Sequential Constant Rank Constraint Qualifications for Nonlinear Semidefinite Programming with Algorithmic Applications

Roberto Andreani[1] · Gabriel Haeser[2] · Leonardo M. Mito[2] · Héctor Ramírez[3]

Abstract

We present new constraint qualification conditions for nonlinear semidefinite programming that extend some of the constant rank-type conditions from nonlinear programming. As an application of these conditions, we provide a unified global convergence proof of a class of algorithms to stationary points without assuming neither uniqueness of the Lagrange multiplier nor boundedness of the Lagrange multipliers set. This class of algorithms includes, for instance, general forms of augmented Lagrangian, sequential quadratic programming, and interior point methods. In particular, we do not assume boundedness of the dual sequence generated by the algorithm. The weaker sequential condition we present is shown to be strictly weaker than Robinson's condition while still implying metric subregularity.

Roberto Andreani, Gabriel Haeser, Leonardo M. Mito and Héctor Ramírez contributed equally to this work.

✉ Gabriel Haeser
ghaeser@ime.usp.br

Roberto Andreani
andreani@unicamp.br

Leonardo M. Mito
leokoto@ime.usp.br

Héctor Ramírez
hramirez@dim.uchile.cl

[1] Department of Applied Mathematics, State University of Campinas, Campinas, Brazil

[2] Department of Applied Mathematics, University of São Paulo, São Paulo, Brazil

[3] Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, Universidad de Chile, Santiago, Chile

1 Introduction

Constraint qualification (CQ) conditions play a crucial role in optimization. They permit to establish first- and second-order necessary optimality conditions for local minima and support the convergence theory of many practical algorithms (see, for instance, a unified convergence analysis for a whole class of algorithms by Andreani et al. [9, Thm. 6]). Some of the well-known CQs in nonlinear programming (NLP) are the *constant-rank constraint qualification* (CRCQ), introduced by Janin [22], and the *constant positive linear dependence* (CPLD) condition. The latter was first conceptualized by Qi and Wei [27], and then proved to be a constraint qualification by Andreani et al. [12]. Moreover, it has been a source of inspiration for other authors to define even weaker constraint qualifications for NLP, such as the *constant rank of the subspace component* (CRSC) [10], and the relaxed versions of CRCQ [24] and CPLD [9]. Our interest in constant rank-type conditions is motivated, mainly, by their applications towards obtaining global convergence results of iterative algorithms to stationary points without relying on boundedness or uniqueness of Lagrange multipliers. However, several other applications that we do not pursue in this paper may be expected to be extended to the conic context, such as the computation of the derivative of the value function [22, 25] and the validity of strong second-order necessary optimality conditions that do not rely on the whole set of Lagrange multipliers [1]. Besides, their ability of dealing with redundant constraints, up to some extent, gives modellers some degree of freedom without losing regularity or convergence guarantees on algorithms. For instance, the standard NLP trick of replacing one nondegenerate equality constraint by two inequalities of opposite sign does not violate CRCQ, while violating the standard *Mangasarian-Fromovitz CQ* (MFCQ).

Constant-rank type CQs have been proposed in conic programming only very recently. The first extension of CRCQ to *nonlinear second-order cone programming* (NSOCP) appeared in [33], but it was shown to be incorrect in [2]. A second proposal [8], which encompasses also *nonlinear semidefinite programming* (NSDP) problems, consists of transforming some of the conic constraints into NLP constraints via a reduction function, whenever it was possible, and then demanding constant linear dependence of the reduced constraints, locally. This was considered by the authors a *naive* extension, since it basically avoids the main difficulties that are expected from a conic framework. What both these works have in common is that they somehow neglected the conic structure of the problem.

In a recent article [5], we introduced weak notions of regularity for *nonlinear semidefinite programming* (NSDP) that were defined in terms of the eigenvectors of the constraints – therein called *weak-nondegeneracy* and *weak-Robinson's CQ*. These conditions take into consideration only the diagonal entries of some particular transformation of the matrix constraint. Noteworthy, weak-nondegeneracy happens to be equivalent to the *linear independence CQ* (LICQ) when an NLP constraint is modeled as a structurally diagonal matrix constraint, unlike the standard *nondegeneracy* condition [31], which in turn is considered the usual extension of LICQ to NSDP. Moreover, the proof technique we employed in [5] induces a direct application in the convergence theory of an external penalty method. In this paper, we use these conditions to derive our extension proposals for CRCQ and CPLD to NSDP. These CQs are called, in this paper, as weak-CRCQ and weak-CPLD, respectively.

However, to provide support for algorithms other than the external penalty method, we present stronger variants of these conditions, called sequential-CRCQ and sequential-CPLD (abbreviated as seq-CRCQ and seq-CPLD, respectively), by incorporating perturbations in their definitions. This makes them robust and easily connectible with algorithms that

keep track of approximate Lagrange multipliers, but also more exigent. Nevertheless, seq-CRCQ is still strictly weaker than nondegeneracy and independent of Robinson's CQ, while seq-CPLD is strictly weaker than Robinson's CQ. On the other hand, weak-CRCQ is strictly weaker than seq-CRCQ, while weak-CPLD is strictly weaker than weak-CRCQ and seq-CPLD. Moreover, we show that seq-CPLD implies the *metric subregularity CQ*. The global convergence results obtained under seq-CPLD, strictly weaker than Robinson's CQ, are the best results known by the authors for the augmented Lagrangian, sequential quadratic programming, and interior point methods discussed in this paper. In particular, we do not assume boundedness of the dual sequence, which is a somewhat stringent but usually employed assumption on the behavior of these algorithms.

The content of this paper is organized as follows: Section 2 introduces notation and some well-known theorems and definitions that will be useful in the sequel. Our main results for NSDP are presented in Sections 3 and 4. Indeed, Section 3 is devoted to the study of weak-CRCQ and weak-CPLD and their properties, which in turn need to invoke weak-nondegeneracy and weak-Robinson's CQ as a motivation. Section 4 studies seq-CRCQ and seq-CPLD – the main CQs of this paper – and some algorithms supported by them. In Section 5, we discuss the relationship between seq-CPLD and the metric subregularity CQ. Lastly, some final remarks are given in Section 6. To abbreviate the paper, we will focus on weak- and seq-CPLD since the extensions of CRCQ follow analogously.

2 A Nonlinear Semidefinite Programming Review

In this section, $\mathbb{S}^m$ denotes the linear space of all $m \times m$ real symmetric matrices equipped with the inner product defined as $\langle M, N \rangle \doteq \text{trace}(MN) = \sum_{i,j=1}^{m} M_{ij} N_{ij}$ for all $M, N \in \mathbb{S}^m$, and $\mathbb{S}_+^m$ is the cone of all positive semidefinite matrices in $\mathbb{S}^m$. Additionally, for every $M \in \mathbb{S}^m$ and every $\tau > 0$, we denote by $B(M, \tau) \doteq \{Z \in \mathbb{S}^m : \|M - Z\| < \tau\}$ the open ball centered at M with radius τ with respect to the Frobenius norm $\|M\| \doteq \sqrt{\langle M, M \rangle}$, and its closure will be denoted by $\text{cl}(B(M, \tau))$.

We consider the NSDP problem in standard (dual) form:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & G(x) \succeq 0, \end{aligned} \qquad \text{(NSDP)}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{S}^m$ are continuously differentiable functions, and $\succeq$ is the partial order induced by $\mathbb{S}_+^m$; that is, $M \succeq N$ if, and only if, $M - N \in \mathbb{S}_+^m$.

Equality constraints are omitted in (NSDP) for simplicity of notation, but our definitions and results are flexible regarding inclusion of such constraints, which should be done in the same way as in [8]. Moreover, throughout the whole paper, we will denote the feasible set of (NSDP) by $\mathcal{F}$.

Let us recall that the orthogonal projection of an element $M \in \mathbb{S}^m$ onto $\mathbb{S}_+^m$, which is defined as

$$\Pi_{\mathbb{S}_+^m}(M) \doteq \underset{N \in \mathbb{S}_+^m}{\text{argmin}} \|M - N\|,$$

is a Lipschitz continuous function of M with modulus 1. Furthermore, since $\mathbb{S}_+^m$ is self-dual, every $M \in \mathbb{S}^m$ has a *Moreau decomposition* [26, Prop. 1] in the form

$$M = \Pi_{\mathbb{S}_+^m}(M) - \Pi_{\mathbb{S}_+^m}(-M)$$

with $\langle \Pi_{\mathbb{S}_+^m}(M), \Pi_{\mathbb{S}_+^m}(-M) \rangle = 0$, and a *spectral decomposition* in the form

$$M = \lambda_1(M)u_1(M)u_1(M)^\top + \ldots + \lambda_m(M)u_m(M)u_m(M)^\top, \tag{1}$$

where $u_1(M), \ldots, u_m(M) \in \mathbb{R}^m$ are arbitrarily chosen orthonormal eigenvectors associated with the eigenvalues $\lambda_1(M), \ldots, \lambda_m(M)$, respectively. In turn, these eigenvalues are assumed to be arranged in non-increasing order. Equivalently, we can write (1) as $M = U\mathcal{D}U^\top$, where U is an orthogonal matrix whose i-th column is $u_i(M)$, and $\mathcal{D} \doteq \mathrm{Diag}(\lambda_1(M), \ldots, \lambda_m(M))$ is a matrix whose diagonal entries are $\lambda_1(M), \ldots, \lambda_m(M)$ and the remaining entries are zero.

A convenient property of the orthogonal projection onto $\mathbb{S}_+^m$ is that, for every $M \in \mathbb{S}^m$, we have

$$\Pi_{\mathbb{S}_+^m}(M) = [\lambda_1(M)]_+ u_1(M)u_1(M)^\top + \ldots + [\lambda_m(M)]_+ u_m(M)u_m(M)^\top,$$

where $[\,\cdot\,]_+ \doteq \max\{\,\cdot\,, 0\}$.

Given a sequence of sets $\{S^k\}_{k \in \mathbb{N}}$, recall its *outer limit* (or *upper limit*) in the sense of Painlevé-Kuratowski (cf. [29, Def. 4.1] or [15, Def. 2.52]), defined as

$$\mathrm{Lim\,sup}_{k \in \mathbb{N}} S^k \doteq \left\{ y \colon \exists I \subseteq_\infty \mathbb{N}, \ \exists \left\{ y^k \right\}_{k \in I} \to y, \ \forall k \in I, \ y^k \in S^k \right\},$$

which is the collection of all cluster points of sequences $\{y^k\}_{k \in \mathbb{N}}$ such that $y^k \in S^k$ for every $k \in \mathbb{N}$. The notation $I \subseteq_\infty \mathbb{N}$ means that I is an infinite subset of the set of natural numbers $\mathbb{N}$.

We denote the *Jacobian* of G at a given point $x \in \mathbb{R}^n$ by $DG(x)$, and the *adjoint* operator of $DG(x)$ will be denoted by $DG(x)^*$. Moreover, the i-th partial derivative of G at x will be denoted by $D_{x_i}G(x)$, and the *gradient* of f at x will be written as $\nabla f(x)$, for every $x \in \mathbb{R}^n$.

2.1 Classical Optimality Conditions and Constraint Qualifications

As usual in continuous optimization, we drive our attention towards local solutions of (NSDP) that satisfy the so-called *Karush-Kuhn-Tucker* (KKT) conditions, defined as follows:

Definition 1 We say that the *Karush-Kuhn-Tucker* conditions hold at $\overline{x} \in \mathcal{F}$ when there exists some $\overline{Y} \succeq 0$ such that

$$\nabla_x L(\overline{x}, \overline{Y}) = 0 \quad \text{and} \quad \langle G(\overline{x}), \overline{Y} \rangle = 0,$$

where $L(x, Y) \doteq f(x) - \langle G(x), Y \rangle$ is the Lagrangian function of (NSDP). The matrix $\overline{Y}$ is called a *Lagrange multiplier* associated with $\overline{x}$, and the set of all Lagrange multipliers associated with $\overline{x}$ will be denoted by $\Lambda(\overline{x})$.

Of course, not every local minimizer satisfies the KKT conditions in the absence of a CQ. In order to recall some classical CQs, it is necessary to use the *(Bouligand) tangent cone* to $\mathbb{S}_+^m$ at a point $M \succeq 0$. This object can be characterized in terms of any matrix $E \in \mathbb{R}^{m \times (m-r)}$, whose columns form an orthonormal basis of $\mathrm{Ker} M$, as follows (e.g., [15, Ex. 2.65]):

$$T_{\mathbb{S}_+^m}(M) = \left\{ N \in \mathbb{S}^m \colon E^\top N E \succeq 0 \right\}, \tag{2}$$

where r denotes the rank of M. So, its *lineality space*, defined as the largest linear space contained in $T_{\mathbb{S}^m_+}(M)$, is computed as follows:

$$\text{lin}\left(T_{\mathbb{S}^m_+}(M)\right) = \left\{N \in \mathbb{S}^m : E^\top N E = 0\right\}. \tag{3}$$

The latter is a direct consequence of the identity $\text{lin}(C) = C \cap (-C)$, satisfied for any closed convex cone C.

One of the most recognized constraint qualifications in NSDP is the *nondegeneracy* (or *transversality*) condition introduced by Shapiro and Fan [31], which can be characterized [15, Eq. 4.172] at a point $\overline{x} \in \mathcal{F}$ when the following relation is satisfied:

$$\text{Im } DG(\overline{x}) + \text{lin}\left(T_{\mathbb{S}^m_+}(G(\overline{x}))\right) = \mathbb{S}^m.$$

If $\overline{x}$ is a local solution of (NSDP) that satisfies nondegeneracy, then $\Lambda(\overline{x})$ is a singleton, but the converse is not necessarily true unless $\text{rank}(G(\overline{x})) + \text{rank}(\overline{Y}) = m$ holds for some $\overline{Y} \in \Lambda(\overline{x})$ [15, Prop. 4.75]. This last condition is known as *strict complementarity* in this NSDP framework. By (2) it is possible to characterize nondegeneracy at $\overline{x}$ by means of any given matrix $\overline{E}$ with orthonormal columns that span $\text{Ker}G(\overline{x})$. Indeed, following [15, Sec. 4.6.1], nondegeneracy holds at $\overline{x}$ if, and only if, either $\text{Ker}G(\overline{x}) = \{0\}$ or the linear mapping $\psi_{\overline{x}} \colon \mathbb{R}^n \to \mathbb{S}^{m-r}$ given by

$$\psi_{\overline{x}}(\,\cdot\,) \doteq \overline{E}^\top DG(\overline{x})[\,\cdot\,]\overline{E} \tag{4}$$

is surjective, which is in turn equivalent to saying that the vectors

$$v_{ij}\left(\overline{x}, \overline{E}\right) \doteq \left[\overline{e}_i^\top D_{x_1} G(\overline{x})\overline{e}_j, \ldots, \overline{e}_i^\top D_{x_n} G(\overline{x})\overline{e}_j\right]^\top, \quad 1 \leq i \leq j \leq m - r, \tag{5}$$

are linearly independent [30, Prop. 6], where $\overline{e}_i$ denotes the i-th column of $\overline{E}$ and r is the rank of $G(\overline{x})$.

Another widespread constraint qualification is Robinson's CQ [28], which can be characterized at $\overline{x} \in \mathcal{F}$ by the existence of some $d \in \mathbb{R}^n$ such that

$$G(\overline{x}) + DG(\overline{x})[d] \in \text{int } \mathbb{S}^m_+ \tag{6}$$

where $\text{int } \mathbb{S}^m_+$ stands for the topological interior of $\mathbb{S}^m_+$. It is known (e.g., [15, Props. 3.9 and 3.17]) that when $\overline{x}$ is a local solution of (NSDP), then $\Lambda(\overline{x})$ is nonempty and compact if, and only if, Robinson's CQ holds at $\overline{x}$.

Given the properties and characterizations recalled above, the nondegeneracy condition is typically considered the natural extension of LICQ from NLP to NSDP, while Robinson's CQ is considered the extension of MFCQ.

2.2 A Sequential Optimality Condition Connected to the External Penalty Method

If we do not assume any CQ, every local minimizer of (NSDP) can still be proved to satisfy at least a *sequential* type of optimality condition that is deeply connected to the classical external penalty method. Namely:

Theorem 1 *Let $\overline{x}$ be a local minimizer of* (NSDP)*, and let $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$. Then, there exists some $\left\{x^k\right\}_{k\in\mathbb{N}} \to \overline{x}$, such that for each $k \in \mathbb{N}$, x^k is a local minimizer of the regularized penalized function*

$$F(x) \doteq f(x) + \frac{1}{2}\|x - \overline{x}\|_2^2 + \frac{\rho_k}{2}\|\Pi_{\mathbb{S}^m_+}(-G(x))\|^2.$$

Proof See [11, Thm. 2]. For a more general proof, see the first part of the proof of [4, Thm. 2]. □

Note that Theorem 1 provides a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that each x^k satisfies, with an error $\varepsilon^k \to 0^+$, the first-order optimality condition of the unconstrained minimization problem

$$\underset{x\in\mathbb{R}^n}{\text{Minimize}}\ f(x) + \frac{\rho_k}{2}\|\Pi_{\mathbb{S}_+^m}(-G(x))\|^2,$$

so $\{x^k\}_{k\in\mathbb{N}}$ characterizes an output sequence of an *external penalty method*. Moreover, the sequence $\{Y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}_+^m$, where

$$Y^k \doteq \rho_k \Pi_{\mathbb{S}_+^m}\left(-G\left(x^k\right)\right)$$

for every $k \in \mathbb{N}$, consists of *approximate Lagrange multipliers* for $\overline{x}$, in the sense that $\nabla_x L\left(x^k, Y^k\right) \to 0$ and complementarity and feasibility are approximately fulfilled, in view of Moreau's decomposition – indeed, note that $\langle G\left(x^k\right)+\Delta^k, Y^k\rangle = 0$ and $G\left(x^k\right)+\Delta^k \succeq 0$, with $\Delta^k = \Pi_{\mathbb{S}_+^m}\left(-G\left(x^k\right)\right) \to 0$, for every $k \in \mathbb{N}$.

These sequences will suffice to obtain the results of the first part of this paper (Section 3), but in order to extend their scope to a larger class of iterative algorithms, in Section 4, we will need a more general sequential optimality condition, which will be presented later on.

2.3 Reviewing Constant Rank-Type Constraint Qualifications for NLP

This section is meant to be a brief review of the main results regarding the classical nonlinear programming problem:

$$\begin{aligned}\underset{x\in\mathbb{R}^n}{\text{Minimize}}\quad & f(x),\\ \text{subject to}\quad & g_1(x) \geq 0, \ldots, g_m(x) \geq 0,\end{aligned} \tag{NLP}$$

where $f, g_1, \ldots, g_m\colon \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable functions.

As far as we know, the first constant rank-type constraint qualification was introduced by Janin [22], to obtain directional derivatives for the optimal value function of a perturbed NLP problem. Janin's condition is defined as follows:

Definition 2 Let $\overline{x} \in \mathcal{F}$. The *constant rank constraint qualification* for (NLP) (CRCQ) holds at $\overline{x}$ if there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every subset $J \subseteq \{i \in \{1, \ldots, m\}\colon g_i(\overline{x}) = 0\}$, the rank of the family $\{\nabla g_i(x)\}_{i\in J}$ remains constant for all $x \in \mathcal{V}$.

As noticed by Qi and Wei [27] it is possible to rephrase Definition 2 in terms of the "constant linear dependence" of $\{\nabla g_i(x)\}_{i\in J}$ for every J. That is, CRCQ holds at $\overline{x}$ if, and only if, there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every $J \subseteq \{i \in \{1, \ldots, m\}\colon g_i(\overline{x}) = 0\}$, if $\{\nabla g_i(\overline{x})\}_{i\in J}$ is linearly dependent, then $\{\nabla g_i(x)\}_{i\in J}$ remains linearly dependent for every $x \in \mathcal{V}$. Based on this characterization, Qi and Wei proposed a relaxed version of CRCQ, called *constant positive linear dependence* (CPLD) condition, to obtain improved global convergence results for a sequential quadratic programming method, but this was only proven to be a constraint qualification a few years later, in [12]. To properly present CPLD, recall that a family of vectors $\{z_i\}_{i\in J}$ of $\mathbb{R}^n$ is said to be *positively linearly independent* when

$$\sum_{i\in J} z_i\alpha_i = 0, \ \ \alpha_i \geq 0, \ \forall i \in J \quad \Rightarrow \quad \alpha_i = 0, \ \forall i \in J.$$

Next, we recall the CPLD constraint qualification:

Definition 3 Let $\overline{x} \in \mathcal{F}$. The *constant positive linear dependence* condition for (NLP) (CPLD) holds at $\overline{x}$ if there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every $J \subseteq \{i \in \{1, \ldots, m\}: g_i(\overline{x}) = 0\}$, if the family $\{\nabla g_i(\overline{x})\}_{i \in J}$ is positively linearly dependent, then $\{\nabla g_i(x)\}_{i \in J}$ remains linearly dependent for all $x \in \mathcal{V}$.

Clearly, CPLD is implied by CRCQ, which is in turn implied by LICQ and is independent of MFCQ. Moreover, CPLD is implied by MFCQ, and all those implications are strict [12, 22]. To prove the main results of this paper (Theorems 2 and 5), we shall take inspiration in [9], where the authors employ Theorem 1 together with the well-known *Carathéodory's Lemma*:

Lemma 1 (Exercise B.1.7 of [13]) *Let $z_1, \ldots, z_p \in \mathbb{R}^n$, and let $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$ be arbitrary. Then, there exist some $J \subseteq \{1, \ldots, p\}$ and some scalars $\tilde{\alpha}_i$ with $i \in J$, such that $\{z_i\}_{i \in J}$ is linearly independent,*

$$\sum_{i=1}^{p} \alpha_i z_i = \sum_{i \in J} \tilde{\alpha}_i z_i,$$

and $\alpha_i \tilde{\alpha}_i > 0$, for all $i \in J$.

If one considers equality constraints in (NSDP) separately, one should employ an adapted version of Carathéodory's Lemma that fixes a particular subset of vectors, which can be found in [9, Lem. 2]. In our current setting, Lemma 1 will suffice as is.

3 Weak Constant Rank Constraint Qualifications for NSDP

The main purpose of this section is to present an extension of the CPLD condition for NSDP, but to do so we first need to know how CRCQ looks like in NSDP, which is a challenging problem on its own. Based on the relation between LICQ and CRCQ in NLP, the most natural candidate for CRCQ in NSDP consists of demanding every subset of

$$\{v_{ij}(x, \overline{E}): 1 \leq i \leq j \leq m - r\}$$

to remain with constant rank (or constant linear dependent) in a neighborhood of $\overline{x}$. However, this candidate cannot be a CQ, as shown in the following counterexample, adapted from [2, Eq. 2]:

Example 1 Consider the problem to minimize $f(x) \doteq -x$ subject to

$$G(x) \doteq \begin{bmatrix} x & x + x^2 \\ x + x^2 & x \end{bmatrix} \succeq 0.$$

For this problem, $\overline{x} \doteq 0$ is the only feasible point and, therefore, the unique global minimizer of the problem. Since $G(\overline{x}) = 0$, the columns of the matrix $\overline{E} \doteq \mathbb{I}_2$ form an orthonormal basis of $\text{Ker} G(\overline{x})$ (the whole space $\mathbb{R}^2$). For this choice of $\overline{E}$, we have

$$v_{11}(x, \overline{E}) = v_{22}(x, \overline{E}) = 1 \quad \text{and} \quad v_{12}(x, \overline{E}) = 1 + 2x.$$

Since they are all bounded away from zero, the rank of every subset of $\{v_{ij}(x, \overline{E}): 1 \leq j \leq 2\}$ remains constant for every x around $\overline{x}$. However, Note that $\overline{x}$ does not satisfy the

KKT conditions because any $\overline{Y} \doteq \begin{bmatrix} \overline{Y}_{11} & \overline{Y}_{12} \\ \overline{Y}_{12} & \overline{Y}_{22} \end{bmatrix} \in \Lambda(\overline{x})$ would necessarily be a solution of the system

$$\begin{aligned}
&\overline{Y}_{11} \geq 0, \\
&\overline{Y}_{22} \geq 0, \\
&\overline{Y}_{11}\overline{Y}_{22} - \overline{Y}_{12}^2 \geq 0, \\
&\overline{Y}_{11} + 2\overline{Y}_{12} + \overline{Y}_{22} = -1,
\end{aligned}$$

which has no solution.

Besides, it is well known that even if G is affine, not all local minimizers of (NSDP) satisfy the KKT conditions, but in this case every subfamily of $\{v_{ij}(x, \overline{E}) \colon 1 \leq i \leq j \leq m - r\}$ remains with constant rank for every $x \in \mathbb{R}^n$.

What Example 1 tells us is that $\overline{E} = \mathbb{I}_2$ may be a bad choice of $\overline{E}$. In fact, let us choose a different $\overline{E}$, namely, denote the columns of $\overline{E}$ by $\overline{e}_1 \doteq [a, b]^\top$ and $\overline{e}_2 \doteq [c, d]^\top$, and take $a = -1/\sqrt{2}$ and $b = c = d = 1/\sqrt{2}$. This election of $\overline{E}$ happens to diagonalize $G(x)$ for every x, but it follows that

$$\begin{aligned}
v_{11}(x, \overline{E}) &= 1 + 2ab(1 + 2x) = -2x; \\
v_{22}(x, \overline{E}) &= 1 + 2cd(1 + 2x) = 2(1 + x); \\
v_{12}(x, \overline{E}) &= (ad + bc)(1 + 2x) = 0,
\end{aligned}$$

and the rank of $\{v_{11}(x, \overline{E})\}$ does not remain constant in a neighborhood of $\overline{x} = 0$.

In light of our previous work [5], the situation presented above is not surprising. Therein, we already noted that identifying the "good" matrices $\overline{E}$ allows us to obtain relaxed versions of nondegeneracy and Robinson's CQ for NSDP. This identification can also be used to extend constant-rank type conditions to NSDP and is the starting point for the results we will present.

For the sake of completeness, let us quickly summarize a discussion raised in [5] before presenting the results of this paper. Consider a feasible point $\overline{x} \in \mathcal{F}$ and denote by r the rank of $G(\overline{x})$. Observe that $\lambda_r(M) > \lambda_{r+1}(M)$ for every $M \in \mathbb{S}^m$ close enough to $G(\overline{x})$. Thus, when $r < m$, define the set

$$\mathcal{E}_r(M) \doteq \left\{ E \in \mathbb{R}^{m \times (m-r)} \colon \begin{array}{l} ME = E\mathrm{Diag}(\lambda_{r+1}(M), \dots, \lambda_m(M)) \\ E^\top E = \mathbb{I}_{m-r} \end{array} \right\}, \tag{7}$$

which consists of all matrices whose columns are orthonormal eigenvectors associated with the $m - r$ smallest eigenvalues of M, which is well defined whenever $\lambda_r(M) > \lambda_{r+1}(M)$. In (7), $\mathrm{Diag}(\lambda_{r+1}(M), \dots, \lambda_m(M))$ denotes the diagonal matrix whose diagonal entries are $\lambda_{r+1}(M), \dots, \lambda_m(M)$. By convention, $\mathcal{E}_r(M) \doteq \emptyset$ when $r = m$. By construction, $\mathcal{E}_r(M)$ is nonempty provided $r < m$ and M is close enough to $G(\overline{x})$. In particular, in this situation, $\mathcal{E}_r(G(\overline{x}))$ is the set of all matrices with orthonormal columns that span $\mathrm{Ker}G(\overline{x})$.

We showed, in [5, Prop. 3.2], that nondegeneracy can be equivalently stated as the linear independence of the smaller family, $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in \{1, \dots, m-r\}}$, as long as this holds for all $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ instead of a fixed one. Similarly, Robinson's CQ can be translated as the positive linear independence of the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in \{1, \dots, m-r\}}$ for every $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ [5, Prop. 5.1]. This characterization suggested a weak form of nondegeneracy (and Robinson's CQ) that takes into account only a particular subset of $\mathcal{E}_r(G(\overline{x}))$ instead of the whole set, which reads as follows:

Definition 4 (Def. 3.2 and Def. 5.1 of [5]) Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. We say that $\overline{x}$ satisfies:

- *Weak-nondegeneracy condition* for NSDP if either $r = m$ or, for each sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \operatorname{Lim\,sup}_{k \in \mathbb{N}} \mathcal{E}_r \left(G \left(x^k \right) \right)^1$ such that the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in \{1, \dots, m-r\}}$ is linearly independent;
- *Weak-Robinson's CQ condition* for NSDP if either $r = m$ or, for each sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \operatorname{Lim\,sup}_{k \in \mathbb{N}} \mathcal{E}_r \left(G \left(x^k \right) \right)$ such that the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in \{1, \dots, m-r\}}$ is positively linearly independent.

Note that, in general, $\operatorname{Lim\,sup}_{k \in \mathbb{N}} \mathcal{E}_r \left(G \left(x^k \right) \right) \subseteq \mathcal{E}_r \left(G(\overline{x}) \right)$, but the reverse inclusion is not always true, meaning $\mathcal{E}_r \left(G(x) \right)$ is not necessarily continuous at $\overline{x}$ as a set-valued mapping. It then follows that weak-nondegeneracy is indeed a strictly weaker CQ than nondegeneracy [5, Ex. 3.1]. Moreover, in contrast with nondegeneracy, weak-nondegeneracy happens to fully recover LICQ when $G(x)$ is a structurally diagonal matrix constraint in the form $G(x) \doteq \operatorname{Diag}(g_1(x), \dots, g_m(x))$ [5, Prop. 3.3]. Similarly, weak-Robinson's CQ is implied by Robinson's CQ and coincides with MFCQ when $G(x)$ is diagonal.

3.1 Weak-CRCQ and Weak-CPLD

A straightforward relaxation of weak-nondegeneracy and weak-Robinson's CQ, likewise NLP, leads to our first extension proposal of CRCQ and CPLD, respectively, to NSDP:

Definition 5 (weak-CRCQ and weak-CPLD) Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. We say that $\overline{x}$ satisfies the:

- *Weak constant rank constraint qualification* for NSDP (weak-CRCQ) if either $r = m$ or, for each sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \operatorname{Lim\,sup}_{k \in \mathbb{N}} \mathcal{E}_r \left(G \left(x^k \right) \right)$ such that, for every subset $J \subseteq \{1, \dots, m - r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\left\{ v_{ii} \left(x^k, E^k \right) \right\}_{i \in J}$ remains linearly dependent, for all $k \in I$ large enough.
- *Weak constant positive linear dependence constraint qualification* for NSDP (weak-CPLD) if either $r = m$ or, for each sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \operatorname{Lim\,sup}_{k \in \mathbb{N}} \mathcal{E}_r \left(G \left(x^k \right) \right)$ such that, for every $J \subseteq \{1, \dots, m - r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is positively linearly dependent, then the family $\left\{ v_{ii} \left(x^k, E^k \right) \right\}_{i \in J}$ remains linearly dependent, for all $k \in I$ large enough.

For both definitions, $I \subseteq_\infty \mathbb{N}$, and $\{E^k\}_{k \in I}$ is a sequence converging to $\overline{E}$ and such that $E^k \in \mathcal{E}_r \left(G \left(x^k \right) \right)$ for every $k \in I$, as required by the Painlevé-Kuratowski outer limit.

Observe that if weak-nondegeneracy holds at $\overline{x}$, then weak-CRCQ is vacuously true (because there is no linearly dependent family of vector $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ when $J \subseteq \{1, \dots, m - r\}$), and weak-CRCQ in turn implies weak-CPLD. Similarly, weak-Robinson's CQ implies weak-CPLD as well. However, Robinson's CQ and its weak variant are both independent of weak-CRCQ. In fact, the next example shows that weak-CRCQ is not implied by either (weak-)Robinson's CQ or weak-CPLD.

---

[1]Without further mention, we will endow $\mathbb{R}^{m \times (m-r)}$ with a norm, which can be the Frobenius norm or a column-wise maximum norm, for instance.

Example 2 Let us consider the constraint

$$G(x) \doteq \begin{bmatrix} 2x_1 + x_2^2 & -x_2^2 \\ -x_2^2 & 2x_1 + x_2^2 \end{bmatrix} \succeq 0$$

and note that, for every orthogonal matrix E in the form

$$E \doteq \begin{bmatrix} a & c \\ b & d \end{bmatrix},$$

we have

$$v_{11}(x, E) = \begin{bmatrix} 2 \\ 2(a-b)^2 x_2 \end{bmatrix} \quad \text{and} \quad v_{22}(x, E) = \begin{bmatrix} 2 \\ 2(c-d)^2 x_2 \end{bmatrix}.$$

Then, at $\overline{x} = 0$, we have $v_{11}(\overline{x}, \overline{E}) = v_{22}(\overline{x}, \overline{E}) = [2, 0]^\top$, so they are linearly dependent, but positively linearly independent for all $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$. However, choosing any sequence $\{x^k\}_{k\in\mathbb{N}} \to 0$ such that $x_2^k \neq 0$ for all k, it follows that the eigenvalues of $G(x_k)$:

$$\lambda_1\left(G\left(x^k\right)\right) = 2\left(x_1^k + \left(x_2^k\right)^2\right) \quad \text{and} \quad \lambda_2\left(G\left(x^k\right)\right) = 2x_1^k,$$

are simple, with associated orthonormal eigenvectors

$$u_1\left(G\left(x^k\right)\right) = \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^\top \quad \text{and} \quad u_2\left(G\left(x^k\right)\right) = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^\top,$$

respectively, for every $k \in \mathbb{N}$. Then, the only sequence $\{E^k\}_{k\in\mathbb{N}}$ such that $E^k \in \mathcal{E}_r\left(G\left(x^k\right)\right)$ for every k, up to sign, is given by $a = -1/\sqrt{2}$ and $b = c = d = 1/\sqrt{2}$. However, keep in mind that $v_{ii}(x, E)$, $i \in \{1, 2\}$, is invariant to the sign of the columns of E, so $v_{22}\left(x^k, E^k\right) = [2, 0]^\top$ and $v_{11}\left(x^k, E^k\right) = [2, 4x_2^k]^\top$ are linearly independent for all large k. Therefore, we conclude that (weak-)Robinson's CQ holds at $\overline{x}$, and consequently weak-CPLD also holds, but weak-CRCQ does not hold at $\overline{x}$.

Conversely, we show with another counterexample, that weak-CRCQ does not imply (weak-)Robinson's CQ, and neither does weak-CPLD.

Example 3 Let us consider the constraint

$$G(x) \doteq \begin{bmatrix} x & x^2 \\ x^2 & -x \end{bmatrix} \succeq 0$$

and the point $\overline{x} = 0$. Take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that $x^k \neq 0$ for every k, and consider two subsequences of it, indexed by I_+ and I_-, such that $x^k > 0$ for every $k \in I_+$, and $x^k < 0$ for every $k \in I_-$. Then, for every $k \in I_+$, we have that:

$$\lambda_1\left(G\left(x^k\right)\right) = x^k\sqrt{\left(x^k\right)^2 + 1} \quad \text{and} \quad \lambda_2\left(G\left(x^k\right)\right) = -x^k\sqrt{\left(x^k\right)^2 + 1},$$

are simple, with associated orthonormal eigenvectors uniquely determined (up to sign) by

$$u_1\left(G\left(x^k\right)\right) = \frac{1}{\eta_1^k}\left[\frac{1 + \sqrt{\left(x^k\right)^2 + 1}}{x^k}, 1\right]^\top, \quad u_2\left(G\left(x^k\right)\right) = \frac{1}{\eta_2^k}\left[\frac{1 - \sqrt{\left(x^k\right)^2 + 1}}{x^k}, 1\right]^\top,$$

where

$$
\eta_1^k \doteq \sqrt{\left(\frac{1 + \sqrt{\left(x^k\right)^2 + 1}}{x^k}\right)^2 + 1} \quad \text{and} \quad \eta_2^k \doteq \sqrt{\left(\frac{1 - \sqrt{\left(x^k\right)^2 + 1}}{x^k}\right)^2 + 1}.
$$

Moreover, one can verify that whenever I_+ is an infinite set,

$$
\lim_{k \in I_+} u_1\left(G\left(x^k\right)\right) = [1, 0]^\top \quad \text{and} \quad \lim_{k \in I_+} u_2\left(G\left(x^k\right)\right) = [0, 1]^\top.
$$

Then, we have that for all $\overline{E} \in \operatorname{Lim\,sup}_{k \in I_+} \mathcal{E}_r\left(G\left(x^k\right)\right)$, the vectors

$$
v_{11}(\overline{x}, \overline{E}) = 1 \quad \text{and} \quad v_{22}(\overline{x}, \overline{E}) = -1
$$

are positively linearly dependent. In addition, since $\eta_1^k \to \infty$ and $\eta_2^k \to 1$, the vectors

$$
v_{11}\left(x^k, E^k\right) = \frac{\left(\eta_1^k\right)^2 + 4\sqrt{\left(x^k\right)^2 + 1} + 2}{\left(\eta_1^k\right)^2}
$$

and

$$
v_{22}\left(x^k, E^k\right) = \frac{\left(\eta_2^k\right)^2 - 4\sqrt{\left(x^k\right)^2 + 1} + 2}{\left(\eta_2^k\right)^2}
$$

are nonzero and have opposite signs; and thus, remain positively linearly dependent, for all large $k \in I_+$.

For the indices $k \in I_-$ the order of $\lambda_1\left(G\left(x^k\right)\right)$ and $\lambda_2\left(G\left(x^k\right)\right)$ is swapped, together with their respective eigenvectors, and we have $\lim_{k \in I_-} u_1\left(G\left(x^k\right)\right) = [0, 1]^\top$ and $\lim_{k \in I_-} u_2\left(G\left(x^k\right)\right) = [-1, 0]^\top$. Hence, for all $\overline{E} \in \operatorname{Lim\,sup}_{k \in I_-} \mathcal{E}_r\left(G\left(x^k\right)\right)$, the vectors

$$
v_{11}(\overline{x}, \overline{E}) = -1 \quad \text{and} \quad v_{22}(\overline{x}, \overline{E}) = 1
$$

are also positively linearly dependent. The order of $v_{11}\left(x^k, E^k\right)$ and $v_{22}\left(x^k, E^k\right)$ is also swapped, so they remain positively linearly dependent for all large $k \in I_-$.

By the above reasoning, observe that any sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$, such that $x^k \neq 0$ for every $k \in \mathbb{N}$, shows that (weak-)Robinson's CQ fails at $\overline{x}$. Moreover, if $x^k = 0$ for infinitely many indices, we may simply take $E^k = \overline{E} = \mathbb{I}_2$ for every k, and then $v_{11}\left(x^k, E^k\right) = v_{11}(\overline{x}, \overline{E}) = 1$ and $v_{22}\left(x^k, E^k\right) = v_{22}(\overline{x}, \overline{E}) = -1$ are positively linearly dependent for every $k \in \mathbb{N}$. This completes checking that weak-CPLD and weak-CRCQ both hold at $\overline{x}$, while (weak-)Robinson's CQ does not.

Just as it happens in NLP, the weak-CPLD condition is strictly weaker than (weak-)Robinson's CQ, and also weaker than weak-CRCQ, which are in turn, independent. Nevertheless, we will show next that weak-CPLD is still sufficient for the existence of Lagrange multipliers at all local solutions of (NSDP). To do so, we get inspiration from the proof of [12, Thm. 3.1]), originally presented for NLP, and the proof of [5, Thm. 3.2]. That is, we analyse the sequence from Theorem 1 in terms of the spectral decomposition of its approximate Lagrange multiplier candidates, under weak-CPLD. Then, we use Carathéodory's Lemma 1 to construct a bounded sequence from it, that converges to a Lagrange multiplier. As an intermediary step, we also obtain a convergence result of the external penalty method to KKT points under weak-CPLD, a fact that is emphasized in the statement of the next theorem. The above statements regarding existence of Lagrange

multipliers and convergence of the external penalty are also true for weak-CRCQ. In fact, because it is understood that almost all properties of weak-CPLD that we study in this paper are naturally carried over to weak-CRCQ, we shall omit the latter from this point onwards to avoid redundancy.

Theorem 2 *Let $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ and $\{x^k\}_{k\in\mathbb{N}} \to \overline{x} \in \mathcal{F}$ be such that*

$$\nabla_x L\left(x^k, \rho_k \Pi_{\mathbb{S}^m_+}\left(-G\left(x^k\right)\right)\right) \to 0.$$

If $\overline{x}$ satisfies weak-CPLD, then $\overline{x}$ satisfies the KKT conditions. In particular, every local minimizer of (NSDP) *that satisfies weak-CPLD also satisfies the KKT conditions.*

Proof Let $Y^k \doteq \rho_k \Pi_{\mathbb{S}^m_+}\left(-G\left(x^k\right)\right)$, for every $k \in \mathbb{N}$. Recall that we assume $\lambda_1\left(-G\left(x^k\right)\right) \geq \ldots \geq \lambda_m\left(-G\left(x^k\right)\right)$, for every k, and denote by r the rank of $\text{Ker}G(\overline{x})$. Note that when k is large enough, say greater than some k_0, we necessarily have $\lambda_i\left(-G\left(x^k\right)\right) = -\lambda_{m-i+1}\left(G\left(x^k\right)\right) < 0$ for all $i \in \{m-r+1,\ldots,m\}$. Let $I \subseteq_\infty \mathbb{N}$, and $\{E^k\}_{k\in I} \to \overline{E}$ be such that $E^k \in \mathcal{E}_r\left(G\left(x^k\right)\right)$ for every $k \in I$, as described in Definition 5. Then, for each $k \in I$ greater than k_0, the spectral decomposition of Y^k is given by

$$Y^k = \sum_{i=1}^{m-r} \alpha_i^k e_i^k \left(e_i^k\right)^\top,$$

where $\alpha_i^k \doteq \left[\rho_k \lambda_i\left(-G\left(x^k\right)\right)\right]_+ \geq 0$ and e_i^k denotes the i-th column of E^k, for every $i \in \{1,\ldots,m-r\}$. Since $\nabla_x L\left(x^k, Y^k\right) \to 0$, we have

$$\nabla f\left(x^k\right) - \sum_{i=1}^{m-r} \alpha_i^k DG\left(x^k\right)^* \left[e_i^k \left(e_i^k\right)^\top\right] \to 0, \tag{8}$$

but note that

$$DG\left(x^k\right)^* \left[e_i^k \left(e_i^k\right)^\top\right] = \begin{bmatrix} \langle D_{x_1} G\left(x^k\right), e_i^k \left(e_i^k\right)^\top \rangle \\ \vdots \\ \langle D_{x_n} G\left(x^k\right), e_i^k \left(e_i^k\right)^\top \rangle \end{bmatrix} = \begin{bmatrix} \left(e_i^k\right)^\top D_{x_1} G\left(x^k\right) e_i^k \\ \vdots \\ \left(e_i^k\right)^\top D_{x_n} G\left(x^k\right) e_i^k \end{bmatrix} = v_{ii}\left(x^k, E^k\right),$$

so we can rewrite (8) as

$$\nabla f\left(x^k\right) - \sum_{i=1}^{m-r} \alpha_i^k v_{ii}\left(x^k, E^k\right) \to 0.$$

Using Carathéodory's Lemma 1 for the family $\left\{v_{ii}\left(x^k, E^k\right)\right\}_{i\in\{1,\ldots,m-r\}}$, for each fixed $k \in I$, we obtain some $J^k \subseteq \{1,\ldots,m-r\}$ such that $\{v_{ii}\left(x^k, E^k\right)\}_{i\in J^k}$ is linearly independent and

$$\nabla f\left(x^k\right) - \sum_{i=1}^{m-r} \alpha_i^k v_{ii}\left(x^k, E^k\right) = \nabla f\left(x^k\right) - \sum_{i\in J} \tilde{\alpha}_i^k v_{ii}\left(x^k, E^k\right), \tag{9}$$

where $\tilde{\alpha}_i^k \geq 0$ for every $k \in I$ and every $i \in J^k$. By the pigeonhole principle, we can assume J^k is the same, say equal to J, for infinitely many $k \in I$. Without loss of generality,

let us say that this holds for all $k \in I$. We claim that the sequences $\{\tilde{\alpha}_i^k\}_{k \in I}$ are all bounded. In order to prove this, suppose that

$$m^k \doteq \max_{i \in J} \left\{ \tilde{\alpha}_i^k \right\}$$

is unbounded with $k \in I$, divide (9) by m^k and note that $m^k \to \infty$ on a subsequence implies that the vectors $v_{ii}(\overline{x}, \overline{E})$, $i \in J$, are positively linearly dependent. On the other hand, the vectors $v_{ii}(x^k, E^k)$, $i \in J$, are linearly independent for all large k, which contradicts weak-CPLD. Finally, note that every collection of limit points $\{\overline{\alpha}_i : i \in J\}$ of their respective sequences $\{\tilde{\alpha}_i\}_{k \in I}$, $i \in J$, generates a Lagrange multiplier associated with $\overline{x}$, which is $\overline{Y} \doteq \sum_{i \in J} \overline{\alpha}_i u_i(G(\overline{x})) u_i(G(\overline{x}))^\top$. Here, $u_i(G(\overline{x}))$ is the eigenvector associated with the i-th eigenvalue of $G(\overline{x})$ (see (1)). Thus, $\overline{x}$ is a KKT point.

The second part of the statement of the theorem follows from Theorem 1. $\square$

Back to Example 1, observe that weak-CPLD does not hold at $\overline{x} = 0$, as expected. Indeed, for any sequence $\{x^k\}_{k \in \mathbb{N}} \to 0$ such that $x^k < 0$ for all k, the matrix $G(x^k)$ has only simple eigenvalues, for all large k, so $E^k \in \mathcal{E}_r(G(x^k))$ is unique up to sign. Without loss of generality, we can assume

$$E^k \doteq \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix},$$

and then we have $v_{11}(x^k, E^k) = -2x^k > 0$, which is linearly independent for all k while $v_{11}(\overline{x}, \overline{E}) = 0$ is positively linearly dependent. Thus weak-CPLD as in Definition 5 is not satisfied.

Remark 1 In [8], we presented a different extension proposal of CPLD to NSDP problems with multiple constraints, which is weaker than Robinson's CQ for a single constraint as in (NSDP) only when the zero eigenvalue of $G(\overline{x})$ is simple. We called this definition the "naive extension of CPLD". We remark that Definition 5 coincides with the naive extension of CPLD when zero is a simple eigenvalue of $G(\overline{x})$, which makes Definition 5 an improvement of it, or a "non-naive variant" of it.

The phrasing of Theorem 2 was chosen to draw the reader's attention to the fact that it is, essentially, a convergence proof of the external penalty method to KKT points, under weak-CPLD. To obtain a more general convergence result, in the next section we introduce new constant rank-type CQs for NSDP that support every algorithm that converges with a more general type of sequential optimality condition. Then, we prove some properties of these new conditions, and we compare them with weak-CPLD and weak-CRCQ.

4 Stronger Sequential-Type Constant Rank CQs for NSDP and Global Convergence of Algorithms

The so-called *Approximate Karush-Kuhn-Tucker* (AKKT) condition, which was brought from NLP to NSDP by Andreani et al. [11], is a general purpose sequential optimality condition that encompasses, for instance, Theorem 1. Such a notion has been recently refined and generalized to optimization problems in Banach spaces [16] and also problems with

geometric constraints [23]. Let us recall one (the most convenient for this paper) of its many characterizations.[2]

Definition 6 (Def. 4 of [4]) We say that a point $\overline{x} \in \mathcal{F}$ satisfies the AKKT condition when there exist sequences $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ and $\{Y^k\}_{k \in \mathbb{N}} \subseteq \mathbb{S}_+^m$, and perturbation sequences $\{\delta^k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ and $\{\Delta^k\}_{k \in \mathbb{N}} \subseteq \mathbb{S}^m$, such that:

1. $\nabla_x L\left(x^k, Y^k\right) = \delta^k$, for every $k \in \mathbb{N}$;
2. $G\left(x^k\right) + \Delta^k \succeq 0$ and $\langle G\left(x^k\right) + \Delta^k, Y^k \rangle = 0$, for every $k \in \mathbb{N}$;
3. $\Delta^k \to 0$ and $\delta^k \to 0$.

Note that $\{Y^k\}_{k \in \mathbb{N}}$ is a sequence of approximate Lagrange multipliers of $\overline{x}$, in the sense that Y^k is an exact Lagrange multiplier, at $x = x^k$, for the perturbed problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad f(x) + \langle \overline{x} - x, \delta^k \rangle,$$
$$\text{subject to} \quad G(x) + \Delta^k \succeq 0.$$

The main goal in enlarging the class of approximate Lagrange multipliers Y^k and perturbations Δ^k as in Definition 6 instead of considering only the ones given by Theorem 1, is to capture the output sequences of a larger class of iterative algorithms. In the next two subsections, we illustrate the previous statement. What is remarkable is that the proof of Theorem 2 can still be somewhat conducted considering this more general class of sequences, arriving at strong global convergence results for such algorithms (Theorem 5).

4.1 Example 1: A Safeguarded Augmented Lagrangian Method

Let us briefly recall a variant of the *Powell-Hestenes-Rockafellar* augmented Lagrangian algorithm that employs a *safeguarding* technique, which is the direct generalization of the one studied in [14]. The variant we use is also a generalization of [9, Pg. 13] and [11, Alg. 1], for instance.

For an arbitrary penalty parameter $\rho > 0$ and a *safeguarded multiplier* $\tilde{Y} \succeq 0$, we define $L_{\rho, \tilde{Y}} : \mathbb{R}^n \to \mathbb{R}$ as the *Augmented Lagrangian function* of (NSDP), which is given by

$$L_{\rho, \tilde{Y}}(x) \doteq f(x) + \frac{\rho}{2} \left\| \Pi_{\mathbb{S}_+^m} \left(-G(x) + \frac{\tilde{Y}}{\rho} \right) \right\|^2 - \frac{1}{2\rho} \left\| \tilde{Y} \right\|^2.$$

Since it will be useful in the convergence proof, we compute the gradient of $L_{\rho, \tilde{Y}}$ at x below:

$$\nabla L_{\rho, \tilde{Y}}(x) = \nabla f(x) - DG(x)^* \left[\rho \Pi_{\mathbb{S}_+^m} \left(-G(x) + \frac{\tilde{Y}}{\rho} \right) \right]. \tag{10}$$

Now, we state the algorithm:

---

[2]Definition 6 coincides with the AKKT condition presented in [11, Def. 3.1]. See, for instance, [4, Prop. 4].

Input: A sequence $\{\varepsilon_k\}_{k \in \mathbb{N}}$ of positive scalars such that $\varepsilon_k \to 0$; a nonempty convex compact set $\mathcal{B} \subset \mathbb{S}_+^m$; real parameters $\tau > 1$, $\sigma \in (0, 1)$, and $\rho_1 > 0$; and initial points $(x^0, \tilde{Y}^1) \in \mathbb{R}^n \times \mathcal{B}$. Also, define $\|V^0\| = \infty$.

Initialize $k \leftarrow 1$. Then:
Step 1 *(Solving the subproblem):* Compute an approximate stationary point x^k of $L_{\rho_k, \tilde{Y}^k}(x)$, that is, a point x^k such that

$$\|\nabla L_{\rho_k, \tilde{Y}^k}\left(x^k\right)\| \leq \varepsilon_k;$$

Step 2 *(Updating the penalty parameter):* Calculate

$$V^k \doteq \Pi_{\mathbb{S}_+^m} \left(-G\left(x^k\right) + \frac{\tilde{Y}^k}{\rho_k}\right) - \frac{\tilde{Y}^k}{\rho_k}; \tag{11}$$

Then,

a.   If $k = 1$ or $\|V^k\| \leq \sigma \|V^{k-1}\|$, set $\rho_{k+1} \doteq \rho_k$;
b.   Otherwise, take ρ_{k+1} such that $\rho_{k+1} \geq \tau \rho_k$.

Step 3 *(Estimating a new safeguarded multiplier):* Choose any $\tilde{Y}^{k+1} \in \mathcal{B}$, set $k \leftarrow k + 1$ and go to Step 1.

Algorithm 1   Safeguarded augmented Lagrangian method.

By the definition of projection we have that $\tilde{Y}^k = \Pi_{\mathbb{S}_+^m}\left(\tilde{Y}^k - \rho_k G\left(x^k\right)\right)$ if, and only if, $\tilde{Y}^k \in \mathbb{S}_+^m$, $G\left(x^k\right) \in \mathbb{S}_+^m$ and $\langle \tilde{Y}^k, G\left(x^k\right)\rangle = 0$, which means that $V^k = 0$ if, and only if, the pair $(x^k, \tilde{Y}^k)$ is primal-dual feasible and complementary. Moreover, note that Algorithm 1 does not necessarily keep a record of the approximate multiplier sequence associated with $\{x^k\}_{k \in \mathbb{N}}$, which is

$$Y^k \doteq \rho_k \Pi_{\mathbb{S}_+^m}\left(-G\left(x^k\right) + \frac{\tilde{Y}^k}{\rho_k}\right). \tag{12}$$

These are usually computed, however, in several practical implementations of it, where $\tilde{Y}^{k+1}$ is chosen as the projection of Y^k onto $\mathcal{B}$. Also, with these multipliers at hand, it is very easy to prove that any feasible limit point $\overline{x}$ of $\{x^k\}_{k \in \mathbb{N}}$ must satisfy the AKKT condition:

Theorem 3 *Fix any choice of parameters in Algorithm 1 and let $\{x^k\}_{k \in \mathbb{N}}$ be the output sequence generated by it. If $\{x^k\}_{k \in \mathbb{N}}$ has a convergent subsequence $\{x^k\}_{k \in I} \to \overline{x}$, then:*

1.   *The point $\overline{x}$ is stationary for the problem of minimizing $\frac{1}{2}\|\Pi_{\mathbb{S}_+^m}(-G(x))\|^2$;*
2.   *If $\overline{x}$ is feasible, then $\overline{x}$ satisfies the AKKT condition.*

Proof Let $\{\varepsilon_k\}_{k \in \mathbb{N}} \to 0$, $\left\{\tilde{Y}^k\right\}_{k \in \mathbb{N}} \subset \mathcal{B} \subset \mathbb{S}_+^m$, $\tau > 1$, $\sigma \in (0, 1)$, and $\rho_1 > 0$ be given as required by Algorithm 1. Moreover, let $\{\rho_k\}_{k \in \mathbb{N}}$ and $\left\{V^k\right\}_{k \in \mathbb{N}}$ be computed as in Step 2. For simplicity, let us also assume that $I = \mathbb{N}$.

1.   This part of the proof is standard; see, for instance, [4, Prop. 4.3];

2. Define $\{Y^k\}_{k \in \mathbb{N}}$ as in (12) and take $\Delta^k \doteq V^k$ for all $k \in \mathbb{N}$, where V^k is as given in (11). Then, it follows from Step 1 that $\nabla_x L \left(x^k, Y^k \right) = \nabla L_{\rho_k, \tilde{Y}^k} \left(x^k \right) \to 0$. We also have

$$G \left(x^k \right) + \Delta^k = \Pi_{\mathbb{S}^m_+} \left(G \left(x^k \right) - \frac{\tilde{Y}^k}{\rho_k} \right)$$

for every $k \in \mathbb{N}$, which yields $\langle Y^k, G \left(x^k \right) + \Delta^k \rangle = 0$ for every k. If $\rho_k \to \infty$, then recall that $\left\{ \tilde{Y}^k \right\}_{k \in \mathbb{N}}$ remains bounded and thus $V^k \to \Pi_{\mathbb{S}^m_+} (-G(\overline{x}))$ by definition and $\Pi_{\mathbb{S}^m_+} (-G(\overline{x})) = 0$ because $\overline{x}$ is assumed to be feasible; on the other hand, if ρ_k remains bounded, then $V^k \to 0$ due to Step 2-a. Therefore, $\Delta^k \to 0$ and $\overline{x}$ satisfies the AKKT condition.

$\square$

Note that when $\tilde{Y}^k$ is set as zero for every k, then Algorithm 1 reduces to the external penalty method, meaning Theorem 3 also covers this method.

4.2 Example 2: A Sequential Quadratic Programming Method

Next, we recall Correa and Ramírez's [17] *sequential quadratic programming* (SQP) method:

Input: A real parameter $\tau > 1$, a pair of initial points $\left(x^1, Y^0 \right) \in \mathbb{R}^n \times \mathbb{S}^m_+$, and an approximation of $\nabla^2_x L \left(x^1, Y^0 \right)$ given by H^1.
Initialize $k \leftarrow 1$. Then:
Step 1 *(Solving the subproblem):* Compute a solution d^k, together with its Lagrange multiplier Y^k, of the problem

$$\begin{aligned} \underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad & d^\top H^k d + \nabla f \left(x^k \right)^\top d, \\ \text{subject to} \quad & G \left(x^k \right) + DG \left(x^k \right) d \in \mathbb{S}^m_+, \end{aligned} \tag{Lin-QP}$$

and if $d^k = 0$, stop;
Step 2 *(Step corrections):* Perform line search to find a steplength $\alpha^k \in (0,1)$ satisfying *Armijo's rule*

$$f \left(x^k + \alpha^k d^k \right) - f \left(x^k \right) \leq \tau \alpha^k \nabla f \left(x^k \right)^\top d^k. \tag{13}$$

Step 3 *(Approximating the Hessian):* Set $x^{k+1} \leftarrow x^k + \alpha^k d^k$, compute a positive definite approximation H^{k+1} of $\nabla^2_x L \left(x^{k+1}, Y^k \right)$, set $k \leftarrow k+1$, and go to Step 1.

Algorithm 2 General SQP method.

The SQP algorithm generates AKKT sequences as well, as it can be seen in the following proposition:

Proposition 4 *Assume that Step 1 of Algorithm 2 is always well defined. If there is an infinite subset $I \subseteq_\infty \mathbb{N}$ such that $\lim_{k \in I} d^k = 0$ and $\{\|H^k\|\}_{k \in I}$ is bounded, then any limit point $\overline{x}$ of $\{x^k\}_{k \in I}$ satisfies the AKKT condition.*

Proof By the KKT conditions for (Lin-QP), there exists some $Y^k \succeq 0$ such that

$$\nabla f\left(x^k\right) + H^k d^k - DG\left(x^k\right)^* [Y^k] = 0, \tag{14}$$

$$\langle G\left(x^k\right) + DG\left(x^k\right) d^k, Y^k \rangle = 0. \tag{15}$$

Set $\Delta^k \doteq DG\left(x^k\right) d^k$ for every $k \in I$ and since $d^k \to 0$, we obtain that $\lim_{k \in I} H^k d^k = 0$ and $\lim_{k \in I} \Delta^k = 0$. Moreover, since d^k is feasible, $G\left(x^k\right) + \Delta^k \succeq 0$. Thus, $\overline{x}$ satisfies the AKKT condition. □

The hypothesis on the convergence of a subsequence of $\{d^k\}_{k \in \mathbb{N}}$ to zero, directly or indirectly, is somewhat common regarding some types of SQP methods, as well as the boundedness of H^k – see, for instance, [10, 17, 27]. Moreover, the constant rank condition that will be presented in the next section ensures that (Lin-QP) is well defined as long as $d = 0$ is feasible for all k such that x^k is sufficiently close to $\overline{x}$ (see Remark 3 for a discussion).

4.3 Sequential Constant Rank CQs for NSDP

Inspired by the AKKT condition, we are led to introduce a small perturbation in weak-CPLD and weak-CRCQ, which makes them stronger, but also brings some useful properties in return. At first, we present it in a form that resembles Definition 5, for comparison purposes. Later, for convenience, we will provide a characterization of it without sequences.

Definition 7 (seq-CRCQ and seq-CPLD) Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. We say that $\overline{x}$ satisfies the

1. *Sequential CRCQ condition for NSDP* (seq-CRCQ) if $r = m$ or, for all sequences $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ and $\{\Delta^k\}_{k \in \mathbb{N}} \subseteq \mathbb{S}^m$ with $\Delta^k \to 0$, there exists $\overline{E} \in \mathrm{Lim\,sup}_{k \in \mathbb{N}} \mathcal{E}_r \left(G\left(x^k\right) + \Delta^k\right)$ such that, for every subset $J \subseteq \{1, \ldots, m-r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\{v_{ii}\left(x^k, E^k\right)\}_{i \in J}$ remains linearly dependent, for all $k \in I$ large enough.
2. *Sequential CPLD condition for NSDP* (seq-CPLD) if $r = m$ or, for all sequences $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ and $\{\Delta^k\}_{k \in \mathbb{N}} \subseteq \mathbb{S}^m$ with $\Delta^k \to 0$, there exists $\overline{E} \in \mathrm{Lim\,sup}_{k \in \mathbb{N}} \mathcal{E}_r \left(G\left(x^k\right) + \Delta^k\right)$ such that, for every subset $J \subseteq \{1, \ldots, m-r\}$: if $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is positively linearly dependent, then $\{v_{ii}\left(x^k, E^k\right)\}_{i \in J}$ remains linearly dependent, for all $k \in I$ large enough.

For both definitions, $I \subseteq_\infty \mathbb{N}$, and $\{E^k\}_{k \in I}$ is a sequence converging to $\overline{E}$ and such that $E^k \in \mathcal{E}_r \left(G\left(x^k\right) + \Delta^k\right)$ for every $k \in I$, as required by the Painlevé-Kuratowski outer limit.

Note that the only difference between Definitions 5 and 7 is the perturbation matrix $\Delta^k \to 0$. In particular, set $\Delta^k \doteq 0$ for every k to see that seq-CRCQ and seq-CPLD imply weak-CRCQ and weak-CPLD, respectively. Moreover, both implications are strict, as we can see in the following example:

Example 4 Consider the constraint

$$G(x) \doteq \begin{bmatrix} x & 0 \\ 0 & -x \end{bmatrix} \succeq 0$$

at the point $\bar{x} = 0$, so in this case $r = 0$. For every sequence $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$, we have

$$\mathcal{E}_r \left(G\left(x^k \right) \right) \subset \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\},$$

for every $k \in \mathbb{N}$ such that $x^k \neq \bar{x}$, whereas if $x^k = \bar{x}$, then $\mathcal{E}_r \left(G\left(x^k \right) \right)$ is the set of all orthogonal 2×2 matrices. Let us assume without loss of generality that $x^k \leq 0$, thus we may take $E^k = \mathbb{I}_2$ for every $k \in \mathbb{N}$ to see that both, weak-CRCQ and weak-CPLD, hold at $\bar{x}$, since

$$v_{11} \left(x^k, E^k \right) = 1 \quad \text{and} \quad v_{22} \left(x^k, E^k \right) = -1$$

are nonzero and (positively) linearly dependent for every $k \in \mathbb{N}$.

On the other hand, take

$$\Delta^k \doteq \begin{bmatrix} 2x^k \left(x^k + 1 \right)^2 & x^k \left(x^k + 1 \right) \\ x^k \left(x^k + 1 \right) & 3x^k + x^k \left(x^k + 1 \right)^2 \end{bmatrix},$$

with $x^k \to \bar{x}$, $x^k > 0$, and note that the eigenvectors of

$$\begin{aligned}
G\left(x^k \right) + \Delta^k &= x^k \begin{bmatrix} 2\left(x^k\right)^2 + 4x^k + 3 & x^k + 1 \\ x^k + 1 & \left(x^k\right)^2 + 2x^k + 3 \end{bmatrix} \\
&= \begin{bmatrix} -1 & x^k + 1 \\ x^k + 1 & 1 \end{bmatrix} \begin{bmatrix} x^k & 0 \\ 0 & 2x^k \end{bmatrix} \begin{bmatrix} -1 & x^k + 1 \\ x^k + 1 & 1 \end{bmatrix}
\end{aligned}$$

are uniquely determined up to sign. Then, because $v_{ii}(x, E)$, $i \in \{1, 2\}$, is invariant to the sign of the columns of E, we can assume without loss of generality that any $E^k \in \mathcal{E}_r \left(G\left(x^k \right) + \Delta^k \right)$ has the form

$$E^k = \frac{1}{\sqrt{1 + (x^k + 1)^2}} \begin{bmatrix} -1 & x^k + 1 \\ x^k + 1 & 1 \end{bmatrix}$$

for every $k \in \mathbb{N}$. Then, for any sequence $\{E^k\}_{k \in \mathbb{N}}$ such that $E^k \in \mathcal{E}_r \left(G\left(x^k \right) + \Delta^k \right)$ for every k, we have

$$v_{11} \left(x^k, E^k \right) = 1 - (x^k + 1)^2 \quad \text{and} \quad v_{22} \left(x^k, E^k \right) = (x^k + 1)^2 - 1,$$

which are both nonzero, but if $\overline{E}$ is a limit point of $\{E^k\}_{k \in \mathbb{N}}$, then $v_{11}(\bar{x}, \overline{E}) = v_{22}(\bar{x}, \overline{E}) = 0$. Thus, neither seq-CRCQ nor seq-CPLD hold at $\bar{x}$.

Furthermore, since nondegeneracy can be characterized as the linear independence of $v_{ii}(\bar{x}, \overline{E})$, $i \in \{1, \ldots, m - r\}$, for every $\overline{E} \in \mathcal{E}_r \left(G(\bar{x}) \right)$ [5, Prop. 3.2], we observe that it implies seq-CRCQ (see also Remark 2 at the end of this section), but this implication is also strict. Let us show this with a counterexample.

Example 5 We analyse the constraint

$$G(x) \doteq \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} \succeq 0$$

at the point $\overline{x} \doteq 0$. For any $x \in \mathbb{R}$ and any arbitrary orthogonal matrix $E \in \mathbb{R}^{2 \times 2}$, note that E has the form

$$E = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \text{ if } \det(E) = 1 \quad \text{or} \quad E = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}, \text{ if } \det(E) = -1 \qquad (16)$$

where $a^2 + b^2 = 1$. In both cases, we have

$$v_{11}(x, E) = v_{22}(x, E) = a^2 + b^2 = 1.$$

That is, $v_{11}(x, E)$ and $v_{22}(x, E)$ are nonzero and linearly dependent, regardless of x and E. Thus, seq-CRCQ holds at $\overline{x}$, although nondegeneracy does not. Note that weak-nondegeneracy also fails at $\overline{x}$, in this example.

In Example 2, we verify that Robinson's CQ does not imply seq-CRCQ; because otherwise, it would also imply weak-CRCQ, contradicting the example. As for the converse, the counterexample below shows that seq-CRCQ does not imply Robinson's CQ either.

Example 6 Consider the constraint

$$G(x) \doteq \begin{bmatrix} x_1 & x_2 \\ x_2 & -x_1 \end{bmatrix} \succeq 0.$$

Clearly, the only feasible point is $\overline{x} = 0$. Then, due to the linearity of G, it is immediate to see that Robinson's CQ does not hold at $\overline{x} = 0$. On the other hand, for any $x \in \mathbb{R}^2$ and any orthogonal matrix $E \in \mathbb{R}^{2 \times 2}$, note that regardless of the form of E as in (16), we have $v_{11}(x, E) \neq 0$, $v_{22}(x, E) \neq 0$, and

$$v_{11}(x, E) = -v_{22}(x, E).$$

Thus, seq-CRCQ holds at $\overline{x} = 0$; see also the characterization of Proposition 6.

The same argument that is used to verify that nondegeneracy implies seq-CRCQ can also be adapted to show that Robinson's CQ implies seq-CPLD in view of [5, Prop. 5.1]. Then, Example 6 tells us that seq-CPLD is actually strictly weaker than Robinson's CQ.

Next, we will show that seq-CPLD (and, consequently, seq-CRCQ) is enough to establish equivalence between AKKT and KKT with a small adaptation of the proof of Theorem 2. Note that in view of Theorem 1, any condition that establishes that an AKKT point is also a KKT point is, in particular, a CQ; in addition, such a CQ necessarily supports the global convergence of any algorithm supported by AKKT to KKT points. This includes the algorithms presented in Sections 4.1 and 4.2, and Yamashita, Yabe, and Harada's primal-dual interior point method for NSDP [32] – for details on the latter, see [3]. We should also stress that this convergence result neither assumes compactness of the Lagrange multiplier set nor that it is a singleton.

Theorem 5 *Let $\overline{x} \in \mathcal{F}$ be an AKKT point that satisfies seq-CPLD. Then, $\overline{x}$ satisfies the KKT conditions.*

Proof Let $\{x^k\}_{k \in \mathbb{N}} \rightarrow \overline{x}$, $\{Y^k\}_{k \in \mathbb{N}} \subseteq \mathbb{S}_+^m$, and $\{\tilde{\Delta}^k\}_{k \in \mathbb{N}} \rightarrow 0$ be the AKKT sequences from Definition 6. Let r be the rank of $G(\overline{x})$ and recall that $\langle G(x^k) + \tilde{\Delta}^k, Y^k \rangle = 0$ implies

that $G\left(x^k\right) + \tilde{\Delta}^k$ and Y^k are simultaneously diagonalizable by some orthogonal matrix $U^k \doteq [u_1^k, \ldots, u_m^k] \in \mathbb{R}^{m \times m}$, such that

$$0 = \langle G\left(x^k\right) + \tilde{\Delta}^k, Y^k \rangle = \left\langle (U^k)^\top (G\left(x^k\right) + \tilde{\Delta}^k)U^k, \ (U^k)^\top Y^k U^k \right\rangle$$

$$= \sum_{i=1}^m \lambda_i \left(G\left(x^k\right) + \tilde{\Delta}^k\right) \lambda_{m-i+1}(Y^k)$$

and because all terms above are nonnegative it follows that, for every $i \in \{1, \ldots, m\}$, the product $\lambda_i(G\left(x^k\right) + \tilde{\Delta}^k)\lambda_{m-i+1}(Y^k)$ is equal to zero. Also, because $\lambda_i\left(G\left(x^k\right)\right) > 0$ for every $i \in \{1, \ldots, r\}$, then $\lambda_i\left(G\left(x^k\right) + \tilde{\Delta}^k\right) > 0$ and $\lambda_{m-i+1}(Y^k) = 0$ for every such i and all k large enough.

Hence, assuming without loss of generality that the columns of U^k are ordered accordingly to the eigenvalues of Y^k, the spectral decomposition of Y^k can be represented in the format

$$Y^k = \sum_{i=1}^{m-r} \lambda_i \left(Y^k\right) u_i^k \left(u_i^k\right)^\top.$$

Defining $E^k = \left[u_1^k, \ldots, u_{m-r}^k\right]$ for every k, we obtain

$$\nabla_x L\left(x^k, Y^k\right) = \nabla f\left(x^k\right) - \sum_{i=1}^{m-r} \lambda_i(Y^k) v_{ii}\left(x^k, E^k\right) \to 0.$$

Also, for each $k \in \mathbb{N}$, construct

$$M^k \doteq U^k \tilde{M}^k \left(U^k\right)^\top, \tag{17}$$

where

$$\tilde{M}^k \doteq U^k \left[\begin{array}{c|c} \text{Diag}\left(\lambda_1\left(G\left(x^k\right)\right), \ldots, \lambda_r\left(G\left(x^k\right)\right)\right) & 0 \\ \hline 0 & \text{Diag}\left((r+1)\|x^k - \overline{x}\|, \ldots, m\|x^k - \overline{x}\|\right) \end{array} \right].$$

Note that $M^k \to G(\overline{x})$ and that the $m - r$ smallest eigenvalues of M^k are simple, if $x^k \neq \overline{x}$, meaning their associated eigenvectors are unique up to sign, when k is large enough. Consequently, $v_{ii}\left(x^k, E^k\right)$ is invariant to the choice of $E^k \in \mathcal{E}_r\left(M^k\right)$, for all such k, and every $i \in \{1, \ldots, m - r\}$. Define $\Delta^k \doteq M^k - G\left(x^k\right)$ for every $k \in \mathbb{N}$ and the rest of this proof follows the exact same lines as the proof of Theorem 2. $\qquad\square$

Remark 2 The "perturbed versions" of weak-nondegeneracy and weak-Robinson's CQ, in the sense of Definition 7, are nondegeneracy and Robinson's CQ, respectively. In other words, nondegeneracy (respectively, Robinson's CQ) holds at $\overline{x} \in \mathcal{F}$ if, and only if, for every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and every $\{\Delta^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m$ such that $\Delta^k \to 0$, there is some $\overline{E} \in \text{Lim}\sup_{k\in\mathbb{N}} \mathcal{E}_r\left(G\left(x^k\right) + \Delta^k\right)$ such that $\{v_{ii}(\overline{x}, \overline{E}): i \in \{1, \ldots, m - r\}\}$ is (positively) linearly independent, where $r = \text{rank}(G(\overline{x}))$. For more details, see [5, Rem. 3.1].

We end this section with a characterization of seq-CRCQ and seq-CPLD without sequences, which may be better suited for other applications.

Proposition 6 *Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$.*

- *seq-CRCQ holds at $\overline{x}$ if, and only if, $r = m$ or, for every $\overline{E} \in \mathcal{E}_r\,(G(\overline{x}))$, there exists some neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that for all $J \subseteq \{1, \ldots, m - r\}$, we have that if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\{v_{ii}(x, E)\}_{i \in J}$ remains linearly dependent for every $(x, E) \in \mathcal{V}$ such that E has orthonormal columns;*
- *seq-CPLD holds at $\overline{x}$ if, and only if, $r = m$ or, for every $\overline{E} \in \mathcal{E}_r\,(G(\overline{x}))$, there exists some neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that for all $J \subseteq \{1, \ldots, m - r\}$, we have that if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is positively linearly dependent, then $\{v_{ii}(x, E)\}_{i \in J}$ remains linearly dependent for every $(x, E) \in \mathcal{V}$ such that E has orthonormal columns.*

Proof We will prove only the first statement since the second one follows analogously. Let $\overline{x}$ satisfy seq-CRCQ; by contradiction: suppose that there exists some $\overline{E} \in \mathcal{E}_r\,(G(\overline{x}))$, some $J \subseteq \{1, \ldots, m - r\}$, and some sequence $\left\{\left(x^k, E^k\right)\right\}_{k \in \mathbb{N}} \to (\overline{x}, \overline{E})$ such that $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, but $\{v_{ii}\left(x^k, E^k\right)\}_{i \in J}$ is linearly independent for every large $k \in \mathbb{N}$. Let $P^k \in \mathbb{R}^{m \times r}$ be a matrix whose columns are orthogonal eigenvectors associated with the r largest eigenvalues of $G\left(x^k\right)$, define $U^k \doteq [P^k, E^k]$, and consider M^k as in (17). Set $\Delta^k \doteq M^k - G\left(x^k\right)$ and note that $v_{ii}\left(x^k, E^k\right)$ is invariant to $E^k \in \mathcal{E}_r\left(\Delta^k + G\left(x^k\right)\right)$ when k is large, provided that $x^k \neq \overline{x}$. This contradicts seq-CRCQ.

Conversely, let $\left\{x^k\right\}_{k \in \mathbb{N}} \to \overline{x}$ and $\Delta^k \to 0$ be any sequences, and let $J \subseteq \{1, \ldots, m - r\}$ be any subset. For each k, pick any $E^k \in \mathcal{E}_r\left(G\left(x^k\right) + \Delta^k\right)$ and consider the sequence $\left\{E^k\right\}_{k \in \mathbb{N}}$, which is bounded. Let $I \subseteq_\infty \mathbb{N}$ and $\overline{E}$ be arbitrary, as long as $\{E^k\}_{k \in I} \to \overline{E}$, so $\overline{E} \in \mathcal{E}_r\,(G(\overline{x}))$. Then, by hypothesis, there exists a neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that if $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\{v_{ii}\left(x^k, E^k\right)\}_{i \in J}$ is also linearly dependent for all large enough $k \in I$, since $\left(x^k, E^k\right) \in \mathcal{V}$ for all such k. $\qquad\square$

5 Relationship with Metric Subregularity CQ

Besides convergence of algorithms, the CQs we present also have implications towards stability and error analysis. We make this link by means of establishing a relationship between seq-CPLD (and seq-CRCQ) and the so-called *metric subregularity CQ* (also known as the *error bound CQ* in NLP), defined in our SDP framework as follows:

Definition 8 (e.g., Def. 1.1 of [19]) We say that a feasible point $\overline{x}$ of (NSDP) satisfies the *metric subregularity CQ* when there exists some $\gamma > 0$ and a neighborhood $\mathcal{V}$ of $\overline{x}$ such that

$$\text{dist}(x, \mathcal{F}) \leq \gamma \|\Pi_{\mathbb{S}_+^m}(-G(x))\|$$

for every $x \in \mathcal{V}$. That is, when the set-valued mapping $\mathcal{G} \colon \mathbb{R}^n \rightrightarrows \mathbb{S}^m$ that maps $x \mapsto G(x) - \mathbb{S}_+^m$ is metrically subregular at $(\overline{x}, 0) \in \text{graph}(\mathcal{G})$. Here $\text{dist}(x, \mathcal{F})$ denotes the distance between x and $\mathcal{F}$, and $\text{graph}(\mathcal{G}) \subset \mathbb{R}^n \times \mathbb{S}^m$ is the graph of $\mathcal{G}$.

The metric subregularity CQ is implied by Robinson's CQ, which in turn coincides with a similar condition called *metric regularity CQ*, and it has relevant implications on the stability analysis of optimization problems – for details, we refer to Ioffe's survey [20, 21]. Besides, there are several works addressing the relationship between constant rank constraint qualifications and the metric subregularity CQ in NLP, such as Minchenko and Stakhovski [24], Andreani et al. [9], and others.

We will use a sufficient condition for metric subregularity CQ to hold, originally proposed by Minchenko and Stakhovski [24, Thm. 2] for NLP problems. We made a simple

extension of it to NSDP, which seems not to have been done before in the literature. It is worth mentioning, nevertheless, that the proof we present is essentially the same as the original one, with some minor adaptations to the NSDP context via Moreau's decomposition.

Proposition 7 *Let $\overline{x} \in \mathcal{F}$ and assume that G is twice continuously differentiable around $\overline{x}$. For every given $x \notin \mathcal{F}$, let $\Lambda_\Pi(x)$ denote the set of Lagrange multipliers of the problem of minimizing $\|z - x\|$ subject to $G(z) \succeq 0$, $z \in \mathbb{R}^n$. If there exist numbers $\tau > 0$ and $\delta > 0$ such that $\Lambda_\Pi(x) \cap \mathrm{cl}(B(0, \tau)) \neq \emptyset$ for every $x \in B(\overline{x}, \delta)$, $x \notin \mathcal{F}$, then $\overline{x}$ satisfies metric subregularity CQ.*

Proof Let τ and δ be as described in the hypothesis. Following the proof of [24, Thm. 2], note that if $\overline{x} \in \mathrm{int}\mathcal{F}$, then it trivially satisfies metric subregularity CQ, so we will assume that $\overline{x} \in \mathrm{bd}\mathcal{F}$. Let $\delta_0 \in (0, \delta)$ be such that

$$\frac{4}{\delta_0}\mathbb{I}_n - D^2G(z)^*[Y] \succeq 0$$

for all $z \in B(\overline{x}, \delta)$ and all $Y \in \mathrm{cl}(B(0, 2\tau))$, where $D^2G(z)$ is the second-order derivative of G at z and $D^2G(z)^*$ is its adjoint as a linear operator $D^2G(z)\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{S}^m$. Let $x \in B(\overline{x}, \delta_0/2)$ be such that $x \notin \mathcal{F}$. Although $\Pi_\mathcal{F}(x)$ may not be well-defined as a function of x, we will use the notation $\Pi_\mathcal{F}(x)$ to denote an arbitrary minimizer of $\|z - x\|$ subject to $G(z) \succeq 0$. Then, by definition, we have that $\|\Pi_\mathcal{F}(x) - x\| \leq \|\overline{x} - x\| < \delta_0/2$, so $\Pi_\mathcal{F}(x) \in B(x, \delta_0/2)$ and, therefore, $\|\Pi_\mathcal{F}(x) - \overline{x}\| \leq \|\Pi_\mathcal{F}(x) - x\| + \|x - \overline{x}\| < \delta_0$. Let $h\colon \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$ be defined as

$$h(z, Y) \doteq \frac{\langle z - x, z - \Pi_\mathcal{F}(x)\rangle}{\|x - \Pi_\mathcal{F}(x)\|} - \langle G(z), Y\rangle$$

and note that

$$\nabla_z h(z, Y) = \frac{2z - x - \Pi_\mathcal{F}(x)}{\|x - \Pi_\mathcal{F}(x)\|} - DG(z)^*[Y]$$

and

$$\nabla_z^2 h(z, Y) = \frac{2}{\|x - \Pi_\mathcal{F}(x)\|}\mathbb{I}_n - D^2G(z)^*[Y] \succeq \frac{4}{\delta_0}\mathbb{I}_n - D^2G(z)^*[Y] \succeq 0$$

whenever $z \in B(\overline{x}, \delta)$ and $Y \in \mathrm{cl}(B(0, 2\tau))$. Thus, $h(z, Y)$ is convex with respect to its first variable $z \in B(\overline{x}, \delta)$, for every $Y \in \mathrm{cl}(B(0, 2\tau))$. Now let us fix an arbitrary $Y \in \Lambda_\Pi(x) \cap \mathrm{cl}(B(0, \tau))$, which is nonempty by hypothesis. Recall that, by definition of the set $\Lambda_\Pi(x)$, we have that Y is a Lagrange multiplier of the projection problem associated with the point $\Pi_\mathcal{F}(x)$. Hence, $2Y$ is a Lagrange multiplier of the problem:

$$\text{Minimize } \tilde{f}_x(z) \doteq \|z - x\| + \frac{\langle z - x, z - \Pi_\mathcal{F}(x)\rangle}{\|x - \Pi_\mathcal{F}(x)\|}, \quad \text{subject to } G(z) \succeq 0 \tag{18}$$

associated with the point $\Pi_\mathcal{F}(x)$, which is a local minimizer of $\tilde{f}_x$ since it is elementary to check that $\tilde{f}_x(\Pi_\mathcal{F}(x)) \leq \|z - x\|$ for every $z \in \mathcal{F}$, by the definition of projection, with equality at $\Pi_\mathcal{F}(x)$. Writing the KKT conditions for problem (18) at $\Pi_\mathcal{F}(x)$ with respect to the Lagrange multiplier $2Y \in \mathrm{cl}(B(0, 2\tau))$, we obtain

$$\frac{2(\Pi_\mathcal{F}(x) - x)}{\|x - \Pi_\mathcal{F}(x)\|} - DG(\Pi_\mathcal{F}(x))^*[2Y] = 0 \tag{19}$$

with $\langle G(\Pi_{\mathcal{F}}(x)), 2Y \rangle = 0$, which yields

$$
\begin{aligned}
\|x - \Pi_{\mathcal{F}}(x)\| &= -\|x - \Pi_{\mathcal{F}}(x)\| - \langle DG(\Pi_{\mathcal{F}}(x))^*[2Y], x - \Pi_{\mathcal{F}}(x) \rangle \\
&\leq \langle G(\Pi_{\mathcal{F}}(x)) - G(x), 2Y \rangle \\
&= -\langle G(x), 2Y \rangle
\end{aligned}
\tag{20}
$$

after taking inner products of both sides of (19) with $x - \Pi_{\mathcal{F}}(x)$. The middle inequality follows from the definition of adjoint and the convexity of $h(z, Y)$ in the first variable. Taking Moreau's decomposition for $G(x)$, we obtain from (20) that

$$
\|x - \Pi_{\mathcal{F}}(x)\| \leq -\langle \Pi_{\mathbb{S}_+^m}(G(x)), 2Y \rangle + \langle \Pi_{\mathbb{S}_+^m}(-G(x)), 2Y \rangle \leq \langle \Pi_{\mathbb{S}_+^m}(-G(x)), 2Y \rangle,
$$

because $Y \succeq 0$, which is self-dual, so $\langle \Pi_{\mathbb{S}_+^m}(G(x)), 2Y \rangle \geq 0$; then

$$
\mathrm{dist}(x, \mathcal{F}) = \|x - \Pi_{\mathcal{F}}(x)\| \leq \|2Y\| \|\Pi_{\mathbb{S}_+^m}(-G(x))\| \leq 2\tau \|\Pi_{\mathbb{S}_+^m}(-G(x))\|.
$$

Since x was chosen arbitrarily, set $\gamma \doteq 2\tau$ and we are done. $\qquad\square$

Now, to compare metric subregularity CQ with seq-CRCQ and seq-CPLD, we first need to show that they are robust, in the sense they are preserved in a neighborhood of the point of interest.

Proposition 8 *If seq-CPLD holds at $\overline{x}$, then there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that seq-CPLD also holds for every $x \in \mathcal{V} \cap \mathcal{F}$. Moreover, the same property holds for seq-CRCQ.*

Proof If $\mathrm{Ker}\, G(\overline{x}) = \{0\}$, the result follows trivially, so let us assume $r < m$, where r is the rank of $G(\overline{x})$. Suppose that the statement of the theorem is false, then there exists a sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ such that seq-CPLD does not hold at any x^k. Let us denote by r^k the rank of $G(x^k)$, and note that we can assume, without loss of generality, that $r^k = \tilde{r} \geq r$ for all k (by the pigeonhole principle). Notice that in order to state seq-CPLD at a point $\overline{x}$, we rely on the vectors defined in (5), which depend on the rank r of $G(\overline{x})$, even though we evaluate these vector also at $x \neq \overline{x}$. Hence, when stating seq-CPLD at x^k, we make this dependence on the rank $\tilde{r}$ of $G(x^k)$ explicit by using $v_{ii}^{\tilde{r}}(\cdot, \cdot)$ instead of $v_{ii}(\cdot, \cdot)$, which is used when stating seq-CPLD at $\overline{x}$. That is, since seq-CPLD does not hold at x^k, there exists some $E^k \in \mathcal{E}_{\tilde{r}}(G(x^k))$, some $J^k \subseteq \{1, \ldots, m - \tilde{r}\}$, and some sequence $\{(x_\ell^k, E_\ell^k)\}_{\ell \in \mathbb{N}} \to (x^k, E^k)$ such that $\{v_{ii}^{\tilde{r}}(x^k, E^k)\}_{i \in J^k}$ is positively linearly dependent, but $\{v_{ii}^{\tilde{r}}(x_\ell^k, E_\ell^k)\}_{i \in J^k}$ is linearly independent for every $\ell \in \mathbb{N}$. We can assume that J^k is the same, say J, for every k, by the pigeonhole principle. For each k large enough so that $\mathcal{E}_r(G(x^k))$ is well-defined, let $e_{m-\tilde{r}+1}^k, \ldots, e_{m-r}^k \in \mathbb{R}^m$ be the orthonormal eigenvectors of $G(x^k)$ such that the following inclusion holds:

$$
E_{\mathrm{ext}}^k \doteq \left[E^k, e_{m-\tilde{r}+1}^k, \ldots, e_{m-r}^k \right] \in \mathcal{E}_r\left(G(x^k)\right).
$$

Since $(E_{\mathrm{ext}}^k)^\top E_{\mathrm{ext}}^k = I_{m-r}$ for every k, we can assume that $E_{\mathrm{ext}}^k \to \overline{E}$ for some $\overline{E} \in \mathbb{R}^{m \times (m-r)}$. Denote by $v_{ii}(\cdot, \cdot)$ the vectors defined by (5) at $\overline{x}$. Hence, since $J \subseteq \{1, \ldots, m - \tilde{r}\}$ (as a direct consequence of $m - \tilde{r} \leq m - r$), for each $k, \ell \in \mathbb{N}$, it holds that $v_{ii}^{\tilde{r}}(x^k, E^k) = v_{ii}(x^k, E_{\mathrm{ext}}^k)$ for all $i \in J$. Consequently, we have that $\{v_{ii}(x^k, E_{\mathrm{ext}}^k)\}_{i \in J}$ is also positively linearly dependent for every large k, which implies that the same holds for $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$.

On the other hand, for each $k, \ell \in \mathbb{N}$, let $\mathcal{P}_\ell^k$ be the projection matrix onto the orthogonal complement of the space spanned by the columns of E_ℓ^k and consider the extended matrix

$$E_{\ell,\text{ext}}^k \doteq \left[E_\ell^k, \text{gramschmidt}\left(\mathcal{P}_\ell^k e_{m-\tilde{r}+1}^k, \ldots, \mathcal{P}_\ell^k e_{m-r}^k \right) \right] \in \mathbb{R}^{m \times (m-r)},$$

where gramschmidt(M) refers to the outcome of the Gram-Schmidt orthonormalization procedure applied to the columns of M. Notice that since $J \subseteq \{1, \ldots, m - \tilde{r}\}$, for each $k, \ell \in \mathbb{N}$, it holds that $v_{ii}^{\tilde{r}}\left(x_\ell^k, E_\ell^k \right) = v_{ii}\left(x_\ell^k, E_{\ell,\text{ext}}^k \right)$ for all $i \in J$.

Now, for each k, let $\ell(k)$ be such that $\lim_{k \to \infty} \ell(k) = +\infty$ and notice that $x_{\ell(k)}^k \to \overline{x}$ and $\left\{ v_{ii}\left(x_{\ell(k)}^k, E_{\ell(k),\text{ext}}^k \right) \right\}_{i \in J} = \left\{ v_{ii}^{\tilde{r}}\left(x_{\ell(k)}^k, E_{\ell(k)}^k \right) \right\}_{i \in J}$ is linearly independent for every k. Moreover, since $\left(E_{\ell(k),\text{ext}}^k \right)^\top E_{\ell(k),\text{ext}}^k = I_{m-r}$, it follows (passing to a subsequence if necessary) that $E_{\ell(k),\text{ext}}^k \to \tilde{E}$ for some $\tilde{E} \in \mathbb{R}^{m \times (m-r)}$ whose $m - \tilde{r}$ first columns coincide with those of $\overline{E}$. Consequently, $\left\{ v_{ii}(\overline{x}, \tilde{E}) \right\}_{i \in J} = \left\{ v_{ii}(\overline{x}, \overline{E}) \right\}_{i \in J}$, which is positively linearly dependent, contradicting seq-CPLD.

The proof for seq-CRCQ is completely analogous.   □

Now, using Proposition 8, it is possible to prove that seq-CPLD (and seq-CRCQ) implies metric subregularity CQ. We shall do this in the same style as Andreani et al. [9]:

Theorem 9 *If seq-CPLD holds at $\overline{x} \in \mathcal{F}$ and G is twice differentiable around $\overline{x}$, then $\overline{x}$ satisfies metric subregularity CQ.*

Proof Suppose that metric subregularity CQ does not hold at $\overline{x}$. In view of Proposition 7, there exist sequences $\{\tau^k\}_{k \in \mathbb{N}} \to \infty$ and $\{y^k\}_{k \in \mathbb{N}} \to \overline{x}$, $y^k \notin \mathcal{F}$, such that $\Lambda(y^k) \cap \text{cl}\left(B(0, \tau^k) \right) = \emptyset$ for every $k \in \mathbb{N}$.

Now let $\{z^k\}_{k \in \mathbb{N}}$ be such that $z^k = \Pi_\mathcal{F}(y^k)$ for each k and note that $z^k \to \overline{x}$. Indeed, even if $\Pi_\mathcal{F}(y^k)$ is not uniquely defined, we have that $\text{dist}(y^k, \mathcal{F}) = \|z^k - y^k\| \to 0$ because $y^k \to \overline{x} \in \mathcal{F}$ and, thus, $\|z^k - \overline{x}\| \leqslant \|z^k - y^k\| + \|y^k - \overline{x}\| \to 0$. By the previous proposition, z^k satisfies seq-CPLD for all k large enough. Consequently, there exists a sequence $\{Y^k\}_{k \in \mathbb{N}} \subseteq \mathbb{S}_+^m$ such that

$$\frac{z^k - y^k}{\|z^k - y^k\|} - DG(z^k)^*[Y^k] = 0$$

and $\langle G(z^k), Y^k \rangle = 0$ for every k, which implies that $\lambda_i(Y^k) = 0$ for every $i \in \{m - r + 1, \ldots, m\}$ and every $k \in \mathbb{N}$. Let U^k be an arbitrary matrix that diagonalizes Y^k and let E^k be the part of it that corresponds to the $m - r$ smallest eigenvalues of $G(z^k)$. So

$$\frac{z^k - y^k}{\|z^k - y^k\|} - \sum_{i=1}^{m-r} \lambda_i(Y^k) v_{ii}\left(z^k, E^k \right) = 0. \tag{21}$$

Again, by Caratheodory' lemma (cf. Lemma 1) and the pigeonhole principle, we obtain a set $J \subseteq \{1, \ldots, m - r\}$ such that $\{v_{ii}(z^k, E^k) \colon i \in J\}$ is linearly independent and $\sum_{i=1}^{m-r} \lambda_i(Y^k) v_{ii}\left(z^k, E^k \right) = \sum_{i \in J} \alpha_i^k v_{ii}\left(z^k, E^k \right)$ for every k where $\alpha_i^k \lambda_i(Y^k) > 0$ for all $i \in J$. Then, recall from the definition that $Y^k \in \Lambda(y^k)$, so $\|Y^k\| > \tau^k \to \infty$. Let $m^k \doteq \max\{\alpha_i^k \colon i \in J\}$ and divide (21) by m^k to obtain that $\{v_{ii}(\overline{x}, \overline{E}) \colon i \in J\}$ is linearly dependent for every limit point $\overline{E}$ of $\{E^k\}_{k \in \mathbb{N}}$, which contradicts seq-CPLD at $\overline{x}$.   □

Remark 3 A consequence of Theorem 5 together with [4, Lem. 3] is that the fulfilment of seq-CPLD at a point $x \in \mathcal{F}$ implies that the set

$$\mathcal{H}(x) \doteq DG(x)^* \left[N_{\mathbb{S}_+^m}(G(x)) \right]$$

is closed, where $N_{\mathbb{S}_+^m}(G(x))$ is the *normal cone* to $\mathbb{S}_+^m$ at $G(x)$. Moreover, following the discussion in [6, Rem. 5.1] we see that the closedness of $\mathcal{H}(x)$ is the weakest condition that ensures existence of Lagrange multipliers for the solutions of (Lin-QP) for any fixed $x \in \mathcal{F}$. Combining these with Proposition 8 leads us to conclude that if seq-CPLD holds at a solution $\overline{x} \in \mathcal{F}$, then $\mathcal{H}(x)$ is closed for every $x \in \mathcal{F}$ in a neighborhood of $\overline{x}$. Thus, if $d = 0$ is a feasible point (not necessarily a solution) of subproblem (Lin-QP) of Algorithm 2, then its Step 1 is well-defined.

6 Conclusion

There are few constraint qualifications available for NSDP, and as far as we know, the use of CQs in the global convergence of algorithms is somewhat limited to nondegeneracy and Robinson's CQ. In contrast, several constraint qualifications have been defined for NLP over the past decades, mostly improving the global convergence of algorithms beyond the case when the set of Lagrange multipliers is bounded. We are in a path to extend these CQs to conic contexts, such as NSDP, that started in [8]. In fact, the results of this paper can be considered a significant improvement of [8] based on our previous developments in [5]. Our main contribution is two-fold: first, we present a new weak CQ, called weak-CPLD, which was used to prove global convergence of an external penalty method to stationary points. Then, by introducing certain perturbations in the definition of weak-CPLD, we obtained a slightly stronger CQ, called seq-CPLD, that is also strictly weaker than Robinson's CQ but can still be used to prove global convergence of a larger class of algorithms (that encompasses, for instance, augmented Lagrangian, sequential quadratic programming, and interior-point methods) to stationary points. Moreover, we showed that seq-CPLD, just as Robinson's CQ, has a property related with the ability to compute error bounds in the vicinity of the points that satisfy it. Similar extensions of CRCQ with equally interesting properties were also obtained along the way. We believe that several other applications of constant rank CQs will appear in the literature, such as the computation of the derivative of the value function of a parameterized NSDP problem and the computation of second-order necessary optimality conditions. In NLP, constant rank CQs are used to define a strong second-order necessary optimality condition that depends a single Lagrange multiplier, rather than on the full set of Lagrange multipliers, which we believe will be the case for conic problems as well. After a first version of this paper appeared, this last objective has been achieved with a different, more geometric, definition of CRCQ [6]. It is also the case that constant rank conditions provide the adequate assumptions for guaranteeing global convergence of algorithms to second-order stationary points, which has been only partially considered in the conic programming literature in [18].

This paper leaves several interesting open questions that can be addressed in future works, such as the use of weak-CRCQ and weak-CPLD in algorithms other than external penalty methods, and the analysis of some stability properties under the conditions introduced in this manuscript. Some interesting results based on this paper were obtained in the context of second-order cone programming, where some applications of weak-CQs were already obtained [7]. It is also worth recalling that although our conditions were defined
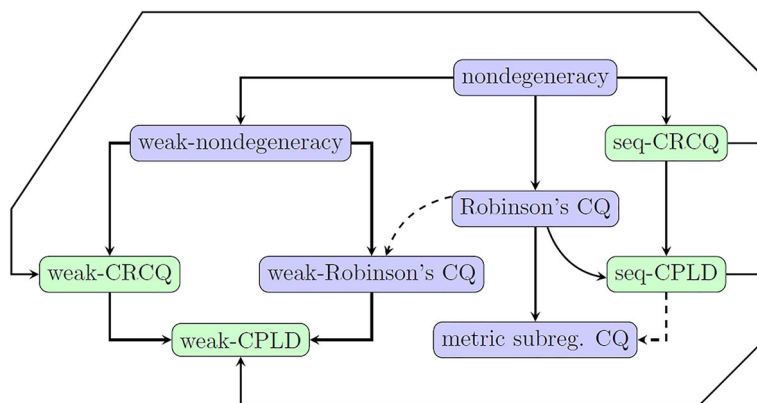
Fig. 1 Relationship among the new constraint qualifications and some of the existing ones

by means of sequences, which seems appropriate when talking about convergence of algorithms, we also provided characterizations of them without sequences, in a more classical way, which should foster new applications.

The relationship among the CQs we presented in this paper, and existing ones, is summarized in the following diagram, where (solid) arrows represent (strict) implications, existing CQs are in blue boxes, and new CQs are in green boxes (Fig. 1).

Declarations

The authors have no relevant financial or non-financial interests to disclose. Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Andreani, R., Echagüe, C.E., Schuverdt, M.L.: Constant-rank condition and second-order constraint qualification. J. Optim. Theory Appl. **146**, 255–266 (2010). https://doi.org/10.1007/s10957-010-9671-8
2. Andreani, R., Fukuda, E.H., Haeser, G., H. Ramírez, C., Santos, D.O., Silva, P.J.S., Silveira, T.P.: Erratum to: New constraint qualifications and optimality conditions for second order cone programs. Set-Valued Variational Anal. **30**, 329–333 (2022). https://doi.org/10.1007/s11228-021-00573-5
3. Andreani, R., Fukuda, E.H., Haeser, G., Santos, D.O., Secchin, L.D.: On the use of Jordan algebras for improving global convergence of an augmented Lagrangian method in nonlinear semidefinite programming. Comput. Optim. Appl. **79**, 633–648 (2021). https://doi.org/10.1007/s10589-021-00281-8
4. Andreani, R., Gómez, W., Haeser, G., Mito, L.M., Ramos, A.: On optimality conditions for nonlinear conic programming. To appear in Mathematics of Operations Research. https://doi.org/10.1287/moor.2021.1203 (2022)
5. Andreani, R., Haeser, G., Mito, L.M., Ramírez, H.: Weak notions of nondegeneracy in nonlinear semidefinite programming, tech. report, arXiv:2012.14810v1 (2022)
6. Andreani, R., Haeser, G., Mito, L.M., Ramírez, H., Silveira, T.P.: First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition, tech. report, arXiv:2107.04693 (2021)
7. Andreani, R., Haeser, G., Mito, L.M., Ramírez, H., Silveira, T.P.: Global convergence of algorithms under constant rank conditions for nonlinear second-order cone programming, To appear in Journal of Optimization Theory and Applications, https://doi.org/10.1007/s10957-022-02056-5 (2022)

8. Andreani, R., Haeser, G., Mito, L.M., Ramírez, H., Santos, D.O., Silveira, T.P.: Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. Optim. Lett. **16**, 589–610 (2021). https://doi.org/10.1007/s11590-021-01737-w

9. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Programm. Ser. A **135**, 255–273 (2012). https://doi.org/10.1007/s10107-011-0456-0

10. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**, 1109–1135 (2012). https://doi.org/10.1137/110843939

11. Andreani, R., Haeser, G., Viana, D.S.: Optimality conditions and global convergence for nonlinear semidefinite programming. Math. Program. Ser. A **180**, 203–235 (2020). https://doi.org/10.1007/s10107-018-1354-5

12. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On the relation between constant positive linear dependence condition and quasinormality constraint qualification. J. Optim. Theory Appl. **125**, 473–485 (2005). https://doi.org/10.1007/s10957-004-1861-9

13. Bertsekas, D.P.: Nonlinear Programming. Athenas Scientific, Belmont Mass (1999)

14. Birgin, E., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. SIAM Publications, Philadelphia (2014)

15. Bonnans, J.F., Shapiro, A.: Pertubation Analysis of Optimization Problems. Springer-Verlag, Berlin (2000)

16. Börgens, E., Kanzow, C., Mehlitz, P., Wachsmuth, G.: New constraint qualifications for optimization problems in Banach spaces based on asymptotic KKT conditions. SIAM J. Optim. **30**, 2956–2982 (2020). https://doi.org/10.1137/19M1306804

17. Correa, R., H. Ramírez, C.: A global algorithm for nonlinear semidefinite programming. SIAM J. Optim. **15**, 303–318 (2004). https://doi.org/10.1137/S1052623402417298

18. Fukuda, E.H., Haeser, G., Mito, L.M.: Second-order analysis for semidefinite and second-order cone programming via sequential optimality conditions, tech. report, https://optimization-online.org/2020/08/7951/ (2021)

19. Gfrerer, H.: First order and second order characterizations of metric subregularity and calmness of constraint set mappings. SIAM J. Optim. **21**, 1439–1474 (2011). https://doi.org/10.1137/100813415

20. Ioffe, A.D.: Metric regularity - A survey. Part I. Theory. J. Aust. Math. Soc. **101**, 188–243 (2016). https://doi.org/10.1017/S1446788715000701

21. Ioffe, A.D.: Metric regularity - A survey. Part II. Applications. J. Aust. Math. Soc. **101**, 376–417 (2016). https://doi.org/10.1017/S1446788715000695

22. Janin, R.: Directional derivative of the marginal function in nonlinear programming. Math. Program. Stud. **21**, 127–138 (1984). https://doi.org/10.1007/BFb0121214

23. Mehlitz, P.: Asymptotic stationarity and regularity for nonsmooth optimization problems. J. Nonsmooth Anal. Optim. https://doi.org/10.46298/jnsao-2020-6575 (2020)

24. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**, 429–440 (2011). https://doi.org/10.1080/02331930902971377

25. Minchenko, L., Stakhovski, S.: Parametric nonlinear programming problems under the relaxed constant rank condition. SIAM J. Optim. **21**, 314–332 (2011). https://doi.org/10.1137/090761318

26. Moreau, J.J.: Décomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires. C. R. Acad. Sci. Paris **255**, 238–240 (1962)

27. Qi, L., Wei, Z.: On the constant positive linear dependence conditions and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000). https://doi.org/10.1137/S1052623497326629

28. Robinson, S.M.: First-order conditions for general nonlinear optimization. SIAM J. Appl. Math. **30**, 597–610 (1976). https://doi.org/10.1137/0130053

29. Rockafellar, R.T., Wets, R.: Variational Analysis, Grundlehren der mathematischen Wissenschaften, vol. 317. Springer-Verlag, Berlin (2009)

30. Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. Math. Program. Ser. B **77**, 301–320 (1997). https://doi.org/10.1007/BF02614439

31. Shapiro, A., Fan, M.K.H.: On eigenvalue optimization. SIAM J. Optim. **5**, 552–569 (1995). https://doi.org/10.1137/0805028

32. Yamashita, H., Yabe, H., Harada, K.: A primal-dual interior point method for nonlinear semidefinite programming. Math. Program. Ser. A **135**, 89–121 (2012). https://doi.org/10.1007/s10107-011-0449-z

33. Zhang, Y., Zhang, L.: New constraint qualifications and optimality conditions for second order cone programs. Set-Valued Variational Anal. **27**, 693–712 (2019). https://doi.org/10.1007/s11228-018-0487-2