Experimentally Testable Noncontextuality Inequalities Via Fourier-Motzkin Elimination

by
Anirudh Krishna

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics - Quantum Information

Waterloo, Ontario, Canada, 2015

Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Generalized noncontextuality as defined by Spekkens is an attempt to make precise the distinction between classical and quantum theories. It is a restriction on models used to reproduce the predictions of quantum mechanics. There has been considerable progress recently in deriving experimentally robust noncontextuality inequalities. The violation of these inequalities is a sufficient condition for an experiment to not admit a generalized noncontextual model.

In this thesis, we present an algorithm to automate the derivation of noncontextuality inequalities. At the heart of this algorithm is a technique called Fourier Motzkin elimination (abbrev. FM elimination).

After a brief overview of the generalized notion of contextuality and FM elimination, we proceed to demonstrate how this algorithm works by using it to derive noncontextuality inequalities for a number of examples.

Belinfante's construction, presented first for the sake of pedagogy, is based on a simple proof by Belinfante to demonstrate the invalidity of von Neumann's notion of noncontextuality. We provide a quantum realization on a single qubit that can violate this inequality.

We then go on to discuss paradigmatic proofs of noncontextuality such as the Peres-Mermin square, Mermin's Star and the 18 vector construction and use them to derive robust noncontextuality inequalities using the algorithm.

We also show how one can generalize the noncontextuality inequalities obtained for the Peres-Mermin square to systems with continuous variables.

Acknowledgements

I would like to begin my thanking my adviser, Dr. Robert Spekkens, for his excellent supervision throughout my masters program. I am grateful to Rob for giving me the freedom to work on my own whilst patiently guiding me when I went astray. In addition to research, I also learned a lot from him about academic writing.

Next, I would like to thank everyone in Joe Emerson's group for many stimulating conversations. In particular, both Dr. Joel Wallman and Dr. Mark Howard were always available when I was struggling to understand something. Thanks Joel and Mark!

I would like to thank Ravi Kunjwal for patiently answering all my questions about contextuality and helping me make sense of what I was supposed to be doing.

A big thank you to all my friends in Waterloo for a wonderful time here.
Lastly, I would like to thank my family for their love and support.

Dedicated to my family

Contents

List of Figures

Chapter 1

Introduction

Quantum mechanics is a strange physical theory. In a typical introduction to the topic, we find abstruse machinery concerning vectors in Hilbert spaces. Upon turning the gears, we can calculate the probabilities of different events. The formalism permits us to ask questions such as, "How likely is it for that photon to be polarized vertically?".

What distinguishes quantum theory from other theories that make statistical predictions? It is not a strange theory merely because we use probabilities. There are plenty of phenomena about which we have limited information and thus resort to a probabilistic description.

Consider for instance the canonical example of a probabilistic phenomenon - a coin toss. Given a restricted amount of information about the coin, we can tell how strongly we believe the coin is going to land facing heads up. However, we do not have an entire industry dedicated to interpreting this phenomenon because it is clear, at least hypothetically, how we might construct a sensible model for it. If we knew the mass, the radius, the moment of inertia of the coin, the force applied to it when flipped and so on, it should be straightforward, using Newtonian mechanics, to calculate the trajectory of the coin and thus predict the outcome of the toss. It is in the face of a lack of knowledge that we opt to use probability theory instead.

There is as yet no consensus on what the underlying guts of quantum mechanics could be although there have been many proposals. Of late, there has been a trend to move away from specific models. Instead we take a bird's eye view and consider entire classes of possible models endowed with properties that we tend to think of as classical. In this sense, these models help us dilineate the space between classical and quantum.

It often comes as a surprise when we discover that models for quantum theory do not permit a certain classical property as they showcase the true innovation of quantum theory over a classical one. These results go under the umbrage of "no-go theorems". Most famous among these results is that of Bell [1], who showed that no model for quantum mechanics can be local. Although this seems to fly in the face of intuition we have garnered from studying special relativity, there is no contradiction at the operational level because we cannot take advantage of these processes to communicate. Yet it seems like a conspiracy theory that Nature might be nonlocal under the hood but that there is no way for us to see it. This is what is unsettling about Bell's theorem.

It is important to point out that these results are not mere exercises in philosophy. Bell's theorem kicked off the study of entanglement and buttresses many key results in quantum information theory. We can use it to quantify the cost of classical simulation complexity [12] and device-independent quantum key distribution [2, 36, 45].

Although it is the most popular, nonlocality is not the only constraint on models for quantum mechanics. In the 1960s, Kochen and Specker put forth certain natural constraints on models called noncontextuality, now referred to as KS noncontextuality. As early as the 1950s, Ernst Specker was thinking about incompatible measurements in quantum mechanics and was curious whether it would be possible that the outcomes of the measurements were somehow determinable perhaps in some deeper theory underlying quantum theory. He was apparently motivated by a question in Christian theology which asks whether God's omniscience extended to the truths of conterfactual claims. [40]

Bell had also independently arrived at the same conclusion as Kochen and Specker [9]. For this reason, KS noncontextuality is also sometimes known as BKS noncontextuality.

Since then, the notion of noncontextuality has been modified and generalized by Spekkens [41]. As will be explained below, quantum theory does not permit models that are noncontextual. Sometimes we will just say that quantum theory is contextual. What we really mean when we say that is that no model for quantum theory can be noncontextual.

In the past, physicists had considered the negativity of "quasi-probability distributions" such as the Wigner function to be a signature of non-classicality. [27, 33, 10, 22, 19]. If the distributions are everywhere positive, it implies that the quasi-probability representations are in fact bonafide probability distributions. This suggests that they can be thought of as arising from an epistemic restriction as has been illustrated in [4]. From the perspective of quantum computation, the outcomes of such experiments can be simulated efficiently on a classical computer by sampling [5, 6, 46]. It has been shown that the negativity of quasiprobability representations is unavoidable [17, 18, 42].

After the development of generalized noncontextuality, Spekkens found a deep connection between noncontextuality and the negativity of quasi-probability representations [42]. An experiment allows a noncontextual model if and only if the elements of its Wigner representation are everywhere positive.

This, together with the above result that experiments that permitted a quasiprobability representation that is everywhere positive can be efficiently simulated implies that contextuality is a necessary condition for the potential speedup that quantum computers could offer. In the context of state injection [11, 23], contextuality is a sufficient condition for universality [21].

Aside from quantum computation, contextuality has also been shown to power another information processing task called parity oblivious multiplexing [44].

An important question that has been the impetus for current research is to see whether we can devise experimental tests to verify if Nature is noncontextual. We do so by studying the sorts of correlations that can arise from a noncontextual model and comparing it with the data from the test. This is important because even if quantum theory itself was to be superseded by a more powerful theory such as a theory of quantum gravity, we will know that that theory must be contextual.

Previous approaches to experimental tests of contextuality have failed to take noise into consideration and have assumed that an experiment can be implemented ideally. The central premise of these works is that measurements in a noncontextual model are deterministic - i.e. that knowledge of the ontic state allows one to determine the outcome of the measurement. However, this assumption is not justified in the presence of noise [43].

In addition, ignoring noise in experiments leads to another problem. One of the central tenets of tests of noncontextuality is the operational equivalence of two or more procedures. Two procedures are said to be operationally equivalent if they are statistically indistinguishable (and is defined precisely below). Exact operational equivalence is never achieved in a real experiment and therefore it is not clear whether an analysis based on this assumption holds water.

For these reasons, it is not clear what to make of their conclusions and their results have been the subject of controversy.

There has been a breakthrough recently in devising experimentally robust tests for noncontextuality [29, 25]. Both papers model noisy measurements by assigning them indeterministic response functions. They construct a single inequality that is robust to noise whose violation is *sufficient* to rule out all noncontextual models underlying the statistics of that experiment.

Furthermore, Mazurek et. al. also demonstrate how to deal with the problem of inexact operational equivalence. Finally, they also performed an experiment which confirmed the violation of the inequality with a high degree of confidence. We direct the interested reader to [29] for details.

Pusey [35] found a set of necessary *and* sufficient conditions for a toy model to admit of a noncontextual model. For this toy model, 2 binary-outcome measurements are tomographically complete. [1]. However, this is not applicable to quantum theory as even the simplest quantum system, a qubit, requires 3 binary-outcome

[1] A set of measurements is said to be tomographically complete if it allows one to pin down the state of the system

measurements to be tomographically complete.

Ideally we would like to find a set of necessary and sufficient conditions for an experiment to admit of a noncontextual model in cases where quantum theory predicts violations. In this thesis, we make progress towards that ideal. We restrict our attention to a subset of the data from the experiment and find necessary and sufficient conditions for a noncontextual model on this subset. We show how to derive these conditions using some illustrative examples. Our derivation involves Fourier-Motzkin elimination [47] (abbrev. FM elimination) which has also recently been used in inferring causal structures and studying Bell inequalities [16, 15].

In this chapter, we shall provide the background material that is sufficient to understand what follows in the rest of this thesis. We begin with an introduction to generalized noncontextuality. We then proceed to provide a brief description of FM elimination.

All the work in this thesis was done together with my adviser, Dr. Spekkens. The results in the second chapter on the Belinfante construction was work done with Ravi Kunjwal and Dr. Spekkens.

The plots in this thesis were made using Mathematica 10.1. [37]

1.1 Generalized noncontextuality

As defined by Spekkens [41], generalized noncontextuality is a constraint on the ontological models used to reproduce the statistics of an operational theory. To the reader who is already familiar with the definition of noncontextuality, we note that our usage of measurement noncontextuality is equivalent to the original definition but expressed in a slightly different fashion.

An operational theory for an experiment is a way of computing the probability $p(k|\mathcal{M},\mathcal{P})$ of an outcome k given that we prepared the system using a preparation procedure $\mathcal{P}$ and measured it using a procedure $\mathcal{M}$. Both the preparation and measurement procedures are specified as lists of instructions to be performed in the laboratory.

An ontological model for an operational theory attempts to reproduce the predictions of the operational theory by modelling the physical states of the system that is being subject to the experiment. All the physical attributes of the system at any given point in time are specified by the *ontic state* of the system. We shall denote this by λ, a point in the space Λ of all possible physical states of the system.

The preparation procedure might not necessarily uniquely pin down the ontic state of the system. Instead, it might only specify the probabilities with which we are likely to find the system in an ontic state λ. This probability distribution is denoted $\mu(\lambda|\mathcal{P})$.

Similarly, the outcome k of a measurement $\mathcal{M}$ might not be completely determined by specifying the ontic state of the system (for instance, the measurement outcome k could also depend on the microscopic degrees of freedom of the measurement apparatus). We associate with every measurement $\mathcal{M}$ a response function $\xi(k|\lambda,\mathcal{M})$ which is the probability that the measurement $\mathcal{M}$ returns a value k if the system is in the ontic state λ.

For the ontological model to be consistent, we require

$$\mathrm{pr}\left\{k|\mathcal{M},\mathcal{P}\right\} = \int_{\lambda\in\Lambda} d\lambda\; \xi(k|\lambda,\mathcal{M})\mu(\lambda|\mathcal{P}) \ , \tag{1.1}$$

that is, the model reproduces the statistics of the experiment.

Let $A(k)$ be some random variable which is a function of the outcome k of the measurement $\mathcal{M}$. We shall define the operational expectation value of A as

$$\langle A\rangle_{\mathcal{P}} = \sum_k A(k)\,\mathrm{pr}\left\{k|\mathcal{M},\mathcal{P}\right\} \ , \tag{1.2}$$

and the ontic expectation value of A as

$$\langle A\rangle_{\lambda} = \sum_k A(k)\xi(k|\mathcal{M},\lambda) \ . \tag{1.3}$$

It follows from the consistency condition, Eq. (1.1), that we can write the operational expectation value of A in terms of its ontic expectation value:

$$\langle A \rangle_{\mathcal{P}} = \int_{\lambda \in \Lambda} d\lambda \, \langle A \rangle_{\lambda} \mu(\lambda|\mathcal{P}) \ . \tag{1.4}$$

Two random variables A and A' are said to be operationally equivalent if

$$\langle A \rangle_{\mathcal{P}} = \langle A' \rangle_{\mathcal{P}} \tag{1.5}$$

for all preparation procedures $\mathcal{P}$. We denote this by

$$A \simeq A' \ . \tag{1.6}$$

The assumption of measurement noncontextuality stipulates that if two random variables A and A' are operationally equivalent, then their ontological counterparts must also be equal, i.e.,

$$\langle A \rangle_{\lambda} = \langle A' \rangle_{\lambda} \tag{1.7}$$

for all $\lambda \in \Lambda$.

Similarly, two preparation procedures $\mathcal{P}$ and $\mathcal{P}'$ are said to be operationally equivalent if

$$\mathrm{pr}\left\{k|\mathcal{M},\mathcal{P}\right\} = \mathrm{pr}\left\{k|\mathcal{M},\mathcal{P}'\right\} \tag{1.8}$$

for all measurement procedures $\mathcal{M}$. We denote this by

$$\mathcal{P} \simeq \mathcal{P}' \ . \tag{1.9}$$

The assumption of preparation noncontextuality states that if two preparations are operationally equivalent, then their ontological representations must also be equivalent, i.e.

$$\mu(\lambda|\mathcal{P}) = \mu(\lambda|\mathcal{P}') \ . \tag{1.10}$$

An operational theory will be said to admit of a universally noncontextual ontological model if it admits of an ontological model that is both preparation and measurement noncontextual.

1.1.1 Kochen Specker contextuality

Prior to [41], noncontextuality often meant only measurement noncontextuality together with the assumption that the outcomes of a projective measurement were deterministic, i.e. that the response function $\xi(k|\lambda,\mathcal{M})$ would either be 0 or 1 everywhere on the ontic state space Λ. The latter assumption is called outcome determinism. Although it was not originally formulated this way, it is equivalent to the definition of noncontextuality introduced by Kochen and Specker in the 60s. Following Kunjwal and Spekkens [25], we shall refer to it in this thesis as KS noncontextuality.

At this point, the reader might ask: "Shouldn't the case where the measurement outcome is assigned a probabilistic value simply be a convex combination of deterministic assignments? If so, shouldn't KS noncontextuality suffice?".

We begin by reemphasizing that KS noncontextuality consists of *two* assumptions - measurement noncontextuality *and* outcome determinism. However, it wraps both of them together and this becomes problematic both conceptually and practically.

When we demonstrate a failure of KS noncontextuality in the lab, we are demonstrating a failure of the following logical conjuction:

$$\begin{aligned}&(\text{Operational equivalences between observables}) \wedge (\text{Measurement Noncontextuality}) \\ &\qquad \wedge (\text{Outcome determinism}) \ .\end{aligned}$$

Since we can verify the operational equivalences in the laboratory, we know that they are valid. However, it is unclear which of the two assumptions of KS contextuality led to a contradiction. By providing a toy example below, we demonstrate how we can salvage measurement noncontextuality by dropping outcome determinism.

Consider a triple of binary variables A, B and C that can take values in the set $\{0, 1\}$. Furthermore, suppose we were able to verify that these three observables were always pairwise anticorrelated. This has been depicted in fig. 1.1 below.

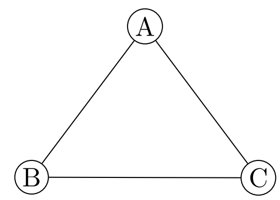


Figure 1.1: Three random variables A, B and C represented by nodes. An edge between two random variables means that they are anticorrelated.

Each node in this figure represents one of the three random variables A, B or C. Suppose the ontic state space of this construction was Λ and a point in this space was denoted λ. If we assume that the ontic states λ always assign these random variables deterministic outcomes, i.e. only 0 or 1, it is quite easy to see that there exists no assignment that can yield perfect anticorrelation for each pair. Equivalently, this could be phrased as the uncolourability of the above graph with just two colours.

On the other hand, if we assume that the ontic state λ can assign these random variables indeterministic outcomes, then the assignment

$$\langle A\rangle_\lambda = \langle B\rangle_\lambda = \langle C\rangle_\lambda = \frac{1}{2} \tag{1.11}$$

is valid while it still preserves measurement noncontextuality.

What does it mean for the ontic state to assign a value $\frac{1}{2}$ to the random variables? The interpretation is that knowing the ontic state of the system tells you nothing whatsoever about the random variables themselves but only about the correlations between them.

For a detailed analysis of the above example, see [26].

Our second critique of the assumptions that go into KS contextuality is that it seems inconsistent to assume only measurement noncontextuality and ignore preparation noncontextuality. Indeed, the latter can be seen as the same assumption, except the direction of time has been reversed. If we want to be even handed, we should assume both preparation *and* measurement noncontextuality.

We refer the interested reader to [41] for a detailed discussion of these points.

1.2 Fourier Motzkin elimination

Fourier Motzkin elimination (abbrev. FM elimination) is a technique to convert a given set of inequalities on n variables $x_1, \cdots, x_n$ to an equivalent set of inequalities on $n-1$ variables $x_2, \cdots, x_n$. We might wish to do so because we do not have access to x_1, and would like to know how the other variables are related to each other in its absence. We have explained how the algorithm works below and then provided our implementation written in python.

Suppose we are given a set of inequalities

$$\sum_{j=1}^{n} A_{ij} x_j \geq b_i \tag{1.12}$$

over n variables $x_1, \cdots, x_n$ and we want to eliminate the first variable, x_1, from this set. FM elimination constructs an equivalent set of inequalities without this variable. It takes as input the matrices A and b and outputs matrices A' and b' which specify a set of inequalities

$$\sum_{j=2}^{n} A'_{ij} x_j \geq b'_i \ . \tag{1.13}$$

We start by partitioning the inequalities into three sets. Let us write the inequalities as

$$A_{i1} x_1 + \sum_{j=2}^{n} A_{ij} x_j \geq b_i \ . \tag{1.14}$$

Define the set G to be the set of indices i for which $A_{i1} < 0$. For $g \in G$,

$$x_1 \leq \frac{\sum_{j=2}^{n} A_{gj} x_j - b_g}{|A_{g1}|} \ . \tag{1.15}$$

Similarly, define the set L to be the set of indices i for which $A_{i1} > 0$. For $l \in L$,

$$x_1 \geq \frac{b_l - \sum_{j=2}^{n} A_{lj} x_j}{A_{l1}} \ . \tag{1.16}$$

Finally, we define the set E that contains the indices i for which $A_{i1} = 0$. We do not need to eliminate x_1 from this set of inequalities as they do not contain x_1. Equations 1.16 and 1.15 state that for all $g \in G$ and all $l \in L$,

$$\frac{b_g - \sum_{j=2}^{n} A_{gj} x_j}{A_{g1}} \geq x_1 \geq \frac{\sum_{j=2}^{n} A_{lj} x_j - b_l}{|A_{l1}|} \ . \tag{1.17}$$

This is equivalent to

$$\frac{b_g - \sum_{j=2}^{n} A_{gj} x_j}{A_{g1}} \geq \frac{\sum_{j=2}^{n} A_{lj} x_j - b_l}{|A_{l1}|} \tag{1.18}$$

because we can always postulate the existence of a variable x_1 such that 1.17 is true. We can re-arrange these inequalities to write

$$\sum_{j=2}^{n} \left(A_{g1} A_{lj} + |A_{l1}| A_{gj} \right) x_j \leq |A_{l1}| b_g + A_{g1} b_l \ .$$

Together with the set of inequalities corresponding to the set E, we can write these inequalities as

$$\sum_{j=2}^{n} A'_{ij} x_j \geq b'_i \tag{1.19}$$

which is an equivalent set of linear equations without x_1. The matrices A' and b' are the output of the algorithm.

If the size of A is $O(n)$, it follows that the size of A' is $O\left(\left(\frac{n}{2}\right)^2\right)$ in the worst case scenario. The drawback of this approach is that many of the inequalities in eq. (1.18) are redundant. If we need to eliminate d variables without getting rid of the redundant inequalities, the size of A', and thus the complexity of the problem, scales as $O\left(\left(\frac{n}{4}\right)^{2^d}\right)$.

To overcome this hurdle, we have to trim the redundant inequalities after eliminating each variable. It is possible to do so using standard software.

The algorithm below was written in python based on the algorithm in [38]. This algorithm will be called as a subroutine in larger algorithms to obtain noncontextuality inequalities.

```python
import numpy as np

#This program performs Fourier-Motzkin elimination. It assumes
#that the linear inequalities are expressed as
#
#    SUM_j (A[i][j] * x[j]) - b[i] geq 0
#
#For the sake of readability, we have assumed that A[i][0] = -b[i]
#It takes as input the matrix A and an index j
#It eliminates x[j] and outputs A_out

#For simplicity, this program has been divided into two parts. The first subroutine
#constructs the matrices G, L and E and drops the column corresponding to x[j]

def Sort(A,j):
   r,c = A.shape
   G,L,E = np.array([np.ones(c)]),np.array([np.ones(c)]),np.array([np.ones(c)])
   for i in np.arange(r):
      x = np.copy(A[i])
      if A[i][j] < 0.0:
         k = np.abs(A[i][j])
         G = np.concatenate((G,[x]/k),axis=0)
      elif A[i][j] > 0.0:
         L = np.concatenate((L,[-1*x]/A[i][j]),axis=0)
      else:
         E = np.concatenate((E,[x]),axis=0)
   G = np.delete(G,0,0)
   L = np.delete(L,0,0)
   E = np.delete(E,0,0)

   G = np.delete(G,j,1)
   L = np.delete(L,j,1)
   E = np.delete(E,j,1)
   return(G,L,E)

#The actual FM elimination is done by the function FMEliminate(A,j)
def FMEliminate(A,j):
   G,L,E = Sort(A,j)
   rG,cG = G.shape
   rL,cL = L.shape
   if rL == 0:
      A_out = G
   elif rG == 0:
      A_out = L
   else:
      A_out = np.array([np.ones(cG)])
      for l in np.arange(rL):
         for g in np.arange(rG):
            A_out = np.append(A_out,[G[g] - L[l]],axis=0)
      A_out = np.delete(A_out,0,0)
   A_out = np.append(A_out,E,axis=0)
   return(A_out)
```

Chapter 2

Belinfante's construction

2.1 Introduction

In his famous textbook on quantum theory [32], von Neumann had shown that certain types of hidden variable models could not reproduce the statistics of quantum theory. At the time, it was assumed that this no-go theorem spelt the end of the hidden variables program, but three decades later, his assumptions faced a lot of critique especially from Bell [1] and Mermin [31] who dismissed his results. Following the development of generalized contextuality, Spekkens was able to return to von Neumann's assumptions and has shown [39] that they can be justified using measurement noncontextuality.

In this section, we start by considering a proof of von Neumann's no-go theorem due to Belinfante [8] and show how to adapt it into an experimentally testable noncontextuality inequality.

Von Neumann's assumptions can be summarized as follows:

1. A Hermitian operator A is associated with a unique function over the ontic states $A(\lambda)$.
2. $A(\lambda) \in \text{spectrum}(A)$
3. If $[A, B] = 0$, then $C = AB$ implies that $C(\lambda) = A(\lambda)B(\lambda)$
4. If $C = \alpha A + \beta B$ for $\alpha, \beta \in \mathbb{R}$, then $C(\lambda) = \alpha A(\lambda) + \beta B(\lambda)$.

Ballentine[3] and Belinfante [8] presented a simple proof of von Neumann's no go theorem in terms of qubits and it proceeds as follows. Define

$$\mathrm{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \mathrm{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad \mathrm{H} \equiv (X + Z)/\sqrt{2} \; . \tag{2.1}$$

The definition of H together with assumption 4 above,

$$\mathrm{H}(\lambda) = \frac{1}{\sqrt{2}} \left(\mathrm{X}(\lambda) + \mathrm{Z}(\lambda) \right) \; . \tag{2.2}$$

Given that each observable has spectrum $\{+1, -1\}$, the LHS of this equation takes its values in $\{+1, -1\}$ whereas the RHS can take values in $\{-\sqrt{2}, 0, \sqrt{2}\}$. Thus we have derived a contradiction from Von Neumann's assumptions.

In the next section, we shall begin our analysis by reformulating the construction from an operational perspective. We then discuss ontological models that could reproduce the statistics of this construction if we assume that the model is noncontextual. We shall see that the ontic assignments are constrained to lie within a polytope. This in turn implies how predictable we can make each of the measurements above.

This is the simplest possible example to understand the techniques that will be used later to obtain contextuality inequalities and is hence the first example we shall tackle. This work was done together with Rob Spekkens and Ravi Kunjwal.

2.2 An operational description of the Belinfante construction

We describe the construction from an operational standpoint because we would like it to be such that even if quantum theory were superceded by a more powerful theory tomorrow, we could conclude that the new theory must be contextual.

Consider three two-outcome measurement procedures $\{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3\}$. The outcome of measurement $\mathcal{M}_i$ is a random variable A_i that can take values in $\{+1, -1\}$. It follows that

$$-1 \leq \langle A_i \rangle_{\mathcal{P}} \leq 1 \ , \tag{2.3}$$

for all preparations $\mathcal{P}$.

We suppose that the three measurements satisfy the following operational equivalences:

$$\langle A_3 \rangle_{\mathcal{P}} = \frac{1}{\sqrt{2}} \left(\langle A_1 \rangle_{\mathcal{P}} + \langle A_2 \rangle_{\mathcal{P}} \right), \tag{2.4}$$

for all preparations $\mathcal{P}$.

Corresponding to each outcome $\alpha \in \{+1, -1\}$ of each measurement $\mathcal{M}_i$, we associate a preparation procedure $P_i^{(\alpha)}$. The reason for doing so shall become evident shortly. Out of these, we shall construct three coarse-grained preparation procedures, $\{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3\}$. $\mathcal{P}_i$ is performed by flipping a fair coin and then implementing $P_i^{(+)}$ or $P_i^{(-)}$ depending on the outcome. We require that the preparations procedures satisfy the following operational equivalences:

$$\mathcal{P}_1 \simeq \mathcal{P}_2 \simeq \mathcal{P}_3 \ . \tag{2.5}$$

In an ideal quantum realization of the Belinfante construction, the three measurements would correspond to measuring the observables Z, X and H respectively. The preparations $P_i^{(\alpha)}$ would correspond to the preparations of the α-th eigenstate of the i-th observable, e.g. $\mathcal{P}_1^{(+)}$ would prepare the quantum state $|0\rangle$, $\mathcal{P}_1^{(-)}$ would prepare the quantum state $|1\rangle$, $\mathcal{P}_2^{(+)}$ would prepare the quantum state $|+\rangle$ and so on. The operational equivalences in eqn. (2.5) follow from the fact that equal mixtures of eigenstates of any qubit observable yields the completely mixed state.

Of course, the six quantum states one assiocates with these preparations can be chosen arbitrarily so long as they obey the operational equivalences above; What distinguishes this particular association of preparations to quantum states and measurements to quantum observables is that it maximally violates the noncontextuality inequality we derive.

Furthermore, we emphasize that this construction is an operational one in the sense that it captures the consequences of noncontextuality for the statistical data one expects to see for *any* six preparations and three measurements one might implement experimentally (assuming these satisfy the appropriate operational equivalences). This is true regardless of whether this statistical data admits of a quantum explanation or not. The inequality we derive can be used to test the assumption of noncontextuality for any operational theory that obeys the operational equivalences we have stipulated above.

2.3 Ontic space geometry

As discussed in the previous chapter, we assume that the true state of affairs of the system is given by its ontic state, which we denote λ. Each ontic state assigns to each of the random variables A_i an expecation value denoted $\langle A_i \rangle_\lambda \in \mathbb{R}$.

At the outset, the three expectation values are constrained to lie within a cube

$$-1 \leq \langle A_1 \rangle_\lambda \leq 1, \tag{2.6}$$

$$-1 \leq \langle A_2 \rangle_\lambda \leq 1, \tag{2.7}$$

$$-1 \leq \langle A_3 \rangle_\lambda \leq 1 \ , \tag{2.8}$$

and are independent otherwise. This follows from eqn. (2.3) via eqn. (1.4).

However, the three ontic expectation values cannot simultaneously be made deterministic, that is, they cannot each be made to take an extremal value of ± 1. Given the operational equivalence between measurement procedures in eqn. (2.4), measurement noncontextuality implies

$$\langle A_3 \rangle_\lambda = \frac{1}{\sqrt{2}} \left(\langle A_1 \rangle_\lambda + \langle A_2 \rangle_\lambda \right), \tag{2.9}$$

for all $\lambda \in \Lambda$. This further constrains the triplet $(\langle A_1 \rangle_\lambda, \langle A_2 \rangle_\lambda, \langle A_3 \rangle_\lambda)$ to lie within a polytope contained within the cube as depicted below in figure 2.1.

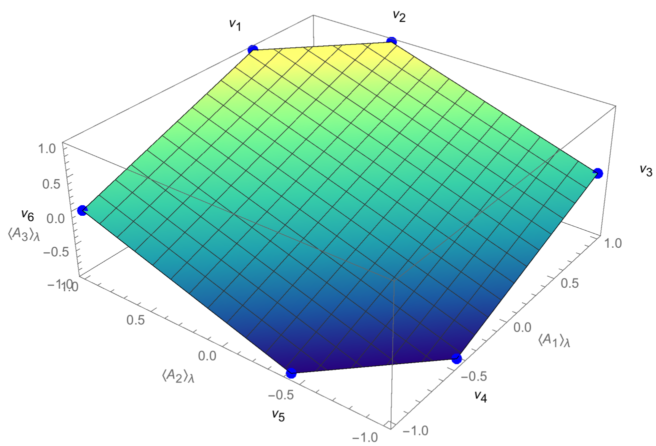


Figure 2.1: The polytope of noncontextual assignmenst to $(\langle \mathrm{A}_1 \rangle_\lambda, \langle \mathrm{A}_2 \rangle_\lambda, \langle \mathrm{A}_3 \rangle_\lambda)$

The vertices of the cube correspond to assignments of expectation values that make all the observables deterministic. The polytope of noncontextual expectation value assignments does not include any of these vertices. As we can see from the diagram, only one observable can be made deterministic for a given ontic state.

2.4 Bounds on the predictability

We define the predictability of measurement $\mathcal{M}_i$ with respect to the preparation $P_i^{(\alpha)}$ as

$$\omega_i^\alpha \equiv |\alpha \langle A_i \rangle_{P_i^{(\alpha)}}| \; . \tag{2.10}$$

The predictability is the modulus of an affine transformation of the probability of obtaining outcome $\mathrm{A}_i = \alpha$ when measuring $\mathcal{M}_i$ on the preparation $\mathcal{P}_i^{(\alpha)}$.

$$\mathrm{pr}\left\{ A_i = \alpha | \mathcal{M}_i, P_i^{(\alpha)} \right\} = \frac{1}{2} \left(1 + \alpha \langle A_i \rangle_{P_i^{(\alpha)}} \right) \; . \tag{2.11}$$

The predictability $\omega_i^{(\alpha)}$ is bounded:

$$0 \le \omega_i^{(\alpha)} \le 1 \; . \tag{2.12}$$

When it attains the value 1, it means that the outcome of the measurement $\mathcal{M}_i$ for preparation $\mathcal{P}_i^{(\alpha)}$ is either $\mathrm{A}_i = \pm\alpha$ or $\mathrm{A}_i = -1$ with certainty. In other words, the probability $\mathrm{pr}\left\{ A_i = \alpha | \mathcal{M}_i, \mathcal{P}_i^{(\alpha)} \right\}$ is either 0 or 1.

When the predictability attains the value 0, it means that the outcome is equally likely to be $\mathrm{A}_i = +1$ or $\mathrm{A}_i = -1$. Equivalently, the probability $\mathrm{pr}\left\{A_i = \alpha|\mathcal{M}_i, \mathcal{P}_i^{(\alpha)}\right\} = 1/2$.

The average predictability of $\mathcal{M}_i$ with respect to the preparation $\mathcal{P}_i$ is

$$\omega_i = \frac{1}{2}\left(\omega_i^{(+)} + \omega_i^{(-)}\right) \tag{2.13}$$

$$= \frac{1}{2}\left(|\langle A_i\rangle_{P_i^{(+)}}| + |\langle A_i\rangle_{P_i^{(-)}}|\right) . \tag{2.14}$$

It follows that the average predictability is an average measure of correlation between the input and output variables.

The average predictabilities are bounded as follows:

$$0 \leq \omega_i \leq 1 . \tag{2.15}$$

The further ω_i is from 0, the higher the predictability for the measurement $\mathcal{M}_i$.

Theorem 1. *The assumption of universal noncontextuality implies the following trade-off relation between* ω_1*,* ω_2 *and* ω_3

$$\frac{1}{3}\left(\omega_1 + \omega_2 + (2-\sqrt{2})\omega_3\right) \leq \frac{2}{3} . \tag{2.16}$$

This is an example of noncontextuality inequality that is experimentally testable and robust to noise. This inequality can be visualized as follows.

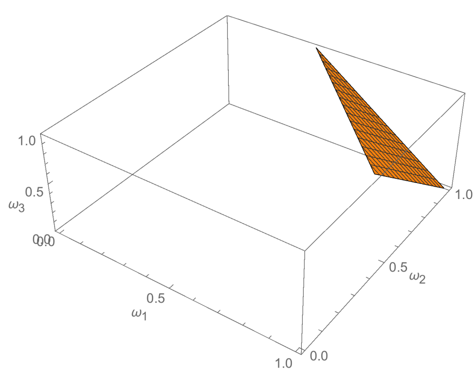


Figure 2.2: The contextuality inequality is the plane depicted in orange. Any triple of $(\omega_1, \omega_2, \omega_3)$ above that plane is not allowed.

As can be seen from the diagram, the noncontextuality inequality shaves off a corner of the set of allowed values of triples $(\omega_1, \omega_2, \omega_3)$.

We shall present the proof of this theorem in two iterations. The first iteration is a straightforward application of FM elimination. The second iteration will improve upon the first by taking advantage of the symmetries of the polytope to reduce the complexity of the problem. Although this is unnecessary for Belinfante's construction, it will become imperative in later chapters and this example will help illustrate the technique.

Proof. **Iteration 1:** We begin by expressing $\omega_i^{(\alpha)}$ in terms of an average over the ontic state space by using equation 1.4. This will permit us to upper bound $\omega_i^{(\alpha)}$.

$$\omega_i^{(\alpha)} = |\alpha\langle A_i\rangle_{P_i^{(\alpha)}}| \tag{2.17}$$

$$= |\alpha \sum_\lambda \langle A_i\rangle_\lambda \mu(\lambda|P_i^{(\alpha)})| \tag{2.18}$$

$$\leq \sum_\lambda |\langle A_i\rangle_\lambda| \mu(\lambda|P_i^{(\alpha)}) \ , \tag{2.19}$$

where the last inequality followed from a triangle-inequality.
Since the three coarse-grained preparations $\mathcal{P}_1, \mathcal{P}_2$ and $\mathcal{P}_3$ are operationally equivalent, eqn. (2.5), the assumption of preparation noncontextuality implies that

$$\mu(\lambda|\mathcal{P}_1) = \mu(\lambda|\mathcal{P}_2) = \mu(\lambda|\mathcal{P}_3) = \nu(\lambda) \ . \tag{2.20}$$

Furthermore, by definition of the coarse-grained preparations, we must have

$$\mu(\lambda|\mathcal{P}_i) = \frac{1}{2}\left(\mu(\lambda|\mathcal{P}_i^{(+)}) + \mu(\lambda|\mathcal{P}_i^{(-)})\right) \ . \tag{2.21}$$

We are thus guaranteed the existence of a distribution $\nu(\lambda)$ such that

$$\frac{1}{2}\left(\mu(\lambda|\mathcal{P}_i^{(+)}) + \mu(\lambda|\mathcal{P}_i^{(-)})\right) = \nu(\lambda) \ , \tag{2.22}$$

for all $i \in \{1, 2, 3\}$. We can use this to deduce that the average predictability can be expressed as

$$\omega_i = \frac{1}{2}\sum_\alpha \omega_i^{(\alpha)} \tag{2.23}$$

$$\leq \frac{1}{2}\sum_\alpha \sum_\lambda |\langle A_i\rangle_\lambda| \mu(\lambda|\mathcal{P}_i^{(\alpha)}) \ . \tag{2.24}$$

We can move the sum over α and use 2.22 to write

$$\omega_i \leq \sum_\lambda |\langle \mathrm{A}_i\rangle_\lambda| \nu(\lambda) \ . \tag{2.25}$$

Since the set of allowed expectation values $(\langle \mathrm{A}_1\rangle_\lambda, \langle \mathrm{A}_2\rangle_\lambda, \langle \mathrm{A}_3\rangle_\lambda)$ lies within a polytope, the vectors can be expressed as convex combinations of the triple of expectation values evaluated on the vertices. Let κ index the vertices of the polytope. There must exist a conditional distribution $\nu(\kappa|\lambda)$ such that

$$\langle \mathrm{A}_i\rangle_\lambda = \sum_\kappa \langle \mathrm{A}_i\rangle_\kappa \nu(\kappa|\lambda) \ . \tag{2.26}$$

Substituting this into the expression above, we obtain

$$\omega_i \leq \sum_\lambda \sum_\kappa \langle \mathrm{A}_i\rangle_\kappa \nu(\kappa|\lambda)\nu(\lambda) \tag{2.27}$$

$$= \sum_\kappa \langle \mathrm{A}_i\rangle_\kappa \eta(\kappa) \ , \tag{2.28}$$

where we defined the probability distribution $\{\eta(\kappa)\}$ over the vertices of the polytope as

$$\eta(\kappa) = \sum_\lambda \nu(\kappa|\lambda)\nu(\lambda) \ . \tag{2.29}$$

At this point, we shall set up some notation that we shall use for the rest of this thesis. We shall express the inequalities as

$$\omega_i \le \sum_{\kappa=1}^{6} Q_{i\kappa}\eta(\kappa) \ , \tag{2.30}$$

where

$$Q_{i\kappa} = |\langle A_i \rangle_\kappa| \ . \tag{2.31}$$

The matrix Q will change from construction to construction but we shall express bounds on the predictabilities in the form 2.30 for all of them. We would like to eliminate $\eta(\kappa)$ from these equations.

We can use brute-force FM elimination to eliminate the distribution $\eta(\kappa)$ from this set of inequalities. We shall define the vector x as

$$x \equiv (\eta(1), \cdots, \eta(6)) \oplus (\omega_1, \omega_2, \omega_3) \tag{2.32}$$

In addition to the constraints imposed by Q, we also need to ensure that η forms a bonafide probability distribution. To this end, we can define the matrices A and b such that

$$A = \begin{bmatrix} Q & -I_3 \\ \mathbf{1}_{1\times 6} & \mathbf{0}_{1\times 3} \\ -\mathbf{1}_{1\times 6} & \mathbf{0}_{1\times 3} \\ 0_{3\times 6} & I_3 \end{bmatrix} \quad b = \begin{bmatrix} \mathbf{0}_{3\times 1} \\ 1 \\ -1 \\ \mathbf{0}_{3\times 1} \end{bmatrix} \ . \tag{2.33}$$

We have used I_3 to represent the 3×3 identity matrix; e to represent a vector in $\mathbb{R}^6$ of all 1s; $\mathbf{0}_{i\times j}$ and $\mathbf{1}_{i\times j}$ to represent a matrix of all 0s and 1s respectively of size $i \times j$.

It is straightforward to verify that $Ax \ge b$. The first row of A encodes (2.30). The next two rows ensure that η is a probability distribution. The last row states that the ω_i are lower bounded by 0. We shall apply FM elimination 6 times to eliminate the probability distribution over the 6 vertices. This yields the inequalities of eqn. (2.16). Furthermore, the inequality 2.30 implies that ω_i are upper bounded by 1.

Iteration 2: In this iteration, we shall exploit the symmetry of the polytope. Looking back, we are not interested in the vector $(\langle \mathrm{A}_i \rangle_\lambda, \langle \mathrm{A}_2 \rangle_\lambda, \langle \mathrm{A}_3 \rangle_\lambda)$ but only in the modulus of its components, i.e. the vector $(|\langle \mathrm{A}_i \rangle_\lambda|, |\langle \mathrm{A}_2 \rangle_\lambda|, |\langle \mathrm{A}_3 \rangle_\lambda|$. The polytope of these vectors has been depicted below in figure 2.3 below.

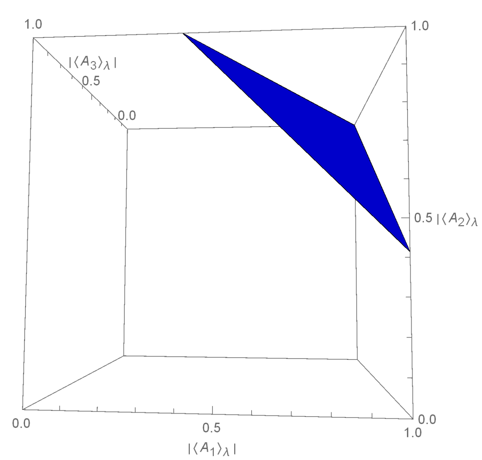


Figure 2.3: The modified polytope with only three vertices

We shall index the vertices of the new polytope using κ'. In inspecting the new polytope, we find that the modulus of the vector of expectation values $(|\langle \mathrm{A}_1 \rangle_\kappa|, |\langle \mathrm{A}_2 \rangle_\kappa|, |\langle \mathrm{A}_3 \rangle_\kappa|)$ takes equal values on vertices 1 and 6, 2 and 5 and finally 3 and 4. This implies that we can write equation 2.26 as

$$|\langle \mathrm{A}_i \rangle| = \sum_{\kappa'=1}^{3} |\langle \mathrm{A}_i \rangle_{\kappa'}| \tilde{\nu}(\kappa'|\lambda) \ . \tag{2.34}$$

Where $\tilde{\nu}(1') = \nu(1) + \nu(6)$, $\tilde{\nu}(2') = \nu(2) + \nu(5)$ and $\tilde{\nu}(3') = \nu(3) + \nu(4)$. In essence, we have formed aggregate vertices by clumping together vertices 1 and 6, 2 and 5, 3 and 4. This reduces the number of vertices by a factor of 2. Henceforth, we shall refer to this process as "aggregating vertices".

We find these vertices as follows. Consider the κ_1-th column of the Q matrix, the vector $(|\langle \mathrm{A}_1 \rangle_{\kappa_1}|, |\langle \mathrm{A}_2 \rangle_{\kappa_1}|, |\langle \mathrm{A}_3 \rangle_{\kappa_1}|)$. If there exists another vertex κ_2 such that $Q_{i\kappa_2} = (|\langle \mathrm{A}_1 \rangle_{\kappa_2}|, |\langle \mathrm{A}_2 \rangle_{\kappa_2}|, |\langle \mathrm{A}_3 \rangle_{\kappa_2}|)$ is the same as $Q_{i\kappa_1}$ then it means that κ_1 and κ_2 assign values to the random variables such that their modulus is equal. Hence, repeated columns of Q correspond to vertices that can be aggregated.

The inequalities in 2.30 will now be expressed as

$$\omega_i \leq \sum_{\kappa'=1}^{3} Q'_{i\kappa'} \tilde{\eta}(\kappa') \ , \tag{2.35}$$

where $\tilde{\eta}(\kappa') = \sum_{\lambda} \tilde{\nu}(\kappa'|\lambda) \nu(\lambda)$.

The number of inequalities and variables to be eliminated has been reduced, thereby reducing the complexity. $\square$

2.5 The algorithm

The algorithm has been written to run on SAGE, a freely available open-source software. We have provided comments below to understand the code.

```
import numpy as np

from FME import*

#Let n denote the number of observables
n = 3

M = np.eye(n + 1)

#The three ontic expectation values are defined as vectors
A_1 = M[1]
A_2 = M[2]
A_3 = M[3]

#They are constrained to lie between -1 and +1
Ineqs = [
M[0] - A_1,
M[0] - A_2,
M[0] - A_3,
M[0] + A_1,
M[0] + A_2,
M[0] + A_3
]

#Contextuality imposes a linear constraint
contextualityConstraint = [
np.sqrt(2)*A_3 - A_1 - A_2
]
```

```
#The SAGE function Polyhedron takes as input the constraints on the variables and
#returns a Polyhedron object

p = Polyhedron(ieqs = Ineqs, eqns=(contextualityConstraint), base_ring=RDF)

#The function ineqs(p,n) performs FM elimination on the vertices of the polytope and
#returns the contextuality inequalities as a matrix

def ineqs(p,n):
  A = [np.array(np.zeros(n))]
  for v in p.vertices():
    z = abs(np.array(v))
    A = np.concatenate((A,[z]),axis=0)
  A = (np.delete(A, 0, 0)).T

  r,c = np.shape(A)
  I = -1*np.eye(n)
  b = np.array([np.zeros(r)])
  A = np.concatenate((b.T,A,I),axis=1)

  e = np.concatenate((np.ones(1),np.ones(c),np.zeros(n)))
  e = np.array([e])
  f = np.concatenate((np.ones(1),-np.ones(c),np.zeros(n)))
  f = np.array([f])

  #|omega_i| are lesser than or equal to 1
      i = np.zeros(shape=(nMsmt,qr + 1))
      i[:,0] = np.ones(nMsmt)
      ii = -1*np.eye(nMsmt)
      I = np.concatenate((i,ii),axis=1)

      #And are greater than 0
      j = np.zeros(shape=(nMsmt,qr + 1))
      jj = np.eye(nMsmt)
      J = np.concatenate((j,jj),axis=1)

  A = np.concatenate((A,H,I,J,e,f),axis=0)

  p = Polyhedron(ieqs=(A), base_ring=RDF)
  A = np.array(p.inequalities())

  for i in range(c):
    F = FMEliminate(A,1)
    p = Polyhedron(ieqs=F, base_ring=RDF)
    A = np.array(p.inequalities())
  return(A)

c = ineqs(p,n)
print(c)
```

Chapter 3

The Peres-Mermin magic square

3.1 Introduction

The Peres-Mermin magic square construction in quantum theory [34, 31] consists of six triples of commuting observables on two qubits, organized into the table shown in Fig. 3.1. I, X, Y and Z are the Pauli matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.1}$$

$$Y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{3.2}$$

The notation PQ is shorthand for $P \otimes Q$. We have eliminated the tensor symbol for the sake of brevity. It can be shown that the observables on any row or column of the square commute pairwise. Furthermore, the product of the observables along each of the rows and along each of the first two columns is II. The product of the observables on the last column is $-II$.

$$\begin{array}{ccc} XI & IX & XX \\ IZ & ZI & ZZ \\ XZ & ZX & YY \end{array} \tag{3.3}$$

Figure 3.1: Observables of the Peres-Mermin Square

3.2 An operational description of the Peres-Mermin square

In order to derive a noncontextuality inequality that ultimately makes no reference to the quantum formalism, we begin by describing this construction operationally. The construction involves six measurements, labelled $\mathcal{M}_1, \ldots, \mathcal{M}_6$. Measurements $1-3$ correspond to measurements of the 3 rows of the square and measurements $4-6$ correspond to measurements of the columns. The measurement outcome can be specified by 3 binary random variables, $\mathrm{A}_i, \mathrm{B}_i$ and C_i that take values in $\{+1, -1\}$. For measurements $\{\mathcal{M}_1, \ldots, \mathcal{M}_5\}$, the triple satisfies the constraint $\mathrm{A}_i \cdot \mathrm{B}_i \cdot \mathrm{C}_i = 1$. For the 6^{th} measurement, they satisfy instead $\mathrm{A}_6 \cdot \mathrm{B}_6 \cdot \mathrm{C}_6 = -1$. The outcome of a measurement can be specified by specifying the value of any two of these variables. We will typically use A and B.

To represent these measurements and their operational equivalences, we use a hypergraph with two types of edges, of the sort introduced in [24] depicted in 3.4. A set of random variables that are defined in terms of the outcome of a single measurement are connected by an edge in the graph, depicted as a loop as in fig. 3.2.

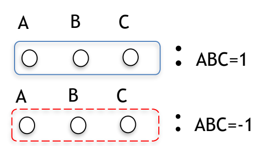


Figure 3.2: Each node represents a random variable taking values in $\{+1, -1\}$. A loop encircling a set of nodes implies that these variables can be computed from the outcome of a single measurement. Blue loops imply that the three variables have product $+1$; a red dashed loop implies that the product is -1.

Two nodes corresponding to operationally equivalent variables are connected in the hypergraph by a second type of edge, denoted by a shaded yellow loop as depicted in fig. 3.3.

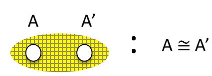


Figure 3.3: A yellow shaded loop enclosing a set of nodes implies that the expectation value of the corresponding variables are equal for all preparation procedures, i.e., the two variables are operationally equivalent.

Fig. 3.4 portrays the 18 random variables in the Peres-Mermin construction, (3 random variables for each of the 6 measurements) as well as the functional relations and operational equivalences that hold between them.

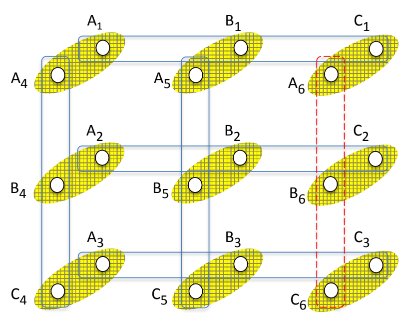


Figure 3.4: The hypergraph representing both the compatibility relations and the operational equivalences among the random variables in the Peres-Mermin square

The duality between expectation values and probabilities One might ask why we do not focus on the probability distribution over the measurement outcomes. For the Peres-Mermin hypergraph, the expectation values of the three random variables $\mathrm{A}_i, \mathrm{B}_i, \mathrm{C}_i$ fully specify the former. To wit,

$$\mathrm{pr}\,\{\mathrm{A}_i = \alpha, \mathrm{B}_i = \beta\} = \frac{1}{4}\left(1 + \alpha\langle \mathrm{A}_i\rangle + \beta\langle \mathrm{B}_i\rangle + (\alpha\beta)\langle C_i\rangle\right) \ . \tag{3.4}$$

This will be shown in the proof of lemma 2.

For each measurement $\mathcal{M}_i$, we shall be concerned with four preparation procedures $\{\mathcal{P}_i^{(\alpha\beta)}\}_{\alpha,\beta}$ for $\alpha, \beta \in \{+1, -1\}$.

In the ideal quantum realization of the Peres-Mermin hypergraph, the $\mathcal{P}_i^{(\alpha\beta)}$ might correspond to preparations of the four eigenstates of the commuting observables in the i-th measurement. For example, consider the first row of the magic-square where the observables are XI, IX and XX. The common eigenbasis is the set $|a\rangle_X \otimes |b\rangle_X$ where $a, b \in \{+,-\}$. $\{\mathcal{P}_1^{+,+}\}$ is the preparation corresponding to the eigenvector $|{+}{+}\rangle$.

From these preparations, we shall construct a course-grained preparation $\mathcal{P}_i$ which is formed by sampling $\alpha, \beta \in \{\pm 1\}$ with equal probability and then implementing $\mathcal{P}_i^{(\alpha\beta)}$. These coarse-grained preparations obey the following operational equivalence relations

$$\mathcal{P}_i \simeq \mathcal{P}_j \qquad \forall i, j \in \{1, \ldots, 6\} \ . \tag{3.5}$$

To illustrate, we refer again to the ideal quantum realization. If we let $\{\mathcal{P}_i^{(\alpha\beta)}\}$ be the set of preparations corresponding to the eigenbasis of the i-th row or column, then the coarse-graining of all these preparations is the maximally mixed state, $\frac{1}{4} I \otimes I$, for any value of i.

3.3 The space of expectation values assigned by an ontic state

3.3.1 Measurement noncontextuality and the impossibility of deterministic assignments

By the assumption of measurement noncontextuality Eq. (1.7), for any pair A, A' such that $A \simeq A'$, we have $\langle \mathrm{A} \rangle_\lambda = \langle \mathrm{A}' \rangle_\lambda \ \forall \lambda \in \Lambda$. Therefore, in the Peres-Mermin construction, the expectation value assigned by the ontic state to A_1 in the context of measurement $\mathcal{M}_1$ must be the same as the expectation value assigned by the ontic state to A_4 in the context of measurement $\mathcal{M}_4$, that is $\langle \mathrm{A}_1 \rangle_\lambda = \langle \mathrm{A}_4 \rangle_\lambda$. Similarly, for the other eight cases of operational equivalence.

A random variable A in measurement $\mathcal{M}$ is said to receive a *deterministic* assignment if $\langle \mathrm{A} \rangle_\lambda \in \{+1, -1\}$. If all the random variables generated by a measurement are deterministic, then the measurement itself is said to be outcome-deterministic.

It is straightforward to see that there is no noncontextual deterministic assignment to the Peres-Mermin construction. The product of the values assigned to the nine variables appearing in the three horizontal loops must be $+1$ (because for each loop, the product is $+1$), while the product of the values assigned to the nine variables appearing in the three vertical loops must be -1 (because for the first two the product is $+1$ and for the third it is -1). Of course, if one had equality for every pair of variables that are operationally equivalent, these two products would have to be equal. Therefore, if one wishes to maintain deterministic assignments these must be contextual which is to say that the values assigned to operationally equivalent variables are not equal. An example of a deterministic but contextual assignment is given in Fig. 3.5.

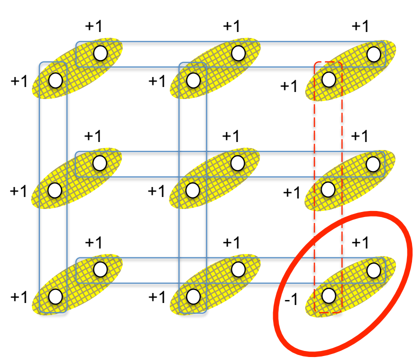


Figure 3.5: An example of a contextual assignment to the Peres-Mermin magic square is given here, with the contextuality highlighted.

It follows, therefore, that the only noncontextual assignments to these variables are assignments that are *in*deterministic for at least some of the measurements. An example of such an assignment is given in Fig. 3.6.

For this example, we have

$$(\langle \mathrm{A}_2\rangle_\lambda, \langle \mathrm{B}_2\rangle_\lambda, \langle \mathrm{C}_2\rangle_\lambda) = (+1, 0, 0) \tag{3.6}$$

which is an equal mixture of the deterministic assignments $(+1,+1,+1)$ and $(+1,-1,-1)$. Note that both the deterministic assignments obey the product relation for measurement $\mathcal{M}_2$. We can make similar statements for measurements $\mathcal{M}_3$ and $\mathcal{M}_5$. For measurement $\mathcal{M}_6$,

$$(\langle \mathrm{A}_6\rangle_\lambda, \langle \mathrm{B}_6\rangle_\lambda, \langle \mathrm{C}_6\rangle_\lambda) = (+1, 0, 0) \ , \tag{3.7}$$

but in this case it is an equal mixture of the deterministic assignments $(+1,+1,-1)$ and $(+1,-1,-1)$, each of which obeys the product relation for measurement $\mathcal{M}_6$.

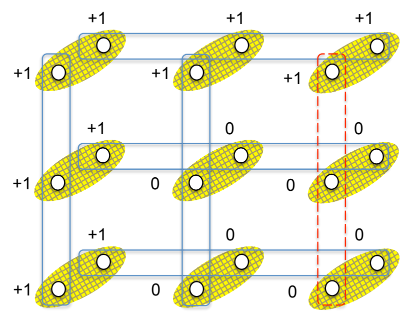


Figure 3.6: An example of an indeterministic noncontextual assignment to the variables in the Peres-Mermin square construction.

The lack of a deterministic assignment to all the random variables implies that there is a bound on how predictable the experiment is. We can see this as follows: Let $\mathcal{P}$ be some preparation whose corresponding distribution over ontic states is $\mu(\lambda|\mathcal{P})$. Suppose A_i in measurement $\mathcal{M}_i$ is indeterministic i.e.

$$-1 < \langle \mathrm{A}_i\rangle_\lambda < 1 \tag{3.8}$$

Since the operational expectation value $\langle \mathrm{A}_i\rangle_\mathcal{P}$ is

$$\langle \mathrm{A}_i\rangle_\mathcal{P} = \int d\lambda \ \langle \mathrm{A}_i\rangle_\lambda \mu(\lambda|\mathcal{P}) \tag{3.9}$$

it must be strictly between $+1$ and -1. From the definition in the previous section, the random variable cannot be perfectly predictable.

3.3.2 The polytope space

It remains to determine the set of possible indeterministic noncontextual assignments to the eighteen variables $\mathrm{A}_i, \mathrm{B}_i, \mathrm{C}_i$ for $i \in \{1, \cdots, 6\}$. To this end, it will be convenient to work with the vector space $\bigoplus_{i=1}^{6} \mathbb{R}^3$. The expectation values of the random variables will be a vector within this space.

For each measurement $\mathcal{M}_i$, we can specify the triple of expectations assigned by a hidden-variable λ as a vector $(\langle \mathrm{A}_i\rangle_\lambda, \langle \mathrm{B}_i\rangle_\lambda, \langle \mathrm{C}_i\rangle_\lambda) \in \mathbb{R}^3$

This vector lies inside a tetrahedron which can be specified by its faces (also known as the half-space representation or the H-representation of the tetrahedron).

Lemma 2. *For measurements $\mathcal{M}_1, \ldots, \mathcal{M}_5$, the expectation values of the random variables are related as* *follows:*

$$0 \leq \frac{1}{4}\left(1+\langle \mathrm{A}_i\rangle_\lambda+\langle \mathrm{B}_i\rangle_\lambda+\langle \mathrm{C}_i\rangle_\lambda\right) \leq 1 \tag{3.10}$$

$$0 \leq \frac{1}{4}\left(1+\langle \mathrm{A}_i\rangle_\lambda-\langle \mathrm{B}_i\rangle_\lambda-\langle \mathrm{C}_i\rangle_\lambda\right) \leq 1 \tag{3.11}$$

$$0 \leq \frac{1}{4}\left(1-\langle \mathrm{A}_i\rangle_\lambda+\langle \mathrm{B}_i\rangle_\lambda-\langle \mathrm{C}_i\rangle_\lambda\right) \leq 1 \tag{3.12}$$

$$0 \leq \frac{1}{4}\left(1-\langle \mathrm{A}_i\rangle_\lambda-\langle \mathrm{B}_i\rangle_\lambda+\langle \mathrm{C}_i\rangle_\lambda\right) \leq 1. \tag{3.13}$$

For measurement $\mathcal{M}_6$, the expectation values of the random variables obey

$$0 \leq \frac{1}{4}\left(1+\langle \mathrm{A}_6\rangle_\lambda+\langle \mathrm{B}_6\rangle_\lambda-\langle \mathrm{C}_6\rangle_\lambda\right) \leq 1 \tag{3.14}$$

$$0 \leq \frac{1}{4}\left(1+\langle \mathrm{A}_6\rangle_\lambda-\langle \mathrm{B}_6\rangle_\lambda+\langle \mathrm{C}_6\rangle_\lambda\right) \leq 1 \tag{3.15}$$

$$0 \leq \frac{1}{4}\left(1-\langle \mathrm{A}_6\rangle_\lambda+\langle \mathrm{B}_6\rangle_\lambda+\langle \mathrm{C}_6\rangle_\lambda\right) \leq 1 \tag{3.16}$$

$$0 \leq \frac{1}{4}\left(1-\langle \mathrm{A}_6\rangle_\lambda-\langle \mathrm{B}_6\rangle_\lambda-\langle \mathrm{C}_6\rangle_\lambda\right) \leq 1. \tag{3.17}$$

The proof illustrates the duality between probabilities and expectations that we alluded to earlier. This duality will also be useful later.

Proof. Let the four outcomes of any of the six measurements be labelled by the values of A and B which are in the set $\{++, +-, -+, --\}$.

We begin by considering measurements $\mathcal{M}_1, \ldots, \mathcal{M}_5$. The random variables associated with these measurements satisfy $\mathrm{A}_i \cdot \mathrm{B}_i \cdot \mathrm{C}_i = 1$. The expectation values of the random variables A, B and C in a measurement can be expressed in terms of the probabilities $\xi\{\mathrm{A}=\alpha; \mathrm{B}=\beta|\mathcal{M}_i, \lambda\}$ which, for the sake of brevity, are written as $\xi\{\alpha\beta\}$.

$$\langle \mathrm{A}\rangle_\lambda = \xi\{++\}+\xi\{+-\}-\xi\{-+\}-\xi\{--\} \tag{3.18}$$

$$\langle \mathrm{B}\rangle_\lambda = \xi\{++\}-\xi\{+-\}+\xi\{-+\}-\xi\{--\} \tag{3.19}$$

$$\langle \mathrm{C}\rangle_\lambda = \xi\{++\}-\xi\{+-\}-\xi\{-+\}+\xi\{--\}\ . \tag{3.20}$$

Together with the normalization rule for probabilities, this relation can be inverted to write down the probabilities in terms of the expectation values. The probability $\xi\{(\alpha\beta)\}$ can be expressed succinctly in terms of the expectation values as

$$\xi\{\alpha\beta\} = \frac{1}{4}\left(1+\alpha\langle \mathrm{A}\rangle+\beta\langle \mathrm{B}\rangle+\alpha\beta\langle \mathrm{C}\rangle\right)\ . \tag{3.21}$$

The inequalities follow from the fact that

$$0 \leq \xi\{\alpha\beta\} \leq 1 \tag{3.22}$$

for all α, β.

Finally we consider measurement $\mathcal{M}_6$. The random variables associated with the measurement satisfy a slightly different constraint, $\mathrm{A}_6 \cdot \mathrm{B}_6 \cdot \mathrm{C}_6 = -1$. We can derive the corresponding inequalities following a similar line of reasoning. □

Equivalently, we could have specified the tetrahedron by specifying its vertices (which is known as the vertex representation or the V-representation). For the first five measurements, the vertices of the tetrahedron are

$$\vec{\mathrm{V}}_{+1,+1} = (+1,+1,+1) \tag{3.23}$$

$$\vec{\mathrm{V}}_{+1,-1} = (+1,-1,-1) \tag{3.24}$$

$$\vec{\mathrm{V}}_{-1,+1} = (-1,+1,-1) \tag{3.25}$$

$$\vec{\mathrm{V}}_{-1,-1} = (-1,-1,+1) \tag{3.26}$$

which correspond to deterministic assignments. These vectors can be parametrized as

$$\vec{\mathrm{V}}_{(\alpha\beta)} = (\alpha, \beta, (\alpha\beta)) \tag{3.27}$$

for $\alpha, \beta \in \{+1, -1\}$. For measurement $\mathcal{M}_6$, the vertices of the tetrahedron are specified by the vectors

$$\vec{\mathrm{U}}_{+1,+1} = (+1, +1, -1) \tag{3.28}$$
$$\vec{\mathrm{U}}_{+1,-1} = (+1, -1, +1) \tag{3.29}$$
$$\vec{\mathrm{U}}_{-1,+1} = (-1, +1, +1) \tag{3.30}$$
$$\vec{\mathrm{U}}_{-1,-1} = (-1, -1, -1) \ . \tag{3.31}$$

They too can be parametrized in a similar fashion

$$\vec{\mathrm{U}}_{(\alpha\beta)} = (\alpha, \beta, -(\alpha\beta)) \ . \tag{3.32}$$

for $\alpha, \beta \in \{+1, -1\}$.

Thus, the vector $\bigoplus_{i=1}^{6} (\langle A_i \rangle_\lambda, \langle B_i \rangle_\lambda, \langle C_i \rangle_\lambda)$ lies inside a polytope in $\bigoplus_{i=1}^{6} \mathbb{R}^3$. The inequalities in lemma 2 can be thought of as "local" constraints within each 3-dimensional subspace. In addition to these, the assumption of measurement noncontextuality constrains them further through a "global" constraint. For instance, the constraint

$$\langle \mathrm{A}_1 \rangle_\lambda = \langle \mathrm{A}_4 \rangle_\lambda \tag{3.33}$$

relates the first component of the first subspace with the first component of the fourth subspace. Since the 9 contextuality constraints are linear, the space of allowed expectation values is still a polytope.

3.4 Bounds on the predictability

In this section, we will define a measure of the predictability.

The *predictability* of the measurement $\mathcal{M}_i$ with respect to the preparation $\mathcal{P}_i^{(\alpha\beta)}$ is

$$\omega_i^{(\alpha\beta)} = \frac{1}{3} \left| \left(\alpha \langle \mathrm{A}_i \rangle_{\mathcal{P}_i^{(\alpha\beta)}} + \beta \langle \mathrm{B}_i \rangle_{\mathcal{P}_i^{(\alpha\beta)}} + \alpha\beta \langle \mathrm{C}_i \rangle_{\mathcal{P}_i^{(\alpha\beta)}} \right) \right| \ . \tag{3.34}$$

Recall that the preparation $\mathcal{P}_i$ is the uniform mixture of the 4 preparations $\mathcal{P}_i^{(\alpha\beta)}$. The predictability of the measurement $\mathcal{M}_i$ with respect to the preparation $\mathcal{P}_i$ is

$$\omega_i = \frac{1}{4} \sum_{\alpha,\beta \in \{+1,-1\}} \omega_i^{(\alpha\beta)} \ . \tag{3.35}$$

The ω_i are bounded:

$$0 \leq \omega_i \leq 1 \ . \tag{3.36}$$

The further ω_i is from 0, the higher the predictability of measurement $\mathcal{M}_i$.

Let us continue with the ideal quantum realization from earlier. Suppose we considered the first row of the Peres-Mermin hypergraph. Recall the preparation $\{\mathcal{P}_1^{(\alpha\beta)}\}$ prepares the eigenstates of the operators XI, IX and XX. Consider the preparation $\mathcal{P}_1^{(++)}$ which prepares the state $|{++}\rangle$.

The predictability of $\mathcal{P}_1^{(++)}$ with respect to measurement $\mathcal{M}_1$ is

$$\omega_1^{(++)} = 1 \ . \tag{3.37}$$

The predictability of the average preparation procedure $\mathcal{P}_1$ with respect to measurement $\mathcal{M}_1$ is

$$\omega_1 = 1 \ . \tag{3.38}$$

Indeed, we have $\omega_i = 1$ for all $i \in \{1, \cdots, 6\}$. However, for a universally noncontextual ontological model, the fact that the only measurement-noncontextual assignments are indeterministic implies, through the assumption of preparation noncontextuality, that the predictability of our six measurements cannot all be arbitrarily high, even when we cater the preparation to the measurement. Formally, this can be stated as follows.

Theorem 3. *Consider an operational theory that admits of a universally noncontextual ontological model. Let $\{\mathcal{M}_i : i \in \{1, \ldots, 6\}\}$ be six four-outcome measurements which obey the equivalences shown in fig. 3.4. Let $\{\mathcal{P}_i^{(\alpha\beta)} : i \in \{1, \ldots, 6\}, \ \alpha, \beta \in \{+1, -1\}\}$ be twenty-four primary preparation procedures, organized into six sets of four according to the value of i. Let $\mathcal{P}_i$ be the average of the four preparations in set labelled by i. Suppose we find that the preparation procedures obey the equivalences*

$$\mathcal{P}_i \simeq \mathcal{P}_j \qquad \forall i, j \in \{1, \ldots, 6\} \ . \tag{3.39}$$

Then the predictabilities are constrained as follows:

$$\begin{aligned}
\frac{1}{2}(\omega_1 + \omega_2) &\leq \frac{2}{3} & \frac{1}{2}(\omega_4 + \omega_5) &\leq \frac{2}{3} \\
\frac{1}{2}(\omega_1 + \omega_3) &\leq \frac{2}{3} & \frac{1}{2}(\omega_4 + \omega_6) &\leq \frac{2}{3} \\
\frac{1}{2}(\omega_2 + \omega_3) &\leq \frac{2}{3} & \frac{1}{2}(\omega_5 + \omega_6) &\leq \frac{2}{3} \\
\frac{1}{3}(\omega_1 + \omega_2 + \omega_3) &\leq \frac{5}{9} & \frac{1}{3}(\omega_4 + \omega_5 + \omega_6) &\leq \frac{5}{9}
\end{aligned}$$

The bounds on the left are on the predictabilities of the measurements associated with the rows whereas those on the right are bounds on the predictabilities associated with the columns.

The noncontextual assignment in figure 3.6 saturates the inequalities. However, the ideal quantum realization of the Peres-Mermin construction violates them. As an example, we shall demonstrate how to violate the inequalities on the left. Recall the preparation $\mathcal{P}_i^{(\alpha\beta)}$ corresponded to the $(\alpha\beta)$-th eigenvector of the i-th row of the magic square and achieves a predictability of 1. Furthermore, it is easy to see that the preparations satisfy the equivalences demanded of them - the average of the 4 preparations along each row yields the maximally mixed state on two-qubits. Each of the three predictabilities ω_1, ω_2 and ω_3 is 1 and hence all the bounds on the left are violated.

We now provide the proof of the theorem 3.

Proof. Albeit a little more complicated, the proof is conceptually identical to that of Belinfante's construction.

Consider the predictability $\omega_i^{(\alpha\beta)}$ for measurement $\mathcal{M}_i$. By expressing it in terms of the ontic expectation values, we can obtain an upper bound on them as follows:

$$\begin{aligned}
\omega_i^{(\alpha\beta)} &= \frac{1}{3}\left|\left(\alpha\langle \mathrm{A}_i\rangle_{\mathcal{P}_i^{(\alpha\beta)}} + \beta\langle \mathrm{B}_i\rangle_{\mathcal{P}_i^{(\alpha\beta)}} + \alpha\beta\langle \mathrm{C}_i\rangle_{\mathcal{P}_i^{(\alpha\beta)}}\right)\right| && (3.40)\\
&= \frac{1}{3}\left|\sum_\lambda \left(\alpha\langle \mathrm{A}_i\rangle_\lambda + \beta\langle \mathrm{B}_i\rangle_\lambda + (\alpha\beta)\langle \mathrm{C}_i\rangle_\lambda\right)\mu(\lambda|\mathcal{P}_i^{(\alpha\beta)})\right| && (3.41)\\
&\leq \frac{1}{3}\sum_\lambda \left(|\langle \mathrm{A}_i\rangle_\lambda| + |\langle \mathrm{B}_i\rangle_\lambda| + |\langle \mathrm{C}_i\rangle_\lambda|\right)\mu(\lambda|\mathcal{P}_i^{(\alpha\beta)}) \ . && (3.42)
\end{aligned}$$

The last step follows from the triangle inequality.

This permits us to upper bound the average predictability ω_i

$$\omega_i = \sum_{\alpha,\beta} \omega_i^{(\alpha\beta)} \tag{3.43}$$

$$\leq \frac{1}{12} \sum_{\lambda} (|\langle \mathrm{A}_i \rangle_\lambda| + |\langle \mathrm{B}_i \rangle_\lambda| + |\langle \mathrm{C}_i \rangle_\lambda|) \sum_{\alpha,\beta} \mu(\lambda|\mathcal{P}_i^{(\alpha\beta)}) \ . \tag{3.44}$$

At this stage, we shall invoke preparation noncontextuality. From the equivalence relations between the preparations and the assumption of preparation noncontextuality, we can deduce that for all $i, j \in \{1, \cdots, 6\}$,

$$\frac{1}{4} \sum_{\alpha,\beta} \mu(\lambda|\mathcal{P}_i^{(\alpha\beta)}) = \frac{1}{4} \sum_{\alpha,\beta} \mu(\lambda|\mathcal{P}_j^{(\alpha\beta)}) \equiv \nu(\lambda) \ , \tag{3.45}$$

where $\nu(\lambda)$ is a distribution independent of the choice of measurement.

Substituting this into the upper-bound for ω_i, we get

$$\omega_i \leq \sum_{\lambda} (|\langle \mathrm{A}_i \rangle_\lambda| + |\langle \mathrm{B}_i \rangle_\lambda| + |\langle \mathrm{C}_i \rangle_\lambda|) \, \nu(\lambda) \ . \tag{3.46}$$

As noted in section III, the set of valid expectation value assignments lies inside a polytope. Therefore, the expectation value assignments $(\langle A_i \rangle_\lambda, \langle B_i \rangle_\lambda, \langle C_i \rangle_\lambda)$ can be expressed as a convex combination of the expectation value assignments on the vertices of the polytope. Let κ be a variable that indexes the vertices so that we may write

$$(\langle A_i \rangle_\lambda, \langle B_i \rangle_\lambda, \langle C_i \rangle_\lambda) = \sum_{\kappa} \vartheta(\kappa|\lambda) \, (\langle A_i \rangle_\kappa, \langle B_i \rangle_\kappa, \langle C_i \rangle_\kappa) \ , \tag{3.47}$$

where $\vartheta(\lambda|\kappa)$ is a conditional distribution. Substituting this into the expression above, we obtain

$$\sum_{\lambda} \sum_{\kappa} (|\langle \mathrm{A}_i \rangle_\lambda| + |\langle \mathrm{B}_i \rangle_\lambda| + |\langle \mathrm{C}_i \rangle_\lambda|) \, \vartheta(\lambda|\kappa) \nu(\lambda) \tag{3.48}$$

$$= \sum_{\kappa} (|\langle \mathrm{A}_i \rangle_\kappa| + |\langle \mathrm{B}_i \rangle_\kappa| + |\langle \mathrm{C}_i \rangle_\kappa|) \, \eta(\kappa) \ , \tag{3.49}$$

where $\eta(\kappa) = \sum_\lambda \vartheta(\lambda|\kappa)\nu(\lambda)$. Thus it is possible to re-write the above bound on ω_i as

$$\omega_i \leq \sum_{\kappa} (|\langle \mathrm{A}_i \rangle_\kappa| + |\langle \mathrm{B}_i \rangle_\kappa| + |\langle \mathrm{C}_i \rangle_\kappa|) \, \eta(\kappa) \ , \tag{3.50}$$

We would now like to eliminate the variables $\{\eta(\kappa)|\kappa$ is a vertex of the polytope $\}$ from the above set of inequalities. As noted in the introduction, FM elimination is a standard algorithm for this task.

Unfortunately, brute-force FM elimination alone did not work for this task. As noted in the section 1.2, the complexity of the problem scales doubly exponentially in the number of vertices n to be eliminated. Given the number of inequalities we have to deal with, it was infeasible with the computational resources we have available. To get around this problem, we used the trick explained in the previous chapter. Let us represent the above set of inequalities as

$$\omega_i \leq \sum_{\kappa} Q_{i\kappa} \eta(\kappa) \ , \tag{3.51}$$

where $Q(i\kappa) = (|\langle \mathrm{A}_i \rangle_\kappa| + |\langle \mathrm{B}_i \rangle_\kappa| + |\langle \mathrm{C}_i \rangle_\kappa|)$. To simplify the problem, we again consider aggregating vertices. As in the Belinfante construction, this leads to a new polytope with fewer vertices.

Let us label the matrix obtained from this process as Q' and index the vertices of the new polytope by κ'. We have the set of inequalities

$$\omega_i \leq \sum_{\kappa'} Q_{i\kappa'} \tilde{\eta}(\kappa') \ . \tag{3.52}$$

FM elimination on the matrix Q' is found to be computationally feasible to obtain the inequalities described in the statement of the theorem. □

The key feature of the magic square that led to a restriction of the predictability is the relation $\mathrm{A}_6 \cdot \mathrm{B}_6 \cdot \mathrm{C}_6 = -1$. If we replace this constraint by $\mathrm{A}_6 \cdot \mathrm{B}_6 \cdot \mathrm{C}_6 = +1$ (this would be equivalent to replacing the red loop with a blue loop in fig 3.4) the predictability achievable in a noncontextual ontological model attains the maximum value of 1. Such a relation holds in Spekkens' toy theory.

At this point, a reader who is familiar with proofs of contextuality might ask whether the proof of contextuality above is a state-dependent proof or a state-independent proof. However, as we elaborate in the appendix, this distinction is not meaningful under the assumption of the generalized notion of contextuality. We have also provided a review of previous attempts to derive noncontextuality inequalities based on the Peres-Mermin square and highlighted their shortcomings.

3.5 Continuous variable noncontextuality inequalities

In this section, we obtain noncontextuality inequalities for continuous variable systems based on the inequalities above. We do so by considering qubits encoded in modes of an oscillator and then demonstrate how to extend the Peres-Mermin construction to this setup. To this end, we employ modular variables. Modular variables were used by Massar and Pironio [28] to construct a continuous variable analogue of the GHZ proof of contextuality for quantum mechanics [31]. In particular, the modular variables allow us to define continuous variable analogues of the Pauli operators.

Alternatively, it is also possible to think of them as logical operators acting on a qubit encoded in an oscillator [20].

These inequalities are interesting because one of the major proposals for constructing fault-tolerant quantum computers using optical elements involve qubits encoded in oscillators [30]. Howard et. al. [21] have shown evidence that contextuality is a resource for quantum information processing on qutrit systems. Our hope is that these inequalities may prove helpful in characterizing resources required for optical implementations of quantum computers.

3.5.1 Modular operators

Let q be the position operator and p be the momentum operator. We shall find it convenient to scale these operators. Let us define

$$\mathrm{Q} = \frac{\mathrm{q}}{\sqrt{\pi}} \quad \mathrm{P} = \frac{\mathrm{p}}{\sqrt{\pi}} \ . \tag{3.53}$$

Define the shift and boost operators as

$$\mathrm{X} \equiv \exp\left(-i\pi\,\mathrm{P}\right) \quad \mathrm{Z} \equiv \exp\left(i\pi\,\mathrm{Q}\right) \ . \tag{3.54}$$

Let $|q\rangle$ define a position eigenstate. For the sake of convenience, we shall be working with the basis $\{\|q\rangle\}_{q\in\mathbb{R}}$ where

$$\|q\rangle \equiv \left|\sqrt{\pi}q\right\rangle \ . \tag{3.55}$$

Similarly, let $\{|p)\}_{p\in\mathbb{R}}$ be a momentum eigenvector. The round brackets here are used to denote the momentum basis in contrast to the angular brackets for the position basis. We define the vector $\|p)$

$$\|p) \equiv \left|\sqrt{\pi}p\right) \ . \tag{3.56}$$

The action of the shift and boost operators can be expressed neatly in these bases.

$$\mathrm{X}\,\|p) = \exp\left(-i\pi p\right)\|p) \quad , \quad \mathrm{Z}\,\|q\rangle = \exp\left(i\pi q\right)\|q\rangle \ . \tag{3.57}$$

We can use this to deduce that these operators anti-commute

$$\mathrm{X}\,\mathrm{Z} = -\,\mathrm{Z}\,\mathrm{X} \ . \tag{3.58}$$

For a real number x, let $\bar{x}$ denote the parity of its integral part.

$$\bar{x} = \begin{cases} 0 & \lfloor x \rfloor \text{ is even} \\ 1 & \lfloor x \rfloor \text{ is odd .} \end{cases} \tag{3.59}$$

We can define the *modular* shift and boost operators as

$$\bar{\mathrm{X}} \, \| p) = (-1)^{\bar{p}} \, \| p) \qquad \bar{\mathrm{Z}} \, \| q \rangle = (-1)^{\bar{q}} \, \| q \rangle \quad . \tag{3.60}$$

These are both binary valued observables. It follows from the definition that

$$\bar{\mathrm{X}}^2 = \bar{\mathrm{Z}}^2 = \mathrm{I} \;\; . \tag{3.61}$$

For $x, z \in [0, 1)$, let us define the vectors

$$\| \bar{0} \rangle_{x,z} = \sum_{j \in \mathbb{Z}} \exp\left(\pi i z (2j)\right) \| x + 2j \rangle \tag{3.62}$$

$$\| \bar{1} \rangle_{x,z} = \sum_{j \in \mathbb{Z}} \exp\left(\pi i z \left(2j+1\right)\right) \| x + (2j+1) \rangle \quad . \tag{3.63}$$

It is easy to see that for all $x, z \in [0, 1)$,

$$\bar{\mathrm{Z}} \, \| \bar{0} \rangle_{x,z} = + \| \bar{0} \rangle_{x,z} \qquad \bar{\mathrm{X}} \, \| \bar{0} \rangle_{x,z} = \| \bar{1} \rangle_{x,z} \tag{3.64}$$

$$\bar{\mathrm{Z}} \, \| \bar{1} \rangle_{x,z} = - \| \bar{1} \rangle_{x,z} \qquad \bar{\mathrm{X}} \, \| \bar{1} \rangle_{x,z} = \| \bar{0} \rangle_{x,z} \quad . \tag{3.65}$$

Thus for fixed $x, z \in [0, 1)$, the $\| \bar{0} \rangle_{x,z}$ and $\| \bar{1} \rangle_{x,z}$ are identical to the qubit algebra.

Proposition 4. *The set* $\{ \| \bar{0} \rangle_{x,z} , \| \bar{1} \rangle_{x,z} | x, z \in [0, 1) \}$ *forms a complete orthonormal basis.*

Proof. Without loss of generality, consider a position eigenvector $\| q \rangle$ such that $\lfloor q \rfloor$ is even. It has no support along the vectors $\| \bar{1} \rangle_{x,z}$. Consider the inner product

$$\langle q | \, \| \bar{0} \rangle_{x,z} = \begin{cases} \exp i \pi z \lfloor q \rfloor & \text{if } x = \{q\} \\ 0 & \text{otherwise .} \end{cases} \tag{3.66}$$

Therefore, we may write

$$\| q \rangle = \int dz \;\; \exp i \pi z \lfloor q \rfloor \, \| \bar{0} \rangle_{\{q\},z} \quad . \tag{3.67}$$

□

From this it follows that $\bar{\mathrm{X}}$ and $\bar{\mathrm{Z}}$ anti-commute.

We note that these states are not physical - they require infinite squeezing and are not normalized.

3.5.2 The PM Square

Instead of two qubits, the proof of CV contextuality shall involve two-modes. The square construction looks very similar to the qubit case and is drawn in fig. 3.7 below.

$$\begin{array}{ccc} \bar{\mathrm{X}}\,\mathrm{I} & \mathrm{I}\,\bar{\mathrm{X}} & \bar{\mathrm{X}}\bar{\mathrm{X}} \\ \mathrm{I}\,\bar{\mathrm{Z}} & \bar{\mathrm{Z}}\,\mathrm{I} & \bar{\mathrm{Z}}\bar{\mathrm{Z}} \\ \bar{\mathrm{X}}\bar{\mathrm{Z}} & \bar{\mathrm{Z}}\bar{\mathrm{X}} & \bar{\mathrm{Y}}\bar{\mathrm{Y}} \end{array}$$

Figure 3.7: Modular operators used to construct the Peres-Mermin square

It can be shown that the product of the observables along the rows and the first two columns is $I\,I$ whereas the product of the observables along the last column is -$I\,I$.

The modular Pauli operators as defined are binary valued. Thus, the operational description of the Peres-Mermin square as described in section I applies to the one made of modular operators above. For this reason, the contextuality inequalities we derived in the previous chapter apply to this construction as well.

Now we shall demonstrate that there exists a quantum violation of the above inequality. Since the algebra of these states is the same as that for qubits, the quantum violation looks identical. For some fixed choice of $x, z \in [0, 1)$, let us choose the preparations $\mathcal{P}_i^{(\alpha\beta)}$ to prepare the $\alpha\beta$-th eigenstate of the i-th set of observables. Each of the $\omega_i^{(\alpha\beta)}$ quantities is 1 and therefore, so are the ω_i quantities.

In a real lab, it is impossible to prepare the states $\|\bar{0}\rangle_{x,z}$ and $\|\bar{1}\rangle_{x,z}$ because they require infinite squeezing and are unnormalized. It is still an open problem to find how well we can violate the inequality if we had a bound on how much we could squeeze these states.

3.6 Appendix

3.6.1 A review of previous attempts to derive noncontextuality inequalities for the Peres-Mermin square

In this section, we shall review previous proposals for experimental tests of noncontextuality based on the Peres-Mermin proof of the Kochen-Specker theorem and discuss their shortcomings. The first such proposal appears in Cabello, Filipp, Rauch and Hasegawa [14]. Cabello [13] subsequently proposed an experimental test of noncontextuality for the state-independent version of Peres-Mermin and we shall begin with a discussion of this proposal.

In this model, the expectation value assigned to a random variable by an ontic state depends only on the equivalence class of the random variable. We introduce labels for the different equivalence classes of variables in the example, as depicted in Fig. 3.8.

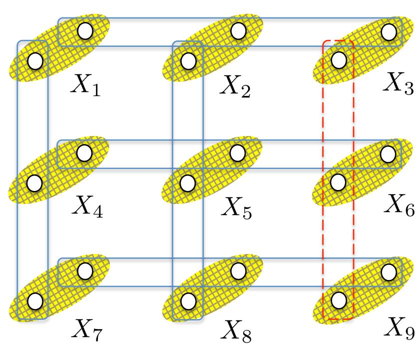


Figure 3.8: Our labelling convention for the equivalence classes of variables in the Peres-Mermin magic square construction.

In Ref [13], it is proposed that the operational quantity of interest is

$$\begin{aligned} R(P) \equiv\ & \langle X_1 X_2 X_3 \rangle_P \\ & + \langle X_4 X_5 X_6 \rangle_P \\ & + \langle X_7 X_8 X_9 \rangle_P \\ & + \langle X_1 X_4 X_7 \rangle_P \\ & + \langle X_2 X_5 X_8 \rangle_P \\ & - \langle X_3 X_6 X_9 \rangle_P, \end{aligned} \tag{3.68}$$

and that it should satisfy the following inequality in a noncontextual model

$$R(P) \leq 4 \ . \tag{3.69}$$

We dispute the claim that this bound delimits the boundary between contextual and noncontextual theories. To begin, we present the derivation of this inequality in the notation introduced here. First, we note that

$$R(P) = \sum_{\lambda} R(\lambda)\mu(\lambda|P). \tag{3.70}$$

The inequality of Eq. (3.69) is derived through an argument wherein it is claimed that in any noncontextual model,

$$R(\lambda) \leq 4 \tag{3.71}$$

and then using Eq. (3.70),

$$R(P) \leq 4 \tag{3.72}$$

The argument presented in favour of Eq. (3.71) in [13] is simply that the possible assignments of values to $(X_1, \ldots, X_9)$ by an ontic state λ are chosen from all 2^9 possible combinations of assignments of an element of $\{-1, +1\}$ to each of the nine X_i. For every such combination, one can verify that $R(\lambda) \leq 4$.

The problem with the argument is as follows. It is *by the definition* of a triple of variables such as X_1, X_2 and X_3 that we expect that any deterministic assignment to these must satisfy $X_3 = X_1X_2$. Recall that each is a coarse-graining of the variable that represents which outcome of the measurement occurred. The case of $X_3 = -X_1X_2$ is, strictly speaking, a *logical contradiction*. But each of the 2^9 different assignments of values to $(X_1, \ldots, X_9)$ implies such a logical contradiction. Indeed, this is precisely the content of the Kochen-Specker theorem for this construction.

It follows that the sort of model that a violation of inequality (3.69) rules out can already be ruled out *by logic alone*; no experiment is required. In fact, because of the definitions, as long as one has the right operational equivalences, then regardless of the operational theory, we will have $R(P) = 1$.

In [14], the authors propose a state-*dependent* proof whose reasoning is very similar to the above. Following the same reasoning as above, we dispute this claim too.

A contradiction with the assumption of a KS-noncontextual model can always be eliminated by rejecting outcome determinism and holding on to the generalized notion of noncontextuality.

As we have shown, there are novel inequalities that do not assume outcome determinism. Our inequality starts with the premise that the noncontextual model assigns indeterministic outcomes to the measurements, and then one can find assignments that do not involve any logical contradiction, such as the one in Fig.3.6.

The second response that one can make is that the proposal did not incorporate enough preparation procedures to verify the operational equivalences. The actual experiment only used a single preparation procedure P_*. But to verify operational equivalencs of measurements, one has to show that the measurements give the same statistics for a tomographically complete set of preparations.

Follow up experiments, such as the one by Bartosik *et al.* [7], consider a different inequality, which is no better. They acknowledge that $\langle X_1X_4X_7\rangle_P = 1$ and $\langle X_2X_5X_8\rangle_P = 1$ must hold for all preparations (and hence also for all λ) and so they conclude that they need not test them. (To be clear, we agree that these two equalities need not be tested, but *only if* one has already verified the operational equivalences that serve to define the Xs, which was not done in the Bartosik *et al.* experiment.) They propose the inequality

$$-\langle X_1X_2\rangle - \langle X_4X_5\rangle - \langle X_7X_8\rangle \leq 1 \ , \tag{3.73}$$

which is based on similar reasoning to 3.72. Consequently, the inequality they derive has the same conceptual issues as the ones above.

3.6.2 A criticism of the distinction between state-dependent and state-independent proofs

In the framework of the generalized version of noncontextuality, the distinction between state-dependent and state-independent proofs of noncontextuality is no longer meaningful. Indeed, to verify the operational equivalences required in theorem 3, one needs to perform a tomographically complete set of preparations so that many states are always required. If we just chose one preparation $\mathcal{P}_i^{(\alpha\beta)}$ for all our measurements then some of our measurements are necessarily going to be unpredictable and therefore we are not going to violate the inequality.

One way to salvage the state-dependent proof is as follows. One could assume that an ontic state assigns a certain expectation value to the random variables of a measurement. In effect, this will restrict us to a sub-polytope within the polytope specified by lemma 2. We can perform FM elimination within this polytope to obtain the new bounds on the predictability quantities.

To illustrate, we shall consider the following example. Consider a specific assignment to measurement $\mathcal{M}_6$ in the ontological model, for example, $(\langle A_6\rangle_\lambda, \langle B_6\rangle_\lambda, \langle C_6\rangle_\lambda) = (-1,-1,-1)$. This deterministic assignment restricts us to lie within a subset of the polytope. Performing FM elimination within this restricted polytope, we find that the inequalities for the predictabilities of the row measurements are unchanged. On the other hand, the predictabilities for the measurements along the columns are

$$\omega_4 \leq \frac{1}{3} \tag{3.74}$$

$$\omega_5 \leq \frac{1}{3}\ , \tag{3.75}$$

which are the inequalities we get from the original theorem if we let $\mathcal{M}_6$ be maximally predictable, i.e. $\omega_6 = 1$.

3.7 The algorithm

```
import numpy as np
from FME import*

#The number of measurements is 6
nMsmt = 6
nOut = 3

#-----------------------------------------------------------------------------#
#--------                Polytope Generation             ------------#
#-----------------------------------------------------------------------------#

M = np.eye(19)

rXI,rIX,rXX = M[1:4]
rIZ,rZI,rZZ = M[4:7]
rXZ,rZX,rYY = M[7:10]

Row = np.array([[rXI,rIX,rXX],
     [rIZ,rZI,rZZ],
     [rXZ,rZX,rYY]])

cXI,cIZ,cXZ = M[10:13]
cIX,cZI,cZX = M[13:16]
cXX,cZZ,cYY = M[16:19]

Col = np.array([[cXI,cIZ,cXZ],
```

```
    [cIX,cZI,cZX],
    [cXX,cZZ,cYY]])

contextualityConstraints = [
#Placeholder for contextuality constraints
]

for i in range(3):
  for j in range(3):
    contextualityConstraints.append(Row[i][j] - Col[j][i])

#Tetrahedron inequalities from Lemma 1
tetrahedronIneqs = [
#Placeholder for tetrahedron inequalities
]

#Sum of expectations GEQ -1
for i in range(3):
  for a in {-1,+1}:
    for b in {-1,+1}:
      tetrahedronIneqs.append(M[0] + a*Row[i][0] + b*Row[i][1] + a*b*Row[i][2])

for i in range(2):
  for a in {-1,+1}:
    for b in {-1,+1}:
      tetrahedronIneqs.append(M[0] + a*Col[i][0] + b*Col[i][1] + a*b*Col[i][2])

#Things work differently for the last column
for a in {-1,+1}:
  for b in {-1,+1}:
    tetrahedronIneqs.append(M[0] + a*Col[2][0] + b*Col[2][1] - a*b*Col[2][2])

#Sum of expectations LEQ 3
for i in range(3):
  for a in {-1,+1}:
    for b in {-1,+1}:
      tetrahedronIneqs.append(3*M[0] - a*Row[i][0] - b*Row[i][1] - a*b*Row[i][2])


for i in range(2):
  for a in {-1,+1}:
    for b in {-1,+1}:
      tetrahedronIneqs.append(3*M[0] - a*Col[i][0] - b*Col[i][1] - a*b*Col[i][2])

#Things work differently for the last column
for a in {-1,+1}:
  for b in {-1,+1}:
    tetrahedronIneqs.append(3*M[0] - a*Col[2][0] - b*Col[2][1] + a*b*Col[2][2])

#Define the polyhedron using these inequalities
p = Polyhedron(ieqs = tetrahedronIneqs, eqns=contextualityConstraints)

#-----------------------------------------------------------------------------#
#-------                 Contextuality inequalities         --------#
#-----------------------------------------------------------------------------#


def unique_rows(a):
   a = np.ascontiguousarray(a)
```

```
    unique_a = np.unique(a.view([('', a.dtype)]*a.shape[1]))
    return unique_a.view(a.dtype).reshape((unique_a.shape[0], a.shape[1]))

def ineqs(p,n):
   nVertices = p.n_vertices()
        M = np.zeros(shape=(nVertices,nMsmt))

        i = 0
        for v in p.vertices():
                z = abs(np.array(v))
                for j in np.arange(nMsmt):
                        M[i][j] = sum(abs(z[j*nOut:(j+1)*nOut]))
         M[i][j] = (1.0/nOut)*M[i][j]
                i += 1

        Q = unique_rows(M)
        qr,qc = Q.shape
        #Q is a 16 x 5 matrix

        #First we impose the inequality w_i \leq sum_j Q_ij v_j
        b = np.array([np.zeros(nMsmt)])
        W = -1*np.eye(nMsmt)
        A = np.concatenate((b.T,Q.T,W),axis=1)

        #Ensure that \nu is a probability distribution
                ##First impose positivity of \nu
        h = np.eye(qr + 1)
   hh = np.zeros(shape=(qr,nMsmt))
        H = np.concatenate((h[1:],hh),axis=1)

                ##Then impose sum to 1
        e = np.array([np.concatenate(([1],-1*np.ones(qr),np.zeros(nMsmt)),axis=1)])
        f = np.array([np.concatenate(([-1],np.ones(qr),np.zeros(nMsmt)),axis=1)])

        #And are greater than 0
        j = np.zeros(shape=(nMsmt,qr + 1))
        jj = np.eye(nMsmt)
        J = np.concatenate((j,jj),axis=1)
        A = np.concatenate((A,H,I,J,e,f),axis=0)

        for m in range(qr):
                print("Done round " + str(m+1) + "/" + str(qr))
                A = FMEliminate(A,1)
                p = Polyhedron(ieqs = A, base_ring = RDF)
                A = np.array(p.inequalities())
        return(A)

o = ineqs(p,n)
print(o)
```

Chapter 4

Mermin's Star

In this chapter, we shall provide another example to demonstrate how FM elimination can be used to automate the derivation of noncontextuality inequalities. We shall consider Mermin's star, a popular proof of the impossibility of noncontextual ontological models [31]. As we did before, we shall begin by providing an operational description of the star construction. We shall then see how the assumption of noncontextuality restricts the set of possible ontological expectation value assignments to the star. We shall see how the predictabilities are bounded if a noncontextual model underlies the experiment.

Mermin's star construction in quantum theory is another proof of impossibility of noncontextual hidden variables. It consists of 5 sets of 4 observables on 3 qubits arranged in the shape of a star depicted below in fig. 4.

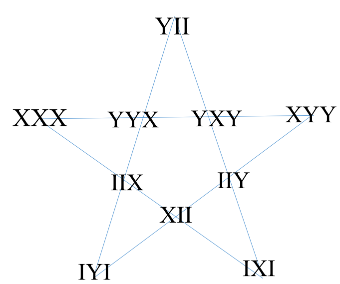


Figure 4.1: Mermin's Star

The notation is the same as that of the Peres-Mermin square in the previous chapter: $\mathrm{I}, \mathrm{X}, \mathrm{Y}$ and Z represent Pauli operators. The observables along each edge of the star commute pairwise and thus can be simultaneously measured. Every observable appears in 2 measurements and can thus be part of two contexts. The product of the observables along any edge is I, except for the horizontal edge where it is $-\mathrm{I}$.

4.1 An operational description of Mermin's Star construction

To derive a noncontextuality inequality, it will be convenient to have an operational description of the star construction. It involves five eight-outcome measurements $\mathcal{M}_1, \cdots, \mathcal{M}_5$. The output of measurement $\mathcal{M}_i$ is specified by 4 binary random variables $\mathrm{A}_i, \mathrm{B}_i, \mathrm{C}_i$ and D_i which take values in the set $\{+1, -1\}$. The random variables associated with measurements $\mathcal{M}_1, \cdots, \mathcal{M}_4$ satisfy the relation $\mathrm{A}_i \cdot \mathrm{B}_i \cdot \mathrm{C}_i \cdot \mathrm{D}_i = 1$. For measurement $\mathcal{M}_5$, the random variables satisfy the product relation $\mathrm{A}_i \cdot \mathrm{B}_i \cdot \mathrm{C}_i \cdot \mathrm{D}_i = -1$. If, upon measurement of observable $\mathcal{M}_i$, we find $\mathrm{A}_i = \alpha$, $\mathrm{B}_i = \beta$ and $\mathrm{C}_i = \gamma$ for $\alpha, \beta, \gamma \in \{+1, -1\}$, we shall say that the outcome of measurement $\mathcal{M}_i$ is $\alpha\beta\gamma$.

In addition to these product relations, there are also some operational equivalences between the random variables. This is illustrated in the hypergraph in fig. 4.1 below.

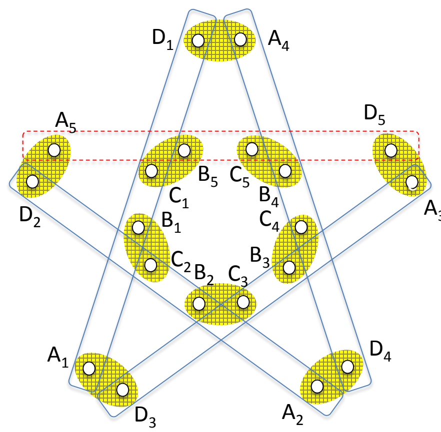


Figure 4.2: The hypergraph corresponding to the star construction.

As in the case of the Peres-Mermin square, there is a duality between probabilities and expectation values of random variables. We may express

$$\mathrm{pr}\left\{\mathrm{A}_i = \alpha, \mathrm{B}_i = \beta, \mathrm{C}_i = \gamma | \mathcal{M}_i, \mathcal{P}\right\} = \sum_{p,q,r \in \{0,1\}} \alpha^p \beta^q \gamma^r \langle \mathrm{A}_i^p \, \mathrm{B}_i^q \, \mathrm{C}_i^r \rangle_{\mathcal{P}} \ . \tag{4.1}$$

We note that some products of the random variables that apppear in the above expression for the probabilities do not appear in the square. For instance, the product $\mathrm{A}_i \, \mathrm{B}_i$ does not appear anywhere in the square.

For each outcome $\alpha\beta\gamma$ of measurement $\mathcal{M}_i$, we shall consider a preparation procedure $\mathcal{P}_i^{(\alpha\beta\gamma)}$. From these, we shall construct the coarse-grained preparation $\mathcal{P}_i$ by sampling α, β and γ with uniform probability and then performing the corresponding preparation. The coarse-grained preparations are assumed to satisfy the equivalence

$$\mathcal{P}_1 \simeq \mathcal{P}_2 \simeq \mathcal{P}_3 \simeq \mathcal{P}_4 \simeq \mathcal{P}_5 \ . \tag{4.2}$$

4.2 Constraints on the ontic assignments

It is straightforward to show that there exists no KS noncontextual model underlying Mermin's star. For the sake of contradiction, let us suppose we have a deterministic assignment. Each of the random variables is assigned a value in $\{+1, -1\}$. Consider first the random variables in measurements $\mathcal{M}_1$ through $\mathcal{M}_4$. Almost all the random variables appear in pairs - for instance, B_1 and C_4 are always assigned the same value and hence their product is going to be $+1$. The only random variables that do not appear in pairs are those that are operationally equivalent to those appearing in measurement $\mathcal{M}_5$. If we were to multiply all the random variables appearing in these measurements, we would have

$$\mathrm{D}_2 \, \mathrm{C}_1 \, \mathrm{B}_4 \, \mathrm{A}_3 \ . \tag{4.3}$$

To evaluate this product, we note that the random variables all lie within the blue loops and hence their product must be $+1$.

Similarly, the product of the random variables which appear in measurement $\mathcal{M}_5$ is -1. However, measurement noncontextuality stipulates that

$$\mathrm{A}_5\,\mathrm{B}_5\,\mathrm{C}_5\,\mathrm{D}_5 = \mathrm{D}_2\,\mathrm{C}_1\,\mathrm{B}_4\,\mathrm{A}_3 \tag{4.4}$$

and hence we run into a contradiction.

Thus all the random variables cannot be made deterministic simultaneously. The following lemma will help us gain a better understanding of the relationship between the random variables and the tradeoffs involved in making some of them deterministic.

Lemma 5. *For measurements $\mathcal{M}_1, \cdots, \mathcal{M}_4$, the ontic expectation values obey*

$$\begin{align} -2 \leq \langle \mathrm{A}_i \rangle_\lambda + \langle \mathrm{B}_i \rangle_\lambda + \langle \mathrm{C}_i \rangle_\lambda - \langle \mathrm{D}_i \rangle_\lambda \leq 2 \tag{4.5}\\ -2 \leq \langle \mathrm{A}_i \rangle_\lambda + \langle \mathrm{B}_i \rangle_\lambda - \langle \mathrm{C}_i \rangle_\lambda + \langle \mathrm{D}_i \rangle_\lambda \leq 2 \tag{4.6}\\ -2 \leq \langle \mathrm{A}_i \rangle_\lambda - \langle \mathrm{B}_i \rangle_\lambda + \langle \mathrm{C}_i \rangle_\lambda + \langle \mathrm{D}_i \rangle_\lambda \leq 2 \tag{4.7}\\ -2 \leq -\langle \mathrm{A}_i \rangle_\lambda + \langle \mathrm{B}_i \rangle_\lambda + \langle \mathrm{C}_i \rangle_\lambda + \langle \mathrm{D}_i \rangle_\lambda \leq 2 \tag{4.8}\\ -1 \leq \langle \mathrm{A}_i \rangle_\lambda \leq 1 \tag{4.9}\\ -1 \leq \langle \mathrm{B}_i \rangle_\lambda \leq 1 \tag{4.10}\\ -1 \leq \langle \mathrm{C}_i \rangle_\lambda \leq 1 \tag{4.11}\\ -1 \leq \langle \mathrm{D}_i \rangle_\lambda \leq 1\ . \tag{4.12} \end{align}$$

For measurement $\mathcal{M}_5$, the ontic expectation values obey

$$\begin{align} -2 \leq \langle \mathrm{A}_i \rangle_\lambda + \langle \mathrm{B}_i \rangle_\lambda + \langle \mathrm{C}_i \rangle_\lambda + \langle \mathrm{D}_i \rangle_\lambda \leq 2 \tag{4.13}\\ -2 \leq \langle \mathrm{A}_i \rangle_\lambda + \langle \mathrm{B}_i \rangle_\lambda - \langle \mathrm{C}_i \rangle_\lambda - \langle \mathrm{D}_i \rangle_\lambda \leq 2 \tag{4.14}\\ -2 \leq \langle \mathrm{A}_i \rangle_\lambda - \langle \mathrm{B}_i \rangle_\lambda + \langle \mathrm{C}_i \rangle_\lambda - \langle \mathrm{D}_i \rangle_\lambda \leq 2 \tag{4.15}\\ -2 \leq -\langle \mathrm{A}_i \rangle_\lambda + \langle \mathrm{B}_i \rangle_\lambda + \langle \mathrm{C}_i \rangle_\lambda - \langle \mathrm{D}_i \rangle_\lambda \leq 2 \tag{4.16}\\ -1 \leq \langle \mathrm{A}_i \rangle_\lambda \leq 1 \tag{4.17}\\ -1 \leq \langle \mathrm{B}_i \rangle_\lambda \leq 1 \tag{4.18}\\ -1 \leq \langle \mathrm{C}_i \rangle_\lambda \leq 1 \tag{4.19}\\ -1 \leq \langle \mathrm{D}_i \rangle_\lambda \leq 1\ . \tag{4.20} \end{align}$$

Proof. From the duality between expectation values and random variables, we can translate the inequalities

$$0 \leq \mathrm{pr}\left\{\mathrm{A}_i = \alpha, \mathrm{B}_i = \beta, \mathrm{C}_i = \gamma\right\} \leq 1 \quad \forall \alpha, \beta, \gamma \in \{+1, -1\} \tag{4.21}$$

into bounds on the expectation values. Some products of random variables do not occur explicitly in Mermin's star. We can use FM elimination to project down to the observables that appear in Mermin's star.

One might argue that in measuring the random variables A_i, B_i and C_i that we could construct the expectation values of any product of them. While this is true we wish to restrict our attention to observables that appear explicitly in the star construction for the sake of simplicity. □

4.3 Bounds on the predictability

The predictability of measurement $\mathcal{M}_i$ with respect to the preparation $\mathcal{P}_i^{(\alpha\beta\gamma)}$ is defined as

$$\omega_i^{(\alpha\beta\gamma)} = \frac{1}{4}\left|\left(\alpha\langle \mathrm{A}_i\rangle_{\mathcal{P}_i}^{(\alpha\beta\gamma)} + \beta\langle \mathrm{B}_i\rangle_{\mathcal{P}_i}^{(\alpha\beta\gamma)} + \gamma\langle \mathrm{C}_i\rangle_{\mathcal{P}_i}^{(\alpha\beta\gamma)} + \alpha\beta\gamma\langle \mathrm{D}_i\rangle_{\mathcal{P}_i}^{(\alpha\beta\gamma)}\right)\right|\ . \tag{4.22}$$

Similar to the examples we have seen previously, this measures the correlation between the input setting $\alpha\beta\gamma$ and the outcome $\alpha\beta\gamma$ of measurement $\mathcal{M}_i$.

The average degree of correlation between the input and the output of measurement $\mathcal{M}_i$ is the average of the predictabilities

$$\omega_i = \frac{1}{8}\sum_{\alpha\beta\gamma}\omega_i^{(\alpha\beta\gamma)} \ . \tag{4.23}$$

These quantities are bounded as follows

Theorem 6. *For a universally noncontextual ontological model with 5 measurements $\mathcal{M}_1, \cdots, \mathcal{M}_5$ that obey the constraints and equivalences in fig. 4.1 and a set of preparation procedures $\{\mathcal{P}_i^{\alpha\beta\gamma}\}$ that obey the equivalences in eqn. (4.2) the predictabilities obey the following bounds:*
Bound on triples of measurements: For any $i \neq j \neq k$,

$$\frac{1}{3}\left(\omega_i + \omega_j + \omega_k\right) \leq \frac{5}{6} \tag{4.24}$$

Bound on quadruples of measurements: For any $i \neq j \neq k \neq l$,

$$\frac{1}{4}\left(\omega_i + \omega_j + \omega_k + \omega_l\right) \leq \frac{3}{4} \ . \tag{4.25}$$

Bound on all measurements:

$$\frac{1}{5}\left(\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5\right) \leq \frac{7}{10} \ . \tag{4.26}$$

Proof. The proof proceeds similarly to the other proofs. As before, we begin by upper bounding the predictability $\omega_i^{(\alpha\beta\gamma)}$.

$$\omega_i^{(\alpha\beta\gamma)} = \frac{1}{4}\left|\left(\alpha\langle \mathrm{A}_i\rangle_{\mathcal{P}_i^{(\alpha\beta\gamma)}} + \beta\langle \mathrm{B}_i\rangle_{\mathcal{P}_i^{(\alpha\beta\gamma)}} + \gamma\langle \mathrm{C}_i\rangle_{\mathcal{P}_i^{(\alpha\beta\gamma)}} + \alpha\beta\gamma\langle \mathrm{D}_i\rangle_{\mathcal{P}_i^{(\alpha\beta\gamma)}}\right)\right| \tag{4.27}$$

$$= \frac{1}{4}\left|\sum_\lambda\left(\alpha\langle \mathrm{A}_i\rangle_\lambda\mu(\lambda|\mathcal{P}_i^{(\alpha\beta\gamma)}) + \beta\langle \mathrm{B}_i\rangle_\lambda\mu(\lambda|\mathcal{P}_i^{(\alpha\beta\gamma)}) + \gamma\langle \mathrm{C}_i\rangle_\lambda\mu(\lambda|\mathcal{P}_i^{(\alpha\beta\gamma)}) + \alpha\beta\gamma\langle \mathrm{D}_i\rangle_\lambda\mu(\lambda|\mathcal{P}_i^{(\alpha\beta\gamma)})\right)\right| \tag{4.28}$$

$$\leq \frac{1}{4}\sum_\lambda\left(|\langle \mathrm{A}_i\rangle|_\lambda + |\langle \mathrm{B}_i\rangle_\lambda| + |\langle \mathrm{C}_i\rangle_\lambda| + |\langle \mathrm{D}_i\rangle_\lambda|\right)\mu(\lambda|\mathcal{P}_i^{(\alpha\beta\gamma)}) \ . \tag{4.29}$$

At this point, we can invoke preparation noncontextuality to deduce that we can define a distribution $\nu(\lambda)$ such that

$$\frac{1}{8}\sum_{\alpha,\beta,\gamma}\mu(\lambda|\mathcal{P}_i^{(\alpha\beta\gamma)}) = \nu(\lambda) \ . \tag{4.30}$$

This implies

$$\omega_i \leq \frac{1}{4}\sum_\lambda\left(|\langle \mathrm{A}_i\rangle_\lambda| + |\langle \mathrm{B}_i\rangle_\lambda| + |\langle \mathrm{C}_i\rangle_\lambda| + |\langle \mathrm{D}_i\rangle_\lambda|\right)\nu(\lambda) \ . \tag{4.31}$$

As before, we can express the ontic expectation values as a convex combination of their valuations on the vertices of the polytope specified by lemma 5. If we index the vertices of this polytope by κ, we can express this inequality as

$$\omega_i \leq \sum_\kappa Q_{i\kappa}\eta(\kappa) \ , \tag{4.32}$$

where $Q_{i\kappa} = \frac{1}{4}\left(|\langle \mathrm{A}_i\rangle_\lambda| + |\langle \mathrm{B}_i\rangle_\lambda| + |\langle \mathrm{C}_i\rangle_\lambda| + |\langle \mathrm{D}_i\rangle_\lambda|\right)$ and η is some probability distribution over the vertices of this polytope. We then aggregate vertices of the polytope and perform FM elimination to obtain the inequalities specified in the statement of the theorem. □

4.4 The algorithm

```
from FME import*
nMsmt = 5
nOut = 4

#--------------------------------------------------------------------------#
#-----                   Polytope setup                ----------#
#--------------------------------------------------------------------------#

#G is the matrix that will encode the geq 0 condition
G = np.array([np.ones(8)],dtype=np.float64)

for j in np.arange(1,8):
   i = '{0:03b}'.format(j)
   G = np.concatenate((G,[np.zeros(8)]))
   G[j][0] = 1.0
   G[j][1] = (-1.0)**int(i[0])
   G[j][2] = (-1.0)**int(i[1])
   G[j][3] = (-1.0)**int(i[2])
   G[j][4] = (-1.0)**(int(i[0])+int(i[1]))
   G[j][5] = (-1.0)**(int(i[1])+int(i[2]))
   G[j][6] = (-1.0)**(int(i[0])+int(i[2]))
   G[j][7] = (-1.0)**(int(i[0])+int(i[1])+int(i[2]))

#L is the matrix that will encode the leq 1 condition
L = (-1.0)*np.copy(G)
L[:,0] = 7*np.ones(8)

#Altogether we have 16 inequalities
Ineq = np.append(G,L,axis=0)
P = Polyhedron(ieqs = Ineq, base_ring=RDF)
I = np.array(P.inequalities())

Ineq4 = FMEliminate(I,4)
P4 = Polyhedron(ieqs = Ineq4, base_ring=RDF)
I4 = np.array(P4.inequalities())

Ineq5 = FMEliminate(I4,4)
P5 = Polyhedron(ieqs = Ineq5, base_ring=RDF)
I5 = np.array(P5.inequalities())

Ineq6 = FMEliminate(I5,4)
P6 = Polyhedron(ieqs = Ineq6, base_ring=RDF)
I6 = np.array(P6.inequalities())

r,c = I6.shape

#4 observables on each row. 5 rows. 20 variables.
M = np.eye(21)

#Row1,...,Row5 represent the expectation values in each row
Row1 = np.concatenate(([M[0]],M[1:5]),axis=0)
Row2 = np.concatenate(([M[0]],M[5:9]),axis=0)
Row3 = np.concatenate(([M[0]],M[9:13]),axis=0)
Row4 = np.concatenate(([M[0]],M[13:17]),axis=0)
Row5 = np.concatenate(([M[0]],M[17:21]),axis=0)
```

```
#A,...,E are the placeholders for the inequalities for each row
A = np.zeros(r*21).reshape(r,21)
B = np.zeros(r*21).reshape(r,21)
C = np.zeros(r*21).reshape(r,21)
D = np.zeros(r*21).reshape(r,21)
E = np.zeros(r*21).reshape(r,21)

for i in range(r):
  for j in range(c):
       A[i] += I6[i][j]*Row1[j]
       B[i] += I6[i][j]*Row2[j]
       C[i] += I6[i][j]*Row3[j]
       D[i] += I6[i][j]*Row4[j]
       E[i] += I6[i][j]*Row5[j]

Ineqs = np.concatenate((A,B,C,D,E))

contextualityConstraints = [
Row1[1] - Row3[4],
Row2[1] - Row4[4],
Row3[1] - Row5[4],
Row4[1] - Row1[4],
Row5[1] + Row2[4],
Row1[2] - Row2[3],
Row2[2] - Row3[3],
Row3[2] - Row4[3],
Row4[2] - Row5[3],
Row5[2] - Row1[3]
]

p = Polyhedron(ieqs = Ineqs, eqns = contextualityConstraints, base_ring=RDF)

#----------------------------------------------------------------------------#
#-------         Contextuality inequalities                -------#
#----------------------------------------------------------------------------#


def unique_rows(a):
   a = np.ascontiguousarray(a)
   unique_a = np.unique(a.view([('', a.dtype)]*a.shape[1]))
   return unique_a.view(a.dtype).reshape((unique_a.shape[0], a.shape[1]))

def ineqs(p):
      nVertices = p.n_vertices()
      M = np.zeros(shape=(nVertices,nMsmt))

      i = 0
      for v in p.vertices():
            z = abs(np.array(v))
            for j in np.arange(nMsmt):
                  M[i][j] = (1.0/nOut)*sum(abs(z[j*nOut:(j+1)*nOut]))
            i += 1

      Q = unique_rows(M)
      qr,qc = Q.shape
  #Q is a 16 x 5 matrix

      #First we impose the inequality w_i \leq sum_j Q_ij v_j
      b = np.array([np.zeros(nMsmt)])
```

```
        W = -1*np.eye(nMsmt)
        A = np.concatenate((b.T,Q.T,W),axis=1)

        #Ensure that \nu is a probability distribution
                ##First impose positivity of \nu
        h = np.eye(qr + 1)
        hh = np.zeros(shape=(qr,nMsmt))
        H = np.concatenate((h[1:],hh),axis=1)

                ##Then impose sum to 1
        e = np.array([np.concatenate(([1],-1*np.ones(qr),np.zeros(nMsmt)),axis=1)])
        f = np.array([np.concatenate(([-1],np.ones(qr),np.zeros(nMsmt)),axis=1)])

        #omega_i are greater than 0
        j = np.zeros(shape=(nMsmt,qr + 1))
        jj = np.eye(nMsmt)
        J = np.concatenate((j,jj),axis=1)
        A = np.concatenate((A,H,I,J,e,f),axis=0)

        for m in range(qr):
                print("Done round " + str(m+1) + "/" + str(qr))
                A = FMEliminate(A,1)
                p = Polyhedron(ieqs = A, base_ring = RDF)
                A = np.array(p.inequalities())
        return(A)

o = ineqs(p)
print(o)
```

Chapter 5

18 vector construction

Recently, Kunjwal and Spekkens [25] considered a proof of KS noncontextuality called the 18 vector construction and were able to obtain a noncontextuality inequality for this construction without assuming outcome determinism. This inequality is a bound on the average of the predictabilities of all the measurements. However, there is no justification a priori to consider the average of the predictabilities. The natural question to ask at this point is whether FM elimination can be used to derive the same inequalities or perhaps even stronger bounds.

In this chapter, we answer this question in the affirmative. The inequalities we obtain are an improvement over the inequality presented in Kunjwal & Spekkens in that they form a tighter set of necessary and sufficient conditions for the 18 vector construction to admit of a noncontextual model. Furthermore, it can be shown that these inequalities imply the Kunjwal and Spekkens bound.

The narrative in this chapter will be fairly similar to the previous chapters apart from one minor change: the predictabilities are expressed in terms of probabilities instead of expectation values. As has been emphasized repeatedly, this is just an affine transformation of the above predictabilities. Hence, there is no difference conceptually between the two definitions.

The figures used in this chapter are from [25].

5.1 The original proof

The original construction consists of 18 vectors in $\mathbb{C}^4$. These vectors are used to construct 9 bases. Each of the 18 vectors appears in exactly 2 bases. The vectors within a basis are orthogonal to each other. These 9 bases can be used to specify 9 measurement procedures with 4 outcomes each.

Suppose we associated an ontological model with this construction. The probabilities assigned by an ontic state λ to the vectors in a basis must add to 1. As per the assumptions of KS noncontextuality, if we assume that these probability assignments are deterministic, then one and only one vector in each basis can be assigned a value of 1, while the rest must be assigned 0.

The 18 ray proof of KS noncontextuality is typically presented as the uncolourability of a particular graph. This can be visualized as done in fig. (5.1). Each of the 18 vectors is depicted as a node. Vectors that constitute a basis are encircled by a blue loop. When a node has been coloured black it means it has been assigned a value 1 and if it is white, it means it has been assigned a value 0. It can be shown that there exists no consistent assignments of colours and hence we run into a contradiction with the assumption of KS noncontextuality. An example of how an attempt at such a colouring leads to an inconsistent assignment is also depicted in fig. 5.1.

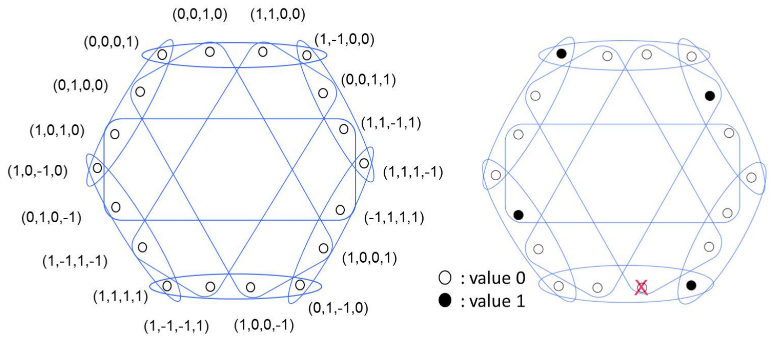


Figure 5.1: Each node in this graph represents one of the 18 vectors in the construction. The blue loop encircles quadruples that constitute a basis. On the right is an example of an impossible assignment. By assigning 1's and 0's as described above, we run into a contradiction depicted by a red cross.

5.2 An operational description of the 18 vector construction

As has become our routine, we begin with an operational description of the 18 vector construction. It has been noted that the construction has 9 4-outcome measurements $\mathcal{M}_1, \cdots, \mathcal{M}_9$. We shall denote by $[k|\mathcal{M}_i]$ the k-th outcome of the measurement $\mathcal{M}_i$. Unlike the square and star constructions, there are no product relations between the random variables. The equivalences between the different outcomes have been depicted in the hypergraph below. If a yellow edge of the hypergraph encircles the nodes $[k|\mathcal{M}_i]$ and $[k'|\mathcal{M}_j]$, it means that $\mathrm{pr}\left\{k|\mathcal{M}_i, \mathcal{P}\right\} = \mathrm{pr}\left\{k'|\mathcal{M}_j, \mathcal{P}\right\}$ for all preparations $\mathcal{P}$.

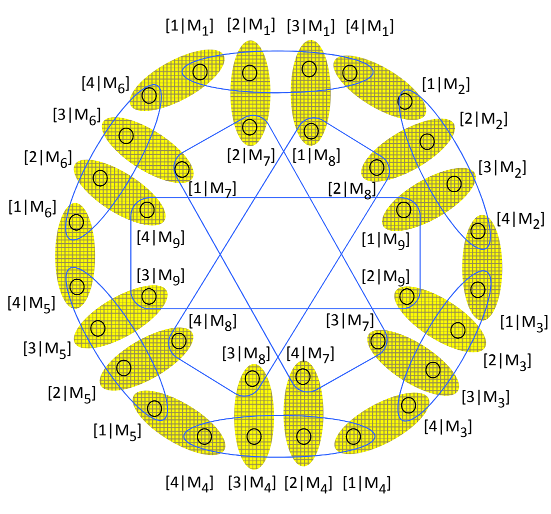


Figure 5.2: Measurement equivalences for the 18 vector construction.

With each outcome k of each measurement $\mathcal{M}_i$, we shall consider a preparation procedure $\mathcal{P}_i^k$. We shall denote by $\mathcal{P}_i$ the coarse-grained preparation formed by sampling $k \in \{1, \cdots, 4\}$ with equal probability and implementing $\mathcal{P}_i^k$. These preparations are assumed to obey the operational equivalences

$$\mathcal{P}_i \simeq \mathcal{P}_j \quad \forall i, j \in \{1, \cdots, 9\} \ . \tag{5.1}$$

5.3 Constraints on the ontic assignments

To understand the tradeoffs between the different measurements, we note that for all measurements $\mathcal{M}_i$, the ontic response functions are bounded by

$$0 \leq \xi \{k|\mathcal{M}_i, \lambda\} \leq 1 \ , \tag{5.2}$$

for every outcome k over the entire ontic space Λ.

Since there are 9 measurements and 4 outcomes, this specifies 36 inequalities.

In addition, measurement noncontextuality imposes 18 linear constraints on the 36 response functions. These follow from the hypergraph in fig. 5.2. Finally, we must ensure that the probabilities assigned by an ontic state to the outcomes of a measurement sum to 1. Since there are 9 measurements, this constitutes 9 more linear constraints on the response functions.

These constraints define a 9 dimensional polytope with a 146 vertices.

5.4 Bounds on the predictability

The predictability of measurement $\mathcal{M}_i$ with respect to preparation $\mathcal{P}_i^k$ is defined as

$$\omega_i^k = \mathrm{pr} \left\{ k|\mathcal{M}_i, \mathcal{P}_i^k \right\} \ . \tag{5.3}$$

Although this is expressed in terms of probabilities, it has the same interpretation as it previously did: it measures how likely it is to see outcome k if we prepare $\mathcal{P}_i^k$ and measure $\mathcal{M}_i$.

The average predictability of $\mathcal{M}_i$ is then defined as

$$\omega_i = \frac{1}{4} \sum_{k=1}^{4} \omega_i^k \ . \tag{5.4}$$

Note that we do not need to consider the modulus of this quantity as

$$0 \leq \omega_i \leq 1 \ , \tag{5.5}$$

by definition.

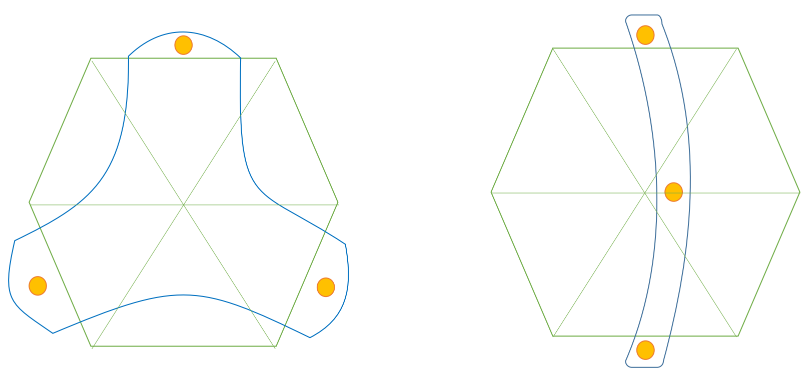
Theorem 7. *Consider an operational theory that admits of a universally noncontextual ontological model. Let $\{\mathcal{M}_1, \cdots, \mathcal{M}_9\}$ be nine four-outcome measurements that obey the measurement equivalences specified in fig. 5.2.*

Let $\{\mathcal{P}_i^k | k \in \{1, \cdots, 4\}, i \in \{1, \cdots, 9\}\}$ be 36 preparation procedures that obey the equivalence specified in equation (5.1). Then the predictabilites are restrained as follows:

$$
\begin{aligned}
\frac{1}{4}\left(\omega_1+\omega_2+\omega_4+\omega_5\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_1+\omega_3+\omega_4+\omega_6\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_1+\omega_3+\omega_8+\omega_9\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_1+\omega_5+\omega_7+\omega_9\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_2+\omega_3+\omega_5+\omega_6\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_2+\omega_4+\omega_7+\omega_9\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_2+\omega_6+\omega_7+\omega_8\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_2+\omega_4+\omega_7+\omega_9\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_3+\omega_5+\omega_7+\omega_8\right) &\leq \frac{7}{8}\\
\frac{1}{4}\left(\omega_4+\omega_6+\omega_8+\omega_9\right) &\leq \frac{7}{8}
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{3}\left(\omega_1+\omega_3+\omega_5\right) &\leq \frac{5}{6}\\
\frac{1}{3}\left(\omega_1+\omega_4+\omega_9\right) &\leq \frac{5}{6}\\
\frac{1}{3}\left(\omega_2+\omega_4+\omega_6\right) &\leq \frac{5}{6}\\
\frac{1}{3}\left(\omega_2+\omega_5+\omega_7\right) &\leq \frac{5}{6}\\
\frac{1}{3}\left(\omega_3+\omega_6+\omega_8\right) &\leq \frac{5}{6}\\
\frac{1}{3}\left(\omega_7+\omega_8+\omega_9\right) &\leq \frac{5}{6}
\end{aligned}
$$

Notice that there are two types of bounds on the predictabilities. The first is a bound on the triples and the second is a bound on quadruples. Although it is not evident initially, there is a certain symmetry to the inequalities which we depict using the following cartoons.



(a) Alternate measurements around the hexagon (b) Inequalities that bound measurements on opposite sides of the hexagon

As depicted in fig. 5.3(a), some inequalities involve predictabilities of alternate measurements around the hexagon in the hypergraph. The inequalities

$$\frac{1}{3}\left(\omega_1+\omega_3+\omega_5\right) \leq \frac{5}{6} \tag{5.6}$$

$$\frac{1}{3}\left(\omega_2+\omega_4+\omega_6\right) \leq \frac{5}{6} \tag{5.7}$$

are examples of such triples.

There are also inequalities on triples that bound measurements on opposite sides of the hexagon and the measurement that is parallel to the two sides as shown in fig. 5.3(b). The inequalities

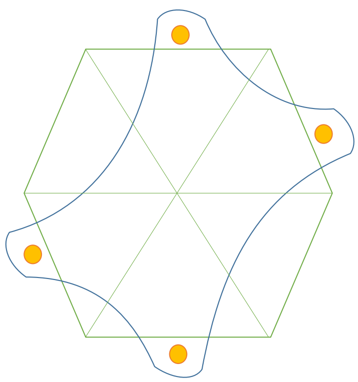
$$\frac{1}{3}\left(\omega_1+\omega_4+\omega_9\right) \leq \frac{5}{6} \tag{5.8}$$
$$\frac{1}{3}\left(\omega_2+\omega_5+\omega_8\right) \leq \frac{5}{6} \tag{5.9}$$

are examples of such triples.

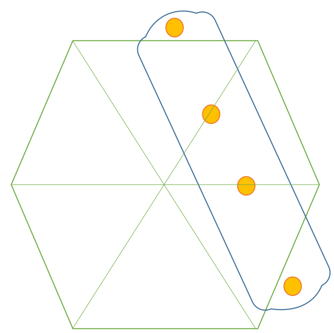
Note that we can put the inequalities on triples together to obtain

$$\frac{1}{9}\sum_{i=1}^{9}\omega_i \leq \frac{5}{6}\ , \tag{5.10}$$

which was the bound obtained by Kunjwal and Spekkens in [25].



(c) Quadruples of predictabilities



(d) Quadruples of predictabilities

Inequalities involving quadruples of predictabilities are depicted in fig. 5.4. The first kind is depicted in 5.3(c). Examples of such inequalities are

$$\frac{1}{4}\left(\omega_1+\omega_2+\omega_4+\omega_5\right) \leq \frac{7}{8} \tag{5.11}$$
$$\frac{1}{4}\left(\omega_1+\omega_3+\omega_4+\omega_6\right) \leq \frac{7}{8} \tag{5.12}$$

Finally, we have quadruples that involve all measurements connected to a side of the hexagon. For example, the inequality

$$\frac{1}{4}\left(\omega_1+\omega_3+\omega_8+\omega_9\right) \leq \frac{7}{8} \tag{5.13}$$

involves all measurements that are connected to measurement 2.

Proof. (**Sketch**) Following the procedure outlined in Chapter 2, we can derive inequalities of the form

$$\omega_i \leq Q_{i\kappa}\eta(\kappa)\ , \tag{5.14}$$

where $Q_{i\kappa} = \max_k \xi(k|\mathcal{M}_i,\kappa)$ where κ is a vertex of the polytope of permitted expectation value assignments. Brute force FM elimination does not work for this problem either. However, by aggregating vertices, we can make the problem feasible. Aggregating vertices reduces the number of vertices from 146 to 61.

We can now perform FM elimination to obtain the inequalities presented in the theorem statement. The algorithm has been presented below. □

5.5 The algorithm

```
import numpy as np
from FME import*

#The number of measurements is 9
nMsmt = 9
nOut = 4

#------------------------------------------------------------------------------#
#-------                 Polytope setup                      ---------#
#------------------------------------------------------------------------------#
M = np.eye(37)

M_11, M_12, M_13, M_14 = M[1:5]
M_21, M_22, M_23, M_24 = M[5:9]
M_31, M_32, M_33, M_34 = M[9:13]
M_41, M_42, M_43, M_44 = M[13:17]
M_51, M_52, M_53, M_54 = M[17:21]
M_61, M_62, M_63, M_64 = M[21:25]
M_71, M_72, M_73, M_74 = M[25:29]
M_81, M_82, M_83, M_84 = M[29:33]
M_91, M_92, M_93, M_94 = M[33:37]

contextualityConstraints = [
M_11 - M_64,
M_12 - M_72,
M_13 - M_81,
M_14 - M_21,
M_22 - M_82,
M_23 - M_91,
M_24 - M_31,
M_32 - M_92,
M_33 - M_73,
M_34 - M_41,
M_42 - M_74,
M_43 - M_83,
M_44 - M_51,
M_52 - M_84,
M_53 - M_93,
M_54 - M_61,
M_62 - M_94,
M_63 - M_71
]

probConstraints = [
M[0] - M_11 - M_12 - M_13 - M_14,
M[0] - M_21 - M_22 - M_23 - M_24,
M[0] - M_31 - M_32 - M_33 - M_34,
M[0] - M_41 - M_42 - M_43 - M_44,
M[0] - M_51 - M_52 - M_53 - M_54,
M[0] - M_61 - M_62 - M_63 - M_64,
M[0] - M_71 - M_72 - M_73 - M_74,
M[0] - M_81 - M_82 - M_83 - M_84,
M[0] - M_91 - M_92 - M_93 - M_94
]

N = -1*np.copy(M)
```

```
N[:,0] = np.ones(37)

ineqs = np.concatenate((M[1:],N[1:]),axis=0)

eqs = np.concatenate((contextualityConstraints,probConstraints),axis=0)

p = Polyhedron(ieqs = ineqs, eqns=eqs, base_ring = RDF)

#-----------------------------------------------------------------------------#
#-------            Contextuality inequalities            ----------#
#-----------------------------------------------------------------------------#

def unique_rows(a):
    a = np.ascontiguousarray(a)
    unique_a = np.unique(a.view([('', a.dtype)]*a.shape[1]))
    return unique_a.view(a.dtype).reshape((unique_a.shape[0], a.shape[1]))

def ineqs(p):
   nVertices = p.n_vertices()
   M = np.zeros(shape=(nVertices,nMsmt))

   i = 0
   for v in p.vertices():
      z = abs(np.array(v))
      for j in np.arange(nMsmt):
         M[i][j] = max(z[j*nOut:(j+1)*nOut])
      i += 1

   Q = unique_rows(M)
   qr, qc = Q.shape
   #qr is 61. There are 61 unique vertices!

   #First we impose the inequality w_i \leq sum_j Q_ij v_j
      b = np.array([np.zeros(nMsmt)])
      W = -1*np.eye(nMsmt)
      A = np.concatenate((b.T,Q.T,W),axis=1)
      #A is an array of size n x (1 + qr + n)

      #Ensure that \nu is a probability distribution
      ##First impose positivity of \nu
      h = np.eye(qr + 1)
      hh = np.zeros(shape=(qr,nMsmt))
      H = np.concatenate((h[1:],hh),axis=1)
      #H is an array of size qr x (1 + qr + n)

      ##Then impose sum to 1
      e = np.array([np.concatenate(([1],-1*np.ones(qr),np.zeros(nMsmt)),axis=1)])
      f = np.array([np.concatenate(([-1],np.ones(qr),np.zeros(nMsmt)),axis=1)])
      #e and f are each arrays of size 1 x (1 + qr + n)

      A = np.concatenate((A,H,e,f),axis=0)

      for m in range(qr):
            print("Done round " + str(m+1) + "/" + str(qr))
            A = FMEliminate(A,1)
            p = Polyhedron(ieqs = A, base_ring = RDF)
            A = np.array(p.inequalities())
      return(A)
```

```
o = ineqs(p)
print(o)
```

Chapter 6

Conclusion

We have presented a way to derive noncontextuality inequalities based on Fourier Motzkin elimination. In each of the experiments that we studied, we began with an operational description of our experimental construction. This is because we wished to make the noncontextuality inequalities independent of quantum theory. In an operational theory, the preparation and measurement procedures are specified as lists of instructions to be performed in the lab.

For each of our constructions, we considered measurement procedures whose outcomes were random variables. We found that there were constraints on the values assigned to *linear* combinations of the random variables by the ontic state.

For each outcome of each measurement, we defined a preparation procedure. These preparation procedures in turn defined coarse-grained preparation procedures. We required that the coarse-grained preparation procedures obey an operational equivalence.

Measurement noncontextuality permitted us to explore the geometry of the ontic expectation values. The fact that the constraints on the random variables were linear was important because this led to a polytope structure for the ontic expectation values. This polytope structure is central to the derivation of the noncontextuality inequalities.

We then proceeded to define the predictability quantities that measure the correlation between the input to the i-th preparation and outcome of the i-th measurement. Using the "Q matrix" construction, we found that we could express the predictabilities in terms of a probability distribution over the vertices of the polytope in the ontic expectation value space and then eliminate these probabilities using FM elimination on the predictabilities alone.

For Belinfante's construction, we obtained a set of necessary and sufficient conditions on the predictability quantities for the experiment to admit of a noncontextual model.

When we moved on to studying the Peres Mermin square construction, brute force FM elimination was not sufficient. The task was computationally difficult and instead we had to resort to exploiting the symmetries of the polytope to make the problem tractable. By observing that the modulus of the ontic expectation values of all the random variables were equal on different vertices, we were able to lump them together to construct aggregate vertices. The redundancy was high enough that we were able to use FM elimination to obtain a set of bounds on the predictability quantities.

Furthermore, using the Peres Mermin square construction as an example, we argued that the distinction between state-dependent and state-independent proofs of contextuality is not meaningful from the perspective of generalized noncontextuality.

Using the modular variables construction, we were able to generalize our noncontextuality inequalities to systems of continuous variables.

We demonstrated that the technique could also be used to study Mermin's star construction so long as we eliminated products of random variables that did not explicitly appear in the star.

Finally we studied the 18 ray construction. Although we modified the definition of the predictability, these quantities also measure the average correlation between the input to the i-th preparation and the output of the i-th measurement. By aggregating vertices, we were able to obtain the noncontextuality inequalities for this construction. Furthermore, it was found that the inequalities we obtained could reproduce the one in [25].

At this juncture, we point out what the algorithm does *not* do. It only considers the correlation between the outcome of measurement $\mathcal{M}_i$ and the choice of preparation $\mathcal{P}_i$. In general we should be able to analyze all the data from an experiment including cases where we have prepared $\mathcal{P}_i^{(\alpha\beta)}$ but measured $\mathcal{M}_j$ where $j \neq i$ and we should look at more refined information than just the average predictabilities to gauge whether or not the experiment permits of an underlying noncontextual model.

In the near future, we hope that these noncontextuality inequalities can be used to clarify resources that are required for quantum information processing.

Bibliography

[1] Omni magazine, May 1988.

[2] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98:230501, Jun 2007.

[3] Leslie E Ballentine. The statistical interpretation of quantum mechanics. *Reviews of Modern Physics*, 42(4):358, 1970.

[4] Stephen D Bartlett, Terry Rudolph, and Robert W Spekkens. Reconstruction of gaussian quantum mechanics from liouville mechanics with an epistemic restriction. *Physical Review A*, 86(1):012103, 2012.

[5] Stephen D Bartlett and Barry C Sanders. Universal continuous-variable quantum computation: Requirement of optical nonlinearity for photon counting. *Physical Review A*, 65(4):042304, 2002.

[6] Stephen D Bartlett, Barry C Sanders, Samuel L Braunstein, and Kae Nemoto. Efficient classical simulation of continuous variable quantum information processes. In *Quantum Information with Continuous Variables*, pages 47–55. Springer, 2003.

[7] Hannes Bartosik, Jürgen Klepp, Claus Schmitzer, Stephan Sponar, Adán Cabello, Helmut Rauch, and Yuji Hasegawa. Experimental test of quantum contextuality in neutron interferometry. *Physical Review Letters*, 103(4):040403, 2009.

[8] F. J. Belinfante. *A survey of hidden variable theories.* Pergamon press, 1973.

[9] John S Bell. On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics*, 38(3):447, 1966.

[10] John Stewart Bell. *Speakable and unspeakable in quantum mechanics: Collected papers on quantum philosophy.* Cambridge university press, 2004.

[11] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Physical Review A*, 71(2):022316, 2005.

[12] Harry Buhrman, Richard Cleve, Serge Massar, and Ronald de Wolf. Nonlocality and communication complexity. *Reviews of modern physics*, 82(1):665, 2010.

[13] Adán Cabello. Experimentally testable state-independent quantum contextuality. *Physical review letters*, 101(21):210401, 2008.

[14] Adán Cabello, Stefan Filipp, Helmut Rauch, and Yuji Hasegawa. Proposed experiment for testing quantum contextuality with neutrons. *Physical Review Letters*, 100(13):130404, 2008.

[15] Rafael Chaves. Polynomial Bell inequalities. *arXiv preprint arXiv:1506.04325*, 2015.

[16] Rafael Chaves, Lukas Luft, and David Gross. Causal structures from entropic information: geometry and novel scenarios. *New Journal of Physics*, 16(4):043001, 2014.

[17] Christopher Ferrie and Joseph Emerson. Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations. *Journal of Physics A: Mathematical and Theoretical*, 41(35):352001, 2008.

[18] Christopher Ferrie, Ryan Morris, and Joseph Emerson. Necessity of negativity in quantum theory. *Physical Review A*, 82(4):044103, 2010.

[19] Richard P Feynman. Negative probability. *Quantum implications: essays in honour of David Bohm*, pages 235–248, 1987.

[20] Daniel Gottesman, Alexei Kitaev, and John Preskill. Encoding a qubit in an oscillator. *Physical Review A*, 64(1):012310, 2001.

[21] Mark Howard, Joel Wallman, Victor Veitch, and Joseph Emerson. Contextuality supplies the 'magic' for quantum computation. *Nature*, 510(7505):351–355, 2014.

[22] Amir Kalev, Ady Mann, Pier A Mello, and Michael Revzen. Inadequacy of a classical interpretation of quantum projective measurements via wigner functions. *Physical Review A*, 79(1):014104, 2009.

[23] Emanuel Knill. Quantum computing with realistically noisy devices. *Nature*, 434(7029):39–44, 2005.

[24] Simon Kochen and Ernst Specker. The problem of hidden variables in quantum mechanics. *Journal of Applied Mathematics and Mechanics*, 17:59–87, 1967.

[25] Ravi Kunjwal and Robert W Spekkens. From the Kochen-Specker theorem to noncontextuality inequalities without assuming determinism. *arXiv preprint arXiv:1506.04150*, 2015.

[26] Yeong-Cherng Liang, Robert W Spekkens, and Howard M Wiseman. Specker s parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. *Physics Reports*, 506(1):1–39, 2011.

[27] L Mandel. Non-classical states of the electromagnetic field. *Physica scripta*, 1986(T12):34, 1986.

[28] Serge Massar and Stefano Pironio. Greenberger-Horne-Zeilinger paradox for continuous variables. *Physical Review A*, 64(6):062108, 2001.

[29] Michael D Mazurek, Matthew F Pusey, Ravi Kunjwal, Kevin J Resch, and Robert W Spekkens. An experimental test of noncontextuality without unwarranted idealizations. *arXiv preprint arXiv:1505.06244*, 2015.

[30] Nicolas C Menicucci. Fault-tolerant measurement-based quantum computing with continuous-variable cluster states. *Physical Review Letters*, 112(12):120504, 2014.

[31] N David Mermin. Hidden variables and the two theorems of John Bell. *Reviews of Modern Physics*, 65(3):803, 1993.

[32] John Von Neumann. *Mathematical foundations of quantum mechanics*. Number 2. Princeton university press, 1955.

[33] Juan Pablo Paz, Salman Habib, and Wojciech H Zurek. Reduction of the wave packet: Preferred observable and decoherence time scale. *Physical Review D*, 47(2):488, 1993.

[34] Asher Peres. Two simple proofs of the Kochen-Specker theorem. *Journal of Physics A: Mathematical and General*, 24(4):L175, 1991.

[35] Matthew F Pusey. The robust noncontextuality inequalities in the simplest scenario. *arXiv preprint arXiv:1506.04178*, 2015.

[36] Ben W Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, 2013.

[37] Wolfram Research. Mathematica 10.1, 2015.

[38] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons ltd., 1986.

[39] Rob Spekkens. Unpublished.

[40] Robert Spekkens. `http://pirsa.org/displayFlash.php?id=14010110`.

[41] Robert W Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A*, 71(5):052108, 2005.

[42] Robert W Spekkens. Negativity and contextuality are equivalent notions of nonclassicality. *Physical review letters*, 101(2):020401, 2008.

[43] Robert W Spekkens. The status of determinism in proofs of the impossibility of a noncontextual model of quantum theory. *Foundations of Physics*, 44(11):1125–1155, 2014.

[44] Robert W Spekkens, Daniel H Buzacott, Anthony J Keehn, Ben Toner, and Geoff J Pryde. Preparation contextuality powers parity-oblivious multiplexing. *Physical review letters*, 102(1):010401, 2009.

[45] Umesh Vazirani and Thomas Vidick. Fully device-independent quantum key distribution. *Physical review letters*, 113(14):140501, 2014.

[46] Victor Veitch, Nathan Wiebe, Christopher Ferrie, and Joseph Emerson. Efficient simulation scheme for a class of quantum optics experiments with non-negative wigner representation. *New Journal of Physics*, 15(1):013037, 2013.

[47] H Paul Williams. Fourier's method of linear programming and its dual. *American Mathematical Monthly*, pages 681–695, 1986.