Restricted Hidden Cardinality Constraints in Causal Models

Beata Zjawin

Perimeter Institute for Theoretical Physics 31 Caroline St. N, Waterloo, Ontario, Canada, N2L 2Y5

International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-308 Gdańsk, Poland

`beata.zjawin@phdstud.ug.edu.pl`

Elie Wolfe   Robert W. Spekkens

Perimeter Institute for Theoretical Physics 31 Caroline St. N, Waterloo, Ontario, Canada, N2L 2Y5

Causal models with unobserved variables impose nontrivial constraints on the distributions over the observed variables. When a common cause of two variables is unobserved, it is impossible to uncover the causal relation between them without making additional assumptions about the model. In this work, we consider causal models with a promise that unobserved variables have known cardinalities. We derive inequality constraints implied by d-separation in such models. Moreover, we explore the possibility of leveraging this result to study causal influence in models that involve quantum systems.

1   Introduction

In causal studies, systems of variables are described by *causal models* [18, 22], which are composed of two elements: (i) the graphical representation of relationships between variables in a model, encoded in a directed acyclic graph, and (ii) the mathematical description of conditional probability distribution of each variable given its causal parents. When a causal model involves *hidden* (i.e., unobserved) variables, any characterization of the model verifiable by observations should only include observed variables. Therefore, one of the objectives of causal inference is to eliminate all hidden variables from inequalities and equalities that describe the model. In principle, this can be achieved using the Tarski–Seidenberg quantifier elimination method [12]. However, its complexity is such that only models with few variables can be solved using this technique, hence the reason for the many attempts to simplify the problem. For example, one can derive a set of incomplete constraints on probability distributions in latent causal models [8, 25, 15, 20, 24]. As these models impose highly nontrivial constraints on the probability distributions, Chaves et al. [6] derive constraints expressed using *entropies*. Interestingly, the entropic framework originates from the foundations of quantum mechanics [5, 11]. Furthermore, causal effects in latent causal structures can be estimated by limiting the causal influence between a confounder and its children [21, 23]. Unfortunately, these methods only provide the necessary, but not sufficient, constraints for complete characterization of a causal model.

The field of causal inference is relevant to a wide range of disciplines. Curiously, quantum correlations seem to have no satisfactory explanation via classical causal models. The most famous example is Bell's theorem [2]. The underlying causal structure of a Bell experiment was investigated in [26]. It was shown that it is impossible to explain Bell-inequality violations using a classical causal model without fine-tuning. As this result only holds for the framework of classical causal inference, it appears that quantum theory calls for a quantum generalization of this formalism. Many works attempt to describe causality in quantum systems by generalizing classical concepts to the quantum formalism, for example: quantum Reichenbach's principle [1], quantum Bayes' theorem [16] and quantum randomized trial [19].

We suggest that studying classical causal models with hidden variables can provide a valuable insight into quantum causality.

In this work, we analyze causal models with hidden variables that exhibit *d*-separation relations, under the assumption that the cardinality of the unobserved variables is known. We derive inequality constraints on the probability distribution over observed variables in such scenarios. We discuss possible applications of these constraints, both in classical causal inference and in quantum causality.

Outline

Section 2 reviews the framework of graphical causal models and the definition of nonnegative rank. Section 3 contains the main results. It begins with description of nonnegative rank as a measure of dependence. Afterwards, it provides constraints implied by *d*-separation in causal models with hidden variables. Section 4 describes applications of constraints derived in Section 3. Section 5 shows that our constraints are sufficient, but not necessary for identifying direct causal influence. Finally, Section 6 discusses possible generalisations of our constraints to the quantum formalism. Lastly, Section 7 concludes the previous sections.

2   Preliminaries

Graphical Causal Models

In this section, we briefly recall basic notions and important results from graph theory and causal inference.

Consider a system of *causally* related events. We represent each event by a random variable and we denote the set of variables that corresponds to the system of events as $\mathbf{V}$. In general, $\mathbf{V}$ can contain both observed and unobserved variables.

The variables in $\mathbf{V}$, and relations between them, can be represented by a graph. Graphs consist of vertices and edges. We call a graph a *causal graph*, if the edges that connect the vertices are directed and correspond to causal relations. For example, a causal graph $X \to Y$ indicates that the variable X has a direct causal influence on Y. Here, X is said to be a parent of Y. We denote all parents of a random variable V in a graph $\mathscr{G}$ by $pa_{\mathscr{G}}(V)$.

In causal inference studies, relationships between variables are illustrated by *Directed Acyclic Graphs* (DAGs) [1]. A *causal model* consists of a DAG $\mathscr{G}$ over a set of variables $\mathbf{V}$ and a joint probability distribution $P_{\mathbf{V}}$. Causal models provide not only a probabilistic, but also a causal interpretation of the relationships between variables in a system.

Properties of a DAG that represents variables in $\mathbf{V}$ imply properties of the distribution $P_{\mathbf{V}}$. The essential one is the Markov condition which states that $P_{\mathbf{V}}$ can be decomposed as

$$P_{\mathbf{V}} = \prod_{V \in \mathbf{V}} P(V \mid pa_{\mathscr{G}}(V)). \tag{1}$$

Naturally, this condition can be tested only if all variables in $\mathbf{V}$ are observed. It means that given a DAG $\mathscr{G}$ and a probability distribution $P_{\mathbf{V}}$, with all variables in $\mathbf{V}$ being observed, the Markov condition enables us to test the *compatibility* of $\mathscr{G}$ and $P_{\mathbf{V}}$. However, when some variables in $\mathbf{V}$ are unobserved, testing compatibility is a hard task, as we already mentioned in the introduction.

---

[1]A graph is directed, if all its edges are directed (they point in a certain direction which is marked with a single arrow). A directed graph is acyclic if there is no directed path from any variable back to itself.

A different property of DAGs that is important in causal inference studies is *d*-separation. Before we formally define it, let us recall the notion of blocking a path.

Let *p* be a path in a DAG $\mathscr{G}$. We say that *p* is blocked by a set of vertices **Z** if and only if

- *p* contains a fork structure $(X \leftarrow Z \rightarrow Y)$ or a chain structure $(X \rightarrow Z \rightarrow Y)$ in which the middle vertex *Z* is in **Z**, or

- *p* contains a collider structure $(X \rightarrow Z \leftarrow Y)$ in which neither the middle vertex *Z* nor any of its descendants is in **Z**.

Definition 1 (*d*-separation) **X** *and* **Y** *are said to be* d-separated *by a set* **Z**, $(\mathbf{X} \perp_d \mathbf{Y}|\mathbf{Z})$, *if and only if* **Z** *blocks every path from a vertex in* **X** *to a vertex in* **Y**.

Any *d*-separation relation $(\mathbf{X} \perp_d \mathbf{Y}|\mathbf{Z})$ in $\mathscr{G}$ implies that **X** and **Y** are conditionally independent given **Z**, $(\mathbf{X} \perp \mathbf{Y}|\mathbf{Z})$, in the corresponding distribution $P_{\mathbf{XYZ}}$.

Notice that testing whether conditional independence relations implied by $\mathscr{G}$ hold is only possible when all variables in **X**, **Y** and **Z** are observed. When dealing with unobserved variables, conditional independence relations are sometimes impossible to evaluate.

Nonnegative factorization

We now review the definitions of nonnegative matrix factorization and nonnegative rank as they will play an important role in our results.

Nonnegative matrix factorization is a group of algorithms used for factorizing a nonnegative matrix (matrix with no negative entries) into two or more nonnegative matrices. The problem of performing a nonnegative matrix factorization is not always tractable, nevertheless numerical approximations of this task exist. This method is used in many fields of science, such as astronomy, bioinformatics and machine learning, as they study data with positive entries [3, 4, 17]. Furthermore, the nonnegative rank, which we define below, plays a critical role in communication theory as it characterizes randomized communication complexity and randomized correlation complexity [14, 27].

Formally, given a nonnegative matrix $\mathbf{M} \in \mathbb{R}_+^{n \times m}$, a nonnegative factorization of **M** means factorizing it into

$$\mathbf{M} = \mathbf{UV},$$

where $\mathbf{U} \in \mathbb{R}_+^{n \times r}$ and $\mathbf{V} \in \mathbb{R}_+^{r \times m}$ are nonnegative. The nonnegative rank of **M**, $rank_+[\mathbf{M}]$, is the smallest number *r* for which such a factorization can be found. Equivalently, the nonnegative rank of **M** is the smallest number for which **M** can be decomposed into a positive sum over *r* nonnegative rank-1 matrices [7]. For the purpose of this paper, we formalize the definition of the nonnegative rank in a following way.

Definition 2 (nonnegative rank) *Let* $\mathbf{P}_{XY}$ *be a* $(|X| \times |Y|)$-*dimensional nonnegative matrix. The non-negative rank of* $\mathbf{P}_{XY}$ *is defined as*

$$rank_+[\mathbf{P}_{XY}] \equiv \min\{r \mid \exists (\mathbf{P}_X^{(1)}, \mathbf{P}_Y^{(1)}), \ldots, (\mathbf{P}_X^{(r)}, \mathbf{P}_Y^{(r)}) \; \mathbf{P}_{XY} = \sum_{z=1}^{r} \mathbf{P}_X^{(z)} \mathbf{P}_Y^{(z)}\}, \qquad (2)$$

where $\mathbf{P}_X^{(z)}$ *is a nonnegative* $|X|$-*dimensional row vector and each* $\mathbf{P}_Y^{(z)}$ *is a nonnegative* $|Y|$-*dimensional column vector.*

The matrices in the set $\{\mathbf{P}_X^{(z)} \mathbf{P}_Y^{(z)}\}_z$ have rank 1, hence this definition coincides with the standard definition of the nonnegative rank.

3   Rank constraints for causal models

In this section, we explore a relation between the nonnegative rank of a probability distribution over two variables, and the causal relation between these variables.

Nonnegative rank as a measure of dependence

We start by making a trivial observation that any joint probability distribution over two variables can be represented as a nonnegative matrix $\mathbf{P}_{XY}$, with entries $\mathbf{P}_{xy} = \mathbf{P}_{X=x,Y=y} = P(X = x, Y = y)$. Notice that although we donote $\mathbf{P}_{xy}$ using the bold font, it is a single element of the matrix $\mathbf{P}_{XY}$. If X and Y are independent, $\mathbf{P}_{XY}$ can be factorized as $\mathbf{P}_{XY} = \mathbf{P}_X \mathbf{P}_Y$. Then, it follows from Definition 2 that $\mathbf{P}_{XY}$ has nonnegative rank 1. Conversely, as the correlation between X and Y increases, so does the nonnegative rank of their distribution. It means that the nonnegative rank of a probability distribution can be interpreted as a measure of dependence between two variables.

    Consider an example in which the dependence between X and Y is only due to a third variable, denoted by Z. Let Z take r different values, i.e., the cardinality of Z is r, which we denote $|Z| = r$. Throughout this paper we take r to be the minimum cardinality of Z for which X and Y are conditionally independent. If X and Y are conditionally independent given Z, the probability distribution must factorize as:

$$\mathbf{P}_{XY} = \sum_{z=1}^{r} \mathbf{P}_{X|Z=z} \mathbf{P}_{Y|Z=z} \mathbf{P}_{Z=z}. \tag{3}$$

Here, $\mathbf{P}_{X|Z=z}$ ($\mathbf{P}_{Y|Z=z}$) is a conditional probability distribution over X (Y) given that Z takes a particular value $Z = z$. This factorization is equivalent to the one given in Eq. (2), with $rank_+[\mathbf{P}_{XY}] = r$. We summarize this result in the following Proposition.

Proposition 1 *Let $\mathbf{P}_{XY} \in \mathbb{R}_+^{n \times m}$ represent a joint probability distribution on random variables X and Y that have finite cardinalities, $|X| = n$ and $|Y| = m$. The following are equivalent:*

   1. There exists a nonnegative factorization of $\mathbf{P}_{XY}$ with $rank_+[\mathbf{P}_{XY}] = r$.

   2. $\mathbf{P}_{XY}$ can be decomposed as in Eq. (3) with $|Z| = r$.

Proof 1 *$1 \implies 2$: Suppose $\mathbf{P}_{XY}$ factorizes as in Eq. (2), with $\mathbf{P}_X^{(z)}$ and $\mathbf{P}_Y^{(z)}$ being unnormalized probability distributions. Define the sums*

$$Q_X^{(z)} = \sum_x \mathbf{P}_{X=x}^{(z)}, \qquad Q_Y^{(z)} = \sum_y \mathbf{P}_{Y=y}^{(z)}.$$

Notice that the set of products $\{Q_X^{(z)} Q_Y^{(z)}\}_z$ form a valid probability distribution. We denote it as

$$\mathbf{P}_{Z=z} = Q_X^{(z)} Q_Y^{(z)},$$
$$\text{where } \sum_z \mathbf{P}_{Z=z} = \sum_{x,y,z} \mathbf{P}_{X=x}^{(z)} \mathbf{P}_{Y=y}^{(z)} = \sum_{x,y} \mathbf{P}_{xy} = 1.$$

Further, define the following vectors:

$$\mathbf{C}_X^{(z)} = \frac{\mathbf{P}_X^{(z)}}{Q_X^{(z)}}, \qquad \mathbf{C}_Y^{(z)} = \frac{\mathbf{P}_Y^{(z)}}{Q_Y^{(z)}}.$$

We can now decompose $\mathbf{P}_{XY}$ *as*

$$\mathbf{P}_{XY} = \sum_z \mathbf{P}_X^{(z)} \mathbf{P}_Y^{(z)} = \sum_z C_X^{(z)} \, \mathbf{C}_Y^{(z)} \, (Q_X^{(z)} Q_Y^{(z)}).$$

It is now clear that one can define the following conditional probabilities

$$\mathbf{P}_{X|Z=z} \equiv \mathbf{C}_X^{(z)}, \qquad \mathbf{P}_{Y|Z=z} \equiv \mathbf{C}_Y^{(z)},$$

and decompose $\mathbf{P}_{XY}$ *as*

$$\mathbf{P}_{XY} = \sum_z \mathbf{P}_{X|Z=z} \mathbf{P}_{Y|Z=z} \mathbf{P}_{Z=z},$$

which is the factorization from Eq. (3).

2 $\implies$ 1: Assume $\mathbf{P}_{XY}$ *can be factorized as in Eq. (3). Then, since* $\mathbf{P}_{Z=z}$ *is just a positive number, it follows trivially that* $rank_+[\mathbf{P}_{XY}] = r$.

We note that Proposition 1 (1 $\implies$ 2) was previously explored in [13].

Recall that we assumed that *r* is the minimum cardinality of *Z* for which *X* and *Y* are conditionally independent. The direction (2 $\implies$ 1) of Proposition 1 is true due to this assumption. In general, the cardinality of *Z* can be greater than $rank_+[\mathbf{P}_{XY}]$, which we exploit in the next Subsection.

Restricted cardinality constraints

Consider the example from the previous Subsection, where *Z* is assumed to be the only source of dependence between *X* and *Y*. If *Z* is an observed variable, it is straightforward to verify whether this hypothesis holds. However, if *Z* is unobserved, constraints implied by conditional independence relations are impossible to directly evaluate. We now show that knowledge about the cardinality of *Z* is enough to test conditional independence relations in some cases.

Let us interpret Proposition 1 from the perspective of causal models. If $\mathbf{P}_{XY}$ can be factorized as in Eq. (3), then there exist a hypothetical variable *Z*, with cardinality $|Z| = r$, and probability distribution $\mathbf{P}_Z$ that makes *X* and *Y* conditionally independent. In graphical causal models, the notion of conditional independence is captured by *d*-separation. If two variables *X* and *Y* are conditionally independent given *Z*, a DAG that illustrates the causal structure of *X*, *Y* and *Z* must satisfy the following two properties. First, there must be no direct causal influence between *X* and *Y* (no direct arrow form *X* to *Y*, or from *Y* to *X*). Second, *Z* must *d*-separate *X* and *Y*. Three examples of causal structures that satisfy these properties are depicted in Figure 1. Circles correspond to observed variables, and dashed circles correspond to unobserved variables. In Figure 1 (a), *Z* is a common cause of *X* and *Y*. In Figure 1 (b), *Z* is a causal mediary between *X* and *Y*. Finally, in Figure 1 (c), a *d*-separating set $\mathbf{Z} = \{Z_1, Z_2\}$ consists of both a common cause and a causal mediary. In the first two examples, the cardinalities of the *d*-separating variables are given by $|Z|$. In the third example, the cardinality of the *d*-separating set is equal to $|Z_1| \times |Z_2|$.

From Proposition 1, the variable that *d*-separates *X* and *Y* must have at least cardinality *r*, which is given by $rank_+[\mathbf{P}_{XY}]$. This result is summarized in the corollary below.

Corollary 1 *The nonnegative rank of the matrix* $\mathbf{P}_{XY}$ *whose components are* $\mathbf{P}_{xy} = P(X = x, Y = y)$ *constitutes a lower bound for the cardinality of a variable* $\mathbf{Z}$ *that d-separates* $\mathbf{X}$ *and* $\mathbf{Y}$,

$$rank_+[\mathbf{P}_{XY}] \leq |\mathbf{Z}|. \tag{4}$$

The constraint given in Eq. (4) enables to verify whether conditional independence relations implied by a DAG hold, even when the *d*-separating variable is unobserved. The knowledge about *Z* required to evaluate the bound is only its cardinality. In the next section, we give examples that use Corollary 1 to test conditional independence relations in DAGs with latent variables.
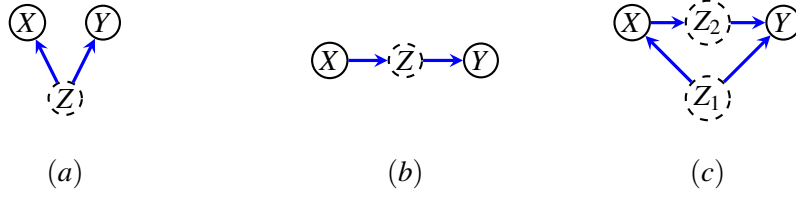
Figure 1: Examples of *d*-separating variables.

4 Applications

In this section, we give two applications of Corollary 1. First, we describe its potential to improve causal discovery algorithms. In particular, we consider a scenario in which our constraint witnesses direct causal influence between two variables in a presence of a confounder. Second, we reprove two results from communication theory: characterization of randomized correlation complexity and randomized communication complexity.

Witnessing Causal Influence

Imagine that two ternary events, X and Y ($|X| = |Y| = 3$), are observed to exhibit some statistical dependence. Additionally, assume they share an unobserved common cause U which is known to have cardinality $|U| = 2$. Two possible hypotheses regarding the causal relations between X, Y and U are illustrated in Fig. 2. In Fig. 2 (a), one believes the statistical dependence is purely due to a common
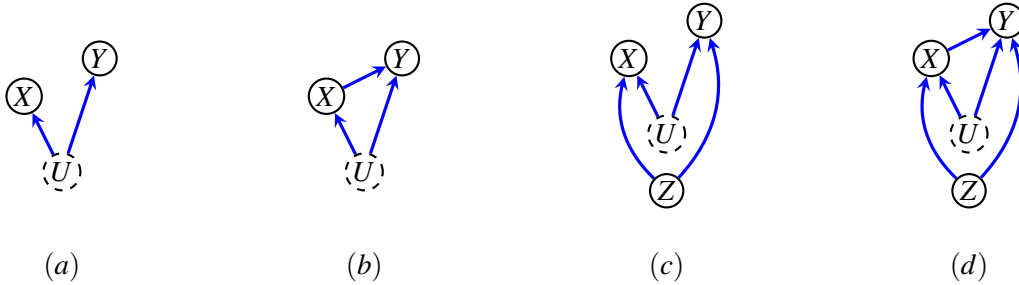

Figure 2: Witnessing direct causal influence in a presence of a confounder.

cause, whereas in Fig. 2 (b), X additionally has a direct causal influence on Y.

In Fig. 2 (a), U *d*-separates X and Y. From Corollary 1, we know that

$$rank_+[\mathbf{P}_{XY}] \leq |U|.$$

Therefore, if $rank_+[\mathbf{P}_{XY}] > 2$, one can discard the hypothesis of Fig. 2 (a) and conclude that there must be a direct causal influence of X on Y as in Fig. 2 (b).

As an example, suppose $\mathbf{P}_{XY}$ is given by the probability distribution from Table 1. In this case, the nonnegative rank of $\mathbf{P}_{XY}$ is equal to 3. One can certify that $rank_+[\mathbf{P}_{XY}]$ is greater than $|U|$, and only the causal hypothesis of Fig. 2 (b) is compatible with this probability distribution. In this example, observing $rank_+[\mathbf{P}_{XY}] = 3$ *witnesses* direct influence between X and Y.

		Y		
		0	1	2
	0	0	$1/6$	$1/6$
X	1	$1/6$	0	$1/6$
	2	$1/6$	$1/6$	0

Table 1: An example of a probability distribution $\mathbf{P}_{XY}$, with $|X| = |Y| = 3$.

The example above shows the essence of Corollary 1. It captures the fact that strong correlation between high cardinality variables, when no direct influence between them exists, must be explained by a high cardinality common cause.

Corollary 1 can also be applied to scenarios where the common cause is *partially* observed. We define a partially observed set as a set in which only some variables are observed. An example of a partially observed common cause $\mathbf{C} = \{Z, U\}$ is illustrated in Fig. 2. In Fig. 2 (c), X and Y share two common causes: an unobserved one, denoted by U, and an observed one, denoted by Z. In Fig. 2 (d), X additionally influences Y.

In Fig. 2 (c), Z is an observed variable, and we can condition the values of X and Y on its particular value, $Z = z$. Let $\mathbf{P}_{XY}^{(z)}$ denote the probability distribution over X and Y conditioned on a particular value z of Z, $\mathbf{P}_{xy}^z = P(X = x, Y = y | Z = z)$. After conditioning, U d-separates X and Y (given $Z = z$) in Fig. 2 (c). The constraint given by Corollary 1 in this case is

$$\forall z: \ rank_+[\mathbf{P}_{XY}^{(z)}] \leq |U|.$$

Given $\mathbf{P}_{XY}^{(z)}$ for all z, any violation of this condition witnesses direct causal influence between X and Y, just like in the example with an unobserved common cause.

Corollary 1 can also be used to solve a converse problem. Suppose a distribution $\mathbf{P}_{XY}$ is observed, where variables X and Y are known to share an unobserved common cause Z (or more generally, are d-separated by Z). If nothing is known about Z, constraint from Eq. (4) can be used to lower bound its cardinality.

Distribution generation

Recall a distribution generation protocol that is often discussed in communication theory [27, 14]. Suppose two parties, Alice and Bob, are responsible for generating random variables X and Y respectively, such that X and Y have a prespecified joint distribution $\mathbf{P}_{XY}$. Our goal is to quantify how much information must be communicated or shared to generate $\mathbf{P}_{XY}$.

In the literature about this problem two scenarios have been developed and studied separately. In the first scenario, Alice and Bob each receive information from a third party before they generate X and Y. We denote the information they share by Z. In this protocol, no communication between Alice and Bob is allowed. The information Z distributed this way is often referred to as a random seed. The minimum size of a seed that must be given to Alice and Bob in order to produce the required distribution $\mathbf{P}_{XY}$ is called the *randomized correlation complexity* of the joint distribution, and is denoted $RCorr[\mathbf{P}_{XY}]$. Here, a size of a seed is the number of bits that is distributed to each party, and is given by $\log_2(|Z|)$. In the second scenario, one-way communication between Alice and Bob is allowed. After generating X, Alice sends a message to Bob through an intermediary, which we denote M. The minimum number of bits Alice

needs to communicate to Bob such that they can produce the target distribution is called the *randomized communication complexity*, and is denoted $RComm[\mathbf{P}_{XY}]$. The number of bits communicated between the parties is given by $\log_2(|M|)$.

It has been shown that $RCorr[\mathbf{P}_{XY}]$ and $RComm[\mathbf{P}_{XY}]$ are equal, and are fully characterized by $\log_2(rank_+[\mathbf{P}_{XY}])$ [27]. We now show that given Corollary 1, this result follows trivially from a causal analysis of the problem. To begin, we note that in the first scenario, the seed Z is distributed to both Alice and Bob, hence it can affect both X and Y. Graphically, this scenario corresponds to Fig. 1 (a), where Z is a common cause of X and Y. In the second scenario, the message M can depend on the random variable X, which in turn affects how Bob generates Y. This scenario is illustrated by a causal structure in Fig. 1 (b), with M playing the role of Z in the figure.

We now observe that X and Y are d-separated by Z and M respectively in the two scenarios. Corollary 1 implies that

$$|Z| \geq rank_+[\mathbf{P}_{XY}] \qquad \text{and} \qquad |M| \geq rank_+[\mathbf{P}_{XY}].$$

It follows that

$$\log_2(|Z|) \geq \log_2(rank_+[\mathbf{P}_{XY}]) \qquad \text{and} \qquad \log_2(|M|) \geq \log_2(rank_+[\mathbf{P}_{XY}]).$$

Note that $RCorr[\mathbf{P}_{XY}]$ and $RComm[\mathbf{P}_{XY}]$ correspond to the *minimum* size of a seed and a message, respectively, needed to generate $\mathbf{P}_{XY}$. Therefore, they are given by the lower bound of the above inequalities. Moreover, Corollary 1 does not distinguish a common cause Z from a causal mediary M, as they both *d*-separate X and Y in this example. This result yields both the equivalence of $RCorr[\mathbf{P}_{XY}]$ and $RComm[\mathbf{P}_{XY}]$, and the fact that they are fully characterized by $\log_2(rank_+[\mathbf{P}_{XY}])$. This example shows that adopting a causal perspective can simplify results in communication theory.

Finally, we analyze a more involved protocol, in which Alice and Bob share a seed Z_1, *and* Alice can communicate to Bob through a mediary Z_2. This scenario is illustrated in Fig. 1 (c). The cardinality of the d-separating set $\mathbf{Z} = \{Z_1, Z_2\}$ is equal to $|Z_1| \times |Z_2|$. It follows from Corollary 1 that

$$|Z_1| \times |Z_2| \geq rank_+[\mathbf{P}_{XY}] \qquad \text{and} \qquad \log_2(|Z_1| \times |Z_2|) \geq \log_2(rank_+[\mathbf{P}_{XY}]).$$

We can intuitively view the constraint on $|Z_1| \times |Z_2|$ as defining a trade-off between the amount of information shared through a seed and the amount transferred using communication. This perspective enables studying relatively complex problems in which the size of the seed or the capacity of the communication channel is constrained. Similarly, if one of these cardinalities is unobserved, information about the other can be used to bound it, using Corollary 1.

5   A Specific Case with Additional Constraints

The inequality given in Corollary 1 gives a sufficient, but not a necessary condition for identifying causal relations between variables. In this Section, we give an example of a correlation that satisfies the constraints implied by the *d*-separation relations in a model, but nevertheless imposes new constraints that require additional causal mechanisms.

Let us revisit the causal hypotheses illustrated in Fig. 2 (c) and Fig. 2 (d). Assume X, Y and Z are binary observed variables, and U is a binary unobserved variable. In this scenario, $\mathbf{P}_{XY}^{(z)}$ is a 2×2 matrix and the constraint given in Corollary 1 is *always* satisfied. This is the case because $rank_+[\mathbf{P}_{XY}^{(z)}] \leq 2$ and $|U| = 2$. However, this is not enough to conclude that the causal relations between X, Y, Z and U are those illustrated by Fig. 2 (c). In this Section, we give an example of a probability distribution $\mathbf{P}_{XY}^{(z)}$ that

satisfies the above conditions, and requires a direct causal influence as in Fig. 2 (d). First, we consider a specific probability distribution $\mathbf{P}_{XY}^{(z)}$, with $|X| = |Y| = |Z| = |U| = 2$, under the causal hypothesis of Fig. 2 (c). Then, we show that the causal structure of Fig. 2 (c) implies some nontrivial constraints that must be satisfied in order to explain the specific form of $\mathbf{P}_{XY}^{(z)}$.

We start with a reminder that an intuitive way of describing an outcome of a random variable is with its expectation value. Let $\mathbf{V} = \{V_1, ..., V_n\}$ be a set of n binary variables that each take a value in $\{-1, 1\}$. The conversion formula between the probability distribution $P(\mathbf{V}) = P(V_1, ...V_n)$ and the expectation values of variables in $\mathbf{V}$ can be written as

$$P(V_1, ...V_n) = \frac{1}{2^n} \left(1 + \sum_{k=1}^{n} \sum_{1 \le i_1 < ... < i_k \le n} v_{i_1} ... v_{i_k} \langle V_{i_1} ... V_{i_k} \rangle \right). \tag{5}$$

Let us consider a special case when X and Y are observed to be perfectly positively correlated for both possible values of Z,

$$\langle XY | Z = z \rangle = 1. \tag{6}$$

Assume that this perfect correlation arises due to the causal mechanism encoded in Fig. 2 (c). Eq. (6) implies that

$$\begin{aligned} &\text{(a)} \quad \langle X|Z,U \rangle = \pm 1; \text{and} \\ &\text{(b)} \quad \langle Y|Z,U \rangle = \langle X|Z,U \rangle, \end{aligned} \tag{7}$$

for both values of Z and U, that is, X and Y need to depend deterministically on Z and U (Eq. (7a)) in the same way (Eq. (7b)).

Our goal is to derive constraints on $\langle X|Z = z \rangle$ that arise due to this condition. If any nontrivial constraints exist, the causal hypothesis of Fig. 2 (c) explains the perfect correlation between X and Y *only if* they are satisfied. Violation of these constraints calls for an additional causal mechanism (additional to ones in Fig. 2 (c)) that explains the dependence between X and Y.

In order for the constraints on $\langle X|Z = z \rangle$ to be verifiable, they can only depend on the observed variables. We perform a quantifier elimination using the known constraints to eliminate the unobserved variable from the set of inequalities and equalities. We find that either

$$\langle X|z = 1 \rangle = \pm \langle X|z = -1 \rangle$$
$$\text{or } \text{Min}[\langle X|z = 1 \rangle, \langle X|z = -1 \rangle] = -1$$
$$\text{or } \text{Max}[\langle X|z = 1 \rangle, \langle X|z = -1 \rangle] = 1.$$

It then follows from $\langle XY|Z = z \rangle = 1$ that

$$\langle Y|z = 1 \rangle = \pm \langle Y|z = -1 \rangle$$
$$\text{or } \text{Min}[\langle Y|z = 1 \rangle, \langle Y|z = -1 \rangle] = -1$$
$$\text{or } \text{Max}[\langle Y|z = 1 \rangle, \langle Y|z = -1 \rangle] = 1.$$

These constraints show that for the special case being considered, one of the following conditions must be true for the marginal expectation values of X conditional on Z:

1. $\langle X|Z = z \rangle$ has the same magnitude for $z = 1$ and $z = -1$ (it is equal up to a sign); or

2. $\langle X|Z = z \rangle$ is deterministic and does not depend on U at all for one value of $Z = z$.

The same holds for $\langle Y|Z = z \rangle$. If any of these conditions on $\langle X|Z = z \rangle$ and $\langle Y|Z = z \rangle$ are violated, then we conclude that perfect positive correlation between X and Y cannot be explained by the causal structure encoded in Fig. 2 (c). An additional causal influence between X and Y is needed to explain the considered correlation, for example the one in Fig. 2 (d).

This is an example of a correlation between X and Y wherein the constraint induced by Corollary 1 is satisfied for the d-separation relation pertinent to the causal hypothesis of Fig. 2 (c), but nonetheless this correlation is incompatible with this causal hypothesis.

6   Towards Quantum Causality

We now give a different perspective on the motivation behind studying causal models with hidden common causes. In this Section, we show that DAGs with unobserved common causes can act as classical analogues of models in which a *quantum* common cause is shared.

Consider a scenario in which two parties share a source of entangled quantum systems and use it to output a classical correlation $\mathbf{P}_{AB}$. The structure of this protocol is illustrated in Fig. 3. The quantum seed that is distributed to Alice and Bob is denoted by ρ_C, quantum systems are represented by double wires and classical variables are represented by single wires. Under the assumption that only one copy of ρ_C is available in this protocol, some form of a quantum copy operation must be performed on it before the distribution. As quantum systems cannot be cloned, the quantum decomposition operation introduced in [1] is applied. It is represented with the white dot symbol and implies the decomposition of the Hilbert space $\mathscr{H}_C$ into a direct sum of tensor products. This operation decomposes ρ_C into factors on $\mathscr{H}_{C_B}$ carried to Bob and $\mathscr{H}_{C_A}$ carried to Alice. The outputs of this operation are denoted by ρ_A and ρ_B. Assume $\mathbf{P}_{AB}$ is generated by one of the two protocols. The first protocol is illustrated in Fig. 3 (a). Here, Bob and Alice perform measurements M_A and M_B on their share of the system ρ_C to generate classical variables A and B, respectively. The second protocol is encoded in Fig. 3 (b). It allows for classical communication from Bob to Alice before Alice outputs A. In this protocol, the value of A is an output of a gate U_{AB} that depends on the value of B.
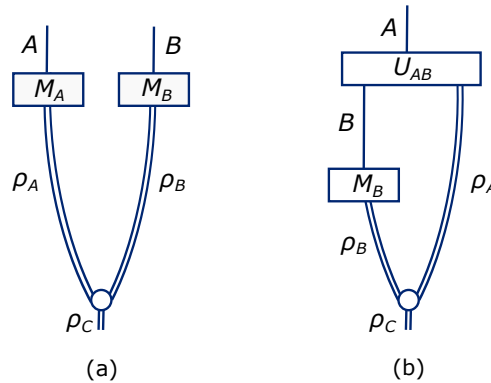

Figure 3: Circuit representation of correlation generation protocols.

The main challenge of analyzing these protocols is the restriction on acquiring information about quantum systems relative to classical systems. Once a quantum state is sharply measured, it is projected to a post-measurement state that corresponds to the choice of measurement. This feature of quantum

theory prevents the observer from making passive observations of quantum common causes. However, the observer can possess some limited information about the quantum system, for example, its dimension. This situation is analogous to a classical causal model involving an unobserved common cause with restricted cardinality. In such models, the observer has limited knowledge about the classical common cause, namely, its cardinality. This perspective on the problem allows to see the similarities between classical causal models illustrated in Figs. 2 (a) and (b), and circuits in Figs. 3 (a) and (b).

Although Corollary 1 only concerns classical causal models, a similar concept is used in communication theory [27, 10], where models with a quantum seed can be characterized using positive semidefinite rank [14]. Therefore, Corollary 1 has the potential to be further extended to models that include quantum systems. Specifically, knowledge about the dimension of a quantum common cause could play a role similar to that of the cardinality of a classical common cause. Formally, the minimum dimension of the local Hilbert space of a shared entangled state needed to generate $\mathbf{P}_{AB}$ is defined as quantum correlation complexity. It is characterized by the positive semidefinite rank of $\mathbf{P}_{AB}$ [14], where positive semidefinite rank is defined as follows. Let $\dim(\rho_A)$ and $\dim(\rho_B)$ be the dimensions of the local Hilbert space of Alice and Bob, respectively.

Definition 3 (positive semidefinite rank) *Let $\mathbf{P}_{AB}$ be a $dim(\rho_A) \times dim(\rho_B)$-dimensional nonnegative matrix. The positive semidefinite rank of $\mathbf{P}_{AB}$ is defined as*

$$
\begin{aligned}
rank_{PSD}[\mathbf{P}_{AB}] &\equiv \min\{r \mid \exists (\mathbf{E}_1, \mathbf{F}_1), ..., (\mathbf{E}_{dim(\rho_A)}, \mathbf{F}_{dim(\rho_B)}) \\
P_{ab} &= \mathrm{tr}\{E_a F_b\} \quad \forall a \in \{1...dim(\rho_A)\}, b \in \{1...dim(\rho_B)\},
\end{aligned}
\tag{8}
$$

where $\mathbf{E}_a$ and $\mathbf{F}_b$ are $r \times r$ positive semidefinite matrices.

Assume $\dim(\rho_A) = \dim(\rho_B) = d$. It follows that for the hypothesis illustrated in Fig. 3 (a) to hold, the relation

$$d \geq rank_{PSD}[\mathbf{P}_{AB}]$$

must be satisfied. This relation enables us to analyse Figs. 3 (a) and (b) in a similar manner to the analysis of Figs. 2 (a) and (b) we conducted in the subsection *Witnessing Causal Influence*.

One interesting property of the positive semidefinite rank is that for any matrix $\mathbf{M}$

$$rank_{PSD}[\mathbf{M}] \leq rank_+[\mathbf{M}].$$

For example, for the probability distribution $\mathbf{P}_{XY}$ defined in Table 1, $rank_+[\mathbf{P}_{XY}] = 3$, while $rank_{PSD}[\mathbf{P}_{XY}] = 2$ (see Example 2.6 in [9]). It means that the dimension of a quantum resource needed to generate $\mathbf{P}_{XY}$ is smaller than the cardinality of a classical resource needed for this task. Examples of this form have been studied before [27], and they show an advantage of quantum resources over classical ones.

We finish by noting that a similar approach can be used to analyse a protocol in which a quantum message is sent from Bob to Alice (the quantum analogue of Fig. 1 (b)). Similarities between classical and quantum protocols for correlation generation suggest that classical results, like the one given in Corollary 1, can be used to set the stage for the quantum equivalent of causal analysis.

7 Conclusions

We derived universal constraints for causal models in which the cardinality of an unobserved d-separating variable is known. While the majority of existing methods of constraining distributions compatible with latent causal structures do not use any information about unobserved variables, our result relies on having

knowledge of the cardinality of an unobserved variable. We showed that our result has applications in causal analysis. Furthermore, we recognized the analogy between classical causal models with hidden common causes and models that involve quantum systems. As we identify a new way of bounding causal effects in models with hidden variables, we set the stage for quantum extensions of this formulation.

Acknowledgments

References

[1] John-Mark A Allen, Jonathan Barrett, Dominic C Horsman, Ciarán M Lee & Robert W Spekkens (2017): *Quantum common causes and quantum causal models*. *Physical Review X* 7(3), p. 031021, doi:10.1103/PhysRevX.7.031021.

[2] John S Bell (1964): *On the Einstein Podolsky Rosen paradox*. *Physics Physique Fizika* 1(3), p. 195, doi:10.1103/PhysicsPhysiqueFizika.1.195.

[3] Olivier Berne, C Joblin, Y Deville, JD Smith, M Rapacioli, JP Bernard, J Thomas, W Reach & A Abergel (2007): *Analysis of the emission of very small dust particles from Spitzer spectro-imagery data using blind signal separation methods*. *Astronomy & Astrophysics* 469(2), p. 575, doi:10.1051/0004-6361:20066282.

[4] Michael W Berry, Murray Browne, Amy N Langville, V Paul Pauca & Robert J Plemmons (2007): *Algorithms and applications for approximate nonnegative matrix factorization*. *Computational statistics & data analysis* 52(1), p. 155, doi:10.1016/j.csda.2006.11.006.

[5] Rafael Chaves (2013): *Entropic inequalities as a necessary and sufficient condition to noncontextuality and locality*. *Physical Review A* 87(2), p. 022102, doi:10.1103/PhysRevA.87.022102.

[6] Rafael Chaves, Lukas Luft, Thiago O Maciel, David Gross, Dominik Janzing & Bernhard Schölkopf (2014): *Inferring latent structures via information inequalities*. *ArXiv e-prints*.

[7] Joel E Cohen & Uriel G Rothblum (1993): *Nonnegative ranks, decompositions, and factorizations of nonnegative matrices*. *Linear Algebra and its Applications* 190, pp. 149–168, doi:10.1016/0024-3795(93)90224-C.

[8] Robin J Evans (2012): *Graphical methods for inequality constraints in marginalized DAGs*. In: *2012 IEEE International Workshop on Machine Learning for Signal Processing*, IEEE, p. 1, doi:10.1109/MLSP.2012.6349796.

[9] Hamza Fawzi, João Gouveia, Pablo A Parrilo, Richard Z Robinson & Rekha R Thomas (2015): *Positive semidefinite rank*. *Mathematical Programming* 153(1), pp. 133–177, doi:10.1007/s10107-015-0922-1.

[10] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary & Ronald De Wolf (2012): *Linear vs. semidefinite extended formulations: exponential separation and strong lower bounds*. In: *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, p. 95, doi:10.1145/2213977.2213988.

[11] Tobias Fritz (2012): *Beyond Bell's theorem: correlation scenarios*. *New Journal of Physics* 14(10), p. 103001, doi:10.1088/1367-2630/14/10/103001.

[12] Dan Geiger & Christopher Meek (2013): *Quantifier elimination for statistical problems*. *ArXiv e-prints*.

[13] Mena Gonzalo (2017): *Conditional Independence as Non-Negative Matrix Factorization*. *Note.* Available at http://stat.columbia.edu/~gonzalo/pubs/CondIndep.pdf.

[14] Rahul Jain, Yaoyun Shi, Zhaohui Wei & Shengyu Zhang (2013): *Efficient protocols for generating bipartite classical distributions and quantum states*. IEEE Transactions on Information Theory 59(8), p. 5171, doi:10.1109/TIT.2013.2258372.

[15] Changsung Kang & Jin Tian (2012): *Polynomial constraints in causal Bayesian networks*. ArXiv e-prints.

[16] Matthew S Leifer & Robert W Spekkens (2013): *Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference*. Physical Review A 88(5), p. 052130, doi:10.1103/PhysRevA.88.052130.

[17] Ben Murrell, Thomas Weighill, Jan Buys, Robert Ketteringham, Sasha Moola, Gerdus Benade, Lise Du Buisson, Daniel Kaliski, Tristan Hands & Konrad Scheffler (2011): *Non-negative matrix factorization for learning alignment-specific models of protein evolution*. PLOS one 6(12), doi:10.1371/journal.pone.0028898.

[18] Judea Pearl (2009): *Causality*. Cambridge University Press, doi:10.1017/CBO9780511803161.

[19] Katja Ried, Megan Agnew, Lydia Vermeyden, Dominik Janzing, Robert W Spekkens & Kevin J Resch (2015): *A quantum advantage for inferring causal structure*. Nature Physics 11(5), p. 414, doi:10.1038/nphys3266.

[20] Denis Rosset, Cyril Branciard, Tomer Jack Barnea, Gilles Pütz, Nicolas Brunner & Nicolas Gisin (2016): *Nonlinear Bell inequalities tailored for quantum networks*. Physical Review Letters 116(1), p. 010403, doi:10.1103/PhysRevLett.116.010403.

[21] Denis Rosset, Nicolas Gisin & Elie Wolfe (2017): *Universal bound on the cardinality of local hidden variables in networks*. ArXiv e-prints, doi:10.26421/QIC18.11-12.

[22] Peter Spirtes, Clark N Glymour, Richard Scheines & David Heckerman (2000): *Causation, prediction, and search*. MIT press, doi:10.1007/978-1-4612-2748-9.

[23] Bastian Steudel & Nihat Ay (2015): *Information-theoretic inference of common ancestors*. Entropy 17(4), p. 2304, doi:10.3390/e17042304.

[24] Armin Tavakoli (2016): *Bell-type inequalities for arbitrary noncyclic networks*. Physical Review A 93(3), p. 030101, doi:10.1103/PhysRevA.93.030101.

[25] Jin Tian & Judea Pearl (2012): *On the testable implications of causal models with hidden variables*. ArXiv e-prints.

[26] Christopher J Wood & Robert W Spekkens (2015): *The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning*. New Journal of Physics 17(3), p. 033002, doi:10.1088/1367-2630/17/3/033002.

[27] Shengyu Zhang (2012): *Quantum strategic game theory*. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, p. 39, doi:10.1145/2090236.2090241.