FULL LENGTH PAPER

Weak notions of nondegeneracy in nonlinear semidefinite programming

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

Abstract

The constraint nondegeneracy condition is one of the most relevant and useful constraint qualifications in nonlinear semidefinite programming. It can be characterized in terms of any fixed orthonormal basis of the, let us say, ℓ-dimensional kernel of the constraint matrix, by the linear independence of a set of $\ell(\ell+1)/2$ derivative vectors. We show that this linear independence requirement can be equivalently formulated in a smaller set, of ℓ derivative vectors, by considering all orthonormal bases of the kernel instead. This allows us to identify that not all bases are relevant for a constraint qualification to be defined, giving rise to a strictly weaker variant of nondegeneracy related to the global convergence of an external penalty method. We use some of these ideas to revisit an approach of Forsgren (Math Program 88, 105–128, 2000) for exploiting the sparsity structure of a transformation of the constraints to define a constraint qualification, which led us to develop another relaxed notion of nondegeneracy using a simpler transformation. If the zeros of the derivatives of the constraint

✉ Leonardo M. Mito
  leonardommito@gmail.com

  Roberto Andreani
  andreani@unicamp.br

  Gabriel Haeser
  ghaeser@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   Department of Mathematical Engineering and Center for Mathematical Modeling (CNRS IRL 2807), University of Chile, Santiago, Chile

function at a given point are considered, instead of the zeros of the function itself in a neighborhood of that point, we obtain an even weaker constraint qualification that connects Forsgren's condition and ours.

Keywords Semidefinite programming · Constraint qualifications · Constraint nondegeneracy

Mathematics Subject Classification 90C46 · 90C30 · 90C26 · 90C22

1 Introduction

The study of *linear* and *nonlinear semidefinite programming* (for short, SDP and NSDP, respectively) problems has been consistently growing over the last decades. There are several models for real world problems that can be reformulated as SDPs or NSDPs (we refer to the handbooks [9, Part 4] and [30, Part 3] for a vast collection of applications), which motivate and are motivated by the development of theoretical results regarding optimality conditions and *constraint qualifications* (CQs) for (N)SDPs. Loosely speaking, CQs are assumptions over the feasible set of an optimization problem that ensure that it can be locally described in terms of its first-order approximation. This leads to the possibility of characterizing all solutions of an (N)SDP problem in terms of the derivatives of the functions that describe it, which gives CQs a pivotal role in building convergence theories for practical algorithms. The standard way to do this is to prove that every feasible limit point of the output sequence of the algorithm satisfies the Karush-Kuhn-Tucker (KKT) conditions under a given CQ. Thus, employing a weaker CQ leads to a more robust convergence theory.

One of the most relevant CQs in the literature of (N)SDP is the so-called *nondegeneracy* (or *transversality) condition*, introduced by Shapiro and Fan in [27, Sec. 2] in the context of eigenvalue optimization, and later reformulated by Shapiro [25, Def. 4] for general NSDPs. This condition has been widely used for characterizing sensitivity results (see, for instance, [14, 17, 19–21, 28]), and also for proving global convergence and the rate of convergence of numerical algorithms (we refer to Yamashita and Yabe [31, Secs. 3, 4, and 5] for a survey on this topic). However, it is known that even in the linear case, the solutions of large scale SDP problems tend to be degenerate, even though nondegeneracy is expected to hold in a generic sense. Besides, when the constraint of an NSDP problem has some sparsity structure near one of its solutions—for instance, a diagonal structure—then nondegeneracy is not satisfied at that solution [25]. This means that the convergence theory of an algorithm supported by nondegeneracy does not cover such points.

The explanation for such kind of issue, in our opinion, is the degree of generality of the nondegeneracy condition. That is, although it was born in NSDP, nondegeneracy does not capture any particularity of the constraints, being straigthforwardly extended for any general conic optimization problem, as long as the cone is closed and convex. However, embedding specific traits of matrix-valued functions into nondegeneracy may be more or less direct, depending on how it is characterized. For example, it is well known that (block-)diagonal problems can be remodelled as multiple potentially

dense constraints, such that the nondegeneracy condition, when applied to this remodelled problem, may hold. But what about other types of sparsity? While this question has once been addressed by Forsgren [13], his approach is somewhat intricate and it was not the main topic of his paper, leaving room for a more dedicated analysis. In this paper, instead of defining nondegeneracy as the transversality of two particular subspaces—which is the most usual definition—we exploit an equivalent characterization by Shapiro [25, Prop. 6], which is phrased in terms of the gradients of the entries of an isolated "active block" of the constraints. One particularly interesting detail about this characterization is that it treats all representations of such an "active block" equally, but we show that some of them are more meaningful than others.

The contributions of this paper revolve around the following results:

- We provide a new characterization of nondegeneracy that induces a weaker variant of it, here called *weak-nondegeneracy*, which uses information of the eigenvectors of the constraints evaluated at nearby points;
- We incorporate a sparsity treatment in [25, Prop. 6], which leads to another weak variant of nondegeneracy, called *sparse-nondegeneracy*.
- We connect sparse-nondegeneracy with Forsgren's CQ by means of replacing, in both conditions, the strucutral zeros of the constraint function in a neighborhood of a point, with the zeros of the gradients of its entries at such point. This new condition happens to be a constraint qualification also, which we call *gradient sparse-nondegeneracy*.

These conditions are designed with the sole goal of assisting in proving global convergence of algorithms by means of sequential optimality conditions [3, 8]; however, we envision that they may be further employed in sensitivity analysis, second-order analysis, among other applications. All variants of nondegeneracy we present are proved to be constraint qualifications strictly weaker than nondegeneracy. We also show that when our conditions are applied to diagonal matrices, they are reduced to the *linear independence constraint qualification* (LICQ) from nonlinear programming (NLP). More generally, the conditions are invariant to block representations of (N)SDP problems as a single semidefinite block diagonal matrix or as multiple semidefinite constraints. Then, we compare our definitions with other CQs from the literature.

This paper is structured as follows: In Sect. 2, we introduce our notation; in Sect. 3 we recall the nondegeneracy condition and we prove a new characterization of it, which is where the definition of weak-nondegeneracy comes from. In Sect. 4, we present our definition of sparse-nondegeneracy and a relaxation of it with distinct properties. Finally, in Sect. 5, we discuss some possibilities of prospective work.

2 Preliminaries

Let $f \colon \mathbb{R}^n \to \mathbb{R}$ and $G \colon \mathbb{R}^n \to \mathbb{S}^m$ be continuously differentiable functions, where $\mathbb{S}^m$ is the linear space of all $m \times m$ symmetric matrices, and let $\mathbb{S}^m_+$ be the closed convex pointed cone of all $m \times m$ positive semidefinite matrices. The problem of interest in

this paper is the following:

$$\text{Minimize}_{x \in \mathbb{R}^n} \quad f(x),$$

$$\text{subject to} \quad G(x) \succeq 0, \tag{NSDP}$$

where $\succeq$ is the partial order induced by $\mathbb{S}_+^m$, characterized by the relation: $M \succeq N$ $\Leftrightarrow M - N \in \mathbb{S}_+^m$, for all $M, N \in \mathbb{S}^m$. It is worth pointing out that all results in this paper can be straightforwardly extended to NSDP problems with separate equality constraints, but we omit them for simplicity. The feasible set of (NSDP) will be denoted by $\mathcal{F} \doteq G^{-1}(\mathbb{S}_+^m)$. It is well-known that $\mathbb{S}^m$ is an Euclidean space when equipped with the (*Frobenius*) inner product $\langle M, N \rangle \doteq \text{trace}(MN) \doteq \sum_{i,j=1}^m M_{ij} N_{ij}$.

The derivative of G at a point $x \in \mathbb{R}^n$ is the linear mapping $DG(x) \colon \mathbb{R}^n \to \mathbb{S}^m$ that can be described (in the canonical basis of $\mathbb{R}^n$) by the action

$$d \mapsto DG(x)[d] \doteq \sum_{i=1}^n D_{x_i} G(x) d_i$$

for all $d = (d_1, \ldots, d_n) \in \mathbb{R}^n$, where $D_{x_i} G(x) \in \mathbb{S}^m$ is the partial derivative of G with respect to the variable x_i at $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Also, for each fixed x, the *adjoint* of $DG(x)$ is the unique linear mapping $DG(x)^* \colon \mathbb{S}^m \to \mathbb{R}^n$ that satisfies $\langle DG(x)[d], M \rangle = \langle d, DG(x)^*[M] \rangle$, for all $(d, M) \in \mathbb{R}^n \times \mathbb{S}^m$. Hence,

$$DG(x)^*[M] = \begin{bmatrix} \langle D_{x_1} G(x), M \rangle \\ \vdots \\ \langle D_{x_n} G(x), M \rangle \end{bmatrix} = \sum_{i,j=1}^m M_{ij} \nabla G_{ij}(x)$$

for all $M \in \mathbb{S}^m$, where $\nabla G_{ij}(x)$ denotes the *gradient* of the (i, j)-th entry of G as a function of x. Similarly, we shall denote the gradient of any real-valued function $F \colon \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ by $\nabla F(x)$.

For any given $M \in \mathbb{S}^m$, we consider its spectral decomposition in the form

$$M = \sum_{i=1}^m \lambda_i(M) u_i(M) u_i(M)^\top,$$

where $\lambda_i(M) \in \mathbb{R}$ denotes the i-th eigenvalue of M arranged in non-increasing order (that is, $\lambda_1(M) \geqslant \lambda_2(M) \geqslant \ldots \geqslant \lambda_m(M)$), and $u_i(M) \in \mathbb{R}^m$ corresponds to any associated eigenvector such that $\{u_i(M) \colon i \in \{1, \ldots, m\}\}$ is an orthonormal basis of $\mathbb{R}^m$ (that is, $u_i(M)^T u_i(M) = 1$ and $u_i(M)^T u_j(M) = 0$ when $i \neq j$, for all $i, j \in \{1, \ldots, m\}$).

A useful fact for our analyses is that the orthogonal projection of M onto $\mathbb{S}_+^m$ with respect to the induced (Frobenius) norm, denoted by $\Pi_{\mathbb{S}_+^m}(M)$, can be characterized in terms of its spectral decomposition as follows:

$$\Pi_{\mathbb{S}_+^m}(M) = \sum_{i=1}^m [\lambda_i(M)]_+ u_i(M) u_i(M)^\top,$$

where $[\lambda]_+ \doteq \max\{0, \lambda\}$ for all $\lambda \in \mathbb{R}$.

Given any $\overline{x} \in \mathcal{F}$ and any orthogonal matrix $\overline{U} \in \mathbb{R}^{m \times m}$ whose columns are eigenvectors of $G(\overline{x})$, we partition $\overline{U} = [\overline{P}, \overline{E}]$ such that the columns of $\overline{P} \in \mathbb{R}^{m \times r}$ correspond to the eigenvectors associated with the positive eigenvalues of $G(\overline{x})$ and the columns of $\overline{E} \in \mathbb{R}^{m \times m-r}$ correspond to the eigenvectors associated with the null eigenvalues of $G(\overline{x})$, where $r = \text{rank}(G(\overline{x}))$. To abbreviate, as an abuse of notation and language, we will say that $\overline{E}$ *spans* $\text{Ker}\,G(\overline{x})$ in this context. That is, $\overline{E}$ spans $\text{Ker}\,G(\overline{x})$ if, and only if, $\overline{E}^\top \overline{E} = \mathbb{I}_{m-r}$ and $G(\overline{x})\overline{E} = 0$, where $\mathbb{I}_{m-r}$ denotes an $(m-r)$-dimensional identity matrix.

There are multiple ways of describing optimality in NSDP problems, but in this paper we direct our attention to necessary optimality conditions that are based on the classical *Karush-Kuhn-Tucker* (KKT) conditions:

Definition 1 (*KKT*) We say that a point $\overline{x} \in \mathcal{F}$ satisfies the *KKT conditions* when there exists some $\overline{Y} \succeq 0$ such that

$$
\begin{aligned}
\nabla_x L(\overline{x}, \overline{Y}) &\doteq \nabla f(\overline{x}) - DG(\overline{x})^*[\overline{Y}] = 0, \\
\langle G(\overline{x}), \overline{Y} \rangle &= 0,
\end{aligned}
\tag{KKT}
$$

where $L : \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$ is the *Lagrangian function* of (NSDP), given by

$$
L(x, Y) \doteq f(x) - \langle G(x), Y \rangle.
$$

As usual, the matrix $\overline{Y}$ is called a *Lagrange multiplier* associated with $\overline{x}$ and we denote the set of all Lagrange multipliers associated with $\overline{x}$ by $\Lambda(\overline{x})$. When $\Lambda(\overline{x}) \neq \emptyset$, $\overline{x}$ is called a *KKT point* of (NSDP). Let r be the rank of $G(\overline{x})$ and let $\overline{E} \in \mathbb{R}^{m \times m-r}$ be a matrix that spans $\text{Ker}\,G(\overline{x})$; then, for any $\overline{Y} \in \Lambda(\overline{x})$, since both $\overline{Y}$ and $G(\overline{x})$ are positive semidefinite, the complementarity relation $\langle G(\overline{x}), \overline{Y} \rangle = 0$ is equivalent to $G(\overline{x})\overline{Y} = 0$, which is in turn equivalent to saying that $\text{Im}\,\overline{Y} \subseteq (\text{Im}\,G(\overline{x}))^\perp = \text{Ker}\,G(\overline{x})$, where $(\text{Im}\,G(\overline{x}))^\perp$ denotes the orthogonal complement of $\text{Im}\,G(\overline{x})$. Therefore, $\overline{Y}$ is complementary to $G(\overline{x})$ if, and only if, it has the form

$$
\overline{Y} = \overline{E} \tilde{Y} \overline{E}^\top,
\tag{1}
$$

where $\tilde{Y} \in \mathbb{S}^{m-r}_+$ is not necessarily a diagonal matrix. Moreover, note that $\tilde{Y}$ is not necessarily positive definite; that is, $\dim(\text{Ker}\,\overline{Y})$ does not necessarily coincide with r. When they do coincide, $\overline{x}$ and $\overline{Y}$ are said to be *strictly complementary* [25].

It is known that the KKT conditions are not necessary for local optimality unless they are paired with a constraint qualification. For instance, one of the most studied constraint qualifications for (NSDP) is *Robinson's CQ* [24, Def. 3], which holds at a point $\overline{x} \in \mathcal{F}$ if there exists $d \in \mathbb{R}^n$ such that

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

where $\text{int}\,\mathbb{S}^m_+$ denotes the topological interior of $\mathbb{S}^m_+$, which in turn coincides with the set of $m \times m$ symmetric positive definite matrices. Alternatively, following Bonnans

and Shapiro [11, Prop. 2.97], it is possible to say that (the dual form of) Robinson's CQ holds at $\overline{x} \in \mathcal{F}$ if, and only if,

$$
\left.\begin{array}{r}
DG(\overline{x})^*[Y] = 0 \\
\langle G(\overline{x}), Y \rangle = 0 \\
Y \succeq 0
\end{array}\right\} \Rightarrow Y = 0.
\tag{2}
$$

Another well-known fact is that, for every local minimizer $\overline{x} \in \mathcal{F}$, the set $\Lambda(\overline{x})$ is nonempty and compact if, and only if, Robinson's CQ holds at $\overline{x}$ (see [11, Props. 3.9 and 3.17] for details). This makes Robinson's CQ the natural analogue of the *Mangasarian-Fromovitz CQ* (MFCQ), from NLP, in NSDP.

3 The nondegeneracy condition for NSDP

In this section, we discuss the well-known nondegeneracy condition introduced by Shapiro and Fan [27, Sec. 2]. We derive a different characterization for it that suggests a way of obtaining a weaker constraint qualification with potentially interesting properties. But firstly, we briefly recall some elements of convex analysis.

The (*Bouligand*) tangent cone to a set C at a point $y \in C$ is defined as

$$
T_C(y) \doteq \left\{ d \colon \begin{array}{l} \exists \{d^k\}_{k \in \mathbb{N}} \to d, \ \exists \{t^k\}_{k \in \mathbb{N}} \to 0, \ t^k > 0, \\ \forall k \in \mathbb{N}, \ y + t^k d^k \in C \end{array} \right\}.
$$

In particular, when $C = \mathbb{S}^m_+$, at a given $M \in \mathbb{S}^m_+$, it can be characterized as follows

$$
T_{\mathbb{S}^m_+}(M) = \left\{ N \in \mathbb{S}^m : d^\top N d \geqslant 0, \ \forall d \in \operatorname{Ker} M \right\}.
$$

Therefore, for every feasible $\overline{x}$ we have

$$
T_{\mathbb{S}^m_+}(G(\overline{x})) = \left\{ N \in \mathbb{S}^m : \overline{E}^\top N \overline{E} \succeq 0 \right\},
\tag{3}
$$

whenever $\overline{E}$ spans $\operatorname{Ker} G(\overline{x})$.

It is clear from (3) that the largest subspace contained in $T_{\mathbb{S}^m_+}(G(\overline{x}))$, that is, its *lineality space*, can be characterized as follows:

$$
\operatorname{lin}(T_{\mathbb{S}^m_+}(G(\overline{x}))) = \left\{ N \in \mathbb{S}^m : \overline{E}^\top N \overline{E} = 0 \right\}.
\tag{4}
$$

The nondegeneracy condition of Shapiro and Fan is verified at $\overline{x}$ when the linear subspaces $\operatorname{Im} DG(\overline{x})$ and $\operatorname{lin}(T_{\mathbb{S}^m_+}(G(\overline{x})))$ of $\mathbb{S}^m$ meet transversally, which is why it was originally called *transversality* in [27]. In mathematical language:

Definition 2 (Def. 4 from [25]) A point $\overline{x} \in \mathcal{F}$ is said to satisfy the *nondegeneracy* condition when the following relation is satisfied:

$$\text{Im } DG(\overline{x}) + \text{lin}(T_{\mathbb{S}^m_+}(G(\overline{x}))) = \mathbb{S}^m. \tag{5}$$

If $\overline{x}$ is a local solution of (NSDP), then nondegeneracy implies that $\Lambda(\overline{x})$ is a singleton; and the converse is also true in the presence of strict complementarity (see [26, Thm. 2.2 and Sect. 3]). Hence, Definition 2 is generally seen as an analogue of LICQ, from NLP, in NSDP. However, this analogy is tied to how the link between NLP and NSDP is made [25]. For example, when an NLP problem with constraints $g_1(x) \geqslant 0, \ldots, g_m(x) \geqslant 0$ is modelled as an NSDP with a single *structurally diagonal* conic constraint; that is, with G in the form

$$G(x) \doteq \begin{bmatrix} g_1(x) & & \\ & \ddots & \\ & & g_m(x) \end{bmatrix} \succeq 0; \tag{6}$$

then Definition 2 fails whenever there is some $\overline{Y} \in \Lambda(\overline{x})$ and some nonzero $H \in \mathbb{S}^m$ with only zeros in its diagonal, such that $H \succeq -\overline{Y}$, regardless of the linear independence of the set $\{\nabla g_1(\overline{x}), \ldots, \nabla g_m(\overline{x})\}$. In fact, structurally diagonal NSDP problems are in general expected to lack uniqueness of the Lagrange multiplier.

On the other hand, it is well-known (cf. [11, Sect. 4.6.1]) that a feasible point $\overline{x}$ satisfies the nondegeneracy condition if, and only if, either $\text{Ker } G(\overline{x}) = \{0\}$ or the linear mapping $\psi_{\overline{x}} : \mathbb{R}^n \to \mathbb{S}^{m-r}$, defined by

$$\psi_{\overline{x}}(d) \doteq \overline{E}^\top DG(\overline{x})[d]\overline{E}, \tag{7}$$

is surjective for any $\overline{E}$ that spans $\text{Ker } G(\overline{x})$. As a direct consequence of the equivalence above, it is possible to characterize Definition 2 as follows:

Proposition 1 (Prop. 6 from [25]) *Let $\overline{x} \in \mathcal{F}$ and let r denote the rank of $G(\overline{x})$. Then, $\overline{x}$ satisfies the nondegeneracy condition if, and only if, either $\text{Ker } G(\overline{x}) = \{0\}$ or the vectors*

$$\begin{aligned} v_{ij}(\overline{x}, \overline{E}) &\doteq \left[\overline{e}_i^\top D_{x_1} G(\overline{x}) \overline{e}_j, \ldots, \overline{e}_i^\top D_{x_n} G(\overline{x}) \overline{e}_j \right]^\top \\ &= DG(\overline{x})^* \left[\frac{\overline{e}_i \overline{e}_j^\top + \overline{e}_j \overline{e}_i^\top}{2} \right], \end{aligned} \quad 1 \leq i \leq j \leq m - r \tag{8}$$

are linearly independent, where $\overline{E} \in \mathbb{R}^{m \times m-r}$ is an arbitrary fixed matrix that spans $\text{Ker } G(\overline{x})$, and $\overline{e}_i$ denotes the i-th column of $\overline{E}$, for all $i \in \{1, \ldots, m - r\}$.

Now, inspired by Proposition 1, we present a similar characterization of nondegeneracy that evaluates the linear independence of a narrower set of vectors at the cost of looking at all possible choices of $\overline{E}$ instead of a fixed one. Since our reasoning can be extended to Robinson's CQ, we also characterize it in a similar manner.

Proposition 2 *Let $\overline{x} \in \mathcal{F}$ and $r = \mathrm{rank}(G(\overline{x}))$. Then, $\overline{x}$ satisfies:*

1. *Nondegeneracy if, and only if, either $r = m$ or the set*

$$\left\{ v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m - r\} \right\} \tag{9}$$

   is linearly independent for every matrix $\overline{E} \in \mathbb{R}^{m \times m - r}$ that spans $\mathrm{Ker}\, G(\overline{x})$.
2. *Robinson's CQ if, and only if, either $r = m$ or (9) is positive linearly independent for every matrix $\overline{E} \in \mathbb{R}^{m \times m - r}$ that spans $\mathrm{Ker}\, G(\overline{x})$.*

Proof Let us assume that $r < m$ since the result follows trivially otherwise. Then, for any fixed $\overline{E} \in \mathbb{R}^{m \times m - r}$ such that $G(\overline{x})\overline{E} = 0$ and $\overline{E}^{\top}\overline{E} = \mathbb{I}_{m-r}$, note that (9) is (positive) linearly independent if, and only if, the following holds: if the scalars $\alpha_1, \ldots, \alpha_{m-r} \in \mathbb{R}$ (with $\alpha_1 \geqslant 0, \ldots, \alpha_{m-r} \geqslant 0$, respectively) satisfy

$$\sum_{i=1}^{m-r} \alpha_i DG(\overline{x})^*[\overline{e}_i \overline{e}_i^{\top}] = \sum_{i=1}^{m-r} \alpha_i v_{ii}(\overline{x}, \overline{E}) = 0, \tag{10}$$

then one must have $\alpha_1 = \ldots = \alpha_{m-r} = 0$. That is, (9) is (positive) linearly independent if, and only if, for every matrix Y of the form

$$Y \doteq \sum_{i=1}^{m-r} \alpha_i \overline{e}_i \overline{e}_i^{\top} = \overline{E} \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_{m-r} \end{bmatrix} \overline{E}^{\top} \tag{11}$$

where $\alpha_1, \ldots, \alpha_{m-r} \in \mathbb{R}$ (with $\alpha_1 \geqslant 0, \ldots, \alpha_{m-r} \geqslant 0$, respectively), we have that

$$DG(\overline{x})^*[Y] = 0 \implies Y = 0. \tag{12}$$

With this in mind, we recall that:

- For any fixed choice of $\overline{E}$ spanning $\mathrm{Ker}\, G(\overline{x})$, nondegeneracy holds at $\overline{x}$ if, and only if, (12) holds for every Y in the form $Y = \overline{E} Z \overline{E}^{\top}$ with $Z \in \mathbb{S}^{m-r}$ (Proposition 1);
- Robinson's CQ holds at $\overline{x}$ if, and only if, (12) holds for every $Y \succeq 0$ such that $\langle G(\overline{x}), Y \rangle = 0$ (see (2));

and it becomes clear that nondegeneracy (respectively, Robinson's CQ) implies that (9) is (positive) linearly independent, for every $\overline{E}$ as described above, because every Y as in (11) satisfies $\langle G(\overline{x}), Y \rangle = 0$.

To prove the converse of item 1, assume that (9) is linearly independent for all $\overline{E}$ that spans $\mathrm{Ker}\, G(\overline{x})$. Let $Y = \overline{E} Z \overline{E}^{\top}$ be such that $Z \in \mathbb{S}^{m-r}$ and let $C \in \mathbb{R}^{m-r \times m-r}$ be an orthogonal matrix such that $C^{\top} Z C = \mathrm{Diag}(z_1, \ldots, z_{m-r})$, where $\mathrm{Diag}(z_1, \ldots, z_{m-r}) \in \mathbb{S}^{m-r}$ is a diagonal matrix whose i-th diagonal entry is z_i, with $i \in \{1, \ldots, m - r\}$. Then, note that $\overline{E} C^{\top}$ also spans $\mathrm{Ker}\, G(\overline{x})$, which puts

$$Y = \overline{E} C^{\top} \mathrm{Diag}(z_1, \ldots, z_{m-r})(\overline{E} C^{\top})^{\top}$$

in format (11); by our previous assumption (12) holds for Y and we conclude that nondegeneracy holds at $\overline{x}$.

Now, to prove the converse of item 2, assume that (9) is positive linearly independent for all $\overline{E}$ that spans $\operatorname{Ker} G(\overline{x})$, and let Y be such that $DG(\overline{x})^*[Y] = 0$, $\langle G(\overline{x}), Y \rangle = 0$ and $Y \succeq 0$. It is elementary to see that there exists some matrix $\overline{E}$ that spans $\operatorname{Ker} G(\overline{x})$, such that Y has the form (11). It follows from our hypothesis that $Y = 0$ and because Y is arbitrary, Robinson's CQ holds at $\overline{x}$. □

The characterizations of nondegeneracy and Robinson's CQ from Proposition 2 may seem less practical than the one from Proposition 1, but it reveals a clear path for defining weaker CQs by ruling out some particular choices of $\overline{E}$, which is the main result of the next subsection.

We recall that Wachsmuth [29] proved for NLPs that LICQ is equivalent to the uniqueness of the Lagrange multiplier for any objective function f (the unique multiplier may vary with f). Thanks to Proposition 2 this characterization can be straightforwardly extended to NSDP replacing LICQ by nondegeneracy, which we omit.

3.1 Sequences of eigenvectors and weak-nondegeneracy

In [8], Andreani et al. introduce a constructive technique for proving the existence of Lagrange multipliers for (NSDP), which is based on the so-called *sequential optimality conditions* from NLP [3]. The core idea of their proof is to apply an external penalty algorithm to (NSDP) after regularizing it around a given local minimizer, to obtain a sequence of approximate KKT points converging to it, as follows:

Theorem 1 (Thm. 3.2 from [8]) *Let $\overline{x}$ be a local minimizer of (NSDP). Then, for any sequence $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there exists some $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, such that for every $k \in \mathbb{N}$, x^k is a local minimizer of the regularized penalty function*

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

In particular, computing derivatives we obtain $\nabla_x L(x^k, Y^k) \to 0$, where $Y^k \doteq \rho_k \Pi_{\mathbb{S}_+^m}(-G(x^k))$.

With this result at hand, the authors prove that the sequence $\{Y^k\}_{k\in\mathbb{N}}$ must be bounded in the presence of Robinson's CQ, and that all of its limit points are Lagrange multipliers associated with $\overline{x}$ [8, Thm. 6.1]. Furthermore, the proof of this fact under nondegeneracy follows easily by contradiction: suppose that $\{Y^k\}_{k\in\mathbb{N}}$ is unbounded, and take any limit point $\overline{Y}$ of the sequence $\{Y^k/\|Y^k\|\}_{k\in\mathbb{N}}$; then:

1. It follows from $\nabla_x L(x^k, Y^k) \to 0$ that $DG(\overline{x})^*[\overline{Y}] = 0$, which means $\overline{Y} \in \operatorname{Ker} DG(\overline{x})^* = \operatorname{Im} DG(\overline{x})^\perp$;
2. By the definition of Y^k, we have $0 \neq \overline{Y} \succeq 0$ and $\langle G(\overline{x}), \overline{Y} \rangle = 0$, so $\overline{Y} \in \operatorname{lin}(T_{\mathbb{S}_+^m}(G(\overline{x})))^\perp$;

Hence, $\overline{Y} \in \text{Im } DG(\overline{x})^{\perp} \cap \text{lin}(T_{\mathbb{S}_+^m}(G(\overline{x})))^{\perp}$, which contradicts nondegeneracy.

With a single extra step, which is to take a spectral decomposition of Y^k for each k, the reasoning of the previous paragraph can be put in the same terms as Proposition 2. Indeed, observe that $\lambda_i(Y^k) = [\rho_k \lambda_i(-G(x^k))]_+ = 0$ for all $i \in \{m-r+1, \ldots, m\}$ and all k large enough, because

$$\lambda_i(-G(x^k)) = -\lambda_{m-i+1}(G(x^k)).$$

So

$$\nabla_x L(x^k, Y^k) = \nabla f(x^k) - \sum_{i=1}^{m-r} [\rho_k \lambda_i(-G(x^k))]_+ v_{ii}(x^k, E^k) \to 0,$$

where $E^k \in \mathbb{R}^{m \times m-r}$ is a matrix whose i-th column is $u_{m-i+1}(G(x^k))$. Then, note that if E^k can be chosen such that at least one of its limit points $\overline{E}$ ensures linear independence of $\{v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m-r\}\}$, then $\{Y^k\}_{k \in \mathbb{N}}$ must be bounded. Although the first clause of the previous sentence resembles nondegeneracy (as in Proposition 2), note that asking for the linear independence of the set $\{v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m-r\}\}$ when $\overline{E}$ is not a limit point of some sequence $\{E^k\}_{k \in \mathbb{N}}$ of eigenvectors of $G(x^k)$ seems unnecessary for defining a constraint qualification. This motivates us to propose a weaker variant of nondegeneracy in a way that can also be extended to Robinson's CQ, which goes as follows:

Definition 3 *(Weak-nondegeneracy and weak-Robinson's CQ)* Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. We say that *weak-nondegeneracy* (respectively, *weak-Robinson's CQ*) holds at $\overline{x}$ if either $\text{Ker } G(\overline{x}) = \{0\}$ or: for every sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$, there exists some sequence of matrices with orthonormal columns $\{E^k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^{m \times m-r}$ such that:

1. The columns of E^k are eigenvectors associated with the $m-r$ smallest eigenvalues of $G(x^k)$, for each $k \in \mathbb{N}$;
2. There exists a limit point $\overline{E}$ of $\{E^k\}_{k \in \mathbb{N}}$ such that the set

$$\{v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m-r\}\},$$

as defined in (8), is (positive) linearly independent.

There are a couple of nuances about Definition 3 that should be properly addressed (see also the discussion after Remark 2). First, we recall that the eigenvector functions $u_i(G(x)), i \in \{m-r+1, \ldots, m\}$ are not necessarily continuous at a given point $\overline{x}$; so weak-nondegeneracy (and weak-Robinson's CQ) relies on the "sequential continuity of eigenvectors" along a given path. Second, for any fixed $\overline{x} \in \mathcal{F}$ and any $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$, the sequence $\{E^k\}_{k \in \mathbb{N}}$ described in Definition 3 is well-defined for k sufficiently large, since the r largest eigenvalues of $G(x^k)$ are necessarily bounded away from zero.

Remark 1 Based on the previous discussion, it is worth mentioning that weak-nondegeneracy (and weak-Robinson's CQ) can be equivalently stated in terms of

a certain notion of continuity of the eigenvectors of $G(x)$. To see why, consider a feasible point $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. Because $r < m$, it follows that $\lambda_r(G(x)) > \lambda_{r+1}(G(x))$ for every x close enough to $\overline{x}$, so the following set is well-defined:

$$\mathcal{B}(x) \doteq \left\{ E \in \mathbb{R}^{m \times (m-r)} \colon \begin{array}{l} G(x)e_i = \lambda_{m-i+1}(G(x))e_i, \ \forall i \in \{1, \ldots, m-r\} \\ E^\top E = \mathbb{I}_{m-r} \end{array} \right\} \tag{13}$$

where $E \doteq [e_1, \ldots, e_{m-r}]$. The set above consists of all matrices whose columns are orthonormal eigenvectors associated with the $m - r$ smallest eigenvalues of $G(x)$. Moreover, for any sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ recall the Painlevé-Kuratowski *upper limit* [11, Def. 2.52]) of the sequence of images $\{\mathcal{B}(x^k)\}_{k \in \mathbb{N}}$, defined as

$$\limsup_{k \to \infty} \mathcal{B}(x^k) \doteq \left\{ z \colon \exists I \subseteq \mathbb{N} \text{ infinite}, \ \exists \{z^k\}_{k \in I} \to z, \ \forall k \in I, \ z^k \in \mathcal{B}(x^k) \right\}.$$

In these terms, it is easy to see that weak-nondegeneracy (respectively, weak-Robinson's CQ) holds at $\overline{x}$ if, and only if, either $\mathrm{Ker}\, G(\overline{x}) = \{0\}$ or, for every sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \limsup_{k \to \infty} \mathcal{B}(x^k)$ such that

$$\left\{ v_{ii}(\overline{x}, \overline{E}) \colon i \in \{1, \ldots, m-r\} \right\}$$

is (positive) linearly independent.

Although the characterization of Remark 1 may shorten notation, in order to check whether weak-nondegeneacry holds or not at a given point $\overline{x}$ requires the computation of the set $\mathcal{B}(\overline{x})$, which may be complicated in practice. Therefore, it is important to emphasize that $\mathcal{B}(\overline{x})$ is not meant to be explicitly computed because weak-nondegeneracy is not meant to be manually checked at any point, except for very specific cases with a convenient eigenvector structure (see Examples 1 and 2); instead, the main purpose of weak-nondegeneracy (and weak-Robinson's CQ) is to serve as a theoretical tool for building the convergence theory of iterative algorithms, as it was presented in the proof of Theorem 2 for the external penalty method. In this context, knowledge of the problem solution is usually limited to an approximation obtained by truncating the method's output sequence, which ends up taking away some of the meaning of checking constraint qualifications in practice.

The discussion that motivated Definition 3 already suggests that it indeed describes a genuine constraint qualification, and it also provides an outline of how to prove it. Nevertheless, we state and prove this fact with appropriate mathematical rigor below. Although we prove the next result for weak-Robinson's CQ, observe that the analogous statement for weak-nondegeneracy follows as a corollary.

Theorem 2 *Every local minimizer $\overline{x} \in \mathcal{F}$ of (NSDP) that satisfies weak-Robinson's CQ also satisfies the KKT conditions. By extension, the same holds for weak-nondegeneracy.*

Proof Let $\overline{x}$ be a local minimizer of (NSDP) that satisfies weak-Robinson's CQ and let $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ and $\{Y^k\}_{k \in \mathbb{N}}$ be the sequences described in Theorem 1, for an arbitrary

sequence $\{\rho_k\}_{k \in \mathbb{N}} \to \infty$. If $r = m$, set $\overline{Y} = 0$ as a Lagrange multiplier associated with $\overline{x}$ and we are done; so let us assume that $r < m$ from now on. From the local optimality of x^k, for each $k \in \mathbb{N}$, we obtain

$$\nabla f(x^k) + (x^k - \overline{x}) + DG(x^k)^*[Y^k] = 0. \tag{14}$$

Recall that we assume, without loss of generality, that $\lambda_1(-G(x^k)) \geqslant \ldots \geqslant \lambda_m(-G(x^k))$, for every k; and note that when k is large enough, say greater than some $k_0 \in \mathbb{N}$, we necessarily have $\lambda_i(-G(x^k)) < 0$ for all $i \in \{m - r + 1, \ldots, m\}$ since $G(x^k) \to G(\overline{x})$ and eigenvalues $\lambda_i(\cdot)$ are continuous mappings. Then, for each $k > k_0$, we have

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

where $\alpha_i^k \doteq [\rho_k \lambda_i(-G(x^k))]_+ \geqslant 0$ and $e_i^k \doteq u_{m-i+1}(G(x^k))$ is an arbitrary unitary eigenvector associated with $\lambda_{m-i+1}(G(x^k))$, for each $i \in \{1, \ldots, m - r\}$. Set $E^k \doteq [e_1^k, \ldots, e_{m-r}^k]$. Since $\{E^k\}_{k \in \mathbb{N}}$ is bounded, we may pick any of its limit points $\overline{E} = [\overline{e}_1, \ldots, \overline{e}_{m-r}]$ and assume, taking a subsequence if necessary, that it converges to $\overline{E}$, which spans $\mathrm{Ker}\, G(\overline{x})$. Then, observe that (14) implies

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

but since $DG(x^k)^*[e_i^k (e_i^k)^\top] = v_{ii}(x^k, E^k)$ (see (8)), we can rewrite it as

$$\nabla f(x^k) - \sum_{i=1}^{m-r} \alpha_i^k v_{ii}(x^k, E^k) \to 0. \tag{15}$$

If $\left\{(\alpha_i^k, \ldots, \alpha_{m-r}^k)\right\}_{k \in \mathbb{N}}$ has any convergent subsequence, denote its limit point by $\overline{\alpha} \doteq (\overline{\alpha}_i, \ldots, \overline{\alpha}_{m-r})$, and note that $\overline{\alpha}$ generates a Lagrange multiplier for $\overline{x}$, which is

$$\overline{Y} \doteq \sum_{i=1}^{m-r} \overline{\alpha}_i \overline{e}_i \overline{e}_i^\top. \tag{16}$$

Hence, it suffices to prove that $\{\alpha_i^k\}_{k \in \mathbb{N}}$, $i \in \{1, \ldots, m - r\}$, must be bounded under weak-Robinson's CQ. Let us assume for a moment that the sequences $\{\alpha_i^k\}_{k \in \mathbb{N}}$ are unbounded, which means

$$m^k \doteq \max \left\{\alpha_i^k : i \in \{1, \ldots, m - r\}\right\} \to \infty.$$

Note that $\left\{(\alpha_1^k, \ldots, \alpha_{m-r}^k)/m^k\right\}_{k \in \mathbb{N}}$ must be bounded and it must also have a nonzero limit point, which we will denote by $(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_{m-r}) \geqslant 0$. We assume without loss of

generality, that $\left\{(\alpha_1^k, \ldots, \alpha_{m-r}^k)/m^k\right\}_{k \in \mathbb{N}} \to (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_{m-r})$. After dividing (15) by m^k for each k and taking limit $k \to +\infty$, we obtain

$$\sum_{i=1}^{m-r} \tilde{\alpha}_i v_{ii}(\overline{x}, \overline{E}) = 0,$$

which means $\left\{v_{ii}(\overline{x}, \overline{E}) \colon i \in \{1, \ldots, m-r\}\right\}$ is positive linearly dependent. However, since our analyses hold for any arbitrary choice of $\{E^k\}_{k \in \mathbb{N}}$ and any $\overline{E}$, this contradicts weak-Robinson's CQ. □

Let us briefly analyse a direct application of weak-Robinson's CQ: As an intermediary step of the proof of Theorem 2, we proved that every feasible limit point of a sequence described in Theorem 1 must satisfy the KKT conditions under weak-Robinson's CQ. However, the sequences $\{x^k\}_{k \in \mathbb{N}}$ and $\{Y^k\}_{k \in \mathbb{N}}$ described in Theorem 2 are precisely the ones that are generated by a standard external penalty method (that is, [8, Algorithm 1] with the parameter $\Omega^{\max}$ fixed at zero). Thus, every feasible limit point of the external penalty method that satisfies weak-Robinson's CQ must also satisfy the KKT conditions. By extension this also holds for weak-nondegeneracy.

Another interesting property of the weak variants of nondegeneracy and Robinson's CQ is that they are equivalent to LICQ and MFCQ, respectively, when G is a structurally diagonal matrix function (as in (6)) that models an NLP problem, which in some sense resolves the inconsistency between nondegeneracy and LICQ noted by Shapiro [25, Page 309].

Remark 2 When G is structurally diagonal, as in (6), then $\overline{x} \in \mathcal{F}$ satisfies weak-nondegeneracy if, and only if, the set $\{\nabla g_i(\overline{x}) \colon g_i(\overline{x}) = 0\}$ is linearly independent. Indeed, if $r = m$ the result follows trivially, so let us assume that $r < m$. Also, suppose that $\{i \in \{1, \ldots, m\} \colon g_i(\overline{x}) = 0\} = \{r+1, \ldots, m\}$, where r is the rank of $G(\overline{x})$. Clearly, if $\{\nabla g_{r+1}(\overline{x}), \ldots, \nabla g_m(\overline{x})\}$ is linearly independent, then we may take

$$E^k \doteq \begin{bmatrix} 0 \\ \mathbb{I}_{m-r} \end{bmatrix} \in \mathbb{R}^{m \times m-r}$$

for all sequences $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ to conclude that $\overline{x}$ satisfies weak-nondegeneracy. Conversely, suppose that weak-nondegeneracy holds at $\overline{x}$, take any sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ and any $\{E^k\}_{k \in \mathbb{N}} \to \overline{E} \doteq [\overline{e}_i, \ldots, \overline{e}_{m-r}]$ such that

$$\left\{v_{ii}(\overline{x}, \overline{E}) \colon i \in \{1, \ldots, m-r\}\right\}$$

is linearly independent. Note that $\overline{E}$ must have the form

$$\overline{E} = \begin{bmatrix} 0 \\ Q \end{bmatrix}, \quad \text{where } Q \in \mathbb{R}^{m-r \times m-r} \text{ is orthonormal,}$$

due to the diagonal structure of G and the fact that $g_i(\overline{x}) \neq 0$ for all $i \in \{1, \ldots, r\}$. Hence,

$$v_{ii}(\overline{x}, \overline{E}) = \sum_{j=r+1}^{m} \nabla g_j(\overline{x}) Q_{i,j-r}^2 = Dg(\overline{x})^\top (Q_i \odot Q_i), \tag{17}$$

where $Dg(\overline{x})$ is the Jacobian matrix of $g(x) \doteq (g_{r+1}(x), \ldots, g_m(x))$ at $\overline{x}$, the operator $\odot$ is the (Hadamard) entry-wise vector product, and Q_i is the i-th column of Q, with $i \in \{1, \ldots, m - r\}$. Then,

$$\text{span}\left\{v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m - r\}\right\} \subseteq \text{Im } Dg(\overline{x})^\top$$

and, consequently,

$$\begin{aligned} m - r &= \dim(\text{span}\left\{v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m - r\}\right\}) \\ &\leqslant \dim(\text{Im } Dg(\overline{x})^\top) \\ &= \text{rank}(Dg(\overline{x})^\top) \leqslant m - r \end{aligned}$$

Hence, $\text{rank}(Dg(\overline{x})^\top) = m - r$, which means that $\{\nabla g_{r+1}(\overline{x}), \ldots, \nabla g_m(\overline{x})\}$ is linearly independent. Using similar arguments, thanks to (17) which states that the vectors $v_{ii}(\overline{x}, \overline{E})$ are nonnegative linear combinations of the columns of $Dg(\overline{x})^\top$, it is possible to prove that $\overline{x} \in \mathcal{F}$ satisfies weak-Robinson's CQ if, and only if, $\{\nabla g_i(\overline{x}) : g_i(\overline{x}) = 0\}$ is positive linearly independent, which is in turn equivalent to Robinson's CQ.

It is clear from Proposition 2 that weak-nondegeneracy is implied by nondegeneracy; and we see in the example below that the converse is not true.

Example 1 Consider the following constraint:

$$G(x) \doteq \begin{bmatrix} x_1 & x_2 \\ x_2 & x_1 \end{bmatrix}$$

at the point $\overline{x} \doteq (0, 0)$, which clearly does not satisfy nondegeneracy. Weak-nondegeneracy, on the other hand, holds at $\overline{x}$ as

$$\mathcal{B}(x) = \begin{cases} \left\{ \frac{\pm 1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}, \frac{\pm 1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \right\}, & \text{if } x_2 \neq 0 \\ \left\{ E \in \mathbb{R}^{2 \times 2} : E^\top E = \mathbb{I}_2 \right\}, & \text{if } x_2 = 0 \end{cases}$$

for every $x \in \mathbb{R}^2$, according to (13), so it suffices to take

$$E^k \doteq \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \in \mathcal{B}(x^k) \quad \text{and} \quad \overline{E} \doteq \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \in \limsup_{k \to \infty} \mathcal{B}(x^k)$$

for all sequences $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ to obtain that $v_{11}(\overline{x}, \overline{E}) = [1, -1]$ and $v_{22}(\overline{x}, \overline{E}) = [1, 1]$ are linearly independent.

This simple example is also important to show that weak-nondegeneracy does not guarantee uniqueness of Lagrange multipliers. For instance, consider the constraint above with the objective function $f(x) \doteq 2x_1$ which has $\bar{x}$ as its global minimizer; then every $\bar{Y}$ in the form

$$\bar{Y} \doteq \begin{bmatrix} 1 - \alpha & 0 \\ 0 & 1 + \alpha \end{bmatrix}$$

for $\alpha \in [-1, 1] \setminus \{0\}$ is a Lagrange multiplier associated with $\bar{x}$.

Another example that serves the same purpose, which can also be used to show how the sparsity structure of the eigenvectors of G is grasped by weak-nondegeneracy is the following:

Example 2 Consider the constraint:

$$G(x) \doteq \begin{bmatrix} x_{11} & 0 & x_{13} \\ 0 & x_{22} & 0 \\ x_{13} & 0 & x_{33} \end{bmatrix} \succeq 0$$

and let $\bar{x} \doteq 0$. Nondegeneracy fails at $\bar{x}$, but weak-nondegeneracy holds. To see this, take any sequence $\{x^k\}_{k \in \mathbb{N}} \to \bar{x}$, if $x_{13}^k \neq 0$ for all k (the other case is trivial, so we will omit it), and

$$E^k \doteq \begin{bmatrix} \dfrac{-\eta_1^k}{\sqrt{(\eta_1^k)^2 + 1}} & 0 & \dfrac{-\eta_2^k}{\sqrt{(\eta_2^k)^2 + 1}} \\ 0 & 1 & 0 \\ \dfrac{1}{\sqrt{(\eta_1^k)^2 + 1}} & 0 & \dfrac{1}{\sqrt{(\eta_2^k)^2 + 1}} \end{bmatrix},$$

where

$$\eta_1^k \doteq \frac{-x_{11}^k + x_{22}^k + \sqrt{(x_{11}^k)^2 - 2x_{22}^k x_{33}^k + (x_{33}^k)^2 + 4(x_{13}^k)^2}}{2x_{13}^k}$$

and

$$\eta_2^k \doteq \frac{-x_{11}^k + x_{22}^k - \sqrt{(x_{11}^k)^2 - 2x_{22}^k x_{33}^k + (x_{33}^k)^2 + 4(x_{13}^k)^2}}{2x_{13}^k}.$$

In this case, assuming that $x_{13}^k > 0$ for all k (which can be done without loss of generality since the other cases are analogous), we have

$$\lim_{k \to \infty} \eta_1^k = \lim_{k \to \infty} \frac{|x_{13}^k|}{x_{13}^k} = 1 \quad \text{and} \quad \lim_{k \to \infty} \eta_2^k = \lim_{k \to \infty} -\frac{|x_{13}^k|}{x_{13}^k} = -1,$$

hence

$$E^k \to \overline{E} \doteq \begin{bmatrix} \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix},$$

and computing the vectors of interest we arrive at

$$v_{11}(\overline{x}, \overline{E}) = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix}, \quad v_{22}(\overline{x}, \overline{E}) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}. \quad v_{33}(\overline{x}, \overline{E}) = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix},$$

which are linearly independent, so weak-nondegeneracy holds at $\overline{x}$. Observe that, in this case, the matrix E^k has the same sparsity structure as G.

Moreover, note that weak-nondegeneracy imposes a less demanding dimensionality constraint over (NSDP); in fact, in order to verify nondegeneracy, one must have $n \geqslant (m-r)(m-r+1)/2$, while weak-nondegeneracy may hold as long as $n \geqslant m-r$ (Remark 2). It is also clear from their definitions that weak-nondegeneracy implies weak-Robinson's CQ; and it is possible to show that the converse is not necessarily true. For instance, consider the constraint defined by:

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

and note that all orthogonal matrices $\overline{E} \in \mathbb{R}^{2 \times 2}$ have in their columns eigenvectors of $G(x)$, for every x. Since $v_{11}(\overline{x}, \overline{E}) = v_{22}(\overline{x}, \overline{E}) = 1$ for every $\overline{E}$, it follows that weak-nondegeneracy and weak-Robinson's CQ are equivalent to their strong counterparts in this case. Thus, from Proposition 2 we see that (weak-)nondegeneracy does not hold, while (weak-)Robinson's CQ does.

It is also clear from Proposition 2 (item 2) that Robinson's CQ implies weak-Robinson's CQ; however, we were not capable of finding a counterexample for the converse. We conjecture that they are equivalent.

Remark 3 If we replace the sequences $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ by matrix sequences $\{M^k\}_{k \in \mathbb{N}} \to G(\overline{x})$ in Definition 3, then we recover the nondegeneracy condition. Indeed, for any $\overline{E} \in \mathbb{R}^{m \times m-r}$ that spans $\operatorname{Ker} G(\overline{x})$, consider

$$M^k \doteq \overline{U} \Lambda^k \overline{U}^\top, \quad \text{with } \overline{U} \doteq [\overline{E}, u_{m-r+1}(G(\overline{x})), \ldots, u_m(G(\overline{x}))],$$

and $\Lambda^k \doteq \operatorname{Diag}(y^k)$ such that $y_i^k \doteq i/k$ for $i \in \{1, \ldots, m-r\}$, and $y_i^k \doteq \lambda_i(G(\overline{x}))$ otherwise. So clearly $M^k \to G(\overline{x})$ and the only convergent sequence E^k to $\overline{E}$ is $\overline{E}$ itself. Consequently, when we assume Definition 3 it necessarily follows that $\{v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m-r\}\}$ is linearly independent. Then, since $\overline{E}$ was chosen arbitrary, Proposition 2 implies that nondegeneracy holds true.

Remark 4 Remark 2 can be straightforwardly extended to structurally block diagonal matrix constraints, such as

$$G(x) \doteq \begin{bmatrix} G_1(x) & & \\ & \ddots & \\ & & G_q(x) \end{bmatrix} \succeq 0, \qquad \text{(Block-NSDP)}$$

where each "block" is defined by a continuously differentiable function $G_\ell \colon \mathbb{R}^n \to \mathbb{S}^{m_\ell}$, with $\ell \in \{1, \dots, q\}$, and $m_1 + \dots + m_q = m$. In fact, let $\overline{x} \in \mathcal{F}$ and $r_\ell \doteq \operatorname{rank}(G_\ell(\overline{x}))$ for each ℓ; and, for simplicity, let us assume that $r_\ell < m_\ell$ for all ℓ. Since $\operatorname{Ker} G(\overline{x}) = \operatorname{Ker} G_1(\overline{x}) \times \dots \times \operatorname{Ker} G_q(\overline{x})$, then $r = r_1 + \dots + r_q$. Then, weak-nondegeneracy (respectively, weak-Robinson's CQ) holds at $\overline{x}$ if, and only if, for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there are sequences of matrices $\{E_\ell^k\}_{k\in\mathbb{N}}$ such that:

- The columns of E_ℓ^k are unitary eigenvectors associated with the $m_\ell - r_\ell$ smallest eigenvalues of $G_\ell(x^k)$, for each $k \in \mathbb{N}$ and each $\ell \in \{1, \dots, q\}$;
- There are limit points $\overline{E}_\ell$ of $\{E_\ell^k\}_{k\in\mathbb{N}}$, $\ell \in \{1, \dots, q\}$, such that the set

$$\bigcup_{\ell=1}^{q} \left\{ v_{ii}^\ell(\overline{x}, \overline{E}_\ell) \colon i \in \{1, \dots, m_\ell - r_\ell\} \right\}$$

is (positive) linearly independent, where

$$v_{ij}^\ell(\overline{x}, \overline{E}_\ell) \doteq \left[\overline{e}_{\ell,i}^\top D_{x_1} G_\ell(\overline{x}) \overline{e}_{\ell,j} , \ \dots , \ \overline{e}_{\ell,i}^\top D_{x_n} G_\ell(\overline{x}) \overline{e}_{\ell,j} \right]^\top, \qquad (18)$$

and $\overline{e}_{\ell,1}, \dots, \overline{e}_{\ell,m_\ell-r_\ell}$ denote the columns of $\overline{E}_\ell$, for each $\ell \in \{1, \dots, q\}$.

The proof of this fact is elementary with [15, Lem. 1.3.10] and (8) at hand. Moreover, note that this is precisely the way weak-nondegeneracy would be defined for an equivalent multifold NSDP with constraints $G_1(x) \succeq 0, \dots, G_q(x) \succeq 0$. Thus, weak-nondegeneracy and weak-Robinson's CQ are invariant to block diagonal and multifold representations of (Block-NSDP). This is especially meaningful in problems that do not present an explicit block-diagonal representation, in which case it is not necessary to have prior knowledge of such a representation to talk about weak-nondegeneracy (or weak-Robinson's CQ), contrary to nondegeneracy.

Recall that the analysis we presented until this point showed, among other things, that some choices of $\overline{E}$ may be more meaningful than others. With this in mind, we are now led to revisit the work of Forsgren [13], who presented a very interesting way of talking about nondegeneracy in the presence of any sparsity structure that appears after applying a particular transformation to the problem. In the next section, we improve some of Forsgren's ideas by presenting a simplified and more general way of dealing with sparsity.

4 Dealing with structural sparsity

In this section, we take inspiration from a regularity condition introduced by Forsgren [13, Sect. 2.3], whose primary goal was to prove second-order optimality conditions for (NSDP). However, what makes Forsgren's condition specially interesting for us is the fact it can benefit from some sparsity structure of a certain Schur complement related to the constraint function. The main objective of this section is to present a more straightforward way of enjoying sparsity, based on Forsgren's results and Sect. 3. But before that, we present some of the notation used by Forsgren.

Given a point $\overline{x}$ and a matrix-valued function $F : \mathbb{R}^n \to \mathbb{S}^\beta$, consider the set $\mathcal{S}(F, \overline{x})$ defined as follows:

$$
\begin{aligned}
\mathcal{S}(F, \overline{x}) &\doteq \left\{ M \in \mathbb{S}^\beta : M_{ij} = 0 \text{ if } F_{ij}(x) \text{ is structurally zero near } \overline{x} \right\} \\
&= \left\{ M \in \mathbb{S}^\beta : M_{ij} = 0 \text{ if } \exists \varepsilon > 0 \text{ such that } F_{ij}(x) = 0, \forall x \in B(\overline{x}, \varepsilon) \right\}.
\end{aligned}
$$

For example, if $\beta = 3$ and for all x close to $\overline{x}$, we are able to identify non trivial mappings F_{ij} such that

$$
F(x) = \begin{bmatrix} F_{11}(x) & 0 & F_{13}(x) \\ 0 & F_{22}(x) & 0 \\ F_{13}(x) & 0 & F_{33}(x) \end{bmatrix}, \text{ then}
$$

$$
M \in \mathcal{S}(F, \overline{x}) \Leftrightarrow M = \begin{bmatrix} M_{11} & 0 & M_{13} \\ 0 & M_{22} & 0 \\ M_{13} & 0 & M_{33} \end{bmatrix},
$$

where M_{11}, M_{13}, M_{22}, and M_{33} may or may not be zero. Also, we define

$$
\mathcal{I}(F, \overline{x}) \doteq \left\{ (i, j) : \forall \varepsilon > 0, \exists x \in B(\overline{x}, \varepsilon) \text{ such that } F_{ij}(x) \neq 0, \ 1 \leqslant i \leqslant j \leqslant \beta \right\}
$$

as the set of indices that define the elements of $\mathcal{S}(F, \overline{x})$.

Forsgren's results are obtained in terms of the function

$$
\tilde{G}(x) \doteq G(x) - G(x)\overline{P}(\overline{P}^\top G(x)\overline{P})^{-1}\overline{P}^\top G(x),
$$

where $\overline{U} = [\overline{P}, \overline{E}]$ has columns that form an orthonormal eigenvector basis for $G(\overline{x})$, such that $\overline{E}$ spans the kernel of $G(\overline{x})$ and $\overline{P}^\top G(\overline{x})\overline{P} \succ 0$. Note that $\overline{E}^\top \tilde{G}(x)\overline{E}$ is the Schur complement of $\overline{P}^\top G(x)\overline{P}$ inside

$$
\overline{U}^\top G(x)\overline{U} = \begin{bmatrix} \overline{P}^\top G(x)\overline{P} & \overline{P}^\top G(x)\overline{E} \\ \overline{E}^\top G(x)\overline{P} & \overline{E}^\top G(x)\overline{E} \end{bmatrix}.
$$

Moreover, following Forsgren [13, Lem. 1], we see that $\tilde{G}(x) \succeq 0$ if, and only if, $G(x) \succeq 0$, for all x sufficiently close to $\overline{x}$, so the original NSDP problem can be locally reformulated as a minimization problem over $\tilde{G}(x) \succeq 0$, around $\overline{x}$. In fact, since

$$\overline{P}(\overline{P}^\top G(\overline{x})\overline{P})^{-1}\overline{P}^\top = \overline{P}\lambda_+(G(\overline{x}))^{-1}\overline{P}^\top$$
$$= \overline{U}\begin{bmatrix} \lambda_+(G(\overline{x}))^{-1} & 0 \\ 0 & 0 \end{bmatrix}\overline{U}^\top$$
$$= G(\overline{x})^\dagger,$$

where $G(\overline{x})^\dagger$ is the Moore-Penrose pseudoinverse of $G(\overline{x})$, it follows that $\tilde{G}(\overline{x}) = 0$ [13, Lem. 2], so $\tilde{G}$ can be considered a reduction to the kernel of $G(\overline{x})$ near $\overline{x}$.

The regularity condition introduced by Forsgren is as follows:

Definition 4 *(Forsgren's CQ)* Let $\overline{x} \in \mathcal{F}$ and let $\overline{U} \doteq [\overline{P}, \overline{E}]$ be an orthogonal matrix that diagonalizes $G(\overline{x})$, such that the columns of $\overline{E}$ span $\mathrm{Ker}\, G(\overline{x})$. Then, *Forsgren's CQ* holds at $\overline{x}$ with respect to $\overline{U}$ when

$$\mathrm{span}\left\{\overline{E}^\top D_{x_i} G(\overline{x})\overline{E} : i \in \{1, \ldots, n\}\right\} = \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E} \tag{F1}$$

and

$$\exists M \in \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E}, \quad \text{such that} \quad M \succ 0. \tag{F2}$$

Forsgren's CQ is indeed a constraint qualification, for when (F1) holds, then (F2) is equivalent to Robinson's CQ [13, Lem. 5]. However, although Forsgren states that any choice of $\overline{U}$ leads to a valid CQ, there is no discussion on the effects of this choice over the condition proposed. Under a specific condition, Forsgren's CQ provides uniqueness of the Lagrange multiplier [13, Thm. 1], but this condition varies with $\overline{U}$. Thus, different choices of $\overline{U}$ are likely to define different variants of Forsgren's CQ. This is not necessarily a negative point, but a comparison among those variants would be appropriate. For instance, from the practical point of view, one may be interested in knowing which choice of $\overline{U}$ defines the weakest CQ, or which one is easier to compute.

A result from Dorsch, Gómez, and Shikhman [12] shows that, ignoring the sparsity treatment, (F1) becomes equivalent to nondegeneracy.

Lemma 1 (Lem. 5 from [12]) *Let $\overline{x} \in \mathcal{F}$ and assume that $\mathcal{S}(\tilde{G}, \overline{x}) = \mathbb{S}^m$. Then, condition (F1) of Forsgren's CQ holds if, and only if, nondegeneracy holds at $\overline{x}$.*

However, similarly to weak-nondegeneracy, Forsgren's CQ also reduces to LICQ from NLP when G is structurally diagonal (as in (6)), contrasting with nondegeneracy. To put Forsgren's CQ in the same terms as the previous sections, we present an elementary characterization of it using the vectors $v_{ij}(\overline{x}, \overline{E})$ defined in Proposition 1:

Proposition 3 *Let $\overline{x} \in \mathcal{F}$ and let $\overline{E} \in \mathbb{R}^{m \times m - r}$ span $\mathrm{Ker}\, G(\overline{x})$. Then, condition (F1) of Forsgren's CQ holds at $\overline{x}$ if, and only if,*

$$\sum_{i=1}^{m-r}\sum_{j=i}^{m-r} M_{ij} v_{ij}(\overline{x}, \overline{E}) = 0, \quad M \in \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E} \quad \Rightarrow \quad M = 0,$$

where $r = \mathrm{rank}(G(\overline{x}))$.

Proof Let us assume that $r < m$, since otherwise the proof is trivial. We employ [13, Lem. 2], which states that

$$\overline{E}^\top D_{x_i} G(\overline{x})\overline{E} = \overline{E}^\top D_{x_i} \tilde{G}(\overline{x})\overline{E}$$

for all $i \in \{1, \ldots, n\}$, to ensure that the linear operator $\psi : \mathbb{R}^n \to \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E}$, defined by the action $\psi(d) \doteq \overline{E}^\top DG(\overline{x})[d]\overline{E}$ is well-defined.

With this in mind, note that

$$\mathrm{Im}\,(\psi) = \mathrm{span}\left\{\overline{E}^\top D_{x_i} G(\overline{x})\overline{E} : i \in \{1, \ldots, n\}\right\} = \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E}$$

if, and only if,

$$
\begin{aligned}
\mathrm{Ker}\,(\psi^*) &= \left\{ M \in \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E} : \langle \overline{E}^\top D_{x_\ell} G(\overline{x})\overline{E}, M \rangle = 0,\ \forall \ell \in \{1, \ldots, n\} \right\} \\
&= \{0\},
\end{aligned}
\tag{19}
$$

whence the result follows since

$$\overline{E}^\top D_{x_\ell} G(\overline{x})\overline{E} = [(v_{ij}(\overline{x}, \overline{E}))_\ell]_{i,j \in \{1, \ldots, m-r\}},$$

where $(v_{ij}(\overline{x}, \overline{E}))_\ell$ is the ℓ-th entry of the vector $v_{ij}(\overline{x}, \overline{E})$. □

As an abuse of language, (F1) consists of the "linear independence" of

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

with respect to the set $\overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E}$. In particular, when $G(\overline{x}) = 0$ and (F2) holds, take $\overline{U} = \overline{E} = \mathbb{I}_m$ and note that Forsgren's CQ holds for this particular choice of $\overline{U}$ if, and only if, the set $\left\{ \nabla G_{ij}(\overline{x}) : (i, j) \in \mathcal{I}(G, \overline{x}) \right\}$ is linearly independent, with $(i, i) \in \mathcal{I}(G, \overline{x})$ for all $i \in \{1, \ldots, m\}$.

Remark 5 As far as we understand, the relation between Forsgren's CQ and nondegeneracy was not formally established in [13]. To clarify this important detail, note that it is clear from Propositions 3 and 1 that nondegeneracy implies Forsgren's CQ. Moreover this implication is clearly strict, as nondegeneracy does not recover LICQ in a diagonal example.

The above discussion leads us to deal with sparsity in a more straightforward way, namely without taking Schur complements, which induces another weak variant of nondegeneracy.

4.1 A sparse variant of nondegeneracy

For any matrix $\overline{E}$ that spans $\mathrm{Ker}\,G(\overline{x})$, consider the function

$$\widehat{G}^{\overline{E}}(x) \doteq \overline{E}^\top G(x)\overline{E}$$

and note that $\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x}) = v_{ij}(\overline{x}, \overline{E})$ for all $i, j \in \{1, \ldots, m - r\}$ with $i \leqslant j$. We incorporate structural sparsity into nondegeneracy directly, but in a similar style of Forsgren's CQ (as characterized in Proposition 3), to introduce a new constraint qualification.

Definition 5 *(Sparse-nondegeneracy)* We say that *sparse-nondegeneracy* holds at $\overline{x} \in \mathcal{F}$ when either $\operatorname{Ker} G(\overline{x}) = \{0\}$ or there exists a matrix $\overline{E} \in \mathbb{R}^{m \times m - r}$ that spans $\operatorname{Ker} G(\overline{x})$ and such that:

1. The set $\left\{ v_{ij}(\overline{x}, \overline{E}) \colon (i, j) \in \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x}), 1 \leqslant i \leqslant j \leqslant m - r \right\}$ is linearly independent;
2. $(i, i) \in \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x})$ for all $i \in \{1, \ldots, m - r\}$.

There are two natural questions about sparse-nondegeneracy that we shall answer in the following paragraphs. The first one consists of knowing whether the sparse-nondegeneracy condition is a genuine constraint qualification; and the second one concerns about the relation between Definition 5 and other constraint qualifications, such as nondegeneracy, Forsgren's CQ, and Robinson's CQ. To address these questions, we first prove an elementary characterization of sparse-nondegeneracy:

Lemma 2 *Let $\overline{x} \in \mathcal{F}$ be such that $\operatorname{Ker} G(\overline{x}) \neq \{0\}$, and let $\overline{E}$ be a matrix that spans $\operatorname{Ker} G(\overline{x})$. Then, item 1 of Definition 5 holds at $\overline{x}$ if, and only if, there is no nonzero $\tilde{Y} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$ such that $DG(\overline{x})^*[\overline{E} \tilde{Y} \overline{E}^\top] = 0$.*

Proof The result follows directly by noticing that

$$\sum_{(i,j) \in \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x})} v_{ij}(\overline{x}, \overline{E}) \tilde{Y}_{ij} = DG(\overline{x})^*[\overline{E} \tilde{Y} \overline{E}^\top] \tag{20}$$

for every $\tilde{Y} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$. □

Next, we prove that sparse-nondegeneracy implies Robinson's CQ, which also shows that it is indeed a constraint qualification.

Proposition 4 *If $\overline{x} \in \mathcal{F}$ satisfies sparse-nondegeneracy, then it also satisfies Robinson's CQ.*

Proof The result follows trivially when $\operatorname{Ker} G(\overline{x}) = \{0\}$, so let us assume that $r = \operatorname{rank}(G(\overline{x})) < m$. Suppose that sparse-nondegeneracy holds at $\overline{x} \in \mathcal{F}$, and take any $Z \succeq 0$ such that $\langle Z, G(\overline{x}) \rangle = 0$ and $DG(\overline{x})^*[Z] = 0$, then there exists some $Y \in \mathbb{S}^{m-r}_+$ such that $Z = \overline{E} Y \overline{E}^\top$. Define the matrix $\tilde{Y} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$ whose (i, j)-th entry is given by

$$\tilde{Y}_{ij} \doteq \begin{cases} Y_{ij}, & \text{if } (i, j) \in \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x}) \\ 0, & \text{otherwise,} \end{cases}$$

and note that

$$DG(\overline{x})^*[Z] = DG(\overline{x})^*[\overline{E} Y \overline{E}^\top] = DG(\overline{x})^*[\overline{E} \tilde{Y} \overline{E}^\top] = 0, \tag{21}$$

so $\tilde{Y} = 0$ due to Lemma 2. Moreover, from item 2 of Definition 5, $(i, i) \in \mathcal{I}(\widehat{G}^E, \overline{x})$ for all $i \in \{1, \dots, m - r\}$, so the diagonal of Y must consist only of zeros, which implies that $Y = 0$ and, consequently, $Z = 0$. Since Z is arbitrary, Robinson's CQ holds. □

We highlight that item 2 of Definition 5 is not superfluous, for removing it may cause us to lose the property of being a constraint qualification. Indeed, the following example illustrates that:

Example 3 Consider the problem:

$$\begin{aligned} \underset{x \in \mathbb{R}^2}{\text{Minimize}} \quad & x_2, \\ \text{subject to} \quad & G(x) \doteq \begin{bmatrix} x_1 & x_2 \\ x_2 & 0 \end{bmatrix} \succeq 0, \end{aligned}$$

which has $\overline{x} \doteq (0, 0)$ as one of its solutions. The point $\overline{x}$ satisfies Definition 5 after removing item 2, with $\overline{E} \doteq \mathbb{I}_2$, because $v_{11}(\overline{x}, \overline{E}) = (1, 0)$ and $v_{12}(\overline{x}, \overline{E}) = (0, 1)$ are linearly independent; but $\overline{x}$ does not satisfy the KKT conditions since there is no $\overline{Y} \succeq 0$ such that $\overline{Y}_{11} = 0$ and $\overline{Y}_{12} = \overline{Y}_{21} = 1/2$. Thus, Definition 5 is not a constraint qualification without item 2.

Remark 6 Let us show that when item 2 fails, the problem can be reformulated such that it holds. Let $\overline{x} \in \mathcal{F}$ and $\overline{E}$ be a matrix that spans $\mathrm{Ker}\, G(\overline{x})$. If item 2 of Definition 5 is not satisfied, then let $J \doteq \{i \in \{1, \dots, m - r\} : (i, i) \notin \mathcal{I}(\widehat{G}^E, \overline{x})\}$ and note that there exists some $\varepsilon > 0$ such that

$$G(x) \in \mathbb{S}_+^m \text{ if, and only if, } G(x) \in \mathbb{S}_+^m \bigcap_{i \in J} \{\overline{e}_i \overline{e}_i^\top\}^\perp,$$

for every $x \in B(\overline{x}, \varepsilon)$, where $\overline{e}_i$ denotes the i-th column of $\overline{E}$. That is, the feasible set $\mathcal{F}$ coincides locally with the preimage of the face $F \doteq \mathbb{S}_+^m \bigcap_{i \in J} \{\overline{e}_i \overline{e}_i^\top\}^\perp$ of $\mathbb{S}_+^m$. Moreover, since F is a face of $\mathbb{S}_+^m$, then there is an orthogonal matrix $V \doteq [V_1, V_2] \in \mathbb{R}^{m \times m}$ such that

$$V^\top F V = \left\{ \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix} : M \in \mathbb{S}_+^{m-\omega} \right\},$$

where ω is the cardinality of J [22, Eq. 2.3]. This means that it is possible to locally replace the original constraint of (NSDP) by the equality constraint $V_2^\top G(x) = 0$ and a smaller semidefinite constraint $\mathcal{G}(x) \doteq V_1^\top G(x) V_1 \in \mathbb{S}_+^{m-\omega}$. If F is minimal, then the new constraint $\mathcal{G}(x) \in \mathbb{S}_+^{m-\omega}$ satisfies item 2 of Definition 5 at $\overline{x}$. Otherwise, this process can be repeated until the minimal face is reached. Thus, every problem

can be equivalently reformulated (reducing dimension if necessary), such that item 2 always holds. In particular, when G is an affine function, then this procedure can be computed via a popular preprocessing technique called *facial reduction* (we refer to Pataki [22] and references therein for more details about it). When $G(\overline{x}) = 0$ and $\overline{E} = \mathbb{I}_m$, this procedure can be done by simply removing the i-th row and the i-th column of G, for every i such that $(i, i) \notin \mathcal{I}(G, \overline{x})$, and including the correspondent equality constraints into the problem. We recall that all of our results can be easily extended to NSDP problems with separate equality constraints.

Let us illustrate this procedure using Example 3. In this case we have $\overline{e}_2^\top G(x)\overline{e}_2 = 0$ for every x; then $x \in \mathcal{F}$ if, and only if, $G(x) \in F$, where

$$F \doteq \mathbb{S}^2_+ \bigcap \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}^\perp = \left\{ \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix} : \alpha \geqslant 0, \right\}$$

which means that the constraint of the problem can be equivalently written as $x_2 = 0$ and $x_1 \geqslant 0$; for which $\overline{x}$ satisfies Definition 5 and the KKT conditions.

Remark 7 If G is structurally diagonal as in (6), then $\overline{x}$ satisfies sparse-nondegeneracy if, and only if, the set $\{\nabla g_i(\overline{x}) : g_i(\overline{x}) = 0\}$ is linearly independent. Moreover, this can be extended to block-diagonal constraints. In this case, assuming the same notation as Remark 4, sparse nondegeneracy holds at a feasible point $\overline{x}$ of (Block-NSDP) if, and only if, for each $\ell \in \{1, \ldots, q\}$ there is some matrix $\overline{E}_\ell$ that spans $\operatorname{Ker} G_\ell(\overline{x})$, such that:

- For all $i \in \{1, \ldots, m_\ell - r_\ell\}$, we have $(i, i) \in \mathcal{I}(G_\ell^{\overline{E}_\ell}, \overline{x})$;
- The set

$$\bigcup_{\ell=1}^q \left\{ v_{ij}^\ell(\overline{x}, \overline{E}_\ell) : (i, j) \in \mathcal{I}(G_\ell^{\overline{E}_\ell}, \overline{x}) \right\}$$

is linearly independent, where $v_{ij}^\ell(\overline{x}, \overline{E}_\ell)$ is defined as in (18).

Note that this is how sparse-nondegeneracy would be defined for a multifold equivalent representation of (Block-NSDP), with constraints $G_1(x) \succeq 0, \ldots, G_q(x) \succeq 0$.

In view of Remark 7, it is easy to build a diagonal counterexample for the converse of Proposition 4. For instance, take $m = 2$ and set $\overline{x} = 0$; then, define the constraint

$$G(x) \doteq \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix}, \tag{22}$$

and note that $v_{11}(\overline{x}, \overline{E}) = v_{22}(\overline{x}, \overline{E}) = 1$ for every matrix $\overline{E}$ that spans $\operatorname{Ker} G(\overline{x})$. Hence, sparse-nondegeneracy does not hold, although Robinson's CQ does.

Furthermore, Remark 7 reveals a similarity among sparse-nondegeneracy, Forsgren's CQ, and weak-nondegeneracy, which is the fact they all reduce to LICQ when considering a diagonal matrix constraint. Moreover, it follows directly from Propositions 1 and 2 that nondegeneracy also strictly implies sparse-nondegeneracy. To make

a rough comparison between Forsgren's CQ and sparse-nondegeneracy, note that both evaluate linear independence of the set $\{v_{ij}(\overline{x}, \overline{E}): 1 \leqslant i \leqslant j \leqslant m - r\}$, but while item 1 of Definition 5 takes coefficients structured as in $\mathcal{S}(\widehat{G}^E, \overline{x})$, condition (F1) takes coefficients structured as in $\overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E}$. This suggests that they are different conditions. In fact, Example 1 can also be used to show that neither weak- nor sparse-nondegeneracy imply Forsgren's CQ.

Example 4 (same as Example 1) Consider the constraint:

$$G(x) \doteq \begin{bmatrix} x_1 & x_2 \\ x_2 & x_1 \end{bmatrix}$$

and the point $\overline{x} \doteq (0, 0)$, which satisfies weak-nondegeneracy and violates nondegeneracy (Example 1). Also:

- **Sparse-nondegeneracy holds at $\overline{x}$**: take the same $\overline{E}$ as above and we have

$$\widehat{G}^{\overline{E}}(x) \doteq \begin{bmatrix} x_1 - x_2 & 0 \\ 0 & x_1 + x_2 \end{bmatrix}$$

  and $\mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x}) = \{(1, 1), (2, 2)\}$;
- **Forsgren's CQ does not hold at $\overline{x}$**: in this case Forsgren's CQ is equivalent to non-degeneracy, which does not hold because if $\overline{E} \doteq \mathbb{I}_m$, then $v_{11}(\overline{x}, \overline{E}) = v_{22}(\overline{x}, \overline{E}) = [1, 0]$.

Thus, neither weak- nor sparse-nondegeneracy imply Forsgren's CQ.

Moreover, if $G(\overline{x}) = 0$ then $\tilde{G} = G$ and in this case Forsgren's CQ implies sparse-nondegeneracy (see Proposition 3 and the discussion afterwards). Whether this still holds or not when $G(\overline{x}) \neq 0$ is an open problem that we are currently unable to address, due to the intricate form of $\tilde{G}$ in the general case.

An elementary consequence of Lemma 2 is that sparse-nondegeneracy guarantees uniqueness of the Lagrange multiplier with respect to a fixed sparsity pattern, which is similar to a result proven for Forsgren's CQ [13].

Proposition 5 *Let $\overline{x}$ be a KKT point of (NSDP) that satisfies item 1 of Definition 5 and let $\overline{E}$ be the matrix that certifies it, which spans* $\text{Ker } G(\overline{x})$. *Then,* $\Lambda(\overline{x}) \bigcap \left(\overline{E}\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})\overline{E}^\top \right)$ *is a singleton.*

Proof Firstly, to see why $\Lambda(\overline{x}) \bigcap \left(\overline{E}\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})\overline{E}^\top \right) \neq \emptyset$ we resort to a result of [8, Thm. 7] which states that under Robinson's CQ any accumulation point $\overline{Y}$ of the sequence

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

must belong to $\Lambda(\overline{x})$. But clearly, for all $k \in \mathbb{N}$ large enough, we see that $Y^k \in \overline{E}\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})\overline{E}^\top$ and so does $\overline{Y}$.

Now let $Y_1, Y_2 \in \Lambda(\overline{x}) \cap \left(\overline{E} \mathcal{S}(\widehat{G}^E, \overline{x}) \overline{E}^\top \right)$ be Lagrange multipliers associated with $\overline{x}$, define $Y \doteq Y_1 - Y_2$, and by definition there exists some $Z \in \mathcal{S}(\widehat{G}^E, \overline{x})$ such that $Y = \overline{E} Z \overline{E}^\top$ and $DG(\overline{x})^*[\overline{E} Z \overline{E}^\top] = 0$. By Lemma 2 we must have $Z = 0$ and, consequently, $Y_1 = Y_2$. □

Another important property of sparse-nondegeneracy is that the number of structural zeros of $\widehat{G}^E$, at points that satisfy it, remains the same regardless of $\overline{E}$.

Proposition 6 *Let $\overline{x} \in \mathcal{F}$ be such that $\mathrm{Ker}\, G(\overline{x}) \neq \{0\}$, and let $\overline{E}$ and $\overline{W}$ be matrices that span $\mathrm{Ker}\, G(\overline{x})$, such that item 1 of Definition 5 holds. Then, $\#\mathcal{I}(\widehat{G}^E, \overline{x}) = \#\mathcal{I}(\widehat{G}^W, \overline{x})$.*

Proof Let $Z \doteq [z_{\ell s}]_{\ell, s \in \{1, \dots, m-r\}}$ be an invertible matrix such that $\overline{E} Z = \overline{W}$ and note that $\widehat{G}^W(\overline{x}) = Z^\top \widehat{G}^E(\overline{x}) Z$, so

$$\widehat{G}_{ij}^W(\overline{x}) = \langle \widehat{G}^E(\overline{x}), z_i z_j^\top \rangle = \sum_{\ell, s=1}^r z_{\ell i} z_{sj} \widehat{G}_{\ell s}^E(\overline{x}),$$

where z_i denotes the i-th column of Z, and

$$\nabla \widehat{G}_{ij}^W(\overline{x}) = \sum_{\ell, s=1}^r z_{\ell i} z_{sj} \nabla \widehat{G}_{\ell s}^E(\overline{x}).$$

Rephrasing,

$$\nabla \widehat{G}_{ij}^W(\overline{x}) = \mathrm{unfold}(D\widehat{G}^E(\overline{x})) \mathrm{vec}(z_i z_j^\top)$$

where $\mathrm{unfold}\colon \mathbb{R}^{m-r \times m-r \times n} \to \mathbb{R}^{n \times (m-r)^2}$ is an *unfolding* operator for the tensor $D\widehat{G}^E(\overline{x})$ when it is seen as an $m - r \times m - r$ matrix with n-dimensional entries. Also, $\mathrm{vec}\colon \mathbb{R}^{m-r \times m-r} \to \mathbb{R}^{(m-r)^2}$ is the usual *vectorization* operator, which transforms a matrix into a vector by stacking up its columns, from left to right. That is,

$$\mathrm{unfold}(D\widehat{G}^E(\overline{x})) \doteq \begin{bmatrix} | & & | & | & & | \\ \nabla \widehat{G}_{11}^E(\overline{x}) & \dots & \nabla \widehat{G}_{m-r,1}^E(\overline{x}) & \nabla \widehat{G}_{12}^E(\overline{x}) & \dots & \nabla \widehat{G}_{m-r,m-r}^E(\overline{x}) \\ | & & | & | & & | \end{bmatrix}$$

and

$$\mathrm{vec}(z_i z_j^\top) \doteq \begin{bmatrix} z_{1i} z_{1j} \\ \vdots \\ z_{m-r,i} z_{1j} \\ z_{1i} z_{2j} \\ \vdots \\ z_{m-r,i} z_{m-r,j} \end{bmatrix}.$$

Consequently,

$$\text{unfold}(D\widehat{G}^{\overline{W}}(\overline{x})) = \text{unfold}(D\widehat{G}^{\overline{E}}(\overline{x}))(Z \otimes Z),$$

where

$$Z \otimes Z \doteq \left[\begin{array}{ccccc} | & & | & | & & | \\ \text{vec}(z_1 z_1^\top) & \ldots & \text{vec}(z_1 z_r^\top) & \text{vec}(z_2 z_1^\top) & \ldots & \text{vec}(z_r z_r^\top) \\ | & & | & | & & | \end{array} \right],$$

is the *Kronecker product* of Z with itself.

But since Z is invertible, $Z \otimes Z$ is also invertible, which means that

$$\text{span}\left(\left\{\nabla \widehat{G}_{ij}^{\overline{W}}(\overline{x}) \colon i \leqslant j\right\}\right) = \text{span}\left(\left\{\nabla \widehat{G}_{ij}^{\overline{E}}(\overline{x}) \colon i \leqslant j\right\}\right).$$

Then, since $\nabla \widehat{G}_{ij}^{\overline{E}}(\overline{x}) = 0$ for all $(i, j) \in \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x})$ (and the same holds for $\overline{W}$), it follows that

$$\text{span}\left(\left\{\nabla \widehat{G}_{ij}^{\overline{W}}(\overline{x}) \colon (i, j) \in \mathcal{I}(\widehat{G}^{\overline{W}}, \overline{x})\right\}\right) = \text{span}\left(\left\{\nabla \widehat{G}_{ij}^{\overline{E}}(\overline{x}) \colon (i, j) \in \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x})\right\}\right).$$

Finally, since item 1 of Definition 5 holds for both $\overline{E}$ and $\overline{W}$, we conclude that $\#\mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x}) = \#\mathcal{I}(\widehat{G}^{\overline{W}}, \overline{x})$. □

Proposition 6 tells us that the strength of sparse-nondegeneracy is invariant with respect to $\overline{E}$. That is, if there are multiple matrices $\overline{E}$ certifying sparse-nondegeneracy at a point $\overline{x}$, then they all induce similar conditions. In our opinion, this is an advantage with respect to Forsgreen's CQ. As for weak-nondegeneracy, we were not able to find any counterexample nor prove any relation between them. In fact, finding this relation seems a challenging task since there is no clear relation between the eigenvectors of $G(x)$ and its sparsity structure, in general.

Remark 8 It is noteworthy that it is also possible to define another variant of Robinson's CQ that enjoys sparsity, by replacing $\tilde{Y} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$ by $\tilde{Y} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x}) \cap \mathbb{S}_+^{m-r}$ in Lemma 2. This definition is strictly implied by sparse-nondegeneracy (see the example given in (22)). Moreover, it is clear that this variant of Robinson's CQ is implied by Robinson's CQ, but the converse is also an open question. The proof that this is a CQ follows similarly to the proof of Theorem 2.

Remark 9 Regarding second-order optimality conditions, we call the reader's attention to the fact that for each $\overline{x} \in \mathcal{F}$ and each $\overline{E}$ that spans $\text{Ker } G(\overline{x})$, there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that

$$G(\mathcal{V}) = G(\mathcal{V}) \cap \underbrace{\overline{E}\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})\overline{E}^\top}_{\doteq H}$$

This means that near $\overline{x}$ we can consider a new space $\mathbb{S}^H \doteq \mathbb{S}^m \cap H$, define a new cone $\mathbb{S}_+^H \doteq \mathbb{S}_+^m \cap H$ which is still closed and convex, and in this setting sparse-nondegeneracy induces a second-order optimality condition, which is inherited from [11, Thm. 3.45]. Namely, for every $d \in DG(\overline{x})^{-1}(T_{\mathbb{S}_+^m \cap H}(G(\overline{x})) \cap \{\nabla f(\overline{x})\}^\perp$ it holds that

$$\sup_{Y \in \Lambda(\overline{x}) \cap H} \left(d^\top \nabla^2 L(\overline{x}, Y)d - \sigma(Y, T^2_{\mathbb{S}_+^m \cap H}(G(\overline{x}), DG(\overline{x})[d])) \right) \geqslant 0, \qquad (23)$$

because sparse-nondegeneracy implies Robinson's CQ, which in turn is carried over to the reduced problem, but since $\Lambda(\overline{x}) \cap H$ is a singleton, we have for $\overline{Y} \in \Lambda(\overline{x}) \cap H$ that

$$d^\top \nabla^2 L(\overline{x}, \overline{Y})d - \sigma(\overline{Y}, T^2_{\mathbb{S}_+^m \cap H}(G(\overline{x}), DG(\overline{x})[d])) \geqslant 0.$$

Although this condition concerns the reduced problem, mostly, it can also bring some information about the original problem, for an inequality analogous to (23) in terms of $\sup_{Y \in \Lambda(\overline{x})}$ is also true. Above, $T^2_{\mathbb{S}_+^m \cap H}(G(\overline{x}), DG(\overline{x})[d])$ denotes the *second-order tangent set* to $\mathbb{S}_+^m \cap H$ at $G(\overline{x})$ along $DG(\overline{x})[d]$ (see [11, Def. 3.28]), and $\sigma(Y, T^2_{\mathbb{S}_+^m \cap H}(G(\overline{x}), DG(\overline{x})[d]))$ denotes its support function.

4.2 Zeros of the gradients and sparse-nondegeneracy

In this short ending section, we discuss how to improve sparse-nondegeneracy even further. This is mainly motivated by the realization that the idea of disregarding "structural zeros" in the study of regularity is actually too conservative. Since nondegeneracy is mainly concerned with the derivative of G at $\overline{x}$ instead of the value of G in a neighborhood of $\overline{x}$, we can in fact ignore all entries of G whose gradients are zero at $\overline{x}$, which is done by considering the following sets:

$$\mathcal{S}_\nabla(F, \overline{x}) \doteq \left\{ M \in \mathbb{S}^\beta : M_{ij} = 0 \text{ if } \nabla F_{ij}(\overline{x}) = 0 \right\}$$

and

$$\mathcal{I}_\nabla(F, \overline{x}) \doteq \left\{ (i, j) : \nabla F_{ij}(\overline{x}) \neq 0, \ 1 \leqslant i \leqslant j \leqslant \beta \right\}.$$

For example, if $n = 1$ and $\beta = 3$, for all x close to $\overline{x} \doteq 0$ we have, as an example,

$$\text{if } F(x) \doteq \begin{bmatrix} x & 0 & x^2 \\ 0 & x & 1 \\ x^2 & 1 & x \end{bmatrix} \text{ then } M \in \mathcal{S}(F, \overline{x}) \Leftrightarrow M = \begin{bmatrix} M_{11} & 0 & 0 \\ 0 & M_{22} & 0 \\ 0 & 0 & M_{33} \end{bmatrix},$$

(24)

where M_{11}, M_{22}, and M_{33} may or may not be zero. Then, we can define a condition similarly to Definition 5 but in terms of $\mathcal{I}_\nabla$:

Definition 6 *(GS-nondegeneracy)* We say that the condition *gradient-sparse -nondegeneracy* (GS-nondegeneracy) holds at $\overline{x} \in \mathcal{F}$ if either $\operatorname{Ker} G(\overline{x}) = \{0\}$ or there exists a matrix $\overline{E} \in \mathbb{R}^{m \times m-r}$ that spans $\operatorname{Ker} G(\overline{x})$ such that:

1. The set $\left\{ v_{ij}(\overline{x}, \overline{E}) \colon (i, j) \in \mathcal{I}_\nabla(\widehat{G^E}, \overline{x}), 1 \leqslant i \leqslant j \leqslant m - r \right\}$ is linearly independent;
2. $(i, i) \in \mathcal{I}_\nabla(\widehat{G^E}, \overline{x})$ for all $i \in \{1, \ldots, m - r\}$.

The interesting properties of GS-nondegeneracy that make it worth an extended comment are twofold. The first one is that sparse-nondegeneracy is strictly stronger than GS-nondegeneracy. Noticing that $\mathcal{I}_\nabla(\widehat{G^E}, \overline{x}) \subseteq \mathcal{I}(\widehat{G^E}, \overline{x})$ is enough to see the implication and the next example shows that the converse is not necessarily true.

Example 5 Let

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

and consider the constraint $G(x) \succeq 0$ at the point $\overline{x} \doteq (0, 0)$. In this case, Forsgren's CQ fails at $\overline{x}$ with $\overline{E} \doteq \mathbb{I}_2$ because

$$\operatorname{span} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \neq \mathcal{S}(\tilde{G}, \overline{x}) = \mathbb{S}^2.$$

In fact, (F1) fails for every orthogonal matrix $\overline{E}$. Furthermore, regardless of $\overline{E}$ the vectors $v_{11}(\overline{x}, \overline{E})$, $v_{22}(\overline{x}, \overline{E})$, and $v_{12}(\overline{x}, \overline{E}) \in \mathbb{R}^2$, are linearly dependent and $\mathcal{I}(\widehat{G^E}, \overline{x}) = \{(1, 1), (1, 2), (2, 2)\}$, hence sparse-nondegeneracy also fails to hold at $\overline{x}$. On the other hand, note that for $\overline{E} = \mathbb{I}_2$, we obtain $\mathcal{I}_\nabla(\widehat{G^E}, \overline{x}) = \{(1, 1), (2, 2)\}$ and

$$v_{11}(\overline{x}, \overline{E}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad v_{22}(\overline{x}, \overline{E}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

are linearly independent, so GS-nondegeneracy holds at $\overline{x}$.

We remark that Lemma 2 and Propositions 4, 5, and 6, can be also stated and proved in terms of GS-nondegeneracy. Moreover, if Forsgren's CQ was defined in terms of $\mathcal{I}_\nabla(\tilde{G}, \overline{x})$ instead of $\mathcal{I}(\tilde{G}, \overline{x})$, we would obtain precisely Definition 6 (due to [13, Lem. 2]), which is quite unexpected. The second interesting aspect of GS-nondegeneracy is that it presents a notion of stability, in the sense of ignoring small perturbations. Formally:

Theorem 3 *Let $\overline{x} \in \mathcal{F}$ and $\delta \colon \mathbb{R}^n \to \mathbb{S}^m$ be any continuously differentiable function such that $\delta(\overline{x}) = 0$ and $D\delta(\overline{x}) = 0$. Then, GS-nondegeneracy holds at $\overline{x}$ for the constraint $G(x) \succeq 0$ if, and only if, it holds for the constraint $G_\delta(x) \doteq G(x) + \delta(x) \succeq 0$ at the same point.*

Proof Direct from the fact $DG(\overline{x}) = DG_\delta(\overline{x})$ and $\mathcal{I}_\nabla(\widehat{G}^{\overline{E}}, \overline{x}) = \mathcal{I}_\nabla(\widehat{G}_\delta^{\overline{E}}, \overline{x})$. $\qquad\square$

Despite the apparent triviality of Theorem 3, observe that it is essentially telling us that any noise of order two can be disregarded, as we could observe in Example 5.

5 Conclusions

In this paper, we studied the nondegeneracy condition of Shapiro and Fan [27] with the purpose of incorporating some matrix structure into it, such as spectral decompositions and structural sparsity. Our work was motivated by a well-known limitation of nondegeneracy, which is the fact it generally fails in the presence of structural sparsity in the constraint function. For example, we recall that a NSDP problem with multiple constraints may be equivalently reformulated as a single block diagonal constraint, but nondegeneracy is not expected to be preserved in the process. This limitation may have important consequences in practice, since many algorithms are theoretically supported by nondegeneracy and, on the other hand, structural sparsity is a very common trait of optimization models of real world problems.

To address this issue, we proposed three variants of nondegeneracy, here called weak-nondegeneracy, sparse-nondegeneracy, and GS-nondegeneracy. They were proven to be strictly weaker than the classical nondegeneracy. In particular, all new constraint qualifications only require the dimension constraint $n \geqslant m - r$, which is considerably less demanding than the constraint $n \geqslant (m-r)(m-r+1)/2$ imposed by nondegeneracy. Also, they are invariant to multifold or block diagonal formulations of (NSDP) and, consequently, they recover the LICQ condition from NLP when the constraint function is structurally diagonal.

All our conditions are inspired by *sequential optimality conditions* [3, 8] which provide simple proofs for the facts that the conditions we define are CQs (the proof for sparse-nondegeneracy and GS-nondegeneracy were not presented but they are left for the reader). Besides the simplicity of the approach, the convergence of an external penalty method to KKT points under these CQs is obtained automatically (see the discussion after Theorem 2), which is a direct application of the new CQs. Also, several other CQs for NLP have been recently (re)invented with sequential optimality conditions in mind. In particular, the so-called *constant rank constraint qualification* (CRCQ) by Janin [16], and the *constant positive linear dependence* (CPLD) of Qi and Wei [23], together with their weaker counterparts [6, 7, 18]. Previous attempts have been made to extend these CQs to the conic context, but they have turned out to be flawed [2] or incomplete [4], since the results in [4] are only relevant for multifold conic problems where at least one block of constraints is such that the zero eigenvalue is simple. The approach we present in this paper gives the proper tools for providing the extension of all mentioned CQs to the context of general NSDPs and, more generally, to optimization over symmetric cones, also extending the global convergence results to more practical algorithms. For instance, in NLP, it is known that the convergence theory of a safeguarded augmented Lagrangian method can be built around CPLD [1], which will also be the case for its NSDP variant [8]. A continuation of this paper will appear shortly with these results.
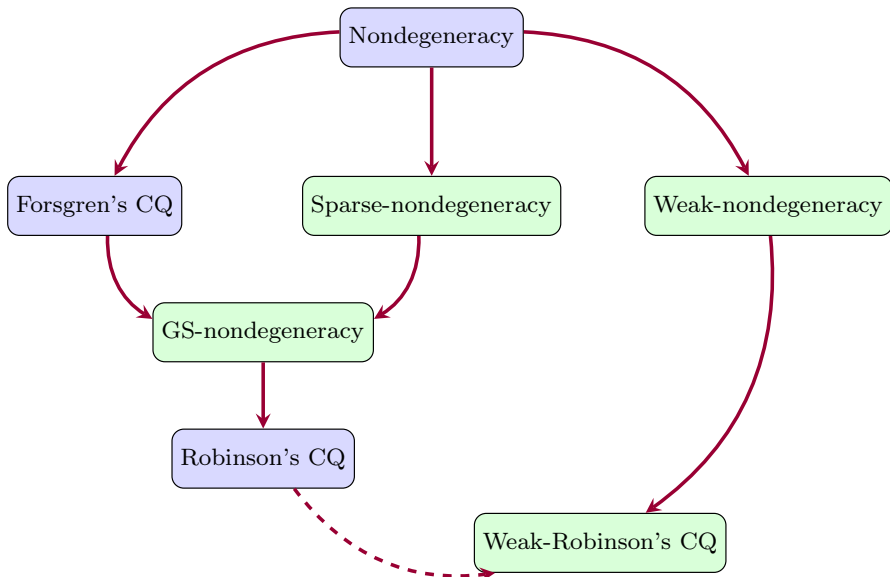
Fig. 1 Relationship among some CQs for NSDP. Classical CQs are in blue boxes, while new CQs are in green boxes. Arrows indicate strict implications, except for the dashed arrow where the reverse implication is unknown

With this in mind, we believe that the concepts introduced in this paper are interesting enough to shed a new light to the classical theme of constraint nondegeneracy for conic programming, showing, in particular, how to redefine it in such a way that linear independence can be replaced by weaker notions. In this process, new and interesting challenging open questions have appeared which we believe should be addressed. In particular, new studies should be conducted to clarify the relationship between weak-nondegeneracy and sparse-nondegeneracy, together with the relationship between weak-Robinson's CQ and Robinson's CQ (see Fig. 1).

References

1. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented lagragrian methods under the constant positive linear dependence constraint qualification. Math. Program. **111**(1–2), 5–32 (2008)
2. Andreani, R., Fukuda, E. H., Haeser, G., Ramírez C. H., Santos, D. O., Silva, P. J. S., Silveira, T. P.: Erratum to: new constraint qualifications and optimality conditions for second order cone programs. To appear in Set Valued Variation. Anal. (2021). https://doi.org/10.1007/s11228-021-00573-5.
3. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**, 627–641 (2011)
4. 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. Opt. Lett. (2021). https://doi.org/10.1007/s11590-021-01737-w
5. Andreani, R., Haeser, G., Mito, L.M., Ramos, A., Secchin, L.D.: On the best achievable quality of limit points of augmented Lagrangian schemes. Numer. Algorithms (2021). https://doi.org/10.1007/s11075-021-01212-8

6. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Opt. **22**, 1109–1135 (2012)
7. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**, 255–273 (2012)
8. Andreani, R., Haeser, G., Viana, D.S.: Optimality conditions and global convergence for nonlinear semidefinite programming. Math. Program. **180**(1–2), 203–235 (2020)
9. Anjos, M. F., Lassere, J. B.: (eds.). Handbook on semidefinite, conic and polynomial optimization. International Series in Operations Research and Management Science. Springer US. New York (2012)
10. Auslender, A., Ramírez, H.: Penalty and barrier methods for convex semidefinite programming. Math. Methods Oper. Res., **63**, Article number: 195 (2006)
11. Bonnans, J.F., Shapiro, A.: Pertubation Analysis of Optimization Problems. Springer-Verlag, Berlin (2000)
12. Dorsch, D., Gómez, W., Shikhman, V.: Sufficient optimality conditions hold for almost all nonlinear semidefinite programs. Math. Program. **158**(1–2), 77–97 (2016)
13. Forsgren, A.: Optimality conditions for nonconvex semidefinite programming. Math. Program. **88**, 105–128 (2000)
14. Fusek, P.: On metric regularity for weakly almost piecewise smooth functions and some applications in nonlinear semidefinite programming. SIAM J. Opt. **23**(2), 1041–1061 (2013)
15. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1985)
16. Janin, R.: Direction derivative of the marginal function in nonlinear programming. Math. Program. Stud. **21**, 127–138 (1984)
17. Klatte, D., Kummer, B.: Aubin property and uniqueness of solutions in cone constrained optimization. Math. Methods Oper. Res. **77**(3), 291–304 (2013)
18. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011)
19. Mordukhovich, B.S., Nghia, T.T.A., Rockafellar, R.T.: Full stability in finite-dimensional optimization. Math. Oper. Res. **40**(1), 226–252 (2015)
20. Mordukhovich, B.S., Outrata, J. V., Ramírez, H.: Graphical derivatives and stability analysis for parameterized equilibria with conic constraints. Set Valued Anal. Appl., pp. 1–18 (2015)
21. Mordukhovich, B.S., Outrata, J.V., Ramírez, H.: Second-order variational analysis in conic programming with applications to optimality and stability. SIAM J. Opt. **25**(1), 76–101 (2015)
22. Pataki, G.: Strong duality in conic linear programming: facial reduction and extended duals. In: Bailey, D., et al. (eds.) Computational and Analytical Mathematics, vol. 50, pp. 613–634. Springer Proceedings in Mathematics and Statistics, New York, NY (2013)
23. Qi, L., Wei, Z.: On the constant positive linear dependence conditions and its application to SQP methods. SIAM J. Opt. **10**, 963–981 (2000)
24. Robinson, S.M.: First-order conditions for general nonlinear optimization. SIAM J. Appl. Math. **30**(4), 597–610 (1976)
25. Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. Math. Program. Ser. B **77**(2), 301–320 (1997)
26. Shapiro, A.: On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints. SIAM J. Opt. **7**(2), 508–518 (1997)
27. Shapiro, A., Fan, M.K.H.: On Eigenvalue optimization. SIAM J. Opt. **5**(3), 552–569 (1995)
28. Sun, D.: The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Math. Oper. Res. **31**(4), 761–776 (2006)
29. Wachsmuth, G.: On LICQ and the uniqueness of Lagrange multipliers. Oper. Res. Lett. **41**(1), 78–80 (2013)
30. Wolkowicz, H., Saigal, R. L.: Vandenberghe (eds.) Handbook of Semidefinite Programming: Theory, Algorithms, an Applications. In: International Series in Operations Research and Management Science. Springer US. New York, (2000)
31. Yamashita, H., Yabe, H.: A survey of numerical methods for nonlinear semidefinite programming. J. Oper. Res. Soc. Japan **58**(1), 24–60 (2015)