_i \subseteq \mathbb{E}_i \) for all \( i \in \{ 1, \ldots, r \} \), 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 \((\text{NCP})\). If \( \bar{x} \) satisfies AKKT, then there are sequences \( \{ x^k \}_{k \in \mathbb{N}} \to \bar{x} \) and \( \{ \Lambda^k \}_{k \in \mathbb{N}} \subset \mathcal{K}^0 \) such that \([9]\) holds and \( \Lambda^k = 0 \) 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 \( \{ (y_1, \ldots, y_r) \}_{k \in \mathbb{N}} \to 0 \), \( \{ \Lambda^k \}_{k \in \mathbb{N}} \to \bar{x} \) and \( \{ (\Lambda_1, \ldots, \Lambda_r)^k \}_{k \in \mathbb{N}} \subset \mathcal{K}^0 \). Proving \([9]\) is trivial. Then, for every index \( i \in \{ 1, \ldots, r \} \) 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 \( k \), there are some \( \delta_i^k > 0 \) such that \( z_i^k + \alpha \Lambda_i^k \in \text{int} \mathcal{K}_i \) for every \( \alpha \in (-\delta_i^k, \delta_i^k) \) when \( G_i(x^k) \in \text{int} \mathcal{K}_i \). Define \( \delta_i := \min \{ \delta_i^k \} \) and we proceed by contradiction: suppose that \( \Lambda_i^k \neq 0 \) for some \( j \) such that \( G_j(x^k) \in \text{int} \mathcal{K}_j \) and some \( k > k_0 \). Then, since \( \langle z_i^k, \Lambda_i^k \rangle = 0 \), we have \( \langle z_j^k + \alpha \Lambda_j^k, \Lambda_i^k \rangle = \alpha \| \Lambda_j^k \|^2 \), whose sign depends on \( \alpha \in (-\delta, \delta) \). On the other hand, it is possible to choose \( \alpha \in (0, \delta) \) such that the point \( w_j := z_j^k + \alpha \Lambda_j^k \in \mathcal{K}_j \). Hence, it satisfies \( \langle w_j, \Lambda_i^k \rangle > 0 \), which means \( \Lambda_i^k \notin \mathcal{K}_i^0 \). This contradiction implies \( \Lambda_i^k = 0 \).

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 \((\text{NCP})\) as well.

---

![Figure 1: Relationship of the new sequential optimality conditions.](image-url)
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:

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

(\text{NLP})

The most straightforward way of viewing (\text{NLP}) as a particular case of (\text{NCP}) is directly phrasing the constraints as \((h(x), g(x)) \in \{0\}^p \times \mathbb{R}^q\), but the conclusions that can be taken from this approach may be weaker than expected. For avoiding this issue, we consider an auxiliary formulation of (\text{NLP}), where equality constraints are treated as two inequalities:

\[
\begin{align*}
\text{Minimize} & \quad f(x), \\
\text{subject to} & \quad G(x) := (h(x), -h(x), g(x)) \in \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^q.
\end{align*}
\]  

(\text{NLP-Auxiliary})

The Lagrange multipliers are then written as

\[\Lambda^k = (\omega^k_1, \ldots, \omega^k_p, \omega^k_{p+1}, \ldots, \omega^k_{p+q}) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q.\]

Our next result concerns about the specializations of CAKKT, AGP, and AKKT, from (\text{NCP}) to (\text{NLP}), via (\text{NLP-Auxiliary}).

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

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

\[
\nabla f(x^k) + \sum_{i=1}^p \omega_i^k \nabla h_i(x^k) + \sum_{j=1}^q \mu_j^k \nabla g_j(x^k) \to 0;
\]  

(36)

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

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

Proof.

(a) AKKT: If AKKT holds for (\text{NLP-Auxiliary}), there exist sequences \(\{x^k\}_{k \in \mathbb{N}} \to \bar{x}\), \(\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{K}^n\), and \(\{Y^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^q\), where \(Y^k = (z^k_+, z^k_-, y^k) \to 0\), 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

\[
G(x^k) + Y^k, \Lambda^k = \sum_{i=1}^p \omega^k_i (h_i(x^k) + z^k_i) + \sum_{i=1}^p \omega^k_i (-h_i(x^k) + z^k_i) + \sum_{j=1}^q \mu_j^k (g_j(x^k) + y_j^k) = 0.
\]  

(37)

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

Conversely, consider the following natural choice: \(Y^k := (\bar{h}(x^k), h(x^k), \bar{g}_j(x^k)) \to 0\), where

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

Then, for sufficiently large \(k\), we have (37) and \(G(x^k) + Y^k \in \mathcal{K}\), AKKT is satisfied with \(\omega^k_+ := \max\{0, \omega^k_+\}\) and \(\omega^k_{i-} := -\min\{0, \omega^k_{i-}\}\).
(b) **AGP:** In this case, \( \langle \Lambda^k, \Pi_K(G(x^k)) \rangle \to 0 \) is equivalent to saying that
\[
\sum_{i=1}^p \omega_{i+}^k \min\{0, h_i(x^k)\} + \sum_{i=1}^p \omega_{i-}^k \min\{0, -h_i(x^k)\} + \sum_{j=1}^q \mu_j^k \min\{0, g_j(x^k)\} \to 0.
\]
Since each part of the sum has the same sign, we get that \( \omega_{i+}^k \min\{0, -h_i(x^k)\} \to 0 \), \( \omega_{i-}^k \min\{0, h_i(x^k)\} \to 0 \), and \( \mu_j^k \min\{0, g_j(x^k)\} \to 0 \). Hence, we have that \( \mu_j^k \min\{0, g_j(x^k)\} \to 0 \) for every \( j \in \{1, \ldots, q\} \) and, defining \( \omega_c^k := \omega_{i+}^k - \omega_{i-}^k \), we observe that
\[
\omega_c^k h_i(x^k) = \omega_{i+}^k \min\{0, h_i(x^k)\} - \omega_{i-}^k \min\{0, h_i(x^k)\} - \omega_{i+}^k \min\{0, -h_i(x^k)\} + \omega_{i-}^k \min\{0, -h_i(x^k)\}.
\]
Then, note that the first and the last term of the right-hand side of \( \text{(38)} \) both vanish, and the middle terms are nonnegative. Hence, \( \liminf_{k \to \infty} \omega_c^k h_i(x^k) \geq 0 \), for every \( i \in \{1, \ldots, p\} \).

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

(c) **CAKKT:** Here, we get
\[
\langle \Lambda^k, \Pi_K(G(x^k)) \rangle = \langle \omega_c^k, \min\{h(x^k), 0\} \rangle + \langle \omega_c^k, \min\{-h(x^k), 0\} \rangle + \langle \mu^k, \min\{g(x^k), 0\} \rangle
\]
and
\[
\langle \Lambda^k, \Pi_{K^0}(G(x^k)) \rangle = \langle \omega_c^k, \max\{h(x^k), 0\} \rangle + \langle \omega_c^k, \max\{-h(x^k), 0\} \rangle + \langle \mu^k, \max\{g(x^k), 0\} \rangle,
\]
so if both tend to zero, define \( \hat{\omega}_c^k := \omega_c^k - \omega_c^k \) and we obtain \( \hat{\omega}_c^k h_i(x^k) \to 0 \) and \( \mu_j^k g_j(x^k) \to 0 \).

The converse is analogous to the previous item, with the same choice of multipliers: \( \omega_{i+}^k := \max\{0, \omega_c^k\} \) and \( \omega_{i-}^k := -\min\{0, \omega_c^k\} \).

Note that from this point of view, our definitions are consistent with the original AKKT from \[3\], the AGP from \[51\] (which follows from Theorem 3.1), and the CAKKT from \[11\], respectively.

We summarize our results in the following table:

<table>
<thead>
<tr>
<th>Approximate complementarity condition</th>
<th>AKKT</th>
<th>AGP</th>
<th>CAKKT</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g_j(x) &lt; 0 \Rightarrow \mu_j^k = 0 ), ( \forall j ) and ( k ) sufficiently large.</td>
<td>( \mu_j^k \min{g_j(x^k), 0} \to 0 ) and ( \liminf_{k \to \infty} \omega_c^k h_i(x^k) \geq 0 ), ( \forall i, \forall j ).</td>
<td>( \mu_j^k g_j(x^k) \to 0 ) and ( \omega_c^k h_i(x^k) \to 0 ), ( \forall i, \forall j ).</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Sequential conditions when specialized to \( \text{NLP} \).

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:

\[
\begin{array}{ll}
\text{Minimize} & f(x), \\
\text{subject to} & G(x) \in S^n, \\
\end{array}
\]

which is \( \text{NCP} \) with \( E = S^n \), \( A, B := \text{tr}(AB) \) is the (Frobenius) inner product given by the trace of \( AB \), and \( K = S^n_+ \). We recall that \( K^0 = -K \) and every symmetric matrix \( A \in E \) has a spectral decomposition, that is, there exits an orthogonal matrix \( U \) such that \( A = UDU^T \), where \( D = \text{Diag}(\lambda_1^U(A), \ldots, \lambda_m^U(A)) \) is a diagonal matrix and \( \lambda_i^U(A) \) are eigenvalues of \( A \) ordered according to the eigenvectors in the columns of \( U \). Moreover, \( \Pi_K(A) = UD(\min\{\lambda_1^U(A), 0\}, \ldots, \min\{\lambda_m^U(A), 0\})U^T \) and a similar relation holds for \( \Pi_{K^0}(A) \) with \( \max \) instead of \( \min \).
When no order is specified, we consider $\lambda_1(A) \leq \cdots \leq \lambda_m(A)$. Note that for every $i \in \{1, \ldots, m\}$ we have $\lambda_i(-A) = -\lambda_{m-i+1}(A)$. Also, the following inequality is important to our analyses: For every $A, B \in S^m_+$, we have the inequality
\[
\sum_{i=1}^{m} \lambda_i(A)\lambda_{m-i+1}(B) \leq \text{tr}(AB) \leq \sum_{i=1}^{m} \lambda_i(A)\lambda_i(B).
\] (39)

For its proof, see [48].

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 $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$, $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{K}^o$, and a sequence of orthogonal matrices $S^k \to U$, where $U$ diagonalizes $G(\bar{x})$ and each $S^k$ diagonalizes $\Lambda^k$, such that [9] holds and $\lambda_{S^k}(\Lambda^k) = 0$, if $\lambda_U(G(\bar{x})) < 0$, for sufficiently large $k$.

(b) AGP implies that, for every $i$, $\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0$, for some sequences $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{K}^o$ such that [9] holds;

(c) CAKKT implies that, for every $i$, $\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0$ and $\min\{0, -\lambda_{m-i+1}(G(x^k))\}\lambda_i(\Lambda^k) \to 0$, for some sequences $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{K}^o$ such that [9] holds;

**Proof.**

(a) **AKKT:** If there is a sequence $\{y^k\}_{k \in \mathbb{N}} \to 0$ such that $G(x^k) + y^k \in \mathbb{K}$ and $(\Lambda^k, G(x^k) + y^k) = 0$ for every $k$, the latter implies that $G(x^k) + y^k$ and $\Lambda^k$ are simultaneously diagonalizable, that is, for every $k$, there is a matrix $\bar{S}^k$ such that $G(x^k) + y^k = \bar{S}^k\Theta^k(\bar{S}^k)^T$ and $\Lambda^k = \bar{S}^k\Theta^k(\bar{S}^k)^T$, where $\Theta^k = \text{Diag}(\lambda_{S^k}(G(x^k) + y^k))$ and $\Gamma^k = \text{Diag}(\lambda_{U}(\Lambda^k))$. The continuity of $G$ and the convergence of $\{x^k\}_{k \in \mathbb{N}}$ imply that $S^k \to U$, for some orthogonal matrix $U$. Then $U$ diagonalizes $G(\bar{x})$ since $G(x^k) + y^k \to G(\bar{x})$, and if $\lambda_U(G(\bar{x})) < 0$, then for sufficiently large $k$ we have $\lambda_{S^k}(G(x^k) + y^k) < 0$, as well. Then $\lambda_{S^k}(\Lambda^k) = 0$ for those $k$.

Conversely, let $\bar{x}$ be a feasible point associated with the sequences $\{\Lambda^k\}_{k \in \mathbb{N}}$, $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$, and $\{S^k\}_{k \in \mathbb{N}} \to U$, where each $S^k$ diagonalize $\Lambda^k$ and $U$ diagonalizes $G(\bar{x})$, such that $\lambda_{S^k}(\Lambda^k) = 0$ whenever $\lambda_U(G(\bar{x})) < 0$. 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 $\alpha$ negative eigenvalues and $\beta$ zero eigenvalues in $G(\bar{x})$, with $\alpha + \beta = m$, and for some indexed matrix $A^k$, define $\bar{A}^k := (S^k)^T A^k S^k$. Then, choose
\[
y^k := S^k\begin{bmatrix} 0 & \bar{G}(x^k)_{\alpha \beta} \\ \bar{G}(x^k)_{\beta \alpha} & \bar{G}(x^k)_{\beta \beta} \end{bmatrix}(S^k)^T, \text{ where } \bar{G}(x^k) = \begin{bmatrix} \bar{G}(x^k)_{\alpha \alpha} & \bar{G}(x^k)_{\alpha \beta} \\ \bar{G}(x^k)_{\beta \alpha} & \bar{G}(x^k)_{\beta \beta} \end{bmatrix}
\]
and the partition $\alpha \alpha$ refers to the first $\alpha$ rows of the matrix, for example. Then, if $\Gamma^k := \bar{\Lambda}^k$, we get
\[
(\bar{A}^k, G(x^k) + y^k) = (\Gamma^k, \bar{G}(x^k) + \bar{y}^k) = 0,
\]
for sufficiently large $k$, because $\Gamma^k_{\alpha \alpha} = 0$ for large enough $k$. Since every block of $\bar{G}(x^k)$ converges to zero, except for $\bar{G}(x^k)_{\alpha \alpha}$, we know that $y^k \to 0$. Also, since $\bar{G}(x^k) + \bar{y}^k \in S^m_+$, we get $G(x^k) + y^k \in S^m_+$.

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

(c) **CAKKT:** Similarly to the previous item, from $\langle \Lambda^k, \Pi_K(G(x^k)) \rangle \to 0$ we get $\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0$. Also, from $\langle \Lambda^k, \Pi_{K^o}(G(x^k)) \rangle \to 0$, using [39], we can obtain $\lambda_i(-\Pi_{K^o}(G(x^k)))\lambda_i(\Lambda^k) \to 0$, for every $i \in \{1, \ldots, m\}$, which is equivalent to $\lambda_i(\Pi_{K^o}(-G(x^k)))\lambda_i(\Lambda^k) \to 0$. Then,
\[
\min\{0, -\lambda_{m-i+1}(G(x^k))\}\lambda_i(\Lambda^k) \to 0.
\]
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 $\Lambda_k^i$ and $G(x^k)$ commute, for every $k$.

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

<table>
<thead>
<tr>
<th>Approximate complementarity condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AKKT</strong></td>
<td>$\lambda_i^U(G(\bar{x})) &lt; 0 \Rightarrow \lambda_i^{U*}(\Lambda^k) = 0,$ for all large, where $U^k$ diag. $\Lambda^k$, $U$ diag. $G(\bar{x})$, and $U^k \rightarrow U$.</td>
<td></td>
</tr>
<tr>
<td><strong>AGP</strong></td>
<td>$\min{0, \lambda_i(G(x^k))} \lambda_i(\Lambda^k) \rightarrow 0,$ for all large.</td>
<td></td>
</tr>
<tr>
<td><strong>CAKKT</strong></td>
<td>$\min{0, \lambda_i(G(x^k))} \lambda_i(\Lambda^k) \rightarrow 0$ and $\max{0, \lambda_{m+1+1}(G(x^k))} \lambda_i(\Lambda^k) \rightarrow 0$, for all large.</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2: Sequential optimality conditions when specialized to (NLSDP). Recall that for $C \in \mathbb{S}^m$, the symbols $\lambda_1(C), \ldots, \lambda_m(C)$ 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 NLDSP, it can have different meanings. As mentioned in the Introduction, the authors of [8] 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 NLDSP and AKKT is consistent with the definition presented in [8, 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 [15].

### 6.3 Nonlinear second-order cone programming

Consider the following particular case of (NSOCP):

**Minimize**

$$f(x),$$

**subject to**

$$G_i(x) \in K_i, i \in \{1, \ldots, r\},$$

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

$$K_i := \{(z_0, \tilde{z}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} \mid \|	ilde{z}\| \leq z_0\}, \ i \in \{1, \ldots, r\}.$$

Denote $G(x) := (G_1(x), \ldots, G_r(x))$ and $K := K_1 \times \cdots \times K_r$. From [32], the interior and the boundary of $K_i$ are described by

$$\text{int} K_i := \{(z_0, \tilde{z}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} \mid \|	ilde{z}\| < z_0\},$$

$$\text{bd} K_i := \{(z_0, \tilde{z}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} \mid \|	ilde{z}\| = z_0\}.$$  

Moreover, consider the following sets of indices:

$$I_{\text{int}} := \{i \in \{1, \ldots, r\} \mid G_i(\bar{x}) \in \text{int} K_i\},$$

$$I_{\text{bd}} := \{i \in \{1, \ldots, r\} \mid G_i(\bar{x}) \in \text{bd} K_i \setminus \{0\}\}. \quad (40)$$

It is well-known that $K_0 := -K_i, \forall i$, and hence $K_0 = -K$. In [32], there is a formula for the projection onto a single Lorentz cone $K_i$. Following [32], every $v = (v_0, \tilde{v}) \in \mathbb{R} \times \mathbb{R}^{m_i-1}$ can be decomposed as

$$v = \mu_1(v)c_1(v) + \mu_2(v)c_2(v), \quad (41)$$

where $\mu_2(v) \in \mathbb{R}$ and $c_\ell(v) \in K_\ell, \ell \in \{1, 2\}$, are given by the following expressions:

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

where $w$ is any unitary vector in $\mathbb{R}^{m-1}$. 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 $v = (v_0, \tilde{v}) \in \mathbb{R} \times \mathbb{R}^{m_i-1}$, we have $\Pi_{K_i}(v) := \max\{\mu_1(v), 0\}c_1(v) + \max\{\mu_2(v), 0\}c_2(v)$.
Proof. Using $\mathcal{K}_i^o = - \mathcal{K}_i$, we get $\Pi_{\mathcal{K}_i^c}(v) := \min\{\mu_1(v), 0\}c_1(v) + \min\{\mu_2(v), 0\}c_2(v)$. Finally, for $v = (v_1, \ldots, v_r) \in \mathbb{R}$ with $v_i \in \mathbb{R} \times \mathbb{R}^{m-1}$, we have that $\Pi_{\mathcal{K}_i^c}(v) = (\Pi_{\mathcal{K}_i(v_1)}, \ldots, \Pi_{\mathcal{K}_i(v_r)})$ and $\Pi_{\mathcal{K}_i^c}(v) = (\Pi_{\mathcal{K}_i^c}(v_1), \ldots, \Pi_{\mathcal{K}_i^c}(v_r))$.

We recall from Lemma 6.2.3 that, for every $z_i, v_i \in \mathbb{R}^m$,

$$\mu(z_i)\mu_2(v_i) + \mu_2(z_i)\mu_1(v_i) \leq 2(z_i, v_i) \leq \mu_1(z_i)\mu_1(v_i) + \mu_2(z_i)\mu_2(v_i).$$

(43)

Now, let $\bar{x}$ be a feasible point of $\text{NSOCP}$, and $(x^k)_{k \in \mathbb{N}}$ be a sequence with $x^k \to \bar{x}$ associated with some sequential optimality condition with multipliers $\{\Lambda^k\}_{k \in \mathbb{N}}$, then (44) can be stated as

$$\nabla f(x^k) + \sum_{i=1}^r DG_i(x^k)^T \lambda^k_i \to 0,$$

(44)

where $\Lambda^k = (\lambda^k_1, \ldots, \lambda^k_r) \in \mathcal{K}_i^o \times \cdots \times \mathcal{K}_i^o$.

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 $\text{NSOCP}$. Then $\bar{x}$ satisfies:

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

$$\frac{\lambda^k_i}{\|\lambda^k_i\|} \to \frac{\bar{G}_i(\bar{x})}{\|\bar{G}_i(\bar{x})\|};$$

(45)

(b) **AGP** implies that $\mu_i(\lambda^k_i) \max\{\mu_i(G_i(x^k)), 0\} \to 0$ for $i \in \{1, 2\}$, and for every $i \in \{1, \ldots, r\}$, for some sequences $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$ and $\{\lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}_i^o$ such that (44) holds;

(c) **CAKKT** implies that $\mu_i(\lambda^k_i) \max\{\mu_i(G_i(x^k)), 0\} \to 0$, $\ell \in \{1, 2\}$ and $\mu_i(\lambda^k_i) \min\{\mu_i(G_i(x^k)), 0\} \to 0$, $\ell, s \in \{1, 2\}$ ($\ell \neq s$), for every $i \in \{1, \ldots, r\}$, for some sequences $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$ and $\{\lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}_i^o$ such that (44) holds.

Proof.

(a) **AKKT:** If there are sequences $\{y^k\}_{k \in \mathbb{N}} \to 0$ such that $z^k_i := G_i(x^k) + y^k_i \in \mathcal{K}_i$ for every $i$ and

$$\sum_{i=1}^r (\lambda^k_i, G_i(x^k) + y^k_i) = 0,$$

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

$$0 = (\lambda^k_i, z^k_i) = (\lambda^k_i, z^k_i) + (\lambda^k_i)_0(z^k_i)_0 \leq \|\lambda^k_i\| \|z^k_i\| + (\lambda^k_i)_0(z^k_i)_0 \leq (\|\lambda^k_i\| + (\lambda^k_i)_0(z^k_i)_0),$$

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

$$1 = \lim_{k \to \infty} \left( \frac{\lambda^k_i}{(\lambda^k_i)_0}, \frac{z^k_i}{(z^k_i)_0} \right) = \lim_{k \to \infty} \left( \frac{\lambda^k_i}{\|\lambda^k_i\|}, \frac{\bar{G}_i(\bar{x})}{\|\bar{G}_i(\bar{x})\|} \right) = \lim_{k \to \infty} \left( \frac{\lambda^k_i}{\|\lambda^k_i\|}, \frac{\bar{G}_i(\bar{x})}{\|\bar{G}_i(\bar{x})\|} \right),$$

which means $\lambda^k_i/\|\lambda^k_i\| \to \bar{G}_i(\bar{x})/\|\bar{G}_i(\bar{x})\|$. 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 $\lambda^k_i$ such that $i \in \mathcal{T}_{\text{bd}^+}$ and $\lambda^k_i \to 0$, is equal to zero for $k$ sufficiently large, and set

$$y^k_i := \begin{cases} -G_i(x^k), & \text{if } G_i(\bar{x}) = 0, \\ -\Pi_{\mathcal{K}_i^c}(G_i(x^k)), & \text{if } i \in \mathcal{T}_{\text{int}} \text{ or } i \in \mathcal{T}_{\text{bd}^+} \text{ with } \lambda^k_i \to 0, \\ \frac{\|\bar{G}_i(\bar{x})\|}{(\lambda^k_i)_0} (\lambda^k_i)_0, -\lambda^k_i) - G_i(x^k), & \text{if } i \in \mathcal{T}_{\text{bd}^+} \text{ with } \lambda^k_i \neq 0. \end{cases}$$
Then, we have $G_i(x^k) + y_i^k \in K_i$, for every $i \in \{1, \ldots, r\}$, because $((\lambda_i^k)_0, -\lambda_i^k) \in K^0$ and $(\lambda_i^k)_0 \leq 0$. If $G_i(\bar{x}) \in \{0\} \cup \text{int } K_i$, we clearly have $\langle G_i(x^k) + y_i^k, \lambda_i^k \rangle = 0$ for $k$ sufficiently large. If $i \in I_{\text{bd}+}$, since $-\lambda_i^k \in \text{bd } K_i$, we get
\[
\langle G_i(x^k) + y_i^k, \lambda_i^k \rangle = \|G_i(x^k)\| \left(\frac{\langle \lambda_i^k, \lambda_i^k \rangle}{\langle \lambda_i^k, \lambda_i^k \rangle}_0 - \frac{\|\lambda_i^k\|^2}{\|\lambda_i^k\|^2}\right) = 0.
\]
Also, note that
\[
(y_i^k)_0 = \|G_i(x^k)\| - (G_i(x^k))_0 \quad \text{and} \quad \tilde{y}_i^k = \frac{\|G_i(x^k)\| \lambda_i^k}{\|\lambda_i^k\|} - \tilde{G}_i(x^k),
\]
so in case $\lambda_i^k \neq 0$ we get $y_i^k \to 0$ from $G_i(\bar{x}) \in \text{bd } K_i \setminus \{0\}$ and (45).

(b) AGP: In this case, since
\[
\langle \lambda_i^k, \Pi_{K_i}(G(x^k)) \rangle = \sum_{i=1}^{r} \langle \lambda_i^k, \Pi_{K_i}(G_i(x_i^k)) \rangle \to 0
\]
and $K^0 = -K$, we obtain $\langle \lambda_i^k, \Pi_{K_i}(G(x^k)) \rangle \to 0$ for every $i \in \{1, \ldots, r\}$. Now, using the spectral decomposition (41) and the right-hand side of (43), we obtain
\[
2(\lambda_i^k, \Pi_{K_i}(G(x^k))) \leq \mu_1(\lambda_i^k) \mu_1(\Pi_{K_i}(G(x^k))) + \mu_2(\lambda_i^k) \mu_2(\Pi_{K_i}(G(x^k))) \leq 0.
\]
Hence, taking $k \to \infty$ in the above expression, we see that:
- $\mu_1(\lambda_i^k) \mu_1(\Pi_{K_i}(G(x^k))) = \mu_1(\lambda_i^k) \max \{\mu_1(G_i(x^k)), 0\} \to 0$;
- $\mu_2(\lambda_i^k) \mu_2(\Pi_{K_i}(G(x^k))) = \mu_2(\lambda_i^k) \max \{\mu_2(G_i(x^k)), 0\} \to 0$.

Furthermore, from above and since $\mu_1(\lambda_i^k) \leq \mu_2(\lambda_i^k) \leq 0$, we see that $\mu_2(\lambda_i^k) \max \{\mu_1(G_i(x^k)), 0\} \to 0$. Thus, $\langle \lambda_i^k, \Pi_{K_i}(G_i(x^k)) \rangle \to 0$ implies $\mu_i(\lambda_i^k) \max \{\mu_i(G_i(x^k)), 0\} \to 0$ for $\ell \in \{1, 2\}$ and $i \in \{1, \ldots, r\}$.

(c) CAKKT: Again, $\langle \lambda_i^k, \Pi_{K_i}(G(x^k)) \rangle \to 0$ and $\langle \lambda_i^k, \Pi_{K_i}(G(x^k)) \rangle \to 0$ are equivalent to $\langle \lambda_i^k, \Pi_{K_i}(G(x^k)) \rangle \to 0$ and $\langle \lambda_i^k, \Pi_{K_i}(G(x^k)) \rangle \to 0$, for every $i \in \{1, \ldots, r\}$.

From $\langle \lambda_i^k, \Pi_{K_i}(G(x^k)) \rangle \to 0$, using (41) and the left-hand side of (43), we have
\[
0 \leq \mu_1(\lambda_i^k) \mu_2(\Pi_{K_i}(G_i(x^k))) + \mu_2(\lambda_i^k) \mu_1(\Pi_{K_i}(G_i(x^k))) \leq 2(\lambda_i^k, \Pi_{K_i}(G_i(x^k))).
\]
Then, from (41) and taking $k \to \infty$ in the above expression, we see that:
- $\mu_1(\lambda_i^k) \mu_2(\Pi_{K_i}(G_i(x^k))) = \mu_1(\lambda_i^k) \min \{\mu_2(G_i(x^k)), 0\} \to 0$;
- $\mu_2(\lambda_i^k) \mu_1(\Pi_{K_i}(G_i(x^k))) = \mu_2(\lambda_i^k) \min \{\mu_1(G_i(x^k)), 0\} \to 0$.

Furthermore, from above, we see that $\mu_2(\lambda_i^k) \min \{\mu_2(G_i(x^k)), 0\} \to 0$. Thus, $\langle \lambda_i^k, \Pi_{K_i}(G_i(x^k)) \rangle \to 0$ implies $\mu_\ell(\lambda_i^k) \min \{\mu_\ell(G_i(x^k)), 0\} \to 0$ for $\ell, s \in \{1, 2\}, s \neq \ell$ and $\forall i \in \{1, \ldots, r\}$. See Table 6.3

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 $i \in I_{\text{int}}, \lambda_i^k \to 0$, and for $i \in I_{\text{bd}+}, -\lambda_i^k \in \text{bd } K_i$ and either $\lambda_i^k \to 0$ or $\lambda_i^k / \|\lambda_i^k\| \to G_i(\bar{x}) / \|G_i(\bar{x})\|$. | AGP $\mu_\ell(\lambda_i^k) \max \{\mu_\ell(G_i(x^k)), 0\} \to 0$ for $\forall \ell \in \{1, 2\}, \forall i$. | CAKKT $\mu_\ell(\lambda_i^k) \max \{\mu_\ell(G_i(x^k)), 0\} \to 0$ and $\mu_\ell(\lambda_i^k) \min \{\mu_\ell(G_i(x^k)), 0\} \to 0$, $\forall \ell, s \in \{1, 2\} (\ell \neq s), \forall i$. |

Table 6.3: Sequential optimality conditions when specialized to (NSOCP). We use $\{\mu_\ell(v), \ell \in \{1, 2\}\}$ to denote spectral values of $v \in K_i$, see (42). For the definitions of $I_{\text{bd}+}$ and $I_{\text{int}}$, 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 \([59]\) holds at some feasible point \(x\) when there exists some \(d \in \mathbb{R}^n\) such that

\[
G(x) + DG(x)d \in \text{int } K.
\]  

(46)

It is widely known that Robinson’s CQ generalizes the Mangasarian-Fromovitz constraint qualification (MFCQ) from NLP (see [47]). 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 \(||\Lambda^k|| \to \infty\), but \(\bar{\Lambda}^k := \Lambda^k/||\Lambda^k|| \to \bar{\Lambda} \in \mathbb{K}^n\). Then from \([19]\) we get \(\nabla f(x^k)/||\Lambda^k|| + DG(x^k)^* \bar{\Lambda}^k \to 0\) and, consequently, \(DG(x^k)^* \bar{\Lambda} = 0\). Moreover,

\[
0 = \langle \Lambda^k, G(x^k) + y^k \rangle = \langle \bar{\Lambda}^k, G(x^k) + y^k \rangle \to \langle \bar{\Lambda}, G(\bar{x}) \rangle = 0
\]

and, by Robinson’s CQ, there exists some \(d \in \mathbb{R}^n\) such that \(G(\bar{x}) + DG(\bar{x})d \in \text{int } K\), then

\[
0 > \langle \bar{\Lambda}, G(\bar{x}) + DG(\bar{x})d \rangle = \langle \bar{\Lambda}, G(\bar{x}) \rangle + \langle \bar{\Lambda}, DG(\bar{x})d \rangle = \langle DG(\bar{x})^* \bar{\Lambda}, d \rangle = 0,
\]

where in the last equality we used the adjoint operator definition. This contradiction implies \(\{\Lambda^k\}_{k \in \mathbb{N}}\) is bounded. Hence, without loss of generality we can assume it converges to \(\bar{\Lambda} \in \mathbb{K}^n\). Trivially, \(\nabla f(\bar{x}) + DG(\bar{x})^* \bar{\Lambda} = 0\) and

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

Also, since \(\mathcal{K}\) is closed and \(\mathcal{K} \supseteq G(x^k) + y^k \to G(\bar{x})\) we have \(G(\bar{x}) \in \mathcal{K}\). Thus \((\bar{x}, \bar{\Lambda})\) is a KKT pair of **(NCP)**. \(\blacksquare\)

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 [10], for each sequential optimality condition (OC) it is possible to define conditions, so-called strict constraint qualifications (SCQ), such that “OC + SCQ ⇒ KKT” and, among them, characterize the weakest one.

We use the same nomenclature style of [10]. 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. \(\mathcal{K}_A(x, r) := \{DG(x)^* \Lambda : |y| \leq r, \ \Lambda \in \mathcal{K}^n, G(x) + y \in \mathcal{K} \cap \{\Lambda^k\}\};\)
2. \(\mathcal{K}_T(x, r) := \{DG(x)^* \Lambda : |\langle \Lambda, G(x) \rangle| \leq r, \ \Lambda \in \mathcal{K}^n\};\)
3. \(\mathcal{K}_{AGP}(x, r) := \{DG(x)^* \Lambda : |\langle \Lambda, \Pi_{\mathcal{K}}(G(x)) \rangle| \leq r, \ \Lambda \in \mathcal{K}^n\};\)
4. \(\mathcal{K}_C(x, r) := \{DG(x)^* \Lambda : \max(|\langle \Lambda, \Pi_{\mathcal{K}}(G(x)) \rangle|, |\langle \Lambda, \Pi_{\mathcal{K}^c}(G(x)) \rangle|) \leq r, \ \Lambda \in \mathcal{K}^n\}.\)

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

**Remark 3.** For a feasible point \(\bar{x}\) of **(NCP)**, we see that, at \((\bar{x}, 0)\), all the sets from Definition 7.1 coincide with \(\{DG(x)^* \Lambda : \langle \Lambda, G(x) \rangle = 0, \ \Lambda \in \mathcal{K}^n\}\). Thus, given an objective function \(f\) for **(NCP)**, the KKT conditions hold at \(\bar{x}\) if, and only if, \(-\nabla f(\bar{x}) \in \mathcal{K}_A(\bar{x}, 0)\). 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 the 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 [9, 10]. 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 $\{\Lambda^k\}_{k \in \mathbb{N}} \subset \mathcal{K}^0$ such that

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

Set $r_k := \max\{|\langle \Lambda^k, G(x^k) \rangle|, |\langle \Lambda^k, G(x^k) \rangle|\}, \forall k \in \mathbb{N}$. Thus, $-\nabla f(x^k) + w^k \in \mathcal{N}_C(x^k, r_k)$. Taking limits in the last expression, we get $-\nabla f(\bar{x}) \in \mathcal{N}_C(\bar{x}, 0)$, and hence the KKT conditions hold at $\bar{x}$.

Now, suppose that 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 $\omega \in \limsup_{(x^k, r_k) \rightarrow (\bar{x}, 0)} \mathcal{N}_C(x^k, r_k)$. Thus, there are sequences $\{x^k\}_{k \in \mathbb{N}} \rightarrow \bar{x}$ and $\{r_k\}_{k \in \mathbb{N}} \rightarrow 0$ with $w^k \in \mathcal{N}_C(x^k, r_k)$ and $w^k \rightarrow w$. Now, define $f(x) = -\langle \omega, x \rangle$. It is not difficult to see, that $\bar{x}$ is a CAKKT point for $f$ having $\{x^k\}_{k \in \mathbb{N}}$ as a associate CAKKT sequence. Then, by assumption, the KKT conditions hold, which is equivalent to $-\nabla f(\bar{x}) = w \in \mathcal{N}_C(\bar{x}, 0)$, and hence CAKKT-regularity hold at $\bar{x}$.

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, $T_{\Omega}(\bar{x}) = L_{\Omega}(\bar{x})$, 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 $\Omega$ at $\bar{x}$ and

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

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

The only affirmation that requires a proof is “CAKKT-regularity implies Abadie’s CQ”. But first, recall the regular normal cone to $\Omega$ at $z \in \Omega$, which is defined as

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

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

Lemma 7.3. We always have that $N_{\Omega}(\bar{x}) \subset \limsup_{(x, r) \rightarrow (\bar{x}, 0)} \mathcal{N}_C(x, r)$.

Proof. Analogous to the proof of [10, Lemma 4.3].

Finally, the result:

Theorem 7.4. 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 $\mathcal{N}_C(x, r)$ and Lemma 7.3 we get that $N_{\Omega}(\bar{x}) \subset \mathcal{N}_C(\bar{x}, 0)$. Now, we observe that $\mathcal{N}_C(\bar{x}, 0) = \{DG(\bar{x})^* \Lambda : \langle \Lambda, G(\bar{x}) \rangle = 0, \Lambda \in \mathcal{K}^0\}$ coincides with the polar of $L_{\Omega}(\bar{x})$. Indeed, by [63, Theorem 2.28], $L_{\Omega}(\bar{x})^o = \{DG(\bar{x})^* \Lambda : \Lambda \in T_K(G(\bar{x}))^o\}$, and for every closed convex cone $\mathcal{K}$, we have that $\Lambda \in T_K(G(\bar{x}))^o$ if, and only if, $\Lambda \in \mathcal{K}^0$ and $\langle \Lambda, G(\bar{x}) \rangle = 0$. Thus, $N_{\Omega}(\bar{x}) \subset \mathcal{N}_C(\bar{x}, 0)$ is just the inclusion $N_{\Omega}(\bar{x}) \subset L_{\Omega}(\bar{x})^o$.

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 will be sufficient to prove the inclusion $L_{\Omega}(\bar{x}) \subset T_{\Omega}(\bar{x})$. Now, from the inclusion $N_{\Omega}(\bar{x}) \subset L_{\Omega}(\bar{x})^o$, we get that $L_{\Omega}(\bar{x}) \subset (L_{\Omega}(\bar{x})^o)^0 \subset N_{\Omega}(\bar{x})^o \subset T_{\Omega}(\bar{x})$, where the first inclusion follows from polarity theorem, since $L_{\Omega}(\bar{x})$ is a closed convex cone due to the fact that $K$ is also a closed convex cone, and the last inclusion comes from [62, Theorems 6.28(b) and 6.26].

For summarizing our results, we illustrate the position of the new SCQs between Robinson’s CQ and Abadie’s CQ in the following diagram:
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 \(\text{NCP}\), 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 [4, 20, 34, 35] for NLP and we intend to extend them to more general contexts as well. Also, interior-point methods were proven to satisfy sequential optimality conditions of first- and second-order in [34], but this has not yet been addressed in NCP.

References


problems. To appear in Computational Optimization and Applications, 2020. DOI: 10.007/s10589-020-
00180-4.

generalized nash equilibrium problems and their practical implications. SIAM Journal of Optimization, 29(1):31–54,
2019.

27. L. Chen and D. Goldfarb. Interior-point 2-penalty methods for nonlinear programming with strong global

28. R. Chill and S. Mildner. The Kurdyka-Lojasiewicz-Simon inequality and stabilisation in nonsmooth infinite-


33. G. Giorgi, B. Jiménez, and V. Novo. Approximate Karush-Kuhn-Tucker condition in multiobjective optimi-

34. G. Haeser. A second-order optimality condition with first- and second-order complementarity associated

35. G. Haeser and A. Ramos. New constraint qualifications with second-order properties in nonlinear optimi-


40. C. Kanzow, D. Steck, and D. Wachsmuth. An augmented Lagrangian method for optimization problems


42. N. Krislock and H. Wolkowicz. Euclidean distance matrices and applications. International Series in

43. K. Kurdyka. On gradients of functions definable in o-minimal structures. Annales de l’institut Fourier

44. G. Li and T. K. Pong. Calculus of the exponent of Kurdyka-Lojasiewicz inequality and its appplications to
linear convergence of first-order methods. Foundations of Computational Mathematics, 18(5):1199–1232,
2018.


46. S. Lojasiewicz. Sur les trajectoires du gradient d’une fonction analytique. in Seminari di Geometria 1982-