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: November 15, 2016)

Quantum correlations are often incompatible with a classical assumption of causal structure. This nonclassicality is often known as quantum nonlocality, and it is witnessed through the violation of causal compatability inequalities, such as Bell inequalities. Such inequalities were recently derived for the Triangle scenario [arXiv:1609.00672], begging the question: can these inequalities be violated by quantum correlations? Here we answer this affirmatively, and discuss specific Triangle scenario inequalities and quantum configurations which manifest nonclassical correlations. We also report the development of novel computational techniques suitable for large causal structures with numerous outcomes. Numerical optimzations reveal quantum resources potentially qualitatively different from those known previously.

* tcfraser@tcfraser.com

CONTENTS

I. INTRODUCTION

Todo (TC Fraser): Mention Bell's work

Todo (TC Fraser): Violations of inequalities correspond to resources

Todo (TC Fraser): Finding inequalities is hard

Todo (TC Fraser): Causal inference is a great lens to answer these questions

Todo (TC Fraser): Demonstrated the usefulness of the inflation technique

Todo (TC Fraser): Definite Extension procedure

Todo (TC Fraser): Optimizations against inequalities

II. COMPATIBILITY, CONTEXTUALITY AND THE MARGINAL PROBLEM

A. Preliminary Notions

A reoccurring motif of these discussions will be a set of marginal probability distributions $\{P_{V_1}, \ldots, P_{V_k}\}$. In agreement with [1] we adopt the following definitions and notation.

Definition 1. A set of random variables $\mathcal{J} = \{v_1, \ldots, v_n\}$ is referred to as the **joint random variables** when denoted with a $\mathcal{J}$. A subset of joint variables $V \subset \mathcal{J}$ is called **marginal context**. A collection of marginal contexts is called the **marginal scenario**[1] and is denoted $\mathcal{M}$[2].

$$\mathcal{M} \equiv \{V_1, \ldots, V_k \mid V_i \subseteq \mathcal{J}\}$$

Additionally, the set of probability distributions $\{P_{V_1}, \ldots, P_{V_k}\}$ over the marginal scenario is called a **marginal model** is compactly denoted $P^{\mathcal{M}}$.

A specific marginal model $P^{\mathcal{M}}$ for a marginal scenario $\mathcal{M}$ abstractly represents a set of measurements or observations made on a physical system. Associated with the physical system are physical models or causal structures.

Definition 2. A **causal structure** is a directed acyclic graph (DAG) $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with the extra distinction between two classes of nodes: the **latent nodes** $\mathcal{N}_L$ and **observed nodes** $\mathcal{N}_O$.

The latent nodes of a causal structure correspond aim to represent completely general hidden variables that are hidden by some fundamental process or cannot be measured due to other limitations. The observed nodes represent random variables that are measurable and in practice will be observed as a collection of marginal contexts $\mathcal{M}$. Any observation over this particular marginal scenario is expressible as a marginal model $P^{\mathcal{M}}$. The field of causal inference aims to answer the looming question: Is the observed marginal model compatible with the hypothesized causal structure?

In order to determine if a given marginal model $P^{\mathcal{M}}$ is compatible with a causal structure $\mathcal{G}$, a formal notion of *compatibility* is warranted; first for a single marginal distribution P_V and then for a complete marginal model $P^{\mathcal{M}} = \{P_{V_1}, \ldots, P_{V_k}\}$.

---

[1] Rigorously speaking as defined by [1], a marginal scenario forms an *abstract simplicial complex* where it is required that $V' \in \mathcal{M}$ whenever $V' \subseteq V \in \mathcal{M}$. Throughout this document, we only consider the *maximal marginal scenario*; restricting our focus to the largest contexts in the marginal scenario $\forall i, j \in 1, \ldots, k : V_i \not\subseteq V_j$.

[2] It is also assumed that the marginal scenario covers the joint variables $\mathcal{J} = \bigcup_{i=1}^{k} V_i$.

Definition 3. 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 its parents in $\mathcal{G}$.

$$\left\{ P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N} \right\} \tag{1}$$

Definition 4. A marginal distribution P_V is **compatible** with a causal structure $\mathcal{G}$ if there exists a *choice* of causal parameters such that P_V can be *recovered* from the following series of operations:

1. Obtain a joint probability distribution over *all* nodes of of the causal structure using the chosen casual parameters: $P_{\mathcal{N}} = \prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)}$.

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

If it is not possible for P_V to be recovered for *any* choice of causal parameters, then P_V is **incompatible**. To generalize, a marginal model $P^{\mathcal{M}}$ is compatible with $\mathcal{G}$ if each distribution $P_V \in P^{\mathcal{M}}$ can be made compatible by the *same* choice of causal parameters; otherwise it is incompatible.

There are two reasons why a given marginal model $P^{\mathcal{M}}$ would be deemed incompatible. One possibility is that the marginal model $P^{\mathcal{M}}$ does not admit a joint distribution $P_{\mathcal{J}}$ such that it is impossible for requirement 3 of theorem 4 to hold[3]. Alternatively, it could be possible that a joint distribution $P_{\mathcal{J}}$ exists but can not factor into causal parameters as per requirement 1 of theorem 4. The former point-of-failure, namely the non-admission of a joint distribution, will be principle focus of this paper. Colloquially, the existence (or non-existence) of a joint distribution is referred to as *the marginal problem.*

Definition 5. **The Marginal Problem:** Given a marginal model $P^{\mathcal{M}} = \{P_{V_1}, \ldots, P_{V_k}\}$ marginal to the joint variables $\mathcal{J} = \bigcup_i V_i$, does there exist a joint distribution $P_{\mathcal{J}}$ such that each given distribution P_{V_i} can be obtained by marginalizing $P_{\mathcal{J}}$?

$$\forall V \in \mathcal{M} : P_V = \sum_{\mathcal{J} \setminus V} P_{\mathcal{J}} \tag{2}$$

Following definition 2.3 in [1], a marginal model is said to be **contextual** if it *does not* admit a joint distribution $P_{\mathcal{J}}$ and **non-contextual** otherwise.

There are two perspectives that can be taken when trying to answer the marginal problem. The first comes as stated by theorem 5; when inured with a particular marginal model $P^{\mathcal{M}}$, can you construct a joint distribution that satisfies theorem 5? Alternatively and more abstractly, when provided a marginal *mscenario* $\mathcal{M}$ what is the set of all marginal models that are (non-)contextual? To facilitate future discussions, let $\mathcal{NC}$ denote the set of marginal models that are non-contextual (admitting a joint distribution) and let $\mathcal{C}$ denote the set of contextual marginal models. It becomes desirable to identifying conditions that constrain either $\mathcal{NC}$ or $\mathcal{C}$. We will continue to refer to a **complete solution marginal problem** as a finite set of necessary and sufficient conditions able to discern if a given marginal model $P^{\mathcal{M}}$ is in $\mathcal{NC}$ or $\mathcal{C}$. These constraints typically take the form of linear inequalities over probabilities $P^{\mathcal{M}}$.

Definition 6. A $\mathcal{NC}$**-inequality** $I_{\mathcal{M}}$ for a marginal scenario $\mathcal{M}$ is a probabilistic inequality that is obeyed for every non-contextual marginal model $P^{\mathcal{M}}$.

In the section II B, we will explore and briefly mention existing techniques for completely (or partially) solving the marginal problem. The basis of understanding being that eq. (2) is a *linear* operation. In certain circumstances, eq. (2) can be represented as a matrix operation.

Todo (TC Fraser): Maybe define the possibilistic marginal problem for later

---

[3] It is typical for $\mathcal{N}_O = \mathcal{J}$. At the very least, it is demanded that $\mathcal{J} \subseteq \mathcal{N}_O$.

B. Casting the Marginal Problem as a Linear Program

Heretofore, our discussions of compatibility, contextuality and the marginal problem have be completely general, making no reference to the nature of the outcomes of the random variables $\mathcal{J}$; whether they be discrete or continuous, finite or unbounded, or otherwise. Todo (TC Fraser): It is entirely possible to obtain complete solutions to the marginal problem for continuous and unbounded measure spaces using entropic inequalities. [Elaborate and reference] In this paper we restrict our attention to finite outcomes. In particular, we will focus on marginal scenarios that are finite but have a large number of outcomes. To assist with the subsequent analysis, we recall some familiar notions of probability theory together with new terminology and notation.

1. Events

To every discrete random variable v there corresponds a prescribed set of **outcomes** O_v. We also define the set of all **events over** v $\mathcal{E}(v)$[4] to be the set of all functions $s : \{v\} \to O_v$ each representing the event that a measurement on v was made where $s(v) \in O_v$ was observed.

$$\mathcal{E}(v) \equiv \{s : \{v\} \to O_v\}$$

Evidently, $\mathcal{E}(v)$ and O_v have a one-to-one correspondence and this distinction can be confounding. There is rarely any harm in referring synonymously to either as outcomes. Nonetheless, a sheaf-theoretic treatment of contextuality [2] demands the distinction. Specifically for this work, the distinction becomes essential for our discussion and exploit of marginal symmetries in section IV. As a natural generalization we define the event over a collection of random variables $V = \{v_1, \ldots, v_n\}$ in a parallel manner,

$$\mathcal{E}(V) \equiv \{s : V \to O_V \mid \forall i : s(v_i) \in O_{v_i}\} \tag{3}$$

Each event s can be compactly represented as a set of mappings over each element of V.

$$s = \{v_1 \mapsto s(v_1), \ldots, v_k \mapsto s(v_k)\} = \{v_i \mapsto s(v_i)\}_{i=1}^k$$

The domain $\mathcal{D}(s)$ of an event s is the set of random variables it valuates, i.e. if $s \in \mathcal{E}(V)$, then $\mathcal{D}(s) = V$.

Example 7. Let $s \in \mathcal{E}(V)$ be an event over V, a probability distribution P_V over V will assign a non-negative probability $P_V(s)$ to s. Let $V = \{v_1, v_2\}$ with each $v \in V$ having 3 distinct outcomes $O_v = \{0, 1, 2\}$. Furthermore let $f \in \mathcal{E}(V)$ be the section $(v_1 \mapsto 2, v_2 \mapsto 1)$. Then the following notations are interchangeable,

$$P(s) = P_{\mathcal{D}(s)}(s) = P_V(s) = P_{v_1, v_2}(s) = P(v_1 \mapsto s_1(v_1), v_2 \mapsto s_2(v_2)) = P(v_1 \mapsto 2, v_2 \mapsto 1) = P_{v_1 v_2}(21)$$

Each of which should be read as "the probability that v_1 got outcome 2 and v_2 got outcome 1"

Definition 8. For every $W \subseteq V$ and $s \in \mathcal{E}(V)$, the **restriction of s onto W** (denoted $s|_W \in \mathcal{E}(W)$) is the event in $\mathcal{E}(W)$ that *agrees* with each of s's assignments for variables in W.

$$\forall v \in W : s|_W(v) = s(v)$$

Definition 9. The events $s_V \in \mathcal{E}(V)$ and $s_W \in \mathcal{E}(W)$ are said to be **compatible** events if they agree on their valuations of $V \cap W$. Compatibility will be denoted with the infix '$\asymp$'.

$$s_V \asymp s_W \iff s_V|_{V \cap W} = s_W|_{V \cap W}$$

Definition 10. It $s_V \in \mathcal{E}(V)$ and $s_W \in \mathcal{E}(W)$ compatible events and $W \subseteq V$ then s_W is said to be **extendable** to s_V. Extendability will be denoted with the infix '$\prec$'.

$$s_W \prec s_V \iff s_W \asymp s_V, W \subseteq V$$

Extendability is dual to the notion of a **restriction** of s_V; s_V can be restricted to s_W ($s_V \succ s_W$) if and only if s_W is extendable to s_V.

---

[4] In the language of sheaf-theory, $\mathcal{E}(v)$ is the *sheaf of events* [2].

Remark 11. The notation '$s \rightthreetimes t$' is used to indicate compatibility between $s \in \mathcal{E}(V)$ and $t \in \mathcal{E}(W)$ because of the following observation: s and t are compatible if and only if they are extensions of some event $r \in \mathcal{E}(V \cap W)$.

$$s \rightthreetimes t \iff \exists r \in \mathcal{E}(V \cap W) \mid s \succ r \prec t$$

In fact r is uniquely defined.

$$\forall v \in V \cap W : r(v) = s(v) = t(v)$$

For every marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$, it is useful to put special emphasis on the **joint events** $\mathcal{E}(\mathcal{J})$. Similarly we define the **context events** $\{\mathcal{E}(V) \mid V \in \mathcal{M}\}$. Finally, we elect to define the disjoint union of all context events are called the **marginal events** and by an abuse of notation we will denote this disjoint union as $\mathcal{E}(\mathcal{M})$.

$$\mathcal{E}(\mathcal{M}) = \coprod_{V \in \mathcal{M}} \mathcal{E}(V)$$

Each marginal section $m \in \mathcal{E}(\mathcal{M})$ has a domain $\mathcal{D}(m) = V$ for some $V \in \mathcal{M}$.

2. Linear Programs

We are now fully equipped to write the marginal problem as a matrix multiplication.

Definition 12. The **incidence matrix** M for a marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$ is a bit-wise matrix where the columns are indexed by *joint* events $j \in \mathcal{E}(\mathcal{J})$ and the rows are events by *marginal* sections $m \in \mathcal{E}(\mathcal{M})$. The entries of M are populated whenever a row index is extendable to a column index.

$$M_{m,j} = \delta_{m \prec j} \equiv \begin{cases} 1 & m \prec j \\ 0 & \text{otherwise} \end{cases}$$

Where $\delta_{m \prec j}$ is termed the **extendability indicator**. The incidence matrix has $|\mathcal{E}(\mathcal{J})|$ columns, $|\mathcal{E}(\mathcal{M})| = \sum_i |\mathcal{E}(V_i)|$ rows and $k|\mathcal{E}(\mathcal{J})|$ non-zero entries.

To illustrate this concretely, consider the following example:

Example 13. Let $\mathcal{J}$ be 3 binary variables $\mathcal{J} = \{a, b, c\}$ and $\mathcal{M}$ be the marginal scenario $\mathcal{M} = \{\{a, b\}, \{b, c\}, \{a, c\}\}$. The incidence matrix becomes:

$$M = \quad$$

$(a,b,c) \mapsto$	(0,0,0)	(0,0,1)	(0,1,0)	(0,1,1)	(1,0,0)	(1,0,1)	(1,1,0)	(1,1,1)
$(a \mapsto 0, b \mapsto 0)$	1	1	0	0	0	0	0	0
$(a \mapsto 0, b \mapsto 1)$	0	0	1	1	0	0	0	0
$(a \mapsto 1, b \mapsto 0)$	0	0	0	0	1	1	0	0
$(a \mapsto 1, b \mapsto 1)$	0	0	0	0	0	0	1	1
$(b \mapsto 0, c \mapsto 0)$	1	0	0	0	1	0	0	0
$(b \mapsto 0, c \mapsto 1)$	0	1	0	0	0	1	0	0
$(b \mapsto 1, c \mapsto 0)$	0	0	1	0	0	0	1	0
$(b \mapsto 1, c \mapsto 1)$	0	0	0	1	0	0	0	1
$(a \mapsto 0, c \mapsto 0)$	1	0	1	0	0	0	0	0
$(a \mapsto 0, c \mapsto 1)$	0	1	0	1	0	0	0	0
$(a \mapsto 1, c \mapsto 0)$	0	0	0	0	1	0	1	0
$(a \mapsto 1, c \mapsto 1)$	0	0	0	0	0	1	0	1

$$(4)$$

Definition 14. The **joint distribution vector** $\mathcal{P}^{\mathcal{J}}$ for a probability distribution $P_{\mathcal{J}}$ is the vector whose entries are the positive, real-valued probabilities that $P_{\mathcal{J}}$ assigns to each joint event $j \in \mathcal{E}(\mathcal{J})$. $\mathcal{P}^{\mathcal{J}}$ shares the same indices as the *column* indices of M.

$$\forall j \in \mathcal{E}(\mathcal{J}) : \mathcal{P}^{\mathcal{J}}_j = P_{\mathcal{J}}(j)$$

Definition 15. Analogously, **marginal distribution vector** $\mathcal{P}^{\mathcal{M}}$ for a marginal model $P^{\mathcal{M}} = \{P_{V_1}, \ldots, P_{V_k}\}$ is the vector whose entries are probabilities over the set of marginal outcomes $\mathcal{E}(\mathcal{M})$. $\mathcal{P}^{\mathcal{M}}$ shares the same indices as the *row* indices of M.

$$\forall m \in \mathcal{E}(\mathcal{M}) : \mathcal{P}^{\mathcal{M}}_m = P_{\mathcal{D}(m)}(m)$$

The marginal and joint distribution vectors are related via the incidence matrix M. Given a joint distribution vector $\mathcal{P}^{\mathcal{J}}$ one can obtain the marginal distribution vector $\mathcal{P}^{\mathcal{M}}$ by multiplying M by $\mathcal{P}^{\mathcal{J}}$.

$$\mathcal{P}^{\mathcal{M}} = M\mathcal{P}^{\mathcal{J}} \tag{5}$$

Henceforth we will use $\mathcal{P}^{\mathcal{J}}$ ($\mathcal{P}^{\mathcal{M}}$) and $P_{\mathcal{J}}$ ($P_{\mathcal{M}}$) interchangeably.

Remark 16. The particular ordering of the rows and columns of M carries no importance, but it *must* be consistent between M, $\mathcal{P}^{\mathcal{J}}$ and $\mathcal{P}^{\mathcal{M}}$.

The marginal problem can now be rephrased in the language of the incidence matrix. Suppose one obtains a marginal distribution vector $\mathcal{P}^{\mathcal{M}}$. The marginal problem becomes equivalent to the question: *Does there exist a joint distribution vector $\mathcal{P}^{\mathcal{J}}$ such that eq.* (5) *holds?*

Definition 17. The Marginal Linear Program is the following linear program:

$$\begin{aligned} \text{minimize:} \quad & \emptyset x \\ \text{subject to:} \quad & x \succeq 0 \\ & Mx = \mathcal{P}^{\mathcal{M}} \end{aligned}$$

If this "optimization"[5] is *feasible*, then there exists a vector x than can satisfy eq. (5) and is a valid joint distribution vector. Therefore feasibility implies that $\mathcal{P}^{\mathcal{J}} = x$. Moreover if the marginal linear program is *infeasible*, then there *does not* exist a joint distribution $\mathcal{P}^{\mathcal{J}}$ over all random variables.

Todo (TC Fraser): Talk about about the linear solvers and how this is easy to implement.

Definition 18. The Dual Marginal Linear Program is the dual of theorem 17 formulated via a procedure similar to [3]:

$$\begin{aligned} \text{minimize:} \quad & y\mathcal{P}^{\mathcal{M}} \\ \text{subject to:} \quad & yM \succeq 0 \end{aligned}$$

Where y is a real valued vector with the same length as $\mathcal{P}^{\mathcal{M}}$.

The dual marginal linear program not only answers the marginal problem for a specific marginal model $\mathcal{P}^{\mathcal{M}}$ but as a by-product provides an inequality that witnesses its contextuality. If this is not obvious, first notice that the dual problem is *never infeasible*; by choosing y to be trivially the null vector $\emptyset$ of appropriate size, all constraints become satisfied. Secondly if the dual constraint $yM \succeq 0$ holds *and* the primal is feasible, then the following must hold:

$$y\mathcal{P}^{\mathcal{M}} = yMx \geq 0 \tag{6}$$

Therefore the *sign* of the dual value $d \equiv \min\left(y\mathcal{P}^{\mathcal{M}}\right)$ classifies a marginal model's contextuality. If $d < 0$ then eq. (6) is violated and therefore $\mathcal{P}^{\mathcal{M}}$ is contextual. Likewise if $d \geq 0$ (satisfying eq. (6)), then $\mathcal{P}^{\mathcal{M}}$ is non-contextual. [6]

Definition 19. An **infeasibility certificate** [5] is any vector y that satisfies the constraint of theorem 18 and also permits violation of eq. (6) for some marginal distribution vector $\mathcal{P}^{\mathcal{M}}$.

$$y \in \mathbb{R}^{|\mathcal{E}(\mathcal{M})|} : yM \succeq 0, \quad y\mathcal{P}^{\mathcal{M}} < 0$$

Furthermore, for every y satisfying $yM \succeq 0$ there corresponds a **certificate inequality** that constraints the set of non-contextual marginal models $\mathcal{NC}$,

$$\forall \mathcal{P}^{\mathcal{M}} \in \mathcal{NC} : y\mathcal{P}^{\mathcal{M}} \geq 0$$

Todo (TC Fraser): Discuss Infeasibility Certificates basis

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

Todo (TC Fraser): Discuss Compatibility, connection to cooperative games/resources, bell incompatibility?

----

[5] "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*.

[6] Actually, 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 [4]. Comment (TC Fraser): Is this really the CSP? Moreover, if $d < 0$, then it is unbounded $d = -\infty$. This latter point becomes clear upon recognizing that for any y with $d < 0$, another $y' = \alpha y$ can be constructed (with $\alpha > 1$) such that $d' = \alpha d < d$. 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 [4]; if the dual is unbounded, then the primal is infeasible. Comment (TC Fraser): Farkas's lemma here?

3. Fourier Motzkin Elimination & Polytope Projection

The complete space of marginal models over $\mathcal{M}$ can be partitioned into two spaces: contextual ($\mathcal{C}$) and non-contextual marginal models ($\mathcal{NC}$). The separation between these spaces is completely characterized by the system of inequalities associated with eq. (5). Explicitly,

$$
\begin{aligned}
\forall m \in \mathcal{E}(\mathcal{M}): \quad & \mathcal{P}_m^{\mathcal{M}} - \sum_j M_{m,j} \mathcal{P}_j^{\mathcal{J}} \geq 0 \\
\forall m \in \mathcal{E}(\mathcal{M}): \quad & -\mathcal{P}_m^{\mathcal{M}} + \sum_j M_{m,j} \mathcal{P}_j^{\mathcal{J}} \geq 0 \\
\forall j \in \mathcal{E}(\mathcal{J}): \quad & \mathcal{P}_j^{\mathcal{J}} \geq 0
\end{aligned}
\tag{7}
$$

Geometrically eq. (7) is a family of $2|\mathcal{E}(\mathcal{M})| + |\mathcal{E}(\mathcal{J})|$ half-planes bounding $\mathbb{R}^{|\mathcal{E}(\mathcal{M})|} \times \mathbb{R}^{|\mathcal{E}(\mathcal{J})|}$. However testing the contextuality of a marginal model $\mathcal{P}^{\mathcal{M}}$ requires a collection of inequalities bounding just $\mathbb{R}^{|\mathcal{E}(\mathcal{M})|}$. Transforming eq. (7) into the desired system of inequalities is the task of linear-quantifier elimination. A popular tool for linear quantifier elimination is Fourier-Motzkin elimination [6–9]. We will avoid expounding the Fourier-Motzkin procedure in this paper and instead reproduce a few of its notable features and consequences.

Definition 20. [10, Section 12.2] The **Fourier-Motzkin elimination** procedure begins with a matrix A and a vector b defining a system of inequalities $S_1 = \{Ax \leq b\}$ constraining a set of variables x and symbolically eliminates a sub-set of theses variables $x \setminus y$. Upon completion, one obtains a new set of constraints $S_2 = \{A'y \leq b'\}$ such that,

$$
\exists y : A'y \leq b' \iff \exists x : Ax \leq b
$$

Any solution y of S_2 will permit at least one solution x in S_1 (and vice versa).

When applied to the marginal problem of eq. (7), the Fourier-Motzkin procedure eliminates all reference to the joint probabilities $\mathcal{P}_j^{\mathcal{J}}$ and returns a system of inequalities over the marginal probabilities $\mathcal{P}_m^{\mathcal{M}}$ representing a *complete* solution to the marginal problem. Applying the Fourier-Motzkin procedure to completely solve the marginal problem is discussed in more detail in [7] and [1]. Todo (TC Fraser): Improve this discussion of FMP.

Todo (TC Fraser): Discuss why FM is slow?

Remark 21. In theory, Fourier-Motzkin offers a complete solution to the marginal problem for all marginal scenarios; its limitations are purely computational. In addition, the entire procedure only utilizes field operations [8] meaning that since eq. (7) is a rational system of inequalities, then a complete solution to the marginal problem only requires a finite set of linear, rational coefficient inequalities.

Theorem 22. *There exists a set of necessary and sufficient conditions*

Todo (TC Fraser): Logical Bell inequalities paper: only need to consider rational inequalities, actually integer inequalities, actually homogeneous inequalities
Todo (TC Fraser): Discuss computational complexity
Todo (TC Fraser): Mention what a complete solution looks like after using Fourier Motzkin

4. Marginal Polytope

Definition 23. Let I_1 and I_2 represent two

Definition 24. Let $\mathcal{I}$

Todo (TC Fraser): Very briefly discuss polytope perspective and facet inequalities vs. non-facet inequalities Todo (TC Fraser): Define facets as irreducible inequalities
Todo (TC Fraser): Maybe put the description of our task here. This will motivate the need for more scalable approaches.

III. TRANSVERSAL INEQUALITIES

Section II B 2 discusses how one can obtain a $\mathcal{NC}$-inequality that witnesses the contextuality of a particular marginal model $\mathcal{P}^{\mathcal{M}}$ by writing the marginal problem as a linear program. Every $\mathcal{P}^{\mathcal{M}} \in \mathcal{NC}$ is witness-able by some certificate inequality. Albeit when a candidate marginal model $\mathcal{P}^{\mathcal{M}}$ is not present, linear optimization offers no utility.

This section aims to develop a procedure for finding *all* certificate inequalities in a computationally efficient manner. Specifically by finding all certificate inequalities that are also facets of the marginal polytope. Before doing so, we will first summarize a method Todo (TC Fraser): Method where this idea came from, original insight from Hardy? that can be used to obtain a complete solution to the *possibilistic* marginal problem [11] and indirectly offer a *partial* solution to the *probabilistic* marginal problem.

A. Hardy Transversals

Let $a \in \mathcal{E}(\mathcal{M})$ be any marginal event and $C = \{c_1, \ldots, c_n\} \subseteq \mathcal{E}(\mathcal{M})$ be a subset of the marginal events. We say that the subset C forms a **logical extension** of a if the following logical implication holds for *all* marginal models $\mathcal{P}^{\mathcal{M}}$.

$$a \implies c_1 \vee \ldots \vee c_n = \bigvee_{c \in C} c \tag{8}$$

Which can be dictated: *Whenever the event a occurs, at least one the events in C occurs.* In this paper, the marginal event a will be referred to as the *antecedent* and C as the *consequent set*. To clarify, a marginal model $\mathcal{P}^{\mathcal{M}}$ satisfies eq. (8) if there always exists at least one $c \in C$ that is possible ($\mathcal{P}_c^{\mathcal{M}} > 0$) whenever a is possible. A marginal model $\mathcal{P}^{\mathcal{M}}$ violates eq. (8) whenever *none* of the sections in C are possible while a is nonetheless still possible. Marginal models that violate logical statements such as eq. (8) are known as **Hardy Paradoxes** [7, 11, 12]. The concept of witnessing quantum incompatibility on a logical level has been analyzed thoroughly for decades [8, 13]; being motivated by a greater sense of robustness compared to probabilistic constraints.

All logical implications of the form eq. (8) can be constructed by first selecting a and then manufacturing a consequent set C persuant to eq. (8).

Definition 25. For each marginal event $m \in \mathcal{E}(\mathcal{M})$ the **extendable set of m into $\mathcal{J}$** is the set of all joint events $j \in \mathcal{E}(\mathcal{J})$ that are extensions of m.

$$\mathsf{Ext}_{\mathcal{J}}(m) = \{j \in \mathcal{E}(\mathcal{J}) \mid m \prec j\}$$

The extendable set of m into $\mathcal{J}$ is the subset of events in $\mathcal{E}(\mathcal{J})$ that *agree* with m about valuations for variables in $\mathcal{D}(m)$. Practically speaking, the extendable set of m into $\mathcal{J}$ is represented by the m-th row of the incidence matrix M.

Lemma 26. *Let $a \in \mathcal{E}(\mathcal{M})$ and $\mathcal{P}^{\mathcal{M}} \in \mathcal{NC}$ be non-contextual. Then,*

$$a \iff \bigvee_{j \in \mathsf{Ext}_{\mathcal{J}}(a)} j \tag{9}$$

The idea being if a joint distribution does exists, then the event a represents partial knowledge of the entire system $\mathcal{J}$; whenever a is measured, exactly one joint event j has occurred in reality and j must be compatible with a. Likewise theorem 26 applies to the consequent set $C \in \mathcal{E}(\mathcal{M})$,

$$\bigvee_{c \in C} c \iff \bigvee_{c \in C} \bigvee_{j \in \mathsf{Ext}_{\mathcal{J}}(c)} j \tag{10}$$

If a subset C can be found such that its extendable set *covers* the extendable set of a,

$$\mathsf{Ext}_{\mathcal{J}}(a) \subseteq \bigcup_{c \in C} \mathsf{Ext}_{\mathcal{J}}(c) \tag{11}$$

Then the combination of eqs. (9) to (11) yields eq. (8). Computationally enumerating all logical extensions $C \subseteq \mathcal{E}(\mathcal{J})$ for a particular antecedent $a \in \mathcal{E}(\mathcal{M})$ corresponds iteratively selecting marginal sections $c \neq a$ for membership in C

until eq. (11) is satisfies. This process is more generally known as a set-covering problem, or a hypergraph transversal problem [7], an idea we return to in the more general picture to follow in section III C. Todo (TC Fraser): Maybe use an example here to define a Hardy transversal.

It is possible to show that for each logical constraint there exists a probabilistic constraint that is tighter. The principle statement of theorem 26 can be taken in a probabilistic form. If the marginal model $\mathcal{P}^{\mathcal{M}}$ is to admit a joint distribution $\mathcal{P}^{\mathcal{J}}$, then it must be that the antecedent probability $\mathcal{P}_a^{\mathcal{M}}$ is equal to a marginalization over its extendable set into $\mathcal{J}$.

$$\mathcal{P}_a^{\mathcal{M}} = \sum_{j \in \mathsf{Ext}_{\mathcal{J}}(a)} \mathcal{P}_j^{\mathcal{J}} \tag{12}$$

This is nothing more than the the a-th row of eq. (5). Likewise for the consequent set,

$$\sum_{c \in C} \mathcal{P}_c^{\mathcal{M}} = \sum_{c \in C} \sum_{j \in \mathsf{Ext}_{\mathcal{J}}(c)} \mathcal{P}_j^{\mathcal{J}} \tag{13}$$

Since C is a logical extension of a, the right-hand side of eq. (12) *algebraically* covers eq. (13)[7]. Therefore one recovers the probabilistic version of eq. (8),

$$\mathcal{P}_a^{\mathcal{M}} \leq \sum_{c \in C} \mathcal{P}_c^{\mathcal{J}} \tag{14}$$

Due to the logical precedent, inequalities of the form of eq. (14) are called **Hardy inequalities**. Marginal models that are contextual on a logical level are **possibilistically contextual**. In [2] it was demonstrated that all possibilistically contextual models have their contextuality witnessed by logical extensions. In this sense, logical implications of the form eq. (8) completely solve the possibilistic marginal problem. Todo (TC Fraser): Discuss logical bell inequalities

Remark 27. The probabilistic logical extension eq. (14) is an *upgraded* version of eq. (8) in the following sense.

- Every $\mathcal{P}^{\mathcal{M}} \in \mathcal{NC}$ satisfies both eq. (14) and eq. (8).

- Every $\mathcal{P}^{\mathcal{M}} \in \mathcal{C}$ that violates eq. (8) will also violate eq. (14); but the converse is not necessarily true.

B. Definite Extensions

We will now generalize the notion of a Hardy transversal by considering multi-sets of antecedents corresponding to multiple marginal events each accompanied by an assigned multiplicity. Afterwards we provide an algorithm to completely solve the marginal problem by transversing generalized hypergraphs. To our benefit, this technique for solving the marginal problem offers considerable advantages over existing methods when dealing with large marginal scenarios.

To facilitate the discussion, some definitions and notation are necessary.

Definition 28. A **multi-set** or **bag** γ for a marginal scenario $\mathcal{M}$ is a vector indexed by marginal events $m \in \mathcal{E}(\mathcal{M})$ whose entries are integers $\gamma_m \in \mathbb{Z}$. The **support** $\sigma(\gamma)$ of an bag is the subset of marginal events that are assigned non-zero multiplicity.

$$\sigma(\gamma) \equiv \{m \in \mathcal{E}(\mathcal{M}) \mid \gamma_m \neq 0\}$$

The **order** $|\gamma|$ of a multi-set γ is entry-wise sum of γ.

$$|\gamma| = \sum_{m \in \mathcal{E}(\mathcal{M})} \gamma_m = \sum_{m \in \sigma(\gamma)} \gamma_m$$

---

[7] If there were a joint event j^* present on the right-hand side of eq. (12) but not the right-hand side of eq. (13), then eq. (14) would be violated by the deterministic joint distribution where $\mathcal{P}_{j^*}^{\mathcal{J}} = 1$

Definition 29. The **marginal indicator** vector δ^m for a marginal event m is the unit vector indexed by marginal outcomes $\mathcal{E}(\mathcal{M})$ whose only non-zero entry is the marginal event m.

$$\forall m' \in \mathcal{E}(\mathcal{M}) : \delta^m_{m'} = \begin{cases} 1 & m' = m \\ 0 & m' \neq m \end{cases}$$

Analogously defined is the **joint indicator** vector δ^j.

Definition 30. The **extension Γ of an multi-set γ** for a marginal scenario $\mathcal{M}$ is a vector indexed by *joint* events $j \in \mathcal{E}(\mathcal{J})$ and can be constructed by left-multiplication of γ into the incidence matrix M.

$$\Gamma \equiv \gamma M$$

For every multi-set γ there corresponds a candidate inequality $I_\mathcal{M} = \{\gamma \mathcal{P}^\mathcal{M} \geq 0\}$ whose validity is completely characterized by Γ.

Theorem 31. *Let γ be a multi-set for the marginal scenario $\mathcal{M}$. Then the inequality $I_\mathcal{M} = \{\gamma \mathcal{P}^\mathcal{M} \geq 0\}$ is a valid non-contextuality inequality if and only if Γ is positive semi-definite.*

$$\forall \mathcal{P}^\mathcal{M} \in \mathcal{NC} : \gamma \mathcal{P}^\mathcal{M} \geq 0 \iff \Gamma \succeq 0 \tag{15}$$

Proof. The proof follows directly from eq. (5). Since $\mathcal{P}^\mathcal{M} = M\mathcal{P}^\mathcal{J}$,

$$\gamma \mathcal{P}^\mathcal{M} = \gamma M \mathcal{P}^\mathcal{J} = \Gamma \mathcal{P}^\mathcal{J}$$

If $\Gamma \succeq 0$, then $\Gamma \mathcal{P}^\mathcal{J} \geq 0$ which gives $\gamma \mathcal{P}^\mathcal{M} \geq 0$. The forward case of eq. (15) can be proven by contradiction. Suppose that $\gamma \mathcal{P}^\mathcal{M} \geq 0$ for all marginal models but $\Gamma \not\succeq 0$. Consequently there exists at least one joint outcome j^* such that $\Gamma_{j^*} < 0$. This negative entry can be exploited by a deterministic joint distribution $\mathcal{P}^\mathcal{J} = \delta^{j^*}$,

$$0 > \Gamma_{j^*} = \Gamma \delta^{j^*} = \gamma M \delta^{j^*}$$

Which is a contradiction: $M\delta^{j^*}$ is a non-contextual marginal model but $\gamma M \delta^{j^*} \not\geq 0$. $\qquad\square$

Clearly if the multi-set itself is positive semi-definite $\gamma \succeq 0$ then $\Gamma \succeq 0$ is also. Corresponding inequalities are then *trivial*: being satisfied by all $\mathcal{P}^\mathcal{M} \in \mathcal{NC}$. As such, it is desired to obtain definite extensions Γ from *indefinite* multi-sets γ.

Corollary 32. *Let γ be a multi-set for the marginal scenario $\mathcal{M}$. Then the equality $I_\mathcal{M} = \{\gamma \mathcal{P}^\mathcal{M} = 0\}$ is satisfied by all non-contextual marginal models if and only if Γ is a null vector (i.e. $\forall j : \Gamma_j = 0$).*

Theorem 31 demonstrates it is possible to generate non-contextuality inequalities by finding multi-sets γ such that $\Gamma \succeq 0$. In light of this, we will formally define the subset of joint outcomes j that are preventing a candidate multi-set γ from having positive semi-definite extensions.

Definition 33. The **inhibiting set** $\mathcal{I}(\gamma)$ for a multi-set γ is the subset of joint outcomes preventing $\Gamma \succeq 0$ from holding true.

$$\mathcal{I}(\gamma) = \left\{ j \in \mathcal{E}(\mathcal{J}) \mid (\gamma M)_j = \Gamma_j < 0 \right\}$$

Procedurally, definite extensions can be built by beginning with a seed multi-set $\gamma^{(0)} = \gamma^A$ composed entirely of non-positive entries and iteratively appending marginal events m $(\gamma^{(n+1)} = \gamma^{(n)} + \delta^m)$ in such a way that progresses $\Gamma^{(n)}$ toward becoming positive semi-definite (i.e. by decreasing the inhibiting set). Taking inspiration from logical, Hardy-type, transversal inequalities discussed in section III A, we can view this process as a weighted hypergraph transversal problem.

C. Definite Extension Procedure

Before discussing the definite extension procedure (DEP) directly, we will motivate its main algorithmic component (weighted hypergraphs transversals) with a short example.

$$M = \begin{array}{c} \\ m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \\ m_7 \\ m_8 \\ m_9 \\ m_{10} \\ m_{11} \\ m_{12} \end{array} \begin{array}{cccccccc} j_1 & j_2 & j_3 & j_4 & j_5 & j_6 & j_7 & j_8 \\ \left(1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \right. \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ \left. 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \right) \end{array}$$

FIG. 1. The incidence matrix for the binary marginal scenario $\mathcal{M} = \{\{a,b\}, \{b,c\}, \{a,c\}\}$ with marginal outcomes $m_1, \dots, m_{12}$ and joint outcomes $j_1, \dots, j_{12}$.
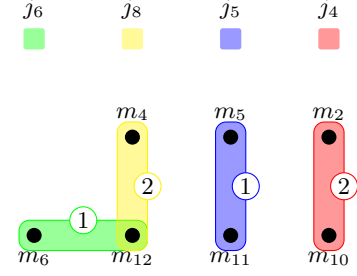

FIG. 2. An edge-weighted hypergraph $\mathcal{H}_\mathcal{W}$ corresponding to the antecedent set γ^A with $\gamma^A_{m_3} = -1$ and $\gamma^A_{m_8} = -2$.

1. Example

Example 34. Recall the exemplary incidence matrix M of eq. (4), theorem 13. Our aim is generate some valid non-contextuality inequalities for marginal models $\mathcal{P}^\mathcal{M}$. To compactify notation, enumerate the marginal outcomes (rows of M) 1 through 12 ($m_1, \dots, m_{12}$) and the joint outcomes (columns of M) 1 through 8 ($j_1, \dots, j_8$) (see fig. 1). Now suppose our antecedent set is γ^A and has two specified entries:

$$\gamma^A_{m_3} = -1 \quad \gamma^A_{m_8} = -2$$

Then the antecedent extension is given by $\Gamma^A = \gamma^A M$:

$$\Gamma^A_{j_4} = \Gamma^A_{j_8} = -2 \quad \Gamma^A_{j_5} = \Gamma^A_{j_6} = -1$$

The objective then becomes to *cover* the joint events of the inhibiting set $\{j_4, j_5, j_6, j_8\}$ with marginal events organized as a consequent multi-set $\gamma^C \succeq 0$ such that,

$$\Gamma^C + \Gamma^A \succeq 0$$

Taking motivation from section III A, this procedure can be casted as a multi-node hitting problem, or an edge-weighted hypergraph problem. Construct a weighted hypergraph $\mathcal{H}_\mathcal{W}$ (see fig. 2) where the edges are labeled by joint outcomes $\mathcal{E} = \{e_{j_4}, e_{j_5}, e_{j_6}, e_{j_8}\}$. These edges are assigned weights w_j precisely equal to the magnitude of Γ^A.

$$\forall j \in \{j_4, j_5, j_6, j_8\} : w_j = -\Gamma^A_j$$

These weights characterize the joint events that are preventing Γ from becoming positive semi-definite. The nodes represent all possible marginal events that have the potential to contribute to γ^C.

$$\mathcal{N} = \{m_2, m_4, m_5, m_6, m_{10}, m_{11}, m_{12}\}$$

A node m is a member of edge e_j if $m \prec j$. By inspection of fig. 2 one can identify numerous multi-sets of γ^C of $\mathcal{N}$ that hit each edge w_j times. As a singular example,

$$\mathcal{T} = \{m_2, m_{10}, m_{11}, m_{12}, m_{12}\} = \begin{array}{ccccccc} m_2 & m_4 & m_5 & m_6 & m_{10} & m_{11} & m_{12} \\ (\ 1 & 0 & 0 & 0 & 1 & 1 & 2 \ ) \end{array}$$

satisfies this condition. We call such $\mathcal{T}$ *transversals* of $\mathcal{H}_\mathcal{W}$. This multi-set enables one to construct the positive semi-definite extension $\Gamma^C + \Gamma^A$ where $\Gamma^C = \mathcal{T}M$. This corresponds to the inequality $-\gamma^A \mathcal{P}^\mathcal{M} \leq \gamma^C \mathcal{P}^\mathcal{M}$,

$$\mathcal{P}^\mathcal{M}_{m_3} + 2\mathcal{P}^\mathcal{M}_{m_8} \leq \mathcal{P}^\mathcal{M}_{m_2} + \mathcal{P}^\mathcal{M}_{m_{10}} + \mathcal{P}^\mathcal{M}_{m_{11}} + 2\mathcal{P}^\mathcal{M}_{m_{12}}$$
$$P(m_3) + 2P(m_8) \leq P(m_2) + P(m_{10}) + P(m_{11}) + 2P(m_{12})$$
$$P_{ab}(10) + 2P_{bc}(11) \leq P_{ab}(01) + P_{ac}(01) + P_{ac}(10) + 2P_{ac}(11)$$

A **hypergraph** $\mathcal{H}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *hyperedges* respectively where the hyperedges are *subsets* of nodes $\forall e \in \mathcal{E} : e \subset \mathcal{N}$. A hypergraph deviates from a traditional graph in that the edges have the capacity to contain more or less than 2 nodes. There is a dual correspondence between the hyperedges $e \in \mathcal{E}$ and nodes $n \in \mathcal{N}$ of a hypergraph: just as a hyperedge e is interpreted as a set of nodes, a node n can be viewed as the set of hyperedges $\{e \mid n \in e\}$ that contain it. 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 a non-empty intersection with every hyperedge in $\mathcal{E}$.

$$\mathcal{T} \subseteq \mathcal{N} \mid \forall e \in \mathcal{E} : \mathcal{T} \cap e \neq \emptyset \tag{16}$$

A transversal $\mathcal{T}$ is said to be **minimal** if there are no redundant nodes n in $\mathcal{T}$, i.e. no node n can be removed from $\mathcal{T}$ such that $\mathcal{T} \setminus n$ remains a transversal. The condition for minimality expressed,

$$\forall n \in \mathcal{T} : \exists e \in \mathcal{E} \mid (\mathcal{T} \setminus n) \cap e = \emptyset \tag{17}$$

When defining a *weighted* hypergraph, there can be some ambiguity as to whether or not the weights $\mathcal{W}$ are associated with the hyperedges or nodes. Since the definite extension procedure concerns itself with weighted hyperedges, we refer to edge-weighted hypergraphs simply as weighted hypergraphs.

Definition 35. A **(edge-)weighted hypergraph** $\mathcal{H}_\mathcal{W}$ is a hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ together with a set of real-valued weights $\mathcal{W} = \{w_e \mid e \in \mathcal{E}\}$ assigned to each hyperedge e.

In the case of the definite extension procedure, these weights will typically be positive integers $w_e \in \mathbb{Z}_{>0}$. As such, we make this the default setting moving forward. Accompanying weighted hypergraphs are the notion of weighted transversals.

Definition 36. A **weighted transversal** $\mathcal{T}$ of a weighted hypergraph $\mathcal{H}_\mathcal{W}$ is a *multi-set* or *vector* of nodes $\forall n \in \mathcal{N} : \mathcal{T}_n \in \mathbb{N}$ such that the total multiplicity of nodes in each hyperedge e surpasses (or matches) the assigned weight w_e.

$$\forall e \in \mathcal{E} : \sum_{n \in e} \mathcal{T}_n \geq w_e \tag{18}$$

Definition 37. A weighted transversal $\mathcal{T}$ of a weighted hypergraph $\mathcal{H}_\mathcal{W}$ is a said to be **minimal** or have **minimal-multiplicities** if all its entries $\mathcal{T}_n : n \in \mathcal{N}$ are minimal in the sense that there does not exist a node $n' \in \sigma(\mathcal{T})$ such that $\mathcal{T} - \delta^{n'}$ is still a transversal.

$$\forall n' \in \mathcal{N} : \exists e \in \mathcal{E}, n' \in e : \sum_{n \in e} \mathcal{T}_n - 1 < w_e \tag{19}$$

Equivalently, $\mathcal{T}$ is minimal if there does not exists another transversal $\mathcal{T}' \neq \mathcal{T}$ such that $\forall j : \mathcal{T}'_j \leq \mathcal{T}_j$.

Remark that if all assigned weights are singular $w_e = 1$, the weighted minimality condition eq. (19) reduces to the unweighted minimality condition eq. (17). The principle component of the definite extension procedure can be summarized as follows:

3. *Transversals of Antecedent Hypergraphs*

Given an antecedent multi-set γ^A where $\gamma^A \preceq 0$, identify the *inhibiting set* $\mathcal{I}(\gamma^A)$ responsible for preventing Γ^A from being classified as positive semi-definite. For an antecedent set γ^A where all entries are negative, there are a few equivalent ways of representing the inhibiting set,

$$\mathcal{I}(\gamma^A) = \sigma(\Gamma^A) = \{j \mid \Gamma_j^A < 0\} = \{j \mid \exists m \in \sigma(\gamma^A) : m \prec j\}$$

The last interpretation corresponding to the extendable set of $\sigma(\gamma^A)$. Using the inhibiting set $\mathcal{I}(\gamma^A)$, construct a weighted hypergraph $\mathcal{H}_{\mathcal{W}}(\gamma^A)$ such that,

$$\mathcal{E}(\gamma^A) = \{e_j \mid j \in \mathcal{I}(\gamma^A)\}$$
$$e_j = \{m \in \mathcal{E}(\mathcal{M}) \mid m \prec j, m \notin \sigma(\gamma^A)\}$$
$$\mathcal{W}(\gamma^A) = \{w_j = -\Gamma_j^A \mid j \in \mathcal{I}(\gamma^A)\}$$
$$\mathcal{N}(\gamma^A) = \bigcup_{j \in \mathcal{I}(\gamma^A)} e_j$$
$$= \{m \in \mathcal{E}(\mathcal{M}) \mid m \notin \sigma(\gamma^A), \exists m' \in \sigma(\gamma^A) : m \succ m'\}$$

The hyperedges $e_j \in \mathcal{E}(\gamma^A)$ are indexed by the inhibiting set. Each hyperedge e_j is the subset of marginal events $e_j \subseteq \mathcal{E}(\mathcal{M})$ that are extendable to j (yet not already included in γ^{A}[8]). The weights $\mathcal{W}(\gamma^A)$ equal the amount that Γ^A needs to recover by, and the nodes $\mathcal{N}(\gamma^A)$ are simply the union of all edges. This weighted hypergraph is termed the **antecedent hypergraph**.

Remark 38. The structure of a hypergraph $\mathcal{H}$ (disregarding weights) is best computationally implemented as a bitwise matrix H where the rows of H are in correspondence with the nodes $n \in \mathcal{N}$ of $\mathcal{H}$. Likewise, the columns of H are in correspondence with the edges $e \in \mathcal{E}$ of $\mathcal{H}$. The entires of H can be populated according to edge inclusion.

$$H_{n,e} = \begin{cases} 1 & n \in e \\ 0 & n \notin e \end{cases}$$

Regarding the antecedent hypergraph $\mathcal{H}(\gamma^A)$, the matrix form $H(\gamma^A)$ of $\mathcal{H}(\gamma^A)$ can be constructed by taking the incidence matrix M and deleting rows not in $\mathcal{N}(\gamma^A)$ and columns not in $\mathcal{E}(\gamma^A)$. As an example, consider theorem 34 and the hypergraph of fig. 2. The corresponding matrix representation of the hypergraph of theorem 34 is a sub-matrix of the the incidence matrix depicted in fig. 1.

$$H(\gamma^A) = \begin{array}{c} \\ m_2 \\ m_4 \\ m_5 \\ m_6 \\ m_{10} \\ m_{11} \\ m_{12} \end{array} \begin{pmatrix} \begin{array}{cccc} j_4 & j_5 & j_6 & j_8 \\ \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & \mathbf{1} \end{array} \end{pmatrix}$$

Although it is possible to generate antecedent hypergraphs in this manner, we recommend against it. One of the advantages of the definite extension procedure (to be discussed in section III C 5) is that typically the size of an antecedent hypergraph H is *far smaller* than the original incidence matrix M. For large enough marginal scenarios, M becomes too resource intensive to hold in memory, while H remains far more manageable.

After generating an antecedent hypergraph $\mathcal{H}(\gamma^A)$, it is possible to find all valid, homogeneous, non-contextuality inequalities $I_{\mathcal{M}} = \{\gamma\mathcal{P}^{\mathcal{M}} \geq 0\}$ by finding all transversals of $\mathcal{H}(\gamma^A)$. As mentioned previously for each transversal $\mathcal{T}$, the marginal multi-set γ^C agreeing with $\mathcal{T}$ can be constructed,

$$\forall m \in \sigma(\mathcal{T}) : \gamma_m^C = \mathcal{T}_m$$

Such that $\gamma^A + \gamma^C$ has definite extension, which by theorem 31 induces the valid inequality,

$$I_{\mathcal{M}} = \{(\gamma^C + \gamma^A)\mathcal{P}^{\mathcal{M}} \geq 0\} \tag{20}$$

More importantly, one only needs to consider *minimal transversals* of $\mathcal{H}(\gamma^A)$. If a transversal $\mathcal{T}$ is non-minimal, then there exists another transversal $\mathcal{T}'$ such that $\mathcal{T}' \preceq \mathcal{T}$. A minimal transversal $\mathcal{T}'$ would have correspondence with a tighter inequality $I'_{\mathcal{M}} = \{(\gamma'^C + \gamma^A)\mathcal{P}^{\mathcal{M}} \geq 0\}$ that is capable of witnessing contextually in a strictly large subset of marginal models $\mathcal{P}^{\mathcal{M}}$.

------

[8] At this point, it might be unclear why we disregard marginal outcomes in $\sigma(\gamma^A)$. Sufficient reasoning will become apparent when presented with a complete discussion of the definite extension procedure.

Definition 39. Let I_1 and I_2 denote two different $\mathcal{NC}$-inequalities.

Definition 40. Let $\mathcal{Y}$ denote the set of all multi-sets γ with positive semi-definite extensions,

$$\mathcal{Y} = \{\gamma \mid \gamma M = \Gamma \succeq 0\}$$

A multi-set γ is said to be **reducible/redundant** if,

$$\exists \gamma^1, \gamma^2 \in \mathcal{Y} : \gamma = \gamma^1 + \gamma^2, \gamma \neq \gamma^1 \neq \gamma^2 \neq 0$$

Todo (TC Fraser): Figure out language and how to delineate the idea of reducible and irreducible extensions
Therefore, it is possible to generate a family of *tight* inequalities for each antecedent choice

4. Completeness

Todo (TC Fraser): FM section would demonstrate that linear, homogeneous, integer inequalities are sufficient
Todo (TC Fraser): Talk about how all facets with antecedent order n would be found by DEP at level n
Todo (TC Fraser): Upper limit of DEP is set by lowest antecedent order of FM result

Theorem 41. *Todo (TC Fraser): main completeness theorem*

5. Advantages

Todo (TC Fraser): Horizontally scalable
Todo (TC Fraser): Low Memory Overhead
Todo (TC Fraser): Effectively depth-first FM
Todo (TC Fraser): Offers partially solution when complete solution intangible
Todo (TC Fraser): Weighted transversal algorithms are natural extensions of regular algorithms
Todo (TC Fraser): inequalities found at lower orders of DEP hierarchy are either facets or are positively separable into one or more facets
Todo (TC Fraser): Symmetries MASSIVELY reduce search space, memory for complete solution (to be discussed)
Todo (TC Fraser): Can be adapted to find correlation inequalities
Todo (TC Fraser): Can be implemented as a relaxation procedure (becomes regular HT problem)
Todo (TC Fraser): Can be adapted to target a specific distribution
Todo (TC Fraser): Seed distributions to avoid trivial inequalities

6. Disadvantages

Todo (TC Fraser): We don't yet know exactly how high the hierarchy goes
Todo (TC Fraser): If complete solution is ultimately desired, FM is in principle far more efficient because it does need to prune anything
Todo (TC Fraser): Trivial equalities show that lower-order transversals are equivalent to higher order transversals (maybe not)

7. Unanswered questions

IV. MARGINAL SYMMETRIES

Exploiting symmetries of a marginal scenario are useful for a few distinct reasons. First, Bancal et. al. [14] discuss computational advantages in considering symmetric versions of marginal polytopes mentioned in section II; the number of extremal points typically grows exponentially in $\mathcal{J}$, but only polynomial for the symmetric polytope. They also note a number of interesting inequalities (such as CHSH [15]) 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. Numerical optimizations against symmetric inequalities lead to one of two interesting cases: either the extremal distribution is symmetric itself or it is not. The latter case generates a family of incompatible distributions obtained by applying symmetry operations on the extremal distribution[9].

In this section we discuss how to exploit symmetries in the definite extension procedure in order to reduce the size of the computational search space. In contrast to [14], we gain advantage in a non-lossy way; symmetries drastically accelerate the rate in which the DEP finds facets of the unsymmetrical polytope.

To demonstrate what types of symmetries we have in mind, we present the following example:

Example 42. First consider the marginal scenario $\mathcal{M} = \{\{a, b, c\}, \{c, d\}, \{a, d\}\}$ where each variable in $\mathcal{J} = \{a, b, c, d\}$ has binary outcomes: $O_a = O_b = O_c = O_d = \{0, 1\}$. Moreover, let $I_\mathcal{M}$ be a $\mathcal{NC}$-inequality,

$$I_\mathcal{M} = \{P_{abc}(010) \leq P_{cd}(00) + P_{ad}(01)\}$$

Intuitively, a the contextuality of $\mathcal{P}^\mathcal{M}$ is independent of the how one labels outcomes. Therefore new $\mathcal{NC}$-inequalities can be generated from $I_\mathcal{M}$ by permuting its outcomes. Let $\omega_v : O_v \to O_v : \{0, 1\} \mapsto \omega_v[\{0, 1\}] = \{1, 0\}$ be a permutation of the outcomes for variable v; i.e. ω_v "flips" the outcomes for v. Let $\omega = \omega_a \times \omega_c$ be the permutation of variables a and c. The action of ω on outcome labels induces an action on the inequality $I_\mathcal{M}$ itself,

$$\omega[I_\mathcal{M}] = \{P_{abc}(\omega[010]) \leq P_{cd}(\omega[00]) + P_{ad}(\omega[01])\}$$
$$= \{P_{abc}(111) \leq P_{cd}(10) + P_{ad}(11)\}$$

The new inequality $\omega[I_\mathcal{M}]$ is also a valid $\mathcal{NC}$-inequality. By design, any permutation ω over outcomes leaves the marginal scenario invariant ($\omega[\mathcal{M}] = \mathcal{M}$) and thus we may write $\omega[I_\mathcal{M}] = \omega[I]_\mathcal{M}$.

Alternatively, permutations of the variables themselves are also permitted. Let $\varphi : \mathcal{J} \to \mathcal{J} : \{a, b, c, d\} \mapsto \varphi[\{a, b, c, d\}] = \{c, a, d, b\}$ be a permutation of the variables $\mathcal{J}$. Intuitively φ acts on $I_\mathcal{M}$ in the following way,

$$\varphi[I_\mathcal{M}] = \{P_{\varphi[abc]}(010) \leq P_{\varphi[cd]}(00) + P_{\varphi[ad]}(01)\}$$
$$= \{P_{cad}(010) \leq P_{db}(00) + P_{cb}(01)\}$$
$$= \{P_{acd}(100) \leq P_{bd}(00) + P_{bc}(10)\}$$

In general (as in this example), permutations over all variables can modify the marginal scenario $\varphi[\mathcal{M}] \neq \mathcal{M}$; in general we write $\varphi[I_\mathcal{M}] = \varphi[I]_{\varphi[\mathcal{M}]}$. Although $\varphi[I_\mathcal{M}]$ is in fact a valid $\mathcal{NC}$-inequality, in constrains marginal models of a different marginal scenario $\varphi[\mathcal{M}]$ and are in general of no interest.

A. Marginal Symmetries

As demonstrated by theorem 42, permutations can be applied to inequalities themselves to generate new inequalities. To formalize this notion, let $\mathsf{Perm}(S)$ denote the **permutation group** acting on the set S consisting of all bijective maps $\pi : S \to S$.

$$\mathsf{Perm}(S) = \{\pi : S \to S\} \tag{21}$$

Together with the $|S| \times |S|$ permutation matrix Λ_π for each permutation π,

$$\forall i, j \in S : \quad \Lambda_{\pi;i,j} \equiv \begin{cases} 1 & \pi(i) = j \\ 0 & \pi(i) \neq j \end{cases}$$

Definition 43. A **marginal symmetry** of $\mathcal{M}$ is a symmetry of the incidence matrix M corresponding to a pair of permutations $(\pi_\mathcal{M} \in \mathsf{Perm}(\mathcal{E}(\mathcal{M})), \pi_\mathcal{J} \in \mathsf{Perm}(\mathcal{E}(\mathcal{J})))$ such that M is invariant under permutation.

$$\Lambda_{\pi_\mathcal{M}} M \Lambda_{\pi_\mathcal{J}} = M$$

———

[9] If the extremal distribution happens to not be symmetric, there is indication that the space of accessible distributions is non-convex.

Identifying the set of all marginal symmetries for a given marginal model is very important; it allows one to generate new inequalities from known ones.

Theorem 44. *For each $\mathcal{NC}$-inequality $I_{\mathcal{M}} = \left\{ \gamma \mathcal{P}^{\mathcal{M}} \geq 0 \right\}$ and marginal symmetry $(\pi_{\mathcal{M}}, \pi_{\mathcal{J}})$, the inequality $\pi_{\mathcal{M}}[I_{\mathcal{M}}]$ defined by,*

$$\pi_{\mathcal{M}}[I_{\mathcal{M}}] \equiv \left\{ \pi_{\mathcal{M}}^{-1}[\gamma] \mathcal{P}^{\mathcal{M}} \geq 0 \right\} \tag{22}$$

is also a valid $\mathcal{NC}$-inequality.

Proof. Since $I_{\mathcal{M}} = \left\{ \gamma \mathcal{P}^{\mathcal{M}} \geq 0 \right\}$ is a $\mathcal{NC}$-inequality, then by theorem 31, $\gamma M = \Gamma \succeq 0$. Also note that $\Lambda_{\pi_{\mathcal{J}}}$ is invertible with $\Lambda_{\pi_{\mathcal{J}}}^{-1} = \Lambda_{\pi_{\mathcal{J}}^{-1}}$. Therefore,

$$\Gamma \Lambda_{\pi_{\mathcal{J}}}^{-1} = \gamma M \Lambda_{\pi_{\mathcal{J}}}^{-1} = \gamma \Lambda_{\pi_{\mathcal{M}}} M$$

But $\Gamma \succeq 0 \implies \Gamma \Lambda_{\pi_{\mathcal{J}}}^{-1} \succeq 0$. Thus,

$$\gamma \Lambda_{\pi_{\mathcal{M}}} M \succeq 0 \tag{23}$$

Interpreting $\gamma \Lambda_{\pi_{\mathcal{M}}}$ as a multi-set for $\mathcal{M}$,

$$\forall m \in \mathcal{E}(\mathcal{M}) : (\gamma \Lambda_{\pi_{\mathcal{M}}})_m = \gamma_{\pi_{\mathcal{M}}^{-1}[m]} \equiv \pi_{\mathcal{M}}^{-1}[\gamma]_m$$

Therefore eq. (23) shows that $\pi_{\mathcal{M}}^{-1}[\gamma]$ has a positive semi-definite extension and thus theorem 31 proves that eq. (22) is also a valid $\mathcal{NC}$-inequality. $\qquad \square$

Remark 45. There are two equivalent interpretations for the permuted inequality $\pi_{\mathcal{M}}[I_{\mathcal{M}}]$ of theorem 44. Notice,

$$\pi_{\mathcal{M}}^{-1}[\gamma] \mathcal{P}^{\mathcal{M}} = \gamma \Lambda_{\pi_{\mathcal{M}}} \mathcal{P}^{\mathcal{M}} = \gamma \pi_{\mathcal{M}}[\mathcal{P}]^{\mathcal{M}}$$

Which means one can either think of $\pi_{\mathcal{M}}[I_{\mathcal{M}}]$ as being $I_{\mathcal{M}}$ with the coefficients rearranged, or with the probabilistic assignments rearranged.

Todo (TC Fraser): Discuss the importance of symmetries to restrict branching in DEP

Diagrammatically, for each marginal event $m \in \mathcal{E}(\mathcal{M})$ and joint event $j \in \mathcal{E}(\mathcal{J})$, the notion of extendability is preserved by $(\pi_{\mathcal{M}}, \pi_{\mathcal{J}})$.

$$
\begin{array}{ccc}
m & \longrightarrow\!\!\!\prec & j \\
{\scriptstyle \pi_{\mathcal{M}}} \Big\downarrow & & \Big\downarrow {\scriptstyle \pi_{\mathcal{J}}} \\
\pi_{\mathcal{M}}[m] & \prec\!\!\!\longrightarrow & \pi_{\mathcal{J}}[j]
\end{array}
$$

Fortunately, every incidence matrix M has a useful **uniqueness property**: there are no duplicated columns or rows in M. This means that in every marginal symmetry $(\pi_{\mathcal{M}}, \pi_{\mathcal{J}})$, $\pi_{\mathcal{J}}$ is uniquely determined by $\pi_{\mathcal{M}}$ and vice versa. Therefore we restrict our attention and discussions to row permutations $\pi_{\mathcal{M}}$, with column permutations being in one-to-one correspondence. Henceforth we will denote to the most general marginal symmetry using π and take it to mean a permutation over marginal events.

B. Outcome Symmetry

Definition 46. The **outcome symmetry group** Ω for the set of joint variables $\mathcal{J} = \{v_1, \dots, v_n\}$ is the direct product of permutations over the outcomes for each variable $v \in \mathcal{J}$,

$$\Omega \equiv \mathsf{Perm}(O_{v_1}) \times \cdots \times \mathsf{Perm}(O_{v_n}) = \prod_{v \in \mathcal{J}} \mathsf{Perm}(O_v)$$

The action of $\omega \in \Omega$ on any event $s \in \mathcal{E}(V)$ (denoted $\omega[s]$) where $V \subseteq \mathcal{J}$ is defined as,

$$\forall v \in \mathcal{D}(s) : (\omega[s])(v) \equiv \omega[s(v)] \tag{24}$$

Theorem 47. *If a marginal event $m \in \mathcal{E}(\mathcal{M})$ is extendable to the joint event $j \in \mathcal{E}(J)$, then for any outcome permutation $\omega \in \Omega$, $\omega[m]$ is extendable to $\omega[j]$.*

$$m \prec j \iff \omega[m] \prec \omega[j] \tag{25}$$

Proof. The proof should be clear by the definition of extendability (theorem 10).

$$m \prec j \iff \forall v \in \mathcal{D}(m) : j(v) = m(v)$$

Every outcome permutation $\omega \in \Omega$ leaves the domain of the event invariant $\mathcal{D}(m) = \mathcal{D}(\omega[m])$. Thus by eq. (24),

$$\forall v \in \mathcal{D}(\omega[m]) : (\omega[j])(v) = \omega[j(v)] = \omega[m(v)] = (\omega[m])(v)$$

Therefore we have that $m \prec j \implies \omega[m] \prec \omega[j]$. The left implication of eq. (25) follows by considering $\omega^{-1} \in \Omega$. $\square$

Corollary 48. *Let M be the incidence matrix for a marginal scenario $\mathcal{M}$ and $\omega \in \Omega$ be an outcome permutation of the joint random variables $\mathcal{J}$. By theorem 47, the action of ω on M leaves M invariant ($\omega[M] = M$),*

$$\omega[M]_{m,j} \equiv M_{\omega[m],\omega[j]} = \delta_{\omega[m] \prec \omega[j]} = \delta_{m \prec j} = M_{m,j} \tag{26}$$

C. Variable Symmetry

In section IV B, permutations of outcomes were introduced as symmetries of the incidence matrix M. However the proof of **??** indicates that any permutation leaving the incidence matrix intact can be used to generate new inequalities.

Following the observations of theorem 42, it is always possible to permute the random variables in a marginal scenario to produce new and valid inequalities. A generic permutation group over a the complete set of variables $\mathcal{J}$ is simply $\mathsf{Perm}(\mathcal{J})$. Unlike elements of Ω, permutations $\varphi \in \mathsf{Perm}(\mathcal{J})$ have the undesired possibility of transforming $\mathcal{M}$ into a different marginal scenario.

$$\varphi[\mathcal{M}] \equiv \{\varphi[V_1], \ldots, \varphi[V_k]\} \neq \mathcal{M}$$

This means that there could exist $V_i \in \mathcal{M}$ such that $\varphi[V_i] \notin \mathcal{M}$. In particular, permutations φ retaining this property are unable to generate $\mathcal{NC}$-inequalities over the same marginal scenario.

Definition 49. The **variable permutation group** Φ for a marginal scenario $\mathcal{M}$ is the joint permutation subgroup that stabilizes the marginal scenario.

$$\Phi \equiv \{\varphi \in \mathsf{Perm}(\mathcal{J}) \mid \forall V \in \mathcal{M} : \varphi[V] \in \mathcal{M}\}$$

The action of $\varphi \in \Phi$ on any outcome $f \in \mathcal{E}(V)$ (denoted $\omega[f]$) is defined as,

$$\forall v \in \mathcal{D}(\varphi[f]) : (\varphi[f])(v) \equiv f(\varphi^{-1}[v]) \tag{27}$$

D. 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 $I_{\mathcal{M}}$, 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 $\mathsf{Perm}(N)$ to be the set of all bijective maps φ from N to N.

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

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

$$\varphi[\mathcal{G}] \equiv (\varphi[\mathcal{N}], \varphi[\mathcal{E}])$$
$$\varphi[\mathcal{N}] \equiv \{\varphi[n] \mid n \in \mathcal{N}\} = \mathcal{N}$$
$$\varphi[\mathcal{E}] \equiv \{\varphi[e] \mid e \in \mathcal{E}\}$$

To motivate a more general treatment of such symmetries, consider the Triangle Scenario of fig. 3. Due to the rotational and reflective symmetries of the Triangle Scenario, any permutation $\varphi \in \mathsf{Perm}(\mathcal{N}_O)$ 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 $I_\mathcal{M}$ 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 $I_\mathcal{M}$ yields $\varphi[I_\mathcal{M}] \equiv \varphi[I]_{\varphi[\mathcal{M}]}$, 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 $I_\mathcal{M}$ to generate new and valid compatibility inequalities. Specifically this alludes to the **graph automorphism group** $\mathsf{Aut}(\mathcal{G})$ of $\mathcal{G}$.

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

In general, $\mathsf{Aut}(\mathcal{G})$ could include elements φ that map latent nodes to observable nodes[10]. 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 50. The **causal symmetry group** for a causal structure $\mathcal{G}$ is the graph automorphism subgroup that stabilizes the observable nodes.

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

Given any compatibility inequality $I_\mathcal{M}$, each element φ of the causal symmetry group defines a new valid compatibility inequality $\varphi[I_\mathcal{M}] = \varphi[I]_{\varphi[\mathcal{M}]}$.

The inflation technique discussed in section VII 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 theorem 57, inequalities I' for $\mathcal{G}'$ are *only* transferable to inequalities I for $\mathcal{G}$ if I' is in terms of distributions over the pre-injectable sets $I'_{\mathsf{PreInj}_\mathcal{G}(\mathcal{G}')}$. As a consequence of this observation, we need to consider a subgroup of the causal symmetry group that preserves the pre-injectable sets,

$$\varphi\big[\mathsf{PreInj}_\mathcal{G}(\mathcal{G}')\big] = \mathsf{PreInj}_\mathcal{G}(\mathcal{G}') \tag{28}$$

Definition 51. The **restricted causal symmetry group** Φ for a causal structure $\mathcal{G}$ and a marginal scenario $\mathcal{M}$ is the graph automorphism subgroup that stabilizes the marginal scenario $\mathcal{M}$.

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

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

The restricted causal symmetry group for the pre-injectable sets of an inflation $\mathcal{G}'$ is the ideal symmetry group for use with the inflation technique. If one obtains a particular inequality $I'_{\mathsf{PreInj}_\mathcal{G}(\mathcal{G}')}$ as well as the restricted causal symmetry group Φ, an entire family of valid compatibility inequalities are also obtained.

$$\left\{\varphi[I']_{\mathsf{PreInj}_\mathcal{G}(\mathcal{G}')} \mid \varphi \in \Phi\right\} \text{ where } \Phi = \mathsf{Aut}_{\mathsf{PreInj}_\mathcal{G}(\mathcal{G}')}(\mathcal{G}')$$

---

[10] 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.

We have obtained Φ for the inflated triangle scenario fig. 4. Φ is an order 48 group with the following 4 generators:

$$
\begin{array}{cccc}
\varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 \\
A_1 \to A_4 & A_1 \to A_1 & A_1 \to C_1 & A_1 \to A_1 \\
A_2 \to A_3 & A_2 \to A_3 & A_2 \to C_2 & A_2 \to A_2 \\
A_3 \to A_2 & A_3 \to A_2 & A_3 \to C_3 & A_3 \to A_3 \\
A_4 \to A_1 & A_4 \to A_4 & A_4 \to C_4 & A_4 \to A_4 \\
B_1 \to B_4 & B_1 \to C_1 & B_1 \to A_1 & B_1 \to B_2 \\
B_2 \to B_3 & B_2 \to C_3 & B_2 \to A_2 & B_2 \to B_1 \\
B_3 \to B_2 & B_3 \to C_2 & B_3 \to A_3 & B_3 \to B_4 \\
B_4 \to B_1 & B_4 \to C_4 & B_4 \to A_4 & B_4 \to B_3 \\
C_1 \to C_4 & C_1 \to B_1 & C_1 \to B_1 & C_1 \to C_3 \\
C_2 \to C_3 & C_2 \to B_3 & C_2 \to B_2 & C_2 \to C_4 \\
C_3 \to C_2 & C_3 \to B_2 & C_3 \to B_3 & C_3 \to C_1 \\
C_4 \to C_1 & C_4 \to B_4 & C_4 \to B_4 & C_4 \to C_2 \\
\end{array}
$$

It is easy to verify that $\varphi_1, \varphi_2, \varphi_3, \varphi_4$ are all automorphisms of the inflated Triangle Scenario $\mathcal{G}'$ of fig. 4. Moreover they stabilize the observable nodes, and the map pre-injectable sets of eq. (35) to the pre-injectable sets. Finally, in the Triangle Scenario we have that,

$$
\Phi = \mathsf{Aut}_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')}(\mathcal{G}') \sim \mathsf{Aut}_{\mathcal{N}_O}(\mathcal{G})
$$

To see this, φ_1 and φ_4 become the identity element $\mathbb{I}$ in $\mathsf{Aut}_{\mathcal{N}_O}(\mathcal{G})$ upon removal of the copy-index, leaving φ_2 to generate reflections and φ_3 to generate rotations.

E. Symmetric Marginal Polytope

Up until this point, we have discussed how new inequalities can be created from known ones by exploiting causal symmetry. Although very useful, we divert our attention to finding *symmetric* inequalities themselves.

$$
\forall \varphi \in \Phi : \varphi[I_{\mathcal{M}}] = I_{\mathcal{M}}
$$

In order to do this, we must first define how elements $\varphi \in \Phi$ act on marginal outcomes $O_{\mathcal{M}}$.

Definition 52. Let $\mathcal{J} = \{v_1, \ldots, v_k\}$ be a collection of random variables and let $\varphi \in \mathsf{Perm}(\mathcal{J})$. $\mathcal{J}$ is said to be **label invariant with respect to** φ if the outcome labels for each variable $v \in \mathcal{J}$ are unchanged under the action of φ.

$$
\forall v \in \mathcal{J} : O_v = O_{\varphi[v]}
$$

Label invariance is a prerequisite for any kind of symmetry amoung random variables.

Definition 53. Let $\mathcal{J} = (v_1, \ldots v_k)$ be a collection of random variables along with a specific outcome $m \in \mathcal{E}(V)$ where $V \subseteq \mathcal{J}$. The action of $\varphi \in \mathsf{Perm}(\mathcal{J})$ on m is a new outcome $\varphi[m] \in \mathcal{E}(\varphi[V])$ over $\varphi[V] \subseteq \mathcal{J}$ defined as follows,

$$
\forall v \in V : (\varphi[m])(\varphi[v]) \equiv m(v)
$$

The idea being that φ modifies m in such a way that permutes its domain while leaving its range invariant.

Example 54. Consider the outcome $m = \{A \mapsto 0, B \mapsto 1, C \mapsto 2\} \in \mathcal{E}(\{A, B, C\})$ where all variables have 4 possible outcomes $O_A = O_B = O_C = \{0, 1, 2, 3\}$ and the permutation $\varphi = \{A \to D, B \to C, C \to B, D \to A\}$. The *action* of φ on m is as follows,

$$
\begin{aligned}
\varphi[m] &= \varphi[\{A \mapsto m(A), B \mapsto m(B), C \mapsto m(C)\}] \\
&= \{\varphi[A] \mapsto m(A), \varphi[B] \mapsto m(B), \varphi[C] \mapsto m(C)\} \\
&= \{D \mapsto m(A), C \mapsto m(B), B \mapsto m(C)\} \\
&= \{D \mapsto 0, C \mapsto 1, B \mapsto 2\} \\
&= \{B \mapsto 2, C \mapsto 1, D \mapsto 0\}
\end{aligned}
$$

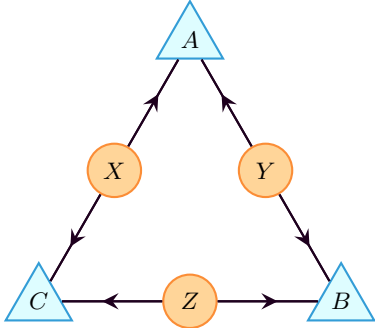Evidently, $\varphi[m] \in \mathcal{E}(\{B, C, D\})$.

FIG. 3. 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.
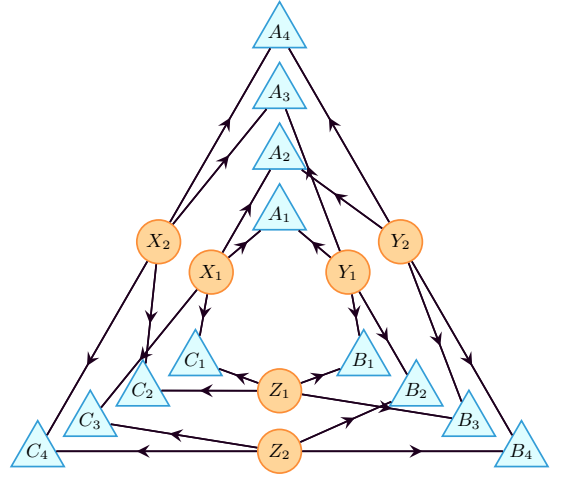

FIG. 4. An inflated causal structure of the Triangle Scenario fig. 3.

Through repeated action of $\varphi \in \Phi$ on marginal outcomes $m \in \mathcal{E}(\mathcal{M})$ and joint outcomes $j \in \mathcal{E}(\mathcal{J})$, one can define group orbits of Φ in $\mathcal{E}(\mathcal{M})$ and $\mathcal{E}(\mathcal{J})$.

$$\mathsf{Orb}_\Phi(m) \equiv \{\varphi[m] \mid \varphi \in \Phi\}$$
$$\mathsf{Orb}_\Phi(j) \equiv \{\varphi[j] \mid \varphi \in \Phi\}$$

Definition 55. The **symmetric incidence matrix** M_Φ for a marginal scenario $\mathcal{M}$ and a particular restricted causal symmetry Φ is a contracted version of the incidence matrix M for $\mathcal{M}$. Each row of M_Φ corresponds to a marginal orbit $\mathsf{Orb}_\Phi(m)$. Analogously each column of M_Φ corresponds to a joint orbit $\mathsf{Orb}_\Phi(j)$. The entries of M_Φ are integers and correspond to summing over the rows and columns of M that belong to each orbit.

$$M_\Phi[\mathsf{Orb}_\Phi(m), \mathsf{Orb}_\Phi(j)] = \sum_{\substack{j' \in \mathsf{Orb}_\Phi(j) \\ m' \in \mathsf{Orb}_\Phi(m)}} M[m', j']$$

F.   Results

Todo (TC Fraser): How we obtained the desired symmetry group Todo (TC Fraser): Group orbits to symmetric marginal description matrix Todo (TC Fraser): representative antecedents completely solve the marginal problem

V.   THE TRIANGLE SCENARIO

The triangle scenario $\mathcal{G}$ is a causal structure composed of 3 parties A, B, C arranged in a triangular configuration while pair-wise sharing latent variables X, Y, Z. It has that has been studied extensively in existing literature (see [16, Fig. 1], [17, Fig. 6], [18, Fig. 8], [19, Fig. 8, Appendix E], [20, Fig. 3], [7, Fig. 1], ...) and is reproduced here in fig. 3 for convenience. At the time of publication for [18] it was noted that characterizing locality in $\mathcal{G}$ remained an open problem and that identifying locality constraints in this configuration seemed challenging. In [19], $\mathcal{G}$ was classified as an "interesting" causal structure.[11] Additionally $\mathcal{G}$ is presented along with a family of entropic inequalities that constrain the set of compatible distributions [19]. Furthermore in [20], Fritz demonstrated that $\mathcal{G}$ is the *smallest* correlation scenario in which their exists quantum incompatible distributions by explicitly mentioning one (which we elect to call the *Fritz distribution* in section VI) and poses an important question asking if other incompatible quantum

---

[11] Indicating that conditional independence relations are not a sufficient characterization of compatibility for $\mathcal{G}$ [19]. Note that in $\mathcal{G}$ there are *zero* conditional independence relations over the observable nodes A, B, C.

distributions exist. Finally in [7], Wolfe et. al. make use of $\mathcal{G}$ along with the Bell Scenario to demonstrate a novel technique for causal inference called the *inflation technique* (summarized in section VII). Wolfe et. al. validate the power and applicability of the inflation technique by proving incompatibility between $\mathcal{G}$ and the W-type distribution; a quantum non-accessible distribution whose incompatibility is not witness-able by *any* entropic inequalities or any other known constraints [7]. Moreover, a family of polynomial, 2-outcome, compatibility inequalities are explicitly derived for an exemplary inflation of $\mathcal{G}$ and 37 non-trivial representatives are printed in [7].

Identifying quantum-accessible distributions that are incompatible with the triangle scenario is of particular importance as it Todo (TC Fraser): Discuss why we are doing this.

As a preliminary search for quantum incompatibility in $\mathcal{G}$, we performed numerical optimizations using bipartite qubit density matrices and 2-outcome POVM measurements (see appendix E for details) against the 37 compatibility inequalities printed in [7] as well as the entropic inequalities presented by [19]. Unfortunately, none of these inequalities could be violated. These early results suggest, although do not prove, that quantum non-locality is not present in $\mathcal{G}$ for *two-outcome* measurements. We make *no claim* that this is a universal truth and simply take it as motivation to explore incompatibility in $\mathcal{G}$ for a larger number of outcomes.

Specifically, we make use of the inflation technique for an inflated triangle scenario $\mathcal{G}'$ depicted in fig. 4 and cast the *causal compatibility* problem as a *marginal problem*. We develop a new technique which we call the *definite extension procedure (DEP)* to find a complete solution to the marginal problem that is suitable for large causal structures where the observable nodes are equipped with numerous outcomes. We use this technique to find numerous compatibility inequalities, some of which witness the incompatibility of the Fritz distribution. Before discussing the DEP, the Fritz distribution will be re-defined explicitly here, followed by a quick summary of the inflation technique of [7]. Following these summaries of existing work, we present the DEP and its connection to Hardy-type paradoxes [7, 11, 21], certificate inequalities and logical Bell inequalities [2]. Moreover, the DEP can be adapted to derive symmetric inequalities or inequalities bounding any subspace of marginal models. To conclude, we subject our found inequalities to non-linear optimizations and discuss the implications of our results.

VI. THE FRITZ DISTRIBUTION

The **Fritz distribution** P_F is a quantum-accessible distribution known to be incompatible with the Triangle Scenario [20]. Explicitly, P_F is a three-party (A, B, C), four-outcome $(1, 2, 3, 4)$ distribution that has form as follows:

$$P_F(000) = P_F(110) = P_F(301) = P_F(211) = P_F(122) = P_F(032) = P_F(233) = P_F(323) = \frac{1}{32}\left(2 + \sqrt{2}\right)$$

$$P_F(010) = P_F(100) = P_F(311) = P_F(201) = P_F(132) = P_F(022) = P_F(223) = P_F(333) = \frac{1}{32}\left(2 - \sqrt{2}\right) \qquad (29)$$

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:

$$\rho_{AB} = |\Psi^+\rangle\langle\Psi^+| \quad \rho_{BC} = \rho_{CA} = |\Phi^+\rangle\langle\Phi^+|$$

$$M_A = \left\{|0\psi_0\rangle\langle 0\psi_0|, |0\psi_\pi\rangle\langle 0\psi_\pi|, |1\psi_{-\pi/2}\rangle\langle 1\psi_{-\pi/2}|, |1\psi_{\pi/2}\rangle\langle 1\psi_{\pi/2}|\right\}$$

$$M_B = \left\{|\psi_{\pi/4}0\rangle\langle\psi_{\pi/4}0|, |\psi_{5\pi/4}0\rangle\langle\psi_{5\pi/4}0|, |\psi_{3\pi/4}1\rangle\langle\psi_{3\pi/4}1|, |\psi_{-\pi/4}1\rangle\langle\psi_{-\pi/4}1|\right\} \qquad (30)$$

$$M_C = \left\{|00\rangle\langle 00|, |01\rangle\langle 01|, |10\rangle\langle 10|, |11\rangle\langle 11|\right\}$$

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

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

Additionally $|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ and $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ are two maximally entangled Bell states.

Fritz first proved it's incompatibility [20] 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 [15],

$$\langle AB|S_A = 1, S_B = 1\rangle + \langle AB|S_A = 1, S_B = 2\rangle + \langle AB|S_A = 2, S_B = 1\rangle - \langle AB|S_A = 2, S_B = 2\rangle \leq 2 \qquad (31)$$

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. (31),

$$\langle AB|C = 1\rangle + \langle AB|C = 2\rangle + \langle AB|C = 3\rangle - \langle AB|C = 0\rangle \leq 2 \qquad (32)$$

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

$$\langle AB \rangle \equiv 2\{P(00) + P(03) + P(30) + P(33) + P(11) + P(12) + P(21) + P(22)\} - 1$$

Which when applied to the Fritz distribution eq. (29) gives the following correlations:

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

$$\langle AB|C=2 \rangle = \langle AB|C=4 \rangle = \frac{\sqrt{2}}{2} \quad \langle AB|C=1 \rangle = -\frac{\sqrt{2}}{2}$$

Which when applied to eq. (32) gives the familiar Tsirelson violation of $2\sqrt{2} \not\leq 2$ [22].

Before continuing it is worth noting that eq. (29) is non-unique. Any distribution that is equal to eq. (29) via a permutation of outcomes or exchange of parties is also referred to as a Fritz distribution. Moreover, the quantum realization of eq. (30) is non-unique. In fact, eq. (30) is not the realization that Fritz originally had in mind. Nonetheless eq. (29) is taken as *the* Fritz distribution for concreteness throughout this paper.

In [20], Fritz posits the following question: *Find an example of non-classical quantum correlations in $\mathcal{G}$ together with a proof of its non-classicality which does not hinge on Bell's Theorem.*

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

VII. SUMMARY OF THE INFLATION TECHNIQUE

The causal inflation technique, first pioneered by Wolfe, Spekkens, and Fritz [7] and inspired by the *do calculus* and *twin networks* of Ref. [23], 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. 3 has copies A_1, A_2, A_3, A_4 in the inflation of fig. 4. Following the notions of Ref. [7], 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.

$$A \sim A_1 \sim A' \not\sim B \sim B_1 \sim B'$$

We preemptively generalize the notion of copy-index equivalence to other mathematical objects like sets, graphs, and groups by saying that $X \sim Y$ if $X = Y$ upon removal of the copy-index. Equipped with the common graph-theoretic terminology and notation of theorem 60, an inflation can be formally defined as follows:

Definition 56. An **inflation** of a causal structure $\mathcal{G}$ is another causal structure $\mathcal{G}'$ such that:

$$\forall n' \in \mathcal{N}' : \mathsf{AnSub}_{\mathcal{G}'}(n') \sim \mathsf{AnSub}_{\mathcal{G}}(n) \tag{33}$$

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 $\mathcal{G}$.

$$\forall n \in \mathcal{N} : P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} = \sum_{\mathcal{N} \backslash n} P_{\mathcal{N}} \tag{34}$$

Now since eq. (33) enforces that $\mathsf{Pa}_{\mathcal{G}'}(n') \sim \mathsf{Pa}_{\mathcal{G}}(n)$, one can create a set of **inflated causal parameters** using eq. (34),

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

Which in turn uniquely defines a compatible joint distribution $P_{\mathcal{N}'}$ on $\mathcal{G}'$[12].

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

---

[12] 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.

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 $\mathcal{G}$.

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

Lemma 57. *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 [7].*

The inflation lemma is the most important result of the causal inflation technique [7]. The contrapositive version of theorem 57 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 [7] proves this explicitly for incompatibility inequalities 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. 3 by considering the marginal problem (see section II) for the set of **maximally pre-injectable sets**[13]. A pre-injectable set V is a subset of $\mathcal{N}'$ that can be decomposed into the disjoint union of injectable sets $V = \coprod_i N_i' \mid N_i' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ that are mutually **ancestrally independent**,

$$\forall i, j : N_i' \perp N_j' \iff \mathsf{An}_{\mathcal{G}'}(N_i') \cap \mathsf{An}_{\mathcal{G}'}(N_j') = \emptyset$$

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 [23],

$$P_V = P_{\cup_i N_i'} = \prod_i P_{N_i'}$$

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.

Focusing on the inflation $\mathcal{G}'$ depicted in fig. 4, we obtained the injectable sets $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ along with the maximally pre-injectable sets through a simple graphical procedure outlined in [7].

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

$$(35)$$

---

[13] 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.

As can be counted, there are 12 maximally pre-injectable sets $\mathcal{G}'$ which will be indexed 1 through 12 in the order seen above ($\{V_1, \ldots, V_{12}\}$). The maximally pre-injectable sets $\{V_i\}_{i \in [12]}$ form a maximal marginal scenario $\mathcal{M}$ over the observable nodes of $\mathcal{G}'$ when equipped with outcome labels. Using theorem 57 inequalities $I_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')}$ constraining the set of contextual marginal models $P^{\mathcal{M}}$ can be deflated into polynomial inequalities over $I_{\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')}$ bounding compatible marginal models over the images of the injectable sets in $\mathcal{G}$.

$$\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}') = \{\{A, B, C\}, \{A\}, \{B\}, \{C\}\}$$

Comment (TC Fraser): This is not all the images. I should clarify this.

VIII. NON-LINEAR OPTIMIZATIONS

Todo (TC Fraser): Mention outlook on te importance of classification and determining interesting inequalities [14] 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 [15] can we viewed as a causal compatibility inequality for the iconic Bell Scenario (Fig. 19 of [24], Fig. 11 of [7], Fig. 1 a) of [25], etc.) corresponding to Bell's notion of local causality [24]. It constrains the set of 2-outcome bipartite distributions over local binary measurement settings for each party $P_{AB|S_A S_B} \equiv \{P_{AB|00}, P_{AB|01}, P_{AB|10}, P_{AB|11}\}$. Numerical optimization *should* obtain the algebraic violation associated with the PR-Box correlations [26]. Maintaining full generality, we simply need to parameterize these 4 distributions using eq. (F2), each requiring 4 real-valued parameters. We define the optimization target for the CHSH inequality to be the left-hand-side of eq. (31),

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

Figure 5 demonstrates this optimization for 5 random seed parameters λ^0, each converging to the expected value of 4. Analogously, [22]

Generic optimization (BFGS, gradient descent) [27]. Basin-hopping [28]

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 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

IX. 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

X. OPEN QUESTIONS & FUTURE WORK

The marginal problem has be well studied in various contexts [1, 7, 29, ...]. Reference [1] contains an excellent summary of its connection to other fields of research including knowledge integration, database theory and coalition games in game theory. Todo (TC Fraser): Lots of stuff
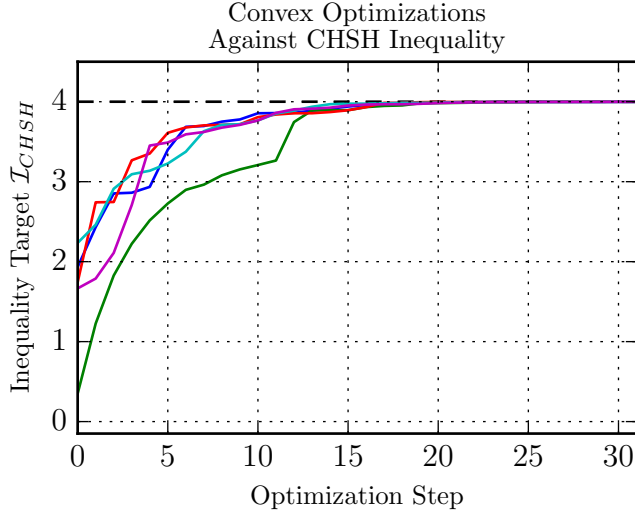
FIG. 5. Convex optimizations against I_{CHSH} recover algebraic violation of 4.
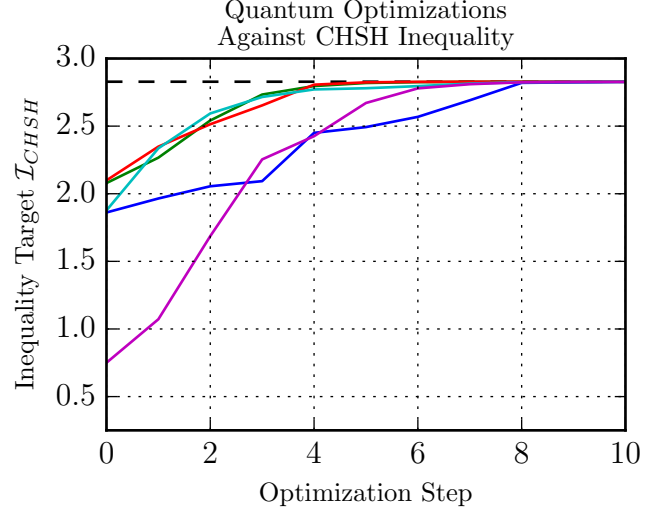
FIG. 6. Quantum optimizations against I_{CHSH} recover maximum violation of $2\sqrt{2}$.

Appendix A: Common Terminology & Notation

1. Graph Theory

Definition 58. A **graph** is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes have the capacity to represent any object while the edges are defined as *pairs* of nodes. Typical, although not necessary, the nodes (edges) are indexed by an index set $S_{\mathcal{N}}$ ($S_{\mathcal{E}}$).

$$\mathcal{N} = \{n_i \mid n_i \in S_{\mathcal{N}}\}$$
$$\mathcal{E} = \{e_i = \{n_j, n_k\} \mid i \in S_{\mathcal{E}}, j, k \in S_{\mathcal{N}}\}$$

Definition 59. A **directed graph** $\mathcal{G}$ is a graph where each edge e has an order or **direction**.

$$\mathcal{N} = \{n_i \mid n_i \in S_{\mathcal{N}}\}$$
$$\mathcal{E} = \{e_i = (n_j \to n_k) \mid i \in S_{\mathcal{E}}, j, k \in S_{\mathcal{N}}\}$$

Definition 60. 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**: $\mathsf{An}_{\mathcal{G}}(n) \equiv \bigcup_{i \in \mathbb{W}} \mathsf{Pa}_{\mathcal{G}}^i(n)$ where $\mathsf{Pa}_{\mathcal{G}}^i(n) \equiv \mathsf{Pa}_{\mathcal{G}}\left(\mathsf{Pa}_{\mathcal{G}}^{i-1}(n)\right)$ and $\mathsf{Pa}_{\mathcal{G}}^0(n) = n$

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**: $\mathsf{An}_{\mathcal{G}}(N) \equiv \bigcup_{n \in N} \mathsf{An}_{\mathcal{G}}(n)$

Definition 61. An **induced subgraph** of $\mathcal{G}$ for a subset of nodes $N \subseteq \mathcal{N}$ is another graph composed of nodes N and all edges e of the original graph that are contained in N.

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

An **ancestral subgraph** of $\mathcal{G}$ for a subset of nodes $N \subseteq \mathcal{N}$ 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 62. A **directed acyclic graph** or **DAG** $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is an directed graph (see theorem 59) with the additional property that no node n is in its set of **ancestors**.

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

It is critical to note the difference in language between *ancestors* from *ancestry*. The ancestry $\mathsf{An}_{\mathcal{G}}(n)$ of a node n is defined in theorem 60 using *whole* numbers $\mathbb{W}$. In contrast, the ancestors of a node are defined similarly with *natural* numbers $\mathbb{N}$.

2. Probability Theory

Definition 63. The **product distribution** of two disjoint probability distributions P_V and P_W (where $V \cap W = \emptyset$) is denoted as $P_V \times P_W$.

$$\forall f \in \mathcal{E}(V \cup W) : (P_V \times P_W)(f) \equiv P_V(f|_V) P_W(f|_W)$$

A product distribution of k mutually disjoint distributions is defined as,

$$\prod_{i \in [k]} P_{V_i} \equiv (P_{V_1} \times \cdots \times P_{V_k})$$

Definition 64. The **marginalization** of a distribution P_V to a distribution over $W \subseteq V$ is denoted $\sum_{V \setminus W} P_V$ and is defined such that,

$$\forall f \in \mathcal{E}(W) : \left(\sum_{V \setminus W} P_V \right)(f) \equiv \sum_{g \in \mathsf{Ext}_V(f)} P_V(g)$$

Appendix B: Exemplary Inequalities

Appendix C: Connections to Sheaf-Theoretic Treatment

Appendix D: Computationally Efficient Parametrization of the Unitary Group

Spengler, Huber and Hiesmayr [30] suggest the parameterization of the unitary group $\mathcal{U}(d)$ using a $d \times d$-matrix of real-valued parameters $\lambda_{n,m}$,

$$U = \left[\prod_{m=1}^{d-1} \left(\prod_{n=m+1}^{d} \exp(iP_n \lambda_{n,m}) \exp(i\sigma_{m,n} \lambda_{m,n}) \right) \right] \left[\prod_{l=1}^{d} \exp(iP_l \lambda_{l,l}) \right] \tag{D1}$$

Where P_l are one-dimensional projective operators,

$$P_l = |l\rangle\langle l| \tag{D2}$$

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 $1 \leq m < n \leq d$. Spengler et. al. proved the validity of eq. (D1) in Ref. [30].

For the sake of reference, let us label the matrix exponential terms in eq. (D1) in a manner that corresponds to their affect on an orthonormal basis $\{|1\rangle, \ldots, |d\rangle\}$.

$$\begin{aligned} GP_l &= \exp(iP_l \lambda_{l,l}) \\ RP_{n,m} &= \exp(iP_n \lambda_{n,m}) \\ R_{m,n} &= \exp(i\sigma_{m,n} \lambda_{m,n}) \end{aligned} \tag{D3}$$

It is possible to remove the reliance on matrix exponential operations in eq. (D1) by utilizing the explicit form of the exponential terms in eq. (D3). As a first step, recognize the defining property of the projective operators eq. (D2),

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

This greatly simplifies the global phase terms GP_l,

$$GP_l = \exp(iP_l\lambda_{l,l}) = \sum_{k=0}^{\infty} \frac{(iP_l\lambda_{l,l})^k}{k!} = \mathbb{I} + \sum_{k=1}^{\infty} \frac{(i\lambda_{l,l})^k}{k!} P_l^k = \mathbb{I} + P_l\left[\sum_{k=1}^{\infty} \frac{(i\lambda_{l,l})^k}{k!}\right] = \mathbb{I} + P_l\left(e^{i\lambda_{l,l}} - 1\right) \quad \text{(D4)}$$

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

$$RP_{n,m} = \cdots = \mathbb{I} + P_n\left(e^{i\lambda_{n,m}} - 1\right) \quad \text{(D5)}$$

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,

$$(|m\rangle\langle n| - |n\rangle\langle m|)^0 = \mathbb{I}$$

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

$$\forall k \in \mathbb{N} : (|m\rangle\langle n| - |n\rangle\langle m|)^{2k+1} = (-1)^k(|m\rangle\langle n| - |n\rangle\langle m|)$$

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

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

$$R_{m,n} = \mathbb{I} + (|m\rangle\langle m| + |n\rangle\langle n|)(\cos\lambda_{n,m} - 1) + (|m\rangle\langle n| - |n\rangle\langle m|)\sin\lambda_{n,m} \quad \text{(D6)}$$

By combining the optimizations of eqs. (D5) to (D4) together we arrive at an equivalent form for eq. (D1) that is computational more efficient.

$$U = \left[\prod_{m=1}^{d-1} \left(\prod_{n=m+1}^{d} RP_{n,m}R_{m,n}\right)\right]\left[\prod_{l=1}^{d} GP_l\right] \quad \text{(D7)}$$

In quantum mechanics, the global phase of a state $|\psi\rangle \in \mathcal{H}^n$ is a *redundant* parameter. Parameterizing unitaries using eq. (D7) 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 [30][14].

$$U_{/GP_l} = \left[\prod_{m=1}^{d-1} \left(\prod_{n=m+1}^{d} RP_{n,m}R_{m,n}\right)\right] \quad \text{(D8)}$$

Todo (TC Fraser): Explanation of Computational Complexity $\mathcal{O}(d^3)$ vs. $\mathcal{O}(1)$ using [31] Todo (TC Fraser): Pre-Caching for Fixed dimension d Todo (TC Fraser): Talk about inverse via haar measure

Appendix E: Parametrization of Quantum States & Measurements

Throughout section VIII, 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 [30, 32–35] with applications Todo (TC Fraser): Finish this sentence. For our purposes, we need to parameterize the space of quantum-accessible distributions P_Q that are *realized* on the Triangle Scenario. We have implemented P_Q under the following description.

$$P_{ABC}(abc) = \text{Tr}[\Pi^\intercal \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi M_{A,a} \otimes M_{B,b} \otimes M_{C,c}] \quad \text{(E1)}$$

———

[14] In our implementation, we accomplish this by explicitly setting $\lambda_{l,l} = 0$ in eq. (D4)

1. Quantum States

The bipartite states $(\rho_{AB}, \rho_{BC}, \rho_{CA})$ of eq. (E1) were taken to be two-qubit density matrices acting on $\mathcal{H}^2 \otimes \mathcal{H}^2$.[15] 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 [35], we can parameterize all such density matrices ρ using **Cholesky Parametrization** [36]. 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 [35].

$$\rho = \frac{T^\dagger T}{\mathrm{Tr}(T^\dagger T)} \quad T = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ \lambda_2 + i\lambda_3 & \lambda_4 & 0 & 0 \\ \lambda_5 + i\lambda_6 & \lambda_7 + i\lambda_8 & \lambda_9 & 0 \\ \lambda_{10} + i\lambda_{11} & \lambda_{12} + i\lambda_{13} & \lambda_{14} + i\lambda_{15} & \lambda_{16} \end{bmatrix} \tag{E2}$$

Our deviation from exclusiving using eq. (E2) is two-fold. First, eq. (E2) 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 [30], they discuss how to parameterization density matrices ρ acting on $\mathcal{H}^d$ of rank k through it's spectral decomposition,

$$\rho = \sum_{i=1}^{k} p_i |\psi_i\rangle\langle\psi_i| \quad p_i \geq 0, \sum_i p_i = 1, k \leq d \tag{E3}$$

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. (E3) 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. (E3) using eq. (F1) and the unitary group $\mathcal{U}(4)$ up to global phase equivalence using eq. (D8). Parameterizing ρ using the Spengler parameterization requires $3 + 12 = 15$ parameters; admitting no degeneracies.

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 [37]. 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. (E1). If each state ρ is decomposable into $|\psi\rangle\langle\psi|$, then eq. (E1) can be written as,

$$P_{ABC}(abc) = \langle\psi_{AB}\psi_{BC}\psi_{CA}|\Pi M_{A,a} \otimes M_{B,b} \otimes M_{C,c}\Pi^\intercal|\psi_{AB}\psi_{BC}\psi_{CA}\rangle \tag{E4}$$

Avoiding the expensive matrix multiplications of eq. (E1).

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 $\{M_i\}_{i=1,\ldots,k}$[16] acting on $\mathcal{H}^d$ summing to the identity,

$$\forall |\phi\rangle \in \mathcal{H}^d : \langle\phi|M_i|\phi\rangle \geq 0 \quad \sum_{i=1}^{k} M_i = \mathbb{I}_{\mathcal{H}^d} \tag{E5}$$

---

[15] 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.

[16] 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 $M_{X,i}$.

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. (E2) without normalizing for trace. Afterwards, M_2 is fully determined by eq. (E5).

$$M_1 = T^\dagger T \quad M_2 = \mathbb{I} - T^\dagger T \tag{E6}$$

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. (E6),

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

Next write a generic $|\phi\rangle \in \mathcal{H}^d$ in terms of a linear combination of the eigenvectors of M_1[17].

$$\langle\phi|M_2|\phi\rangle = \sum_j \left|\Big\langle\phi\Big|m_1^{(j)}\Big\rangle\right|^2 - \sum_i m_1^{(i)}\left|\Big\langle\phi\Big|m_1^{(i)}\Big\rangle\right|^2 = \sum_i \left(1 - m_1^{(i)}\right)\left|\Big\langle\phi\Big|m_1^{(i)}\Big\rangle\right|^2 \tag{E7}$$

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

$$\Big\langle m_1^{(i)}\Big|M_2\Big|m_1^{(i)}\Big\rangle = \left(1 - m_1^{(i)}\right) \geq 0 \implies m_1^{(i)} \leq 1 \tag{E8}$$

By eq. (E7) is not difficult to see that eq. (E8) is a sufficient condition. During optimization, eq. (E8) 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 [38] however a valid parameterization with little to no degeneracy was not found. Upon making this observation, a necessary departure to **projective-valued measures (PVMs)** is warranted[18]. With loss of generality, consider the set of PVMs M satisfying eq. (E5) in addition to orthogonal and projective properties,

$$M_i M_j = \delta_{ij} M_i \quad M_i = |m_i\rangle\langle m_i| \tag{E9}$$

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. (D8). This method was inspired by the *measurement seeding* method of Pál and Vértesi's [40] iterative optimization technique.

Analogously to eq. (E4), projective measurements offer considerable computational advantage as eq. (E1) can be rewritten as,

$$P_{ABC}(abc) = \langle m_{A,a} m_{B,b} m_{C,c}|\Pi^\intercal \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA}\Pi|m_{A,a} m_{B,b} m_{C,c}\rangle \tag{E10}$$

---

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

[18] Strictly speaking, when the number of outcomes (k) *matches* the Hilbert space dimension (d), eq. (E5) implies eq. (E9) 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 [39].
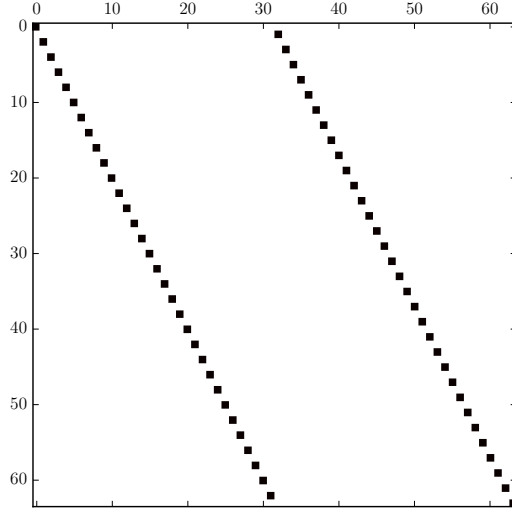
FIG. 7. The network permutation matrix Π for $\left(\mathcal{H}^2\right)^{\otimes 6}$ realized on the triangle scenario. Black represents a value of 1 and 0 otherwise.

3. Network Permutation Matrix

Finally, we introduce the **network permutation matrix** Π for the Triangle Scenario of fig. 3. For bipartite qubit states, Π becomes a 64×64 bit-wise matrix that acts on the measurements M and is depicted in fig. 7. To illuminate its necessity, consider eq. (E1) without Π.

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

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 Π 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 $|q_1\rangle \otimes \cdots \otimes |q_6\rangle = |q_1 q_2 q_3 q_4 q_5 q_6\rangle$ where $\forall i : |q_i\rangle \in \mathcal{H}^2$.

$$
\Pi|q_1 q_2 q_3 q_4 q_5 q_6\rangle = |q_2 q_3 q_4 q_5 q_6 q_1\rangle
$$

Π 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 $\left(\mathcal{H}^2\right)^{\otimes 6}$,

$$
\Pi \equiv \sum_{|q_i\rangle \in \{|0\rangle, |1\rangle\}} |q_2 q_3 q_4 q_5 q_6 q_1\rangle\langle q_1 q_2 q_3 q_4 q_5 q_6|
$$

4. Degeneracy

Todo (TC Fraser): Discuss local unitary degeneracy

Appendix F: Convex Parametrization of Finite Probability Distributions

As discussed in section VIII, 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 $\mathbb{R}^n_{\geq 0}$ [41] 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 n probability values p_j using hyperspherical coordinates [30, 33],

$$p_j = \cos^2 \lambda_j \prod_{i=1}^{j-1} \sin^2 \lambda_i \quad \forall j \in 1, \ldots, n - 1$$

$$p_n = \prod_{i=1}^{n-1} \sin^2 \lambda_i$$

(F1)

Furthermore due to the periodicity of the parameter space λ, eq. (F1) 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[19].

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

(F2)

For various convex optimization tasks sensitive to initial conditions outlined section VIII, the latter parameterization of eq. (F2) generally performed better than the former eq. (F1).

---

[1] Tobias Fritz and Rafael Chaves, "Entropic inequalities and marginal problems," (2011), 10.1109/TIT.2012.2222863, arXiv:1112.4788.
[2] Samson Abramsky and Adam Brandenburger, "The sheaf-theoretic structure of non-locality and contextuality," (2011), 10.1088/1367-2630/13/11/113036, arXiv:1102.0264.
[3] Sébastien Lahaie, "How to take the dual of a linear program," (2008).
[4] Bradley, Hax, and Magnanti, *Applied Mathematical Programming* (Addison-Wesley, 1977) Chap. 4, pp. 143–144.
[5] Erling D. Andersen, Computational Optimization and Applications **20**, 171–183 (2001).
[6] George B Dantzig and B Curtis Eaves, "Fourier-motzkin elimination and its dual," Journal of Combinatorial Theory, Series A **14**, 288–297 (1973).
[7] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz, "The inflation technique for causal inference with latent variables," (2016), arXiv:1609.00672.
[8] Samson Abramsky and Lucien Hardy, "Logical bell inequalities," (2012), 10.1103/PhysRevA.85.062114, arXiv:1203.1352.
[9] Colin Jones, Eric C Kerrigan, and Jan Maciejowski, *Equality set projection: A new algorithm for the projection of polytopes in halfspace representation*, Tech. Rep. (Cambridge University Engineering Dept, 2004).
[10] Alexander Schrijver, *Theory of Linear and Integer Programming* (Wiley, 1998).
[11] 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.
[12] L Mančinska and S Wehner, "A unified view on hardy's paradox and the clauser–horne–shimony–holt inequality," Journal of Physics A: Mathematical and Theoretical **47**, 424027 (2014).
[13] Daniel M. Greenberger, "Bell's theorem without inequalities," Am. J. Phys. **58**, 1131 (1990).
[14] Jean-Daniel Bancal, Nicolas Gisin, and Stefano Pironio, "Looking for symmetric bell inequalities," Journal of Physics A: Mathematical and Theoretical **43**, 385303 (2010).
[15] 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).
[16] Bastian Steudel and Nihat Ay, "Information-theoretic inference of common ancestors," (2010), arXiv:1010.5720.
[17] Rafael Chaves, Lukas Luft, and David Gross, "Causal structures from entropic information: geometry and novel scenarios," New Journal of Physics **16**, 043001 (2014).
[18] Cyril Branciard, Denis Rosset, Nicolas Gisin, and Stefano Pironio, "Bilocal versus nonbilocal correlations in entanglement-swapping experiments," Phys. Rev. A **85** (2012), 10.1103/physreva.85.032119.

---

[19] Strictly speaking, eq. (F2) *also* suffers from non-uniformity; being biased toward uniform probability distributions P_V. Todo (TC Fraser): Discuss rejection sampling simplex algorithms

[19] Joe Henson, Raymond Lal, and Matthew F Pusey, "Theory-independent limits on correlations from generalized bayesian networks," New Journal of Physics **16**, 113043 (2014), arXiv:1405.2572.

[20] Tobias Fritz, "Beyond bell's theorem: Correlation scenarios," (2012), 10.1088/1367-2630/14/10/103001, arXiv:1206.5115.

[21] 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**, 1–39 (2011).

[22] B. S. Cirel'son, "Quantum generalizations of bell's inequality," Lett Math Phys **4**, 93–100 (1980).

[23] Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009).

[24] 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.

[25] Armin Tavakoli, "Quantum correlations in connected multipartite bell experiments," (2015), 10.1088/1751-8113/49/14/145304, arXiv:1509.08491.

[26] D. Rohrlich and S. Popescu, "Nonlocality as an axiom for quantum theory," (1995), arXiv:quant-ph/9508009.

[27] Jorge Nocedal and Stephen Wright, *Numerical Optimization (Springer Series in Operations Research and Financial Engineering)* (Springer, 2000).

[28] David J. Wales and Jonathan P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," J. Phys. Chem. A **101**, 5111–5116 (1997).

[29] N. N. Vorob'ev, "Consistent families of measures and their extensions," Theory of Probability & Its Applications **7**, 147–163 (1962).

[30] 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.

[31] 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).

[32] Dénes Petz and László Ruppert, "Optimal quantum-state tomography with known parameters," (2015), 10.1088/1751-8113/45/8/085306, arXiv:1511.06666.

[33] S. R. Hedemann, "Hyperspherical parameterization of unitary matrices," (2013), arXiv:1303.5904.

[34] Kazuyuki Fujii, Kunio Funahashi, and Takayuki Kobayashi, "Jarlskog's parametrization of unitary matrices and qudit theory," (2005), arXiv:quant-ph/0508006.

[35] 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.

[36] Markus Grasmair, "On the existence of a cholesky factorization," (2014).

[37] 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.

[38] Dénes Petz and László Ruppert, "Optimal quantum-state tomography with known parameters," Journal of Physics A: Mathematical and Theoretical **45**, 085306 (2012).

[39] Asher Peres, *Quantum Theory: Concepts and Methods (Fundamental Theories of Physics)* (Springer, 1995) Chap. 9.6, p. 285.

[40] 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.

[41] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner, "Bell nonlocality," (2013), 10.1103/RevModPhys.86.419, arXiv:1303.2849.