PHYSICAL REVIEW A **101**, 012120 (2020)

Contextual robustness: An operational measure of contextuality

Lu Li,[1,*] Kaifeng Bu,[2,1,†] and Junde Wu[1,‡]

[1]*School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, People's Republic of China*

[2]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 16 June 2019; published 23 January 2020)

The operational characterization of contextuality is the cornerstone in the development of resource theory of contextuality. Here, we introduce contextual robustness as an operational measure of contextuality on empirical models. We first show that it satisfies all the properties a proper contextual measure should satisfy. Besides, we derive the linear programing of contextual robustness and prove that contextual robustness exactly characterizes the maximal violation of certain types of Bell inequalities. Moreover, we show that contextual robustness quantifies the accessible advantage of empiric models in measurement-based quantum computation, and provide a tighter upper bound on accessible advantage in certain cases.

DOI: 10.1103/PhysRevA.101.012120

I. INTRODUCTION

Nonlocality is one of the distinctive features of quantum theory, which exhibits the correlations by local measurement on separate subsystems. It has been proved that nonlocality is a fundamental resource in a variety of practical applications, ranging from quantum key distribution [1] to quantum communication complexity [2]. However, there is another phenomenon called quantum contextuality older than nonlocality. It states that the outcomes cannot be assigned to the measurements independent of the contexts of the measurements; otherwise, a logical contradiction would be obtained, which is known as the Kochen-Specker paradox [3]. Recently, the phenomenon of contextuality has been investigated in depth [4–9]. One critical observation about contextuality and nonlocality is that nonlocality is a special case of contextuality: the compatibility of the measurement outcomes are given by the measurement of observables on separable subsystems, and contextuality can even hold in a single system.

It has been shown that contextuality plays an important role in quantum computation [10–17], such as increasing the power of quantum computation [10,11] and realizing the universal quantum computation [12]. Identifying the relevant resource that enables the advantage of quantum speedup is one of the central problems in quantum computation, for which the quantification of contextuality is required to reveal the role of contextuality in quantum computation. Thus the resource theory of contextuality has attracted lots of attention in recent years [18–21]. A resource theory consists of two elements: free states and free operations. One of the well-known resource theories is that of quantum entanglement [22], which acts as a basic resource in a variety of quantum information processing protocols such as superdense coding [23], remote state preparation [24,25], and quantum teleportation [26]. In the resource theory of entanglement, the free states are separable states and the free operations are local operations and classical communication. Other notable examples include the resource theories of quantum coherence [27], asymmetry [28–31], thermodynamics [32], and steering [33]. One of the main advantages that a resource theory offers is the lucid quantitative and operational description. Recently, a lot of effort has been put into developing a resource theory of contextuality, and several operational measures of contextuality have been proposed, namely relative entropy of contextuality [19], contextual cost [19], and contextual fraction [21], subject to the physical requirements such as monotonicity under the free operations. For example, relative entropy of contextuality, defined by the relative entropy from information theory, has been proved to equal the maximal accessible correlation in certain communication scenarios [19]. Contextual fraction, defined as the minimum amount of contextuality contained in a given empiricial model, has been proposed and proved to play a critical role in some measurement-based quantum computation (MBQC) [21]. There is growing concern about the operational characterization of contextuality and further investigations are needed to provide an explicit and rigorous operational interpretation of contextuality.

Several useful frameworks to study contextuality have been proposed, including graph-theoretic framework [34], combinatorial approach [35], and sheaf-theoretical framework [36]. In this work, we consider the resource theory of contextuality in the sheaf-theoretical framework [36], which consists of measurement scenario and the (nonsignaling) empirical models. Here we introduce a measure of contextuality, contextual robustness, and focus on its operational interpretation. The contextual robustness is defined to be the minimal mixing required to erase the contextuality of an empirical model on a measurement scenario. We prove several properties of contextual robustness, which implies that it is a proper measure of contextuality. Besides, we derive the linear programing of the contextual robustness, and obtain the value for some special examples such as the Popescu-Rohrlich (PR) box. Based on the linear programing form of this contextual measure, we

*lilu93@zju.edu.cn

†kfbu@fas.harvard.edu

‡wjd@zju.edu.cn

find that the contextual robustness of any empirical model exactly characterizes the maximal violation of all positive Bell inequalities on this model. Moreover, we prove that the contextual robustness provides an upper bound for the probability of success in the l_2-MQBCs for evaluating nonlinear function and the upper bound obtained by contextual robustness is tighter than that obtained by contextual fraction [21] when the function becomes more nonlinear.

II. PRELIMINARIES

Before the main result, let us recall some basic facts about contextuality in the sheaf-theoretical approach [21,36]. A measurement scenario is represented by a triple $\langle X, \mathcal{M}, O\rangle$ with X being a finite set of measurements, $\mathcal{M}$ being a set of subsets of X, and O being a finite set of outcomes for each measurement in X. Any element $C \in \mathcal{M}$ is called a measurement context, i.e., a set of measurements which can be preformed together. Given a measurement scenario $\langle X, \mathcal{M}, O\rangle$, the (no-signaling) empirical model e is defined as follows: for each context $C \in \mathcal{M}$, e_C is a probability distribution on the set O^C with O^C being the set of all assignments to each measurement in C. Besides, it requires that this family of probability distribution is compatible on the overlap of the measurement contexts, i.e., for any two contexts $C, C' \in \mathcal{M}$, $e_C|_{C\cap C'} = e_{C'}|_{C\cap C'}$. Any empirical model e is called noncontextual if there is a probability distribution d on O^X such that $d|_C = e_C$ for any $C \in \mathcal{M}$ with O^X being the set of global assignments to all measurements. The empirical model e is called contextual if such global probability distribution on O^X does not exist.

Besides, we consider two combing operations, which combine two empirical models into a new one [21]. The first one is the controlled choice, i.e., given two models e_1 and e_2 on scenarios $\langle X_1, \mathcal{M}_1, O\rangle$ and $\langle X_2, \mathcal{M}_2, O\rangle$, respectively, then the model $e_1\&e_2$ is defined on the scenario $\langle X_1 \sqcup X_2, \mathcal{M}_1 \sqcup \mathcal{M}_2, O\rangle$ as follows, for any $C_i \in \mathcal{M}_i$:

$$(e_1\&e_2)_{C_i} := (e_i)_{C_i}, \tag{1}$$

where $\mathcal{M}_1 \sqcup \mathcal{M}_2 = \{C \mid C \in \mathcal{M}_1 \text{or } C \in \mathcal{M}_2\}$. The second one is the product, i.e., the model $e_1 \otimes e_2$ is defined on the scenario $\langle X_1 \sqcup X_2, \mathcal{M}_1 * \mathcal{M}_2, O\rangle$ as follows:

$$(e_1 \otimes e_2)_{C_1\sqcup C_2}\langle t_1, t_2\rangle := (e_1)_{C_1}(t_1)(e_2)_{C_2}(t_2), \tag{2}$$

for any $C_i \in \mathcal{M}_i$, $t_i \in O^{C_i}$, where $\mathcal{M}_1 * \mathcal{M}_2 = \{C_1 \sqcup C_2 \mid C_1 \in \mathcal{M}_1, C_2 \in \mathcal{M}_2\}$.

III. MAIN RESULTS

Given an empirical model e on a measurement scenario $\langle X, \mathcal{M}, O\rangle$, the contextual robustness $\mathrm{CR}(e)$ is defined as follows:

$$\mathrm{CR}(e) = \min\left\{\lambda \in \mathbb{R}_+ \middle| \frac{e + \lambda e'}{1+\lambda} \text{is noncontextual}, \; e' \text{ is a no-signaling empirical model}\right\}, \tag{3}$$

which quantifies the minimal mixing to erase the contextuality. If there exists λ such that $(1+\lambda)e^{NC} = e + \lambda e'$, then $(1+\lambda)e^{NC}|_C \geqslant e_C$ for any $C \in \mathcal{M}$; thus the contextual robustness can also be written as

$$\mathrm{CR}(e) = \min\{\lambda \in \mathbb{R}_+ | (1+\lambda)e^{NC}|_C \geqslant e_C, \forall C \in \mathcal{M}, \; e^{NC} \text{ is a noncontextual model}\}.$$

First, let us show some properties of contextual robustness, which implies that contextual robustness is a proper measure of contextuality.

Theorem 1. Given a measurement scenario $\langle X, \mathcal{M}, O\rangle$ and an empirical model e, the contextual robustness satisfies the following properties: (i) positivity, i.e., $\mathrm{CR}(e) \geqslant 0$ and $\mathrm{CR}(e) = 0$ iff e is noncontextual; (ii) monotonicity under translation of measurement and coarse-graining of outcomes; (iii) convexity, i.e., $\mathrm{CR}[pe + (1-p)e_2] \leqslant p\mathrm{CR}(e_1) + (1-p)\mathrm{CR}(e_2)$ for any $p \in [0, 1]$; (iv) $\mathrm{CR}(e_1\&e_2) = \max\{\mathrm{CR}(e_1), \mathrm{CR}(e_2)\}$; (v) $\mathrm{CR}(e_1 \otimes e_2) = \mathrm{CR}(e_1) + \mathrm{CR}(e_2) + \mathrm{CR}(e_1)\mathrm{CR}(e_2)$.

For simplicity, we put the definition of translation of measurements and coarse graining of outcomes in Appendix A. The properties (i)–(iii) hae been proved in Refs. [37,38]. The proof of properties (iv) and (v) is presented in Appendix A. Note that, based on property (v), we know that $\mathrm{CR}(e_1 \otimes e_2) \geqslant \mathrm{CR}(e_1) + \mathrm{CR}(e_2)$, which means contextual robustness will increase under tensor product. However, it has proved that $\mathrm{CF}(e_1 \otimes e_2) = \mathrm{CF}(e_1) + \mathrm{CF}(e_2) - \mathrm{CF}(e_1)\mathrm{CF}(e_2) \leqslant \mathrm{CF}(e_1) + \mathrm{CF}(e_2)$, which means contextual fraction will decrease under tensor product. Thus contextual robustness and contextual fraction behave differently under product [21].

For the given measurement scenario $\langle X, \mathcal{M}, O\rangle$, let us denote $n = |O^X|$, $m = \sum_{C\in\mathcal{M}} |O^C| = |\{\langle C, t\rangle, C \in \mathcal{M}, t \in C\}|$; then we can define an $m \times n$ (0,1) matrix M, called incidence matrix [21,36], as follows:

$$M[\langle C, t\rangle, g] = \begin{cases} 1, & g|_C = t, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Any empirical model e can be represented by a vector $v_e \in \mathbb{R}^m$ with the component $v_e[\langle C, t\rangle] = e_C(t)$ and the corresponding vector $v_{e^{NC}}$ of a noncontextual model e^{NC} can be written as $v_{e^{NC}} = M\vec{d}$, where $\vec{d}$ is a normalized vector (i.e., the components are nonnegative and the sum is equal to 1) [21,36]. Besides, a global superprobability distribution a can also be represented by a vector $\vec{a} \in \mathbb{R}^n$ with nonnegative components and the weight $w(a) = \vec{1}\cdot\vec{a}$, where $\vec{1} \in \mathbb{R}^n$ is a vector with each component being 1. Then $1 + \mathrm{CR}(e)$ is the solution of the following linear program (LP):

$$\begin{aligned} \min \quad & \vec{1}\cdot\vec{a}, \\ \text{such that} \quad & M\vec{a} \geqslant v_e, \\ & \vec{a} \geqslant 0, \vec{a} \in \mathbb{R}^n. \end{aligned} \tag{5}$$

Since O^X is the set of all global assignments, then for any $C \in \mathcal{M}$ and $t \in O^C$, there exists some global assignment $g \in O^X$ such that $g|_C = t$, which implies $\sum_{g\in O^X} M[\langle C, t\rangle, g] \geqslant 1$, i.e., the sum of each row is larger than 1. Let $\vec{a} = \mu\vec{1}$ with $\mu = \max_{C,t\in O^C} e_C(t) \in [0, 1]$; then $M\vec{a} \geqslant v_e$, i.e., $\vec{a}$ is a feasible solution of the LP (5) and $1 + \mathrm{CR}(e) \leqslant w(a) = n\mu \leqslant n < +\infty$. Thus, due to the strong duality of the linear program [39], $1 + \mathrm{CR}(e)$ is also the solution of the following linear

TABLE I. CHSH empirical model on (2,2,2) Bell scenario [36].

Alice	Bob	00	01	10	11
a	b	$\frac{1}{2}$	0	0	$\frac{1}{2}$
a	b'	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
a'	b	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
a'	b'	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

program:

$$\begin{aligned} \max \quad & \vec{b}\cdot v_e, \\ \text{such that} \quad & M^T\vec{b} \leqslant \vec{1}, \\ & \vec{b} \geqslant 0, \vec{b} \in \mathbb{R}^m. \end{aligned} \tag{6}$$

Since there are lots of efficient algorithms calculating the linear program [40,41], then the robustness of contextuality can be evaluated efficiently. For example, let us consider (2,2,2) Bell scenario, where Alice can choose one of the measurements a and a', Bob can also choose the measurement b or b', and each of the measurements has two possible outcomes, 0 or 1. In this scenario, $X = \{a, a', b, b'\}$, $\mathcal{M} = \{\{a, b\}, \{a, b'\}, \{a', b\}, \{a', b'\}\}$, and $O = \{0, 1\}$. Here, we consider the CHSH model e_{CHSH} and the PR model e_{PR} on the (2,2,2) Bell scenario [36], which come from the local projective measurement on the maximally entangled state on two qubits [42] and PR box [43], respectively. (See the data in Tables I and II.) Then the contextual robustness of these two models are $\text{CR}(e_{CHSH}) = 1/12$ and $\text{CR}(e_{PR}) = 1/3$ using the solvers for linear programs in Refs. [40,41]. The corresponding incidence matrix M and the vectors $v_{e_{CHSH}}, v_{e_{PR}}$ are provided in Appendix B.

In the following contexts, we focus on the operational interpretation and application of the contextual robustness, and its connection with other contextual measures.

Maximum violation of positive generalized Bell inequality. Any nontrivial inequality for a scenario $\langle X, \mathcal{M}, O\rangle$ is represented by a real vector $\vec{b} \in \mathbb{R}^m$ with at least one non-negative component. For a model e, the following inequality:

$$\vec{b}\cdot v_e \leqslant L(\vec{b}), \tag{7}$$

is called a generalized Bell inequality if it is satisfied by all noncontextual models [21], where $\vec{b}\cdot v_e := \sum_{C\in\mathcal{M}, t\in O^C} b[\langle C, t\rangle]e_C(t)$ and $L(\vec{b}) \in \mathbb{R}$ is non-negative. Let us take $L(\vec{b}) = \max_{e^{NC}} \vec{b}\cdot v_{e^{NC}}$, where the maximization is taken over all noncontextual models. (Note that such type of inequality is a generalization of Bell inequality defined for Bell-type scenarios for nonlocality; see [44,45] for the framework of Bell inequality in Bell-type scenarios.) Then the violation of the generalized Bell inequality $\vec{b}$ by the model e is quantified by

$$\frac{\vec{b}\cdot v_e}{L(\vec{b})}. \tag{8}$$

Here, we find that CR(e) captures the maximal violation of all positive Bell inequalities, where positive Bell inequality means each component of $\vec{b}$ is non-negative, i.e., $b[\langle C, t\rangle] \geqslant 0$ for any $C \in \mathcal{M}, t \in O^C$ and we denote it by $\vec{b} \geqslant 0$.

Theorem 2. Given an empirical model e on a measurement scenario $\langle X, \mathcal{M}, O\rangle$, the maximal violation over all positive Bell inequality is equal to $1 + \text{CR}(e)$, i.e.,

$$1 + \text{CR}(e) = \max_{\vec{b}\geqslant 0} \frac{\vec{b}\cdot v_e}{L(\vec{b})}. \tag{9}$$

The proof of this result is based on the linear programs (5) and (6) of CR(e) and the details of proof are presented in Appendix C. Based on Theorem 2, the violations of positive generalized Bell inequality are able to demonstrate contextuality.

Upper bound on average success possibility of l_2-MBQC and nonlocal games. Now, let us consider the role of contextuality in the following l_2-MBQC, which is proposed in [11] and also studied in [21]. Here we make a short introduction of the process of l_2-MBQC in [11] for the completeness of the work. An l_2-MBQC on an n-particle system with l bits of classical inputs and m bits of classical outputs consists of the following elements: (1) a preprocessing $n \times l$ $\mathbb{Z}_2$-matrix Q; (2) a postprocessing $m \times n$ $\mathbb{Z}_2$-matrix Z; (3) a $n \times n$ lower triangular $\mathbb{Z}_2$-matrix T with vanishing diagonal; (4) an empirical model e on the $(n, 2, 2)$ Bell scenario, where the measurement context and the set of joint outcomes can be represented by the set $\mathbb{Z}_2^n$ and thus the empirical model e determines a function from $\mathbb{Z}_2^n$ to $D(\mathbb{Z}_2^n)$, i.e., for any measurement context $\vec{q} \in \mathbb{Z}_2^n$, $e_{\vec{q}}$ is a distribution on the set of joint outcomes.

For each input $\vec{i} \in \mathbb{Z}_2^l$, the output $\vec{o} \in \mathbb{Z}_2^m$ can be obtained with $\vec{q}, \vec{s} \in \mathbb{Z}_2^n$ as follows:

$$\vec{q} = Q\vec{i} + T\vec{s}, \vec{o} = Z\vec{s}. \tag{10}$$

Such l_2-MBQC is denoted by $\langle K, e\rangle$, where $K = \langle Q, Z, T\rangle$ describes the classical processing. The l_2-MQBC determines a map $[[K, e]] : \mathbb{Z}_2^l \to D(\mathbb{Z}_2^m)$, where for any input $\vec{i} \in \mathbb{Z}_2^l$ and output $\vec{o} \in \mathbb{Z}_2^m$, $[[K, e]](\vec{i})(\vec{o})$ is the probability to obtain the outcome $\vec{o}$ with input $\vec{i}$ by the l_2-MBQC $\langle K, e\rangle$. For any function $f : \mathbb{Z}_2^l \to \mathbb{Z}_2^m$, the average success probability to evaluate f by l_2-MBQC $\langle K, e\rangle$ is

$$\bar{p}_s^{\langle K,e\rangle,f} = \frac{1}{2^l}\sum_{\vec{i}\in\mathbb{Z}_2^l} [[K, e]](\vec{i})[f(\vec{i})]. \tag{11}$$

We denote $\bar{p}_s^{\langle K,e\rangle,f}$ by $\bar{p}_s$ for convenience. In [21], it has been proved that the contextual fraction can provide an upper bound to the average success probability $\bar{p}_s$. Here, we find that contextual robustness can also provide an upper bound on $\bar{p}_s$,

TABLE II. PR empirical model on (2,2,2) Bell scenario [36].

Alice	Bob	00	01	10	11
a	b	$\frac{1}{2}$	0	0	$\frac{1}{2}$
a	b'	$\frac{1}{2}$	0	0	$\frac{1}{2}$
a'	b	$\frac{1}{2}$	0	0	$\frac{1}{2}$
a'	b'	0	$\frac{1}{2}$	$\frac{1}{2}$	0

and such an upper bound is much tighter when the function f is more nonlinear.

Theorem 3. Given a Boolean function $f:\mathbb{Z}_2^l\to\mathbb{Z}_2^m$ and an empirical model e, for any l_2-MBQC $\langle K,e\rangle$ that evaluates the function f with average success probability $\bar{p}_s$, it holds that

$$\bar{p}_s \leqslant [1+\mathrm{CR}(e)][1-\tilde{\upsilon}(f)], \tag{12}$$

where the average distance of f to the closest $\mathbb{Z}_2$-linear function $\tilde{\upsilon}(f)$ is defined as

$$\tilde{\upsilon}(f) := \min\left\{ dis(f,h) \mid h:\mathbb{Z}_2^l\to\mathbb{Z}_2^m \mathrm{is}\mathbb{Z}_2-\mathrm{linear}\right\}$$

and

$$\mathrm{dis}(f,g) := \tfrac{1}{2^l}\left|\left\{\vec{i}\in\mathbb{Z}_2^l \mid f(\vec{i})\neq g(\vec{i})\right\}\right|.$$

The proof of this theorem is presented in Appendix C. This theorem proves that the average probability to evaluate the nonlinear function correctly by l_2 MBQC is upper bounded by the contextual robustness of the empirical model, which quantitatively reveals the role of contextuality in the l_2-MQBC. Now, let us show that contextual robustness can provide a tighter bound on the success probability than that of the contextual fraction [21] for more nonlinear function. Let us consider the CHSH model e_{CHSH} in a (2,2,2) Bell scenario. First, based on Theorem 3, the average success probability $\bar{p}_s$ to calculate a nonlinear function f by the l_2-MBQC $\langle K,e_{CHSH}\rangle$ is upper bounded by

$$\bar{p}_s \leqslant [1+\mathrm{CR}(e_{CHSH})][1-\tilde{\upsilon}(f)] = \tfrac{13}{12}[1-\tilde{\upsilon}(f)],$$

where $\mathrm{CR}(e_{CHSH})=1/12$. Besides, based on [21], the success probability $\bar{p}_s$ is also upper bounded as

$$\bar{p}_s \leqslant 1-[1-\mathrm{CF}(e_{CHSH})]\tilde{\upsilon}(f) = 1-\tfrac{3}{4}\tilde{\upsilon}(f),$$

where $\mathrm{CF}(e_{CHSH})=1/4$. It is easy to verify that $\frac{13}{12}[1-\tilde{\upsilon}(f)]\leqslant 1-\frac{3}{4}\tilde{\upsilon}(f)$ if $\tilde{\upsilon}(f)\geqslant 1/4$, which means that the upper bound on the average success probability in (12) provides a tighter bound for more nonlinear function in the (2,2,2) Bell scenario.

Relationship with other contextual measures. Now, we investigate the relationship between contextual robustness and other measures of contextuality such as relative entropy of contextuality $X_{\max}$ [19] and contextual fraction [21] (or contextual cost [19]). First, due to the definition of relative entropy of contextuality $X_{\max}$ [19], relative entropy of contextuality of an empirical model e on a measurement scenario $\langle X,\mathcal{M},O\rangle$ can be written as

$$X_{\max}(e) = \max_{p_C}\min_{e^{NC}}\sum_{C\in\mathcal{M}} p_C D\big(e_C\|e_C^{NC}\big),$$

where $D(e_C\|e_C^{NC})=\sum_{t\in O^C} e_C(t)\log_2[e_C(t)/e_C^{NC}(t)]$, the minimization is taken over all the noncontextual empirical models, and the maximization is taken over the probability distributions $\{p_C\}$ on the set of contexts $\mathcal{M}$. In terms of the definition of contextual robustness, there exists a noncontextual empirical model e^{NC} such that $e\leqslant[1+\mathrm{CR}(e)]e^{NC}$, i.e., $e_C(t)\leqslant[1+\mathrm{CR}(e)]e_C^{NC}(t)$ for any context $C\in\mathcal{M}$ and any $t\in O^C$. Thus $D(e_C\|e_C^{NC})\leqslant\log_2[1+\mathrm{CR}(e)]$ for any $C\in\mathcal{M}$, that is,

$$X_{\max}(e)\leqslant\log_2[1+\mathrm{CR}(e)],$$

for any empirical model e. Besides, due to the definition of contextual fraction [21], any empirical model e can be decomposed as $e=[1-\mathrm{CF}(e)]e^{NC}+\mathrm{CF}(e)e'$, where e' is strongly contextual. Since contextual robustness is convex, then $\mathrm{CR}(e)\leqslant[1-\mathrm{CF}(e)]\mathrm{CR}(e^{NC})+\mathrm{CF}(e)\mathrm{CR}(e')\leqslant \mathrm{CF}(e)(|O|^{|X|}-1)$, where $\mathrm{CR}(e^{NC})=0$ and the contextual robustness is upper bounded by $|O|^{|X|}-1$. Therefore, given a nonsignaling empirical model e on a measurement scenario $\langle X,\mathcal{M},O\rangle$, we have the following relationship among these three contextual measures,

$$2^{X_{\max}(e)}-1\leqslant\mathrm{CR}(e)\leqslant(|O|^{|X|}-1)\mathrm{CF}(e). \tag{13}$$

Let us consider the (2,2,2) Bell scenario, where the only nonsignaling and strongly contextual model is PR model e_{PR} [36] and the corresponding contextual robustness is $\mathrm{CR}(e_{PR})=1/3$. Then, for any (nonsignaling) empirical model e on the (2,2,2) Bell scenario, it holds that

$$\mathrm{CR}(e)\leqslant\tfrac{1}{3}\mathrm{CF}(e). \tag{14}$$

IV. CONCLUSION

To summarize, we have investigated the properties of contextual robustness, especially the operational interpretation and application of contextual robustness. It has been proved that contextual robustness is monotone under the free operations in a resource theory of contextuality, i.e., contextual robustness is a proper measure of contextuality. Besides, contextual robustness is equal to the maximal violation of positive Bell inequalities, which shows the nonlocal feature of contextuality in single quantum systems. Moreover, the contextual robustness provides an upper bound for the probability of success in l_2-MBQCs for evaluating the nonlinear function, and such an upper bound is much tighter than that of the contextual fraction when the function is more nonlinear. Furthermore, the relationship between contextual robustness and other known contextual measures has been investigated. These result may highlight the understanding to an operational resource theory of contextuality and pave the road to the further investigation of contextuality in quantum computation.

There are also some problems left. In this paper, we only consider the positive generalized Bell inequality $\vec{b}$ in Theorem 2. In fact, we can also consider nonpositive generalized Bell inequality $\vec{b}$, and define a new measure as follows:

$$\mathrm{CB}(e)=\max_{\vec{b}\neq 0}\frac{\vec{b}\cdot v_e}{L(\vec{b})}-1.$$

We leave this measure of contextuality for a future work.

ACKNOWLEDGMENTS

K.F.B. acknowledges the support of a grant from the Templeton Religion. This work is supported by the National Natural Science Foundation of China (Grant No. 61877054).

APPENDIX A: PROPERTIES OF CONTEXTUAL ROBUSTNESS

There are two possible free operations in the resource theory of contextuality proposed in Ref. [21]. The first one is the translation of measurements: given two measurement scenarios $\langle X', \mathcal{M}', O\rangle$ and $\langle X, \mathcal{M}, O\rangle$ and an empirical model e on $\langle X', \mathcal{M}', O\rangle$, then for any context-preserving function $f : X \to X'$ [i.e., for any $C \in \mathcal{M}$, there exists $C' \in \mathcal{M}'$ such that $f(C) \subset C'$], the empirical model f^*e on $\langle X, \mathcal{M}, O\rangle$ is defined as follows, for any $C \in \mathcal{M}$ and $t \in O^C$:

$$(f^*e)_C(t) := \sum_{s\in O^{f(C)}, s\circ f|_C = t} e_{f(C)}(s). \tag{A1}$$

The second one is the coarse graining of the outcomes, i.e., given an empirical model e on the scenario $\langle X, \mathcal{M}, O'\rangle$, then for any $h : O' \to O$, the model e/h on the scenario $\langle X, \mathcal{M}, O\rangle$ is defined as follows: for each $C \in \mathcal{M}$ and $t \in O^C$,

$$(e/h)_C(t) := \sum_{s\in O'^C, h\circ s = t} e_C(s). \tag{A2}$$

To prove the properties of contextual robustness, we need several lemmas first, where the lemmas can be easily obtained based on Lemmas 5 and 6 in the Supplemental Material of [21].

Lemma 4. (Reference [21].) Given two superprobability distributions a_S and a_T on the sets S and T, respectively, with the same weight w, then there exists a superprobability distribution "a" on $S \times T$ with weight w and $a|_S = a_S$, $a|_T = a_T$.

Note that this lemma can be directly obtained from Lemma 5 in the Supplemental Material of [21] as the proof in [21] does not use the condition that the summation of the components is less than 1.

Lemma 5. Given two superprobability distributions a_S and a_T on the sets S and T, respectively, there exists a superprobability distribution a on $S \times T$ such that $a|_S \geqslant a_S$, $a|_T \geqslant a_T$ and $w(a) = \max\{\, w(a_S), w(a_T)\,\}$.

Proof. Since a_S and a_T are superprobability distributions, then the weight is larger than 1. Without loss of generality, suppose $w(a_T) = \max\{\, w(a_S), w(a_T)\,\} \geqslant 1$; then let us define a new superprobability distribution $a'|_S = \frac{w(a|_T)}{w(a|_S)} a_S$. Thus a'_S and a_T have the same weight $\max\{\, w(a_S), w(a_T)\,\}$. Due to Lemma 4, we get the result. ■

Proof of Theorem 1. Let us prove the properties (iv) and (v).

(iv) Given two empirical models e_1 and e_2 on $\langle X_1, \mathcal{M}_1, O\rangle$ and $\langle X_2, \mathcal{M}_2, O\rangle$, respectively, the model $e_1\&e_2$ is defined on the scenario $\langle X_1 \sqcup X_2, \mathcal{M}_1 \sqcup \mathcal{M}_2, O\rangle$. In view of the definition of contextual robustness, there exists an optimal superprobabilty distribution $a_{e_1\&e_2}$ on $O^{X_1\sqcup X_2} \cong O^{X_1} \times O^{X_2}$, such that $a_{e_1\&e_2}|_C \geqslant (e_1\&e_2)_C$ for any $C \in \mathcal{M}_1 \sqcup \mathcal{M}_2$.

Let $a_i := a_{e_1\&e_2}|_{X_i}$ be a superprobability distribution on O^{X_i}; then $w(a_i) = w(a_{e_1\&e_2})$ for $i = 1, 2$. For any $C \in \mathcal{M}_i$ and $t \in O^C$,

$$\begin{aligned} a_i|_C(t) &= (a_{e_1\&e_2}|_{X_i})_C(t) \\ &= a_{e_1\&e_2}|_C(t) \\ &\geqslant (e_1\&e_2)_C(t) \\ &= (e_i)_C(t), \end{aligned}$$

where the inequality comes from the fact that $a_{e_1\&e_2}|_C \geqslant (e_1\&e_2)_C$ for any $C \in \mathcal{M}_i$. Thus $1 + \mathrm{CR}(e_i) \leqslant w(a_i) = w(a_{e_1\&e_2}) = 1 + \mathrm{CR}(e_1\&e_2)$, which implies that $\mathrm{CR}(e_1\&e_2) \geqslant \max\{\, \mathrm{CR}(e_1), \mathrm{CR}(e_2)\,\}$.

To prove the converse direction, let a_{e_i} be the optimal superprobability distribution such that $w(a_{e_i}) = 1 + \mathrm{CR}(e_i)$ and $a_{e_i}|_C \geqslant (e_i)_C$ for any $C \in \mathcal{M}_i$. According to Lemma 5, there exists a superprobability distribution on $O^{X_1\sqcup X_2} \cong O^{X_1} \times O^{X_2}$ such that $w(a) = \max\{\, w(a_{e_1}), w(a_{e_2})\,\}$ and $a|_{X_i} \geqslant a_{e_i}$ for $i = 1, 2$. Thus, for any $C \in \mathcal{M}_1 \sqcup \mathcal{M}_2$ with $\mathcal{M}_i$ being the component which C belongs to, then

$$\begin{aligned} a|_C &= a|_{X_i}|_C \\ &\geqslant a_{e_i}|_C \\ &\geqslant (e_i)_C \\ &= (e_1\&e_2)_C, \end{aligned}$$

where the first inequality comes from the condition $a|_{X_i} \geqslant a_{e_i}$ and the second inequality comes from the fact $(a_{e_i})|_C \geqslant (e_i)_C$ for any $C \in \mathcal{M}_i$. Thus $1 + \mathrm{CR}(e_1\&e_2) \leqslant w(a) = \max\{\, w(a_{e_1}), w(a_{e_2})\,\} = \max\{\, 1 + \mathrm{CR}(e_1), 1 + \mathrm{CR}(e_2)\,\}$.

(v) Given two empirical models e_1 and e_2 on $\langle X_1, \mathcal{M}_1, O\rangle$ and $\langle X_2, \mathcal{M}_2, O\rangle$, respectively, it has been shown that the incidence matrix M for the scenario $\langle X_1 \sqcup X_2, \mathcal{M}_1 * \mathcal{M}_2, O\rangle$ is $M = M_1 \otimes M_2$, with M_i being the incidence matrix for $\langle X_i, \mathcal{M}_i, O\rangle$ for $i = 1, 2$ and the vector representation of $e_1 \otimes e_2$ is $v_{e_1\otimes e_2} = v_{e_1} \otimes v_{e_2}$. Note that the global assignments for $\langle X_1 \sqcup X_2, \mathcal{M}_1 * \mathcal{M}_2, O\rangle$ correspond to the points $\langle g_1, g_2\rangle$ bijectively with g_i being a global assignment for $\langle X_i, \mathcal{M}_i, O\rangle$ and the context $C \in \mathcal{M}_1 * \mathcal{M}_2$ corresponding to the pairs $\langle C_1, C_2\rangle$ with $C_i \in \mathcal{M}_i$ [21].

Let $\vec{a}_i$ be the optimal solution with respect to the model e_i; then for $\vec{a} = \vec{a}_1 \otimes \vec{a}_2$, we have

$$\begin{aligned} M\vec{a} &= (M_1 \otimes M_2)(\vec{a}_1 \otimes \vec{a}_2) \\ &= M_1\vec{a}_1 \otimes M_2\vec{a}_2 \\ &\geqslant v_{e_1} \otimes v_{e_2}. \end{aligned}$$

Thus $1 + \mathrm{CR}(e_1 \otimes e_2) \leqslant \vec{1} \cdot \vec{a} = (\vec{1} \cdot \vec{a}_1)(\vec{1} \cdot \vec{a}_2) = [1 + \mathrm{CR}(e_1)][1 + \mathrm{CR}(e_2)]$.

To prove the converse direction, let $\vec{b}_i$ be the optimal solution of dual LP with respect to the model e_i for $i = 1, 2$; then for $\vec{b} = \vec{b}_1 \otimes \vec{b}_2$,

$$\begin{aligned} M^T\vec{b} &= \left(M_1^T \otimes M_2^T\right)(\vec{b}_1 \otimes \vec{b}_2) \\ &= \left(M_1^T\vec{b}_1\right) \otimes (M^T\vec{b}_2) \\ &\leqslant \vec{1} \otimes \vec{1}. \end{aligned}$$

Thus $1 + \mathrm{CR}(e_1 \otimes e_2) \geqslant \vec{b} \cdot v_{e_1\otimes e_2} = (\vec{b}_1 \cdot v_{e_1})(\vec{b}_2 \cdot v_{e_2}) = [1 + \mathrm{CR}(e_1)][1 + \mathrm{CR}(e_2)]$. ■

APPENDIX B: CONTEXTUAL ROBUSTNESS OF CHSH MODEL AND PR MODEL

Here, we consider the (2,2,2) Bell scenario, where $X = \{a, a', b, b'\}$, $\mathcal{M} = \{\{a, b\}, \{a, b'\}, \{a', b\}, \{a', b'\}\}$, and $O = \{0, 1\}$. In this case, the incidence matrix M is a 16×16 (0,1) matrix, which can be written as follows:

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{B1}$$

The corresponding vectors of CHSH model e_{CHSH} and PR model e_{PR} can be presented as follows:

$$\vec{v}_{e_{CHSH}} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{3}{8} & \frac{1}{8} & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{bmatrix}^T$$

and

$$\vec{v}_{e_{PR}} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}^T.$$

Using the solver in Refs. [40,41], we obtain the contextual robustness and context fraction for e_{CHSH} and e_{PR} as follows:

$$\mathrm{CR}(e_{CHSH}) = 1/12, \quad \mathrm{CR}(e_{PR}) = 1/3,$$
$$\mathrm{CF}(e_{CHSH}) = 1/4, \quad \mathrm{CF}(e_{PR}) = 1.$$

APPENDIX C: DETAILS ABOUT PROOF OF THEOREMS 2 AND 3

Proof of Theorem 2. First, due to the definition of contextual robustness, there exists a noncontextual model e^{NC} such that $[1 + \mathrm{CR}(e)]e^{NC} \geqslant e$, which implies $[1 + \mathrm{CR}(e)]v_{e^{NC}} \geqslant v_e$; then for any positive Bell inequality $\vec{b}$, we have

$$[1 + \mathrm{CR}(e)]\vec{b} \cdot v_{e^{NC}} \geqslant \vec{b} \cdot v_e.$$

Hence

$$1 + \mathrm{CR}(e) \geqslant \max_{\vec{b} \geqslant 0} \frac{\vec{b} \cdot v_e}{L(\vec{b})},$$

where $L(\vec{b}) = \max_{e^{NC}} \vec{b} \cdot v_{e^{NC}}$.

Next, we prove the converse direction based on the linear program (6). Recall that all noncontextual models are convex combinations of deterministic noncontextual models and the columns of the incidence matrix M (and the rows of M^T) are the vectors corresponding to these deterministic models. In view of the dual linear program (4), there exists an optimal vector $\vec{b}_* \geqslant 0$ such that $\vec{b}_* \cdot v_e = 1 + \mathrm{CR}(e)$ and $M^T \vec{b}_* \leqslant \vec{1}$, where $M^T \vec{b} \leqslant \vec{1}$ implies that $L(\vec{b}_*) \leqslant 1$. Thus, for the positive generalized Bell inequality $\vec{b}_*$,

$$\frac{\vec{b}_* \cdot v_e}{L(\vec{b}_*)} \geqslant 1 + \mathrm{CR}(e).$$

Therefore, we obtain the result. ■

Proof of Theorem 3. Due to the definition of contextual robustness, there exists a noncontextual model e^{NC} such that $[1 + \mathrm{CR}(e)]e^{NC} \geqslant e$. Then for the l_2-MBQCs $\langle K, e^{NC}\rangle$, which use the some classical processing K, we have

$$[1 + \mathrm{CR}(e)][[K, e^{NC}]](\vec{i}) \geqslant [[K, e]](\vec{i}),$$

for any $\vec{i} \in \mathbb{Z}_2^l$. Hence

$$[1 + \mathrm{CR}(e)]\bar{p}_s^{\langle K, e^{NC}\rangle, f} \geqslant \bar{p}_s^{\langle K, e\rangle, f}.$$

Besides, it has been proved that $1 - \bar{p}_s \geqslant \tilde{\nu}(f)$ [21], which is equivalent to $\bar{p}_s^{\langle K, e^{NC}\rangle, f} \leqslant 1 - \tilde{\nu}(f)$. Thus

$$[1 + \mathrm{CR}(e)][1 - \tilde{\nu}(f)] \geqslant \bar{p}_s^{\langle K, e\rangle, f}.$$

[1] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[2] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Rev. Mod. Phys. **82**, 665 (2010).

[3] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).

[4] A. Cabello, Phys. Rev. Lett. **101**, 210401 (2008).

[5] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Phys. Rev. Lett. **101**, 020403 (2008).

[6] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos, Nature (London) **460**, 494 (2009).
[7] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, Nature (London) **474**, 490 (2011).
[8] R. Ramanathan and P. Horodecki, Phys. Rev. Lett. **112**, 040404 (2014).
[9] A. Cabello, M. Kleinmann, and C. Budroni, Phys. Rev. Lett. **114**, 250402 (2015).
[10] J. Anders and D. E. Browne, Phys. Rev. Lett. **102**, 050502 (2009).
[11] R. Raussendorf, Phys. Rev. A **88**, 022322 (2013).
[12] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Nature (London) **510**, 351 (2014).
[13] N. Delfosse, P. Allard Guerin, J. Bian, and R. Raussendorf, Phys. Rev. X **5**, 021003 (2015).
[14] R. Raussendorf, D. E. Browne, N. Delfosse, C. Okay, and J. Bermejo-Vega, Phys. Rev. A **95**, 052334 (2017).
[15] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf, Phys. Rev. Lett. **119**, 120505 (2017).
[16] A. Karanjai, J. J. Wallman, and S. D. Bartlett, arXiv:1802.07744.
[17] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurel, arXiv:1905.05374.
[18] M. Kleinmann, O. Gühne, J. R. Portillo, J.-Å. Larsson, and A. Cabello, New J. Phys. **13**, 113011 (2011).
[19] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik, Phys. Rev. Lett. **112**, 120401 (2014).
[20] K. Horodecki, A. Grudka, P. Joshi, W. Kłobus, and J. Łodyga, Phys. Rev. A **92**, 032104 (2015).
[21] S. Abramsky, R. S. Barbosa, and S. Mansfield, Phys. Rev. Lett. **119**, 050504 (2017).
[22] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[23] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992).
[24] A. K. Pati, Phys. Rev. A **63**, 014302 (2000).
[25] C. H. Bennett, D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, and W. K. Wootters, Phys. Rev. Lett. **87**, 077902 (2001).
[26] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).
[27] T. Baumgratz, M. Cramer, and M. B. Plenio, Phys. Rev. Lett. **113**, 140401 (2014).
[28] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Rev. Mod. Phys. **79**, 555 (2007).
[29] G. Gour and R. W. Spekkens, New J. Phys. **10**, 033023 (2008).
[30] I. Marvian, Ph.D. thesis, University of Waterloo, 2012.
[31] I. Marvian and R. W. Spekkens, Nat. Commun. **5**, 3821 (2014).
[32] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Phys. Rev. Lett. **111**, 250404 (2013).
[33] R. Gallego and L. Aolita, Phys. Rev. X **5**, 041008 (2015).
[34] A. Cabello, S. Severini, and A. Winter, Phys. Rev. Lett. **112**, 040401 (2014).
[35] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, Commun. Math. Phys. **334**, 533 (2015).
[36] S. Abramsky and A. Brandenburger, New J. Phys. **13**, 113036 (2011).
[37] H. Meng, H. Cao, and W. Wang, Sci. China: Phys., Mech. Astron. **59**, 640303 (2016).
[38] B. Amaral and M. T. Cunha, arXiv:1709.04812.
[39] G. B. Dantzig and M. N. Thapa, *Linear Programming 2: Theory and Extensions*, Springer Series in Operations Research and Financial Engineering (Springer, Berlin, 2003).
[40] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, 2014, http://cvxr.com/cvx.
[41] M. Grant and S. Boyd, in *Recent Advances in Learning and Control*, edited by V. Blondel, S. Boyd, and H. Kimura, Lecture Notes in Control and Information Sciences (Springer-Verlag Limited, Berlin, 2008), pp. 95–110.
[42] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[43] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[44] M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf, Phys. Rev. Lett. **104**, 170405 (2010).
[45] C. Palazuelos and T. Vidick, J. Math. Phys. **57**, 015220 (2016).