Triangle Scenario Manuscript Title

Thomas C. Fraser*
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada
University of Waterloo, Waterloo, Ontario, Canada
(Dated: August 13, 2016)

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

* tcfraser@tcfraser.com

I. INTRODUCTION

II. DEFINITIONS & NOTATION

Definition 1. Borrowing the notation from [1], each random variable v has a set of all possible outcomes called the **outcome space** or sample space and is denoted O_v. When referencing a *specific* element of O_v or outcome of v, the notation $o[v]$ is used. This notation generalizes to set of random variables $V = \{v_1, \dots, v_{|V|}\}$; the outcome space O_V for a set of random variables V is the Cartesian product of the individual outcome spaces,

$$O_V \equiv O_{v_1} \times \dots \times O_{v_{|V|}}$$

Similarly, a specific outcome of $o[V] \in O_V$ is used to reference a particular collection of outcomes,

$$o[V] \equiv \{o[v_1], o[v_2], \dots, o[v_{|V|}]\}$$

Definition 2. An outcome $o[V] \in O_V$ over the set V is said to be **extendable** to an outcome $o[W] \in O_W$ over the set W if $o[V]$ is contained in $o[W]$:

$$o[V] \subseteq o[W]$$

This also implies the necessary condition that $V \subseteq W$. The idea being that a *less specific* outcome $o[V]$ can be made *more specific* by assigning outcomes to the remaining random variables in $W \setminus V$.

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

$$o[V] \times O_{W \setminus V} = \{o[W] \in O_W \mid o[V] \subseteq o[W]\} \subseteq O_W$$

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

$$o[V] \times O_{W \setminus V} = o[\{a, b\}] \times O_c = \{\{a = 1, b = 2, c = 1\}, \{a = 1, b = 2, c = 2\}, \{a = 1, b = 2, c = 3\}\}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Definition 11. A **weighted hypergraph** $\mathcal{H}_{\mathcal{W}}$ is a regular hypergraph satisfying definition 9 equipped with a set of weights $\mathcal{W}$ ascribed to each node such that a weighted hypergraph is written as a triplet $(\mathcal{W}, \mathcal{N}, \mathcal{E})$.

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

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

Definition 12. A **bounded transversal** of a weighted hypergraph $\mathcal{H}_{\mathcal{W}}$ is a transversal $\mathcal{T}$ of the unweighted hypergraph $\mathcal{H}$ and a real number t (denoted $\mathcal{T}_{\leq t}$) such that the sum of the node weights of the transversal is bounded by t.

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

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

Definition 13. A **causal structure** is simply a DAG with the extra classification of each node into one of two categories; the **latent nodes** and **observed nodes** denoted $\mathcal{N}_L$ and $\mathcal{N}_O$. The latent nodes correspond to random variables that are either hidden through some fundamental process or cannot/will not be measured. The observed nodes are random variables that are measurable. Every node is either latent or observed and no node is both:

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

TODO: How many definitions do I need to write??

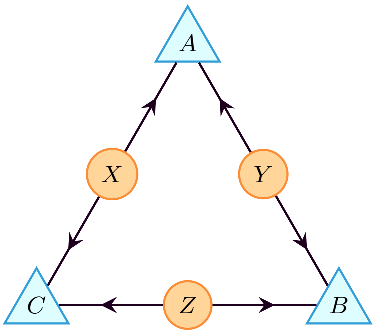

FIG. 1. The casual structure of the triangle scenario. Three variables A, B, C are observable and illustrated as triangles, while X, Y, Z are latent variables illustrated as circles.

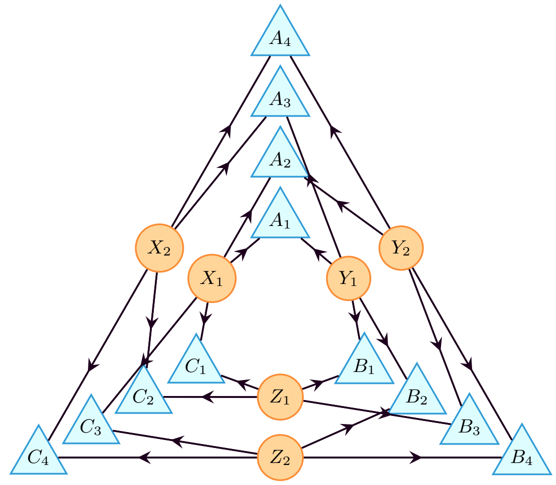

FIG. 2. An inflated causal structure of the triangle scenario fig. 1.

III. SUMMARY OF THE INFLATION TECHNIQUE

The causal inflation technique, first pioneered by Wolfe, Spekkens, and Fritz [2] and inspired by the *do calculus* and *twin networks* of Ref. [3], 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 nodes, called the 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. Equipped with the common graph-theoretic terminology and notation of definition 7, an inflation can be formally defined as follows:

Definition 14. An **inflation** of a causal structure $\mathcal{G}$ is another causal structure $\mathcal{G}'$ such that:

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

Where '$\sim$' is notation for equivalence up to removal of the copy-index. To clarify, each node in an inflated causal structure $n' \in \mathcal{N}'$ shares a *label* assigned to a node $n \in \mathcal{N}$ in the original causal structure together with an additional index called the **copy-index**.

TODO: Clean up what is meant by copy index, example maybe?
TODO: Define injectable sets
TODO: Define pre-injectable sets and then it's connection to probabilities
TODO: Define pre-injectable sets
TODO: State the main Compatibility lemma of inflation

IV. COMPATIBILITY, CONTEXTUALITY AND THE MARGINAL PROBLEM

V. TRIANGLE SCENARIO

Figure 2 in [4]

VI. THE FRITZ DISTRIBUTION

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

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

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:

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

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

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

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

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

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 condition settings in eq. (1),

$$\langle AB|C = 2\rangle + \langle AB|C = 3\rangle + \langle AB|C = 4\rangle - \langle AB|C = 1\rangle \leq 2$$

Finally, specifying the correlation between A and B to be defined in terms of a $\{1, 2, 3, 4\} \to \{(1, 4), (2, 3)\}$ coarse-graining,

$$\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} - \frac{-\sqrt{2}}{2} \leq 2$$

$$2\sqrt{2} \leq 2$$

Which corresponds to the maximum quantum violation of the CHSH inequality eq. (1)

TODO: Discuss non-uniqueness and relabeling

VII. INFEASIBILITY CERTIFICATES

VIII. LOGICAL IMPLICATIONS OF NON-CONTEXTUALITY

Following the definition of contextuality given as Definition 2.3 in [4]

Definition 15. The **pre-injectable marginal distribution vector** denoted b_Π or simply just the **marginal distribution vector** is a list of probability values over outcomes of the pre-injectable sets Π.

Principle assumption: The probability distribution P over the pre-injectable variables admits a compatible joint distribution over the observable inflated variables. **Conclusions:** Given a particular event $A \in$

A. Logical Implications & Inequalities

Following the work conducted by Mansfield and Fritz [7], we consider a possibilistic implications of the $(1, n)$**-type** to be all implications of the form,

$$A \implies C_1 \vee \cdots \vee C_n = \bigvee_{i=1}^{n} C_i \tag{2}$$

Where A and each of the C_i's are events or outcomes of a particular set of variables. The letter 'A' is chosen for the event A since it takes the place of the logical **antecedent** of eq. (2). Likewise, the letter 'C' is chosen to represent logical **consequents**. The implication eq. (2) can be read as *whenever A occurs, at least one of the C_i's also occurs.* A particular scenario where eq. (2) is anticipated to hold true, but is nonetheless violated is called a **Hardy paradox**.

It is possible to turn possibilistic $(1, n)$-type implications into probabilistic inequalities by recognizing that the logical implication of eq. (2) induces the inequality,

$$P(A) \leq P\left(\bigvee_{i=1}^{n} C_i\right) = \sum_{i=1}^{n} P(C_i) - P\left(\bigvee_{\substack{j,k=1 \\ j \neq k}}^{n} [C_j \wedge C_k]\right) \leq \sum_{i=1}^{n} P(C_i) \tag{3}$$

Such that whenever the inequality eq. (3) is violated, the implication in eq. (2) is violated as well. Note that the converse is *not* true; if the inequality eq. (3) holds true, it is still possible for there to be a violation of eq. (2).

Remark 16. Notice that the $P(C \wedge C) \equiv P\left(\bigvee_{\substack{j,k=1 \\ j \neq k}}^{n} [C_j \wedge C_k]\right)$ term in eq. (3) can be read as *the probability that at least two of events in C occurred.* It is omitted from eq. (3) due to it's non-negativity $P(C \wedge C) \geq 0$. However it is possible that $P(C \wedge C)$ vanishes exactly if the elements of C are pair-wise mutually exclusive such that $P(C_i \wedge C_j) = 0, \forall i, j \in 1, \ldots, n$.

B. Tautologies of The Marginal Problem

IX. HARDY TRANSVERSALS

X. DERIVING SYMMETRIC INEQUALITIES

XI. NON-LINEAR OPTIMIZATIONS

XII. RESULTS

XIII. CONCLUSIONS

XIV. OPEN QUESTIONS & FUTURE WORK

Appendix A: Computationally Efficient Parametrization of the Unitary Group

Spengler, Huber and Hiesmayr [8] suggest the parameterization of the unitary group $\mathcal{U}(d)$ using a $d \times d$-matrix of real-valued parameters $\lambda_{n,m}$

$$U = \left[\prod_{m=1}^{d-1} \left(\prod_{n=m+1}^{d} \exp\left(i P_n \lambda_{n,m}\right) \exp\left(i \sigma_{m,n} \lambda_{m,n}\right)\right)\right] \cdot \left[\prod_{l=1}^{d} \exp\left(i P_l \lambda_{l,l}\right)\right] \tag{A1}$$

Where P_l are one-dimensional projective operators,

$$P_l = |l\rangle \langle l| \tag{A2}$$

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 $1 \leq m < n \leq d$. For the sake of reference, let us label the matrix exponential terms in eq. (A1) in a manner that corresponds to their affect on a orthonormal basis $\{|1\rangle, \ldots, |d\rangle\}$.

$$\begin{aligned} GP_l &= \exp\left(iP_l\lambda_{l,l}\right) \\ RP_{n,m} &= \exp\left(iP_n\lambda_{n,m}\right) \\ R_{n,m} &= \exp\left(i\sigma_{m,n}\lambda_{m,n}\right) \end{aligned} \tag{A3}$$

It is possible to remove the reliance on matrix exponential operations in eq. (A1) by utilizing the explicit form of the exponential terms in eq. (A3). As a first step, recognize the defining property of the projective operators eq. (A2),

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

This greatly simplifies the global phase terms GP_l,

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

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

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

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

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

One can verify that the following properties hold,

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

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

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

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

TODO: Explanation of Computational Complexity $\mathcal{O}\left(d^3\right)$ vs. $\mathcal{O}\left(1\right)$ using [9]
TODO: Pre-Caching for Fixed dimension d

Appendix B: Parametrization of Quantum States & Measurements

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

[2] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz, "The inflation technique for causal inference with latent variables," (2016).

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

[4] Tobias Fritz and Rafael Chaves, "Entropic inequalities and marginal problems," (2011), 10.1109/TIT.2012.2222863, arXiv:1112.4788.

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

[6] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, "Proposed experiment to test local hidden-variable theories," Phys. Rev. Lett. **23**, 880–884 (1969).
[7] 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.
[8] 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.
[9] 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).