Contextuality and noncontextuality measures and generalized Bell inequalities for cyclic systems

Ehtibar N. Dzhafarov,[1,*] Janne V. Kujala,[2,†] and Víctor H. Cervantes[1,‡]

[1] *Purdue University, USA*

[2] *University of Turku, Finland*

Cyclic systems of dichotomous random variables have played a prominent role in contextuality research, describing such experimental paradigms as the Klyachko-Can-Binicioğlu-Shumovsky, Einstein-Podolsky-Rosen-Bell, and Leggett-Garg ones in physics, as well as conjoint binary choices in human decision making. Here, we understand contextuality within the framework of the Contextuality-by-Default (CbD) theory, based on the notion of probabilistic couplings satisfying certain constraints. CbD allows us to drop the commonly made assumption that systems of random variables are consistently connected (i.e., it allows for all possible forms of "disturbance" or "signaling" in them). Consistently connected systems constitute a special case in which CbD essentially reduces to the conventional understanding of contextuality. We present a theoretical analysis of the degree of contextuality in cyclic systems (if they are contextual) and the degree of noncontextuality in them (if they are not). By contrast, all previously proposed measures of contextuality are confined to consistently connected systems, and most of them cannot be extended to measures of noncontextuality. Our measures of (non)contextuality are defined by the L_1-distance between a point representing a cyclic system and the surface of the polytope representing all possible noncontextual cyclic systems with the same single-variable marginals. We completely characterize this polytope, as well as the polytope of all possible probabilistic couplings for cyclic systems with given single-variable marginals. We establish that, in relation to the maximally tight Bell-type CbD inequality for (generally, inconsistently connected) cyclic systems, the measure of contextuality is proportional to the absolute value of the difference between its two sides. For noncontextual cyclic systems, the measure of noncontextuality is shown to be proportional to the smaller of the same difference and the L_1-distance to the surface of the box circumscribing the noncontextuality polytope. These simple relations, however, do not generally hold beyond the class of cyclic systems, and noncontextuality of a system does not follow from noncontextuality of its cyclic subsystems.

I. INTRODUCTION

A *cyclic system of rank* $n = 2, 3, \ldots$, is a system

$$\mathcal{R} = \left\{ \left\{ R_i^i, R_{i\oplus 1}^i \right\} : i = 1, \ldots, n \right\}, \tag{1}$$

where $i \oplus 1 = i + 1$ for $i < n$, and $n \oplus 1 = 1$; R_j^i denotes a Bernoulli (0/1) random variable measuring *content* q_j in *context* c_i ($j = i, i \oplus 1$). A content is any property that can be present or absent (e.g., spin of a half-spin particle in a given direction), a context here is defined by which two contents are measured together (simultaneously or in a specific order). A cyclic system of rank n has n contexts containing two *jointly distributed* random variables each, $\left\{R_i^i, R_{i\oplus 1}^i\right\}$. Each of such pairs is referred to as a *bunch* (of random variables). The system also has n *connections* $\left\{R_i^{i\ominus 1}, R_i^i\right\}$ (where $i \ominus 1 = i - 1$ for $i > 1$, and $1 \ominus 1 = n$), each of which contains two *stochastically unrelated* (i.e., possessing no joint distribution) random variables measuring the same content in two different contexts.

Cyclic systems have played a prominent role in contextuality studies [1, 2]. The matrices below represent cyclic systems of rank 5 (describing, e.g., the Klyachko-Can-Binicioğlu-Shumovsky experiment [3, 4]), rank 4 (describing, e.g., Bell's "Alice-Bob" experiments [5–8]), rank 3 (describing, e.g., the Leggett-Garg experiments [9–12]), and rank 2 (of primary interest outside quantum physics, e.g., describing the question-order experiment in human decision making [13, 14]).

$$\begin{array}{|c|c|c|c|c||c|}\hline R_1^1 & R_2^1 & & & & c^1 \\ \hline & R_2^2 & R_3^2 & & & c^2 \\ \hline & & R_3^3 & R_4^3 & & c^3 \\ \hline & & & R_4^4 & R_5^4 & c^4 \\ \hline R_1^5 & & & & R_5^5 & c^5 \\ \hline\hline q_1 & q_2 & q_3 & q_4 & q_5 & \mathcal{R}_5 \\ \hline \end{array} \quad \begin{array}{|c|c|c|c||c|}\hline R_1^1 & R_2^1 & & & c^1 \\ \hline & R_2^2 & R_3^2 & & c^2 \\ \hline & & R_3^3 & R_4^3 & c^3 \\ \hline R_1^4 & & & R_4^4 & c^4 \\ \hline\hline q_1 & q_2 & q_3 & q_4 & \mathcal{R}_4 \\ \hline \end{array}$$

$$\begin{array}{|c|c|c||c|}\hline R_1^1 & R_2^1 & & c^1 \\ \hline & R_2^2 & R_3^2 & c^2 \\ \hline R_1^3 & & R_3^3 & c^3 \\ \hline\hline q_1 & q_2 & q_3 & \mathcal{R}_3 \\ \hline \end{array} \quad \begin{array}{|c|c||c|}\hline R_1^1 & R_2^1 & c^1 \\ \hline R_1^2 & R_2^2 & c^2 \\ \hline\hline q_1 & q_2 & \mathcal{R}_2 \\ \hline \end{array} \tag{2}$$

A cyclic system is *consistently connected* (satisfies the "no-disturbance" or "no-signaling" condition) if R_i^i and $R_i^{i\ominus 1}$ are identically distributed for $i = 1, \ldots, n$. This assumption is commonly made in quantum physical applications. The present paper, however, is based on the Contextuality-by-Default (CbD) theory [15–17], which is not predicated on this assumption, that is, the systems of random variables we consider are generally *inconsistently connected.* Cyclic systems have been intensively analyzed within the framework of CbD [2, 9, 13, 18, 19]. In this paper they are studied in relation to the measures of contextuality and noncontextuality considered in Ref. [17].

The familiarity of the reader with CbD (e.g., Refs. [15, 17]) for understanding this paper is not necessary, even if desirable. We recapitulate here all relevant definitions and results, although they are presented in the

* To whom correspondence should be addressed. E-mail: ehtibar@purdue.edu

† E-mail: jvk@iki.fi

‡ E-mail: cervantv@purdue.edu

form specialized to cyclic systems rather than in complete generality, so the broader motivation behind the constructs may not always be apparent. In particular, we take it for granted in this paper that it is important not to be constrained by the confines of consistent connectedness [2, 18]. The simplest reason for this is that if a consistently connected system is contextual or noncontextual by one's definition, then it is reasonable to require from this definition that the system's contextuality status should not change under sufficiently small perturbations rendering it inconsistently connected. Another reason is that inconsistent connectedness is ubiquitous. Thus, in accordance with the quantum-mechanical laws, consistent connectedness does not generally hold for sequential measurements, e.g., for the Leggett-Garg system [9, 10]. In other experimental paradigms it is often violated due to unavoidable or inadvertent design biases [4]. In all such cases, use of CbD to analyze data has proved to be useful [2, 9, 20–24]. At the same time, all contextuality measures proposed outside CbD are confined to consistent connectedness [25–29].

We also take for granted in this paper that it is desirable to seek principled and unified ways of measuring both contextuality and noncontextuality [17]. Degree of contextuality has been related to such concepts as quantum advantage in computation and communication complexity [30–32], and generally is viewed as a measure of nonclassicality of a system. Moreover, it is intrinsically interesting to compare different contextual systems in terms of which of them can be more easily rendered noncontextual by perturbing its random variables (see Ref. [26] for an overview). Intrinsic interest in measures of noncontextuality can be justified similarly. It is too uninformative to simply view noncontextual systems as having zero contextuality: some of them would be easier than others to render contextual by perturbing their random variables. Remarkably, there seem to be no measures of noncontextuality proposed in the literature prior to Ref. [17], and most of the proposed measures of contextuality (e.g., the Contextual Fraction measure proposed in Ref. [25] and generalized to inconsistently connected systems in Ref. [17]) do not naturally extend to measures of noncontextuality. By "natural extension" we mean the extension to noncontextual systems using the same principles as in constructing a contextuality measure being extended.

Note that the term "degree of noncontextuality" in this paper always applies to noncontextual systems only, in the same way as "degree of contextuality" only applies to contextual systems. This is useful to mention because "degree of noncontextuality" has been used in the literature in a different meaning: as a measure complementary to the degree of contextuality in contextual systems. Thus, the Noncontextual Fraction measure in Ref. [25] is unity minus Contextual Fraction measure. Both are defined for contextual systems, while noncontextual ones all have Noncontextual Fraction equal to unity.

Of the several measures of contextuality considered in Ref. [17] we focus here on two, labeled CNT_1 and CNT_2. The former is the oldest measure introduced within the framework of CbD [2, 18, 19], whereas CNT_2 is the newest one, discussed in Ref. [17]. A detailed description of these measures will have to wait until we have introduced the necessary definitions and results. In a nutshell, however, a cyclic system (with the distribution of each of the random variables R_j^i being fixed) is represented in CbD by two vectors of product expectations, $\mathbf{p_b}$ and $\mathbf{p_c}$, conventionally referred to as vectors of "correlations." (We will only use this term, strictly speaking, incorrect, in this informal introduction, due to its familiarity in the contextuality literature.) The subscripts $\mathbf{b}$ and $\mathbf{c}$ stand for the just-defined CbD terms "bunch" and "connection." The vector $\mathbf{p_b}$ encodes the correlations within the bunches $\left\{R_i^i, R_{i\oplus 1}^i\right\}$, $i = 1, \ldots, n$. The vector $\mathbf{p_c}$ encodes the correlations imposed on the within-connection pairs $\left\{R_i^{i\ominus 1}, R_i^i\right\}$, $i = 1, \ldots, n$, defining thereby so-called *couplings* of the connections (recall that the connections themselves do not possess joint distributions). A cyclic system whose (non)contextuality we measure is represented by vectors $\mathbf{p_b^*}, \mathbf{p_c^*}$, where $\mathbf{p_b^*}$ consists of the observed bunch correlations, and $\mathbf{p_c^*}$ consists of the correlations computed for the connections in a special way (the *maximal coupling*s of the connections). In the case of CNT_1, the L_1-distance is measured between $\mathbf{p_c^*}$ and the *feasibility polytope* $\mathbb{P}_\mathbf{c}$ comprising all possible $\mathbf{p_c}$-vectors compatible with $\mathbf{p_b^*}$. In the case of CNT_2, L_1-distance is computed between $\mathbf{p_b^*}$ and the *noncontextuality polytope* $\mathbb{P}_\mathbf{b}$ comprising all $\mathbf{p_b}$-vectors compatible with $\mathbf{p_c^*}$. The two measures therefore are, in a well-defined sense, mirror images of each other.

In this paper, we provide a complete characterization of the noncontextuality polytope, and show that the L_1-distance between this polytope and the observed vector $\mathbf{p_b^*}$ is a single-coordinate distance, i.e. it can be computed along a single coordinate of $\mathbf{p_b}$. Moreover, when $\mathbf{p_b^*}$ is outside this polytope, this distance is the same along all coordinates of $\mathbf{p_b}$ (see Fig. 1A), and it is proportional to the amount of violation of the generalized Bell criterion derived in Ref. [19] for noncontextuality of (generally inconsistently connected) cyclic systems.[33] In other words, if we schematically present the Bell criterion as stating that a system is noncontextual if and only if some expression E does not exceed a constant k, then CNT_2 is proportional to $E - k$ when this value is positive. Since precisely the same is true for CNT_1 [19], with the same proportionality coefficient, we have

$$\mathrm{CNT}_2 = \mathrm{CNT}_1. \tag{3}$$

To understand why this is the case, we characterize the polytope $\mathbb{P}$ of all possible vectors $(\mathbf{p_b}, \mathbf{p_c})$, and show that its L_1-distance from the vector $(\mathbf{p_b^*}, \mathbf{p_c^*})$ representing the observed contextual cyclic system has the same properties as above: it is a single-coordinate distance, the same along any of the coordinates of $(\mathbf{p_b}, \mathbf{p_c})$. The equality of the two measures follows from this immediately.

Despite the fact that CNT_1 and CNT_2 are "mirror images" of each other, only one of them, CNT_2, was shown in Ref. [17] to be naturally extendable to a measure of the degree of noncontextuality in noncontextual systems, NCNT_2. Geometrically, this measure is the L_1-distance between a point $\mathbf{p_b^*}$ inside the noncontextuality polytope $\mathbb{P}_\mathbf{b}$ and the polytope's surface. It is, too, a single-coordinate distance (as is the case for any internal point of any convex region [34]), but its properties

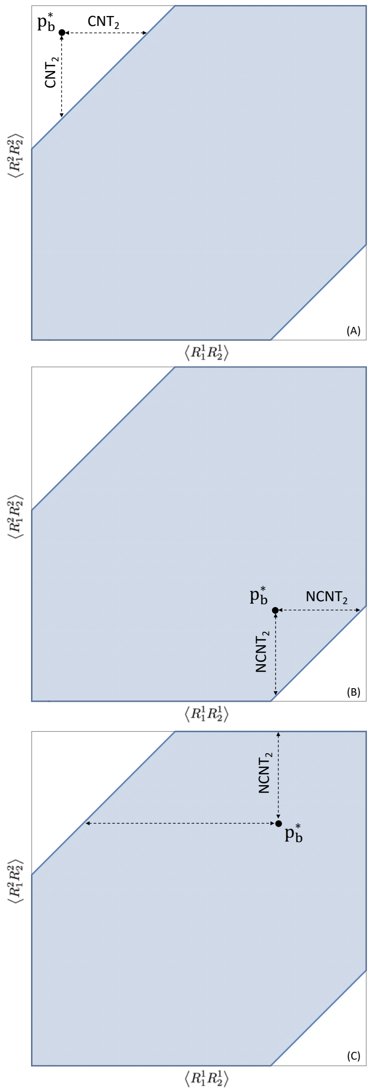

Figure 1. The relationship between a noncontextuality polytope $\mathbb{P}_{\mathbf{b}}$ (light-colored hexagon), the circumscribing box $\mathbb{R}_{\mathbf{b}}$ (rectangle), and a point $\mathbf{p}_{\mathbf{b}}^*$ representing a system (small circle). The illustration uses cyclic systems of rank 2, shown as $\mathcal{R}_2$ in (2). The point $\mathbf{p}_{\mathbf{b}}^*$ in (A) represents a contextual system, and its L_1-distance to the polytope is CNT_2; it can be measured along any of the (here, two) coordinates. The point $\mathbf{p}_{\mathbf{b}}^*$ in (B) represents a noncontextual system, and since its L_1-distance to an internal face of the polytope (here, edge of the hexagon) is smaller than that to the surface of $\mathbb{R}_{\mathbf{b}}$, this distance is NCNT_2, and it can be measured along any single coordinate. The point $\mathbf{p}_{\mathbf{b}}^*$ in (C) represents a noncontextual system which is L_1-closer to the surface of $\mathbb{R}_{\mathbf{b}}$, and this (also single-coordinate) distance from the surface is NCNT_2.

are somewhat more complicated due to the structure of $\mathbb{P}_{\mathbf{b}}$. The polytope $\mathbb{P}_{\mathbf{b}}$ is circumscribed by an n-box $\mathbb{R}_{\mathbf{b}}$, so that some of the faces of $\mathbb{P}_{\mathbf{b}}$ lie within the box's interior, while others lie within its surface. If the point $\mathbf{p}_{\mathbf{b}}^*$ is L_1-closer to an internal face of $\mathbb{P}_{\mathbf{b}}$ than to the surface of the box, NCNT_2 can be measured along any single coordinate of $\mathbf{p}_{\mathbf{b}}$ (see Fig. 1B), and it is proportional to the amount of compliance of the system with the generalized Bell criteria of noncontextuality [19]. In other words, in this case NCNT_2 is proportional to $k - E$ if the criterion is written as $E \leq k$. However, NCNT_2 becomes the L_1-distance between $\mathbf{p}_{\mathbf{b}}^*$ and the surface of the box $\mathbb{R}_{\mathbf{b}}$ when this distance is smaller than that to any internal face of $\mathbb{P}_{\mathbf{b}}$ (Fig. 1C). In this case, NCNT_2 is not related to the Bell inequalities.

One might wonder why we could not simply define the degree of contextuality by the amount of violation of the appropriate Bell criterion (and, by extension, define the degree of noncontextuality by the amount of compliance with it). Brunner and coauthors address this approach in Ref. [26], where they discuss contextuality in the special form of nonlocality. They call this approach "a common choice for quantifying nonlocality," and correctly point out that it is untenable, because there can be a potential infinity of the alternatives $E' \leq k'$ to $E \leq k$ such that the two inequalities are equivalent but $E - k$ and $E' - k'$ are grossly different. Our approach is to define contextuality and noncontextuality as certain distances in the space of points representing cyclic systems, and then to see how these distances are related to specific forms of the generalized Bell criteria of noncontextuality.

The choice of L_1-distances is natural and convenient when dealing with probabilities, because of their additivity. However, due to the special structure of the noncontextuality polytope, any L_p-distance ($p \geq 1$), including the Euclidean (L_2) and supremal (L_∞) ones, are simply scaled versions of L_1:

$$L_p \equiv n^{\frac{1-p}{p}} L_1, \tag{4}$$

where n is the rank of the cyclic system. The consequences of replacing L_1-distances with other L_p-distances in our measures of contextuality and noncontextuality are discussed in Sec. VIII.

In the concluding section we consider the question of whether the regularities established in this paper for cyclic systems extend to noncyclic systems as well. We answer this question in the negative: in particular, CNT_1 and CNT_2 are not generally equal, nor is one of them any function of the other.

II. CYCLIC SYSTEMS

In each context $i = 1, \ldots, n$ of the cyclic system (1), the joint distribution of the bunch $\left\{R_i^i, R_{i\oplus 1}^i\right\}$ is described by three numbers,

$$\begin{aligned} \left\langle R_i^i \right\rangle &= p_i^i = \Pr\left[R_i^i = 1\right], \\ \left\langle R_{i\oplus 1}^i \right\rangle &= p_{i\oplus 1}^i = \Pr\left[R_{i\oplus 1}^i = 1\right], \\ \left\langle R_i^i R_{i\oplus 1}^i \right\rangle &= p_{i,i\oplus 1} = \Pr\left[R_i^i = R_{i\oplus 1}^i = 1\right]. \end{aligned} \tag{5}$$

(One does not need a superscript for the product expectation because the context is uniquely determined by the two contents measured in this context.) For instance, a cyclic system of rank 4 has all bunch distributions in it described as shown in Fig. 2.

A cyclic system therefore can be represented by two column vectors:

$$\mathbf{p}_{\mathbf{l}} = \left(1, p_1^1, p_2^1 \ldots, p_n^n, p_1^n\right)^\intercal, \tag{6}$$

which is the vector of single-variable expectations preceded by $\langle\rangle = 1$ (the index $\mathbf{l}$ stands for "low-level marginals"), and

$$\mathbf{p}_{\mathbf{b}} = \left(p_{12}, p_{23}, \ldots, p_{n-1,n}, p_{n1}\right)^\intercal, \tag{7}$$

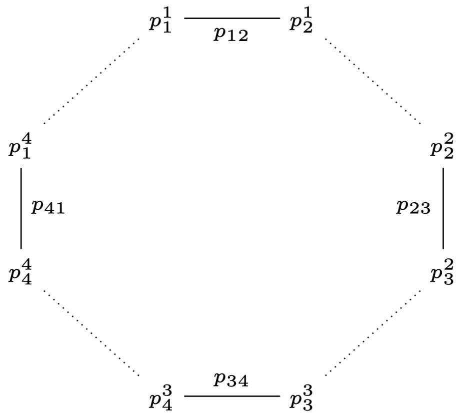

Figure 2. Diagram of all bunch distributions for a rank-4 cyclic system (e.g., the distribution in context c_1 is described by $p_1^1 = \Pr\left[R_1^1 = 1\right]$, $p_2^1 = \Pr\left[R_2^1 = 1\right]$, and $p_{12} = \Pr\left[R_1^1 = R_2^1 = 1\right]$.)

the vector of all bunch *product expectations*.

A *coupling of a connection* $\left\{R_i^i, R_i^{i\ominus 1}\right\}$ is a pair of *jointly distributed* random variables $\left\{T_i^i, T_i^{i\ominus 1}\right\}$ with the same 1-marginals:

$$\begin{array}{c}\left\langle T_i^i\right\rangle = \left\langle R_i^i\right\rangle = p_i^i,\\ \left\langle T_i^{i\ominus 1}\right\rangle = \left\langle R_i^{i\ominus 1}\right\rangle = p_i^{i\ominus 1}.\end{array} \tag{8}$$

In other words, a coupling adds to each pair $p_i^i, p_i^{i\ominus 1}$ describing the connection a product expectation

$$\left\langle T_i^i T_i^{i\ominus 1}\right\rangle = p^{i,i\ominus 1} = \Pr\left[T_i^i = T_i^{i\ominus 1} = 1\right], \tag{9}$$

as it is shown in Fig. 3. This can generally be done in an infinity of ways, constrained only by

$$\max\left(0, p_i^i + p_i^{i\ominus 1} - 1\right) \le p^{i,i\ominus 1} \le \min\left(p_i^i, p_i^{i\ominus 1}\right). \tag{10}$$

If couplings are constructed for all connections, they are represented by a vector of *connection product expectations*,

$$\mathbf{p_c} = \left(p^{1n}, p^{21}, p^{32}, \ldots, p^{n,n-1}\right)^{\intercal}. \tag{11}$$

An (overall) *coupling of the entire system* $\mathcal{R}$ is a set

$$\mathcal{S} = \left\{S_j^i : j = i, i\oplus 1; i = 1, \ldots, n\right\} \tag{12}$$

of jointly distributed random variables such that, for $i = 1, \ldots, n$,

$$\begin{array}{c}\left\langle S_i^i\right\rangle = \left\langle R_i^i\right\rangle = p_i^i,\\ \left\langle S_{i\oplus 1}^i\right\rangle = \left\langle R_{i\oplus 1}^i\right\rangle = p_{i\oplus 1}^i,\\ \left\langle S_i^i S_{i\oplus 1}^i\right\rangle = \left\langle R_i^i R_{i\oplus 1}^i\right\rangle = p_{i,i\oplus 1}.\end{array} \tag{13}$$

In other words, a coupling $\mathcal{S}$ induces as its 1-marginals and 2-marginals the same vectors $\mathbf{p_l}, \mathbf{p_b}$ as those representing $\mathcal{R}$. An overall coupling also induces couplings of all connections as its 2-marginals $\left(S_i^i, S_i^{i\ominus 1}\right)$, which means that it induces a vector $\mathbf{p_c}$ of connection product expectations.

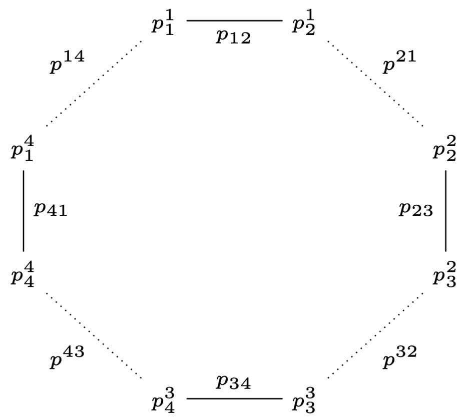

Figure 3. The same diagram as in Fig. 2, complemented with coupling distributions for the connections (e.g., for the coupling of the connection corresponding to q_1 the distribution is described by $p_1^1 = \Pr\left[T_1^1 = 1\right] = \Pr\left[R_1^1 = 1\right]$, $p_1^4 = \Pr\left[T_1^4 = 1\right] = \Pr\left[R_1^4 = 1\right]$, and $p^{14} = \Pr\left[T_1^1 = T_1^4 = 1\right]$).

III. (NON)CONTEXTUALITY

In the following it is convenient to speak of cyclic systems as *represented* by vectors

$$\mathbf{p} = \begin{pmatrix}\mathbf{p_l}\\ \mathbf{p_b}\\ \mathbf{p_c}\end{pmatrix}, \tag{14}$$

even though $\mathbf{p_c}$ is computed and added to a given system. Since this can be done in multiple ways, one and the same system is represented by multiple vectors $\mathbf{p}$.

If in a vector $\mathbf{p_c}$,

$$p^{i,i\ominus 1} = \min\left(p_i^i, p_i^{i\ominus 1}\right), i = 1, \ldots, n, \tag{15}$$

then the values of $p^{i,i\ominus 1}$ are maximal possible ones, and the couplings of the connections used to compute these product expectations are called *maximal couplings.* In particular, if the system is consistently connected, i.e.,

$$p_i^i = p_i^{i\ominus 1} = p_i, i = 1, \ldots, n, \tag{16}$$

then the joint and marginal probabilities in the maximal coupling are as shown,

probability of	$T_i^i = 1$	$T_i^i = 0$	
$T_i^{i\ominus 1} = 1$	p_i	0	p_i
$T_i^{i\ominus 1} = 0$	0	$1 - p_i$	$1 - p_i$
	p_i	$1 - p_i$	

(17)

whence

$$\Pr\left[T_i^i \ne T_i^{i\ominus 1}\right] = 0, i = 1, \ldots, n. \tag{18}$$

In other words, in a consistently connected system the random variables in each connection are treated as if they were essentially the same random variable. In the general case, with p_i^i and $p_i^{i\ominus 1}$ not necessarily equal, it is easy to show that

$$\Pr\left[T_i^i \ne T_i^{i\ominus 1}\right] = \left|p_i^i - p_i^{i\ominus 1}\right|, i = 1, \ldots, n. \tag{19}$$

That is, the maximal coupling $\left\{T_i^i, T_i^{i\ominus 1}\right\}$ of $\left\{R_i^i, R_i^{i\ominus 1}\right\}$ provides a natural measure of difference between the two variables (in fact, it is the *total variation distance* between them). The intuitive meaning of contextuality can be presented in the form of the following counterfactual: if all the random variables in the system containing pairs R_i^i and $R_i^{i\ominus 1}$ were jointly distributed, it would force some of these pairs (measuring "the same thing" in different contexts) to be more dissimilar than they are in isolation.

Let us agree that an *observed*, or *target* system $\mathcal{R}$ (one being investigated) is represented by the vector

$$\mathbf{p}^* = \begin{pmatrix} \mathbf{p}_\mathbf{l}^* \\ \mathbf{p}_\mathbf{b}^* \\ \mathbf{p}_\mathbf{c}^* \end{pmatrix}, \tag{20}$$

where $\mathbf{p}_\mathbf{l}^*$ and $\mathbf{p}_\mathbf{b}^*$ are as they are observed, and $\mathbf{p}_\mathbf{c}^*$ is the vector of the maximal connection product expectations.

Definition 1. A target system $\mathcal{R}$ represented by vector $(\mathbf{p}_\mathbf{l}^*, \mathbf{p}_\mathbf{b}^*, \mathbf{p}_\mathbf{c}^*)^\intercal$ is *noncontextual* if it has a coupling $\mathcal{S}$ that induces as its marginals the vector $\mathbf{p}_\mathbf{c}^*$ (of maximal connection product expectations). If no such coupling exists, the system is contextual.

In other words, if a system is noncontextual it has an overall coupling that (by definition) satisfies (13), and also

$$\left\langle S_i^i S_i^{i\ominus 1}\right\rangle = p^{i,i\ominus 1} = \min\left(p_i^i, p_i^{i\ominus 1}\right), i = 1, \ldots, n. \tag{21}$$

In the case of consistent connectedness, CbD essentially reduces to the conventional contextuality analysis (see Refs. [16] and [35] for logical ramifications of this reduction). As an example, for a consistently connected cyclic system of rank 3,

$$\begin{array}{|c|c|c||c|}\hline R_1^1 & R_2^1 & & c^1 \\ \hline & R_2^2 & R_3^2 & c^2 \\ \hline R_1^3 & & R_3^3 & c^3 \\ \hline\hline q_1 & q_2 & q_3 & \mathcal{R}_3 \\ \hline \end{array}, \tag{22}$$

if it is noncontextual, its coupling satisfying Definition 1, due to (18), can be presented as

$$\begin{array}{|c|c|c||c|}\hline S_1 & S_2 & & c^1 \\ \hline & S_2 & S_3 & c^2 \\ \hline S_1 & & S_3 & c^3 \\ \hline\hline q_1 & q_2 & q_3 & \mathcal{S}_3 \\ \hline \end{array}, \tag{23}$$

involving just three random variables recorded two at a time.

Let

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_\mathbf{l} \\ \mathbf{M}_\mathbf{b} \\ \mathbf{M}_\mathbf{c} \end{pmatrix} \tag{24}$$

be a Boolean (incidence) matrix with 0/1 cells. The 2^{2n} columns of $\mathbf{M}$ are indexed by events

$$S_1^1 = r_1^1, S_2^1 = r_2^1, \ldots, S_n^n = r_n^n, S_1^n = r_1^n, \tag{25}$$

while its rows are indexed by the elements of $\mathbf{p}$ (with $\mathbf{M}_\mathbf{l}$ corresponding to $\mathbf{p}_\mathbf{l}$, $\mathbf{M}_\mathbf{b}$ to $\mathbf{p}_\mathbf{b}$, and $\mathbf{M}_\mathbf{c}$ to $\mathbf{p}_\mathbf{c}$). A cell (l, m) of $\mathbf{M}$ is filled with 1 if the following is satisfied: for each random variable S_j^i entering the expectation that indexes the lth row of $\mathbf{M}$, the value of S_j^i in the event indexing the mth column of $\mathbf{M}$ is equal to 1. Otherwise the cell is filled with zero. For instance, if the lth row of $\mathbf{M}$ corresponds to the expectation $\left\langle S_1^1 S_1^2\right\rangle$ in $\mathbf{p}$, we put 1 in the cell (l, m) if both r_1^1 and r_1^2 in the event (25) corresponding to the mth column of $\mathbf{M}$ are 1; otherwise the cell is filled with zero.

Once $\mathbf{p}^*$and $\mathbf{M}$ are defined, one can reformulate the definition of (non)contextuality as follows.

Definition 2 (equivalent to Definition 1)**.** A target system $\mathcal{R}$ represented by vector $\mathbf{p}^* = (\mathbf{p}_\mathbf{l}^*, \mathbf{p}_\mathbf{b}^*, \mathbf{p}_\mathbf{c}^*)^\intercal$ is noncontextual if and only if there is a vector $\mathbf{h} \geq 0$ (componentwise) such that

$$\mathbf{M}\mathbf{h} = \mathbf{p}^*. \tag{26}$$

Otherwise the system is contextual.

It is easy to show that if such a vector $\mathbf{h}$ exists, then it can always be interpreted as the column-vector of probabilities

$$\Pr\left[S_1^1 = r_1^1, S_2^1 = r_2^1, \ldots, S_n^n = r_n^n, S_1^n = r_1^n\right]$$

for some overall coupling $\mathcal{S}$ of a system, across all 2^{2n} combinations of $r_j^i = 0/1$. In particular, the elements of $\mathbf{h}$ sum to 1, because the first row of $\mathbf{M}$ and the first element of $\mathbf{p}^*$ consist of 1's only.

IV. RELABELING FROM 0/1 TO ±1

For many aspects of cyclic systems it is more convenient to label the values of the random variables ± 1 rather than consider them Bernoulli, 0/1. This amounts to switching from R_j^i variables to $A_j^i = 2R_j^i - 1$. In the case of the connection couplings (8), this means switching from T_j^i to $U_j^i = 2T_j^i - 1$. A cyclic system $\mathcal{R}$ with Bernoulli variables will then be renamed into a cyclic system $\mathcal{A}$ with ± 1-variables. We have, for $i = 1, \ldots, n$,

$$\begin{aligned} \left\langle A_j^i\right\rangle &= e_j^i = 2p_j^i - 1, j = i, i \oplus 1, \\ \left\langle A_i^i A_{i\oplus 1}^i\right\rangle &= e_{i,i\oplus 1} = 4p_{i,i\oplus 1} - 2p_i^i - 2p_{i\oplus 1}^i + 1, \\ \left\langle U_i^i U_i^{i\ominus 1}\right\rangle &= e^{i,i\ominus 1} = 4p^{i,i\ominus 1} - 2p_i^i - 2p_i^{i\ominus 1} + 1, \end{aligned} \tag{27}$$

and this defines componentwise the transformation of the expectation vectors

$$\begin{pmatrix} \mathbf{e}_\mathbf{l} \\ \mathbf{e}_\mathbf{b} \\ \mathbf{e}_\mathbf{c} \end{pmatrix} = \phi \begin{pmatrix} \mathbf{p}_\mathbf{l} \\ \mathbf{p}_\mathbf{b} \\ \mathbf{p}_\mathbf{c} \end{pmatrix}. \tag{28}$$

The relabeling in question is useful in the formulation of the *Bell-type criterion of noncontextuality*. Let us denote

$$s_1\left(\mathbf{e}_\mathbf{b}\right) = \max_{\substack{\lambda_i = \pm 1, i = 1, \ldots, n \\ \prod_{i=1}^n \lambda_i = -1}} \sum \lambda_i e_{i,i\oplus 1}, \tag{29}$$

$$\delta\left(\mathbf{e_l}\right)=\sum_{i=1}^{n}\left|e_{i}^{i}-e_{i}^{i\ominus1}\right|. \tag{30}$$

and

$$\Delta\left(\mathbf{e_l}\right)=\min\left(n-2+\delta\left(\mathbf{e_l}\right),n\right). \tag{31}$$

Note that δ and Δ depend on $\mathbf{e_l}$, but since this vector is fixed, we may (and will henceforth) consider δ and Δ as constants [36].

Theorem 3 (Kujala-Dzhafarov [19])**.** *A cyclic system $\mathcal{A}$ represented by vector $\left(\mathbf{e}_\mathbf{l}^*,\mathbf{e}_\mathbf{b}^*,\mathbf{e}_\mathbf{c}^*\right)^\intercal$ is noncontextual if and only if*

$$s_1\left(\mathbf{e}_\mathbf{b}^*\right)-\Delta\leq 0. \tag{32}$$

This result generalizes the criterion derived in Ref. [1] for consistently connected cyclic systems (those with $\delta = 0$).

V. MEASURES OF CONTEXTUALITY AND A MEASURE OF NONCONTEXTUALITY

The idea of the two measures of contextuality considered in Ref. [17], CNT_1 and CNT_2, is as follows. First we think of the space of all $\mathbf{p}=(\mathbf{p_l},\mathbf{p_b},\mathbf{p_c})^\intercal$ obtainable as $\mathbf{p}=\mathbf{Mh}$ with $\mathbf{h}\geq 0$. In this space, we fix the 1-marginals $\mathbf{p_l}$ at $\mathbf{p}_\mathbf{l}^*$ (observed values), and define the polytope

$$\mathbb{P}=\left\{\left(\begin{array}{c}\mathbf{p_b}\\ \mathbf{p_c}\end{array}\right)\middle|\exists\mathbf{h}\geq0:\left(\begin{array}{c}\mathbf{p}_\mathbf{l}^*\\ \mathbf{p_b}\\ \mathbf{p_c}\end{array}\right)=\left(\begin{array}{c}\mathbf{M_l}\\ \mathbf{M_b}\\ \mathbf{M_c}\end{array}\right)\mathbf{h}\right\}. \tag{33}$$

This polytope describes all possible couplings of all systems with low-marginals $\mathbf{p}_\mathbf{l}^*$. Then we do one of the two: either we fix $\mathbf{M_b h}=\mathbf{p}_\mathbf{b}^*$ and see how close $\mathbf{p_c}=\mathbf{M_c h}$ can be made to $\mathbf{p}_\mathbf{c}^*$ by changing $\mathbf{h}$; or we fix $\mathbf{M_c h}=\mathbf{p}_\mathbf{c}^*$ and see how close $\mathbf{p_b}=\mathbf{M_b h}$ can be made to $\mathbf{p}_\mathbf{b}^*$. These two procedures define two polytopes that we use to define CNT_1 and CNT_2, respectively.

Definition 4. If a system $\mathcal{R}$ represented by vector $\mathbf{p}^*$ is contextual,

$$\mathrm{CNT}_1=L_1\left(\mathbf{p}_\mathbf{c}^*,\mathbb{P}_\mathbf{c}\right) \tag{34}$$

the L_1-distance between $\mathbf{p}_\mathbf{c}^*$ and the *feasibility polytope*

$$\mathbb{P}_\mathbf{c}=\left\{\mathbf{p_c}\middle|\exists\mathbf{h}\geq0:\left(\begin{array}{c}\mathbf{p}_\mathbf{l}^*\\ \mathbf{p}_\mathbf{b}^*\\ \mathbf{p_c}\end{array}\right)=\left(\begin{array}{c}\mathbf{M_l}\\ \mathbf{M_b}\\ \mathbf{M_c}\end{array}\right)\mathbf{h}\right\}. \tag{35}$$

Written in extenso,

$$\mathrm{CNT}_1=\min_{\mathbf{p_c}\in\mathbb{P}_\mathbf{c}}\left\Vert\mathbf{p}_\mathbf{c}^*-\mathbf{p_c}\right\Vert_1=\mathbf{1}\cdot\mathbf{p}_\mathbf{c}^*-\max_{\mathbf{p_c}\in\mathbb{P}_\mathbf{c}}\left(\mathbf{1}\cdot\mathbf{p_c}\right). \tag{36}$$

Because $\mathbf{p_l}$ is fixed at $\mathbf{p}_\mathbf{l}^*$, the transformation ϕ in (28) has the form $4p^{i,i\ominus1}+\mathrm{const}$ for each component of $\mathbf{p_c}$, and we have

$$\left\Vert\mathbf{p}_\mathbf{c}^*-\mathbf{p_c}\right\Vert_1=\frac{\left\Vert\mathbf{e}_\mathbf{c}^*-\mathbf{e_c}\right\Vert_1}{4}. \tag{37}$$

This allows us to redefine the measure in the way more convenient for our purposes,

$$\mathrm{CNT}_1=\frac{1}{4}L_1\left(\mathbf{e}_\mathbf{c}^*,\mathbb{E}_\mathbf{c}\right), \tag{38}$$

where (pointwise)

$$\mathbb{E}_\mathbf{c}=\phi\left(\mathbb{P}_\mathbf{c}\right). \tag{39}$$

Definition 5. If a system $\mathcal{R}$ represented by vector $\mathbf{p}^*$ is contextual,

$$\mathrm{CNT}_2=L_1\left(\mathbf{p}_\mathbf{b}^*,\mathbb{P}_\mathbf{b}\right), \tag{40}$$

the L_1-distance between $\mathbf{p}_\mathbf{b}^*$ and the *noncontextuality polytope*

$$\mathbb{P}_\mathbf{b}=\left\{\mathbf{p_b}\middle|\exists\mathbf{h}\geq0:\left(\begin{array}{c}\mathbf{p}_\mathbf{l}^*\\ \mathbf{p_b}\\ \mathbf{p}_\mathbf{c}^*\end{array}\right)=\left(\begin{array}{c}\mathbf{M_l}\\ \mathbf{M_b}\\ \mathbf{M_c}\end{array}\right)\mathbf{h}\right\}. \tag{41}$$

Here,

$$\mathrm{CNT}_2=\min_{\mathbf{p_b}\in\mathbb{P}_\mathbf{b}}\left\Vert\mathbf{p}_\mathbf{b}^*-\mathbf{p_b}\right\Vert_1. \tag{42}$$

For the same reason as above, the transformation ϕ in (28) has the form $4p_{i,i\oplus1}+\mathrm{const}$ for each component of $\mathbf{p_b}$. We have therefore

$$\mathrm{CNT}_2=\frac{1}{4}L_1\left(\mathbf{e}_\mathbf{b}^*,\mathbb{E}_\mathbf{b}\right), \tag{43}$$

the L_1-distance between $\mathbf{e}_\mathbf{b}^*=\phi\left(\mathbf{p}_\mathbf{b}^*\right)$ and the polytope

$$\mathbb{E}_\mathbf{b}=\phi\left(\mathbb{P}_\mathbf{b}\right). \tag{44}$$

For convenience, we will use the same term, "feasibility polytope," for both $\mathbb{P}_\mathbf{c}$ and $\mathbb{E}_\mathbf{c}$. Analogously, both $\mathbb{P}_\mathbf{b}$ and $\mathbb{E}_\mathbf{b}$ can be referred to as "noncontextuality polytope."

As for any two ±1-random variables, we have

$$\left|e_i^i+e_{i\oplus1}^i\right|-1\leq e_{i,i\oplus1}\leq1-\left|e_i^i-e_{i\oplus1}^i\right|, i=1,\ldots,n. \tag{45}$$

Therefore the convex polytope $\mathbb{E}_\mathbf{b}$ is circumscribed by the n-box

$$\mathbb{R}_\mathbf{b}=\prod_{i=1}^{n}\left[\left|e_i^i+e_{i\oplus1}^i\right|-1,1-\left|e_i^i-e_{i\oplus1}^i\right|\right]. \tag{46}$$

We can analogously define the n-box circumscribing $\mathbb{E}_c$, but we do not need this notion.

The idea of the noncontextuality measure NCNT_2 extending CNT_2 to noncontextual systems is as follows.

Definition 6. If a system $\mathcal{R}$ represented by vector $\mathbf{p}^*$ is noncontextual,

$$\mathrm{NCNT}_2=L_1\left(\mathbf{p}_\mathbf{b}^*,\partial\mathbb{P}_\mathbf{b}\right)=\frac{1}{4}L_1\left(\mathbf{e}_\mathbf{b}^*,\partial\mathbb{E}_\mathbf{b}\right), \tag{47}$$

the L_1 distance between $\mathbf{p}_\mathbf{b}^*$ and the surface $\partial\mathbb{P}_\mathbf{b}$ of the noncontextuality polytope $\mathbb{P}_\mathbf{b}$.

Note that CNT_2, too, could be defined as the distance from a point to $\partial\mathbb{P}_{\mathbf{b}}$, so the definition is the same for both CNT_2 and NCNT_2, only the position of the $\mathbf{p}^*_{\mathbf{b}}$ changes from the outside to the inside of the polytope. In extenso,

$$\mathrm{NCNT}_2 = \frac{1}{4}\min_{\mathbf{e_b}\in\partial\mathbb{E}_{\mathbf{b}}} \|\mathbf{e}^*_{\mathbf{b}} - \mathbf{e_b}\|_1 = \frac{1}{4}\inf_{\mathbf{x}\in\mathbf{R}^n-\mathbb{E}_{\mathbf{b}}} \|\mathbf{e}^*_{\mathbf{b}} - \mathbf{x}\|_1\,, \tag{48}$$

where $\mathbf{R}$ is the set of reals. As shown in Ref. [17], no such extension to a noncontextuality measure exists for CNT_1 (see Sec. IX for the argument by which this is established).

VI. ADDITIONAL TERMINOLOGY AND CONVENTIONS

To focus now on CNT_2 and NCNT_2, we need a few additional terms and conventions. We confine our consideration to the space of all possible points $\mathbf{e_b}$, which is the n-cube

$$\mathbb{C}_{\mathbf{b}} = [-1,1]^n\,. \tag{49}$$

Given an arbitrary n-box

$$\mathbb{X} = \prod_{i=1}^{n}[\min x_i, \max x_i] \subseteq \mathbb{C}_{\mathbf{b}}, \tag{50}$$

a vertex V of $\mathbb{X}$ is called *odd* if its coordinates contain an odd number of $\min x_i$'s; otherwise the vertex is *even.* A hyperplane is said to be *pocket-forming* at vertex V if it cuts each of the n edges emanating from V, i.e., if it intersects each of them between V and the edge's other end. The region within $\mathbb{X}$ strictly above the pocket-forming hyperplane at V is called a *pocket* at V. This pocket is said to be *regular* if the pocket-forming hyperplane cuts all n edges emanating from V at an equal distance from V. We apply this terminology to two special n-boxes: the n-box $\mathbb{R}_{\mathbf{b}}$ circumscribing the noncontextuality polytope (46), and the ambient n-cube $\mathbb{C}_{\mathbf{b}}$ itself.

We will assume in the following that no context in the system contains a deterministic variable. If such a context exists, the n-box $\mathbb{R}_{\mathbf{b}}$ is degenerate (has lower dimensionality than n), and

$$\mathbb{E}_{\mathbf{b}} = \mathbb{R}_{\mathbf{b}},$$

making the system trivially noncontextual. Indeed, assume, e.g., that A_1^1 is a deterministic variable. We know that any deterministic variable can be removed from a system without affecting its (non)contextuality [37]. The system therefore can be presented as a noncyclic chain

$$A_2^1, A_2^2, A_3^2, \ldots, A_n^n, A_1^n.$$

Whatever the joint distributions of adjacent pairs in such a chain, there is always a global joint distribution that agrees with these pairwise distributions as its marginals: for any assignment of values to the links of the chain, the coupling probability is obtained as the product of the chained conditional probabilities.

A cyclic system $\mathcal{A}$ is called a *variant* of a cyclic system $\mathcal{B}$ of the same rank if

$$\left\{A_i^i, A_i^{i\ominus 1}\right\} = \pm 1 \cdot \left\{B_i^i, B_i^{i\ominus 1}\right\}, \tag{51}$$

for $i = 1, \ldots, n$.

Lemma 7 (Kujala-Dzhafarov [19])**.** *All variants of a system have the same values of $s_1(\mathbf{e_b})$ and $\left|e_i^i - e_i^{i\ominus 1}\right|$, $i = 1, \ldots, n$ (hence also they have the same value of Δ).*

Lemma 8 (Kujala-Dzhafarov [19])**.** *Among the 2^n variants of a cyclic system there is one, called canonical, in which (following a circular permutation of indices)*

$$|e_{n1}| \le e_{i,i+1}, i = 1, \ldots, n-1. \tag{52}$$

Clearly, a canonical variant of a system is a canonical variant of any variant of the system, including itself. In a canonical variant of a system,

$$s_1(\mathbf{e_b}) = \sum_{i=1}^{n-1} e_{i,i+1} - e_{n1}. \tag{53}$$

VII. PROPERTIES OF THE NONCONTEXTUALITY POLYTOPE

In this section we present a series of lemmas establishing the remarkably simple structure of the noncontextuality polytope. The proofs of these results are relegated to the Appendix.

Lemma 9. *For each odd vertex $V = \{\lambda_i : i = 1, \ldots, n\}$ of $\mathbb{C}_{\mathbf{b}}$, the inequality $\sum \lambda_i e_{i,i\oplus 1} > \Delta$ describes a regular pocket at V. The distance at which the hyperplane segment $\sum \lambda_i e_{i,i\oplus 1} = \Delta$ cuts each of the edges of the cube emanating from V is $n - \Delta$. (See Fig. 4.)*

Lemma 10. *For a given Δ, no two pockets $\sum \lambda_i e_{i,i\oplus 1} > \Delta$ and $\sum \lambda'_i e_{i,i\oplus 1} > \Delta$ formed by the hyperplanes at different odd vertices of $\mathbb{C}_{\mathbf{b}}$ intersect. The pocket-forming hyperplanes at the odd vertices are also disjoint within $\mathbb{C}_{\mathbf{b}}$ unless $\Delta = n-2$. (See Figs. 5 and 6.)*

Lemma 11. *If a point $\mathbf{x}$ is within the pocket formed at an odd vertex $V = \{\lambda_i : i = 1, \ldots, n\}$* of *$\mathbb{C}_{\mathbf{b}}$ by a hyperplane segment $\sum \lambda_i e_{i,i\oplus 1} = \Delta$, then*

$$s_1(\mathbf{x}) = \sum \lambda_i x_{i,i\oplus 1} = \Delta_{\mathbf{x}} > \Delta,$$

and $s_1(\mathbf{x}) - \Delta$ is the distance between the points at which the two hyperplane segments $\sum \lambda_i e_{i,i\oplus 1} = \Delta_{\mathbf{x}}$ and $\sum \lambda_i e_{i,i\oplus 1} = \Delta$ cut any of the edges emanating from V. (See Fig. 7.)

The *extended noncontextuality polytope* $\mathbb{N}_{\mathbf{b}} \subseteq \mathbb{C}_{\mathbf{b}}$ is defined by 2^{n-1} half-space inequalities

$$\sum_{i=1}^{n} \lambda_i e_{i,i\oplus 1} \le \Delta, i = 1, \ldots, n, \tag{54}$$

where $n-2 \le \Delta \le n$ and $\{\lambda : i = 1, \ldots, n\}$ are odd vertices of $\mathbb{C}_{\mathbf{b}}$. Therefore we can identify $\mathbb{N}_{\mathbf{b}}$ by the value of Δ, and write $\mathbb{N}_{\mathbf{b}} = \mathbb{N}_{\mathbf{b}}(\Delta)$. (See Fig. 8.)

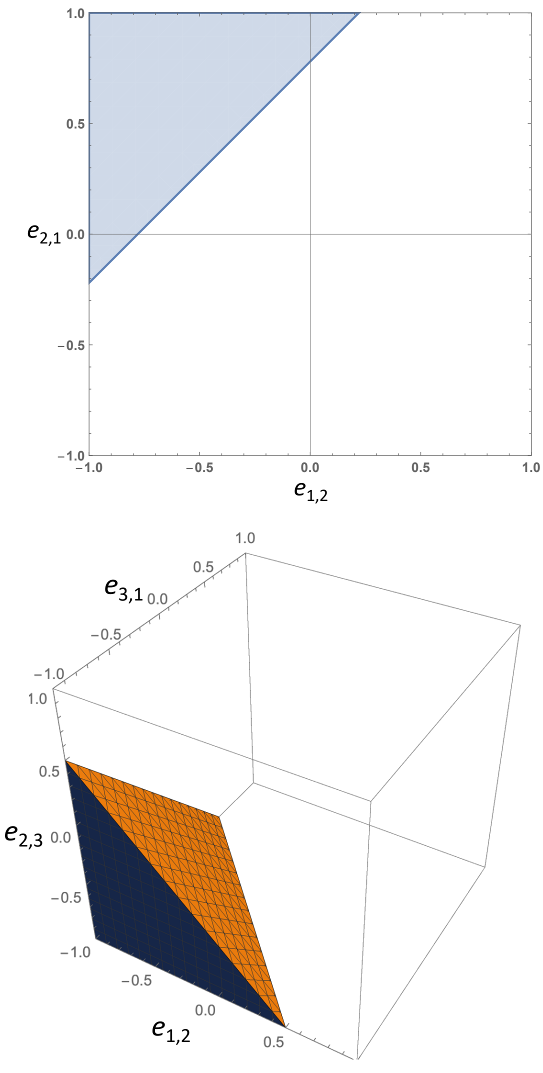

Figure 4. Illustration for Lemma 9, $n = 2$ and $n = 3$.

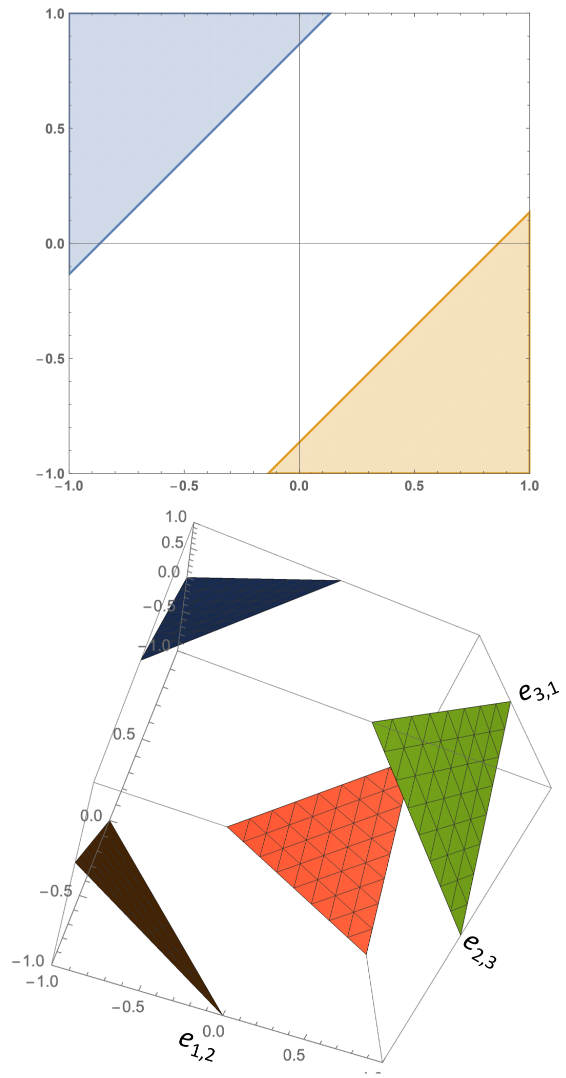

Figure 5. Illustration for Lemma 10, $n = 2$ and $n = 3$. The pocket-forming hyperplanes at different odd vertices of $\mathbb{C}_{\mathbf{b}}$ do not touch within $\mathbb{C}_{\mathbf{b}}$ when $\Delta > n - 2$.

Lemma 12. *If a point* $\mathbf{x}$ *is within the extended noncontextuality polytope* $\mathbb{N}_{\mathbf{b}}(\Delta)$*, then*

$$s_1(\mathbf{x}) = \sum \lambda_i x_{i,i\oplus 1} = \Delta_{\mathbf{x}} \leq \Delta,$$

where $V = \{\lambda : i = 1, \ldots, n\}$ *is an odd vertex of* $\mathbb{C}_{\mathbf{b}}$ *(unique if* $\Delta > n - 2$*) at which the hyperplane segment* $\sum \lambda_i e_{i,i\oplus 1} = \Delta_{\mathbf{x}}$ *forms a pocket. The difference* $\Delta - \Delta_{\mathbf{x}}$ *is the distance between the points at which the two hyperplanes cut any of the edges emanating from* V*. (See Fig. 9.)*

The proof of Lemma 12 is obvious, in view of the previous results.

We know that $\mathbb{E}_{\mathbf{b}}$ is the intersection of $\mathbb{R}_{\mathbf{b}}$ and the polytope $\mathbb{N}_{\mathbf{b}}(\Delta)$. The following lemma stipulates an important property of this intersection.

Lemma 13. *All even vertices of* $\mathbb{R}_{\mathbf{b}}$ *are within* $\mathbb{E}_{\mathbf{b}}$*. (See Fig. 10.)*

Corollary 14. *A point* $\mathbf{x}$ *represents a contextual system if and only if it belongs to a pocket formed by a pocket-forming hyperplane segment* $\sum \lambda_i e_{i,i\oplus 1} = \Delta$ *at an odd vertex* $V = \{\lambda_i : i = 1, \ldots, n\}$ *of* $\mathbb{R}_{\mathbf{b}}$*. These pockets are regular and their number is* $0 \leq k \leq 2^{n-1}$*. (See Fig. 11.)*

VIII. MAIN THEOREMS

The following two theorems now are simple corollaries of the previous results. Consider a noncontextuality polytope

$$\mathbb{E}_{\mathbf{b}} = \mathbb{R}_{\mathbf{b}} \cap \mathbb{N}_{\mathbf{b}}(\Delta). \tag{55}$$

Theorem 15. *The* L_1*-distance between* $\mathbb{E}_{\mathbf{b}}$ *and a point* $\mathbf{e}^*_{\mathbf{b}}$ *representing a contextual system is a single-coordinate distance, equal to* $s_1(\mathbf{e}^*_{\mathbf{b}}) - \Delta$ *for all coordinates. This is the value of* CNT_2*. (See Fig. 12.)*

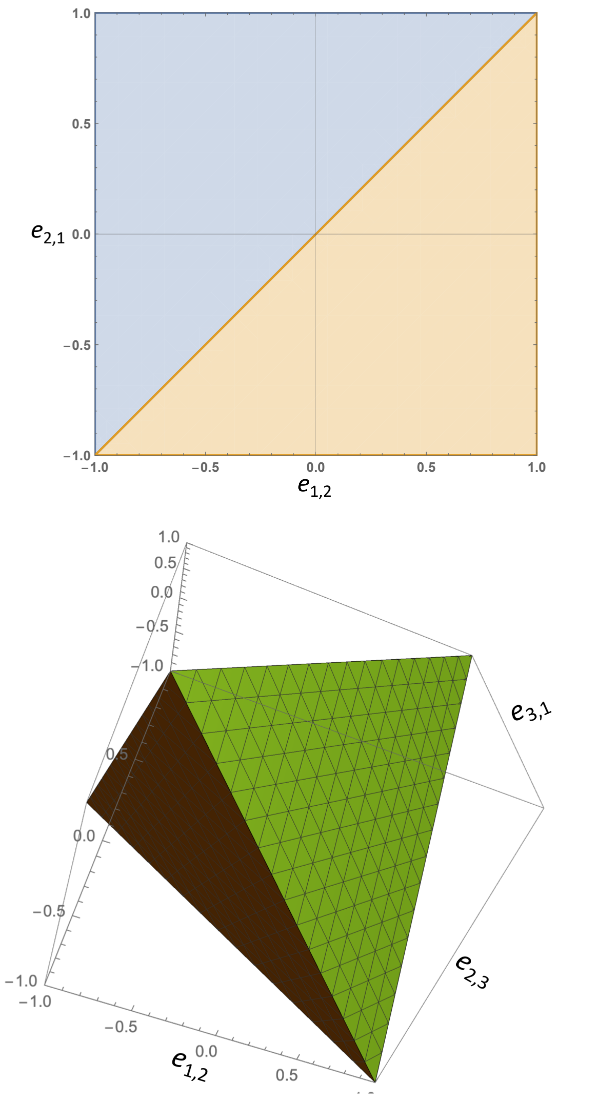

Figure 6. Illustration for Lemma 10 (continued), $n = 2$ and $n = 3$. When $\Delta = n - 2$, the pocket-forming hyperplanes at different odd vertices of $\mathbb{C}_{\mathbf{b}}$ form an n-demicube, the convex hull of the 2^{n-1} even vertices of $\mathbb{C}_{\mathbf{b}}$.

It is easy to show that for any $p \geq 1$, the L_p-distance between $\mathbf{e}^*_{\mathbf{b}}$ and $\mathbb{E}_{\mathbf{b}}$ (let us call it $\mathrm{CNT}_2^{(p)}$) is simply

$$\mathrm{CNT}_2^{(p)} = n^{\frac{1-p}{p}} \mathrm{CNT}_2, \tag{56}$$

where n is the rank of the cyclic system. This means that in the case of contextual cyclic systems L_1-distance can be, if one so wishes, replaced by any L_p-distance with no nontrivial changes in the theory. However, this may not be possible for noncyclic systems, where the faces of the noncontextuality polytope need not have the simple structure of $\mathbb{E}_{\mathbf{b}}$.

Let us define a new measure now, the L_1-distance between the box $\mathbb{R}_{\mathbf{b}}$ and a point $\mathbf{e}_{\mathbf{b}}$ within the box:

$$m\left(\mathbf{e}_{\mathbf{b}}\right) = \min_{i=1,\dots,n} \left(\min \left(\begin{array}{c} e_{i,i\oplus 1} - \left|e_i + e_{i\oplus 1}\right| + 1, \\ 1 - \left|e_i^i - e_{i\oplus 1}^i\right| - e_{i,i\oplus 1} \end{array} \right) \right). \tag{57}$$

Theorem 16. *The L_1-distance between the surface of $\mathbb{E}_{\mathbf{b}}$ and a point $\mathbf{e}^*_{\mathbf{b}}$ representing a noncontextual system is a single-coordinate distance, equal to* $\min\left(\Delta - s_1\left(\mathbf{e}^*_{\mathbf{b}}\right), m\left(\mathbf{e}^*_{\mathbf{b}}\right)\right)$. *This is the value of* NCNT_2. *If this value equals $s_1\left(\mathbf{e}^*\right) - \Delta$, it is the same for all coordinates. (See Fig. 13.)*

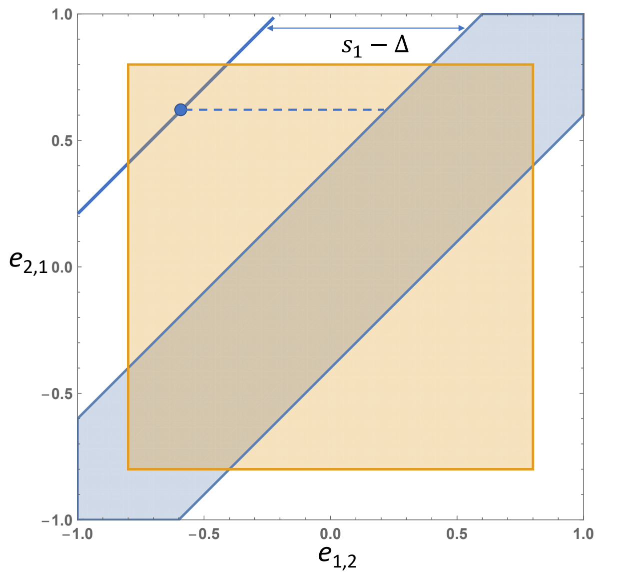

Figure 7. Illustration for Lemma 11, $n = 2$, and for subsequent development. The hyperplane segment $\sum \lambda_i e_{i,i\oplus 1} = \Delta$ is shown as the left boundary of the extended noncontextuality polytope (here, hexagon) $\mathbb{N}_{\mathbf{b}}$. The smaller internal rectangle represents $\mathbb{R}_{\mathbf{b}}$.

By geometric considerations, $m\left(\mathbf{e}_{\mathbf{b}}\right)$ is also an L_p-distance between $\mathbb{R}_{\mathbf{b}}$ and $\mathbf{e}^*_{\mathbf{b}}$, for any $p \geq 1$. Because of this, using the same reasoning as in the case of $\mathrm{CNT}_2^{(p)}$, the L_p-distance from $\mathbf{e}^*_{\mathbf{b}}$ to the surface of $\mathbb{E}_{\mathbf{b}}$ is

$$\mathrm{NCNT}_2^{(p)} = \min\left(n^{\frac{1-p}{p}}\left(\Delta - s_1\left(\mathbf{e}^*_{\mathbf{b}}\right)\right), m\left(\mathbf{e}^*_{\mathbf{b}}\right)\right). \tag{58}$$

As we see, unlike in the case of $\mathrm{CNT}_2^{(p)}$, this is not simply a scaled version of NCNT_2, indicating that replacing the latter with $\mathrm{NCNT}_2^{(p)}$ is not inconsequential for the theory.

Figures 14 and 15 illustrate the dynamics of CNT_2 and NCNT_2 as point $\mathbf{e}^*_{\mathbf{b}}$ moves along the diagonal connecting two opposite vertices of $\mathbb{R}_{\mathbf{b}}$ for cyclic systems of several ranks. To emphasize that NCNT_2 is an extension of CNT_2 (and vice versa), we plot NCNT_2 with minus sign: as $\mathbf{e}^*_{\mathbf{b}}$ moves closer to the surface of $\mathbb{E}_{\mathbf{b}}$, CNT_2 decreases from a positive value to zero, the system becomes noncontextual, and as the point continues to move inside the polytope, the value of $-\mathrm{NCNT}_2$ proceeds to decrease continuously.

IX. POLYTOPE OF ALL POSSIBLE COUPLINGS

We now need to gain insight into why CNT_1 and CNT_2 are the same for cyclic systems. Is it a peculiar coincidence? Does CNT_1, if interpreted geometrically, have the same "nice" properties as CNT_2? The answer to the first question turns out to be negative, and to second one affirmative.

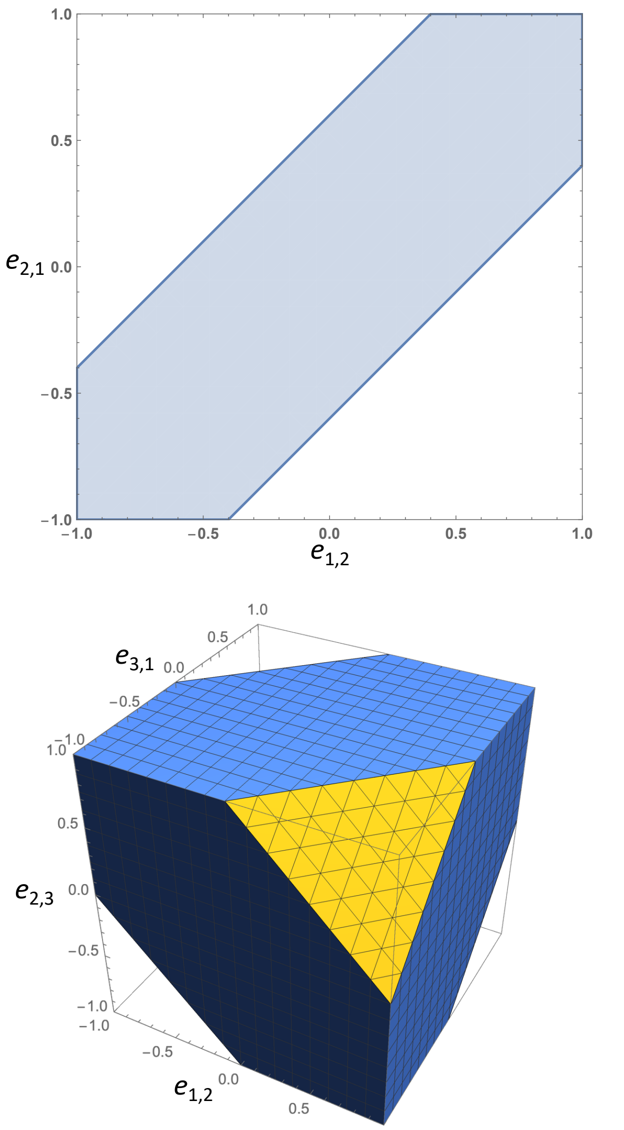

Figure 8. Examples of polytope $\mathbb{N}_{\mathbf{b}}$ within cube $\mathbb{C}_{\mathbf{b}}$, for $n = 2$ and $n = 3$.

In Sec. V we introduced in (33) the polytope $\mathbb{P}$ of all possible couplings for a system with the low-marginals $\mathbf{p}_{\mathbf{l}}^{*}$. As in the cases of $\mathbb{P}_{\mathbf{b}}$ and $\mathbb{P}_{\mathbf{c}}$, we redefine this polytope in terms of ± 1-variables,

$$\mathbb{E} = \phi\left(\mathbb{P}\right), \tag{59}$$

and use it to define a measure of contextuality

$$\mathrm{CNT}_0 = L_1\left(\left(\begin{pmatrix}\mathbf{p}_{\mathbf{b}}^{*}\\ \mathbf{p}_{\mathbf{c}}^{*}\end{pmatrix}, \mathbb{P}\right)\right) = \frac{1}{4}L_1\left(\left(\begin{pmatrix}\mathbf{e}_{\mathbf{b}}^{*}\\ \mathbf{e}_{\mathbf{c}}^{*}\end{pmatrix}, \mathbb{E}\right)\right). \tag{60}$$

To investigate the properties of $\mathbb{E}$ and CNT_0 we use the following result:

Theorem 17 (Kujala-Dzhafarov-Larsson [2]). *A system represented by* $\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right)^{\intercal}$ *is noncontextual if and only if*

$$s_1\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right) \leq 2n - 2.$$

This can be understood as a special case of Theorem 3 if one uses the procedure of treating connections as if they were additional contexts, rendering thereby any system consistently connected [17, 38]. Here and in the following we write $s_1\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right)$ instead of the more correct $s_1\left(\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right)^{\intercal}\right)$.

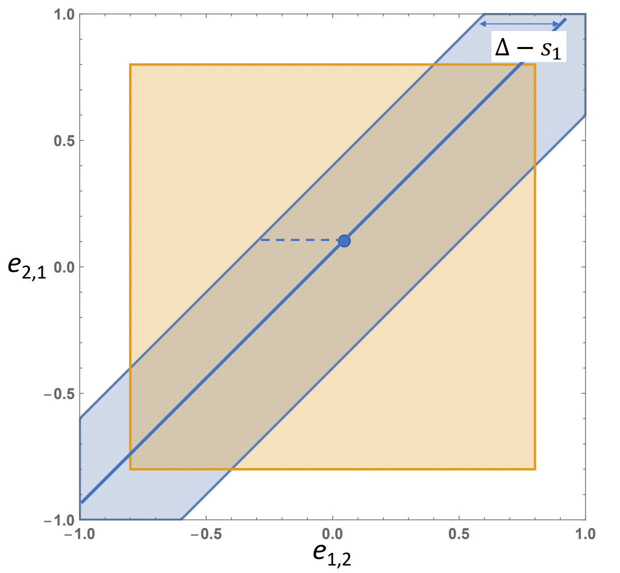

Figure 9. Illustration for Lemma 12, $n = 2$. The polytope and the box $\mathbb{R}_{\mathbf{b}}$ are as in Fig. 7.

It is evident now that the entire development in Secs. VI and V can be repeated with $\mathbb{E}$ replacing $\mathbb{E}_{\mathbf{b}}$, except that the ambient cube $\mathbb{C}$, extended noncontextuality polytope $\mathbb{N}$, and the box $\mathbb{R}$ circumscribing $\mathbb{E}$ (replacing, respectively, $\mathbb{C}_{\mathbf{b}}$, $\mathbb{N}_{\mathbf{b}}$, and $\mathbb{R}_{\mathbf{b}}$) are $2n$-dimensional rather than n-dimensional, and the value of Δ that defines the polytope is $2n-2$. In particular, the shape of the polytope $\mathbb{N}$ is always a $2n$-demicube, the convex hull of the 2^{2n-1} even vertices of $\mathbb{C}$, similar to the n-demicubes shown in Fig. 6, except that the minimal meaningful number of dimensions has to be 4 (representing a cyclic system of rank 2). The following analog of Theorem 15 then holds.

Theorem 18. *The* L_1*-distance between* $\mathbb{E}$ *and a point* $\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right)^{\intercal}$ *representing a contextual system is a single-coordinate distance, equal to* $s_1\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right) - (2n-2)$ *for all coordinates. This is the value of* CNT_0.

It is easy to see now that

$$\mathrm{CNT}_0 = \mathrm{CNT}_1 = \mathrm{CNT}_2. \tag{61}$$

Indeed, the single-coordinate L_1-distance mentioned in the theorem can be taken along an $\mathbf{e}_{\mathbf{b}}$-coordinate or along an $\mathbf{e}_{\mathbf{c}}$-coordinate, and with all other coordinates being fixed at appropriate values, this will be a single-coordinate L_1-distance from, respectively, $\mathbb{E}_{\mathbf{b}}$ or $\mathbb{E}_{\mathbf{c}}$. Since we know that

$$\mathrm{CNT}_1 = \mathrm{CNT}_2 = s_1\left(\mathbf{e}_{\mathbf{b}}^{*}\right) - \Delta, \tag{62}$$

and that

$$\mathrm{CNT}_0 = s_1\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right) - (2n-2), \tag{63}$$

we have an indirect proof that when $s_1\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right) > (2n-2)$ (i.e., the system is contextual),

$$s_1\left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right) = s_1\left(\mathbf{e}_{\mathbf{b}}^{*}\right) + n - \delta. \tag{64}$$

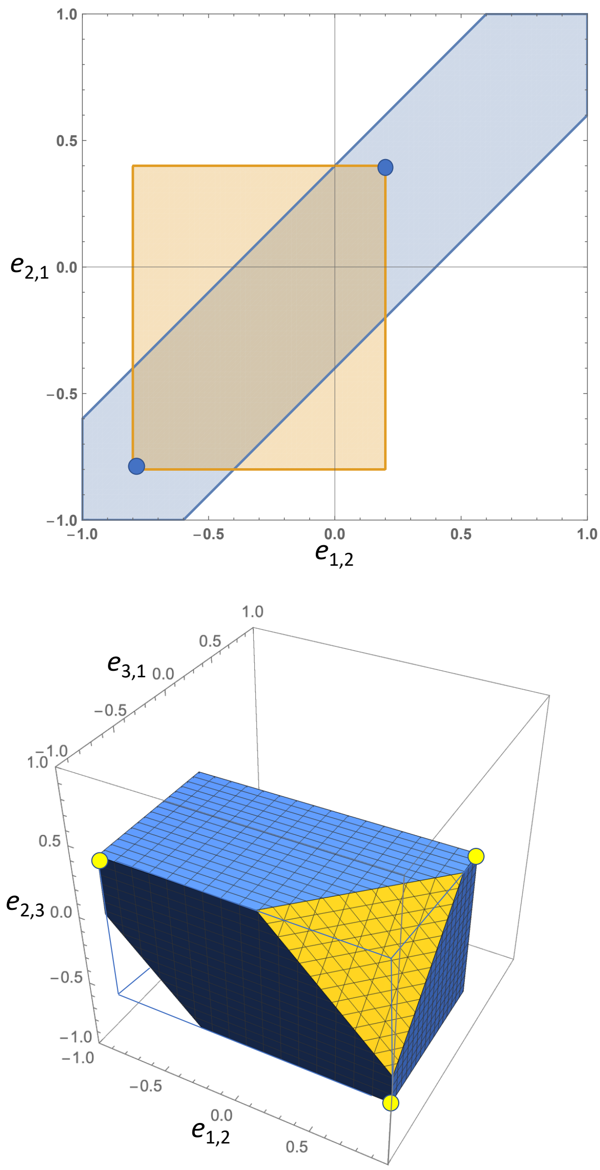

Figure 10. Illustration for Lemma 13, $n = 2$ and $n = 3$: even vertices are shown by small circles.

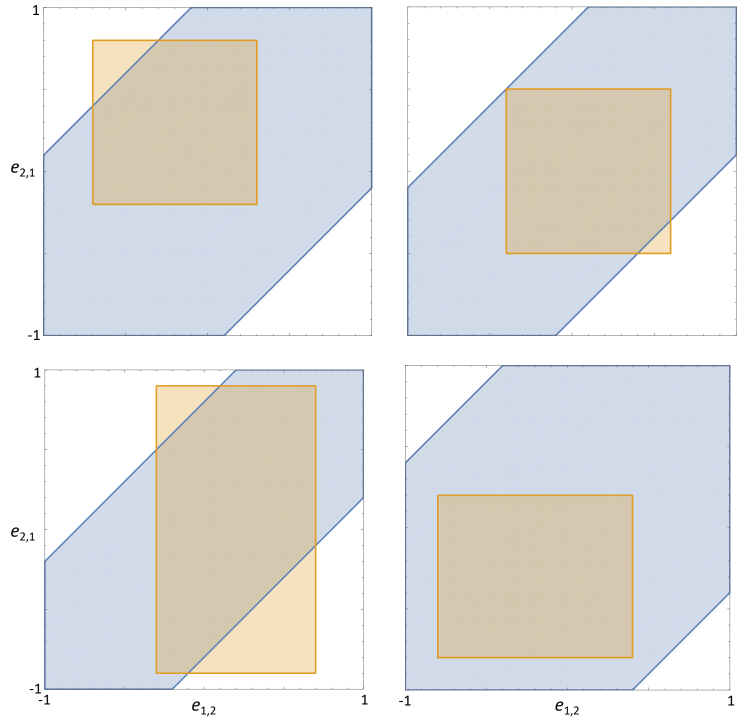

Figure 11. Illustration for Corollary 14, $n = 2$. The number of regular pockets formed at odd vertices of box $\mathbb{R}_{\mathbf{b}}$ can be 1 (upper panels), 2 (left lower), or zero (right lower).

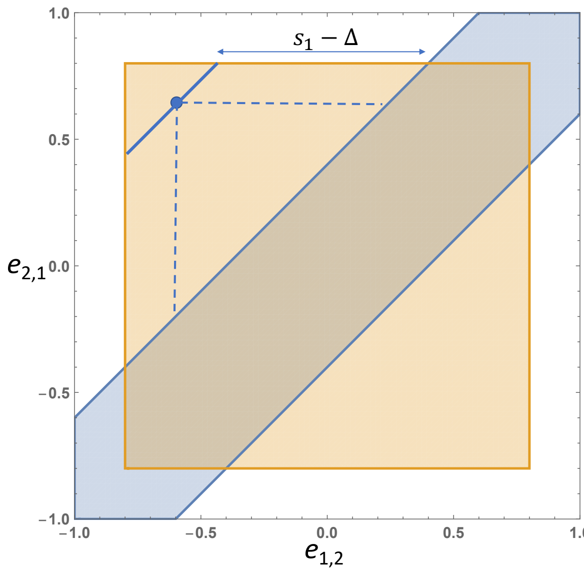

Figure 12. Illustration for Theorem 15, $n = 2$, a detailed analog of Fig. 1A.

Note that CNT_0, like CNT_1 and unlike CNT_2, cannot be naturally extended to a noncontextuality measure. Because $\mathbf{e}_{\mathbf{c}}^{*}$ consists of the maximal possible values of $e^{i,i\oplus 1}$ ($i = 1, \ldots, n$), any point $(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*})^{\intercal}$ representing a noncontextual system should lie on the surface of the polytope $\mathbb{E}$, yielding

$$s_1 \left(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*}\right) - (2n - 2) = 0. \tag{65}$$

The argument leading to this conclusion was presented in Ref. [17] for $\mathbb{E}_{\mathbf{c}}$. When applied to $\mathbb{E}$, it goes as follows: if $(\mathbf{e}_{\mathbf{b}}^{*}, \mathbf{e}_{\mathbf{c}}^{*})^{\intercal}$ were an interior point of $\mathbb{E}$, it would be surrounded by a $2n$-ball entirely within $\mathbb{E}$, and one would be able to increase any component of $\mathbf{e}_{\mathbf{c}}^{*}$ while remaining within this ball, which is not possible. CNT_2 remains the only one of the contextuality measures considered in the literature that can be naturally extended into a noncontextuality measure.

X. CONCLUSION WITH A GLIMPSE INTO NONCYCLIC SYSTEMS

Most of the regularities established in this paper do not generalize to noncyclic systems. In particular, CNT_1 and CNT_2 do not generally coincide, nor is one of them any function of the other. This can be seen in Fig. 16 that presents the values of CNT_1 and CNT_2 for several systems obtained by modifying the noncyclic PR3 box system described in Ref [39]. The PR3 box system is

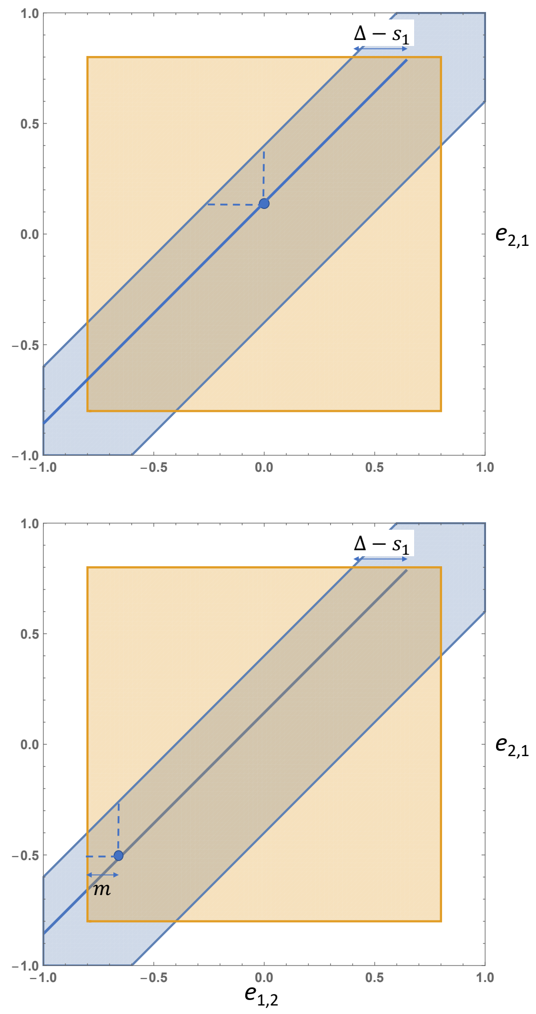

Figure 13. Illustration for Theorem 16, $n = 2$, a detailed analog of Figs. 1B and 1C. Upper panel: the case $\Delta - s_1(\mathbf{e}_\mathfrak{b}^*) \le m(\mathbf{e}_\mathfrak{b}^*)$. Lower panel: the case $\Delta - s_1(\mathbf{e}_\mathfrak{b}^*) > m(\mathbf{e}_\mathfrak{b}^*)$.

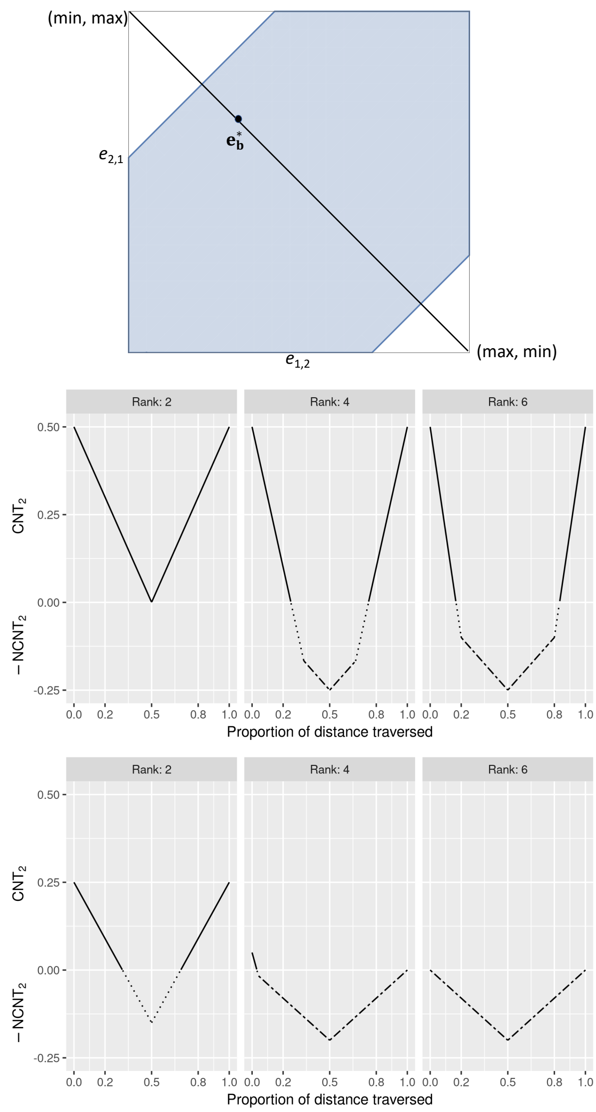

Figure 14. CNT$_2$ (solid lines, positive values) and $-$NCNT$_2$ (dotted or dashed lines, nonpositive values) as a function of the position of $\mathbf{e}_\mathfrak{b}^*$ on the diagonal connecting two opposite vertices of $\mathbb{R}_\mathfrak{b}$ for cyclic systems of ranks 2, 4, 6. The dashed lines show the cases $\Delta - s_1(\mathbf{e}_\mathfrak{b}^*) > m(\mathbf{e}_\mathfrak{b}^*)$, when NCNT$_2 = m(\mathbf{e}_\mathfrak{b}^*)$; the dotted lines show the case $\Delta - s_1(\mathbf{e}_\mathfrak{b}^*) \le m(\mathbf{e}_\mathfrak{b}^*)$, when NCNT$_2 = \Delta - s_1(\mathbf{e}_\mathfrak{b}^*)$. The lower set of graphs represents inconsistently connected systems, with $\langle A_i^i \rangle = -0.2, \langle A_i^{i\ominus 1} \rangle = 0.1$, for $i = 1, \ldots, n$. The upper graphs represent consistently connected systems, with the expectations of all random variables equal to zero. As indicated in the illustration on the top, for the even-ranked systems, it connects the vertex whose last coordinate is its single min-coordinate to the vertex whose last coordinate is its single max-coordinate.

given by

$$
\begin{array}{|c|c|c|c|c|c||c}
\hline
A_1^1 & A_2^1 & & & & & c_1 \\ \hline
 & A_2^2 & A_3^2 & & & & c_2 \\ \hline
 & & A_3^3 & A_4^3 & & & c_3 \\ \hline
 & & & A_4^4 & A_5^4 & & c_4 \\ \hline
 & & & & A_5^5 & A_6^5 & c_5 \\ \hline
A_1^6 & & & A_4^6 & & & c_6 \\ \hline
 & A_2^7 & & & A_5^7 & & c_7 \\ \hline
 & & A_3^8 & & & A_6^8 & c_8 \\ \hline
A_1^9 & & & & & A_6^9 & c_9 \\ \hline\hline
q_1 & q_2 & q_3 & q_4 & q_5 & q_6 &
\end{array}
\tag{66}
$$

where A_i^k are ± 1-random variables with $e_i^k = 0$ for all $i = 1, \ldots, 6, k = 1, \ldots, 9$. In the original system $e_{23} = e_{45} = -1$, and $e_{ij} = 1$ in all other contexts. We have looked at the changes in the values of CNT$_1$ and CNT$_2$ in response to two ways of modifying these parameters, as described in the legend of Fig. 16. For each combination of the parameters, CNT$_1$ and CNT$_2$ were computed by means of linear programming [17], provided the corresponding system was contextual. We see that in none of these cases CNT$_1$ and CNT$_2$ were equal to each other. Moreover, we can see that no functional relation between the two is satisfied either: for either of the measures, there are pairs of systems with different values of this measure at a fixed value of the other.

It might be tempting to think that cyclic systems could help one in at least detecting if not measuring (non)contextuality of a system. Clearly, if a system contains a contextual cyclic subsystem, then it is contextual.

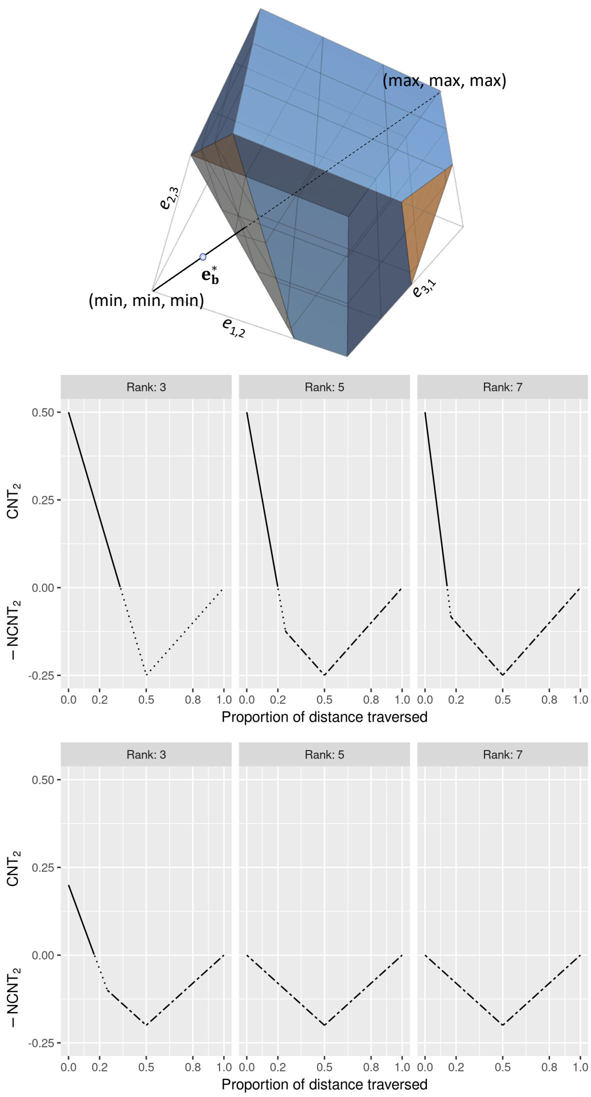

Figure 15. The same as Fig. 14, but for cyclic systems of ranks 3, 5, 7. As indicated in the illustration on the top, for the odd-ranked systems, the diagonal connects the vertex with all min-coordinates to the vertex with all max-coordinates.

This is not surprising, however, because this is true for any contextual subsystem, cyclic or not [15, 16]. Could it be, one might wonder, that a system is always noncontextual if it does not contain a contextual cyclic subsystem? The answer is negative, as we see from the following counterexample. Let a system of dichotomous random variables be

$$\begin{array}{|c|c|c|c||c}
\hline
R_1^1 & R_2^1 & R_3^1 & & c_1 \\
\hline
 & R_2^2 & R_3^2 & R_4^2 & c_2 \\
\hline
R_1^3 & & R_3^3 & R_4^3 & c_3 \\
\hline
q_1 & q_2 & q_3 & q_4 &
\end{array}, \tag{67}$$

with four contents measured in three contexts. Let the joint distributions of the three bunches be

$$\begin{array}{|c|c|c|c|}
\hline
R_1^1 & R_2^1 & R_3^1 & \\
\hline
-1 & -1 & +1 & 1/4 \\
\hline
-1 & +1 & -1 & 1/4 \\
\hline
+1 & -1 & -1 & 1/4 \\
\hline
+1 & +1 & +1 & 1/4 \\
\hline
\end{array},
\begin{array}{|c|c|c|c|}
\hline
R_2^2 & R_3^2 & R_4^2 & \\
\hline
-1 & -1 & +1 & 1/4 \\
\hline
-1 & +1 & -1 & 1/4 \\
\hline
+1 & -1 & -1 & 1/4 \\
\hline
+1 & +1 & +1 & 1/4 \\
\hline
\end{array},
\begin{array}{|c|c|c|c|}
\hline
R_1^3 & R_3^3 & R_4^3 & \\
\hline
+1 & +1 & -1 & 1/4 \\
\hline
+1 & -1 & +1 & 1/4 \\
\hline
-1 & +1 & +1 & 1/4 \\
\hline
-1 & -1 & -1 & 1/4 \\
\hline
\end{array}. \tag{68}$$

Here, the probabilities of the triples of values in each bunch are shown in the rightmost columns, with all remaining triples having probability zero. One can check that all random variables are distributed uniformly,

$$\Pr\left[R_i^k = -1\right] = \Pr\left[R_i^k = +1\right] = \frac{1}{2}, \tag{69}$$

so the system is consistently connected. All pairs $\left(R_i^k, R_j^k\right)$ are also uniformly distributed,

$$\begin{array}{|c|c|c|}
\hline
R_i^k & R_j^k & \\
\hline
-1 & -1 & 1/4 \\
\hline
-1 & +1 & 1/4 \\
\hline
+1 & -1 & 1/4 \\
\hline
+1 & +1 & 1/4 \\
\hline
\end{array}. \tag{70}$$

This means that the system is *strongly consistently connected*: whenever a set of contents is measured in two contexts, their joint (here, pairwise) distributions coincide. Because the variables in each bunch are pairwise independent, any cyclic subsystem of this system is noncontextual. The entire system, however is contextual. Indeed,

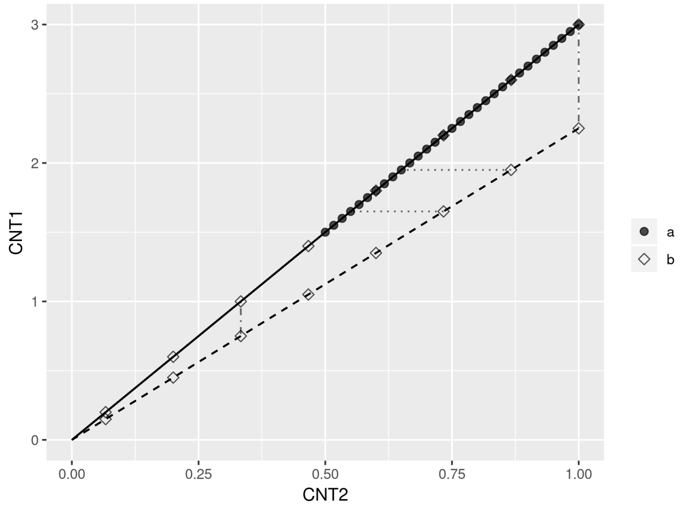

Figure 16. CNT_1 vs CNT_2 for the modifications of the PR3 box system described in the text. The symbols represent the contextual systems for which both CNT_1 and CNT_2 were computed by means of linear programming (for noncontextual systems, $\mathrm{CNT}_1 = \mathrm{CNT}_2 = 0$). Circles and diamonds represent two ways of modifying the original system: (a) increasing e_{23} from -1 to 1 by $1/15$ steps, with all other parameters remaining as in the original PR3; (b) changing $e_{23} = e_{45} = -e_{ij}$ (for all (i, j) other than $(2, 3)$ and $(4, 5)$) from -1 to 1 by $1/15$ steps, with all e_i^k-expectations fixed at zero. The symbols are aligned along two straight lines, with slopes 2.25 and 3. The horizontal and vertical segments between these lines represent some pairs of observed systems in which one of the measures was constant while the other varied.

in the hypothetical coupling satisfying the definition of noncontextuality, if $\left(S_1^1, S_2^1, S_3^1\right) = (-1,-1,1)$, then $\left(S_2^2, S_3^2, S_4^2\right)$ can only be $(-1,1,-1)$, and $\left(S_1^3, S_3^3, S_4^3\right)$ can only be $(-1,1,1)$. The reason for this is that in this hypothetical coupling we should have $\left(S_1^1, S_3^1\right) = \left(S_1^3, S_3^3\right)$, and $\left(S_2^1, S_3^1\right) = \left(S_2^2, S_3^2\right)$. However it should also be true that $\left(S_3^2, S_4^2\right) = \left(S_3^3, S_4^3\right)$, and this is not the case in the above triples: $\left(S_3^2, S_4^2\right) = (1,-1)$ while $\left(S_3^3, S_4^3\right) = (1,1)$. This completes the counterexample.

To summarize, we know now that the regular way in which the noncontextuality polytope $\mathbb{P}_{\mathbf{b}}$ (or $\mathbb{E}_{\mathbf{b}}$) and the polytope of all possible couplings $\mathbb{P}$ (or $\mathbb{E}$) create pockets at the vertices of the circumscribing boxes makes CNT_1 and CNT_2 single-coordinate distances that are equal to each other. Both of them are proportional to the degree of violation of the generalized Bell criterion derived in Ref. [19], $s_1\left(\mathbf{e}_{\mathbf{b}}^*\right) - \Delta$. We have known from Ref. [17], that CNT_2, unlike CNT_1, naturally extends to a measure of noncontextuality, NCNT_2, and this can be taken as a reason for preferring CNT_2 to CNT_1. NCNT_2 is a single-coordinate distance, and in the case of cyclic systems, the properties of the noncontextuality polytope make NCNT_2 proportional, with the same proportionality coefficient as for CNT_2 and CNT_1, to the smaller of two quantities: the degree of compliance with the generalized Bell inequality, $\Delta - s_1\left(\mathbf{e}_{\mathbf{b}}^*\right)$, and the distance $m\left(\mathbf{e}_{\mathbf{b}}^*\right)$ of $\mathbf{e}_{\mathbf{b}}^*$ from the surface of the circumscribing box $\mathbb{R}_{\mathbf{b}}$. We also know that none of these regularities extend beyond the class of cyclic systems, so the general theory of the relationship between the measures considered in this paper has much left to develop.

APPENDIX: PROOFS OF THE FORMAL STATEMENTS

Proof of Lemma 9. Verify that, for any Δ, each of the n points

$$\mathbf{x}_k = \left\{\lambda_1, \ldots, \lambda_k\left(1-n+\Delta\right), \ldots, \lambda_n\right\}, k=1,\ldots,n,$$

satisfies

$$\sum \lambda_i x_{i,i\oplus 1} = n-1+\lambda_k^2\left(1-n+\Delta\right) = \Delta,$$

whence so does the hyperplane passing through these points. Since $n-2 \le \Delta \le n$, the distance $n-\Delta$ is between 0 and 2, so that the hyperplane does cut each of the edges joined at the vertex.

Proof of Lemma 10. Two odd vertices have nonoverlapping sets of edges emanating from them, and each of the two hyperplanes cuts its own set. The only case when an axis from one set is cut at the same point as an axis from another set is when the cuts are at the ends of the emanating edges, and this means that $\Delta = n-2$.

Proof of Lemma 11. We need to show that for any other odd vertex $V' = \left\{\lambda_i' : i=1,\ldots,n\right\}$ of $\mathbb{C}_{\mathbf{b}}$, $\sum \lambda' x_{i,i\oplus 1} \le \Delta_{\mathbf{x}}$. This is indeed the case because for any value $\Delta' \ge n-2$, the hyperplane segment $\sum \lambda' e_{i,i\oplus 1} = \Delta'$ does not cut any of the edges emanating from vertex V, except, possibly, at their other ends (if $\Delta' = n-2$). Consequently, $\sum \lambda' x_{i,i\oplus 1} < n-2 \le \Delta < \Delta_{\mathbf{x}}$.

Proof of Lemma 13. By induction. For $n=2$ we have to show that the even vertices $\left(1-\left|e_1^1-e_2^1\right|, 1-\left|e_1^2-e_2^2\right|\right)$ and $\left(\left|e_1^1+e_2^1\right|-1, \left|e_1^2+e_2^2\right|-1\right)$ are within $\mathbb{E}_{\mathbf{b}}$. For the former vertex this means that

$$\left|\left(1-\left|e_1^1-e_2^1\right|\right)-\left(1-\left|e_1^2-e_2^2\right|\right)\right| \le \left|e_1^1-e_1^2\right|+\left|e_2^1-e_2^2\right|.$$

Without loss of generality, let the left-hand side be $\left(1-\left|e_1^1-e_2^1\right|\right)-\left(1-\left|e_1^2-e_2^2\right|\right)$. The inequality then is equivalent to

$$\left|e_1^2-e_2^2\right| \le \left|e_1^2-e_1^1\right|+\left|e_1^1-e_2^1\right|+\left|e_2^1-e_2^2\right|,$$

which is true by the triangle inequality. For the second even vertex we have to show that

$$\left|\left(\left|e_1^1+e_2^1\right|-1\right)-\left(\left|e_1^2+e_2^2\right|-1\right)\right| \le \left|e_1^1-e_1^2\right|+\left|e_2^1-e_2^2\right|.$$

Again, without loss of generality, let the left-hand side be $\left(\left|e_1^1+e_2^1\right|-1\right)-\left(\left|e_1^2+e_2^2\right|-1\right)$.

(1) If $e_1^1+e_2^1 \ge 0$, $e_1^2+e_2^2 \ge 0$, the inequality acquires the form $\left(e_1^1-e_1^2\right)+\left(e_2^1-e_2^2\right) \le \left|e_1^1-e_1^2\right|+\left|e_2^1-e_2^2\right|$, which is true.

(2) If $e_1^1+e_2^1 < 0$, $e_1^2+e_2^2 < 0$, the inequality acquires the form $\left(e_1^2-e_1^1\right)+\left(e_2^2-e_2^1\right) \le \left|e_1^1-e_1^2\right|+\left|e_2^1-e_2^2\right|$, which is true.

(3) If $e_1^1+e_2^1 \ge 0$, $e_1^2+e_2^2 < 0$, we have

$$\begin{aligned}\left(e_1^1+e_1^2\right)+\left(e_2^1+e_2^2\right) &= \left(e_1^1-e_1^2\right)+\left(e_2^1-e_2^2\right)+2\left(e_1^2+e_2^2\right)\\ &\le \left(e_1^1-e_1^2\right)+\left(e_2^1-e_2^2\right) \le \left|e_1^1-e_1^2\right|+\left|e_2^1-e_2^2\right|,\end{aligned}$$

which is true. The fourth case is analogous.

Assume now that the statement of the theorem holds for all $2 \le k < n$. We have to show that

$$s_1\left(\mathbf{x}\right) \le n-2+\delta^{(n)}$$

for any even vertex of $\mathbb{R}_{\mathbf{b}}$. Without changing the values of s_1 and any of the summands in $\delta^{(n)}$, we can put the inequality in the canonical form (Lemma 8),

$$\sum_{i=1}^{n-1} x_{i,i\oplus 1} - x_{n1} \le n-2+\delta^{(n)}.$$

Consider two cases.

(Case 1) At least one of the coordinates $x_{i,i\oplus 1}$ ($i=1,\ldots,n-1$) is a max-coordinate. Let it be $x_{12} = 1-\left|x_1^1-x_2^1\right|$. We can rewrite the inequality as

$$\begin{aligned}\left(\sum_{i=2}^{n-1} x_{i,i\oplus 1} - x_{n1}\right)+1-\left|x_1^1-x_2^1\right| \le \left(n-3+\delta^{(n-1)}\right)\\ +1+\left|x_1^1-x_1^n\right|+\left|x_2^1-x_2^2\right|-\left|x_2^2-x_1^n\right|\end{aligned}. \quad (*)$$

The value of $\sum_{i=2}^{n-1} x_{i,i\oplus 1} - x_{n1}$ is equal to the s_1 of some system of rank $n-1$, and since the vector $\left\{x_{23},\ldots,x_{n-1,n},x_{n1}\right\}$ contains an even number of min-coordinates,

$$\left(\sum_{i=2}^{n-1} x_{i,i\oplus 1} - x_{n1}\right) \le n-3+\delta^{(n-1)}$$

holds by the induction hypothesis. At the same time, obviously,

$$1-\left|x_1^1-x_2^1\right| \le 1+\left|x_1^1-x_1^n\right|+\left|x_2^1-x_2^2\right|-\left|x_2^2-x_1^n\right|,$$

and this establishes (*) for this case.

(Case 2) All coordinates $x_{i,i\oplus 1}$ ($i = 1, \ldots, n-1$) are min-coordinates. Let us then replace two of them (which is possible since $n - 1 \geq 2$) with the corresponding max-coordinates — this will leave the number of the min-coordinates even. The left-hand side of (*) can only increase, but we can use the argument of the previous case to show that it is still less than the (unchanged) right-hand side of (*).

This completes the proof.

---

[1] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, *All noncontextuality inequalities for the n-cycle scenario*, Phys. Rev. A 88, 022118 (2013).

[2] J. V. Kujala, E. N. Dzhafarov, and J.-Å. Larsson, *Necessary and sufficient conditions for extended noncontextuality in a broad class of quantum mechanical systems*, Phys. Rev. Lett. 115, 150401 (2015).

[3] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, *A simple test for hidden variables in spin-1 system*, Phys. Rev. Lett. 101, 020403 (2008).

[4] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, *Experimental non-classicality of an indivisible quantum system*, Nature 474, 490 (2011).

[5] J. Bell, *On the Einstein-Podolsky-Rosen paradox*, Phys. 1, 195 (1964).

[6] J. Bell, *On the problem of hidden variables in quantum mechanics*, Rev. Modern Phys. 38, 447 (1966).

[7] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Proposed experiment to test local hidden-variable theories*, Phys. Rev. Lett. 23, 880 (1969).

[8] A. Fine, *Joint distributions, quantum correlations, and commuting observables*, J. Math. Phys. 23, 1306 (1982).

[9] G. Bacciagaluppi, *Leggett-Garg inequalities, pilot waves and contextuality*, Int. J. Quant. Found. 1, 1 (2015).

[10] J. Kofler and C. Brukner, *Condition for macroscopic realism beyond the Leggett-Garg inequalities*, Phys. Rev. A 87, 052115 (2013).

[11] A. J. Leggett and A. Garg, *Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks?* Phys. Rev. Lett. 54, 857 (1985).

[12] P. Suppes and M. Zanotti, *When are probabilistic explanations possible?* Synth. 48, 191 (1981).

[13] E. N. Dzhafarov, R. Zhang, and J. V. Kujala, *Is there contextuality in behavioral and social systems?* Phil. Trans. Roy. Soc. A 374, 20150099 (2016).

[14] Z. Wang, T. Solloway, R. M. Shiffrin, and J. R. Busemeyer, *Context effects produced by question orders reveal quantum nature of human judgments*, Proc. Natl. Acad. Sci. 111, 9431 (2014).

[15] E. N. Dzhafarov, V. H. Cervantes, and J. V. Kujala, *Contextuality in canonical systems of random variables*, Phil. Trans. Roy. Soc. A 375, 20160389 (2017).

[16] E. N. Dzhafarov and J. V. Kujala, *Probabilistic foundations of contextuality*, Fortsch. Phys. - Prog. Phys. 65, 1600040 (1-11) (2017).

[17] J. V. Kujala and E. N. Dzhafarov, *Measures of contextuality and noncontextuality*, Phil. Trans. Roy. Soc. A 377, 20190149 (2019).

[18] E. N. Dzhafarov, J. V. Kujala, and J.-Å. Larsson, *Contextuality in three types of quantum-mechanical systems*, Found. Phys. 45, 762 (2015).

[19] J. V. Kujala and E. N. Dzhafarov, *Proof of a conjecture on contextuality in cyclic systems with binary variables*, Found. Phys. 46, 282 (2016).

[20] M. Malinowski, C. Zhang, F. M. Leupold, J. Alonso, J. P. Home, and A. Cabello, A, *Probing the limits of correlations in an indivisible quantum system*, arXiv:1712.06494v2 (2018).

[21] M. Arias, G. Canas, E. S. Gomez, J. B. Barra, G. B. Xavier, G. Lima, V. D'Ambrosio, F. Baccari, F. Sciarrino, and A. Cabello, *Testing noncontextuality inequalities that are building blocks of quantum correlations,* Phys. Rev. A 92, 032126 (2015)

[22] A. Crespi, M. Bentivegna, I. Pitsios, D. Rusc, D. Poderini, G. Carvacho, V. D'Ambrosio, A. Cabello, F. Sciarrino, and R., Osellame, *Single-photon quantum contextuality on a chip,* ACS Photonics 4, 2807-2812. (2017).

[23] C. Flühmann, V. Negnevitsky, M. Marinelli, and J. P. Home, *Sequential modular position and momentum measurements of a trapped ion mechanical oscillator,* Phys. Rev. X 8, 021001 (2018).

[24] X. Zhan, P. Kurzyński, D. Kaszlikowski, K. Wang, Z. Bian, Y. Zhang, and P. Xue, *Experimental detection of information deficit in a photonic contextuality scenario,* Phys. Rev. Lett. 119, 220403 (2017).

[25] S. Abramsky, R. S. Barbosa, and S. Mansfield, *The contextual fraction as a measure of contextuality,* Phys. Rev. Lett. 119, 050504 (2017).

[26] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Bell nonlocality*, Rev. Mod. Phys. 86, 419 (2014).

[27] B. Amaral, M. T. Cunha, and A. Cabello, *Quantum theory allows for absolute maximal contextuality,* Phys. Rev. A 92, 062125 (2015).

[28] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik, *Quantifying Contextuality*, Phys. Rev. Lett. 112, 120401 (2014).

[29] M. Kleinmann, O. Gühne, J.-R. Portillo, J.-Å. Larsson, and A. Cabello, *Memory cost of quantum contextuality*, New J. Phys. 13, 113011 (2011).

[30] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf, *Contextuality as a resource for models of quantum computation with qubits,* Phys. Rev. Lett. 119, 120505 (2017).

[31] M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Contextuality supplies the 'magic' for quantum computation,* Nature 510, 351–355 (2014).

[32] C. Brukner, M. Zukowski, J.-W. Pan, and A. Zeilinger, *Bell's inequalities and quantum communication complexity,* Phys. Rev. Lett. 92, 127901 (2004).

[33] As Ref. [19] plays an important role in the present paper (see Theorem 3 and Lemmas 7-8), we should mention that we have noticed two unfortunate typos in the introduction to that paper (in Eq. 7 and Sec. 1.4). They are corrected in the arXiv'ed version of the paper, arXiv:1503.02181, and on the authors' websites.

[34] H. J. H. Tuenter, *Minimum L1-distance projection onto the boundary of a convex set: Simple characterization*, J. Optim. Theor. Appl., 112, 441 (2002).

[35] E. N. Dzhafarov, *On joint distributions, counterfactual values, and hidden variables in understanding contextuality,* Phil. Trans. Roy. Soc. A 377, 20190144 (2019).

[36] In all our previous publications Δ was simply $n-2+\delta$. However, at $\Delta \geq n$ any cyclic system of rank n is noncontextual. This will become apparent in Sec. VII: at $\Delta = n$ the noncontextuality polytope fills the entire hypercube of possible values of $\mathbf{e_b}$, and a further increase in Δ does not change this.

[37] E. N. Dzhafarov, *Replacing nothing with something special: Contextuality-by-Default and dummy measure-*

ments, in *Quantum Foundations, Probability and Information.* (Springer, Berlin, 2017), pp 39-44.

[38] B. Amaral, C. Duarte, and R. I. Oliveira, *Necessary conditions for extended noncontextuality in general sets of random variables,* J. Math. Phys. 59, 072202 (2018).

[39] N. Brunner, V. Scarani, and N. Gisin, *Bell-type inequalities for non-local resources*, J. Math. Phys. 47, 112101 (2006).