PHYSICAL REVIEW A **89**, 012103 (2014)

Information-theoretic metric as a tool to investigate nonclassical correlations

Paweł Kurzyński[1,2,*] and Dagomir Kaszlikowski[1,3,†]

[1]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543 Singapore, Singapore*

[2]*Faculty of Physics, Adam Mickiewicz University, Umultowska 85, 61-614 Poznań, Poland*

[3]*Department of Physics, National University of Singapore, 3 Science Drive 2, 117543 Singapore, Singapore*

(Received 22 October 2013; published 6 January 2014)

In this paper we study an application of an information-theoretic distance between two measurements to investigate nonclassical correlations. We postulate the triangle principle, which states that any information-theoretic distance is valid for any sets of measurements, regardless if they can be jointly measured or not. As a consequence, the triangle inequality for this distance is obeyed for any three measurements. This simple principle is valid in any classical realistic theory, however, it may not hold in quantum theory. It leads to the derivation of certain inequalities whose violations are an indicator of nonclassicality. Some of these inequalities formally look the same as those found in the literature on local realism and noncontextuality, but we also derive completely different inequalities. We also show that our geometrical approach naturally implies monogamy of nonclassical correlations.

DOI: 10.1103/PhysRevA.89.012103 PACS number(s): 03.65.Ta, 03.65.Ud

I. INTRODUCTION

Since the seminal papers by Bell [1] and Kochen-Specker [2] we know that quantum mechanics is incompatible with the assumptions of local realism (LR) and noncontextuality (NC). The NC-LR hypothesis states that all measurable properties of a physical system do not depend on the context in which they are measured. More precisely, suppose a given physical system has properties A, B, and C that yield outcomes a, b, and c with some probability distributions $p(a)$, $p(b)$, and $p(c)$. Suppose that the property A can be comeasured with the property B giving a probability distribution $p(a,b)$ or that it can be comeasured with the property C giving a probability distribution $p(a,c)$. We say that A can be measured in the context of B or C. The NC-LR hypothesis states that there exists a joint probability distribution $p(a,b,c)$ such that $p(a,b)$ and $p(a,c)$ are recovered as marginals.

Note that it might be impossible to measure $p(a,b,c)$ for some reason. For instance, in quantum theory if B and C are two orthogonal components of spin one cannot measure them jointly. The NC-LR hypothesis can be extended to more properties $A,B,C,D,\dots$ and it is equivalent to the existence of a joint probability distribution (JPD) for all the properties $p(a,b,c,d,\dots)$ [3–6].

The NC-LR hypothesis is very plausible based on our everyday experience. The color of your car, defined by its spectral profile, would be the same regardless if you looked at it together with Kochen or Specker. All relativistic classical theories of matter are compatible with NC-LR. The NC-LR hypothesis has been so far the only tool to investigate the borderline between classical and nonclassical correlations. In this paper we offer an alternative.

In NC-LR tests, a context can be established in two different ways. The most common one, which we call the Bell scenario, is to assume that A is measured in Alice's laboratory whereas B and C, who provide context for A, are measured in the spatially separated Bob's laboratory. We guarantee the lack of mutual influence between measurements in different laboratories by invoking the fact that information cannot propagate faster than light. The less common but more general scenario, which we call the Kochen-Specker one, is where there is no division into subsystems and all observables are measured on the same system. The context for A is provided by B and C whose lack of influence on A is imposed by the so-called no-disturbance assumption first mentioned by Gleason [7]. It is therefore clear that LR is a special case of NC and no-signaling is a special case of no-disturbance.

No-disturbance has not been justified by any general principle such as the finite speed of information's propagation but it can be tested experimentally. Simply measure A alone and then measure A followed by a measurement of B. Repeat the whole procedure in the reverse order. Do the same for A and C. If the obtained statistics for A, i.e., the probability distribution $p(a)$, is the same in all scenarios then you conclude that no-disturbance holds. This way we can verify that quantum mechanics is a no-disturbance theory. All relativistic and nonrelativistic classical theories of matter are, by their very foundations, no-disturbance theories as well.

It was Bell who showed that LR can be tested experimentally in a Bell scenario [1]. Experiments followed [8], clearly demonstrating that quantum mechanics violates LR. The Kochen and Specker [2] proof that indivisible quantum mechanical systems violate noncontextuality seemed to be impossible to test experimentally in a Kochen-Specker scenario until the paper by Klyachko-Can-Binicioglu-Schumovsky (KCBS) [6] whose KCBS inequality was tested experimentally [9,10]. It is interesting from the sociological point of view that it took 50 years to experimentally test the Kochen-Specker theorem whereas Bell scenarios were tested within 20 years of their formulation.

More formally, the Bell and Kochen-Specker scenarios can be tested via violations of noncontextuality inequalities. Many such inequalities have been derived for both scenarios. All derivations start from assuming that there is a JPD for all observables used to test a given system. One then manipulates this hypothetical JPD to obtain expressions that

*cqtpkk@nus.edu.sg

†phykd@nus.edu.sg

involve only measurable marginal probability distributions. Next, one derives upper and lower bounds for these expressions resulting from the assumption about the existence of a JPD.

However, inequalities that test LR and NC can be formulated in many ways. For example, the Clauser-Horne-Shimony-Holt (CHSH) inequality [11] uses probabilities of two comeasureable events from which one construct correlation functions between four pairs of measurements. The same measurement scenario leads to an equivalent inequality by Clauser and Horne (CH) [12], where, in addition to probabilities used in the CHSH case, one also uses probabilities of single local events. Yet another inequality involving entropies of these measurements can be derived [13–15]. Each of these inequalities requires a different approach to the JPD problem. This is because one has to use different mathematical methods to derive classical bounds for these inequalities. It is therefore natural to ask whether there exists a different approach to draw the line between classical and nonclassical correlations. In this paper, we modify and expand an idea of Schumacher [14] that properties of classical correlations can be expressed in geometrical terms.

We postulate the triangle principle that allows us to derive a large class of inequalities that must be obeyed by all classical correlations and show that they can be violated by quantum correlations. Some of these inequalities formally look like the known Bell-Kochen-Specker ones. Moreover, we show that the tradeoffs between their violations, known in the literature as monogamy relations, are a straightforward consequence of the properties of information-theoretic distance measures.

II. DISTANCES AND TRIANGLES

In this section we postulate a new principle that we call *the triangle principle*. It allows us to derive a large class of inequalities separating classical from nonclassical correlations. The inequalities we obtain have a simple geometrical meaning.

A. Shortest distance from A to B

We start with an abstract metric space with a distance function $d(X,Y)$ between any two points X and Y. Consider an arbitrary discrete subset of points X_1, X_2, ... that belongs to this space. It can be easily shown that the distance $d(X_1,X_2)$ between X_1 and X_2 is shorter than or equal to any other path from X_1 to X_2 that goes through $N-2$ other points in this subset $X_1 \to X_N \to X_{N-1} \to \cdots \to X_3 \to X_2$. More formally,

$$d(X_1,X_2) \leqslant \sum_{i=2}^{N-1} d(X_i,X_{i+1}) + d(X_N,X_1). \tag{1}$$

To prove this one simply starts with a triangle inequality

$$d(X_1,X_2) \leqslant d(X_2,X_N) + d(X_N,X_1). \tag{2}$$

Since $d(X_2,X_N)$ does not appear in Eq. (1), one uses another triangle inequality to bound this element from above

$$d(X_2,X_N) \leqslant d(X_{N-1},X_N) + d(X_2,X_{N-1}), \tag{3}$$

which after substitution to (2) gives

$$d(X_1,X_2) \leqslant d(X_2,X_{N-1}) + d(X_{N-1},X_n) + d(X_N,X_1). \tag{4}$$

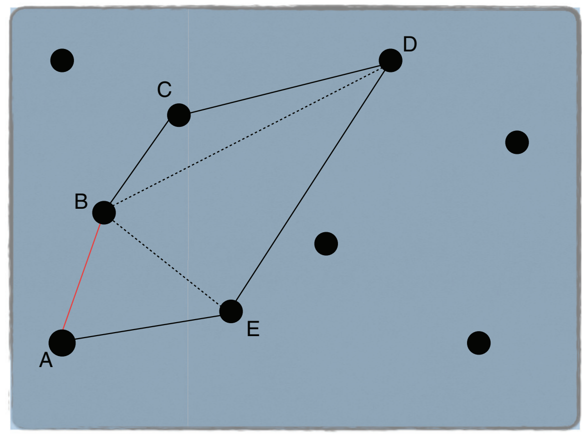

FIG. 1. (Color online) In a metric space the shortest distance from A to B is always $d(A,B)$ (red line). This can be shown via multiple applications of the triangle inequality. Dashed lines represent distances that do not occur in an alternative path from A to B (solid black lines), but are used to show that the length of this path is never shorter than $d(A,B)$.

The above procedure, wgich utilizes triangle inequality, is applied until all elements from (1) are obtained. A schematic picture corresponding to the case of five points is presented in Fig. 1.

B. Information-theoretic distance

Let us consider an experiment where we measure some properties of a physical system denoted as $A,B,C,\ldots$. Each property X yields an outcome x with probability $p(x)$. We further assume that only certain pairs of properties can be jointly measured. For instance, it is possible to obtain the probability distribution $p(a,b)$ for A and B but not for A and C, and so on. We are not interested in probability distributions involving more than two properties although they might be measurable in the experiment.

We introduce an information-theoretic distance measure $d(X,Y)$ between two probability distributions $p(x)$ and $p(y)$ having a joint probability distribution $p(x,y)$. This function must be (1) nonnegative and $d(X,Y)=0$ if and only if $X=Y$, (2) symmetric $d(X,Y)=d(Y,X)$, and (3) obey the triangle inequality $d(X,Y)+d(Y,Z) \geqslant d(X,Z)$ for arbitrary probability distributions $p(x,y)$, $p(y,z)$, and $p(x,z)$.

It is instructive to give a few examples of such distances.

Covariance distance [14]. It is defined for binary random variables ($x=\pm 1$) as $C(X,Y)=1-\sum_{x,y=\pm 1} xyp(x,y) = 1-\langle XY\rangle$.

Entropic distance [16]. Definition is as follows: $E(X,Y)=H(X|Y)+H(Y|X)$ where $H(X|Y)=H(XY)-H(Y)$ is the Shannon conditional entropy, therefore $E(X,Y)=2H(XY)-H(X)-H(Y)$. Interestingly, it is also a distance measure if one replaces the Shannon entropy by algorithmic entropy [16] or Tsallis entropy [17,18].

Kolmogorov distance. It reads $K(X,Y)=P(x_0)+P(y_0)-2P(x_0,y_0)$ where x_0,y_0 denote some particular events (not random variables as in the previous definitions), for instance, for binary variables we could have $x_0=-1, y_0=1$. It is a simple exercise in Venn diagrams to prove that K is a proper distance measure [19]. For binary variables there is a one-to-one correspondence between the Kolmogorov distance and the covariant distance $C(X,Y) = 2K(X=+1, Y=+1) = 2K(X=-1,Y=-1) = 2 - 2K(X=+1,Y=-1) = 2 - 2K(X=-1,Y=+1)$.

C. Triangle principle

We are now ready to formulate the triangle principle.

Any information-theoretic distance $d(X,Y)$ is valid for any sets of measurements, regardless if they can be jointly measured or not.

The formulation of the triangle principle is based on weaker assumptions than LR-NC. In particular JPD for three measurements X_1,X_2,X_3 imply that the distances $d(X_1,X_2),d(X_1,X_3),d(X_2,X_3)$ obey the triangle inequality. On the other hand, if the triangle inequality for these three distances is obeyed, a JPD may not exist. Here is a simple example

$$\begin{aligned} p(X_1=+1,X_2=-1) &= p(X_1=-1,X_2=+1) = 1/2,\\ p(X_2=+1,X_3=+1) &= p(X_2=-1,X_3=-1) = 1/2,\\ p(X_1=+1,X_3=+1) &= p(X_1=-1,X_3=-1) = 1/2, \end{aligned} \tag{5}$$

with all the unlisted probabilities equal to zero. It is straightforward to see that this probability distribution does not have a JPD and it does not violate the triangle inequality for the entropic distance. This example is revisited in more detail in the next section, but it already shows that the triangle principle is more general than the LR-NC hypothesis.

III. APPLICATIONS

To illustrate how this principle works we consider N binary measurements $X_1,\ldots,X_N$ (each X_i takes only two values) that are cyclically compatible, i.e., measurement X_i can be jointly measured with X_{i+1} (modulo N) [20]. First, we explicitly present the cases of $N=3,4,5$ followed by an arbitrary N.

A. $N = 3$

The case $N=3$ is particularly interesting since if X_1 is comeasurable with X_2, X_2 with X_3, and X_3 with X_1, one may think that all three observables are jointly measurable. This is true in quantum mechanics, however, one can consider generalized probabilistic theories (GPT) in which pairwise compatible measurements are not jointly compatible [5].

A special version of this problem was studied by Specker, who considered X_i to be three exclusive events [21]. We use notation $X_i = 1$ to denote that X_i occurred, and $X_i = -1$ to denote that it did not occur. Due to exclusivity the following holds:

$$p(X_1=1) + p(X_2=1) + p(X_3=1) \leqslant 1. \tag{6}$$

We refer to this inequality as the Specker inequality or the Specker principle. It is obeyed in classical and quantum theory but it can be violated in GPT [5,22].

Let us consider three binary measurements $X_i = \pm 1$ and invoke the triangle principle, i.e., assume that there exists an information distance between all three measurements. Therefore, the inequality (1) gives

$$d(X_1,X_2) \leqslant d(X_2,X_3) + d(X_3,X_1). \tag{7}$$

We show that this distance inequality yields a correlation inequality, an entropic inequality, or a probability inequality (Specker's inequality), if one chooses a proper distance function. Although these three inequalities are satisfied in quantum theory, we show that they can be violated in general probabilistic theories (GPT). Therefore, GPT do not obey the triangle principle.

Covariance distance. The inequality (7) becomes

$$1 - \langle X_1X_2\rangle \leqslant 1 - \langle X_2X_3\rangle + 1 - \langle X_3X_1\rangle, \tag{8}$$

which gives

$$-\langle X_1X_2\rangle + \langle X_2X_3\rangle + \langle X_3X_1\rangle \leqslant 1. \tag{9}$$

The above inequality is clearly the correlation noncontextuality inequality discussed in Ref. [20]. Although it is obeyed in quantum theory, it can be violated up to 3 in GPT by the following no-disturbance probability distribution

$$\begin{aligned} p(X_1=+1,X_2=-1) &= p(X_1=-1,X_2=+1) = 1/2,\\ p(X_2=+1,X_3=+1) &= p(X_2=-1,X_3=-1) = 1/2,\\ p(X_1=+1,X_3=+1) &= p(X_1=-1,X_3=-1) = 1/2, \end{aligned} \tag{10}$$

with all the remaining probabilities equal to zero. In all further examples we only list nonzero probabilities in any given probability distribution.

Entropic distance. The entropic version of (7) gives

$$\begin{aligned} H(X_1|X_2) + H(X_2|X_1) \leqslant{}& H(X_3|X_2) + H(X_2|X_3)\\ &+ H(X_1|X_3) + H(X_3|X_1). \end{aligned} \tag{11}$$

It is also obeyed in quantum theory, however, the GPT no-disturbance distribution

$$\begin{aligned} p(X_1=+1,X_2=+1) &= p(X_1=-1,X_2=-1) = 1/4,\\ p(X_1=+1,X_2=-1) &= p(X_1=-1,X_2=+1) = 1/4,\\ p(X_2=+1,X_3=+1) &= p(X_2=-1,X_3=-1) = 1/2,\\ p(X_1=+1,X_3=+1) &= p(X_1=-1,X_3=-1) = 1/2. \end{aligned} \tag{12}$$

leads to its violation, i.e., one gets a contradiction that $2 \leqslant 0$.

Note that the distribution (12) violates the inequality (9), but the distribution (10) does not violate (11). However, as was shown by Chaves [23], distributions that violate correlation inequalities can be mixed with some noncontextual distributions to give distributions that violate entropic inequalities. For example, (12) can be obtained from an even mixture of (10) and a noncontextual distribution that obeys both (9) and (11)

$$\begin{aligned} p(X_1=+1,X_2=+1) &= p(X_1=-1,X_2=-1) = 1/2,\\ p(X_2=+1,X_3=+1) &= p(X_2=-1,X_3=-1) = 1/2,\\ p(X_1=+1,X_3=+1) &= p(X_1=-1,X_3=-1) = 1/2. \end{aligned} \tag{13}$$

Kolmogorov distance. We define three events $A = (X_1 = 1, X_2 = -1)$, $B = (X_2 = 1, X_3 = 1)$, and $C = (X_3 = -1, X_1 = -1)$ with the previous notation that if $A = 1$ the event occurs and if $A = -1$ it does not. The same holds for B and C. Note that these events are pairwise exclusive.

Let us consider the following version of (7)

$$K(A = -1, B = -1) \leqslant K(B = -1, C = 1) + K(C = 1, A = -1). \quad (14)$$

Due to exclusivity $K(X = -1, Y = -1) = p(X = 1) + p(Y = 1)$, where we used the fact that $p(X = -1, Y = -1) = 1 - p(X = 1) - p(Y = 1)$ and $p(X = -1) + p(X = 1) = 1$. Also, $K(X = 1, Y = -1) = p(Y = -1) - p(X = 1)$ because $p(X = 1) = p(X = 1, Y = -1)$. We get

$$p(A = 1) + p(B = 1) \leqslant p(B = -1) + p(A = -1) - 2p(C = 1), \quad (15)$$

which after substitution of $p(X = -1) = 1 - p(X = 1)$ and division by 2 leads to the Specker's inequality

$$p(A = 1) + p(B = 1) + p(C = 1) \leqslant 1. \quad (16)$$

This inequality is violated up to 3/2 by the GPT distribution (10).

B. $N = 4$

This case naturally describes a bipartite Bell scenario in which Alice measures X_1 and X_3 while Bob measures X_2 and X_4. The corresponding inequality (1) reads

$$d(X_1, X_2) \leqslant d(X_1, X_4) + d(X_4, X_3) + d(X_3, X_2). \quad (17)$$

The distances $d(X_1, X_3)$ and $d(X_2, X_4)$ cannot be evaluated due to the lack of comeasurability. However, we use the triangle principle, which assumes that these distances, although unaccessible, exist.

We show that depending on the distance function the inequality (17) becomes the Clauser-Horne-Shimony-Holt (CHSH) inequality [11], the Schumacher inequality [14], or the Clauser-Horne (CH) inequality [12]. All three inequalities can be violated in qunatum theory provided a quantum state and measurement setups are properly chosen.

Covariance distance. For covariance distance the inequality (17) becomes

$$1 - \langle X_1 X_2 \rangle \leqslant 3 - \langle X_1 X_4 \rangle - \langle X_4 X_3 \rangle - \langle X_3 X_2 \rangle, \quad (18)$$

which has the form of the CHSH inequality

$$\langle X_1 X_4 \rangle + \langle X_4 X_3 \rangle + \langle X_3 X_2 \rangle - \langle X_1 X_2 \rangle \leqslant 2. \quad (19)$$

This observation was already made by Schumacher [14].

Entropic distance. The application of the entropic distance to (17) gives the Schumacher inequality

$$E(X_1, X_2) \leqslant E(X_1, X_4) + E(X_4, X_3) + E(X_3, X_2). \quad (20)$$

It is also important to mention similar entropic inequalities by Braunstein-Caves (BC) [13] and Cerf-Adami (CA) [15]. In particular the latter inequality is similar to the inequality we derive using the following distance

$$CA(X, Y) = 1 - \frac{I(X : Y)}{\max(X, Y)}, \quad (21)$$

where $I(X : Y) = H(X) + H(Y) - H(X, Y)$ is mutual information between X, Y and $\max(X, Y) = \max[H(X), H(Y)]$. We get

$$\frac{I(X_1 : X_2)}{\max(X_1, X_2)} + \frac{I(X_2 : X_3)}{\max(X_2, X_3)} + \frac{I(X_3 : X_4)}{\max(X_3, X_4)} - \frac{I(X_4 : X_1)}{\max(X_4, X_1)} \leqslant 2. \quad (22)$$

If all entropies $H(X_i) = 1$ we recover inequalities in Ref. [15].

Kolmogorov distance. We define four events $X_1 = 1$, $X_2 = 1$, $X_3 = 1$, and $X_4 = 1$. The Kolmogorov distance and the corresponding inequality (17) yield

$$K(X_1 = 1, X_2 = 1) \leqslant K(X_2 = 1, X_3 = 1) + K(X_3 = 1, X_4 = 1) + K(X_4 = 1, X_1 = 1), \quad (23)$$

which, after the substitution of $K(X = 1, Y = 1) = p(X = 1) + p(Y = 1) - 2P(X = 1, Y = 1)$ and division by 2, takes the form of the CH inequality [12]

$$-p(X_1 = 1, X_2 = 1) + p(X_2 = 1, X_3 = 1) + p(X_3 = 1, X_4 = 1) + p(X_4 = 1, X_1 = 1) - p(X_3 = 1) - p(X_4 = 1) \leqslant 0. \quad (24)$$

C. $N = 5$

The case of five cyclically compatible measurements is often related to the noncontextuality tests because $N = 5$ is the smallest number of measurements that can reveal contextuality in a three-level quantum system. Moreover, these measurements cannot be naturally distributed between two observers, therefore they are applied to a single indivisible system.

This time there are ten possible distances with only five that can be measured

$$d(X_1, X_2) \leqslant d(X_1, X_5) + d(X_5, X_4) + d(X_4, X_3) + d(X_3, X_2). \quad (25)$$

The remaining five distances cannot be measured, however, the triangle principle assumes that they exist and that the inequality (25) holds. We show that this inequality gives rise to correlation and probability versions of the Klyachko-Can-Binicioglu-Shumovsky (KCBS) inequalities [6] and to the entropic inequality that is similar to the one studied in Refs. [24,25].

Covariance distance. Plugging in the covariance distance into (25) results in

$$1 - \langle X_1 X_2 \rangle \leqslant 4 - \langle X_1 X_5 \rangle - \langle X_5 X_4 \rangle - \langle X_4 X_3 \rangle - \langle X_3 X_2 \rangle, \quad (26)$$

which is equivalent to a form of the KCBS inequality [6,20]

$$\langle X_1 X_5 \rangle + \langle X_5 X_4 \rangle + \langle X_1 X_5 \rangle + \langle X_5 X_4 \rangle - \langle X_1 X_2 \rangle \leqslant 3. \quad (27)$$

Entropic distance. For this distance one gets an inequality that is a five-measurement version of the Schumacher inequality. This inequality resembles the inequality studied in [24,25] where instead of $E(X, Y)$ one uses conditional entropy

$H(X|Y)$. Moreover, the distance $CA(X,Y)$ (21) can also be applied giving a completely new inequality.

Kolmogorov distance. Application of the Kolmogorov distance to the case $N=5$ resembles the one of $N=3$. Consider five pairwise exclusive events $A=(X_1=1,X_2=-1)$, $B=(X_2=1,X_3=-1)$, $C=(X_3=1,X_4=-1)$, $D=(X_4=1,X_5=-1)$, and $E=(X_5=1,X_1=-1)$. As before, $p(X=1,Y=1)=0$ (for $X,Y=A,\ldots,E$) and as a consequence $P(X=1,Y=-1)=P(X=1)$.

We apply the Kolmogorov distance to (25) and follow exactly the same steps as for $N=3$. We arrive at

$$P(A=1)+P(B=1)+P(C=1)+P(D=1) + P(E=1)\leqslant 2, \quad (28)$$

which is a version of the KCBS inequality expresed in terms of probabilities [6].

D. General N

The discussion of cases $N=3,4,5$ shows that the application of the covariance and the entropic distances readily generates correlation and entropic inequalities for general N. The correlation inequalities that are generated are of the form

$$-\langle X_1X_2\rangle+\sum_{i=2}^{N}\langle X_iX_{i+1}\rangle\leqslant N-2, \quad (29)$$

where $X_{N+1}\equiv X_1$. They exactly correspond to the inequalities discussed in Ref. [20]. The entropic inequalities

$$E(X_1,X_2)\leqslant\sum_{i=2}^{N}E(X_i,X_{i+1}), \quad (30)$$

are N element versions of Schumacher inequalities [14] and resemble N-cycle conditional entropic inequalities studied in Refs. [5,25]. In fact, these inequalities are symmetrized versions of conditional entropic inequalities.

Note that our model also applies to a bipartite Bell scenario if N is even and measurements X_{2i+1} are performed by Alice whereas X_{2i} are performed by Bob. In this case there is an additional number of distances that can be evaluated from the experimental data. Namely, every distance between Alice's and Bob's measurements exists. In this Bell scenario the inequalities (29) and (30) correspond to the chained Bell inequalities [26] and to the symmetric version of the multisetting BC inequalities [13], respectively.

On the other hand, the form of probability inequalities that are generated via application of the Kolmogorov distance depend on whether N is even or odd. For odd N one can generate inequalities that involve N cyclically exclusive events $A_1,\ldots,A_N$ defined as $A_i=(X_i=+1,X_{i+1}=-1)$. Next, one considers the Kolmogorov distance for $A_1=-1$, $A_i=1$ for $i=3,5,\ldots,N$ and $A_i=-1$ for $i=2,4,\ldots,N-1$. The following inequalities are obtained

$$\sum_{i=1}^{N}p(A_1=1)\leqslant\frac{N-1}{2}. \quad (31)$$

They correspond to the inequalities studied in Ref. [22].

The case of even N has to be explored in more detail. The inequalities involving N cyclicaly exclusive events can be violated in quantum theory and in GPT only for odd N [22]. We showed that for $N=4$ one can obtain a different type of inequality, namely the CH inequality. However, the application of Kolmogorov distance to scenarios with $N>4$ (even) remains to be explored.

IV. MONOGAMY RELATIONS

The monogamy relation between two inequalities is defined as a tradeoff between the violations of these inequalities. The more the first inequality is violated the less the second one is. In the most interesting case, if one inequality is violated the other one is satisfied and vice versa.

The monogamy relation can be studied either on a general level of GPT [27–29], or within quantum theory [29–31]. Monogamies in GPT stem from general principles such as no-signaling and no-disturbance. In quantum theory they originate from properties of operators in the Hilbert space.

Here we focus on the GPT case. In particular, we show that instead of referring directly to probabilities (no-signaling and no-disturbance), one uses the triangle principle to derive monogamy relations. The advantage of this approach is that once a monogamy relation is derived for a general distance measure, it automatically applies to every distance measure. We show how to derive a monogamy relation between the Bell inequality ($N=4$) and the noncontextuality inequality ($N=5$) [29]. We also derive a monogamy relation between two bipartite Bell inequalities ($N=4$) [30]. Finally, we speculate that our method can be easily applied to more general cases.

A. Monogamy between nonlocality and contextuality

It was shown in Ref. [29] that there exists a monogamy tradeoff between KCBS and CHSH inequalities (corresponding to the covariance distance). Here we show that this result can be generalized to an arbitrary distance measure $d(X,Y)$.

Consider two parties Alice and Bob sharing a bipartite system. Alice has five cyclically compatible measurements on her subsystem $\{A_1,\ldots,A_5\}$ and she randomly chooses to perform A_i and A_{i+1} (modulo 5). Bob can perform one of the two measurements B_1 or B_2 on his subsystem.

Alice's measurements can be used to test the following distance inequality (noncontextuality inequality if one assumes JPD)

$$d(A_1,A_5)\leqslant d(A_1,A_2)+d(A_2,A_3)+d(A_3,A_4)+d(A_4,A_5). \quad (32)$$

On the other hand, Bob's measurements and two incompatible measurements of Alice (say A_1 and A_3) can be used to test another distance inequality (Bell inequality if one assumes JPD)

$$d(A_1,B_1)\leqslant d(A_1,B_2)+d(B_2,A_3)+d(A_3,B_1). \quad (33)$$

Next, we show that if one inequality is violated, the other one is necessarily obeyed. In particular, it is enough to use the fact that the triangle inequality is always obeyed for compatible measurements to show that the following must hold

$$d(A_1,A_5)+d(A_1,B_1)\leqslant d(A_1,A_2)+d(A_2,A_3) + d(A_3,A_4)+d(A_4,A_5)+d(A_1,B_2) + d(B_2,A_3)+d(A_3,B_1). \quad (34)$$

We start with the triangle inequality

$$d(A_1,A_5) \leqslant d(A_1,B_2) + d(A_5,B_2), \quad (35)$$

and then we expand the last term using another triangle inequality

$$d(A_1,A_5) \leqslant d(A_1,B_2) + d(A_4,A_5) + d(A_4,B_2). \quad (36)$$

We repeat this procedure one more time to obtain

$$d(A_1,A_5) \leqslant d(A_1,B_2) + d(A_4,A_5) + d(A_3,A_4) + d(A_3,B_2). \quad (37)$$

Next, we follow similar steps to obtain

$$\begin{aligned} d(A_1,B_1) &\leqslant d(A_1,A_2) + d(A_2,B_1) \\ &\leqslant d(A_1,A_2) + d(A_2,A_3) + d(A_3,B_1). \end{aligned} \quad (38)$$

Finaly, we sum (37) and (38) to get (34).

B. Monogamy between two Bell inequalities

Consider three parties Alice, Bob, and Charlie who share a tripartite system. Each of them performs one of the two measurements on their subsystems $A_{1,2}$, $B_{1,2}$, and $C_{1,2}$. Note, that due to space-like separation the measurements A_i, B_j, and C_k $(i,j,k = 1,2)$ are mutually compatible.

Next, consider two Bell inequalities

$$d(A_1,B_1) \leqslant d(A_1,B_2) + d(B_2,A_2) + d(A_2,B_1), \quad (39)$$

$$d(A_1,C_1) \leqslant d(A_1,C_2) + d(C_2,A_2) + d(A_2,C_1). \quad (40)$$

Using the same methods as in the previous example we can show that

$$d(A_1,B_1) \leqslant d(A_1,C_2) + d(C_2,A_2) + d(A_2,B_1) \quad (41)$$

and

$$d(A_1,C_1) \leqslant d(A_1,B_2) + d(B_2,A_2) + d(A_2,C_1). \quad (42)$$

The sum of these two inequalities gives the monogamy relation

$$\begin{aligned} d(A_1,B_1) + d(A_1,C_1) \leqslant\ & d(A_1,B_2) + d(B_2,A_2) \\ & + d(A_2,B_1) + d(A_1,C_2) \\ & + d(C_2,A_2) + d(A_2,C_1). \end{aligned} \quad (43)$$

V. CONCLUSION

The triangle principle is an assumption about nature on the same footing as the NC-LR assumption. Both assume some mathematical properties of observed and unobserved probability distributions in nature. The NC-LR hypothesis assumes an existence of a hypothetical JPD that cannot be obtained in an experiment. The triangle principle assumes that any information-theoretic distance is valid for any sets of measurements, regardless if they can be jointly measured or not.

Mathematically speaking, the NC-LR hypothesis treats measurements and outcomes as points in a space with a measure (probability). The triangle principle introduces a metric on this space, which allows us to study the relation between these points using a geometric intuition. As we showed, this allows us to unify different types of non-contextuality and Bell inequalities in a more general framework and to derive more general monogamy relations.

We would like to highlight once more that *the triangle principle* leads to distance inequalities that are mathematically identical to known Bell and noncontextual inequalities, but they are based on different assumptions. Simply speaking, they imply that the assumption about the validity of the triangle inequality for nonobserved probability distributions is false.

This work leaves an important open problem. As shown, there are probability distributions that do not have JPD, but they satisfy the triangle inequality for some of the information theoretic distances, violating it for other distances. Are there probability distributions, perhaps signaling ones, that do not have JPD but satisfy the triangle inequality for *all* information-theoretic distances?

ACKNOWLEDGMENTS

This work is supported by the Foundational Questions Institute (FQXi) and by the National Research Foundation and Ministry of Education in Singapore. The authors acknowledge discussions with Marek Żukowski, Antoni Wójcik, Andrzej Grudka, Jayne Thompson, and Luigi Vacanti.

[1] J. S. Bell, Physics **1**, 195 (1964).
[2] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).
[3] A. Fine, Phys. Rev. Lett. **48**, 291 (1982).
[4] S. Abramsky and A. Brandenburger, New. J. Phys. **13**, 113036 (2011).
[5] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Phys. Rep. **506**, 1 (2011).
[6] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Phys. Rev. Lett. **101**, 020403 (2008).
[7] A. Gleason, J. Math. Mech. **6**, 885 (1957).
[8] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).
[9] R. Łapkiewicz, P. Li, C. Schaeff, N. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, Nature (London) **474**, 490 (2011).
[10] J. Ahrens, E. Amselem, A. Cabello, and M. Bourennane, Sci. Rep. **3**, 2170 (2013).
[11] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[12] J. F. Clauser and M. A. Horne, Phys. Rev. D **10**, 526 (1974).
[13] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).
[14] B. W. Schumacher, Phys. Rev. A **44**, 7047 (1991).
[15] N. J. Cerf and C. Adami, Phys. Rev. A **55**, 3371 (1997).
[16] W. H. Zurek, Nature (London) **341**, 119 (1989).
[17] S. Furuichi, J. Math. Phys. **47**, 023302 (2006).
[18] J. Modławska, P. Kurzyński, and D. Kaszlikowski (unpublished).

[19] A. Kolmogorov, *VI Zjazd Matematyków Polskich: Warszawa 20-23.IX.1948*, Dodatek do Rocznika Polskiego Towarzystwa Matematycznego (Instytut Matematyczny Uniwersytetu Jagiellońskiego, 1950).
[20] M. Araujo, M. T. Quintino, C. Budroni, M. T. Cuhna, and A. Cabello, Phys. Rev. A **88**, 022118 (2013).
[21] E. P. Specker, Dialectica **14**, 239 (1960).
[22] A. Cabello, S. Severini, and A. Winter, arXiv:1010.2163.
[23] R. Chaves, Phys. Rev. A **87**, 022102 (2013).
[24] P. Kurzyński, R. Ramanathan, and D. Kaszlikowski, Phys. Rev. Lett. **109**, 020404 (2012).
[25] R. Chaves and T. Fritz, Phys. Rev. A **85**, 032113 (2012).
[26] S. L. Braunstein and C. M. Caves, Ann. Phys. (NY) **202**, 22 (1990).
[27] R. Ramanathan, A. Soeda, P. Kurzyński, and D. Kaszlikowski, Phys. Rev. Lett. **109**, 050404 (2012).
[28] M. Pawlowski and Č. Brukner, Phys. Rev. Lett. **102**, 030403 (2009).
[29] P. Kurzynski, A. Cabello, and D. Kaszlikowski, arXiv:1307.6710.
[30] B. Toner and F. Verstraete, arXiv:quant-ph/0611001.
[31] P. Kurzyński, T. Paterek, R. Ramanathan, W. Laskowski, and D. Kaszlikowski, Phys. Rev. Lett. **106**, 180402 (2011).