Inequalities Witnessing Quantum Incompatibility in The Triangle Scenario

Thomas C. Fraser*
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada
University of Waterloo, Waterloo, Ontario, Canada
(Dated: August 24, 2016)

This document is my current working draft of a paper to do with causal inference, inflation, incompatibility inequalities, hypergraph transversals and quantum correlations.

* tcfraser@tcfraser.com

CONTENTS

I. INTRODUCTION

Todo (TC Fraser): Discuss why the marginal problem is important.

II. DEFINITIONS & NOTATION

Definition 1. Borrowing the notation from [1], each random variable v has a set of all possible outcomes called the **outcome space** or **valuation space** and is denoted O_v. When referencing a *specific* element of O_v or *valuation* of v, the notation $o[v]$ is used. This notation generalizes to set of random variables $V = \{v_1, \ldots, v_{|V|}\}$; a specific outcome of $o[V] \in O_V$ is used to reference a particular tuple or vector of outcomes,

$$o[V] \equiv \left(o[v_1], o[v_2], \ldots, o\left[v_{|V|}\right]\right) = (o[v])_{v \in V} \tag{1}$$

It is important to note that the ordering of events in eq. (1) is irrelevant. Similarly, the outcome space O_V for a set of random variables V is the **valuation product** '$\circ$' of the individual outcome spaces,

$$O_V \equiv O_{v_1} \circ \cdots \circ O_{v_{|V|}}$$

Where '$\circ$' is defined such that for $A \cap B = \emptyset$,

$$o[A] \circ o[B] \equiv \left(o[a_1], \ldots, o\left[a_{|A|}\right], o[b_1], \ldots, o\left[b_{|B|}\right]\right) = (o[v])_{v \in A \cup B}$$

Often '$\circ$' will be omitted where there is little ambiguity. When referencing the probability that a set of variables V has outcome $o[V]$ (say $v_1 = 0, v_2 = 3$), we intend for the following class of notations to be used interchangeably.

$$P_V(o[V]) = P(o[V]) = P(o[v_1]\, o[v_2]) = P(v_1 = 0, v_2 = 3) = P(v_2 = 3, v_1 = 0) = P_{v_1,v_2}(03)$$

Definition 2. An outcome $o[V]$ is said to be **extendable** to an outcome $o[W]$ (where $V \subseteq W$) if there exists an outcome $o[W \setminus V]$ such that $o[W] = o[W \setminus V] \circ o[V]$.

$$\mathsf{Ext}(o[V] \to o[W]) \iff \exists o[W \setminus V] \mid o[W] = o[W \setminus V] \circ o[V]$$

The idea being that a *less specific* outcome $o[V]$ can be made *more specific* by assigning outcomes to the remaining random variables in $W \setminus V$. If such an outcome $o[W \setminus V]$ exists, it is unique.

Definition 3. The set of all extendable outcomes of $o[V]$ in O_W is called the **extendable set** and can be written as,

$$\mathsf{Ext}_W(o[V]) \equiv o[V] \circ O_{W \setminus V} \equiv \{o[W] \in O_W \mid \mathsf{Ext}(o[V] \to o[W])\}$$

The extendable set of $o[V]$ in W is the set of all outcomes of O_W that *agree* with $o[V]$ about valuations for variables in V. In this language, two outcomes $o[V]$ and $o[W]$ are **compatible** if they are both extendable to some outcome $o[V \cup W]$. Equivalently, $o[V]$ and $o[W]$ agree on their valuations of $V \cap W$.

$$\mathsf{Com}(o[V], o[W]) \iff \exists o[V \cup W] \mid \mathsf{Ext}(o[V] \to o[V \cup W]), \mathsf{Ext}(o[W] \to o[V \cup W])$$

Example 4. Consider two sets of random variables $V = \{a, b\}$ and $W = \{a, b, c\}$. Clearly $V \subseteq W$; a prerequisite for extendability. Also take all individual outcome spaces to be finite and of order 3: $O_a = O_b = O_c = \{1, 2, 3\}$. Then $o[V] = o[\{a, b\}] = (a = 1, b = 2)$ is extendable to the outcome $o[W] = (a = 1, b = 2, c = 1)$, and the extendable set of $o[V]$ in O_W is,

$$\mathsf{Ext}_{a,b,c}(a = 1, b = 2) = \{(a = 1, b = 2, c = 1), (a = 1, b = 2, c = 2), (a = 1, b = 2, c = 3)\}$$

Definition 5. A **graph** is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes can represent any object and the edges are pairs of nodes. For convenience of notation, one defines an index set over the nodes denoted $\mathcal{I}_{\mathcal{N}}$.

$$\mathcal{N} = \{n_i \mid i \in \mathcal{I}_{\mathcal{N}}\} \quad \mathcal{E} = \{\{n_j, n_k\} \mid j, k \in \mathcal{I}_{\mathcal{N}}\}$$

Definition 6. A **directed graph** $\mathcal{G}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes can represent any object and the edges are *ordered* pairs of nodes. For convenience of notation, one defines an index set over the nodes denoted $\mathcal{I}_{\mathcal{N}}$.

$$\mathcal{N} = \{n_i \mid i \in \mathcal{I}_{\mathcal{N}}\} \quad \mathcal{E} = \{n_j \to n_k \mid j, k \in \mathcal{I}_{\mathcal{N}}\}$$

Definition 7. The following definitions are common language in directed graph theory. Let $n, m \in \mathcal{N}$ be example nodes of the graph $\mathcal{G}$.

- The **parents of a node**: $\mathsf{Pa}_{\mathcal{G}}(n) \equiv \{m \mid m \to n\}$
- The **children of a node**: $\mathsf{Ch}_{\mathcal{G}}(n) \equiv \{m \mid n \to m\}$
- The **ancestry of a node**: $\mathsf{An}_{\mathcal{G}}(n) \equiv \bigcup_{i\in\mathbb{W}} \mathsf{Pa}^{i}_{\mathcal{G}}(n)$ where $\mathsf{Pa}^{i}_{\mathcal{G}}(n) \equiv \mathsf{Pa}_{\mathcal{G}}\big(\mathsf{Pa}^{i-1}_{\mathcal{G}}(n)\big)$ and $\mathsf{Pa}^{0}_{\mathcal{G}}(n) = n$

All of these terms can be generalized to sets of nodes $N \subseteq \mathcal{N}$ through union over the elements,

- The **parents of a node set**: $\mathsf{Pa}_{\mathcal{G}}(N) \equiv \bigcup_{n\in N} \mathsf{Pa}_{\mathcal{G}}(n)$
- The **children of a node set**: $\mathsf{Ch}_{\mathcal{G}}(N) \equiv \bigcup_{n\in N} \mathsf{Ch}_{\mathcal{G}}(n)$
- The **ancestry of a node set**: $\mathsf{An}_{\mathcal{G}}(N) \equiv \bigcup_{n\in N} \mathsf{An}_{\mathcal{G}}(n)$

Moreover, an **induced subgraph** of $\mathcal{G}$ due to a set of nodes $N \subseteq \mathcal{N}$ is the graph composed of N and all edges $e \in \mathcal{E}$ of the original graph that are contained in N.

$$\mathsf{Sub}_{\mathcal{G}}(N) \equiv (N, \{e_i \mid i \in \mathcal{I}_{\mathcal{E}}, e_i \subseteq N\})$$

An **ancestral subgraph** of $\mathcal{G}$ due to $N \subseteq \mathcal{N}$ is the induced subgraph due to the ancestry of N.

$$\mathsf{AnSub}_{\mathcal{G}}(N) \equiv \mathsf{Sub}_{\mathcal{G}}\big(\mathsf{An}_{\mathcal{G}}(N)\big)$$

Definition 8. A **directed acyclic graph** or **DAG** $\mathcal{G}$ is an directed graph definition 6 with the additional property that no node n is in its set of **ancestors**.

$$\forall n \in \mathcal{N} : n \notin \bigcup_{i\in\mathbb{N}} \mathsf{Pa}^{i}_{\mathcal{G}}(n)$$

Notice the difference between using the natural numbers $\mathbb{N}$ to distinguish *ancestors* from *ancestry*.

Definition 9. A **hypergraph** denoted $\mathcal{H}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *hyperedges* respectively where the nodes can represent any object and the hyperedges are *subsets* of nodes. For convenience of notation, one defines an index set over the nodes and hyperedges of a hypergraph $\mathcal{H}$ denoted $\mathcal{I}_{\mathcal{N}}$ and $\mathcal{I}_{\mathcal{E}}$ respectively.

$$\mathcal{H} = (\mathcal{N}, \mathcal{E}) \quad \mathcal{N} = \{n_i \mid i \in \mathcal{I}_{\mathcal{N}}\} \quad \mathcal{E} = \{e_i \mid i \in \mathcal{I}_{\mathcal{E}}, e_i \subseteq \mathcal{N}\}$$

Note that whenever the hyperedge or node index is arbitrary, it will be omitted. There is a dual correspondence between hyperedges $e \in \mathcal{E}$ and nodes $n \in \mathcal{N}$ in a Hypergraph. A hyperedge e is viewed as a set of nodes $\{n_i\}$, and a node n can be viewed as the set of hyperedges $\{e_i\}$ that contain it.

Definition 10. A **hypergraph transversal** (or edge hitting set) $\mathcal{T}$ of a hypergraph $\mathcal{H}$ is a set of nodes $\mathcal{T} \subseteq \mathcal{N}$ that have non-empty intersections with every hyperedge in $\mathcal{E}$.

$$\mathcal{T} = \{n_i \in \mathcal{N} \mid i \in \mathcal{I}_{\mathcal{T}}\} \quad \forall e \in \mathcal{E} : \mathcal{T} \cap e \neq \emptyset$$

Definition 11. A **necessary node** of a transversal $\mathcal{T}$ is a node n such that $\mathcal{T} \setminus n$ is no longer a valid transversal. The set of all necessary nodes is denoted $\mathsf{Nec}(\mathcal{T})$,

$$\mathsf{Nec}(\mathcal{T}) = \{n \in \mathcal{T} \mid \exists e \in \mathcal{E} : (\mathcal{T} \setminus n) \cap e = \emptyset\}$$

An **unnecessary node** of a transversal $\mathcal{T}$ is any node that is *not* necessary. The set of all unnecessary nodes is denoted $\mathsf{UnNec}(\mathcal{T})$,

$$\mathsf{UnNec}(\mathcal{T}) = \mathcal{T} \setminus \mathsf{Nec}(\mathcal{T})$$

A **minimal hypergraph transversal** $\mathcal{T}$ is any valid transversal of $\mathcal{H}$ where every node n is *necessary*.

$$\mathcal{T} = \mathsf{Nec}(\mathcal{T})$$

Definition 12. A **weighted hypergraph** $\mathcal{H}_{\mathcal{W}}$ is a regular hypergraph satisfying definition 9 equipped with a set of weights $\mathcal{W}$ ascribed to each node such that a weighted hypergraph is written as a triplet $(\mathcal{W}, \mathcal{N}, \mathcal{E})$.

$$\mathcal{W} = \{w_i \mid i \in \mathcal{I}_{\mathcal{N}}, w_i \in \mathbb{R}\}$$

One would say that a particular node n_i carries weight w_i for each $i \in \mathcal{I}_{\mathcal{N}}$.

Definition 13. A **bounded transversal** of a weighted hypergraph $\mathcal{H}_{\mathcal{W}}$ is a transversal $\mathcal{T}$ of the unweighted hypergraph $\mathcal{H}$ and a real number t (denoted $\mathcal{T}_{\leq t}$) such that the sum of the node weights of the transversal is bounded by t.

$$\mathcal{T}_{\leq t} = \{n_i \mid i \in \mathcal{I}_{\mathcal{T}}\} \quad \text{s.t.} \sum_{j \in \mathcal{I}_{\mathcal{T}}} w_j \leq t$$

One can definte analogous **(strictly) upper/lower bounded transversals** by considering modifications of the notation: $\mathcal{T}_{<t}, \mathcal{T}_{\geq t}, \mathcal{T}_{>t}$.

Definition 14. A **causal structure** is simply a DAG with the extra classification of each node into one of two categories; the **latent nodes** and **observed nodes** denoted $\mathcal{N}_L$ and $\mathcal{N}_O$. The latent nodes correspond to random variables that are either hidden through some fundamental process or cannot/will not be measured. The observed nodes are random variables that are measurable. Every node is either latent or observed and no node is both:

$$\mathcal{N}_L \cap \mathcal{N}_O = \emptyset \qquad \mathcal{N}_L \cup \mathcal{N}_O = \mathcal{N}$$

Definition 15. The **product distribution** two distributions is denoted as usual with $\times$ and is defined as,

$$(P_v \times P_w)\left(o[v], o[w]\right) \equiv P_v(o[v])\, P_w(o[w])$$

A product distribution of k distributions is defined recursively,

$$\prod_{i=1}^{k} P_{v_i} \equiv \left(P_{v_1} \times \cdots \times P_{v_k}\right)$$

Definition 16. The **marginalization** of a distribution $P_{v\cup w}$ to the distribution P_v is denoted $\sum_w P_{v\cup w} = P_v$ and is defined such that,

$$\forall o[v] \in O_v : \left(\sum_w P_{v,w}\right)(o[v]) \equiv \sum_{o[w]\in O_w} P_{v,w}(o[v], o[w])$$

Todo (TC Fraser): How many definitions do I need to write??

III. TRIANGLE SCENARIO

Todo (TC Fraser): Discuss the Triangle Scenario, previous work done on it, etc. Focusing on the inflation depicted in fig. 2, we obtained the maximally pre-injectable sets through the procedure outlined in [2].

Maximal Pre-injectable Sets	**Ancestral Independences**
$\{A_1, B_1, C_1, A_4, B_4, C_4\}$	$\{A_1, B_1, C_1\} \perp \{A_4, B_4, C_4\}$
$\{A_1, B_2, C_3, A_4, B_3, C_2\}$	$\{A_1, B_2, C_3\} \perp \{A_4, B_3, C_2\}$
$\{A_2, B_3, C_1, A_3, B_2, C_4\}$	$\{A_2, B_3, C_1\} \perp \{A_3, B_2, C_4\}$
$\{A_2, B_4, C_3, A_3, B_1, C_2\}$	$\{A_2, B_4, C_3\} \perp \{A_3, B_1, C_2\}$
$\{A_1, B_3, C_4\}$	$\{A_1\} \perp \{B_3\} \perp \{C_4\}$
$\{A_1, B_4, C_2\}$	$\{A_1\} \perp \{B_4\} \perp \{C_2\}$
$\{A_2, B_1, C_4\}$	$\{A_2\} \perp \{B_1\} \perp \{C_4\}$
$\{A_2, B_2, C_2\}$	$\{A_2\} \perp \{B_2\} \perp \{C_2\}$
$\{A_3, B_3, C_3\}$	$\{A_3\} \perp \{B_3\} \perp \{C_3\}$
$\{A_3, B_4, C_1\}$	$\{A_3\} \perp \{B_4\} \perp \{C_1\}$
$\{A_4, B_1, C_3\}$	$\{A_4\} \perp \{B_1\} \perp \{C_3\}$
$\{A_4, B_2, C_1\}$	$\{A_4\} \perp \{B_2\} \perp \{C_1\}$

(2)

As can be counted, there are 12 maximally pre-injectable sets which will be indexed 1 through 12 in the order seen above ($\{V_1, \ldots, V_{12}\}$)

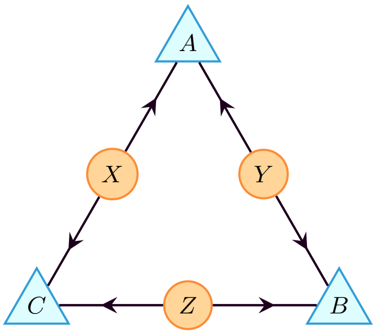

FIG. 1. The casual structure of the Triangle Scenario. Three variables A, B, C are observable and illustrated as triangles, while X, Y, Z are latent variables illustrated as circles.

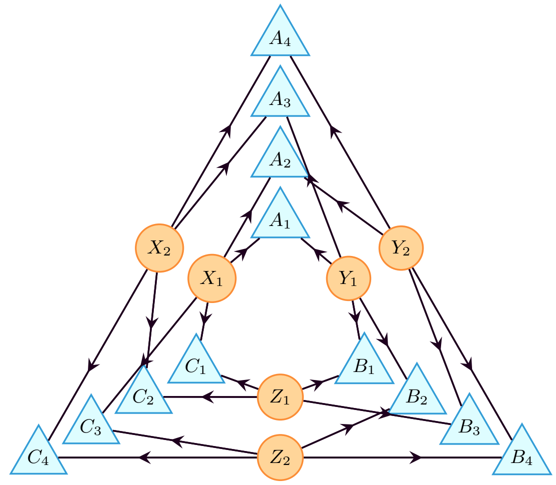

FIG. 2. An inflated causal structure of the Triangle Scenario fig. 1.

IV. COMPATIBILITY, CONTEXTUALITY AND THE MARGINAL PROBLEM

In order to determine if a given marginal distribution P_V or set of marginal distributions $\{P_{V_1}, \dots, P_{V_k}\}$ is compatible with a causal structure $\mathcal{G}$, one should first formalize what is meant by *compatible*.

Definition 17. A set of **causal parameters** for a particular causal structure $\mathcal{G}$ is the specification of a conditional distribution for every node $n \in \mathcal{N}$ given it's parents in $\mathcal{G}$.

$$\left\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\right\}$$

Definition 18. A marginal distribution P_V is **compatible** with a causal structure $\mathcal{G}$ (where it is assumed that $V \subseteq \mathcal{N}_O$) if there exists a *choice* of causal parameters $\left\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\right\}$ such that P_V can be *recovered* from the following series of operations:

Todo (TC Fraser): Define this notation here

1. First obtain a joint distribution over *all* nodes of of the causal structure,

$$P_{\mathcal{N}} = \prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)}$$

2. Then marginalize over the latent nodes of $\mathcal{G}$,

$$P_{\mathcal{N}_O} = \sum_{\mathcal{N}_L} P_{\mathcal{N}}$$

3. Finally marginalize over the observed nodes not in V to obtain P_V,

$$P_V = \sum_{\mathcal{N}_O \setminus V} P_{\mathcal{N}_O}$$

A set of marginal distributions $\{P_{V_1}, \dots, P_{V_k}\}$ is compatible with $\mathcal{G}$ if each of the distributions can be made compatible by the *same* choice of causal parameters. A distribution P_V or set of distributions $\{P_{V_1}, \dots, P_{V_k}\}$ is said to be **incompatible** with a causal structure if there *does not exist* a set of causal parameters with the above mentioned property.

Todo (TC Fraser): Source this?

Operations 2 and 3 of definition 18 are related to the *marginal problem*.

Definition 19. **The Marginal Problem:** Given a set of distributions $\{P_{V_1}, \ldots, P_{V_k}\}$ where $V_i \subseteq \mathcal{V}$ for some set of random variables $\mathcal{V}$ and $k \geq 2$, does there exist a joint distribution $P_{\mathcal{V}}$ such that each given distribution P_{V_i} can be obtained from marginalizing $P_{\mathcal{V}}$?

$$\forall i \in \{1, \ldots, k\} : P_{V_i} = \sum_{\mathcal{V} \setminus V_i} P_{\mathcal{V}}$$

Typically (although not strictly necessary), $\mathcal{V}$ is taken to mean the union of all V_i's.

$$\mathcal{V} = V_1 \cup \cdots \cup V_k = \bigcup_{i=1}^{k} V_i \tag{3}$$

Todo (TC Fraser): Mention polytope perspective and existing methods

Definition 20. A reoccurring motif of these discussions will be the set of distributions $\{P_{V_1}, \ldots, P_{V_k}\}$ mentioned in definition 19. In agreement with [3] we will call this set of distributions a **marginal model** and denote it $P^{\mathcal{M}}$ provided that they are *compatible*:

$$\forall i \neq j \text{ if } V_i \cap V_j \neq \emptyset \text{ then } \sum_{V_i \setminus V_j} P_{V_i} = \sum_{V_j \setminus V_i} P_{V_j}$$

We call the set of subsets $\{V_1, \ldots, V_k\}$ the marginal contexts or the **maximal marginal scenario** $\mathcal{M}$[1] and an individual V_i a **marginal context**. The union of all contexts is denoted $\mathcal{V}$ and define it exactly as in eq. (3).

Following definition 2.3 in [3], a marginal model is said to be **contextual** if it *does not* admit a solution to the marginal problem and **non-contextual** otherwise.

In addition, we elect to define the **marginal outcomes** $O^{\mathcal{M}}$ to be the set of all outcomes belonging to outcomes of the marginal contexts.

$$O^{\mathcal{M}} \equiv O_{V_1} \cup \cdots \cup O_{V_k} = \bigcup_{j=1}^{k} O_{V_j}$$

Note that $O^{\mathcal{M}}$ is *not* a valid outcome space, analogous to the fact that $P^{\mathcal{M}}$ is *not* a probability distribution. Instead $O^{\mathcal{M}}$ is a collection of outcome spaces just as $P^{\mathcal{M}}$ is a collection of distributions.

Definition 21. A **causal compatibility inequality** $\mathcal{I}_{\mathcal{M}}$ for a marginal scenario $\mathcal{M}$ is a probabilistic inequality that is obeyed for every compatible marginal model $P^{\mathcal{M}}$. Whenever the marginal scenario $\mathcal{M}$ in question is evident by context, the subscript of $\mathcal{I}_{\mathcal{M}}$ will be dropped leaving $\mathcal{I}$.

Todo (TC Fraser): Discuss Compatibility, connection to cooperative games/resources, bell incompatibility? Todo (TC Fraser): Connection between contextuality and Compatibility via the marginal problem for causal parameters Todo (TC Fraser): Discuss what is meant by a 'complete' solution to the marginal problem Todo (TC Fraser): Maybe define the possibilistic marginal problem for later

V. SUMMARY OF THE INFLATION TECHNIQUE

The causal inflation technique, first pioneered by Wolfe, Spekkens, and Fritz [2] and inspired by the *do calculus* and *twin networks* of Ref. [4], is a family of causal inference techniques that can be used to determine if a probability distribution is compatible or incompatible with a given causal structure. As a preliminary summary, the inflation technique begins by *augmenting* a causal structure with additional copies of its nodes, producing an *inflated* causal structure, and then exposes how causal inference tasks on the inflated causal structure can be used to make inferences on the original causal structure. Copies of the original nodes are distinguished by an additional subscript called the **copy-index**. For example node A of fig. 1 has copies A_1, A_2, A_3, A_4 in the inflation of fig. 2. Following the notions

[1] Rigorously speaking as defined by [3], a marginal scenario forms an *abstract simplicial complex* where $V' \in \mathcal{M}$ whenever $V' \subseteq V_i$ for some i. "Maximal" refers to the restriction that $\forall i, j : V_i \not\subseteq V_j$.

of Ref. [2], all such copies are deemed equivalent via a **copy-index equivalence** relation denoted '$\sim$'. A copy-index is effectively arbitrary, so A' will refer to an arbitrary inflated copy of A.

$$A \sim A_1 \sim A' \nsim B \sim B_1 \sim B'$$

Equipped with the common graph-theoretic terminology and notation of definition 7, an inflation can be formally defined as follows:

Definition 22. An **inflation** of a causal structure $\mathcal{G}$ is another causal structure $\mathcal{G}'$ such that:

$$\forall n' \in \mathcal{N}' : \mathsf{AnSub}_{\mathcal{G}'}(n') \sim \mathsf{AnSub}_{\mathcal{G}}(n) \tag{4}$$

Where the notion of copy-index equivalence '$\sim$' is *extended* to two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ if they are equal upon removal of copy-index.

The motivation being that since the ancestry of each node n' in $\mathcal{G}'$ plays the *exact same* role as the ancestry of its source copy n in $\mathcal{G}$, then every compatible joint distribution $P_{\mathcal{N}}$ on $\mathcal{G}$ can be used to *create* compatible joint distributions on $\mathcal{G}'$. To do this, first notice that every compatible joint distribution $P_{\mathcal{N}}$ uniquely defines a set of causal parameters for $\mathcal{G}$.

$$\forall n \in \mathcal{N} : P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} = \sum_{\mathcal{N} \setminus n} P_{\mathcal{N}} \tag{5}$$

Now since eq. (4) enforces that $\mathsf{Pa}_{\mathcal{G}'}(n') \sim \mathsf{Pa}_{\mathcal{G}}(n)$, one can create a set of **inflated causal parameters** using eq. (5),

$$\forall n' \in \mathcal{N}' : P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \equiv P_{n|\mathsf{Pa}_{\mathcal{G}}(n)}$$

Which in turn uniquely defines a compatible joint distribution $P_{\mathcal{N}'}$ on $\mathcal{G}'$[2].

$$P_{\mathcal{N}'} = \prod_{n' \in \mathcal{N}'} P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')}$$

This is known as the **weak inflation lemma**; compatible joint distributions over $\mathcal{N}$ induce compatible joint distributions over $\mathcal{N}'$. Before generalizing to the inflation lemma, a careful observation needs to be made. For any pair of subsets of nodes $N \subseteq \mathcal{N}$ and $N' \subseteq \mathcal{N}'$ that are equivalent up to copy-index $N \sim N'$, if $\mathsf{AnSub}_{\mathcal{G}}(N) \sim \mathsf{AnSub}_{\mathcal{G}'}(N')$ then any compatible *marginal* distribution P_N over N induces a compatible marginal distribution $P_{N'}$ over N'. In fact since N' contains no duplicate nodes up to copy index (simply because $N \sim N'$ and N cannot contain duplicate nodes), $P_N = P_{N'}$. We refer such sets N' as the **injectable sets** of $\mathcal{G}'$ and N the **images of the injectable sets** of $\mathcal{G}$.

$$\begin{aligned} \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}') &\equiv \{N' \subseteq \mathcal{N}' \mid \exists N \subseteq \mathcal{N} : N \sim N'\} \\ \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}') &\equiv \{N \subseteq \mathcal{N} \mid \exists N' \subseteq \mathcal{N}' : N \sim N'\} \end{aligned}$$

Lemma 23. ***The Inflation Lemma*** *Given a particular inflation $\mathcal{G}'$ of $\mathcal{G}$, if a marginal model $\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$ is compatible with $\mathcal{G}$ then all marginal models $\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\}$ are compatible with $\mathcal{G}'$ provided that $P_N = P_{N'}$ for all instances where $N \sim N'$. This is lemma 3 of [2].*

The inflation lemma is the most important result of the causal inflation technique [2]. The contrapositive version of lemma 23 is a powerful tool for determining compatibility. Any compatibility constraint on marginal models $\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\}$ of the inflated causal structure $\mathcal{G}'$ correspond to valid compatibility constraints on marginal models $\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$ of the original causal structure. Corollary 5 of [2] proves this explicitly for incompatibility inequalities $\mathcal{I}$; which the remainder of this work focuses on. Additionally, the inflation lemma holds when considering any subset of $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ (analogously $\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')$). Therefore, in situations where latent nodes are present in $\mathcal{G}$, one only needs to consider injectable sets that are composed of *observable nodes*.

In this work, we obtain inequalities that constrain the set of injectable marginal models $\{P_{N'} \mid N' \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$ for the inflated Triangle Scenario of fig. 1 by considering the marginal problem (see section IV) for the set of **maximally**

[2] Of course not all compatible joint distributions on $\mathcal{N}'$ are constructed in this way; all that is demonstrated is that all joint distributions constructed in this way are compatible.

pre-injectable sets[3]. As defined by [2], a pre-injectable set V is a subset of $\mathcal{N}'$ that can be written as the decomposed into the union of injectable sets $V = \bigcup_i N'_i$ where $\left\{N'_i \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\right\}$ that are mutually **ancestrally independent**,

$$\forall i, j : N'_i \perp N'_j \iff \mathsf{An}_{\mathcal{G}'}(N'_i) \cap \mathsf{An}_{\mathcal{G}'}\left(N'_j\right) = \emptyset$$

In doing so, any distribution over all nodes of a pre-injectable set $P_{\cup_i N'_i}$ will factorize according to graphical d-separation conditions [4],

$$P_V = P_{\cup_i N'_i} = \prod_i P_{N'_i}$$

This turns linear inequalities over the pre-injectable distributions into polynomial inequalities over the injectable distributions, allowing one to replace all such distributions with equivalent distributions over the original random variables of $\mathcal{N}$. $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')$ will denote the set of all pre-injectable sets.

VI. THE FRITZ DISTRIBUTION

The **Fritz distribution** P_F is a quantum-accessible distribution known to be incompatible with the Triangle Scenario. Explicitly, P_F is a three-party (A, B, C), four-outcome $(1, 2, 3, 4)$ distribution that has form as follows:

$$\begin{aligned} P_F(111) = P_F(221) = P_F(412) = P_F(322) = P_F(233) = P_F(143) = P_F(344) = P_F(434) &= \frac{1}{32}\left(2+\sqrt{2}\right) \\ P_F(121) = P_F(211) = P_F(422) = P_F(312) = P_F(243) = P_F(133) = P_F(334) = P_F(444) &= \frac{1}{32}\left(2-\sqrt{2}\right) \end{aligned} \tag{6}$$

Here the notation $P_F(abc) = P_{ABC}(abc) = P(A = a, B = b, C = c)$ is used. The Fritz distribution P_F can be realized with the following quantum configuration:

$$\begin{gathered} \rho_{AB} = \left|\Psi^+\right\rangle\left\langle\Psi^+\right| \quad \rho_{BC} = \rho_{CA} = \left|\Phi^+\right\rangle\left\langle\Phi^+\right| \\ M_A = \left\{\left|0\psi_0\right\rangle\left\langle 0\psi_0\right|, \left|0\psi_\pi\right\rangle\left\langle 0\psi_\pi\right|, \left|1\psi_{-\pi/2}\right\rangle\left\langle 1\psi_{-\pi/2}\right|, \left|1\psi_{\pi/2}\right\rangle\left\langle 1\psi_{\pi/2}\right|\right\} \\ M_B = \left\{\left|\psi_{\pi/4}0\right\rangle\left\langle\psi_{\pi/4}0\right|, \left|\psi_{5\pi/4}0\right\rangle\left\langle\psi_{5\pi/4}0\right|, \left|\psi_{3\pi/4}1\right\rangle\left\langle\psi_{3\pi/4}1\right|, \left|\psi_{-\pi/4}1\right\rangle\left\langle\psi_{-\pi/4}1\right|\right\} \\ M_C = \left\{\left|00\right\rangle\left\langle 00\right|, \left|01\right\rangle\left\langle 01\right|, \left|10\right\rangle\left\langle 10\right|, \left|11\right\rangle\left\langle 11\right|\right\} \end{gathered} \tag{7}$$

Where for convenience of notation ψ_x is used to denote the superposition,

$$\left|\psi_x\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0\right\rangle + e^{ix}\left|1\right\rangle\right)$$

Additionally $\left|\Psi^+\right\rangle = \frac{1}{\sqrt{2}}\left(\left|01\right\rangle + \left|10\right\rangle\right)$ and $\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}\left(\left|00\right\rangle + \left|11\right\rangle\right)$ are two maximally entangled Bell states. Fritz first proved it's incompatibility [5] by showing C acts a moderator to ensure measurement pseudo-settings for A and B are independent, satisfying non-broadcasting requirements for the standard Bell scenario. In fact, by coarse-graining outcomes for A and B and treating C as a measurement-setting moderator, P_F maximally violates the CHSH inequality. To illustrate this, begin with the CHSH inequality [6],

$$\langle AB|S_A = 1, S_B = 1\rangle + \langle AB|S_A = 1, S_B = 2\rangle + \langle AB|S_A = 2, S_B = 1\rangle - \langle AB|S_A = 2, S_B = 2\rangle \leq 2 \tag{8}$$

Where $\langle AB|S_A = i, S_B = j\rangle$ is the correlation between A and B given the measurement settings for A (B) is i (j) respectively. Next each of C's outcomes become the conditioned settings in eq. (8),

$$\langle AB|C = 2\rangle + \langle AB|C = 3\rangle + \langle AB|C = 4\rangle - \langle AB|C = 1\rangle \leq 2 \tag{9}$$

Finally, specifying the two outcome coarse-graining as $\{1, 2, 3, 4\} \to \{(1, 4), (2, 3)\}$ gives a definition of correlation,

$$\langle AB\rangle \equiv 2\left\{P(11) + P(14) + P(41) + P(44) + P(22) + P(23) + P(32) + P(33)\right\} - 1$$

[3] The set of all pre-injectable sets forms a topological *simplicial conplex*. 'Maximal' refers to the pre-injectable sets that are no proper subset of another.

Which when applied to the Fritz distribution eq. (6) gives the following correlations:

$$\langle AB|C=2\rangle = 2\left\{\frac{1}{32}\left[2\left(2+\sqrt{2}\right)+2\left(2-\sqrt{2}\right)\right]\right\}-1=\frac{\sqrt{2}}{2}$$
$$\langle AB|C=3\rangle = \langle AB|C=4\rangle = \frac{\sqrt{2}}{2} \quad \langle AB|C=1\rangle = -\frac{\sqrt{2}}{2}$$

Which when applied to eq. (9) gives the familiar Tsirelson violation of $2\sqrt{2} \not\leq 2$ [7].

Before continuing it is worth noting that eq. (6) is non-unique. Any distribution that is equal to eq. (6) via a permutation of outcomes or exchange of parties is also referred to as a Fritz distribution. Moreover, the quantum realization of eq. (7) is non-unique. In fact, eq. (7) is not the realization that Fritz originally had in mind. Nonetheless eq. (6) is taken as *the* Fritz distribution for concreteness throughout this paper.

Todo (TC Fraser): Summarize Problem 2.17 in fritz BBT, make it more formal

VII. CERTIFICATE INEQUALITIES

A. Casting the Inflated Marginal Problem as a Linear Program

After obtaining the maximal pre-injectable sets associated with a particular inflation, one can write the marginal problem of definition 19 as a linear program. The key observation is that marginalization is a *linear* operator that can be performed via a matrix multiplication. To do this, we will define the *marginalization matrix*.

Definition 24. The **marginalization matrix** M for a marginal scenario $\{V_1,\dots,V_k\}$ is a bit-wise matrix where the columns are indexed by *joint* outcomes $o[\mathcal{V}] \in O_{\mathcal{V}}$ and the rows are indexed by *marginal* outcomes corresponding to all outcomes $o[V_i] \in O_{V_i} \subseteq O^{\mathcal{M}}$. The entries of M are populated whenever a row index is extendable to a column index.

$$M_{(o[V_i],o[\mathcal{V}])} = \begin{cases} 1 & \mathsf{Ext}(o[V_i] \to o[\mathcal{V}]) \\ 0 & \text{otherwise} \end{cases}$$

The marginalization matrix has $|O_{\mathcal{V}}|$ columns and $\left|O^{\mathcal{M}}\right| = \sum_{i=1}^{k} |O_{V_i}|$ rows. The number of non-zero entries of M is a simple expression,

$$\sum_{i=1}^{k} |O_{V_i}|\left|O_{\mathcal{V}\setminus V_i}\right| = \sum_{i=1}^{k} |O_{\mathcal{V}}| = k\,|O_{\mathcal{V}}|$$

Each of the k elements of $\{V_1,\dots,V_k\}$ contributes a single non-zero entry to each column of M, resulting in $k\,|O_{\mathcal{V}}|$ total non-zero entries.

Note that the row and column indices of the marginalization matrix will be referred to very frequently. We will refer to the

Todo (TC Fraser): Computationally Efficient generation?

To illustrate this concretely, consider the following example:

Example 25. Suppose one has 4 binary random variables $\mathcal{V} = \{a,b,c\}$ in mind and 2 subsets $\{\{a,c\},\{b\}\}$. Then the marginalization matrix is:

$$M = \begin{array}{c} (a,b,c)= \\ (a{=}0,c{=}0) \\ (a{=}0,c{=}1) \\ (a{=}1,c{=}0) \\ (a{=}1,c{=}1) \\ (b{=}0) \\ (b{=}1) \end{array} \begin{array}{c} \begin{array}{cccccccc} (0,0,0) & (0,0,1) & (0,1,0) & (0,1,1) & (1,0,0) & (1,0,1) & (1,1,0) & (1,1,1) \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

In order to describe how marginalization can be written as matrix multiplication $M \cdot x = b$, we need to describe how to define two more quantities:

Definition 26. The **joint distribution vector** $\mathcal{P}_\mathcal{V}$ for a probability distribution $P_\mathcal{V}$ is the vector whose entries are the positive, real-valued probabilities that $P_\mathcal{V}$ assigns to each outcome of $o[\mathcal{V}]$ of $O_\mathcal{V}$. $\mathcal{P}_\mathcal{V}$ shares the same indices as the *column* indices of M.

$$\mathcal{P}_\mathcal{V}^\intercal = [P_\mathcal{V}(o[\mathcal{V}])]_{o[\mathcal{V}]\in O_\mathcal{V}}$$

Definition 27. The **marginal distribution vector** $\mathcal{P}_{\{V_1,\dots,V_k\}}$ for a marginal model $\{P_{V_1},\dots,P_{V_k}\}$ is the vector whose entries are probabilities over the set of marginal outcomes $O^\mathcal{M}$. $\mathcal{P}_{\{V_1,\dots,V_k\}}$ shares the same indices as the *row* indices of M.

$$\mathcal{P}_{\{V_1,\dots,V_k\}}^\intercal = [P_{V_i}(o[V_i])]_{o[V_i]\in O^\mathcal{M}}$$

The marginal and joint distribution vectors are related via the marginalization matrix M. Given a joint distribution vector $\mathcal{P}_\mathcal{V}$ one can obtain the marginal distribution vector $\mathcal{P}_{\{V_1,\dots,V_k\}}$ by multiplying M by $\mathcal{P}_\mathcal{V}$.

$$\mathcal{P}_{\{V_1,\dots,V_k\}} = M \cdot \mathcal{P}_\mathcal{V} \tag{10}$$

Todo (TC Fraser): Discuss non-unique but consistent ordering of M, $\mathcal{P}_\mathcal{V}$ and $\mathcal{P}_{\{V_1,\dots,V_k\}}$

The marginal problem can now be rephrased in the language of the marginalization matrix. Suppose one obtains a marginal distribution vector $\mathcal{P}_{\{V_1,\dots,V_k\}}$. The marginal problem becomes equivalent to the question: *Does there exist a joint distribution vector $\mathcal{P}_\mathcal{V}$ such that eq. (10) holds?*

Definition 28. **The Marginal Linear Program** is the following linear program:

$$\begin{aligned} \text{minimize:} \quad & \emptyset \cdot x \\ \text{subject to:} \quad & x \succeq 0 \\ & M \cdot x = \mathcal{P}_{\{V_1,\dots,V_k\}} \end{aligned}$$

If this "optimization"[4] is *feasible*, then there exists a vector x than can satisfy eq. (10) and is a valid joint distribution vector. Therefore feasibility implies that $P_\mathcal{V} = x$, solving the marginal problem with positive result. Moreover if the marginal linear program is *infeasible*, then there *does not* exist a joint distribution $P_\mathcal{V}$ over all random variables.

Definition 29. **The Dual Marginal Linear Program** is the dual of definition 28 formulated via a procedure similar to [8]:

$$\begin{aligned} \text{minimize:} \quad & y \cdot \mathcal{P}_{\{V_1,\dots,V_k\}} \\ \text{subject to:} \quad & y \cdot M \succeq 0 \end{aligned}$$

Where y is a real valued vector with the same length as $\mathcal{P}_{\{V_1,\dots,V_k\}}$.

B. Infeasibility Certificates

The dual marginal linear program also provides an answer to the marginal problem. To prove this, first notice that the dual problem is *never infeasible*; by choosing y to be trivially the null vector $\emptyset$ of appropriate size, all constraints are satisfied. Secondly if $y \cdot M \succeq 0$ and $x \succeq 0$, then the following must hold if the primal is feasible:

$$y \cdot \mathcal{P}_{\{V_1,\dots,V_k\}} = y \cdot M \cdot x \geq 0 \tag{11}$$

Therefore the *sign* of the dual value $d \equiv \min\left(y \cdot \mathcal{P}_{\{V_1,\dots,V_k\}}\right)$ solves the marginal problem. If $d < 0$ then eq. (11) is violated and therefore the marginal problem has negative result. Likewise if d satisfies eq. (11), then a joint distribution $P_\mathcal{V}$ exists. Before continuing, an important observation needs to be made. If $d \geq 0$, then it is exactly $d = 0$, due to the existence of the trivial $y = \emptyset$. This observation is an instance of the *Complementary Slackness Property* of [9]. Comment (TC Fraser): Is this really the CSP? Moreover, if $d < 0$, then it is unbounded $d = -\infty$.

[4] "Optimization" is presented in quotes here because the minimization objective is trivially always zero ($\emptyset$ denotes the null vector of all zero entries). The primal value of the linear program is of no interest, all that matters is its *feasibility*.

This latter point becomes clear upon recognizing that for any y such that $d < 0$, another y' can be constructed by multiplying y by a real constant α greater than one such that,

$$y' = \alpha y \mid \alpha > 1 \implies d' = \alpha d < d$$

Since a more negative d' can always be found, it must be that d is unbounded. This is a demonstration of the fundamental *Unboundedness Property* of [9]; if the dual is unbounded, then the primal is infeasible.

Comment (TC Fraser): Farkas's lemma here?

Definition 30. An **infeasibility certificate** is any vector y that satisfies the constraints of definition 29 and also permits violation of eq. (11) for some marginal distribution vector $\mathcal{P}_{\{V_1,\dots,V_k\}}$.

$$y \in \mathbb{R}^{\left|O^{\mathcal{M}}\right|} : y \cdot M \succeq 0, \quad y \cdot \mathcal{P}_{\{V_1,\dots,V_k\}} < 0$$

Furthermore, any y satisfying $y \cdot M \succeq 0$ induces a **certificate inequality** that constraints the space of marginal distribution vectors which takes the symbolic form of eq. (11),

$$y \cdot \mathcal{P}_{\{V_1,\dots,V_k\}} \geq 0$$

Where the entries of the certificate y act as coefficients for the entries of $\mathcal{P}_{\{V_1,\dots,V_k\}}$.

Todo (TC Fraser): Discuss Infeasibility Certificates basis

Example 31. The marginal problem for the inflated causal structure $\mathcal{G}'$ depicted in fig. 2 concerns itself with whether or not distributions over the pre-injectable sets $\{P_{\Pi_1}, \dots, P_{\Pi_{12}}\}$ admit a joint distribution $P_{\mathcal{N}'}$ over all nodes of $\mathcal{G}'$ where $\mathcal{N}' = \bigcup_{i=1}^{12} \Pi_i$. The marginalization matrix M has $16,896$ rows and $16,777,216$ columns.

$$\begin{aligned}
\#\text{ Rows} &= \sum_{i=1}^{12} |O_{\Pi_i}| = \sum_{i=1}^{12} 4^{|\Pi_i|} = 4 \cdot 4^6 + 8 \cdot 4^3 = 16,896 \\
\#\text{ Columns} &= |O_{\mathcal{N}'}| = 4^{\left|\mathcal{N}'\right|} = 4^{12} = 16,777,216
\end{aligned}$$

With $12 \cdot 4^{12} = 201,326,592$ non-zero entries.

Todo (TC Fraser): Discuss the certificate inequalities we found.

$$\mathcal{I}_{\text{Mosek-Cert}}, \mathcal{I}_{\text{Cert}}$$

VIII. LOGICAL IMPLICATIONS OF NON-CONTEXTUALITY

Section VII discusses how one can obtain a valid compatibility inequality that witnesses incompatibility for a particular marginal model $P^{\mathcal{M}}$ by writing the marginal problem as a linear program. Todo (TC Fraser): Motivate why certificates are not enough, want many solutions, want logical foundation

A. Logical Implications & Inequalities

Following the work conducted by Mansfield and Fritz [10], we consider a possibilistic implications of the form,

$$a \implies c_1 \vee \dots \vee c_m = \bigvee_{i=1}^{m} c_i \tag{12}$$

Where a and each of the c_i's are simply events or outcomes of a particular set of variables. The letter 'a' is chosen for the event a since it takes the place of the logical **antecedent** of eq. (12). Likewise, the letter 'c' is chosen to represent logical **consequents**. We refer to the set of all c_i's simply as $C = \{c_i \mid i \in 1, \dots, n\}$. The implication eq. (12) can be read as *whenever a occurs, at least one element of c also occurs.*

It is possible to turn possibilistic implications into probabilistic inequalities by recognizing that the logical implication of eq. (12) induces the inequality,

$$P(a) \leq P\left(\bigvee_{i=1}^{m} c_i\right)$$

Furthermore utilizing Boole's inequality,

$$P\left(\bigvee_{i=1}^{m} c_i\right) \leq \sum_{i=1}^{m} P(c_i) \tag{13}$$

Gives,

$$P(a) \leq \sum_{i=1}^{m} P(c_i) \tag{14}$$

Such that whenever the inequality eq. (14) is violated, the implication in eq. (12) is violated as well. Note that the converse is *not* true; if the inequality eq. (14) holds true, it is still possible for there to be a violation of eq. (12).

Remark 32. An important result of Boole's inequality is that eq. (13) becomes an exact *equality* whenever elements of C are pairwise disjoint. Therefore, finding a set C of pairwise disjoint events satisfying eq. (12) will give rise to tighter inequalities.

B. Implications in the Marginal Problem

The question then remains, *how does non-contextuality give rise to implications of the form of eq.* (12)*?* Begin by making the principle assumption that a joint distribution *does* exist for a marginal model $\{P_{V_1}, \dots, P_{V_k}\}$. This assumption induces logical *tautologies* that take the form of eq. (12), which we call **marginal implications**, using the following train of logic.

1. Suppose a particular marginal outcome $a \in O^{\mathcal{M}}$ happens to occur.

2. Since a joint distribution exists, then a refers to some *incomplete* knowledge about a *joint* event $j = o[\mathcal{V}] \in O_{\mathcal{V}}$ that actually occured. More precisely, this set of possible j's is the extendable set of a in $O_{\mathcal{V}}$.

$$j \in \mathsf{Ext}_{\mathcal{V}}(a)$$

3. Therefore whenever a occurs, one of the elements j of $\mathsf{Ext}_{\mathcal{V}}(a)$ *has* to occur.

$$a \implies \bigvee_{j \in \mathsf{Ext}_{\mathcal{V}}(a)} j \tag{15}$$

4. Now suppose we could obtain a set of marginal outcomes $C = \{c_1, \dots, c_m\}$ each different from a such that for every j, $\mathsf{Ext}(c \to j)$ for at least one element $c \in C$. If such a C can be found, then the possibility of at least one j occurring implies the possibility of at least one c occurring.

$$\bigvee_{j \in \mathsf{Ext}_{\mathcal{V}}(a)} j \implies c_1 \vee \cdots \vee c_m = \bigvee_{i=1}^{m} c_i \tag{16}$$

5. Combing eq. (15) with eq. (16), one obtains eq. (12).

Todo (TC Fraser): Talk about why this is called hardy paradox

Finding all marginal implications for a chosen marginal outcome a corresponds to finding a set of marginal outcomes $C = \{c_1, \dots, c_m\}$ whose extendable sets *cover* $\mathsf{Ext}_{\mathcal{V}}(a)$. Formally this corresponds to a **set covering** problem Todo (TC Fraser): cite which we elected to cast as the equivalent **hypergraph transversal problem**.

Todo (TC Fraser): Mention sufficient solution to the possibilistic marginal problem Todo (TC Fraser): Illustrate how it can distinguish more than possibilistic differences

Given an antecedent a, we can construct a hypergraph $\mathcal{H}_a$ called the **marginal hypergraph** with nodes $\mathcal{N}_a$ and $\mathcal{E}_a$. The nodes of this hypergraph are the subset of marginal outcomes $\mathcal{N}_a \subseteq O^{\mathcal{M}}$ *compatible* with a. The edges are labeled by outcomes in $\mathsf{Ext}_{\mathcal{V}}(a) \subseteq O_{\mathcal{V}}$, namely the set of j's, and contain all nodes compatible with j.

$$\mathcal{N}_a = \left\{n \in O^{\mathcal{M}} \mid \mathsf{Com}(n, a)\right\} \tag{17}$$

$$\mathcal{E}_a = \left\{\left\{n \in O^{\mathcal{M}} \mid \mathsf{Com}(n, j)\right\}, j \in \mathsf{Ext}_{\mathcal{V}}(a)\right\} \tag{18}$$

Remark 33. Note that every node is contained in some edge. To prove this, consider the definition of compatibility between outcomes. If every n is compatible with a, then there exists some outcome $j \in \mathsf{Ext}_{\mathcal{V}}(A)$ such that $\mathsf{Ext}(n \to j)$. Therefore since $\mathsf{Ext}(n \to j)$, $\mathsf{Com}(n, j)$; satisfying the central condition of eq. (18).

C. Higher Order Marginal Implications

Todo (TC Fraser): Discuss $(m, n) - type$ implications and the non-triviality Todo (TC Fraser): Link to logical bell inequalities/completeness or not?

D. Hypergraph Transversals

Section VIII B demonstrates how finding possibilistic marginal implications can be casted as a hypergraph transversal problem. This section aims to summarize the general idea behind existing algorithms and also discusses some of the inequalities found using these algorithms. Todo (TC Fraser): References here

Definition 34. A **hypergraph transversal generation** is any algorithm that correctly generates the complete set of all minimal transversals of $\mathcal{H}$.

Definition 35. Any hypergraph with strictly non-empty edges $\forall e \neq \emptyset$ will always admit the **trivial transversal** $\mathcal{T}^*$ where all nodes are considered as members $\mathcal{T}^* = \mathcal{N}$. Any hypergraph with empty edges will be called a **degenerate hypergraph** as it admits no transversals[5].

Remark 33 guarantees that all marginal hypergraphs are non-degenerate. There are two distinct approaches to hypergraph transversal generation: *top-down* and *bottom-up*.

Definition 36. A **top-down** hypergraph transversal generation refers to any algorithm that begins with the trivial transversal $\mathcal{T}^*$ and iteratively removes unnecessary nodes from $\mathcal{T}^*$.

Definition 37. A **bottom-up** hypergraph transversal generation refers to any algorithm that begins with the empty set $\emptyset$ and iteratively adds nodes.

One should select a top-down method if the typical size of minimal transversals $|\mathcal{T}|$ is comparable to the number of nodes $|\mathcal{N}|$, otherwise a bottom-up method will perform better. For our purposes we implemented a deep-first transversal algorithm similar to Recalling remark 33

E. Weighted Hypergraph Transversals

In section VII is was demonstrated that the Fritz distribution is witness-able via a certificate inequality. It is also possible to witness Todo (TC Fraser): Discuss the Inequalities Derived/ Trivial and non-trivial

Todo (TC Fraser): Weighted transversals and Optimizations Todo (TC Fraser): Seeding inequalities (huge advantage here)

[5] Most authors require that all hypergraphs be non-degenerate [11].

IX. DERIVING SYMMETRIC INEQUALITIES

Symmetric compatibility inequalities are useful for a number of reasons. First, Bancal et. al. [12] discuss computational advantages in considering symmetric versions of marginal polytopes mentioned in section IV; the number of extremal points typically grows exponentially in $\mathcal{V}$, but only polynomial for the symmetric polytope. They also note a number of interesting inequalities (such as CHSH [6]) can be written in a way that is symmetric under the exchange of parties, demonstrating that quantum-non-trivial inequalities can be recovered from facets of a symmetric polytope. Second, numerical optimizations against inequalities invariant under exchange of parties will lead to one of two interesting cases: either the extremal distribution is invariant under exchange of parties or it is not. The latter case generates a family of incompatible distributions obtained by exchanging parties of the found extremal distribution[6]. In this section we discuss how to achieve computational advantage and its application to the inequality techniques mentioned in section VII and section VIII A.

A. Causal Symmetry

The aim of this section is to formalize what types of symmetries are present in a particular causal structure $\mathcal{G}$. The group of these causal symmetries will define, for each compatibility inequality $\mathcal{I}_\mathcal{M}$, a family of inequalities that are also valid incompatibility inequalities.

First consider the permutation group acting on a set of nodes $N \subseteq \mathcal{N}$ denoted $\Phi(N)$ to be the set of all bijective maps φ from N to N.

$$\Phi(N) = \{\varphi \mid \varphi : N \to N\}$$

The action of $\varphi \in \Phi(N)$ on a causal structure $\mathcal{G}$ is defined via an the extension of φ to the corresponding element in $\Phi(\mathcal{N})$ that leaves nodes $n \in \mathcal{N} \setminus N$ invariant,

$$\begin{aligned}\varphi(\mathcal{G}) &\equiv (\varphi(\mathcal{N}), \varphi(\mathcal{E}))\\ \varphi(\mathcal{N}) &\equiv \{\varphi(n) \mid n \in \mathcal{N}\} = \mathcal{N}\\ \varphi(\mathcal{E}) &\equiv \{\varphi(e) \mid e \in \mathcal{E}\}\end{aligned}$$

To motivate a more general treatment of such symmetries, consider the Triangle Scenario of fig. 1. Due to the rotational and reflective symmetries of the Triangle Scenario, any permutation $\varphi \in \Phi(\mathcal{N}_O)$ of the observable nodes $\mathcal{N}_O = \{A, B, C\}$ creates a new causal structure $\varphi(\mathcal{G})$ that is *equivalent* to $\mathcal{G}$ up to a relabeling of *latent* nodes. By construction, any valid compatibility inequality $\mathcal{I}_\mathcal{M}$ for the marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$ is *independent* of the labeling of latent nodes. Therefore applying φ to all distributions in $\mathcal{I}_\mathcal{M}$ yields $\varphi(\mathcal{I}_\mathcal{M}) \equiv \varphi(\mathcal{I})_{\varphi(\mathcal{M})}$, another valid compatibility inequality for the permuted marginal scenario $\varphi(\mathcal{M})$ defined as,

$$\varphi(\mathcal{M}) = \{\varphi(V_1), \ldots, \varphi(V_k)\}$$

More generically, any permutation φ that preserves the graphical structure of $\mathcal{G}$ can be applied to $\mathcal{I}_\mathcal{M}$ to generate new and valid compatibility inequalities. Specifically this alludes to the **graph automorphism group** $\mathsf{Aut}(\mathcal{G})$ of $\mathcal{G}$.

$$\mathsf{Aut}(\mathcal{G}) \equiv \{\varphi \in \Phi(\mathcal{N}) \mid \varphi(\mathcal{G}) = \mathcal{G}\}$$

In general, $\mathsf{Aut}(\mathcal{G})$ could include elements φ that map latent nodes to observable nodes[7]. This behaviour is *undesired* for causal inference where the latent nodes will never appear in a marginal scenario. Instead, the subgroup of $\mathsf{Aut}(\mathcal{G})$ that never maps observable nodes to latent nodes is the **set-wise stabilizer** of $\mathsf{Aut}(\mathcal{G})$ for $\mathcal{N}_O$ and will be expressed with a subscript.

Definition 38. The **causal symmetry group** for a causal structure $\mathcal{G}$ is the graph automorphism subgroup that stabilizes the observable nodes.

$$\mathsf{Aut}_{\mathcal{N}_O}(\mathcal{G}) \equiv \{\varphi \in \Phi(\mathcal{N}) \mid \varphi(\mathcal{G}) = \mathcal{G}, \forall n \in \mathcal{N}_O : \varphi(n) \in \mathcal{N}_O\}$$

Given any compatibility inequality $\mathcal{I}_\mathcal{M}$, each element φ of the causal symmetry group defines a new valid compatibility inequality $\varphi(\mathcal{I}_\mathcal{M}) \equiv \varphi(\mathcal{I})_{\varphi(\mathcal{M})}$.

---

[6] If the extremal distribution is not invariant under exchange of parties, there is indication that the space of accessible distributions is non-convex.

[7] In practice however, it is rare to be considering a causal structure $\mathcal{G}$ where some latent node $n_L \in \mathcal{N}_L$ takes on an role indistinguishable from some observable node $n_O \in \mathcal{N}_O$, so this won't be an issue.

The inflation technique discussed in section V allows one to derive compatibility inequalities for a causal structure $\mathcal{G}$ by considering the marginal problem over the pre-injectable sets of $\mathcal{G}'$ denoted $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}') = \{V_1, \ldots, V_k\}$. It is important to recall that due to lemma 23, inequalities $\mathcal{I}'$ for $\mathcal{G}'$ are *only* transferable to inequalities $\mathcal{I}$ for $\mathcal{G}$ if $\mathcal{I}'$ is in terms of distributions over the pre-injectable sets $\mathcal{I}'_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')}$. As a consequence of this observation, we need to consider a subgroup of the causal symmetry group that preserves the pre-injectable sets,

$$\varphi\left(\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')\right) = \mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}') \tag{19}$$

Definition 39. The **restricted causal symmetry group** for a causal structure $\mathcal{G}$ and a marginal scenario $\mathcal{M}$ is the graph automorphism subgroup that simultaneously stabilizes each marginal context V_i of $\mathcal{M}$.

$$\mathsf{Aut}_{\mathcal{M}}(\mathcal{G}) \equiv \{\varphi \in \Phi(\mathcal{N}) \mid \varphi(\mathcal{G}) = \mathcal{G}, \forall V \in \mathcal{M} : \forall n \in V : \varphi(n) \in V\}$$

Given any compatibility inequality $\mathcal{I}_{\mathcal{M}}$ each element φ of the causal symmetry group defines a new valid compatibility inequality $\varphi(\mathcal{I}_{\mathcal{M}}) \equiv \varphi(\mathcal{I})_{\mathcal{M}}$ over the *same* marginal context $\mathcal{M}$.

The restricted causal symmetry group is essential for determining finding symmetry inequalities for the inflated causal structure $\mathcal{G}'$ that can be deflated to symmetry inequalities on $\mathcal{G}$.

B. Symmetric Marginal Polytope

C. Symmetric Logical Implications

D. Results

Todo (TC Fraser): Identify the desired symmetry group Todo (TC Fraser): How we obtained the desired symmetry group Todo (TC Fraser): Group orbits to symmetric marginal description matrix Todo (TC Fraser): Infeasibility on symmetric marginal problem Todo (TC Fraser): Hardy Transversals can't work on the symmetric marginal problem Todo (TC Fraser): Symmetrizing non-symmetric inequalities through avoiding orbits Todo (TC Fraser): higher order transversals on mutually impossible events

X. NON-LINEAR OPTIMIZATIONS

Compatibility inequalities for a given causal structure are fantastic for finding incompatible distributions. In the inflation technique, this is no exception. Parameterizing a space distributions using a set of real-valued parameters λ, enables us to perform numerical optimizations against these inequalities in hopes that a particular set of parameters λ is able to generate an incompatible distribution P. To illustrate this generic procedure and it's reliability, we will first examine the popular CHSH inequality.

A. Numerical Violations of The CHSH Inequality

The CHSH inequality [6] can we viewed as a causal compatibility inequality for the iconic Bell Scenario (Fig. 19 of [13], Fig. 11 of [2], Fig. 1 a) of [14], etc.) corresponding to Bell's notion of local causality [13]. It constrains the set of 2-outcome bipartite distributions over local binary measurement settings for each party $P_{AB|S_AS_B} \equiv \left\{P_{AB|00}, P_{AB|01}, P_{AB|10}, P_{AB|11}\right\}$. Numerical optimization *should* obtain the algebraic violation associated with the PR-Box correlations [15]. Maintaining full generality, we simply need to parameterize these 4 distributions using eq. (E2), each requiring 4 real-valued parameters. We define the optimization target for the CHSH inequality to be the left-hand-side of eq. (8),

$$\mathcal{I}_{\text{CHSH}} = \langle AB|11\rangle + \langle AB|12\rangle + \langle AB|21\rangle - \langle AB|22\rangle$$

Figure 3 demonstrates this optimization for 5 random seed parameters λ^0, each converging to the expected value of 4. Analogously, [7]

Todo (TC Fraser): Demonstrate Quantum, Convexity Todo (TC Fraser): Why Inequalities are great for optimizations Todo (TC Fraser): Non-linearity Todo (TC Fraser): Techniques Used Todo (TC Fraser): Finding maximum

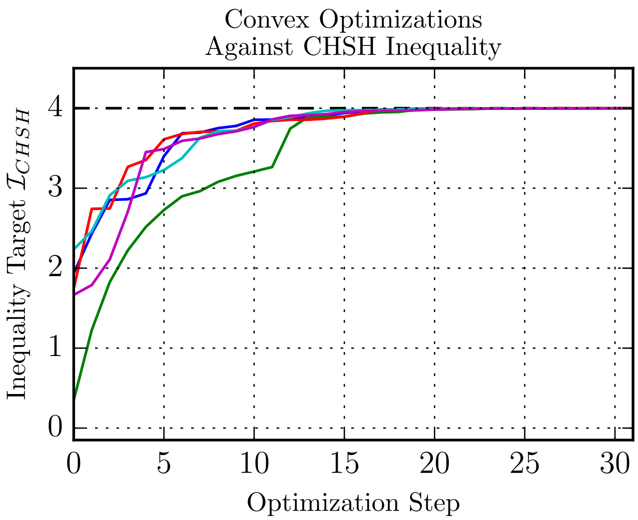

FIG. 3. Convex optimizations against $\mathcal{I}_{\text{CHSH}}$ recover algebraic violation of 4.

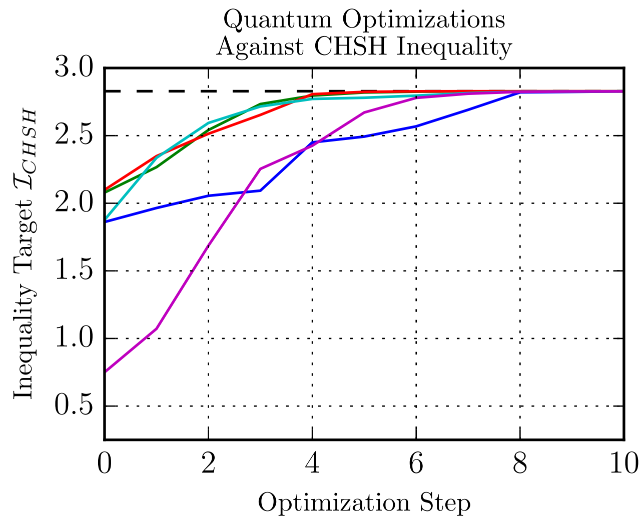

FIG. 4. Quantum optimizations against $\mathcal{I}_{\text{CHSH}}$ recover maximum violation of $2\sqrt{2}$.

violation of CHSH easily Todo (TC Fraser): Unreliance when number of parameters increases Todo (TC Fraser): Issues with local minimum Todo (TC Fraser): Using initial conditions close to fritz, obtain greater violation Todo (TC Fraser): Greater violation shares possibilistic structure of fritz and violates CHSH under definition Todo (TC Fraser): Not realizable with maximally entangled qubit states Todo (TC Fraser): Not realizable with separable measurements Todo (TC Fraser): Many non-trivial inequalities to be tested Todo (TC Fraser): inequality -> dist -> inequality evolution

XI. CONCLUSIONS

Todo (TC Fraser): Inflation technique allows one to witness fritz incompatibility Todo (TC Fraser): Linear optimization induces certificates which are incompatibility witnesses Todo (TC Fraser): There are quantum distributions in the Triangle Scenario that are incompatible and different from fritz in terms of entanglement but not possibilistic structure

XII. OPEN QUESTIONS & FUTURE WORK

Todo (TC Fraser): Lots of stuff

Appendix A: Exemplary Inequalities

Appendix B: Connections to Sheaf-Theoretic Treatment

Appendix C: Computationally Efficient Parametrization of the Unitary Group

Spengler, Huber and Hiesmayr [16] suggest the parameterization of the unitary group $\mathcal{U}(d)$ using a $d \times d$-matrix of real-valued parameters $\lambda_{n,m}$,

$$U = \left[\prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d} \exp\left(iP_n\lambda_{n,m}\right)\exp\left(i\sigma_{m,n}\lambda_{m,n}\right)\right)\right] \cdot \left[\prod_{l=1}^{d}\exp\left(iP_l\lambda_{l,l}\right)\right] \tag{C1}$$

Where P_l are one-dimensional projective operators,

$$P_l = |l\rangle\langle l| \tag{C2}$$

and the $\sigma_{m,n}$ are generalized anti-symmetric σ-matrices,

$$\sigma_{m,n} = -i\,|m\rangle\langle n| + i\,|n\rangle\langle m|$$

Where $1 \le m < n \le d$. Spengler et. al. proved the validity of eq. (C1) in Ref. [16].

For the sake of reference, let us label the matrix exponential terms in eq. (C1) in a manner that corresponds to their affect on an orthonormal basis $\{|1\rangle, \dots, |d\rangle\}$.

$$\begin{aligned} GP_l &= \exp(iP_l\lambda_{l,l}) \\ RP_{n,m} &= \exp(iP_n\lambda_{n,m}) \\ R_{m,n} &= \exp(i\sigma_{m,n}\lambda_{m,n}) \end{aligned} \tag{C3}$$

It is possible to remove the reliance on matrix exponential operations in eq. (C1) by utilizing the explicit form of the exponential terms in eq. (C3). As a first step, recognize the defining property of the projective operators eq. (C2),

$$P_l^k = (|l\rangle\langle l|)^k = |l\rangle\langle l| = P_l$$

This greatly simplifies the global phase terms GP_l,

$$GP_l = \exp(iP_l\lambda_{l,l}) = \sum_{k=0}^{\infty}\frac{(iP_l\lambda_{l,l})^k}{k!} = \mathbb{I} + \sum_{k=1}^{\infty}\frac{(i\lambda_{l,l})^k}{k!}P_l^k = \mathbb{I} + P_l\left[\sum_{k=1}^{\infty}\frac{(i\lambda_{l,l})^k}{k!}\right] = \mathbb{I} + P_l\left(e^{i\lambda_{l,l}} - 1\right) \tag{C4}$$

Analogously for the relative phase terms $RP_{n,m}$,

$$RP_{n,m} = \cdots = \mathbb{I} + P_n\left(e^{i\lambda_{n,m}} - 1\right) \tag{C5}$$

Finally, the rotation terms $R_{m,n}$ can also be simplified by examining powers of $i\sigma_{n,m}$,

$$R_{m,n} = \exp(i\sigma_{m,n}\lambda_{m,n}) = \sum_{k=0}^{\infty}\frac{(|m\rangle\langle n| - |n\rangle\langle m|)^k\,\lambda_{m,n}^k}{k!}$$

One can verify that the following properties hold,

$$\begin{aligned} (|m\rangle\langle n| - |n\rangle\langle m|)^0 &= \mathbb{I} \\ \forall k \in \mathbb{N}, k \neq 0 : (|m\rangle\langle n| - |n\rangle\langle m|)^{2k} &= (-1)^k(|m\rangle\langle m| + |n\rangle\langle n|) \\ \forall k \in \mathbb{N} : (|m\rangle\langle n| - |n\rangle\langle m|)^{2k+1} &= (-1)^k(|m\rangle\langle n| - |n\rangle\langle m|) \end{aligned}$$

Revealing the simplified form of $R_{m,n}$,

$$R_{m,n} = \mathbb{I} + (|m\rangle\langle m| + |n\rangle\langle n|)\sum_{j=1}^{\infty}(-1)^j\frac{\lambda_{n,m}^{2j}}{(2j)!} + (|m\rangle\langle n| - |n\rangle\langle m|)\sum_{j=0}^{\infty}(-1)^j\frac{\lambda_{n,m}^{2j+1}}{(2j+1)!}$$

$$R_{m,n} = \mathbb{I} + (|m\rangle\langle m| + |n\rangle\langle n|)(\cos\lambda_{n,m} - 1) + (|m\rangle\langle n| - |n\rangle\langle m|)\sin\lambda_{n,m} \tag{C6}$$

By combining the optimizations of eqs. (C5) to (C4) together we arrive at an equivalent form for eq. (C1) that is computational more efficient.

$$U = \left[\prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d} RP_{n,m}R_{m,n}\right)\right]\cdot\left[\prod_{l=1}^{d} GP_l\right] \tag{C7}$$

In quantum mechanics, the global phase of a state $|\psi\rangle \in \mathcal{H}^n$ is a *redundant* parameter. Parameterizing unitaries using eq. (C7) is especially attractive since the global phase terms GP_l can be dropped, allowing one to parameterize all unitaries in $\mathcal{U}(d)$ up to this degeneracy [16][8].

$$U_{/GP_l} = \left[\prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d} RP_{n,m}R_{m,n}\right)\right] \tag{C8}$$

Todo (TC Fraser): Explanation of Computational Complexity $\mathcal{O}\left(d^3\right)$ vs. $\mathcal{O}(1)$ using [17] Todo (TC Fraser): Pre-Caching for Fixed dimension d Todo (TC Fraser): Talk about inverse via haar measure

Appendix D: Parametrization of Quantum States & Measurements

Throughout section X, we utilize a variety of parameterizations of quantum states and measurements in order to generate quantum-accessible probability distributions. There are numerous techniques that can used when parameterizing quantum states and measurements [16, 18–21] with applications Todo (TC Fraser): Finish this sentence. For our purposes, we need to parameterize the space of quantum-accessible distributions $P_{\mathcal{Q}}$ that are *realized* on the Triangle Scenario. We have implemented $P_{\mathcal{Q}}$ under the following description.

$$P_{ABC}(abc) = \mathrm{Tr}\left[\Omega^{\intercal}\rho_{AB}\otimes\rho_{BC}\otimes\rho_{CA}\Omega M_{A,a}\otimes M_{B,b}\otimes M_{C,c}\right] \tag{D1}$$

1. Quantum States

The bipartite states $(\rho_{AB}, \rho_{BC}, \rho_{CA})$ of eq. (D1) were taken to be two-qubit density matrices acting on $\mathcal{H}^2\otimes\mathcal{H}^2$.[9] The space of all such states corresponds to the space of all 4×4 positive semi-definite hermitian matrices with unitary trace. Throughout this section, we refer to these bipartite states simply as ρ unless otherwise indicated. There are three distinct techniques that we have considered.

Taking inspiration from [21], we can parameterize all such density matrices ρ using **Cholesky Parametrization** [22]. The Cholesky decomposition allows one to write any hermitian positive semi-definite matrix ρ in terms of a lower (or upper) triangular matrix T using $\rho = T^{\dagger}T$. Our Cholesky parameterization consists of assigning 16 real-valued parameters λ to the entires of T and generating a unitary trace ρ similar to eq. (4.4) of [21].

$$\rho = \frac{T^{\dagger}T}{\mathrm{Tr}\left(T^{\dagger}T\right)} \quad T = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ \lambda_2 + i\lambda_3 & \lambda_4 & 0 & 0 \\ \lambda_5 + i\lambda_6 & \lambda_7 + i\lambda_8 & \lambda_9 & 0 \\ \lambda_{10} + i\lambda_{11} & \lambda_{12} + i\lambda_{13} & \lambda_{14} + i\lambda_{15} & \lambda_{16} \end{bmatrix} \tag{D2}$$

Our deviation from exclusiving using eq. (D2) is two-fold. First, eq. (D2) is degenerate in that the normalization indicates only $16 - 1 = 15$ parameters are required for fully generic parameterization of all such states ρ. Removing this degeneracy is possible although difficult. Second, the parameters λ_i carry no physical meaning associated with the state ρ, unlike our next parameterization.

In Spengler, Huber and Hiesmayr's work [16], they discuss how to parameterization density matrices ρ acting on $\mathcal{H}^d$ of rank k through it's spectral decomposition,

$$\rho = \sum_{i=1}^{k} p_i |\psi_i\rangle\langle\psi_i| \quad p_i \geq 0, \sum_i p_i = 1, k \leq d \tag{D3}$$

Where any orthonormal basis $\{|\psi_i\rangle\}$ of $\mathcal{H}^d$ can be transformed into a computational basis $\{|i\rangle\}$ by a unitary $U \in \mathcal{U}(d)$ such that $|\psi_i\rangle = U|i\rangle$. We refer to eq. (D3) as the **Spengler Parametrization**. Without loss of generality we parameterize all full-rank $(k = d)$ matrices by simultaneously parameterizing the $d = 4$ eigenvalues p_i of eq. (D3) using eq. (E1) and the unitary group $\mathcal{U}(4)$ up to global phase equivalence using eq. (C8). Parameterizing ρ using the Spengler parameterization requires $3 + 12 = 15$ parameters; admitting no degeneracies.

[8] In our implementation, we accomplish this by explicitly setting $\lambda_{l,l} = 0$ in eq. (C4)

[9] We also considered qutrit $\mathcal{H}^3$ qutit $\mathcal{H}^4$ states. However for 6 d-dimensional $\mathcal{H}^d$ states, the joint density matrix ρ acts on $\left(\mathcal{H}^d\right)^{\otimes 6}$ making it a $\left(d^6, d^6\right)$ matrix with d^{12} entries. Computationally only $d = 2$ was feasible for our optimization tasks.

Finally in cases where we wish to restrict ourselves to *pure* bipartite states $\rho = |\psi\rangle\langle\psi|$, we have the luxury to use a **Schmidt Parametrization**. This is accomplished via a Schmidt decomposition $|\psi_{AB}\rangle = \sum_i \sigma_i |i_A\rangle \otimes |i_B\rangle$ where normalization demands that $\sum_i \sigma_i^2 = 1$, $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ are orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively [23]. Additionally for qubit sources we can *choose* our orthonormal bases to be the computational basis $\{|0\rangle, |1\rangle\}$ and write,

$$|\psi\rangle = \cos^2(\lambda_1)|0\rangle \otimes |0\rangle + \sin^2(\lambda_1)|1\rangle \otimes |1\rangle$$

Where only 1 real-valued parameter $\lambda = \{\lambda_1\}$ is required to parameterize all pure states up to local unitaries. Pure states are also attractive due to their computational advantage in computing eq. (D1). If each state ρ is decomposable into $|\psi\rangle\langle\psi|$, then eq. (D1) can be written as,

$$P_{ABC}(abc) = \langle \psi_{AB}\psi_{BC}\psi_{CA}|\Omega M_{A,a} \otimes M_{B,b} \otimes M_{C,c}\Omega^{\intercal}|\psi_{AB}\psi_{BC}\psi_{CA}\rangle \tag{D4}$$

Avoiding the expensive matrix multiplications of eq. (D1).

2. Measurements

With full generality, we consider a measurement M to be a **projective-operator valued measure (POVM)** represented by a set of hermitian, positive semi-definite operators $\{M_i\}_{i=1,\dots,k}$[10] acting on $\mathcal{H}^d$ summing to the identity,

$$\forall |\phi\rangle \in \mathcal{H}^d : \langle\phi| M_i |\phi\rangle \geq 0 \quad \sum_{i=1}^{k} M_i = \mathbb{I}_{\mathcal{H}^d} \tag{D5}$$

When considering $k = 2$ outcome measurements acting on $\mathcal{H}^4$ we parameterize the first POVM element M_1 by using a Cholesky parameterization similar to eq. (D2) without normalizing for trace. Afterwards, M_2 is fully determined by eq. (D5).

$$M_1 = T^{\dagger}T \quad M_2 = \mathbb{I} - T^{\dagger}T \tag{D6}$$

However, in order for M_2 to be positive semi-definite, the largest eigenvalue of M_1 has to be less than 1. To see this is a necessary and sufficient constraint, first expand out eq. (D6),

$$\langle\phi|M_2|\phi\rangle = \langle\phi|\mathbb{I} - M_1|\phi\rangle = |\phi|^2 - \langle\phi| \left(\sum_{i=1}^{d} m_1^{(i)} \left|m_1^{(i)}\right\rangle \left\langle m_1^{(i)}\right|\right) |\phi\rangle$$

Next write a generic $|\phi\rangle \in \mathcal{H}^d$ in terms of a linear combination of the eigenvectors of M_1[11].

$$\langle\phi|M_2|\phi\rangle = \sum_j \left|\left\langle \phi\middle|m_1^{(j)}\right\rangle\right|^2 - \sum_i m_1^{(i)} \left|\left\langle \phi\middle|m_1^{(i)}\right\rangle\right|^2 = \sum_i \left(1 - m_1^{(i)}\right) \left|\left\langle \phi\middle|m_1^{(i)}\right\rangle\right|^2 \tag{D7}$$

Since $|\phi\rangle$ is arbitrary, for each i set $|\phi\rangle = \left|m_1^{(i)}\right\rangle$ to see that each eigenvalue of M_1 needs to be less than 1.

$$\left\langle m_1^{(i)}\middle|M_2\middle|m_1^{(i)}\right\rangle = \left(1 - m_1^{(i)}\right) \geq 0 \implies m_1^{(i)} \leq 1 \tag{D8}$$

By eq. (D7) is not difficult to see that eq. (D8) is a sufficient condition. During optimization, eq. (D8) can either be enforced passively as a constraint or directly by normalizing M_1 by its largest eigenvalue $\max\left(m_1^{(i)}\right)$ whenever necessary.

When generalizing the above parameterization to more than 2 outcomes, only necessary conditions were found. Generating k-outcome POVM measurements is doable using rejection sampling techniques such as those used in [24] however a valid parameterization with little to no degeneracy was not found. Upon making this observation, a

---

[10] The i-th element of M is referenced using a subscript M_i. The set of measurement elements for a particular party X will be written M_X. When both party and element are to be referenced, we write $M_{X,i}$.

[11] The eigenvectors of M_1 form an orthonormal basis for $\mathcal{H}^d$ because M_1 is Hermitian

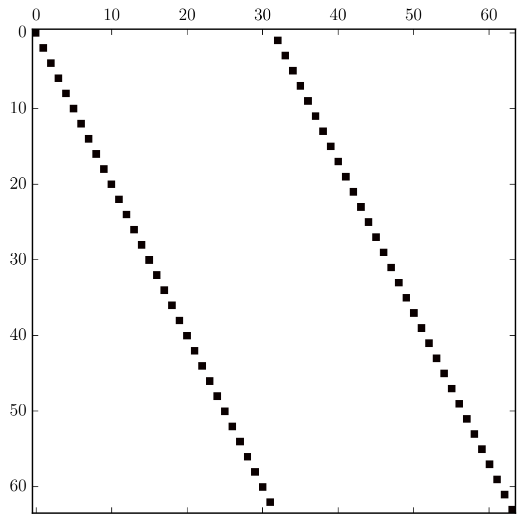

FIG. 5. The network permutation matrix Ω for $\left(\mathcal{H}^2\right)^{\otimes 6}$ realized on the triangle scenario. Black represents a value of 1 and 0 otherwise.

necessary departure to **projective-valued measures (PVMs)** is warranted[12]. With loss of generality, consider the set of PVMs M satisfying eq. (D5) in addition to orthogonal and projective properties,

$$M_i M_j = \delta_{ij} M_i \quad M_i = |m_i\rangle \langle m_i| \tag{D9}$$

Parameterizing M for k-outcome measurements corresponds to parameterizing the set of all k-th order orthonormal sub-bases of $\mathcal{H}^d$. First note that any such basis $\{|\psi_1\rangle, \ldots, |\psi_k\rangle\}$ can be transformed into the computational basis $\{|1\rangle, \ldots, |k\rangle\}$ by a unitary denoted $U \in \mathcal{U}(d)$,

$$U|\psi_i\rangle = |i\rangle$$

With this observation we just need to parameterize the set of all unitaries $\mathcal{U}(d)$,

$$M = \left\{U|i\rangle \langle i| U^\dagger\right\}_{i \in 1,\ldots,k}$$

Specifically, the projective property each M_i means that the global phase of U is completely arbitrary; one only needs to consider parameterizing unitaries up to global phase eq. (C8). This method was inspired by the *measurement seeding* method of Pál and Vértesi's [26] iterative optimization technique.

Analogously to eq. (D4), projective measurements offer considerable computational advantage as eq. (D1) can be rewritten as,

$$P_{ABC}(abc) = \langle m_{A,a} m_{B,b} m_{C,c} | \Omega^\intercal \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Omega | m_{A,a} m_{B,b} m_{C,c}\rangle \tag{D10}$$

3. Network Permutation Matrix

Finally, we introduce the **network permutation matrix** Ω for the Triangle Scenario of fig. 1. For bipartite qubit states, Ω becomes a 64×64 bit-wise matrix that acts on the measurements M and is depicted in fig. 5. To illuminate its necessity, consider eq. (D1) without Ω.

$$\begin{aligned}
P_{ABC}(abc) &\stackrel{?}{=} \operatorname{Tr}\left[(\rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA})\left(M_A^a \otimes M_B^b \otimes M_C^c\right)\right] \\
&= \operatorname{Tr}\left[(\rho_{AB} M_A^a) \otimes \left(\rho_{BC} M_B^b\right) \otimes (\rho_{CA} M_C^c)\right] \\
&= \operatorname{Tr}(\rho_{AB} M_A^a) \operatorname{Tr}\left(\rho_{BC} M_B^b\right) \operatorname{Tr}(\rho_{CA} M_C^c) \\
&= P_{A|\rho_{AB}}(a)\, P_{B|\rho_{BC}}(b)\, P_{C|\rho_{CA}}(c)
\end{aligned}$$

---

[12] Strictly speaking, when the number of outcomes (k) *matches* the Hilbert space dimension (d), eq. (D5) implies eq. (D9) by completeness. When considering 4 outcome measurements on bipartite qubit states in $\mathcal{H}^4$, PVMs are completely general. Moreover, Naimark's dilation theorem guarantees that PVMs acting on $\mathcal{H}^q$ can *emulate* the behaviour of any POVM acting on $\mathcal{H}^d$ provided that q is sufficiently larger than d [25].

On an operational level, this corresponds to A making a measurement on *both* subsystems of ρ_{AB} and *not* on any component of ρ_{CA}. This is analogously troubling for B and C as well. The network permutation matrix Ω corresponds to *aligning* the underlying 6-qubit joint state ρ with the joint measurement M. To understand its effect, consider its effect on 6-qubit pure state $|q_1\rangle \otimes \cdots \otimes |q_6\rangle = |q_1q_2q_3q_4q_5q_6\rangle$ where $\forall i : |q_i\rangle \in \mathcal{H}^2$.

$$\Omega\, |q_1q_2q_3q_4q_5q_6\rangle = |q_2q_3q_4q_5q_6q_1\rangle$$

Ω acts as a *partial transpose* on $\left(\mathcal{H}^2\right)^{\otimes 6}$ by shifting the underlying tensor structure one subsystem to the "left". It is uniquely defined by its action on all 2^6 orthonormal basis elements of $\left(\mathcal{H}^2\right)^{\otimes 6}$,

$$\Omega \equiv \sum_{|q_i\rangle\in\{|0\rangle,|1\rangle\}} |q_2q_3q_4q_5q_6q_1\rangle\langle q_1q_2q_3q_4q_5q_6|$$

4. Degeneracy

Todo (TC Fraser): Discuss local unitary degeneracy

Appendix E: Convex Parametrization of Finite Probability Distributions

As discussed in section X, there is a need to parameterize the family of all probability distributions P_V over a given set of variables $V = \left(v_1, \ldots, v_{|V|}\right)$. If the cardinality of O_V is finite, then this computationally feasible. The space of probability distributions over $n = |O_V|$ distinct outcomes forms a $n-1$ dimensional convex polytope naturally embedded in $\mathbb{R}^n_{\geq 0}$ [27] that is parameterizable by $n-1$ real value parameters; normalization $\sum_{o[V]\in O_V} P_V(o[V]) = 1$ accounts for the '-1'. An example of a non-degenerate parameterization of P_V consists of $n-1$ parameters $\lambda = (\lambda_1, \ldots, \lambda_{n-1})\,, \lambda_i \in [0, \pi/2]$ which generate the probabilities n probability values p_j using hyperspherical coordinates [16, 19],

$$\begin{aligned} p_j &= \cos^2\lambda_j \prod_{i=1}^{j-1} \sin^2\lambda_i \quad \forall j \in 1, \ldots, n-1 \\ p_n &= \prod_{i=1}^{n-1} \sin^2\lambda_i \end{aligned} \tag{E1}$$

Furthermore due to the periodicity of the parameter space λ, eq. (E1) can be used for either constrained or unconstrained optimization problems. For continuity reasons, unconstrained optimizations are performed whenever possible.

Although non-degenerate, this parameterization suffers from uniformity; a randomly sampled vector of parameters λ *does not* translate to a randomly sampled probability P_V. An easy-to-implement, degenerate parameterization of P_V can be constructed by simply beginning with n real parameters $\lambda = (\lambda_1, \ldots, \lambda_n)$, then making them positive and normalized by their sum[13].

$$p_j = \frac{|\lambda_j|}{\sum_{i=1}^{n} |\lambda_i|} \quad \forall j \in 1, \ldots, n \tag{E2}$$

For various convex optimization tasks sensitive to initial conditions outlined section X, the latter parameterization of eq. (E2) generally performed better than the former eq. (E1).

[1] Tobias Fritz, "Beyond bell's theorem ii: Scenarios with arbitrary causal structure," (2014), 10.1007/s00220-015-2495-5, arXiv:1404.4812.

[13] Strictly speaking, eq. (E2) *also* suffers from non-uniformity; being biased toward uniform probability distributions P_V. Todo (TC Fraser): Discuss rejection sampling simplex algorithms

[2] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz, "The inflation technique for causal inference with latent variables," (2016).
[3] Tobias Fritz and Rafael Chaves, "Entropic inequalities and marginal problems," (2011), 10.1109/TIT.2012.2222863, arXiv:1112.4788.
[4] Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009).
[5] Tobias Fritz, "Beyond bell's theorem: Correlation scenarios," (2012), 10.1088/1367-2630/14/10/103001, arXiv:1206.5115.
[6] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, "Proposed experiment to test local hidden-variable theories," Phys. Rev. Lett. **23**, 880–884 (1969).
[7] B. S. Cirel'son, "Quantum generalizations of bell's inequality," Lett Math Phys **4**, 93–100 (1980).
[8] Sébastien Lahaie, "How to take the dual of a linear program," (2008).
[9] Bradley, Hax, and Magnanti, *Applied Mathematical Programming* (Addison-Wesley, 1977) Chap. 4, pp. 143–144.
[10] Shane Mansfield and Tobias Fritz, "Hardy's non-locality paradox and possibilistic conditions for non-locality," (2011), 10.1007/s10701-012-9640-1, arXiv:1105.1819.
[11] Dimitris J. Kavvadias and Elias C. Stavropoulos, "An efficient algorithm for the transversal hypergraph generation," JGAA **9**, 239–264 (2005).
[12] Jean-Daniel Bancal, Nicolas Gisin, and Stefano Pironio, "Looking for symmetric bell inequalities," Journal of Physics A: Mathematical and Theoretical **43**, 385303 (2010).
[13] Christopher J. Wood and Robert W. Spekkens, "The lesson of causal discovery algorithms for quantum correlations: Causal explanations of bell-inequality violations require fine-tuning," (2012), 10.1088/1367-2630/17/3/033002, arXiv:1208.4119.
[14] Armin Tavakoli, "Quantum correlations in connected multipartite bell experiments," (2015), 10.1088/1751-8113/49/14/145304, arXiv:1509.08491.
[15] D. Rohrlich and S. Popescu, "Nonlocality as an axiom for quantum theory," (1995), arXiv:quant-ph/9508009.
[16] Christoph Spengler, Marcus Huber, and Beatrix C. Hiesmayr, "A composite parameterization of unitary groups, density matrices and subspaces," (2010), 10.1088/1751-8113/43/38/385306, arXiv:1004.5252.
[17] Cleve Moler and Charles Van Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," SIAM Rev. **45**, 3–49 (2003).
[18] Dénes Petz and László Ruppert, "Optimal quantum-state tomography with known parameters," (2015), 10.1088/1751-8113/45/8/085306, arXiv:1511.06666.
[19] S. R. Hedemann, "Hyperspherical parameterization of unitary matrices," (2013), arXiv:1303.5904.
[20] Kazuyuki Fujii, Kunio Funahashi, and Takayuki Kobayashi, "Jarlskog's parametrization of unitary matrices and qudit theory," (2005), arXiv:quant-ph/0508006.
[21] Daniel F. V. James, Paul G. Kwiat, William J. Munro, and Andrew G. White, "Measurement of qubits," Phys. Rev. A **64** (2001), 10.1103/physreva.64.052312.
[22] Markus Grasmair, "On the existence of a cholesky factorization," (2014).
[23] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2011) Chap. The Schmidt decomposition and purifications, p. 109.
[24] Dénes Petz and László Ruppert, "Optimal quantum-state tomography with known parameters," Journal of Physics A: Mathematical and Theoretical **45**, 085306 (2012).
[25] Asher Peres, *Quantum Theory: Concepts and Methods (Fundamental Theories of Physics)* (Springer, 1995) Chap. 9.6, p. 285.
[26] Károly F. Pál and Tamás Vértesi, "Maximal violation of a bipartite three-setting, two-outcome bell inequality using infinite-dimensional quantum systems," Phys. Rev. A **82** (2010), 10.1103/physreva.82.022116.
[27] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner, "Bell nonlocality," (2013), 10.1103/RevModPhys.86.419, arXiv:1303.2849.