

# A Method to Derive Polynomial Inequalities for Causal Structures

Tobias Fritz,<sup>1,2,\*</sup> Robert W. Spekkens,<sup>1,†</sup> and Elie Wolfe<sup>1,‡</sup>

<sup>1</sup>*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5*

<sup>2</sup>*Max Planck Institute for Mathematics in the Sciences, Leipzig, Saxony, Germany, 04103*

(Dated: August 4, 2016)

The fundamental task of causal inference is to ascertain which observable probability distributions are compatible with a given causal structure, especially when the structure includes latent variables which cannot be directly observed. While causal compatibility criteria are important in many fields, our interest is motivated by their use in quantum theory, where they distinguish non-classical from classical correlation. Bell inequalities are a type of causal compatibility criteria which apply only to very special causal structures. For more general causal structure one must use different types of causal compatibility criteria, such as testing for conditional independence between observed variables, or by checking the mutual information between variables against entropic upper bounds implied by the causal structure. These more general techniques, however, are often incapable of recognizing uniquely quantum correlations. We here introduce a method for deriving polynomial inequalities constraining compatibility with general causal structures. While this method was originally motivated by desiderata from quantum theory it should nevertheless be valuable in causal inference tasks more broadly.

---

\* [tfritz@perimeterinstitute.ca](mailto:tfritz@perimeterinstitute.ca)

† [rspekkens@perimeterinstitute.ca](mailto:rspekkens@perimeterinstitute.ca)

‡ [ewolfe@perimeterinstitute.ca](mailto:ewolfe@perimeterinstitute.ca)

## I. INTRODUCTION & NOTATION

Given some hypothesis of causal structure it is desirable to determine the set of observable probability distributions compatible with the hypothesis. Causal structure compatibility criteria are leveraged in a wide variety of statistics application, from sussing out biological pathways to enabling artificially intelligent machine learning [? ? ? ?]. The foundational role of causal structure in quantum information theory has only recently been appreciated [? ? ? ?]: classical correlations on a causal structure are those probability distributions compatible with restricting the latent variables to be hidden ontic states; quantum correlations are those which are uniquely realizable if the latent variables in the causal structure are allowed to be quantum states.

Tightly characterizing the set of observable probability distributions compatible with a causal structure is therefore physically critical, in order to recognize and exploit uniquely quantum distributions. Practical techniques for generically constraining causal compatibility include the use of conditional independence relations (easy) [? ? ? ?] and entropic inequalities (more advanced) [? ? ? ?]. These criteria, however, only rarely provide a tight characterization, and frequently fail to ascertain the non-classicality of quantum correlation.

Distinguishing quantum from classical correlations has historically been achieved through to use of Bell inequalities [? ? ? ?]. Bell inequalities however, and their convex hull derivation technique, are limited to causal scenarios involving only one latent common cause variable. New techniques are required to derive quantum-useful causal compatibility criteria for more general causal scenarios [? ? ? ?].

To this end, we introduce a technique applicable to general causal structures for deriving polynomial inequalities constraining observable probability distributions. A causal structure, for the purpose of this article, means a directed acyclic graph (DAG): each node in the DAG corresponds to a random variable, each edge represents a causal influence between variables. Any pair of DAGs which are equivalent under relabelling of latent variables are considered identical causal structures.

Without loss of generality, we choose to consider exclusively **deterministic** DAGs. A deterministic DAG is one where root nodes are genuinely random but all other nodes are presumed to be deterministic functions of their parents. Entropically,  $H(A|\text{pa}[A]) = 0$ , knowing  $\text{pa}[A]$  allows one to predict  $A$  with perfect certainty. If a reader wishes to imagine a causal structure which is nondeterministic (as is typical), the causal structure can be expressed as a deterministic DAG by judicious use of latent variables.

Note that in deterministic DAGs latent variables serve two purposes: Firstly, latent variables with more than one child account for correlations which are mediated by an unknowable common cause. Secondly, latent variables account for why their children are *not* deterministic functions of their observable parents.

If a latent variable has only one child, and moreover the child has no observable parents, then the latent variable is useless. Without loss of generality we therefore choose to consider exclusively DAGs which do not have any such useless latent variables.

A kind of DAG with special importance to use here is a DAG which happens to be in **purely common cause (PCC)** form. A PCC scenario is a two-layer causal structure in which every node is either a “cause” or an “effect”, such that the only edges in the DAG are those which connect cause nodes to effect nodes. All causal pathways in a PCC DAG have length 1, i.e. there is precisely one edge in every connected path between any root node and any terminal node in a PCC DAG.

PCC DAGs are important to use because our technique works by first mapping the target causal structure  $G$  to a PCC representation,  $G' = \text{ReduceToPCC}[G]$  whenever  $G$  is not already in PCC form. For pedagogical clarity we choose to delay discussing the details of the `ReduceToPCC` until ???. We begin, rather, by illustrating sample derivations of polynomial inequalities on a few *innately* PCC scenarios.

We follow the convention that upper-case letters indicate random variables while lower-case letters indicate some particular value associated with the corresponding random variable. In this convention, for example, a student’s score on some exam  $X$  might depend probabilistically on the extent of sleep  $S$ . The Boolean proposition, or event,  $X=x|S=s$  should be understood as “the students scores  $x$  on the exam given a duration of sleep equal to  $s$ ”. As conditional propositions form the basis of much of what follows, we **represent conditioning via subscript notation**, such as  $x_s$  to indicate the event  $X=x|S=s$ .

**NEW CONVENTION: Need to update ALL the figures.** Throughout this article latent random variables are represented by Greek letters  $\Lambda=\lambda, \Gamma=\gamma, \dots$ , whereas observable variables are represented by Roman letters  $A=a, B=b, \dots$  etc. We reserve letters from the *end* of the Roman alphabet to represent variables which are root nodes in the DAG. In graphical depictions we follow the convention of representing latent variables by circles and observable variables by triangles [? ?].

We choose to indicate logical negation by  $\text{Not}[x] \equiv \bar{x}$  such that  $\bar{x}$  references the possibility of *any* outcome other than  $x$ , i.e.  $p(\bar{x}) \equiv p(X \neq x) = 1 - p(X=x)$ . The negation of a conditional event should be interpreted as any other outcomes given the same settings, such that  $p(\overline{x_{sa}}) = p(X \neq x|S=s, A=a)$ .

Note that **logical conjunction is herein represented by default**, such that  $p(xy)$  is the probability of  $X=x$  and  $Y=y$ . Logical conjunction is also implicit among all variables appearing in subscript.

Finally, note that **superscript notation indicates a dummy index**, such that  $x_{s^2a^1}^2$  is shorthand for  $(X=x^{(2)}|S=s^{(2)}, A=a^{(1)})$ .

## II. BELL SCENARIO (NEW)

Consider the causal structure associated to the Bell/CHSH [??] experiment [? (Fig. E#2), ? (Fig. 19), ? (Fig. 1), ? (Fig. 1), ? (Fig. 2b), ? (Fig. 2)], depicted here in ???.  $\{A, B, X, Y\}$  are all observable variables, and  $\Lambda$  is the latent common cause of  $A$  and  $B$ .

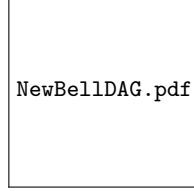


FIG. 1. The causal structure of the Bell scenario. The local outcomes of Alice's and Bob's experimental probing is assumed to be a function of some latent common cause, in addition to their independent local experimental settings.

The assumption of this causal structure dictates that

$$p(a_{\lambda xy}) = p(a_{\lambda x}), \quad p(b_{\lambda xy}) = p(b_{\lambda y}), \quad p(\lambda xy) = p(\lambda)p(x)p(y), \quad (1)$$

and hence

$$p(ab|\lambda xy) = p(a_{\lambda x}b_{\lambda y}), \quad \text{and accordingly} \quad p(ab|xy) = \sum_{\lambda \in \|\Lambda\|} p(a_{\lambda x}b_{\lambda y})p(\lambda). \quad (2)$$

**Proposition 1.** *The Bell causal structure (??) implies that*

$$p(a^1b^1|x^1y^1)p(a^2b^2|x^2y^2) \leq p(\overline{a^3b^3}|x^1y^1) + p(a^1|x^1)p(a^2b^3|x^2y^1) + p(b^1|y^1)p(a^3b^2|x^1y^2)$$

*Proof.* We use the causal structure to consider various counterfactual propositions, which we collect into a logical tautology as follows.

$$\text{And}[a_{x^1\lambda^1}^1, b_{y^1\lambda^1}^1, a_{x^2\lambda^2}^2, b_{y^2\lambda^2}^2] \implies \text{Or} \left[ \begin{array}{l} \text{And}[\overline{a_{x^1\lambda^2}^3}, \overline{b_{y^1\lambda^2}^3}], \\ \text{And}[b_{y^1\lambda^1}^1, a_{x^1\lambda^2}^3, b_{y^2\lambda^2}^2], \\ \text{And}[a_{x^1\lambda^1}^1, a_{x^2\lambda^2}^2, \overline{b_{y^1\lambda^2}^3}] \end{array} \right] \quad (3)$$

Next, we convert the tautology to an inequality via two rules:

1. As the antecedent always implies the consequent, the probability of the antecedent is necessarily less-than-or-equal-to the probability of the consequent. If  $j \implies k$  then  $p(j) \leq p(k)$ .
2. The probability of a disjunction of events is less-than-or-equal-to the sum of the probabilities of the individual events, i.e.  $p(j \vee k) = p(j) + p(k) - p(j, k) \leq p(j) + p(k)$ .

The inequality which corresponds to ?? is

$$p(a_{x^1\lambda^1}^1, b_{y^1\lambda^1}^1)p(a_{x^2\lambda^2}^2, b_{y^2\lambda^2}^2) \leq p(\overline{a_{x^1\lambda^2}^3}, \overline{b_{y^1\lambda^2}^3}) + p(b_{y^1\lambda^1}^1)p(a_{x^1\lambda^2}^3, b_{y^2\lambda^2}^2) + p(a_{x^1\lambda^1}^1)p(a_{x^2\lambda^2}^2, b_{y^1\lambda^2}^3) \quad (4)$$

Note that we have factored terms in ?? according to distinct counterfactual instances of *causes*. This is the key step in deriving polynomial inequalities. The distinct counterfactual instances are perhaps best seen by rewriting ?? in a manner which does not assume a particular causal structure, namely

$$p(a^1b^1|x^1y^1\lambda^1)p(a^2b^2|x^2y^2\lambda^2) \leq p(\overline{a^3b^3}|x^1y^1\lambda^2) + p(a^1|x^1\lambda^1)p(a^2b^3|x^2y^1\lambda^2) + p(b^1|y^1\lambda^1)p(a^3b^2|x^1y^2\lambda^2) \quad (5)$$

which we can marginalize both sides over all  $\lambda^1$  and  $\lambda^2$  to obtain ??.

□

As both ???? can be inferred from ??, we shall include only inequalities written in terms of single-effect counterfactuals in subsequent proofs of polynomial inequalities.

Note that as a general matter, polynomial inequalities can be relaxed into interesting special cases by *disregarding* some outcome possibilities, either by imagining the outcome to be *impossible* or *inevitable*. For example, the special

case  $a^3 \rightarrow \text{False}$  means we imagine  $a^3 \notin \|A\|$ , the ramifications being  $p(\overline{a^3 b^3} | x^1 y^1) \rightarrow p(\overline{b^3} | x^1 y^1)$  and  $p(a^3 b^2 | x^1 y^2) \rightarrow 0$  etc. Similarly, the special case  $\{a^1, b^1\} \rightarrow \text{True}$  reduces ?? into

$$p(a^2 b^2 | x^2 y^2) \leq p(\overline{a^3 b^3} | x^1 y^1) + p(a^2 b^3 | x^2 y^1) + p(a^3 b^2 | x^1 y^2). \quad (6)$$

A further special case of ?? is achieved by selecting  $a^2 \rightarrow a$ ,  $a^3 \rightarrow \bar{a}$ ,  $b^2 \rightarrow b$ ,  $b^3 \rightarrow \bar{b}$ , yielding

$$p(ab | x^2 y^2) \leq p(ab | x^1 y^1) + p(a\bar{b} | x^2 y^1) + p(\bar{a}b | x^1 y^2). \quad (7)$$

We may eliminate negation notation entirely from ?? by noting that  $p(a\bar{b} | xy) = p(a | xy) - p(ab | xy) = p(a | x) - p(ab | xy)$  etc, where the last equality makes use of the causal structure per ??. Hence

**Proposition 2.** *The Bell causal structure (??) implies that*

$$p(ab | x^2 y^2) + p(ab | x^2 y^1) + p(ab | x^1 y^2) \leq p(ab | x^1 y^1) + p(a | x^2) + p(b | y^2)$$

which is simply the Clauser-Horne (CH) inequality [?] for the Bell scenario. The CH inequality is the *unique* Bell inequality (up to permutations) for the Bell scenario if  $\{A, B, X, Y\}$  are all binary, and hence the CH inequality is a necessary and sufficient criterion to ascertain if correlations are compatible with that Bell scenario variant.

### III. TRIANGLE AND S15 SCENARIOS NEED REAL NAME? (NEW)

Consider the causal structure depicted here in ??, which corresponds to “interesting” DAG #15 in Ref. [? ].  $\{A, B, C, X\}$  are all observable variables,  $\{A, B, C\}$  are all pairwise-correlated by a common cause although only the common cause of  $A$  and  $B$  is observable.

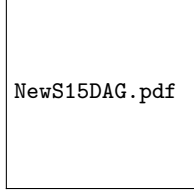


FIG. 2. The causal structure of the S15 scenario.

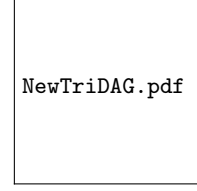


FIG. 3. The causal structure of the Triangle scenario.

The assumption of causal structure per ?? dictates that

$$p(a_{x\lambda\omega}) = p(a_{x\lambda}), p(b_{x\lambda\omega}) = p(b_{x\omega}), p(c_{x\lambda\omega}) = p(c_{\lambda\omega}), \quad \text{and hence} \quad p(abc|x\lambda\omega) = p(a_{x\lambda}b_{x\omega}c_{\lambda\omega}), \quad (8)$$

and accordingly,

$$p(abc|x) = \sum_{\omega \in \|\Omega\|} \sum_{\lambda \in \|\Lambda\|} p(a_{x\lambda}b_{x\omega}c_{\lambda\omega})p(\lambda)p(\omega). \quad (9)$$

The Triangle scenario in ?? is clearly just the S15 scenario under ignorance about the distribution of  $X$ ; i.e.  $X \rightarrow \Delta$ , an observable variable becomes hidden. This mapping is useful to us, as we can derive polynomial causal compatibility criteria for both scenarios simultaneously. Indeed, ???? apply just as well to the Triangle scenario under the substitution  $x \rightarrow \delta$ . The assumption of causal structure per ?? dictates that  $p(abc|\delta\lambda\omega) = p(a_{\delta\lambda}b_{\delta\omega}c_{\lambda\omega})$ , and accordingly that

$$p(abc) = \sum_{\delta \in \|\Delta\|} p(abc|\delta) = \sum_{\delta \in \|\Delta\|} \sum_{\omega \in \|\Omega\|} \sum_{\lambda \in \|\Lambda\|} p(a_{\delta\lambda}b_{\delta\omega}c_{\lambda\omega})p(\lambda)p(\omega)p(\delta). \quad (10)$$

**Proposition 3.** *The S15 causal structure (??) implies that*

$$p(a^1b^1c^1|x^1)p(a^2b^2c^2|x^2) \leq p(c^1)p(\overline{a^3b^3}c^2|x^1) + p(a^2|x^2)p(a^1b^3|x^1) + p(b^2|x^2)p(a^3b^1|x^1)$$

*Proof.*

$$p(a_{x^1\lambda^1}^1b_{x^1\omega^1}^1c_{\omega^1\lambda^1}^1)p(a_{x^2\lambda^2}^2b_{x^2\omega^2}^2c_{\omega^2\lambda^2}^2) \leq \left( \begin{array}{l} p(\overline{a_{x^1\lambda^2}^3b_{x^1\omega^2}^3}c_{\omega^2\lambda^2}^2)p(c_{\omega^1\lambda^1}^1) \\ + p(a_{x^1\lambda^2}^3b_{x^1\omega^1}^1)p(b_{x^2\omega^2}^2) \\ + p(a_{x^1\lambda^1}^1b_{x^1\omega^2}^3)p(a_{x^2\lambda^2}^2) \end{array} \right) \square \quad (11)$$

In turn, and for free, we also have

**Proposition 4.** *The Triangle causal structure (??) implies that*

$$p(a^1b^1c^1)p(a^2b^2c^2) \leq p(c^1)p(\overline{a^3b^3}c^2) + p(a^2)p(a^1b^3) + p(b^2)p(a^3b^1)$$

which follows from ?? by marginalization over  $x$ , which is not observable in the Triangle scenario.

Similarly, we have

**Proposition 5.** *The S15 causal structure (??) implies that*

$$p(a^1 b^1 c^1 | x^1) p(a^2 b^2 c^2 | x^2) p(a^3 b^3 c^3 | x^3) \leq \left( \begin{array}{l} p(\overline{a^4 b^4 c^4} | x^1) p(\overline{a^5 b^5 c^5} | x^2) \\ + p(b^3 | x^3) p(b^2 c^4 | x^2) p(a^1 b^1 c^1 | x^1) \\ + p(c^2) p(b^5 c^3 | x^2) p(a^1 b^1 c^1 | x^1) \\ + p(a^3 | x^3) p(a^1 c^5 | x^1) p(a^2 b^2 c^2 | x^2) \\ + p(c^1) p(a^4 c^3 | x^1) p(a^2 b^2 c^2 | x^2) \\ + p(a^2 | x^2) p(a^1 b^4 | x^1) p(a^3 b^3 c^3 | x^3) \\ + p(b^1 | x^1) p(a^5 b^2 | x^2) p(a^3 b^3 c^3 | x^3) \end{array} \right)$$

*Proof.*

$$p(a_{x^1 \lambda^1}^1 b_{x^1 \omega^1}^1 c_{\omega^1 \lambda^1}^1) p(a_{x^2 \lambda^2}^2 b_{x^2 \omega^2}^2 c_{\omega^2 \lambda^2}^2) p(a_{x^3 \lambda^3}^3 b_{x^3 \omega^3}^3 c_{\omega^3 \lambda^3}^3) \leq \left( \begin{array}{l} p(\overline{a_{x^1 \lambda^3}^4 b_{x^1 \omega^2}^4 c_{\omega^2 \lambda^3}^4}) p(\overline{a_{x^2 \lambda^1}^5 b_{x^2 \omega^3}^5 c_{\omega^3 \lambda^1}^5}) \\ + p(a_{x^1 \lambda^1}^1 b_{x^1 \omega^2}^2) p(a_{x^3 \lambda^3}^3 b_{x^3 \omega^3}^3 c_{\omega^3 \lambda^3}^3) p(a_{x^2 \lambda^2}^2) \\ + p(a_{x^1 \lambda^1}^1 c_{\omega^3 \lambda^1}^3) p(a_{x^2 \lambda^2}^2 b_{x^2 \omega^2}^2 c_{\omega^2 \lambda^2}^2) p(a_{x^3 \lambda^3}^3) \\ + p(a_{x^2 \lambda^1}^2 b_{x^2 \omega^2}^2) p(a_{x^3 \lambda^3}^3 b_{x^3 \omega^3}^3 c_{\omega^3 \lambda^3}^3) p(b_{x^1 \omega^1}^1) \\ + p(b_{x^2 \omega^2}^2 c_{\omega^3 \lambda^3}^3) p(a_{x^1 \lambda^1}^1 b_{x^1 \omega^1}^1 c_{\omega^1 \lambda^1}^1) p(b_{x^3 \omega^3}^3) \\ + p(a_{x^1 \lambda^3}^4 c_{\omega^3 \lambda^3}^3) p(a_{x^2 \lambda^2}^2 b_{x^2 \omega^2}^2 c_{\omega^2 \lambda^2}^2) p(c_{\omega^1 \lambda^1}^1) \\ + p(b_{x^2 \omega^3}^3 c_{\omega^3 \lambda^3}^3) p(a_{x^1 \lambda^1}^1 b_{x^1 \omega^1}^1 c_{\omega^1 \lambda^1}^1) p(c_{\omega^2 \lambda^2}^2) \end{array} \right) \quad \square \quad (12)$$

Again, in turn, and for free, we also have

**Proposition 6.** *The Triangle causal structure (??) implies that*

$$p(a^1 b^1 c^1) p(a^2 b^2 c^2) p(a^3 b^3 c^3) \leq \left( \begin{array}{l} p(\overline{a^4 b^4 c^4}) p(\overline{a^5 b^5 c^5}) \\ + p(b^3) p(b^2 c^4) p(a^1 b^1 c^1) \\ + p(c^2) p(b^5 c^3) p(a^1 b^1 c^1) \\ + p(a^3) p(a^1 c^5) p(a^2 b^2 c^2) \\ + p(c^1) p(a^4 c^3) p(a^2 b^2 c^2) \\ + p(a^2) p(a^1 b^4) p(a^3 b^3 c^3) \\ + p(b^1) p(a^5 b^2) p(a^3 b^3 c^3) \end{array} \right)$$

which again follows by marginalization over  $x$ , this time via from ??.

A special case of ?? is

**Proposition 7.** *The Triangle causal structure (??) implies that*

$$p(a)p(b)p(c) \leq p(\bar{a}, \bar{b}, \bar{c}) + p(a)p(b, c) + p(b)p(c, a) + p(c)p(a, b).$$

*Proof.* First, let  $\{a^5, b^5, c^5\} \rightarrow \text{False}$  and  $\{a^2, a^3, b^1, b^3, c^1, c^2\} \rightarrow \text{True}$ . This yields

$$p(a^1) p(b^2) p(c^3) \leq p(\overline{a^4 b^4 c^4}) + p(b^2 c^4) p(a^1) + p(a^4 c^3) p(b^2) + p(a^1 b^4) p(c^3) \quad (13)$$

which reduces to ?? by further substituting  $a^4 \rightarrow a^1 \rightarrow a$ ,  $b^4 \rightarrow b^2 \rightarrow b$ , and  $c^4 \rightarrow c^3 \rightarrow c$ .  $\square$

A consequence of ?? is that the W-type distribution

$$p_W(a, b, c) = \begin{cases} \frac{1}{3} & \text{if } a + b + c = 1 \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

is found to be incompatible with the Triangle scenario, where  $a, b, c \in \{0, 1\}$ . The W-distribution states that the in any event in which  $A, B, C$  are observed, precisely one of them will be found to equal 1 while the other two will equal 0. The identity of the variable which takes the value 1 is uniformly random. In informal but intuitive notation, the W-type distribution is  $1/3[100] + 1/3[010] + 1/3[001]$ . To see how this distribution is incompatible with ??, note that for three *identically distributed* (but not independent) binary variables a further special case of ?? is

$$p(A=1)^3 \leq p(A=B=C=0) + 3 \times p(A=1, B=1) p(C=1).$$

For the W-distribution  $p(A=B=C=0) = 0$ , and also  $p(A=1, B=1) = 0$ , yet  $p(A=1) = 1/3$ . As  $(1/3)^3 \not\leq 0$  we have proven that the W-type distribution is incompatible with the Triangle scenario.

#### IV. FAILURE OF ENTROPIC INEQUALITIES SWITCH NOTATION

It is interesting to note that entropic inequalities [10] fail to recognize the W-type distribution as incompatible with the Triangle scenario, the polynomial inequalities derivable through our method are capable of doing so. To reiterate from the abstract, enumeration of entropic inequalities is considered state-of-the-art derivation of necessary albeit insufficient causal structure compatibility criteria [10]. The insufficiency is a pressing concern in quantum information theory as there are uniquely-quantum distributions which cannot be certified as non-classical by means of entropic inequalities, for instance in the Triangle scenario [10, Prob. 2.17]. Many interesting quantum correlations are identified by entropic inequalities, however [10].

The entropic inequalities associated with the Triangle scenario are given by

$$\begin{aligned}
 & I(A : B) + I(A : C) \leq H(A) \\
 \text{and} \quad & I(A : B) + I(A : C) + I(B : C) \leq H(A, B) - I(A : B : C) \\
 \text{and} \quad & I(A : B) + I(A : C) + I(B : C) \leq \frac{H(A) + H(B) + H(C)}{2} - I(A : B : C)
 \end{aligned} \tag{15}$$

and their permutations [10].

Note that bipartite mutual information may be understood as  $I(A : B) \equiv H(A) + H(B) - H(A, B)$  and tripartite mutual information is defined as  $I(A, B, C) \equiv H(A) + H(B) + H(C) - H(A, B) - H(A, C) - H(B, C) + H(A, B, C)$ . It is straightforward to demonstrate that the W-type distribution given in [10] satisfy [10].



## V. DERIVING INEQUALITIES ALGORITHMICALLY

As evidenced, we derive polynomial inequalities in the observable variables by constructing logical tautologies (and their associated inequalities) in terms of conditional propositions. From the previous examples it should be clear why the tautologies make use of PCC format. Therefore, the preliminary step for deriving polynomial inequalities from  $G$  is **ReduceToPCC** if the scenario isn't already innately PCC.

In algorithmically deriving polynomial inequalities we elect to sacrifice some generality in exchange for simplicity. In particular, we herein limit our consideration to tautologies of the “supplemented excluded middle” (SEM) form. An excluded-middle (EM) tautology follows the following format:

$$\text{Or}[ Q1, Q2, Q3, \dots, \text{And}[ \overline{Q1}, \overline{Q2}, \overline{Q3}, \dots ] ] \quad (16)$$

An SEM tautology intersperses an EM tautology with a bunch of “known” proposition, such as the following:

$$\text{And}[ P1, P2, P3, \dots ] \implies \text{Or} \left[ \begin{array}{l} \text{And}[ Q1, P2, P3, \dots ] \\ \text{And}[ P1, Q2, P3, \dots ] \\ \text{And}[ P1, P2, Q3, \dots ] \\ \text{And}[ \overline{Q1}, \overline{Q2}, \overline{Q3}, \dots ] \end{array} \right] \quad (17)$$

The inequalities in terms of conditional events (such as ??????) must satisfy the following requirements in order to correspond to SEM tautologies and still admit marginalization over the latent variables:

- Every event on the right hand side of the inequality but not on the left hand side should appear negated in the final term on the left hand side. Such events are the “unknowns” in the supplemented tautology of the excluded middle. We call the final term on the right hand side the “closure” term, as it is closes the excluded middle tautology.
- Every “unknown” event (i.e. every event on the right hand side of the inequality but not on the left hand side) should appear in precisely *one* non-closure term on the left hand side. No two such events are allowed in any single term, aside from the closure term, as this would (typically) break the tautology.<sup>1</sup>
- In every term, on both sides of the inequality, all instances of a latent variable with common dummy index must be restricted to the same joint probability. This is necessary in order to marginalize over the latent variables pursuant to the causal structure.

While not strictly necessary, in order to make the inequality as compelling as possible one should ensure that every event on the left hand side should also be referenced on the right hand side. If one finds an inequality lacking this property, one can tighten it by using the special in which all such events are set to **True**. Such choice substitution increases the probability of the left hand side without affecting the right hand side.

Note that these polynomial inequalities are not related whatsoever to the polynomial Bell inequalities recently introduced by ? ] and ? ], nor to the interventional inequalities of ? ] and ? ].

---

<sup>1</sup> Any tautology of the form  $\text{And}[ \dots ] \implies \text{Or}[ \text{And}[ \dots ], \text{And}[ \dots ], \dots ]$  corresponds to a legitimate inequality. The SEM tautology is only the easiest to construct, but others are also possible.

## VI. REDUCTION TO PURELY COMMON CAUSE (NEW)

Reducing a DAG  $G$  to purely common cause form consists of two transformations in succession.

1. Every causal pathway is replaced with a direct causal connection. This transformation adds edges to  $G$  until every root node is directly connection to every one of its causal descendants.  
 $A$ -directed-path-to- $B$  becomes  $A \rightarrow B$  whenever  $A \in \text{RootVars}[G]$ .
2. Edges which do not initiate from root nodes are purged. This transformation deletes edges from  $G$  until the only remaining edges are those connected nodes from the “causes” layer to nodes in the “effects” layer.  
 $U \rightarrow V$  becomes  $U \nrightarrow V$  whenever  $U \notin \text{RootVars}[G]$ .

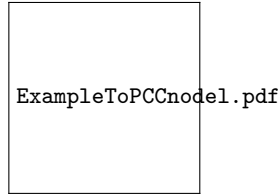


FIG. 4. An example of forcing a DAG into purely common cause (PCC) form. The DAG is relaxed under this transformation. Note the addition of edges initiating from  $X$  and  $\Lambda$ , and the delete of an edge between the two non-root nodes  $A \rightarrow B$ .

We will shortly prove that reducing to PCC is generally a relaxation of  $G$ ; any causal compatibility criteria applicable to  $G' = \text{ReduceToPCC}[G]$  is therefore also applicable to  $G$ .

Many observationally equivalent causal structures can have the same PCC reduction, for example  $G \neq H$  but  $\text{ReduceToPCC}[G] = \text{ReduceToPCC}[H]$ . If two structures have the same PCC reduction we call them **equivalent under PCC reduction**. If a structure is observationally equivalent to its PCC reduction then we deem that structure *observationally invariant under PCC reduction*, or **PCC-lossless** for short. Determining if a DAG is PCC-lossless is addressed in ??.

Mapping genuine causal structures into their PCC reductions has previously been recognized as an effective technique in causal inference, see Refs. [?, Thm. 2.4] and [?, Sec. 5]. Our inequalities apply to the PCC reductions, and hence only indirectly to their generating causal structures. Causal structures which are isomorphic to their PCC reductions, including the Triangle scenario, have been studied at length by ?; the connection to more general causal structures is further elucidated in Ref. [?].

Innately PCC causal structures where the root nodes are all latent variables are known as “correlation scenarios” [?]. Quantum physicists may wish to note the following result [?, Thm. 3.8]: Correlation scenarios which are star graphs [?, Fig. 6] do not admit uniquely-quantum (non-classical) probability distributions. That is, replacing classical latent variables with quantum states does not expand the set of possibly-observable distributions for such correlation structures. Consequently, all general causal structures which are star graphs under PCC reduction also do not admit quantum probability distributions **no longer 100% obvious**. The classification of causal structures which admit non-classical probability distributions has been addressed in further detail by ?].

## VII. RECOGNIZING OBSERVATIONALLY EQUIVALENT DAGS

Notes to self: Discuss add-edge theorem, then subtract edge theorem, then why the second step of ReduceToPCC is observationally invariant, then give the criterion for PCC-lossless, then comment about matching-up latent variables between causal structures.

We intuitively expect that an edge  $A \rightarrow B$  can be added to DAG  $G$  while leaving  $G$  observationally invariant if the new connection does introduce any new information about observable variables to  $B$ . If the added connection does inform  $B$  about some observable  $C$ , that's still ok so long as if the new informational cannot be exploited to increase the correlation between  $B$  and  $C$ . We formalize this notion now.

**Theorem 1.** *Let  $\mathbf{O}$  be the set of observable nodes in a DAG  $G$ . Let  $\mathbf{X} = ( X \text{ s.t. } \text{pa}[B] \in \text{SS}[A|X] )$  be the set of nodes in  $G$  relative to which  $\text{pa}[B]$  comprise a sufficient statistic for node  $A$ . Let  $\mathbf{Y} = \text{UC}_B[G]$  be the set of nodes in  $G$  which have potentially unlimited correlation with  $B$ . Then, an edge  $A \rightarrow B$  can be added to DAG  $G$  while leaving  $G$  observationally invariant if and only if (prior to adding the edge)  $\mathbf{O} \subseteq \mathbf{X} \cup \mathbf{Y}$ .*

Let's define the concepts in ??, and make them operationally useful.  $Z$  is (or "are") a sufficient statistic for  $A$  relative to  $X$ , hereafter  $Z \in \text{SS}[A|X]$  if all inferences about  $X$  knowing  $A$  are also possible without knowing  $A$  but with knowing  $Z$  instead. In other words, learning  $A$  can never teach anything new about  $X$  if  $Z$  is already known.

$Z \in \text{SS}[A|X]$  if  $A$  or  $X$  are completely predictable given  $Z$ , or if  $Z$  contains a Markov Blanket (MB) for  $V$  such that  $A \in V$  while  $X \notin V$ , which screens off  $X$  from  $A$ . These need further definition.

In a deterministic DAG,  $V$  is perfectly predictable given its parents, that is,  $\text{pa}[V] \in \text{SS}[V|\text{anything}]$ . Moreover, the parents of a latent variable are perfectly predictable "given" the latent variable, that is,  $\Lambda \in \text{SS}[\text{pa}[\Lambda]|\text{anything}]$ .

By composing these two rules we can identify the complete set of nodes which are perfectly predicable from  $Z$  alone. **do so!**

It is convenient to assume that all latent variables are parentless in the DAG. (This can actually be done without loss of generality; a DAG with non-root latent nodes can be transformed into an observationally equivalent one by rerouting all parents of the latent node directly to all its children. **proof...**) In a latent-parentless deterministic DAG the node  $V$  is completely predictable from  $Z$  alone if and only if  $V$  is an *exclusive descendant* of  $Z$ .  $V$  is an exclusive descendant of  $Z$  if the intersection of [the ancestors of  $V$ ] with [the non-ancestors of  $Z$ ] is a subset of [the descendants of  $Z$ ].

The Markov Blanket for  $V$ ,  $\text{MB}[V]$  is the set of all of  $V$ 's children, parents, and co-parents. The Markov Blanket is so defined because  $V$  is conditionally independent *everything* given  $\text{MB}[V]$ . For our purposes,  $\text{MB}[V] \in \text{SS}[V|\text{anything-but-}V]$ .

Nodes  $A$  and  $B$  have potentially unlimited correlation in a DAG if and only if  $A \rightsquigarrow B$  (and  $A$  has at least one latent parent), or  $A \rightsquigarrow B$  (and  $B$  has at least one latent parent), or the exists a latent  $\Lambda$  such that  $\Lambda \rightsquigarrow A$  and  $\Lambda \rightsquigarrow B$ . The notation  $A \rightsquigarrow B$  denotes the existence of a directed path from node  $A$  to node  $B$ , where any intermediate nodes are latent [? ].

Using these definitions, we can now consider special cases of ??.

**Corollary 2.** *An edge  $A \rightarrow B$  can be added to some DAG  $G$  while leaving  $G$  observationally invariant if  $\text{pa}[B]$  are a sufficient statistic for  $A$  relative to all observable nodes. As such, the edge  $A \rightarrow B$  can always be added whenever  $A$  is perfectly predictable given  $\text{pa}[B]$ . Furthermore, the edge  $\Lambda \rightarrow B$  can be also always be added when  $\Lambda$  is latent if  $\text{pa}[B]$  contains a Markov Blanket (MB) for some node-set  $V$  such that  $\Lambda \in V$  while no observable variables are inside  $V$ .*

We can also define an analogous condition for when an edge can be removed from a DAG without impacting it observationally.

**Theorem 3.** *An edge  $A \rightarrow B$  can be dropped from DAG  $G$  to form  $G'$  such that  $G$  and  $G'$  are observationally equivalent if and only if the edge  $A \rightarrow B$  can be added (back) to  $G'$  while leaving  $G'$  observationally invariant per ??.*

Recall now the two steps of the transformation ReduceToPCC. Imagine after the first step is finished, that an edge  $A \rightarrow B$  remains, where  $A$  is not a root node. Have directly connected all causal pathways, we know that  $\text{pa}[A] \subseteq \text{pa}[B]$ . As such,  $\text{pa}[B] \in \text{SS}[A|\text{anything}]$ , that is to say,  $A$  is perfectly predictable given the parents of  $B$ . By ??, therefore, if the edge  $A \rightarrow B$  were not in the DAG, we would be able to observationally invariantly add it. By ??, therefore, removing the edge does not observational impact the DAG. Indeed, the second step of ReduceToPCC leaves the post-first-step DAG observationally invariant. This allows us to quickly determine if a DAG is PCC-lossless.

**Proposition 8.** *A DAG  $G$  is PCC-lossless if every new edge in  $\text{ReduceToPCC}[G]$  relative to  $G$  can be accounted for by adding edges to  $G$  while leaving  $G$  observationally invariant, pursuant to ??.*

## ACKNOWLEDGMENTS

Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation.

## Appendix A: Tobias's Original 7 Inequalities

“I present several inequalities... together with a method of proof which has a combinatorial flavour. No quantum violations of any of these inequalities has been found to date.

In the following the complement of a value is marked by an empty circle accent, so  $\overset{\circ}{b}$  means “anything but  $b$ ”, and accordingly  $P(\overset{\circ}{a}\overset{\circ}{b}) = P(A \neq a, B \neq b)$ .

**Theorem 4.** *The following inequalities hold for all classical correlations in the triangle scenario:*

- (a)  $P(a)P(c) \leq P(ab) + P(\overset{\circ}{b}c)$
- (b)  $P(ab\overset{\circ}{c})P(\overset{\circ}{a}b\overset{\circ}{c})P(\overset{\circ}{a}bc) \leq P(abc) + P(\overset{\circ}{a}\overset{\circ}{b})P(ab\overset{\circ}{c})P(a) + P(\overset{\circ}{a}\overset{\circ}{c})P(\overset{\circ}{a}b\overset{\circ}{c})P(c) + P(\overset{\circ}{b}\overset{\circ}{c})P(\overset{\circ}{a}bc)P(b)$
- (c)  $P(ab\overset{\circ}{c})P(\overset{\circ}{a}b\overset{\circ}{c})P(\overset{\circ}{a}bc) \leq P(abc)^2 + 2\left(P(\overset{\circ}{a}\overset{\circ}{b})P(ab\overset{\circ}{c}) + P(\overset{\circ}{a}\overset{\circ}{c})P(\overset{\circ}{a}b\overset{\circ}{c}) + P(\overset{\circ}{b}\overset{\circ}{c})P(\overset{\circ}{a}bc)\right)$
- (d)  $P(abc)^2P(\overset{\circ}{a}\overset{\circ}{b}\overset{\circ}{c}) \leq P(ab\overset{\circ}{c})P(\overset{\circ}{a}b\overset{\circ}{c})P(\overset{\circ}{a}bc) + \left(2P(abc) + P(\overset{\circ}{a}\overset{\circ}{b}\overset{\circ}{c})\right)\left(1 - P(abc) - P(\overset{\circ}{a}\overset{\circ}{b}\overset{\circ}{c})\right)$
- (e)  $P(abc)^2P(\overset{\circ}{a}\overset{\circ}{b}\overset{\circ}{c}) \leq P(abc)^3 + \left(2P(abc) + P(\overset{\circ}{a}\overset{\circ}{b}\overset{\circ}{c})\right)\left(1 - P(abc) - P(\overset{\circ}{a}\overset{\circ}{b}\overset{\circ}{c})\right)$
- (f)  $P(a)P(b)P(c) \leq P(\overset{\circ}{a}\overset{\circ}{b}\overset{\circ}{c}) + P(ab)P(c) + P(ac)P(b) + P(bc)P(a)$
- (g)  $P(a)P(b)P(c) \leq P(\overset{\circ}{a}\overset{\circ}{b}\overset{\circ}{c})^2 + 2\left(P(ab)P(c) + P(ac)P(b) + P(bc)P(a)\right)$

It is quite likely that some of these inequalities are dominated by the others, but I do not know for sure whether any of them are actually redundant.”

Note that ?? implies inequalities (a), (b), and (f). I haven't checked the others yet. ~EW