

Quantifying contextuality via linear programming

Samson Abramsky

Rui Soares Barbosa

Department of Computer Science
University of Oxford

{samson.abramsky,rui.soares.barbosa}@cs.ox.ac.uk

Shane Mansfield

Institut de Recherche en Informatique Fondamentale
Université Paris Diderot – Paris 7

shane.mansfield@univ-paris-diderot.fr

We explore how the qualitative hierarchy of strengths of contextuality introduced in [3] relates to the quantitative notion of the non-contextual fraction of an empirical model. We show that this provides a fully general quantitative measure of contextuality, applicable to any empirical model, which is related to the maximal possible violation of generalised Bell inequalities. The calculation of this measure, as well as finding such witnessing inequalities, can be formulated as linear optimisation problems. We have implemented these methods and used them to explore the degree of non-locality of empirical models arising from local equatorial measurements on the $|\phi^+\rangle$ Bell state and the n -partite GHZ states.

1 Introduction

The sheaf theoretic framework introduced in [3] provides a unified approach to non-locality [9] and contextuality [10, 15], in which the latter is seen to be a generalisation of the former, in the general setting of no-signalling empirical models. In this setting, a natural, qualitative hierarchy of contextuality was also shown to arise, which can be succinctly defined with reference to the set $S_e(X)$ of global assignments (of outcomes to the global set of measurements X) that are consistent with a given empirical model e .

The highest level in the hierarchy is **strong contextuality**, which was shown to be equivalent to a straightforward generalisation of the notion of maximal non-locality [13] which we refer to as maximal contextuality [3, Proposition 6.3]. A maximally contextual model admits no non-degenerate decomposition into a convex combination of a non-contextual (or local) model and another model. In this sense, it is meaningful to talk about the non-contextual fraction of any no-signalling empirical model.

In this paper, we shed further light on how the notion of non-contextual fraction relates to the sheaf-theoretic description of contextuality, with strong contextuality arising naturally as the (possibilistic) limit of this notion. It can be seen that any model can be decomposed as a convex combination of a non-contextual and a strongly contextual model¹ – the proportion of this mixture serves as a quantitative grading between non-contextuality and strong contextuality. This provides a (quantitative) **measure of contextuality** which, as we will show, is related to the **maximal violation** of generalised Bell inequalities. This can be understood as providing a substantial generalisation, to empirical models in any measurement scenario, of the amount by which a (2, 2, 2) Bell model violates the CHSH inequality [11].

We show that **linear programming** methods may be used to calculate this measure of contextuality for any empirical model, as well as to find Bell inequalities for which the maximal violation is attained. We stress that these methods are fully general and apply in any measurement scenario, including of course all Bell-type scenarios. The methods described have been implemented as part of a *Mathematica* package with computational tools for analysing empirical models. Computational exploration using these tools can be useful, for example, in attempting to classify non-local states [4], which is a goal of future research.

As a demonstration of how the package works, we use it to explore the non-locality of empirical models arising from the $|\phi^+\rangle$ Bell state and the n -partite GHZ states. In this way, we find new sets of measurements on the state $|\phi^+\rangle$ which give rise to empirical models that achieve the maximum violation of the CHSH

¹This implies, in particular, that only models of these kinds may be vertices of the polytope of no-signalling models for a given scenario. We study the combinatorial structure of such no-signalling polytopes more comprehensively in [2].
Submitted to:

inequality, as well as new sets of measurements on the n -partite GHZ states that lead to similar logical proofs of non-locality.

Outline Section 2 summarises the main aspects of the sheaf-theoretic framework, introducing empirical models and the hierarchy of levels of contextuality. Section 3 discusses the idea of non-contextual fraction in analogy to that of consistent global assignments. In Section 4, the task of calculating the non-contextual fraction is phrased as a linear programming problem. In Section 5, it is shown how this measure of contextuality relates to maximal violation of general Bell inequalities, using the dual linear program. Finally, Section 6 presents the results of computational explorations of some quantum empirical models. Some of the proofs are in Appendix A; see also [16, 7].

2 Empirical models and contextuality

We briefly summarise some of the main aspects of [3].

Measurement scenarios are abstract descriptions of a set of possible experiments. A **measurement scenario** is a triple $\langle X, \mathcal{M}, O \rangle$ where: X is a finite set of allowed measurements, O is a finite set of outcomes, and \mathcal{M} is a cover of X representing the compatibility of measurements. Each $C \in \mathcal{M}$ is called a **measurement context** and corresponds to a set of measurements that can be performed together. Examples of measurement scenarios include multipartite Bell-type scenarios familiar from discussions of non-locality, Kochen-Specker configurations, and more. For example, the usual (2,2,2) Bell scenario, where two experiments, Alice and Bob, can each choose between performing one of two different measurements, say a_1 or a_2 for Alice and b_1 or b_2 for Bob, is represented as follows:

$$\begin{aligned} X &= \{a_1, a_2, b_1, b_2\} & O &= \{0, 1\} \\ \mathcal{M} &= \{\{a_1, b_1\}, \{a_1, b_2\}, \{a_2, b_1\}, \{a_2, b_2\}\}. \end{aligned}$$

The event sheaf $\mathcal{E}: \mathcal{P}(X)^{\text{op}} \rightarrow \text{Set}$ is defined by $\mathcal{E}(U) := O^U$ for each $U \subseteq X$, i.e. $\mathcal{E}(U)$ contains all functional assignments of outcomes to the measurements in U , and the maps $\mathcal{E}(U' \subseteq U): \mathcal{E}(U) \rightarrow \mathcal{E}(U')$ are the obvious restrictions.

An empirical model represents a set of empirical observations for a given measurement scenario. For each valid choice of jointly performable measurements (i.e. measurement context), it specifies the probabilities of obtaining each joint outcome.

Given a commutative semiring R , we write $\mathcal{D}_R: \text{Set} \rightarrow \text{Set}$ for the R -distribution functor, which takes a set X to the set of R -valued distributions on X . The composition $\mathcal{D}_R \cdot \mathcal{E}$ is a presheaf such that $\mathcal{D}_R \mathcal{E}(U)$ is the set of R -distributions over the joint assignments of outcomes to the measurements in U and restriction $\mathcal{D}_R \mathcal{E}(U' \subseteq U)$ is given by marginalisation of distributions. A (probabilistic) **empirical model** is then a compatible family on $\mathcal{D}_{\mathbb{R}_{\geq 0}} \mathcal{E}$ for the cover \mathcal{M} . More explicitly, it is a family of probability distributions $\{e_C \in \mathcal{D}_{\mathbb{R}_{\geq 0}} X\}_{C \in \mathcal{M}}$ whose marginals agree wherever contexts overlap, i.e.

$$\forall_{C, D \in \mathcal{M}} \cdot e_C|_{C \cap D} = e_D|_{C \cap D}.$$

This is a generalisation of the usual **no-signalling** condition. An example of an empirical model for the (2,2,2) Bell scenario mentioned above is represented in the following table of probabilities.

A	B	00	01	10	11
a_1	b_1	1/2	0	0	1/2
a_1	b_2	3/8	1/8	1/8	3/8
a_2	b_1	3/8	1/8	1/8	3/8
a_2	b_2	1/8	3/8	3/8	1/8

(1)

This model arises from quantum mechanics by choosing the following two-qubit Bell state

$$|\phi^+\rangle = \frac{|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle}{\sqrt{2}}$$

and the following local measurements available to each experimenter on their respective qubit:

$$a_1 = b_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad a_2 = b_2 = \begin{pmatrix} 0 & e^{-i\frac{\pi}{3}} \\ e^{i\frac{\pi}{3}} & 0 \end{pmatrix}.$$

Fix a measurement scenario $\langle X, \mathcal{M}, O \rangle$. Given two empirical models e and e' and a value $r \in [0, 1]$, we can define a convex sum

$$re + (1-r)e'$$

contextwise by setting $(re' + (1-r)e'')_C := re'_C + (1-r)e''_C$ for each $C \in \mathcal{M}$. It is easy to verify that compatibility is preserved by such convex combination, hence this yields a well-defined empirical model. A empirical model e is said to be **extremal** if it does not admit a non-degenerate convex decomposition, i.e. if it cannot be written as $e = re' + (1-r)e''$ with e' and e'' two distinct empirical models and $r \in (0, 1)$.

We now present the hierarchy of contextuality (and of non-locality, as a special case) introduced in [3] (see also [17, 1]).

Definition 2.1. Let e be an empirical model. For any $U \subseteq X$, we write $S_e(U)$ for the set of assignments of outcomes to the measurements in U that are **consistent** with the model e :

$$S_e(U) := \{s \in \mathcal{E}(U) \mid \forall C \in \mathcal{M}. s|_{C \cap U} \in \text{supp}(e_C|_{C \cap U})\},$$

where supp gives the support of a distribution. In particular, $S_e(X) = \{g \in \mathcal{E}(X) \mid \forall C \in \mathcal{M}. g|_C \in \text{supp } e_C\}$ is the set of global assignments consistent with the model e .

Definition 2.2. An empirical model e is said to be:

- **(probabilistically) contextual** if there the family of distributions cannot be extended to a global probability distribution on $\mathcal{E}(X)$, i.e. there is no $d \in \mathcal{D}_{\mathbb{R}_{\geq 0}} \mathcal{E}(X)$ such that $\forall C \in \mathcal{M}. d_C = e_C$.
- **logically contextual** if there is a possible local assignment that cannot be extended to global assignment consistent with e , i.e. $\exists C \in \mathcal{M}. s \in \text{supp } e_C. \forall g \in S_e(X). g|_C \neq s$.²
- **strongly contextual** if there is no global assignment consistent with e , i.e. $S_e(X) = \emptyset$.

Note that the non-contextual models are those that can be written as a convex combination of deterministic empirical models of the form δ_g for a global assignment $g \in \mathcal{E}(X)$, where δ_g is given at each context $C \in \mathcal{M}$ by the delta-distributions $(\delta_g)_C := \delta_{g|_C}$.

3 Quantifying contextuality

3.1 Global assignments consistent with a fraction of events

In dealing with possibilistic versions of contextuality (both logical contextuality and strong contextuality), an important rôle was played by the set $S_e(X)$ of consistent global assignments, comprising those global assignments whose restriction to each measurement context is an event deemed possible according to the empirical model e . By possible, one means that the probability of that assignment of outcomes is not zero. But if one wants to take into account the actual probabilities with which these restrictions occur, and so to consider global assignments that are consistent with at least a certain percentage of occurring events, then the following generalisation seems natural.

Definition 3.1. Given an empirical model e on a scenario $\langle X, O, \mathcal{M} \rangle$, for any $U \subseteq X$ and $r \in (0, 1]$ we write $S_e^{\geq r}(U)$ for the set of assignments of outcomes to the measurements in U that are consistent with (at least) a fraction r of the events of e :

$$S_e^{\geq r}(U) := \{t \in \mathcal{E}(U) \mid \forall C \in \mathcal{M}. e_C|_{C \cap U}(t|_{C \cap U}) \geq r\}.$$

²This is equivalent to saying that the possibilistic collapse of the $\{e_C\}$ is not extendable to a global possibilistic (i.e. boolean) distribution on $\mathcal{E}(X)$.

As is clear from the discussion above, we shall mostly care about the case when $U = X$,³ in which the defining expression above reduces to

$$S_e^{\geq r}(X) = \{g \in \mathcal{E}(X) \mid \forall C \in \mathcal{M}. e_C(g|_C) \geq r\} .$$

It is clear that the sets $(S_e^{\geq r}(X))_{r \in (0,1]}$ form an anti-monotone family: if $r \leq r'$ then $S_e^{\geq r}(X) \supseteq S_e^{\geq r'}(X)$. Moreover, their union recovers the usual possibilistic notion of consistent assignment:

$$\bigcup_{r \in (0,1]} S_e^{\geq r}(X) = S_e(X) . \quad (2)$$

This union can be seen as a limit towards $r = 0$. At the opposite extreme, $S_e^{\geq 1}(X)$ contains at most one assignment g , and it is non-empty exactly when the model e is deterministic, i.e. of the form δ_g .⁴

The point is that a global assignment in $S_e^{\geq r}(X)$ can explain away a fraction r of the model, as is made precise by the proposition below. This simple observation can be seen as the contextualised version of the trivial fact that if p is a distribution on a set Y and $p(y) \geq r$ for some $y \in Y$, then p can be decomposed as a mixture of the form $p = r\delta_y + (1-r)p'$ for some other distribution p' .

Proposition 3.2. *Let e be an empirical model on a measurement scenario $\langle X, O_m, \Sigma \rangle$. Given $r \in (0, 1]$ and $g \in S_e^{\geq r}(X)$ a global assignment consistent with a fraction r of events of e , then the model e has a convex decomposition of the form $e = r\delta_g + (1-r)e'$ for some other model e' .*

Proof. We use the observation above at each context, obtaining for each $\sigma \in \Sigma$ a distribution e'_σ on $\mathcal{E}(\sigma)$ satisfying

$$e_\sigma = r\delta_{g|_\sigma} + (1-r)e'_\sigma = r(\delta_g)_\sigma + (1-r)e'_\sigma .$$

The fact that both δ_g and e satisfy the compatibility (no-signalling) condition implies, by virtue of it being a linear condition, that so does this new family of distributions, making e' an empirical model. \square

3.2 Consistent subdistributions

However, the notion of $S_e^{\geq r}$ above falls short of the full rôle of $S_e(X)$ in the possibilistic case because the explanations provided are not cumulative, in the sense we now describe. If g_1 and g_2 are two (possibilistically) consistent global assignments, and so provide a possibilistic explanation for two parts of the model, then jointly they explain both parts together⁵. However, things are different in the probabilistic realm: if $g_1 \in S_e^{\geq r_1}$ and $g_2 \in S_e^{\geq r_2}$, so that g_i explains a fraction r_i of the model for $i \in \{1, 2\}$, then it is not necessarily the case that we can consider these two explanations *at the same time*, taking g_1 and g_2 together to explain a fraction $r_1 + r_2$ of the model; that is, that we can write

$$e = r_1\delta_{g_1} + r_2\delta_{g_2} + (1-r_1-r_2)e' .$$

The problem is that the parts of the model for which each global assignment provides an explanation may overlap, and hence be counted twice in the purported explanation. This problem does not arise in the possibilistic case since $1 \vee 1 = 1$, making double-counting harmless.

In order to overcome this limitation, we need to use probability subdistributions, a relaxation of the notion of distribution where the weights are allowed to add up to less than one.

³Incidentally, the set $S_e(U)$ for a general $U \subseteq X$ was used in Mansfield & Barbosa [17] to discuss contextuality of sub-models, and to construct canonical extensions of possibilistic models to scenarios with more compatibility, which were then used to transform Kochen–Specker models into equivalent Bell-type models. It remains to be investigated whether a probabilistic version of the results of that paper can be achieved using the quantitative generalisations of $S_e(U)$, namely $S_e^{\geq r}(U)$ and $C_e(U)$, presented in this section.

⁴The remarks regarding monotonicity and the recovery of the possibilistic notion also hold if one replaces X by any other subset $U \subseteq X$. As for $S_e^{\geq r}(U)$, it also contains at most one assignment $t \in \mathcal{E}(U)$, exactly when the submodel restricted to the measurements in U is deterministic, of the form δ_t .

⁵Recall the observation that if there is a possibilistic extension, then $S_e(X)$ provides such an extension since it makes no difference to throw in extra consistent assignments.

Definition 3.3. Let R be an ordered semiring. An R -**subdistribution** on a set S is a function $c: S \rightarrow R$ such that its support, $\text{supp } c = \{s \in S \mid c(s) \neq 0\}$, is finite and such that $\sum_{s \in S} c(s) \leq 1$. The value of this sum is called the **weight** of c , and denoted $\omega(c)$. The R -**subdistributions functor** $\mathcal{S}\mathcal{D}_R: \text{Set} \rightarrow \text{Set}$ assigns to a set S the set $\mathcal{S}\mathcal{D}_R S$ of R -subdistributions on S , and to a function $f: S \rightarrow T$ the function

$$\mathcal{S}\mathcal{D}_R f: \mathcal{S}\mathcal{D}_R S \rightarrow \mathcal{S}\mathcal{D}_R T :: c \mapsto \lambda t \in T. \sum_{s \in S, f(s)=t} c(s).$$

The set $\mathcal{S}\mathcal{D}_R S$ is a poset with the order defined pointwise from the order in R .

In the case $R = \mathbb{B}$, the only difference compared to $\mathcal{D}_{\mathbb{B}}$ is that the zero function is now allowed; or in terms of sets, the empty set is now allowed, thus $\mathcal{S}\mathcal{D}_{\mathbb{B}}$ is the covariant finite powerset functor. Of course, we are interested primarily in the case $R = \mathbb{R}_{\geq 0}$ of probability subdistributions.

A more informative version of S_e , behaving better in the probabilistic case, can be introduced by employing the notion of subdistribution.

Definition 3.4. Given an empirical model e on a scenario $\langle X, O_m, \Sigma \rangle$, for any $U \subseteq X$ we write $C_e(U)$ for the set of subdistributions on the set of assignments of outcomes to the measurements in U that are consistent with e :

$$C_e(U) := \{c \in \mathcal{S}\mathcal{D}_{\mathbb{R}_{\geq 0}} \mathcal{E}(U) \mid \forall c \in \mathcal{M}. \forall s \in \mathcal{E}(c \cap U). e_C|_{c \cap U}(s) \geq c|_{c \cap U}(s)\}.$$

Again, we mostly care about the case $U = X$, in which the above reduces to⁶

$$C_e(X) = \{c \in \mathcal{S}\mathcal{D}_{\mathbb{R}_{\geq 0}} \mathcal{E}(X) \mid \forall c \in \mathcal{M}. \forall s \in \mathcal{E}(c). e_C(s) \geq c|_C(s)\}.$$

Let us add two obvious remarks regarding the relationship between $C_e(U)$ and the sets discussed in Section 3.1. First, the sets $S^{\geq r}(U)$ are embedded in $C_e(U)$ by taking an assignment $t \in S^{\geq r}(U)$ to the probability subdistribution $r\delta_t$. Conversely, given a probability subdistribution $c \in C_e(U)$, any assignment $t \in \mathcal{E}(U)$ in its support belongs to the set $S^{\geq c(t)}(U)$.

There is an observation analogous to (or rather, in light of the above remarks, generalising) Proposition 3.2. The gain is that we can now use more than a single global assignment, *simultaneously*, to explain part of the model in a non-contextual way. The observation is the contextualised version of the obvious fact that if d is a distribution on a set Y and c is a subdistribution with $c \leq d$, then d can be written as a mixture of the form $\sum_{y \in Y} c(y)\delta_y + (1 - \omega(c))p'$ for some other distribution p' .

Proposition 3.5. *Let e be a (probabilistic) empirical model on a measurement scenario $\langle X, O_m, \Sigma \rangle$. Given a consistent global subprobability distribution $c \in C_e(X)$, then the model e has a convex decomposition of the form*

$$e = \left(\sum_{g \in \mathcal{E}(X)} c(g)\delta_g \right) + (1 - \omega(c))e'$$

for some other model e' .

Proof. Analogous to that of Proposition 3.2. □

3.3 Non-contextual and strongly contextual fractions

The convex decomposition in Proposition 3.5 can be rewritten as

$$\left(\sum_{g \in \mathcal{E}(X)} c(g)\delta_g \right) + (1 - \omega(c))e' = \omega(c) \left(\sum_{g \in \mathcal{E}(X)} \frac{c(g)}{\omega(c)} \delta_g \right) + (1 - \omega(c))e',$$

⁶Note that a similar definition could be written using R -subdistributions for any ordered semiring R and empirical models defined over R , in particular possibilistic empirical models. But in the possibilistic realm, $R = \mathbb{B}$, $C_e(X)$ simply consists of all the sets of consistent global assignments, i.e. all the subsets of $S_e(X)$. So, it contains the same information as $S_e(X)$ and this new definition is rather pointless. This observation hinges on the fact that there is a maximal set of consistent (global) assignments, which essentially boils down to the cumulative property discussed at the start of this section. In the probabilistic realm, $R = \mathbb{R}_{\geq 0}$, however, this is no longer the case: a maximum consistent probability subdistribution does not exist for a general empirical model.

making it clear that this is a decomposition of e into a non-contextual model, with weight $\omega(c)$, and another model, e' .

A natural question to ask of a given empirical model e is what fraction of it can be explained non-contextually, or locally. That is, what is the maximum value of the weight r over all convex decompositions of the form

$$e = re_1 + (1 - r)e_2 ,$$

where e_1 is a non-contextual model and e_2 is a (no-signalling) empirical model. This maximum is called the **non-contextual fraction** of e [3], by analogy with the terminology **local fraction** introduced earlier for models on Bell-type scenarios [13] (see also [8, 6] where the term ‘local fraction’ is actually used). In light of the observation above, this is the same as the maximum weight of a consistent subdistribution:

$$\max \{ \omega(c) \mid c \in C_e(X) \} .$$

Note that such a maximum exists, i.e. the supremum of these weights is attained. This is because, if one thinks of the subdistributions as vectors with $|\mathcal{E}(X)|$ components, they form a bounded subset of $\mathbb{R}^{|\mathcal{E}(X)|}$, and moreover $C_e(X)$ is a closed subset, a fact that is obvious from its definition (note the use of \geq , not $>$). Hence, by the Heine–Borel theorem, the set $C_e(X)$, being bounded and closed, is compact. Consequently, the real-valued weight function ω , being continuous, attains its supremum, by the extreme value theorem.

The following proposition amounts to showing that the remainder of such a maximal consistent subdistribution, or maximal non-contextual explanation, is a strongly contextual model.

Proposition 3.6. *Let $\langle X, O_m, \Sigma \rangle$ be a measurement scenario. Then any empirical model e admits a convex decomposition, $e = re^{NC} + (1 - r)e^{SC}$, into a non-contextual and a strongly contextual model.*

This result has an immediate consequence concerning extremal points of the no-signalling polytope: that they must either be non-contextual (in fact, deterministic non-contextual) or strongly contextual. We have further characterised these in [2].

The decomposition of Proposition 3.6 is not unique: there might be more than one probability subdistribution consistent with e achieving the maximal possible weight. This can happen when there is a face of the no-signalling polytope of dimension at least 1, i.e. not a vertex but at least an edge, consisting only of strongly contextual models (equivalently, whose vertices are all strongly contextual) that lies parallel to a face of the non-contextual polytope. In particular, such non-uniqueness cannot arise when the scenario is such that there are no two adjacent strongly contextual vertices of the no-signalling polytope, hence no face of the polytope consisting solely of strongly contextual models. This is the case, for example, for the (2, 2, 2) Bell-type scenario.

From Proposition 3.6, one sees that the non-contextual fraction, or maximum weight of a consistent probability subdistribution, provides a grading between non-contextuality and strong contextuality: it is equal to 1 iff the empirical model is non-contextual, and it is equal to 0 iff it is strongly contextual. This suggests that it can serve as a measure of how contextual a given model is. We shall see in Section 5 that this measure relates to the maximal violation of any Bell inequality.

4 Linear programming for contextuality

The task of finding a consistent probability subdistribution with maximum weight for a given empirical model can be phrased as a linear optimisation problem. This is based on the phrasing of contextuality in linear algebraic terms [3].

Fix a measurement scenario $\langle X, \mathcal{M}, O \rangle$. An empirical model e admits an explicit representation as a vector \mathbf{v}^e in \mathbb{R}^m , where $m = \sum_{C \in \mathcal{M}} |\mathcal{E}(C)|$, the components being indexed by the local assignments: pairs $\langle C, s \rangle$ with $C \in \mathcal{M}$ a maximal context and $s \in \mathcal{E}(C)$ an assignment of outcomes to the measurements in that context. For each such $\langle C, s \rangle$, the component $\mathbf{v}^e[\langle C, s \rangle]$ of this vector is the probability given to s by the model at context C ,

$$\mathbf{v}^e[\langle C, s \rangle] := e_C(s) .$$

These vectors can be thought of as streamlined versions of the tables used to represent empirical models: for example, the Bell–CHSH model from table (1) can be represented by the vector

$$[1/2, 0, 0, 1/2, \quad 3/8, 1/8, 1/8, 3/8, \quad 3/8, 1/8, 1/8, 3/8, \quad 1/8, 3/8, 3/8, 1/8] .$$

Similarly, a probability subdistribution (in particular, also, a probability distribution) on global assignments, $c \in \mathcal{SD}_{\mathbb{R}_{\geq 0}} \mathcal{E}(X)$, can be represented as a vector \mathbf{c} in \mathbb{R}^n with $n := |\mathcal{E}(C)|$. It is clear that the vector satisfies $\mathbf{c} \geq \mathbf{0}$, meaning that each component is non-negative, since we are taking subdistributions on $\mathbb{R}_{\geq 0}$. Moreover, the weight of the subdistribution, $\omega(c)$, is given in vector terms by the dot product $\mathbf{1} \cdot \mathbf{c}$, where $\mathbf{1}$ stands for the vector (in \mathbb{R}^n) whose coordinates are all equal to 1.

The measurement scenario also determines an **incidence matrix** that records the relationship between global assignments and local assignments. This is an $m \times n$ matrix \mathbf{M} whose components are given by

$$\mathbf{M}[\langle C, s \rangle, g] := \begin{cases} 1 & \text{if } g|_C = s \\ 0 & \text{otherwise} \end{cases} .$$

Note that the columns of this matrix, $\mathbf{M}[-, g]$, are the vectors representing the deterministic models, δ_g . On the other hand, its rows, $\mathbf{M}[\langle C, s \rangle, -]$, are the (row) vectors that indicate the global assignments that restrict to a given local assignment. So, the dot product of a row and (the vector representing) a probability subdistribution c on global assignments yields the weight that this subdistribution gives to the local assignment corresponding to the row, since

$$\mathbf{M}[\langle C, s \rangle, -] \cdot \mathbf{c} = \sum_{g \in \mathcal{E}(X), g|_C = s} c(g) = c|_C(s) .$$

The condition for c to be a global probability distribution extending the model e (as in Definition 2.2) is that, for all local assignments $s \in \mathcal{E}(C)$, this value be the probability attributed by e to the assignment s , namely $e_C(s)$. This can be phrased in terms of the vector representation as follows:

$$\mathbf{M}\mathbf{c} = \mathbf{v}^e .$$

The empirical model e is non-contextual if and only if this system of linear equations can be solved for \mathbf{c} subject to the constraint $\mathbf{c} \geq \mathbf{0}$. Similarly, the condition, given in Definition 3.4, for a subdistribution $c \in \mathcal{SD}_{\mathbb{R}_{\geq 0}} \mathcal{E}(X)$ to be consistent with a model e is that, for all local assignments $s \in \mathcal{E}(C)$, the value $c|_C(g)$ be at most $e_C(s)$. In terms of the vector representation, this corresponds to the inequality

$$\mathbf{M}\mathbf{c} \leq \mathbf{v}^e .$$

We are now in possession of all the ingredients to translate the problem of finding a probability subdistribution of maximum weight consistent with a given empirical model as the following linear programming problem:

$$\begin{array}{ll} \text{Find} & \mathbf{c} \in \mathbb{R}^n \\ \text{maximising} & \mathbf{1} \cdot \mathbf{c} \\ \text{subject to} & \mathbf{M}\mathbf{c} \leq \mathbf{v}^e \\ \text{and} & \mathbf{c} \geq \mathbf{0} \end{array} . \quad (3)$$

5 Violation of Bell inequalities

We make more precise the idea that the non-contextual fraction yields a measure of contextuality by relating it to the violation of Bell inequalities.

An **inequality** for a measurement scenario $\langle X, \mathcal{M}, \mathcal{O} \rangle$ is determined by a set of coefficients $\alpha = \{\alpha(C, s)\}_{C \in \mathcal{M}, s \in \mathcal{E}(C)}$ and a bound R . For a model e , the inequality reads as

$$\mathcal{B}_\alpha(e) \leq R ,$$

where the left-hand side is given by

$$\mathcal{B}_\alpha(e) := \sum_{C \in \mathcal{M}, s \in \mathcal{E}(C)} \alpha(C, s) e_C(s) .$$

We often refer to such inequality as the tuple $\langle \alpha, R \rangle$. We restrict our attention to the case when the bound R is non-negative as any inequality is equivalent to one with that property. It is called a **Bell inequality** if it is satisfied by any non-contextual (or local) model. If, moreover, equality is attained by some non-contextual (or local) model, the Bell inequality is said to be **tight**.

Whereas a Bell inequality establishes a limit for the value of $\mathcal{B}_\alpha(e)$ amongst non-contextual models, for a general (no-signalling) model e , this quantity is limited only by

$$\|\alpha\| := \sum_{C \in \mathcal{M}} \max \{ \alpha(C, s) \mid s \in \mathcal{E}(C) \} ,$$

since

$$\mathcal{B}_\alpha(e) = \sum_{C \in \mathcal{M}} \sum_{s \in \mathcal{E}(C)} \alpha(C, s) e_C(s) \leq \sum_{C \in \mathcal{M}} \left(\max \{ \alpha(C, s) \mid s \in \mathcal{E}(C) \} \sum_{s \in \mathcal{E}(C)} e_C(s) \right) = \|\alpha\| ,$$

where the last equality follows from the fact that each e_C is a probability distribution. Note that we are implicitly restricting our attention to inequalities $\langle \alpha, R \rangle$ where $R \leq \|\alpha\|$, since other cases are evidently uninteresting. In fact, we further assume that $R < \|\alpha\|$, in this way excluding inequalities that are trivially satisfied by all no-signalling models. This will avoid cluttering the presentation with special caveats about division by 0.

A model e violates the Bell inequality by $\max\{0, \mathcal{B}_\alpha(e) - R\}$. However, it is more appropriate to talk about the value normalised by the maximal theoretical violation, as this gives a better idea of the extent to which the model violates the inequality.

Definition 5.1. The **normalised violation** of a Bell inequality $\langle \alpha, R \rangle$ by an empirical model e is the value

$$\frac{\max\{0, \mathcal{B}_\alpha(e) - R\}}{\|\alpha\| - R} .$$

Proposition 5.2. *Let e be an empirical model with non-contextual fraction k . Then its normalised violation of any Bell inequality is at most $1 - k$ (its strongly contextual fraction).*

In fact, for any empirical model, there exists a specific Bell inequality such that this maximal violation is attained. This inequality is necessarily tight, for otherwise there would exist another inequality (with the same coefficients but a smaller bound) that would still be a Bell inequality and that would have a larger violation by the model e , contradicting the previous result. We now show that it is in fact saturated by the non-contextual fraction of the model in question.

As is detailed in the proof of the following proposition, this Bell inequality can be found using the dual linear programming problem to (3).

Proposition 5.3. *Let e be an empirical model with non-contextual fraction k . Then there is a Bell inequality whose normalised violation by e is exactly $1 - k$. Moreover, this Bell inequality is tight at the non-contextual model e^{NC} .*

Proof. Recall the linear program (3) which calculates a maximum-weight probability subdistribution on global assignments consistent with the model C . To say that e has non-contextual fraction k is to say that an optimal solution, \mathbf{c}^* , to the linear programming problem attains the value k for its objective function, i.e. $\mathbf{1} \cdot \mathbf{c}^* = k$.

Let us consider the symmetric dual problem to (3):

$$\begin{aligned} \text{Find} \quad & \mathbf{b} \in \mathbb{R}^m \\ \text{minimising} \quad & \mathbf{v}^e \cdot \mathbf{b} \\ \text{subject to} \quad & \mathbf{M}^\top \mathbf{b} \geq \mathbf{1} \\ \text{and} \quad & \mathbf{b} \geq \mathbf{0} \end{aligned} . \tag{4}$$

Given a solution \mathbf{b} to the linear constraints of this problem, we define a vector

$$\mathbf{a} := \mathbf{1} - (T - R)\mathbf{b} \tag{5}$$

where $T := |\mathcal{M}|$ is the number of maximal contexts in the scenario and R is a constant satisfying $0 \leq R < T$ (in fact, perhaps the easiest is to set $R = 0$). The idea is that the components of the vector are to be taken as the coefficients of an inequality,

$$\alpha(C, s) := \mathbf{a}[\langle C, s \rangle],$$

and R as its bound. For f an empirical model, the inequality, which reads $\mathcal{B}_\alpha(f) \leq R$, can be phrased in terms of vectors as

$$\mathbf{v}^f \cdot \mathbf{a} \leq R. \quad (6)$$

It remains to show: that a solution to (4) yields a Bell inequality; that the optimal solution yields a Bell inequality whose normalised violation by e is exactly $1 - k$; and that the non-contextual part of the model, e^{NC} , saturates this inequality. See Appendix A for the remainder of the proof. \square

6 Computational explorations

Computational tools in the form of a *Mathematica* package have been developed for:

1. calculating quantum empirical models from any state and any sets of compatible measurements;
2. calculating the incidence matrix for any measurement scenario;
3. quantifying the degree of contextuality of any empirical model using the linear programming method (3) of Section 4.
4. calculating the Bell inequality of Proposition 5.3, using the dual linear program (4).

We stress that these tools are completely general: they can be applied to any pure or mixed quantum state in any Hilbert space and to any sets of compatible observables in that space, including Bell scenarios as a special case.

6.1 Equatorial measurements on the Bell state $|\phi^+\rangle$

As an example of how the package can be used, we consider a family of empirical models that can be obtained by considering local measurements on the two-qubit Bell state

$$|\phi^+\rangle = \frac{|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle}{\sqrt{2}}$$

Recall that projective measurements on a qubit can equivalently be represented by a point on the Bloch sphere. Suppose that we allow the same two local measurements on each qubit, and that these are equatorial on the Bloch sphere, parametrised by angles ϕ_1 and ϕ_2 as in Figure 1. One such model is the Bell-CHSH model from table (1), which is obtained when

$$(\phi_1, \phi_2) = (0, \pi/3).$$

We can use the package to plot the degree of contextuality (non-contextual fraction) of the resulting models as a function of ϕ_1 and ϕ_2 (Figure 2). It is interesting to note that the Bell-CHSH model does not achieve the maximum degree of contextuality. The minima of the plot (which correspond to maximum contextuality) occur when

$$\{\phi_1, \phi_2\} \in \left\{ \left\{ \frac{\pi}{8}, \frac{5\pi}{8} \right\}, \left\{ \frac{7\pi}{8}, \frac{3\pi}{8} \right\} \right\}.$$

All of the corresponding empirical models take the form of the following table

A	B	00	01	10	11
a_1	b_1	p	$(1/2 - p)$	$(1/2 - p)$	p
a_1	b_2	$(1/2 - p)$	p	p	$(1/2 - p)$
a_2	b_1	$(1/2 - p)$	p	p	$(1/2 - p)$
a_2	b_2	$(1/2 - p)$	p	p	$(1/2 - p)$

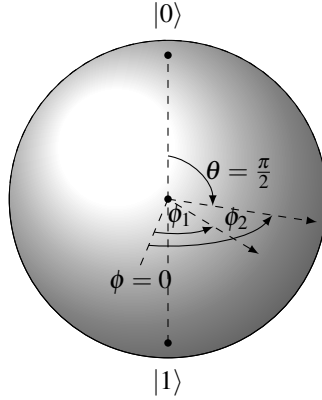


Figure 1: Equatorial measurements at ϕ_1 and ϕ_2 on the Bloch sphere.

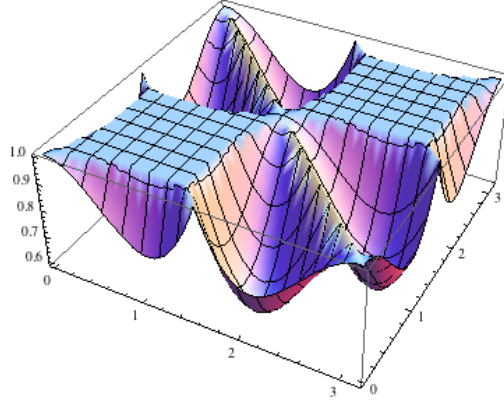


Figure 2: Contextuality of empirical models obtained with equatorial measurements at ϕ_1 and ϕ_2 on each qubit of $|\phi^+\rangle$.

where

$$p = \frac{\sqrt{2}+2}{8}.$$

These can easily be shown to achieve the Tsirelson violation of the CHSH inequality. Note that none of these models are strongly contextual: this provided one motivation to look for proofs that Bell states cannot witness logical forms of non-locality [16, 5].

It may seem surprising at first that the empirical models are not constant with respect to the relative angle $(\phi_2 - \phi_1)$ between measurements; a fact that is apparent from figure 2. For example, the empirical model obtained when $(\phi_1, \phi_2) = (0, \pi/4)$ is local, but if these values are shifted by $\pi/8$ the resulting model achieves the maximum violation of the CHSH inequality. Nevertheless, this must be the case since a rotation by ϕ around the Z-axis for each of the qubits is described by

$$\begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix} \otimes \begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix} = \begin{pmatrix} e^{-i\phi} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\phi} \end{pmatrix} \quad (7)$$

and thus introduces a relative phase of 2ϕ between the terms in $|\phi^+\rangle$.

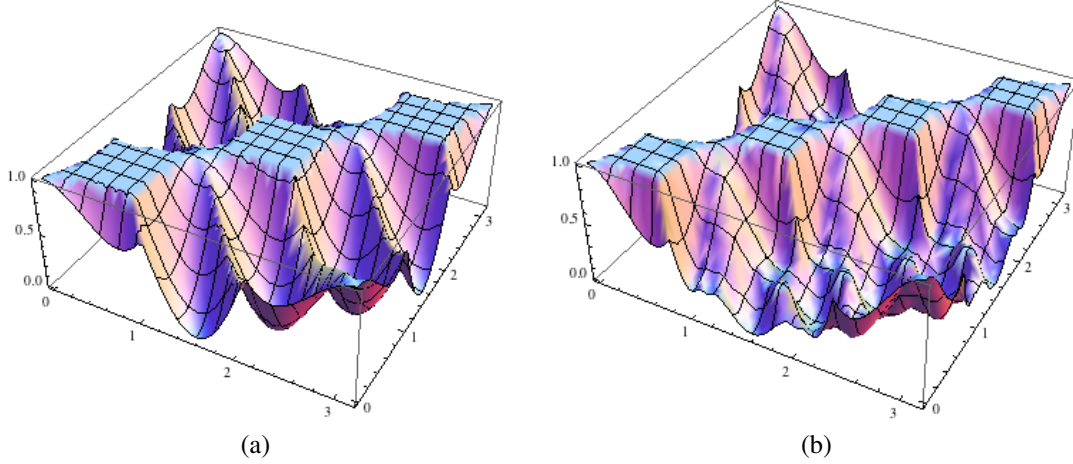


Figure 3: Contextuality of empirical models obtained with equatorial measurements at ϕ_1 and ϕ_2 on each qubit of $|\psi_{\text{GHZ}(n)}\rangle$ with: (a) $n = 3$; (b) $n = 4$.

6.2 Equatorial measurements on n -partite GHZ states

We can consider similar families of models for the n -partite GHZ states [14], given for each $n > 2$ by:

$$|\psi_{\text{GHZ}(n)}\rangle = \frac{|\uparrow\rangle^{\otimes n} + |\downarrow\rangle^{\otimes n}}{\sqrt{2}} \quad (8)$$

Note that for $n = 2$ the state obtained is the $|\phi^+\rangle$ Bell state. For $n > 2$, Mermin considered models in which each party can perform Pauli X or Y measurements to provide logical proofs of non-locality [18].

Again, we allow the same two local measurements on each qubit and assume that these are equatorial on the Bloch sphere. For $|\psi_{\text{GHZ}(3)}\rangle$ and $|\psi_{\text{GHZ}(4)}\rangle$ we obtain the plots shown in figure 3. The minima of the plot for the tripartite state reach 0, indicating strong contextuality, and occur when

$$\{\phi_1, \phi_2\} \in \left\{ \left\{ \frac{\pi}{2}, 0 \right\}, \left\{ \frac{2\pi}{3}, \frac{\pi}{6} \right\}, \left\{ \frac{5\pi}{6}, \frac{\pi}{3} \right\} \right\}. \quad (9)$$

Of course, $(\phi_1, \phi_2) = (\pi/2, 0)$ correspond to the Pauli measurements Y and X , yielding the usual GHZ model. The empirical models corresponding to other minima are identical up to re-labelling, so these provide alternative sets of measurements that can be made on the GHZ state and still lead to the familiar parity argument for non-locality [18]. The situation is similar for $n = 4$, in which minima of 0 are seen to occur at

$$\{\phi_1, \phi_2\} \in \left\{ \left\{ \frac{\pi}{2}, 0 \right\}, \left\{ \frac{5\pi}{8}, \frac{\pi}{8} \right\}, \left\{ \frac{3\pi}{4}, \frac{\pi}{4} \right\}, \left\{ \frac{7\pi}{8}, \frac{3\pi}{8} \right\} \right\}. \quad (10)$$

We can see a pattern beginning to emerge in (9) and (10), which leads to the following proposition.

Proposition 6.1. *Equatorial measurements at*

$$(\phi_1, \phi_2) \in \left\{ \left(\frac{(n+k)\pi}{2n}, \frac{k\pi}{2n} \right) \mid 0 \leq k < n \right\}$$

on each qubit of the n -partite GHZ state give rise to the strongly contextual GHZ(n) model described in [3, Section 6].

7 Conclusion

An important feature of the measure of contextuality and the associated computational tools described in this paper is that they are applicable to empirical models on any measurement scenario. The linear programming approach to finding the degree of contextuality of a model, for example, works in full generality

and provides a means of finding a general analogue of a Bell inequality witnessing contextuality given any contextual or non-local model. This is especially relevant for experimental verification of contextuality, where it can be used to ensure robustness of contextuality with respect to inaccuracies in state preparation and measurements in scenarios where the CHSH or other inequalities are not applicable.

It is also an interesting development in itself, which is worthy of further investigation. For example, similar methods can be used to define a quantitative grading of the notion of logical contextuality, if one regards it as being about the extension of a particular local assignment to a global one. Similarly, one can use $C_e(U)$ to discuss the contextuality of submodels as in [17], and this might provide a way to achieve probabilistic analogues of the results in that paper.

The tools described here have served as a useful companion to research on the sheaf-theoretic programme, as a means to calculate examples and to inform and test conjectures, such as some of the results of [4, 5]. It is hoped that they can continue to play an important rôle in guiding future results and developments. For example, these tools can be especially useful in the continuing effort to classify non-locality of states.

References

- [1] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal & Shane Mansfield (2015): **Contextuality, cohomology and paradox**. In Stephan Kreutzer, editor: *24th EACSL Annual Conference on Computer Science Logic (CSL 2015), Leibniz International Proceedings in Informatics (LIPIcs)* 41, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, pp. 211–228, doi:10.4230/LIPIcs.CSL.2015.211. Available at <http://drops.dagstuhl.de/opus/volltexte/2015/5416>. Eprint available at arXiv:1502.03097 [quant-ph].
- [2] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal & Shane Mansfield (2016): **Possibilities determine the combinatorial structure of probability polytopes**. Submitted to Journal of Mathematical Psychology, Special Issue on Foundations of Probability.
- [3] Samson Abramsky & Adam Brandenburger (2011): **The sheaf-theoretic structure of non-locality and contextuality**. *New Journal of Physics* 13(11), p. 113036, doi:10.1088/1367-2630/13/11/113036. Eprint available at arXiv:1102.0264 [quant-ph].
- [4] Samson Abramsky & Carmen Constantin (2014): **A classification of multipartite states by degree of non-locality**. In Bob Coecke & Matty Hoban, editors: *Proceedings 10th International Workshop on Quantum Physics and Logic, Castelldefels (Barcelona), Spain, 17th to 19th July 2013, Electronic Proceedings in Theoretical Computer Science* 171, Open Publishing Association, pp. 10–25, doi:10.4204/EPTCS.171.2. Eprint available at arXiv:1412.5213 [quant-ph].
- [5] Samson Abramsky, Carmen M. Constantin & Shenggang Ying (2015): **Hardy is (almost) everywhere: non-locality without inequalities for almost all entangled multipartite states**. Private communication, to appear.
- [6] Leandro Aolita, Rodrigo Gallego, Antonio Acín, Andrea Chiuri, Giuseppe Vallone, Paolo Mataloni & Adán Cabello (2012): **Fully nonlocal quantum correlations**. *Physical Review A* 85(3), p. 032107, doi:10.1103/PhysRevA.85.032107. Eprint available at arXiv:1105.3598 [quant-ph].
- [7] Rui Soares Barbosa (2015): **Contextuality in quantum mechanics and beyond**. DPhil thesis, University of Oxford.
- [8] Jonathan Barrett, Adrian Kent & Stefano Pironio (2006): **Maximally nonlocal and monogamous quantum correlations**. *Physical Review Letters* 97(17), p. 170409, doi:10.1103/PhysRevLett.97.170409. Eprint available at arXiv:quant-ph/0605182.
- [9] John S. Bell (1964): **On the Einstein-Podolsky-Rosen paradox**. *Physics* 1(3), pp. 195–200.
- [10] John S. Bell (1966): **On the problem of hidden variables in quantum mechanics**. *Reviews of Modern Physics* 38(3), pp. 447–452, doi:10.1103/RevModPhys.38.447.
- [11] John F. Clauser, Michael A. Horne, Abner Shimony & Richard A. Holt (1969): **Proposed experiment to test local hidden-variable theories**. *Physical Review Letters* 23(15), pp. 880–884, doi:10.1103/PhysRevLett.23.880.
- [12] George B. Dantzig & Mukund N. Thapa (1997): **Linear programming 1**. Springer Series in Operations Research and Financial Engineering, Springer Verlag.
- [13] Avshalom C. Elitzur, Sandu Popescu & Daniel Rohrlich (1992): **Quantum nonlocality for each pair in an ensemble**. *Physics Letters A* 162(1), pp. 25–28, doi:10.1016/0375-9601(92)90952-I.

- [14] Daniel M. Greenberger, Michael A. Horne & Anton Zeilinger (1989): **Going beyond Bell's theorem**. In M. Kafatos, editor: *Bell's theorem, quantum theory, and conceptions of the universe*, Kluwer, pp. 69–72.
- [15] Simon Kochen & Ernst P. Specker (1967): **The problem of hidden variables in quantum mechanics**. *Journal of Mathematics and Mechanics* 17(1), pp. 59–87.
- [16] Shane Mansfield (2013): **The mathematical structure of non-locality and contextuality**. DPhil thesis, University of Oxford.
- [17] Shane Mansfield & Rui Soares Barbosa (2013): **Extendability in the sheaf-theoretic approach: Construction of Bell models from Kochen-Specker models**. In: *(Pre)Proceedings of 10th Workshop on Quantum Physics and Logic (QPL X), ICFo Barcelona*. Eprint available at arXiv:1402.4827 [quant-ph].
- [18] N. David Mermin (1990): **Extreme quantum entanglement in a superposition of macroscopically distinct states**. *Physical Review Letters* 65(15), pp. 1838–1840.

A Proofs

Proof of Proposition 3.6. Let c be a probability subdistribution in $C_e(X)$ with maximal weight, and use Proposition 3.5 to get a decomposition of e of the form

$$e = \omega(c) \left(\sum_{g \in \mathcal{E}(X)} \frac{c(g)}{\omega(c)} \delta_g \right) + (1 - \omega(c))e'.$$

We show that e' is strongly contextual. Suppose for a contradiction that it has a consistent global assignment $h \in S_{e'}(X)$. Then, by equation (2), there is an $r \in (0, 1]$ such that $h \in S_{e'}^{\geq r}(X)$. But then $c' := c + (1 - \omega(c))r\delta_h$ is a probability subdistribution consistent with e , i.e. $c' \in C_e(X)$. This is because for any $\sigma \in \Sigma$,

$$\begin{aligned} & e_\sigma \\ = & \quad \{ \text{by the convex decomposition above} \} \\ & \left(\sum_{g \in \mathcal{E}(X)} c(g) \delta_g |_\sigma \right) + (1 - \omega(c))e'_\sigma \\ \geq & \quad \{ \text{since } h \in S_{e'}^{\geq r}(X), \text{ we have } e' = r\delta_h + (1-r)e'' \text{ by Proposition 3.2} \} \\ & \left(\sum_{g \in \mathcal{E}(X)} c(g) \delta_g |_\sigma \right) + (1 - \omega(c))r\delta_h \\ = & c'_\sigma \end{aligned}$$

This contradicts the maximality of the subdistribution c in $C_e(X)$, implying that there can be no such h , i.e. that $S_{e'}(X)$ is empty. \square

Proof of Proposition 5.2. This follows from the decomposition of e into a non-contextual and a strongly contextual models

$$e = ke^{NC} + (1-k)e^{SC}.$$

The left-hand side of the Bell inequality adds up to

$$\begin{aligned}
& \mathcal{B}_\alpha(e) \\
= & \quad \{ \text{using the decomposition above and the obvious linearity of } \mathcal{B} \} \\
& k\mathcal{B}_\alpha(e^{NC}) + (1-k)\mathcal{B}_\alpha(e^{SC}) \\
\leq & \quad \{ \text{since } \mathcal{B}_\alpha(e^{NC}) \leq R \text{ by non-contextuality of } e^{NC}, \text{ and } \mathcal{B}_\alpha(e^{SC}) \leq \|\alpha\| \} \\
& kR + (1-k)\|\alpha\| \\
= & \\
& kR + (1-k)R + (1-k)(\|\alpha\| - R) \\
= & \\
& R + (1-k)(\|\alpha\| - R).
\end{aligned}$$

□

Continuation of proof of Proposition 5.3. Let us first show that this is a Bell inequality, i.e. that it is satisfied by all non-contextual models. It suffices to show it for the deterministic models, δ_g with $g \in \mathcal{E}(X)$. Recalling that the columns of \mathbf{M} (and so the rows of \mathbf{M}^\top) are exactly the vectors representing these models, and bearing equation (6) in mind, the fact that $\langle \alpha, R \rangle$ is a Bell inequality is concisely expressed by the system of linear inequalities

$$\mathbf{M}^\top \mathbf{a} \leq R \mathbf{1}.$$

Let us see why this follows from the fact that \mathbf{b} is a solution to the linear programming problem (4).

$$\begin{aligned}
& \mathbf{M}^\top \mathbf{a} \leq R \mathbf{1} \\
\Leftrightarrow & \quad \{ \text{definition of } \mathbf{a}, \text{ equation (5)} \} \\
& \mathbf{M}^\top (\mathbf{1} - (T-R)\mathbf{b}) \leq R \mathbf{1} \\
\Leftrightarrow & \quad \{ \text{linearity} \} \\
& \mathbf{M}^\top \mathbf{1} - (T-R)\mathbf{M}^\top \mathbf{b} \leq R \mathbf{1} \\
\Leftrightarrow & \quad \{ \mathbf{M}^\top \mathbf{1} = T \mathbf{1} \text{ since each row of } \mathbf{M}^\top \text{ has exactly a 1 entry per maximal context} \} \\
& T \mathbf{1} - (T-R)\mathbf{M}^\top \mathbf{b} \leq R \mathbf{1} \\
\Leftrightarrow & \quad \{ \text{basic algebra} \} \\
& (T-R)\mathbf{M}^\top \mathbf{b} \geq (T-R) \mathbf{1} \\
\Leftrightarrow & \quad \{ R < T, \text{ hence } T-R \text{ is positive} \} \\
& \mathbf{M}^\top \mathbf{b} \geq \mathbf{1} \\
\Leftrightarrow & \quad \{ \mathbf{b} \text{ is a solution to the LP problem (4)} \} \\
& \text{true}
\end{aligned}$$

Moreover, the non-negativity of \mathbf{b} translates to a bound on the components of \mathbf{a} :

$$\begin{aligned}
& \mathbf{a} \leq \mathbf{1} \\
\Leftrightarrow & \quad \{ \text{definition of } \mathbf{a}, \text{ equation (5)} \} \\
& \mathbf{1} - (T-R)\mathbf{b} \leq \mathbf{1} \\
\Leftrightarrow & \\
& (T-R)\mathbf{b} \geq \mathbf{0} \\
\Leftrightarrow & \quad \{ R < T, \text{ hence } T-R \text{ is positive} \}
\end{aligned}$$

$$\begin{aligned}
& \mathbf{b} \geq \mathbf{0} \\
\Leftrightarrow & \quad \{ \mathbf{b} \text{ is a solution to the LP problem (4) } \} \\
& \text{true}
\end{aligned}$$

This has as an immediate consequence a bound on the maximal violation of the inequality α represented by the vector \mathbf{a} , namely

$$\|\alpha\| = \sum_{C \in \mathcal{M}} \max \{ \mathbf{a}[\langle C, s \rangle] \mid s \in \mathcal{E}(C) \} \leq \sum_{C \in \mathcal{M}} 1 = T.$$

Now, we know that the primal LP problem, (3), has an optimal solution \mathbf{c}^* . The strong duality theorem for linear programming (see e.g. [12]) then says that the dual LP problem, (4), also admits an optimal solution \mathbf{b}^* , and moreover that these two optimal solutions are related by

$$\mathbf{v}^e \cdot \mathbf{b}^* = \mathbf{1} \cdot \mathbf{c}^*. \quad (11)$$

Using this fact, one can calculate the value attained by the Bell inequality $\langle \alpha^*, R \rangle$ represented by the vector \mathbf{a}^* that in turn corresponds to the optimal solution \mathbf{b}^* via equation (5).

$$\begin{aligned}
& \mathcal{B}_{\alpha^*}(e) \\
= & \quad \{ \text{vector representation} \} \\
& \mathbf{v}^e \cdot \mathbf{a}^* \\
= & \quad \{ \text{definition of } \mathbf{a}^*, \text{ as in (5), and linearity} \} \\
& \mathbf{v}^e \cdot \mathbf{1} - (T - R) \mathbf{v}^e \cdot \mathbf{b}^* \\
= & \quad \{ \mathbf{v}^e \cdot \mathbf{b}^* = \mathbf{1} \cdot \mathbf{c}^* = k, \text{ the non-contextual fraction} \} \\
& \mathbf{v}^e \cdot \mathbf{1} - k(T - R) \\
= & \quad \{ \mathbf{v}^e \cdot \mathbf{1} = T \text{ for any model, since } \mathbf{v}^e \text{ has } T \text{ probability distributions} \} \\
& T - k(T - R) \\
= & \quad \{ \text{arithmetic (adding } R - R) \} \\
& R + (1 - k)(T - R) \\
\geq & \quad \{ T \geq \|\alpha^*\| \text{ as seen above and } 1 - k \geq 0 \} \\
& R + (1 - k)(\|\alpha^*\| - R)
\end{aligned}$$

This shows that the normalised violation of the inequality by the model e is at least $1 - k$. Since the opposite inequality follows from Proposition 5.2, this concludes the proof that the model e attains a normalised violation of $1 - k$ of the Bell inequality $\langle \alpha^*, R \rangle$.

Now, let us show that the non-contextual part of the model, e^{NC} , saturates this inequality. Recall that e^{NC} is the model determined by the probability subdistribution represented by the vector \mathbf{c}^* that was an optimal solution to the primal LP. As such, it is represented as $1/k \mathbf{M} \mathbf{c}^*$, where the scalar serves to normalise by $k = \mathbf{1} \cdot \mathbf{c}^*$. By virtue of being non-contextual, we know that this model satisfies the inequality, i.e. that

$$\left(\frac{1}{k} \mathbf{M} \mathbf{c}^* \right) \cdot \mathbf{a}^* \leq R.$$

We thus need to show the opposite inequality. We have:

$$\begin{aligned}
& \left(\frac{1}{k} \mathbf{M} \mathbf{c}^* \right) \cdot \mathbf{a}^* \geq R \\
\Leftrightarrow & (\mathbf{M} \mathbf{c}^*) \cdot \mathbf{a}^* \geq kR \\
\Leftrightarrow & \{ \text{definition of } \mathbf{a}^*, \text{ as in (5), and linearity} \} \\
& (\mathbf{M} \mathbf{c}^*) \cdot \mathbf{1} - (T - R)(\mathbf{M} \mathbf{c}^*) \cdot \mathbf{b}^* \geq kR \\
\Leftrightarrow & \{ (\mathbf{M} \mathbf{c}^*) \cdot \mathbf{1} = kT, \text{ since the vector } (\mathbf{M} \mathbf{c}^*) \text{ has weight } k \text{ at each context} \} \\
& (T - R)k - (T - R)(\mathbf{M} \mathbf{c}^*) \cdot \mathbf{b}^* \geq 0 \\
\Leftrightarrow & \{ T - R > 0 \text{ since } R < T \} \\
& k - (\mathbf{M} \mathbf{c}^*) \cdot \mathbf{b}^* \geq 0 \\
\Leftrightarrow & \\
& (\mathbf{M} \mathbf{c}^*) \cdot \mathbf{b}^* \leq k \\
\Leftarrow & \{ \mathbf{c}^* \text{ and } \mathbf{b}^* \text{ are solutions to (3) and (4), resp., hence } \mathbf{M} \mathbf{c}^* \leq \mathbf{v}^e \text{ and } \mathbf{b}^* \geq 0 \} \\
& \mathbf{v}^e \cdot \mathbf{b}^* \leq k \\
\Leftarrow & \{ \text{by equation (11)} \} \\
& \text{true}
\end{aligned}$$

□

Proof of Proposition 6.1. First, we know that this holds for $k = 0$, since in that case we simply have Pauli X and Y measurements, which were the measurements prescribed for obtaining the usual n -partite GHZ models considered by Mermin [18]. Mermin gave logical non-locality proofs for n -partite generalisations of the GHZ state for all $n > 2$, where each party can perform Pauli X or Y measurements. With a little calculation, it is possible to provide a very concise description of the resulting empirical models.

Recall that the GHZ states are:

$$|\Psi_{\text{GHZ}(n)}\rangle := \frac{|\uparrow\rangle^{\otimes n} + |\downarrow\rangle^{\otimes n}}{\sqrt{2}}. \quad (12)$$

where n is the number of qubits. The eigenvectors of the X operator are

$$|0_x\rangle = \frac{|0\rangle + e^{i0}|1\rangle}{\sqrt{2}} \quad \text{and} \quad |1_x\rangle = \frac{|0\rangle + e^{i\pi}|1\rangle}{\sqrt{2}}. \quad (13)$$

The vector $|0_x\rangle$ has eigenvalue $+1$ and the vector $|1_x\rangle$ has eigenvalue -1 . These are more usually denoted $|+\rangle$ and $|-\rangle$, respectively, but we use an alternative notation to match our usual $\{0, 1\}$ labelling for outcomes. Similarly, the $+1$ and -1 eigenvectors of the Y operator are

$$|0_y\rangle = \frac{|0\rangle + e^{i\pi/2}|1\rangle}{\sqrt{2}} \quad \text{and} \quad |1_y\rangle = \frac{|0\rangle + e^{-i\pi/2}|1\rangle}{\sqrt{2}}. \quad (14)$$

The phases have been made explicit since they will play the crucial role in the following calculations. The various probabilities for the empirical model predicted by quantum mechanics can be calculated as

$$|\langle \Psi_{\text{GHZ}(n)} | v_1 \dots v_n \rangle|^2$$

where the v_i are the appropriate eigenvectors. This evaluates to

$$\left| \frac{1 + e^{i\phi}}{\sqrt{2^{n+1}}} \right|^2 = \frac{1 + \cos \phi}{2^n}, \quad (15)$$

where ϕ is the sum of the phases of the v_i . From the phases of the possible eigenvectors, (13) and (14), it is clear that we must have $\phi = l\pi/2$ for some $l \in \mathbb{Z}_4$, the four element cyclic group. For $l \equiv 0 \pmod{4}$, the probability will be $\frac{1}{\sqrt{2^{n-1}}}$; for $l \equiv 2 \pmod{4}$ the probability will be 0; and for $l \equiv 1$ or $3 \pmod{4}$ the probability will be $\frac{1}{\sqrt{2^n}}$.

We can now reduce the calculation of probabilities for any such model into a simple counting argument. If l_{0_x} is the number of $|0_x\rangle$ eigenvectors, l_{1_x} is the number of $|1_x\rangle$ eigenvectors, and so on, then

$$\begin{aligned} l &\equiv l_{0_y} + 2 \cdot l_{1_x} + 3 \cdot l_{1_y} \pmod{4} \\ &\equiv (l_{0_y} + l_{1_y}) + 2 \cdot (l_{1_x} + l_{1_y}) \pmod{4}. \end{aligned}$$

Hence, the empirical model is given as follows:

- For contexts containing an odd number of Y s, every outcome is possible with equal probability $\frac{1}{\sqrt{2^n}}$, since $l = 1$ or $3 \pmod{4}$.
- For contexts containing a number of Y s divisible by four (equal to $0 \pmod{4}$), outcomes are possible if and only if they contain an even number of 1 's. For these outcomes, $l \equiv 0 \pmod{4}$ and the probabilities are $\frac{1}{\sqrt{2^{n-1}}}$. If there were an odd number of 0 's in the outcomes then $l \equiv 2 \pmod{4}$ and the probability would be 0.
- Similarly, for contexts that contain a number of Y s equal to $2 \pmod{4}$, outcomes are possible if and only if they contain an odd number of 1 's. Again, the non-zero probabilities are $\frac{1}{\sqrt{2^{n-1}}}$.

Now consider the case $0 < k < n$. We can rotate each qubit by the phase $\bar{\phi} = k\pi/n$, so that we continue to deal with X and Y measurements. It is necessary, however, to take account of the relative phase introduced by this operation on the overall state. By generalising (7) it is clear that the state after rotations will be

$$|\Psi_{\text{GHZ}(n, \bar{\phi})}\rangle = \frac{|\uparrow\rangle^{\otimes n} + e^{i2n\bar{\phi}}|\downarrow\rangle^{\otimes n}}{\sqrt{2}}.$$

Notice that for the relevant values of $\bar{\phi}$ the relative phase vanishes and we are left with the state $|\Psi_{\text{GHZ}(n)}\rangle$ from (8). Then the probabilities of the various outcomes can simply be calculated using equation (15), as before, and it is clear that we must obtain the same strongly contextual GHZ models described above. \square