THE d-SEPARATION CRITERION IN CATEGORICAL PROBABILITY

EXTENDED ABSTRACT

TOBIAS FRITZ AND ANDREAS KLINGLER

The d-separation criterion [3] is a method for testing the compatibility of a probability distribution with a given causal structure. It states that a joint distribution of a collection of random variables is compatible with a causal structure if and only if it satisfies a list of conditional independence relations encoded in the causal structure. While causal models are mainly studied for probability distributions which are discrete or Gaussian, we here introduce a way of treating causal models and the d-separation criterion independently of the underlying types of variables. To the best of our knowledge, this yields the first proof of a d-separation criterion for continuous variables.

Our approach belongs to the field of *categorical probability theory*. This framework constitutes an abstract way of treating probability independently of measure theory. It allows for a transparent treatment that already contains all structural properties necessary for developing the d-separation criterion. The main objects of study are *Markov categories* [1], which model Markov kernels of certain classes (i.e. arising from discrete, Gaussian, or measure-theoretic probability distributions) as morphisms. Markov categories allow for proofs of statements independently of whether random variables are discrete, Gaussian or continuous. More concretely, we study causal models and causal compatibility in the recent framework of generalized causal models introduced in [2] by defining a categorical notion of d-separation and proving an abstract version of the d-separation criterion.

By definition, generalized causal models are the morphisms of a free Markov category, or equivalently certain kinds of string diagrams. The wires in such a string diagram correspond to local random variables and the boxes to causal mechanisms. We call a Markov kernel f *compatible* with a given causal structure if f arises as the diagram by filling in concrete Markov kernels into its boxes. Therefore, compatibility is a functorial property. This framework is more general than the traditional notion of causal structure defined as a directed acyclic graph in several ways: for example, it models Markov kernels rather than mere probability distributions; through string diagrams in which the same box appears several times, it allows for causal structures with the constraint of multiple identical causal mechanisms.

Based on the categorical definition of causal models, we define a categorical notion of d-separation. A causal model displays d-separation of a collection of wires $\mathcal{Z}$ if the string diagram becomes disconnected upon removing the wires in $\mathcal{Z}$. Our first main result is that categorical d-separation is equivalent to the classical d-separation whenever the latter makes sense (Proposition 19). Finally, we prove an abstract version of the d-separation criterion as our main result:

Department of Mathematics, Technikerstr. 13, A-6020 Innsbruck, Austria
Institute for Theoretical Physics, Technikerstr. 21a, A-6020 Innsbruck, Austria
E-mail addresses: tobias.fritz@uibk.ac.at, andreas.klingler@uibk.ac.at.
Date: July 11, 2022.

Theorem 27. *Let* $\mathbf{C}$ *be a strict Markov category with conditionals,* φ *a pure bloom causal model in which each type of box appears at most once in* φ*, and* f *a morphism in* $\mathbf{C}$*. Then the following statements are equivalent:*

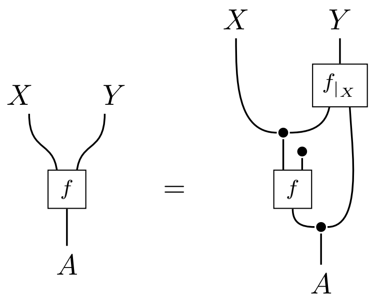
(a) f *is compatible with* φ*.*

(b) f *satisfies the global Markov property, i.e. for every three disjoint subsets* $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ *of wires in* φ*,*

$$\mathcal{Z} \text{ d-separates } \mathcal{X} \text{ and } \mathcal{Y} \quad \Longrightarrow \quad \mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}.$$

(c) f *satisfies the local Markov property, i.e. every output is conditionally independent from its non-descendants given its inputs.*

A central assumption in this result is the existence of conditionals [1]. Pictorially, a Markov category has conditionals if every $f : A \to X \otimes Y$ factorizes like this:



In other words, the outputs in f are generated successively while having access to all prior information. Conditionals exist for discrete random variables as well as continuous random variables on standard Borel spaces and in Gaussian probability, and this facilitates the application of our results to all of these cases.

Our abstract approach to the d-separation criterion has several attractive features:

- In contrast to classical d-separation, the formulation in terms of string diagram connectedness makes our categorical d-separation criterion is very simple to understand.
- It provides a uniform treatment of the d-separation criterion for discrete variables, Gaussian variables and continuous variables, where the latter seems to be entirely new.
- We may hope for future extensions to more general kinds of causal models to which classical d-separation does not apply.

References

[1] T. Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:1–98, 2020. https://doi.org/10.1016/j.aim.2020.107239, https://arxiv.org/abs/1908.07021.

[2] T. Fritz and W. Liang. Free gs-monoidal categories and free Markov categories. 2022. https://arxiv.org/abs/2204.02284.

[3] J. Pearl. *Causality: Models, reasoning, and inference*. Cambridge University Press, 2009. Second edition.

THE D-SEPARATION CRITERION IN CATEGORICAL PROBABILITY

TOBIAS FRITZ AND ANDREAS KLINGLER

Abstract. The d-separation criterion detects the compatibility of a joint probability distribution with a directed acyclic graph through certain conditional independences. In this work, we study this problem in the context of categorical probability theory by introducing a categorical definition of causal models, a categorical notion of d-separation, and proving an abstract version of the d-separation criterion. This approach has two main benefits. First, categorical d-separation is a very intuitive criterion based on topological connectedness. Second, our results apply in measure-theoretic probability (with standard Borel spaces), and therefore provide a clean proof of the equivalence of local and global Markov properties with causal compatibility for continuous and mixed variables.

Contents

1. Introduction

The d-separation criterion [18] is a necessary and sufficient condition for the compatibility of a probability distribution with a causal structure in the form of a directed acyclic graph (DAG). It states that a joint probability distribution of a collection of random variables is compatible with the DAG—in the sense that each of its nodes is one of the given variables, and each arrow denotes the possibility of causal influence—if and only if the distribution satisfies a list of conditional independence relations encoded in the structure of the DAG.

In this paper, we study this causal compatibility problem in the framework of *categorical probability theory*. We elaborate on the framework of generalized causal models recently proposed in [10], introduce a categorical notion of d-separation, and prove a categorical generalization of the d-separation criterion.

Department of Mathematics, Technikerstr. 13, A-6020 Innsbruck, Austria
Institute for Theoretical Physics, Technikerstr. 21a, A-6020 Innsbruck, Austria
E-mail addresses: `tobias.fritz@uibk.ac.at, andreas.klingler@uibk.ac.at`.
Date: July 11, 2022.
Acknowledgments. We thank Rob Spekkens for the original idea behind the definition of categorical d-separation and Patrick Forré for discussion and pointers to the literature. AK acknowledges support from the Stand Alone project P33122-N of the Austrian Science Fund (FWF).

The framework of generalized causal models involves freely generated categories called *free Markov categories* [10]. Starting from a set of morphisms as building blocks (representing the basic causal mechanisms), we construct a morphism in this category by assembling these blocks consistently. More precisely, a morphism in such a free category is a *string diagram* consisting of wires and boxes. In our context, these diagrams represent the causal models. Each wire corresponds to a local random variable and each box to a causal mechanism generating one or several new output variables from its input ones. A morphism in any Markov category is *compatible* with such a causal model if it can be decomposed in the form specified by the causal model; this generalizes the standard factorization definition of Bayesian networks [6].

Further, we define a categorical notion of d-separation in terms of the string diagrams. A causal structure displays d-separation of a collection of outputs $\mathcal{Z}$ if it becomes disconnected upon removing the corresponding wires in $\mathcal{Z}$. We show that this is equivalent to classical d-separation for the class of causal models for which the latter is defined, namely those that correspond to DAGs.

Finally, we prove an abstract version of the d-separation criterion for our categorical notions of causal model and d-separation. We show that a given distribution, possibly depending on additional input variables, is compatible with a causal structure if and only if it displays conditional independence between sets of variables whenever the causal model displays d-separation for the corresponding sets of wires. This result holds not just for "distributions" in the usual sense, but in the general sense of morphisms in any suitable Markov category $\mathbf{C}$. A central structural ingredient in the proof is the assumption that $\mathbf{C}$ has conditionals [9]. Intuitively, this property states that every Markov kernel with multiple output variables arises by generating one output variable after the other, namely as a function of the input variables and previously generated outputs. Such conditionals exist in the Markov category **FinStoch** which describes discrete random variables, but also for continuous or mixed random variables (technically for variables taking values in standard Borel spaces, but not arbitrary measurable spaces) as well as Gaussian variables. This implies that our definitions and results apply equally easily to all of these cases.

This approach leads to new insights into the structure of d-separation and generalizes the classical result, to the best of our knowledge, in two notable directions:

– It gives a criterion for compatibility of *Markov kernels* with a causal structure rather than probability distributions.[1] This is because Markov categories describing probabilities have Markov kernels as their morphisms, which therefore become the basic primitives of our formalism.

– It provides a uniform proof of the equivalence between d-separation and causal compatibility for discrete variables, continuous variables (or rather arbitrary variables taking values in standard Borel spaces) and Gaussian variables. This follows from the fact that we reason abstractly and even more generally in any Markov category with conditionals. This approach generalizes the d-separation criterion on the pure DAG-setting for continuous variables shown for distribution with a density [15, 14], which was later extended to arbitrary variables with values in standard Borel spaces [8].

This paper is organized as follows. In Section 2, we give a more detailed non-technical overview of the main concepts of this paper, including causal models as string diagrams and categorical d-separation. In Section 3, we recall and explain the definitions of Markov categories and gs-monoidal categories. In Section 4, we review the construction of free

[1] For a different approach to causal compatibility of Markov kernels we refer to [7].

gs-monoidal and free Markov categories, leading to generalized causal models and causal compatibility for morphisms in Markov categories. In Section 5, we present the notions of conditionals and conditional independence. In Section 6 we introduce a categorical notion of d-separation and prove the main results of this paper, namely the equivalence with classical d-separation (Proposition 21) and our abstract version of the d-separation criterion for causal compatibility (Theorem 27).

2. Causal models and Markov categories

In this paper, we study causal models, in the sense of Bayesian networks, from a categorical perspective. In order to make this accessible to readers without a formal background in category theory, we outline this paper's main concepts and results in this section.

Section 2.1 gives a non-technical introduction to the string-diagrammatic formalism representing causal models and its relation to the classical DAG formalism. We then present the concept of causal compatibility in Section 2.2 as a functorial property. In Section 2.3, we present categorical d-separation as a statement about the connectedness of the string diagram and explain our result on the d-separation criterion in Section 2.4. Therefore, the string diagrammatic approach opens the door for a new perspective on d-separation and its connection to causal compatibility.

2.1. Extending Bayesian networks with string diagrams.

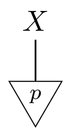
Traditionally, the definition of Bayesian networks relies on the concept of directed acyclic graphs (DAG), since such a graph encodes the underlying causal structure. To each node $v \in V(G)$ of a DAG G is associated a random variable X_v, and each directed arrow $w \to v$ is associated a direct possible causal dependence of the variable X_v on X_w. Formally, if we index the nodes by $1, \ldots, n$, then this means that a joint probability distribution P is compatible with the causal structure G if P factorizes into

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathsf{Pa}(X_i))$$

where $\mathsf{Pa}(X_i) = \{X_j : G \text{ contains the arrow } j \to i\}$ is the set of parents of X_i relative to G.

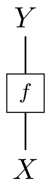
In the categorical framework, we can refine the notion of Bayesian networks via string diagrams, an idea that extends the categorical approach developed by Fong [6] and others [19, 20, 12, 13]. These diagrams arise naturally in *categorical probability*, which refers more generally to recent work on axiomatizing probability theory with simple, algebraic rules which avoid the low-level machinery of measure theory by hiding it in the proofs of the relevant axioms [9].

In categorical probability, a probability measure or probability distribution on a set or measurable space X is a morphism $p : I \to X$, depicted as



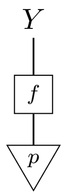
Here I is a singleton set that is not drawn in the diagram, the upper wire represents the probability space X, and the box p a probability distribution, considered now as a function with no input and one (random) output in X. Depending on the particular context, p can be a finite probability distribution, a Gaussian, or a probability measure on an arbitrary measurable space. A conditional probability distribution—also known as a

Markov kernel—is represented as a general morphism $f : X \to Y$, depicted as



where the wire X represents the input and Y the output space. The basic primitives in categorical probability are Markov kernels, of which probability distributions as above are special cases obtained by setting $X = I$.

These morphisms can be composed into new ones. For example, composing f with p, pictorially



gives rise to a new distribution of a random variable with values in Y. In the concrete setting, the composition is given by the Chapman-Kolmogorov equation, which reads for discrete distributions as

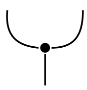
$$(f \circ p)(y) \coloneqq \sum_{x \in X} f(y|x) \cdot p(x)$$

and for Markov kernels on arbitrary measurable spaces as

$$(f \circ p)(A) \coloneqq \int_X f(A|x)\, p(\mathrm{d}x).$$

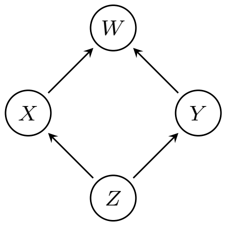
where $A \in \Sigma_Y$, with Σ_Y the σ-algebra of the measurable space Y.

Each random variable can be copied or marginalized over, for which there exists special string-diagrammatic notation:

 and 

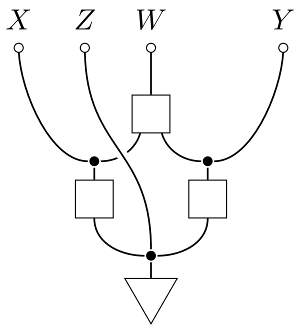
We formally axiomatize the semantics of the resulting string diagram calculus in Definition 1 as the definition of a *Markov category*. This definition consists of a set of rules which axiomatize the behaviour of Markov kernels.

Each flavour of probability has its own Markov category. There is a Markov category for discrete probabilities (called **FinStoch**), one for Gaussian probability (**Gauss**), one for probability theory on standard Borel spaces (**BorelStoch**), and one for probability theory on arbitrary measurable spaces (**Stoch**). But there are also other Markov categories in which the morphisms are not Markov kernels. These include the so-called *free Markov categories*, which form our framework for causal structures which generalize DAGs. Rather than Markov kernels, the morphisms in a free Markov category are the string diagrams themselves, i.e. all ”networks“ that can be built by wiring together a set of boxes, similar to how an electrical circuit is obtained by wiring together electrical components. In this way, string diagrams constitute generalized causal models. In particular, we will see that string diagrams can represent arbitrary DAG causal models. Consider for example this DAG:



(1)

As a string diagram, this causal structure looks like this:



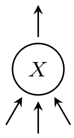
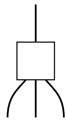
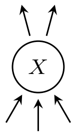
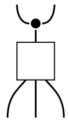
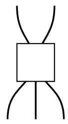
(2)

Each loose wire[2] represents a random variable where the wire's name indicates the variable's name. Further, each variable has a corresponding type, a placeholder for the measurable space in which the variable takes values. Unless necessary, we will not explicitly mention the type of each variable.

In our setting, some wires are connected to an ″output‶ representing a variable that is ″observed‶ rather than marginalized over, as indicated by the white dots. Note that every variable then becomes an output in at most one way. In certain situations we consider causal models where every wire is connected to an output. Throughout the paper, we call such diagrams *pure blooms* (see Definition 7). Unless stated otherwise we will denote the output by the name of the wire connected to the output. We will define certain operations, like d-separations, only with respect to output wires which highlights that we cannot address latent variables (i.e. wires that are not connected to an output) in general.

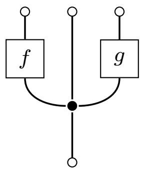
Causal models given by a DAG correspond to pure bloom string diagrams where each box has precisely one output wire, and the input wires represent the parents of the node in the DAG. The following table explains the relation between nodes in a DAG, boxes in the string diagram and the corresponding conditional probability distribution.

[2]Note that we use the term ″wire‶ as referring to an entire connected piece of circuitry, i.e. traversing a black dot in the diagram does not leave the wire.

	DAG	string diagram	conditional distribution
one output			$P(X\|ABC)$
ident. outputs			$P(XX\|ABC)$
diff. outputs	✗		$P(XY\|ABC)$

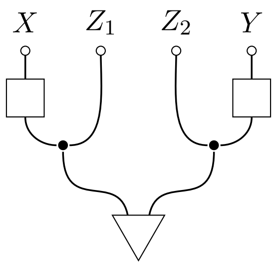
Using string diagrams themselves as generalized causal models allows us to go beyond the DAG approach in several directions:

– String diagrams in Markov categories describe Markov kernels instead of just probability distributions. Therefore, the string diagram language allows for modeling causal structures with inputs, such as



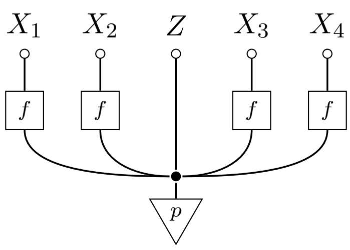
This describes a causal structure in which the input variable at the bottom does not have any particular distribution itself.

– As indicated in the table, boxes in the string diagram can have more than a single output wire. This allows for causal structures like



which are not native to the DAG framework (see Example 29(i) for a detailed discussion of this structure).

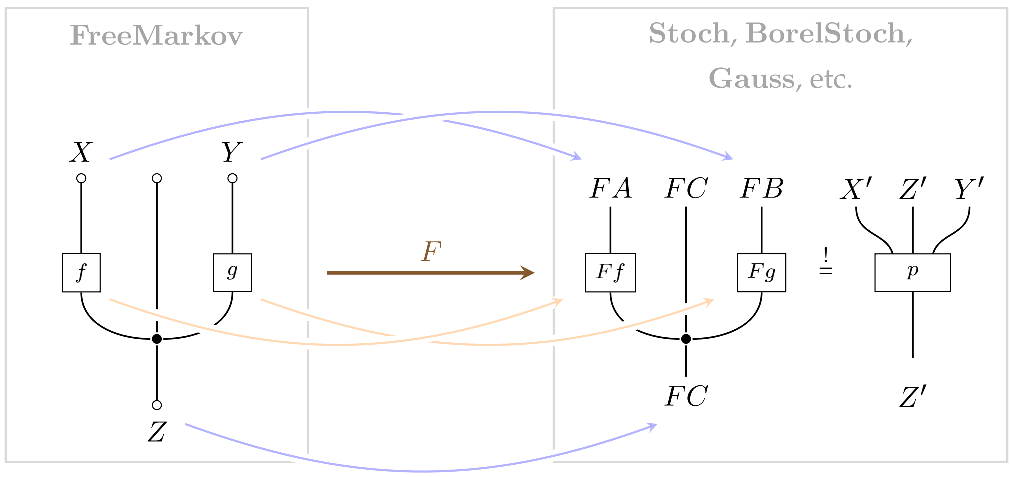
– String diagrams allow for the use of identical boxes multiple times. In particular, we can represent symmetric causal structures, for example,



represents a causal structure in which one random variable with distribution p causally influences four others with the additional constraint that the causal mechanism must be the same for all four. Further, in this situation the types of the variables $X_1, \ldots, X_4$ must be the same.

2.2. **Causal compatibility for Markov categories.** A distribution is compatible with a causal model if it can be written as a composition in precisely the way specified by the corresponding string diagram. In other words, every type W in the string diagram must be mapped to a concrete measurable space FW and every box f to a concrete Markov kernel Ff.

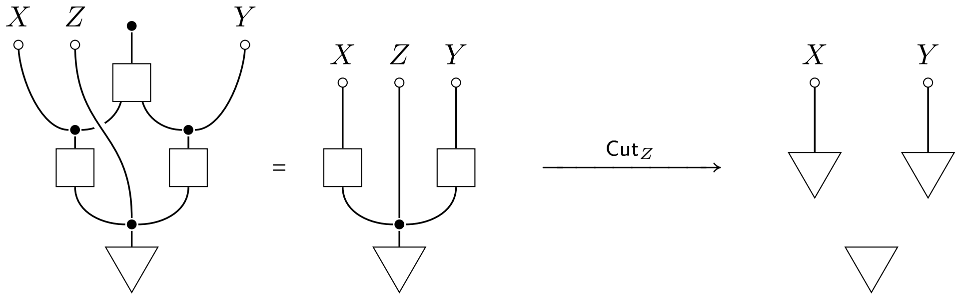
In the category-theoretic language, this is captured in the following way [6]. A morphism p in a concrete Markov category is compatible with a causal structure φ if and only if there exists a Markov functor F such that $p = F\varphi$. Intuitively, this functor acts in the following way:



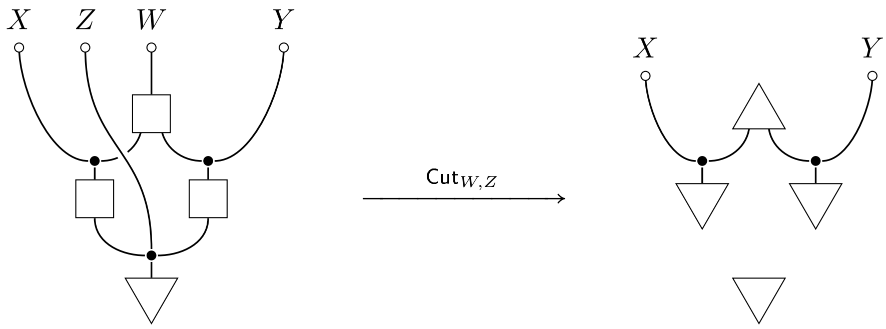
where p is the given morphism in a concrete Markov category and **FreeMarkov** the free Markov category whose morphisms capture the causal models. Further X has type A, Y has type B and Z has type C. So if the original p takes input from a measurable space Z' and outputs values in measurable spaces X', Z' and Y', then in particular the types must match in the sense that $FA = X'$ etc.

2.3. **Categorical d-separation.** The notion of d-separation for DAGs is a criterion relating conditional independence in causal models to the causal compatibility with a DAG. In Section 6.1, we introduce a categorical notion of d-separation from a different perspective. Although this notion looks different and much simpler than classical d-separation, we prove that it coincides with the classical one when considering causal models on DAGs.

An output wire Z *categorically d-separates* the output X from output Y if X and Y become disconnected upon *marginalizing* over all wires that are not involved in the d-separation relation and *removing* the wire Z. Consider, for example, the DAG in Equation (1). Z classically d-separates X from Y, as one can see based on the fact that the only paths between X and Y are the collider $X \to W \leftarrow Y$ and the fork $X \leftarrow Z \to Y$. In the corresponding string diagram, Equation (2), we witness categorical d-separation by first marginalizing over W, then removing the Z wire, and finally observing that X and Y are disconnected, pictorially:



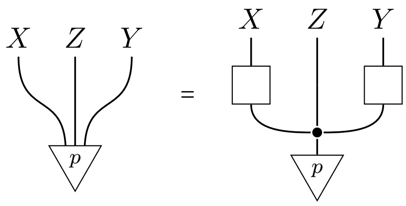
On the other hand, X is not d-separated from Y by W and Z due to the collider $X \to W \leftarrow Y$. In the string diagram, this is apparent since upon removing the wires Z and W,



X and Y are still connected.

2.4. ***d*-separation and causal compatibility.** For Bayesian networks, d-separation detects conditional independences for any compatible probability distribution. In particular, if two output wires X and Y are d-separated by the output wire Z, we have $X \perp Y \mid Z$.

Conditional independence in Markov categories is defined string diagrammatically. A probability distribution p on a product space $X \times Y \times Z$ displays the conditional independence $X \perp Y \mid Z$ if it can be written in the form



This reflects the classical notion of conditional independence; in the situation of finite probability distributions, it encodes the equation

$$P(X = x, Y = y, Z = z) \;=\; P(X = x|Z = z) \cdot P(Y = y|Z = z) \cdot P(Z = z).$$

But also for Gaussian random variables as well as measures on standard Borel spaces, one recovers the intuitive notions of conditional independence [9, Section 12]. Moreover, the diagrammatic definition is even sufficient to derive the semigraphoid properties [9, Lemma 12.5]. It also generalizes to a notion of conditional independence for morphisms with inputs (see Definition 14).

In Section 6.2, we prove that the categorical d-separation criterion applies to generalized causal models in Markov categories. For this reason, we define a notion of conditional independence which applies to arbitrary Markov kernels (Definition 14). We first prove the soundness of the d-separation, namely, if Z categorically d-separates X from Y, then $X \perp Y \mid Z$ (Corollary 25). Second, we prove the completeness of d-separation for causal compatibility: if a Markov kernel satisfies the global Markov property for a causal structure (i.e. every d-separated triple shows conditional independence), then the Markov kernel is compatible with the structure (see Theorem 27 for the precise assumptions and also the equivalence with the local Markov property).

A central assumption for the proof is the existence of conditionals (see Definition 11). Intuitively it says that the outputs of any morphism f can be generated successively while having access to all prior information. Conditionals exist in discrete probability, measure-theoretic probability on standard Borel spaces and in Gaussian probability, and this facilitates the application of our results to all of these cases, where the second case includes continuous random variables.

3. gs-monoidal and Markov categories

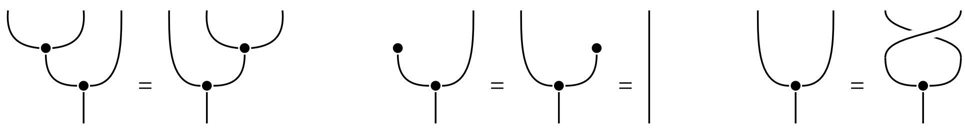
In the following, we recall the definitions of Markov categories and gs-monoidal categories. Markov categories are the basic structure modeling different flavors of probability, including discrete random variables, Gaussian random variables, continuous random variables on standard Borel spaces, or random variables on arbitrary measurable spaces. We assume some familiarity with symmetric monoidal categories and string diagrams.

The notion of gs-monoidal categories goes back to Gadducci's thesis [11, Definition 3.9] and an associated paper by Corradini and Gadducci [5]. There, it was considered with a different motivation in the context of term graphs and term graph rewriting.

Definition 1 (gs-monoidal category [11] and Markov category [9]).

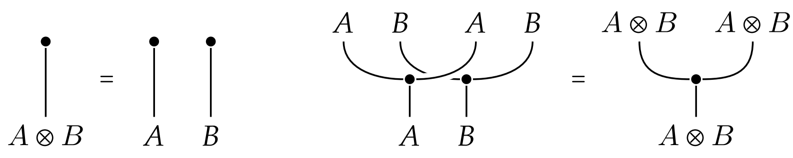
(*i*) *A* gs-monoidal category $\mathbf{C}$ *is a symmetric monoidal category with monoidal unit* I *equipped with a commutative comonoid structure for every object* $X \in \mathbf{C}$ *given by a counit* $\mathsf{del}_X : X \to I$ *and a comultiplication* $\mathsf{copy}_X : X \to X \otimes X$*. In the string diagrammatic notation, these operations are depicted as*

They are required to satisfy the commutative comonoid equations, diagrammatically given by



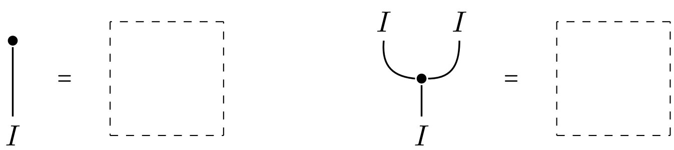
(3)

and to be compatible with the monoidal structure, i.e.



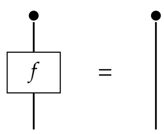
(4)

as well as



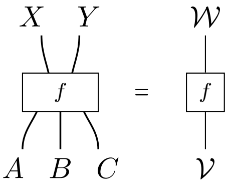
(5)

(*ii*) *A gs-monoidal category is called a* Markov category *if* del *is in addition natural, i.e. if for all morphisms* f,



(6)

We refer to [10, Remark 2.2] for more details on the multifarious history of these notions. When considering morphisms with multiple outputs, we often denote the collective inputs and outputs as sets instead of tensor products. For example, we write



where $\mathcal{V} = \{A, B, C\}$ and $\mathcal{W} = \{X, Y\}$. Modulo some abuse of notation, the order of the inputs and outputs is irrelevant since we can always permute the wires, and therefore it is enough to consider $\mathcal{V}$ and $\mathcal{W}$ as sets rather than lists or totally ordered sets.

The most important examples of Markov categories for probability theory are the following:

(a) **FinStoch** is the Markov category of finite stochastic maps. The objects are finite sets X, morphisms $p : I \to X$ are probability distributions and general morphisms $f : X \to Y$ are stochastic matrices.
(b) **Stoch** is the Markov category of arbitrary Markov kernels on measurable spaces. The objects are measurable spaces, morphisms $p : I \to X$ are probability measures, and general morphisms $f : X \to Y$ are measurable Markov kernels.
(c) **BorelStoch** is given by **Stoch** restricted to standard Borel spaces as objects.
(d) **Gauss** is the Markov category of Gaussian probability distributions. The objects are the spaces $\mathbb{R}^n$, morphisms $p : I \to \mathbb{R}^n$ are Gaussian probability measures and general morphisms $f : \mathbb{R}^n \to \mathbb{R}^m$ can be understood as stochastic maps
$$x \mapsto Ax + \xi$$
where A is any real $m \times n$ matrix and ξ is Gaussian noise. **Gauss** is a (non-full) subcategory of **BorelStoch**.

In each case, composition, the symmetric monoidal structure and the Markov category structure are the obvious ones. For more details and further examples we refer to [9]. In our context, gs-monoidal categories that are not Markov categories play more of an auxiliary role which we will detail below.

4. Free Markov categories and generalized causal models

Causal models are a framework for studying and modeling dependencies between random variables. In this section, we introduce such a framework in the language of Markov categories. We therefore investigate causal relationships independently of the particular notion of probability behind it (discrete, measure-theoretic, Gaussian, etc).[3]

Free Markov categories are the tailored notion for these purposes. These categories contain precisely all blueprints for causal networks that can be built from a bunch of given causal mechanisms. A morphism in this category is then what we call a *generalized causal model*.

The remainder of this section explains this in technical detail based on the formalism of free gs-monoidal categories and free Markov categories from [10][4]. This part is structured as follows: In Section 4.1, we introduce the category of hypergraphs. In Section 4.2 we use hypergraphs to define gs-monoidal string diagrams, free gs-monoidal categories,

[3]It is also possible to instantiate all of our definitions and results in a "possibilistic" rather than probabilistic setting. While this would be interesting e.g. in the context of non-determinism in computer science, the focus of the present work is on the probabilistic case.

[4]Another very similar construction of free gs-monoidal categories has been given independently in [16].

and subsequently free Markov categories. In Section 4.3 we introduce generalized causal models and causal compatibility of morphisms in arbitrary Markov categories.

4.1. **The category of hypergraphs.** A gs-monoidal string diagram, and therefore also a generalized causal model is defined as a hypergraph with extra structure. To define the relevant notion of hypergraph following [1], let first $\mathbf{I}$ be the category defined in the following way:

(i) The set of objects is given by $\{(k,\ell) \mid k,\ell \in \mathbb{N}\} \cup \{\star\}$.
(ii) Besides the identity morphisms, for every (k,ℓ) there are $k+\ell$ different morphisms

$$\mathsf{in}_1,\dots,\mathsf{in}_k,\mathsf{out}_1,\dots,\mathsf{out}_\ell : (k,\ell) \to \star,$$

and no other morphisms.

It is not necessary to specify a composition rule in $\mathbf{I}$, since no compositions except the trivial ones can be formed.

Definition 2. *A functor* $G : \mathbf{I} \to \mathbf{Set}$ *is called a* hypergraph. *Accordingly, we define the functor category*

$$\mathbf{Hyp} \coloneqq \mathbf{Set}^{\mathbf{I}}$$

to be the category of hypergraphs.

Intuitively, the functor G characterizes our common interpretation of (directed) hypergraphs in the following way:

(i) $W(G) \coloneqq G(\star)$ is the set of vertices, which we will call *wires*.
(ii) $B_{k,\ell}(G) \coloneqq G((k,\ell))$ is the set of hyperedges, which we will call *boxes*, with k inputs and ℓ outputs.
(iii) $G(\mathsf{in}_i)$ specifies the ith input wire of every box.
(iv) $G(\mathsf{out}_j)$ specifies the jth output wire of every box.

While the set of wires and boxes of a hypergraph may be infinite, the number of inputs and outputs of a box is always finite. We present a pictorial representation of a hypergraph in Figure 1.

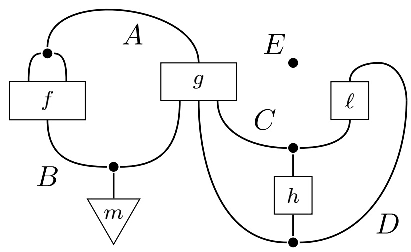


FIGURE 1. Pictorial representation of a hypergraph with wire set $\{A,B,C,D,E\}$ and box set $\{f,g,h,m,n\}$. For example, the box f has one input incident to the wire B and two outputs both incident to the wire A. The wire E is not incident to any box.

Given a box $b \in B_{k,\ell}(G)$ and a wire $A \in W(G)$, we define

$$\mathsf{in}(b,A) \coloneqq \left|\{j \in \{1,\dots,\ell\} : \mathsf{in}_j(b) = A\}\right|,$$
$$\mathsf{out}(b,A) \coloneqq \left|\{i \in \{1,\dots,k\} : \mathsf{out}_i(b) = A\}\right|.$$

Thus $\mathsf{in}(b, A)$ and $\mathsf{out}(b, A)$ counts the number of incoming or outgoing wires of type A in the box b. We also define the sets of inputs and outputs as

$$\mathsf{in}(b) \coloneqq \left\{\mathsf{in}_i(b) : i \in \{1, \dots, \ell\}\right\},$$
$$\mathsf{out}(b) \coloneqq \left\{\mathsf{out}_i(b) : i \in \{1, \dots, k\}\right\},$$

where repeated wires are counted only once.

Next, we analyze the morphisms in $\mathbf{Hyp}$. Since $\mathbf{Hyp}$ is a functor category, a morphism $\alpha : F \to G$ is precisely a natural transformation $\alpha : F \Rightarrow G$. Such a natural transformation is fully determined by its components

$$\alpha_\star : W(F) \to W(G) \quad \text{and} \quad \alpha_{(k,l)} : B_{(k,\ell)}(F) \to B_{(k,\ell)}(G) \quad \text{for all } k, \ell \in \mathbb{N}$$

satisfying naturality, i.e. commutativity of all diagrams of the form

$$\begin{array}{ccc} B_{k,\ell}(F) & \xrightarrow{\mathsf{in}_i} & W(F) \\ \downarrow{\scriptstyle \alpha_{k,\ell}} & & \downarrow{\scriptstyle \alpha_{k,\ell}} \\ B_{k,\ell}(G) & \xrightarrow{\mathsf{in}_i} & W(G) \end{array} \qquad \begin{array}{ccc} B_{k,\ell}(F) & \xrightarrow{\mathsf{out}_j} & W(F) \\ \downarrow{\scriptstyle \alpha_{k,\ell}} & & \downarrow{\scriptstyle \alpha_{k,\ell}} \\ B_{k,\ell}(G) & \xrightarrow{\mathsf{out}_j} & W(G). \end{array}$$

In other words, every natural transformation is a structure-preserving map between hypergraphs, i.e. if box f is incident to wire A in its ith input in the hypergraph F, then the same applies to their images with respect to α in the hypergraph G.

A hypergraph can contain an infinite set of wires and boxes. In the following we mainly restrict to *finite hypergraphs*, i.e. functors $G : \mathbf{I} \to \mathbf{Set}$ where $W(G)$ and

$$B(G) \coloneqq \coprod_{k,\ell \in \mathbb{N}} B_{k,\ell}(G).$$

are finite sets. We denote the corresponding full subcategory of $\mathbf{Hyp}$ by $\mathbf{FinHyp}$.

4.2. **gs-monoidal string diagrams and free Markov categories.** The pictorial representation of hypergraphs indicates their use for modeling causal structures in a categorical framework. In this subsection, we use hypergraphs in order to construct free Markov categories generated by a fixed hypergraph Σ. These are Markov categories in which the morphisms are string diagrams formed out of the boxes in Σ, or equivalently generalized causal models.

However, three apparent problems make hypergraphs not directly applicable to represent string diagrams:

(1) General hypergraphs might contain loops, for example in the sense that an output wire of a box may directly feed back as an input.
(2) While the splitting of a wire into two represents the copying of values and makes sense in any Markov category, the merging of wires as in Figure 1 does not make sense.
(3) A hypergraph in itself does not include any information about which wires are inputs or outputs of the overall diagram.

We resolve these issues by restricting to acyclic and left monogamous hypergraphs and by representing gs-monoidal string diagrams by cospans thereof:

Definition 3. *Let Σ be a hypergraph. A* gs-monoidal string diagram *for Σ is a cospan in the slice category $\mathbf{FinHyp}/\Sigma$ of the form*

$$\begin{array}{ccccc} & & G & & \\ & \overset{p}{\nearrow} & & \overset{q}{\nwarrow} & \\ \underline{n} & & & & \underline{m} \end{array}$$

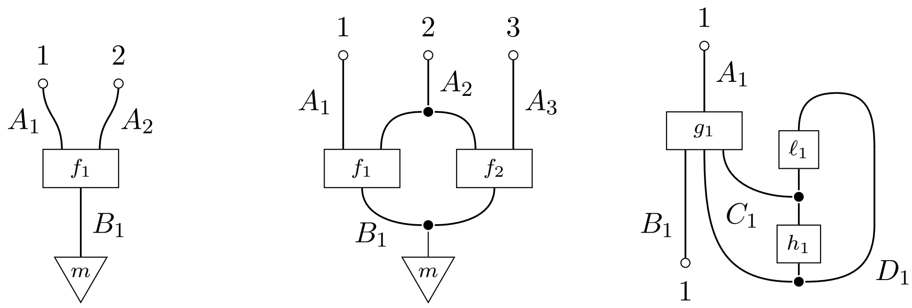


FIGURE 2. An example and two non-examples of gs-monoidal string diagrams where Σ is the hypergraph from Figure 1. The hypergraph morphisms to Σ are given by mapping the wires A_i to A in both examples as well as mapping the boxes labelled f_i to the f in Σ.

satisfying that:

(*i*) *G is* acyclic, *i.e. there is* no *family of wires* $A_0, \dots, A_{k-1} \in W(G)$ *and boxes* $f_0, \dots, f_{k-1} \in B(G)$ *such that*

$$\mathsf{in}(f_i, A_i) \geq 1 \quad \textit{and} \quad \mathsf{out}(f_i, A_{i+1}) \geq 1,$$

where the subscripts are modulo k.

(*ii*) *G is* left monogamous, *i.e. for every wire* $W \in W(G)$ *we have*

$$|p^{-1}(W)| + \sum_{f \in E(G)} \mathsf{out}(f, W) \leq 1.$$

By abuse of notation, we also write G for the underlying hypergraph of the object G in **FinHyp**/Σ, *and we write* $\mathsf{type} : G \to \Sigma$ *for the morphism that makes it into an object of* **FinHyp**/Σ.

In this definition, the *discrete hypergraph* $\underline{n}$ is defined to have $W(\underline{n}) = \{1, \dots, n\}$ and contains no boxes. Thus the morphism $p : \underline{n} \to G$ simply equips some wires with labels from $1, \dots, n$, thereby telling us which wires of G are input wires of the overall diagram and in which order. The other cospan leg $q : \underline{m} \to G$ similarly encodes the m output wires.

Pictorially, an acyclic hypergraph does not contain a family of wires which form a loop. Further, left monogamy requires that every wire in the hypergraph arises as either a global input or as an output of a box in precisely one way, ensuring that no "merging" of wires occurs. See Figure 2 for an illustration of all of this. The hypergraph morphisms $\mathsf{type} : G \to \Sigma$ are given by mapping the wires A_i to A in both examples as well as mapping the two distinct boxes f_i to the only morphism f in Σ, etc. The first and second hypergraphs are acyclic, while the third one is not. The first and third hypergraphs are left monogamous, while the second one is not since A_2 is an output of two boxes. Finally, we have $n = 0$ in the first two cases, so that the left cospan leg p is trivial, while the right leg q maps each number i to the ith overall output wire as counted from left to right.

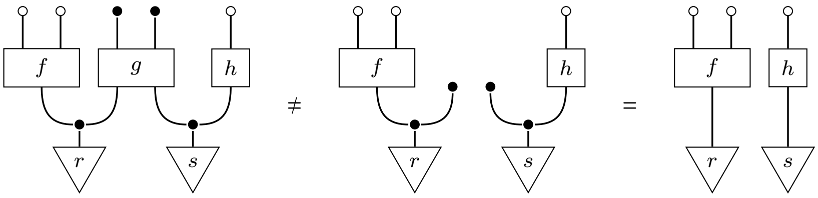
The notion of gs-monoidal string diagram is the main ingredient for constructing a gs-monoidal category whose morphisms are freely generated by the wires and boxes in a fixed hypergraph Σ. Indeed we can now define the category $\mathbf{FreeGS}_\Sigma$ as follows:

(i) Objects are all hypergraph morphisms $\sigma : \underline{n} \to \Sigma$ for $n \in \mathbb{N}$, or equivalently finite sequences of wires in Σ.

(ii) Morphisms are the isomorphism classes of gs-monoidal string diagrams.

Composition in $\mathbf{FreeGS}_\Sigma$ is defined by a pushout which which coincides with the way of stacking and connecting up drawings of string diagrams. The gs-monoidal structure likewise corresponds to the obvious operations in terms of string diagrams. We refer to [10] for details. In the following, we will not distinguish between a gs-monoidal string diagram and its isomorphism class.

$\mathbf{FreeGS}_\Sigma$ is typically not a Markov category. For example, the first step in the following simplification does not hold since the (cospans of) hypergraphs are not isomorphic, while the second equation does hold:



(7)

In the following, we define the free Markov category $\mathbf{FreeMarkov}_\Sigma$ by taking a quotient of $\mathbf{FreeGS}_\Sigma$ which enforces Equation (6), so that also the first equation above becomes true.

Definition 4. *Let*

$$\varphi \coloneqq \quad n \xrightarrow{\;p\;} G \xleftarrow{\;q\;} m \tag{8}$$

be a gs-monoidal string diagram.

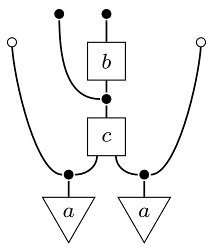
(*i*) *A box $b \in B(G)$ is called* eliminable *if each output of b gets discarded, i.e. if for every $W \in W(G)$ such that $\mathsf{out}(b, W) > 0$ we have*
  (*a*) $q^{-1}(W) = \emptyset$.
  (*b*) $\mathsf{in}(b', W) = 0$ *for every box $b' \in B(G)$.*

(*ii*) *φ is called* normalized *if it contains no eliminable boxes.*

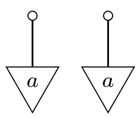
Every gs-monoidal string diagram has a normalized version obtained by iteratively applying the rule of Equation (6) to any eliminable box. Since every diagram is finite, this procedure terminates after finitely many steps, and we reach the normalized version. In addition, this diagram is unique since the order of elimination does not matter.

The free Markov category $\mathbf{FreeMarkov}_\Sigma$ is now defined just as $\mathbf{FreeGS}_\Sigma$, but with morphisms restricted to the normalized gs-monoidal string diagrams. The composition of morphisms is then defined as composition in $\mathbf{FreeGS}_\Sigma$ followed by normalization since the composition of two normalized diagrams need not be normalized. See [10] for details.

Example 5. The morphism

φ =


is not normalized, since the output of b gets discarded. Applying Equation (6), also the output of c get discarded. Therefore the normalization of φ is

$\mathsf{norm}(\varphi)$ =


△

In general, normalizing a gs-monoidal string diagram defines a strict gs-monoidal functor

$$\text{norm} : \mathbf{FreeGS}_\Sigma \to \mathbf{FreeMarkov}_\Sigma$$

that is identity-on-objects.

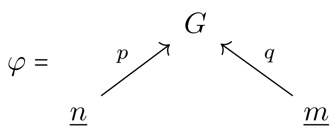
4.3. **Causal models and causal compatibility.** We now introduce the notion of a generalized causal model and define when a morphism in a Markov category is considered compatible with a generalized causal model.

Definition 6 ([10, Definition 7.1])**.** *Given a hypergraph Σ, a* generalized causal model *is a normalized gs-monoidal string diagram* (8) *over Σ such that q is injective.*

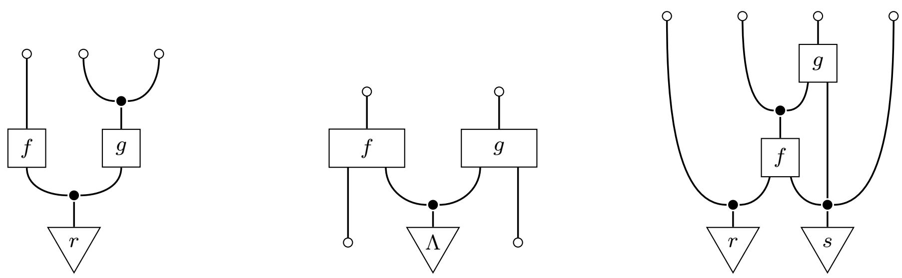
Intuitively, a generalized causal model is a morphism in $\mathbf{FreeMarkov}_\Sigma$ where the injectivity of q ensures that each wire is connected to at most one output. This lets us identify the global inputs and outputs with the wires in $W(G)$ (see Notation 9). In the traditional terminology of random variables, the injectivity of q guarantees that different outputs correspond to different variables. Figure 3 shows examples of generalized causal models.

One relevant subclass of generalized causal models are pure blooms. These morphisms represent causal models in which all variables are observed, i.e. every wire is an output in exactly one way, such as in Figure 3(C).

Definition 7 ([10])**.** *Let φ be a generalized causal model represented by a gs-monoidal string diagram*



Then φ is called pure bloom *if q is a bijection on wires.*



(A) Non-example of a generalized causal model.

(B) The Bell scenario.

(C) The instrumental scenario with every variable being observed.

FIGURE 3. (Non-)Examples of generalized causal models. While (B), (C) are generalized causal models, the string diagram (A) is not since the output wire of g is connected to two global outputs. Concerning Definition 7, the generalized causal model (B) is not a pure bloom since the output of Λ is not connected to a global output. The generalized causal model (C) is a pure bloom without inputs. For a further analysis of (C) regarding d-separation, see Example 29(i). In all three examples, we have $\Sigma = G$ and assume type to be the identity map for simplicity.

For an example of a pure bloom morphism we refer to Figure 3. We will show that the soundness of the d-separation criterion holds for arbitrary generalized causal models (see Corollary 25) while the completeness holds for pure bloom morphisms (see Theorem 27).

To define causal compatibility, we make the following assumption for the rest of the paper for convenience:

Assumption 8. *Throughout,* $\mathbf{C}$ *is a strict Markov category.*

Although most examples like **FinStoch**, **BorelStoch** or **Stoch** already fail strictness, this does not exclude these examples since we can always work with a strictification instead [9, Theorem 10.17], which satisfies Assumption 8. On the other hand, our free Markov categories $\mathbf{FreeMarkov}_\Sigma$ already satisfy this condition "on the nose". In any case, Assumption 8 is a useful convenience that holds without loss of generality.

Notation 9. *For the rest of the paper, we will assume that* φ *is a generalized causal model with*

$$\varphi \quad \coloneqq \quad \underline{n} \xrightarrow{\;p\;} G \xleftarrow{\;q\;} \underline{m}$$

which becomes a cospan in $\mathbf{FinHyp}/\Sigma$ *through* $\mathsf{type}: G \to \Sigma$.

We identity inputs and outputs with the wires they map to under p *and* q *and refer to them as such. In particular, we define*

$$\mathsf{in}(\varphi) \coloneqq p(\underline{n}) \subseteq W(\varphi) \tag{9}$$

$$\mathsf{out}(\varphi) \coloneqq q(\underline{m}) \subseteq W(\varphi) \tag{10}$$

for the set of all input/output wires. If φ *is a pure bloom morphism, then* $\mathsf{out}(\varphi) = W(\varphi)$.

Note that φ is a morphism

$$\varphi : \bigotimes_{i=1}^{n} \mathsf{type}(p(i)) \longrightarrow \bigotimes_{j=1}^{m} \mathsf{type}(q(j))$$

in $\mathbf{FreeMarkov}_\Sigma$.

In the following, we present the notion of causal compatibility for a generalized causal model φ. Intuitively, a morphism f in any Markov category $\mathbf{C}$ is compatible with φ if we can plug in a morphism from $\mathbf{C}$ into every box in $B(\Sigma)$ in such a way that the composite is exactly f, and such that the global input and output wires of φ correspond to a given tensor factorization of the domain and codomain of f:

Definition 10 (Compatibility)**.** *For* Σ *a hypergraph, let* φ *be a generalized causal model as in* (9)*. Let further*

$$f : \bigotimes_{i=1}^{n} W'_i \to \bigotimes_{j=1}^{m} V'_j$$

be a morphism in any Markov category $\mathbf{C}$ *satisfying Assumption 8, equipped with a fixed tensor decomposition of its domain and codomain as indicated.*

We call f compatible with φ *if there exists a strict Markov functor*[5] $F : \mathbf{FreeMarkov}_\Sigma \to \mathbf{C}$ *such that:*

(*i*) *We have*

$$W'_i = F(\mathsf{type}(p(i))), \qquad V'_j = F(\mathsf{type}(q(j))) \tag{11}$$

for all input indices $i = 1, \ldots, n$ *and output indices* $j = 1, \ldots, k$.

(*ii*) $f = F(\varphi)$.

[5]i.e. a strict symmetric monoidal functor which preserves the comonoid structure.

This generalizes the functorial definition of causal model as first studied by Fong [6].

Note that the functor F must assign to every wire type (i.e., wire in Σ) a corresponding object in the category $\mathbf{C}$. This implies that wires in $W(G)$ with identical types must map to the same object in $\mathbf{C}$. For example, one may consider a situation in which f is a probability distribution with no inputs and all output variables are real-valued. In this case, we would have $V'_j = \mathbb{R}$ for all j, and one may want to consider a causal model φ in which all wires are likewise of the same type.

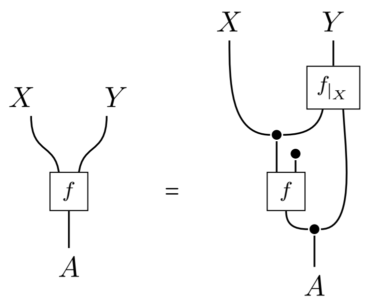
Similarly, the hypergraph morphism $\mathsf{type} : G \to \Sigma$ assigns to each box in G a specific "type" box in Σ. This means that under F, any two boxes with the same type must map to the same morphism in $\mathbf{C}$. This is why generalized causal models in our sense can naturally incorporate the condition that several causal mechanism must be the same, namely when one chooses the types in a way which enforces this.

In the following, we denote for every wire $X \in W(G)$ in φ the corresponding object $F(\mathsf{type}(X))$ in $\mathbf{C}$ by X'. Similarly, for every set of wires $\mathcal{W} \subseteq W(G)$ in φ, we denote the corresponding multiset of wires in $\mathbf{C}$ by $\mathcal{W}'$. For the rest of the paper, we will associate this multiset with the corresponding tensor product in $\mathbf{C}$ obtained by tensoring its elements, where we ignore the question of how to order the factors.

5. Conditionals and conditional independence

5.1. **Existence of conditionals.** To use the d-separation criterion to detect causal compatibility, we need in addition the existence of conditionals. This notion has been studied in categorical terms in [3] in a special case, where the authors call it *admitting disintegration*, and subsequently in [9] in general. In the following, we briefly review the definition following [9, Section 11].

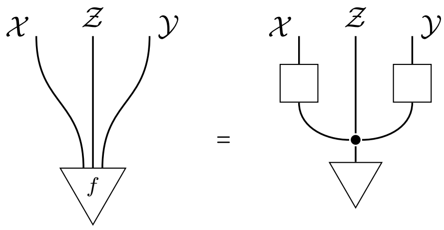
Definition 11. *Let* $\mathbf{C}$ *be a Markov category. We say that* $\mathbf{C}$ *has* conditionals *if for every morphism* $f : A \to X \otimes Y$, *there is a morphism* $f_{|X}$ *such that*



Examples of categories having conditionals are **FinStoch**, **Gauss** as well as **BorelStoch**. In contrast, **Stoch** does not have conditionals (see [9, Examples 11.6–11.8] and references therein).

5.2. **Conditional independence.** A second ingredient of the d-separation criterion is the definition of conditional independence. The following definition has been introduced in several works (see for example [3, Section 6] or [4] in a different setup) and shown to satisfy the well-known semigraphoid properties. In addition, [9] shows that it is still meaningful to define conditional independence in the absence of conditionals.

Definition 12. *A morphism* $f : I \to \mathcal{X} \otimes \mathcal{Z} \otimes \mathcal{Y}$ *in* $\mathbf{C}$ *displays the* conditional independence $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ *if it can be written as*



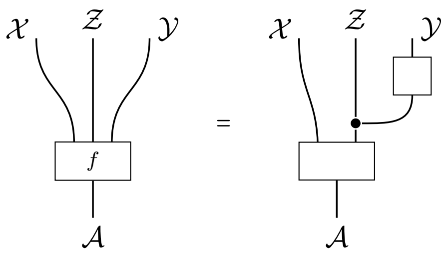
In other words, $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ holds if f is compatible with the generalized causal model which corresponds to the string diagram on the right-hand side (where all boxes are of distinct type, and we leave labels off for simplicity).

Remark 13. The conditional independence $\mathcal{X} \perp \emptyset \mid \mathcal{Z}$ is equivalent to the existence of the conditional $r_{|\mathcal{Z}}$. Therefore, if $\mathbf{C}$ has conditionals, then every state r satisfies $\mathcal{X} \perp \emptyset \mid \mathcal{Z}$ for every tensor factorization of its codomain $\mathcal{X} \otimes \mathcal{Z}$. △

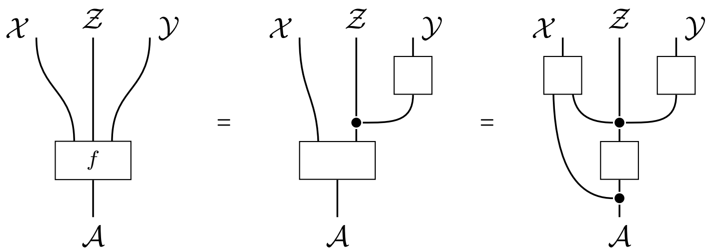
As we have already done in the previous definition, we write $\mathcal{X}, \mathcal{Y}, \ldots$ for arbitrary lists of objects in $\mathbf{C}$. We also allow for implicit reordering of these lists—effectively identifying these lists with multisets—and omit mention of the relevant compositions of f by swap morphisms.[6] This allows us to talk about conditional independence with respect to any tripartition of the tensor factors in the codomain of any state f.

With this in mind, we now introduce a notion of conditional independence for morphisms with inputs. This notion is the key ingredient of the d-separation criterion for pure bloom causal models presented in Section 6.2, and it is the categorical generalization of the *transitional conditional independence* introduced recently by Forré [7, Definition 3.1][7].

Definition 14. *A morphism* $f : \mathcal{A} \to \mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}$ *in* $\mathbf{C}$ *displays the* conditional independence $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ *if there exists a factorization of the form*



Remark 15. Note that the above definition of conditional independence is not symmetric, i.e. $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ does not necessarily imply $\mathcal{Y} \perp \mathcal{X} \mid \mathcal{Z}$. If $\mathbf{C}$ has conditionals then $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ rewrites to



which highlights the asymmetry. Moreover, if $\mathcal{A}$ is trivial, then the conditional independence coincides with Definition 12.

[6] Of course, these kinds of bookkeeping mechanisms are also present in the traditional notation of probability distributions and measures, though rarely made explicit in that context.

[7] Note that Forré's definition has the roles of $\mathcal{X}$ and $\mathcal{Y}$ swapped.

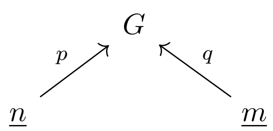
Due to the asymmetry, the outputs in $\mathcal{X}$ might contain information about the inputs $\mathcal{A}$ which cannot be retrieved just from the outputs in $\mathcal{Z}$. On the other hand, the outputs in $\mathcal{Y}$ are generated using only the information from the outputs in $\mathcal{Z}$. The local Markov property that we will use in Definition 26 explicitly highlights this asymmetry: the output of a box (corresponding to $\mathcal{Y}$) is independent of its non-descendants ($\mathcal{X}$) given its inputs ($\mathcal{Z}$). Every global input is non-descendant to any box; however, not every global input wire is directly an input of the box itself. △

6. Deciding causal compatibility with d-separation

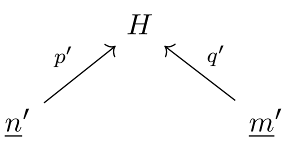
The main goal of this section is to prove that the d-separation criterion [18, Section 1.2.3] correctly detects causal compatibility not just in discrete probability but in all Markov categories with conditionals. To this end, we introduce a categorical notion of d-separation phrased in terms of connectedness of the gs-monoidal string diagram representing the causal model. We then show that this notion coincides with the classical notion of d-separation whenever the latter applies.

This part is structured as follows. In Section 6.1, we introduce the categorical notion of d-separation on generalized causal models. Moreover, we show in Proposition 21 that this notion coincides with the classical notion of d-separation for all those generalized causal models that correspond to DAGs. In Section 6.2, we first show that d-separation implies conditional independences for compatible morphisms in Markov categories with conditionals. We then show in Theorem 27 that d-separation fully characterizes causal compatibility.

6.1. **Categorical d-separation.** For a gs-monoidal string diagram

$\varphi =$ 

and a set of output wires $\mathcal{Z} \subseteq \mathsf{out}(\varphi)$, we define a new gs-monoidal string diagram $\mathsf{Cut}_{\mathcal{Z}}(\varphi)$ obtained by removing the wires in $\mathcal{Z}$ in the following sense. Its underlying hypergraph H is such that the set of boxes is the same, $B(H) \coloneqq B(G)$, while the set of wires is $W(H) \coloneqq W(G) \setminus \mathcal{Z}$. Each box has the same input and output wires as before, expect in that those in $\mathcal{Z}$ are simply removed, which lowers the arities of the boxes correspondingly. We also remove all occurrences of wires in $\mathcal{Z}$ from the global inputs and outputs, and this results in a gs-monoidal string diagram

$\mathsf{Cut}_{\mathcal{Z}}(\varphi) \coloneqq$ 

Note that $\mathsf{Cut}_{\mathcal{Z}}(\varphi)$ is generally not a morphism in $\mathbf{FreeMarkov}_{\Sigma}$ anymore since it is not normalized. However, it can be understood as a morphism in $\mathbf{FreeGS}_{H}$. Example 18 will present a few examples.

We next introduce some notation for paths of wires.

Definition 16. *Let φ be a gs-monoidal string diagram in* $\mathbf{FreeGS}_{\Sigma}$.

(*i*) *An* undirected path *between two wires $X, Y \in W(G)$ is a sequence of wires*

$$X = W_1, W_2, \dots, W_n, W_{n+1} = Y$$

together with a sequence of boxes $b_1, \dots, b_n \in B(G)$ *such that*

$$\mathsf{in}(b_i, W_i) + \mathsf{out}(b_i, W_i) \geq 1 \quad \textit{and} \quad \mathsf{in}(b_i, W_{i+1}) + \mathsf{out}(b_i, W_{i+1}) \geq 1.$$

If there exists an undirected path between X *and* Y*, then we write* $X - Y$*.*

(*ii*) *For two wires* $A, B \in W(G)$*, we write* $A \to B$ *if there exists a box* $b \in B(G)$ *such that*

$$\mathsf{in}(b, A) = 1 \quad \textit{and} \quad \mathsf{out}(b, B) = 1. \tag{12}$$

(*iii*) *For two wires* $A, B \in W(G)$*, we write* $A \twoheadrightarrow B$ *if there exists a sequence of wires* $W_1, \dots, W_n \in W(G)$ *such that*

$$A \to W_1 \to \dots \to W_n \to B. \tag{13}$$

Thus, an undirected path in φ may traverse a box not just from input or output or vice versa, but also from input to input or output to output.

The intuitive idea behind the following definition of d-separation, as already briefly discussed at [10, Remark 7.2], was communicated to us by Rob Spekkens.

Definition 17 (Categorical d-separation)**.** *Let* φ *be a generalized causal model. For three disjoint sets of output wires* $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathsf{out}(\varphi)$*, we say that* $\mathcal{Z}$ d-separates $\mathcal{X}$ *and* $\mathcal{Y}$ *if*

$$\mathsf{Cut}_{\mathcal{Z}}(\varphi_{\mathcal{X},\mathcal{Y},\mathcal{Z}})$$

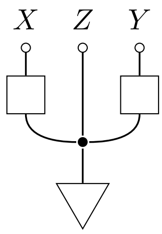
has no undirected path between any output in $\mathcal{X}$ *and any output in* $\mathcal{Y}$*.*

Here, $\varphi_{\mathcal{W}} \coloneqq \mathsf{norm}(\mathsf{del}_{\mathcal{W}^c} \circ \varphi)$ denotes the marginal on $\mathcal{W} \subseteq \mathsf{out}(\varphi)$ in $\mathbf{FreeMarkov}_\Sigma$. The absence of an undirected path as in the definition manifests itself in the string diagrams simply as topological disconnectedness.

Example 18. The following examples constitute the basic components of "classical" d-separation and illustrate the simplicity of categorical d-separation. In all cases, the unlabeled boxes denote distinct generators, i.e. distinct boxes in the generating hypergraph Σ.

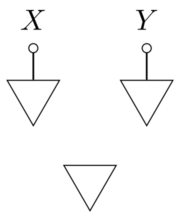
(i) Fork: consider the morphism

φ =



Then Z d-separates X from Y, since

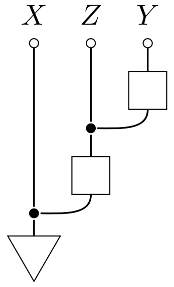
$\mathsf{Cut}_Z(\varphi)$ =



has disconnected X and Y.

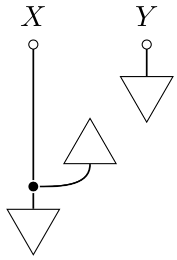
(ii) Chain: consider the morphism

φ =



Then Z d-separates X from Y, since

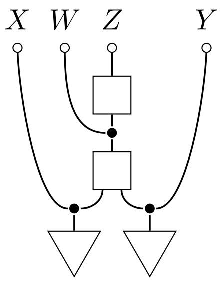
$\mathsf{Cut}_Z(\varphi)$ =



has disconnected X and Y.

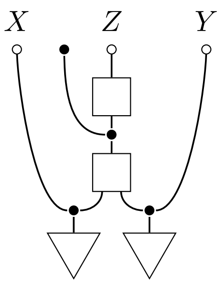
(iii) Collider: consider the morphism

φ =

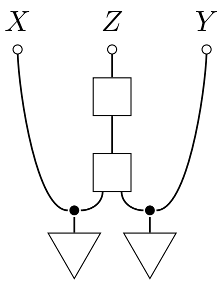


Then Z does *not* d-separate X from Y, since we have

$\varphi_{X,Z,Y}$ =

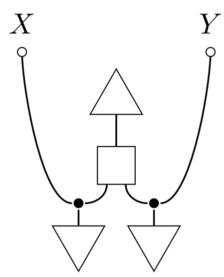


=

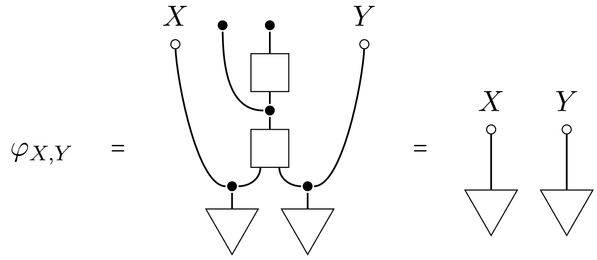


and therefore

$\mathsf{Cut}_Z(\varphi_{X,Z,Y})$ =

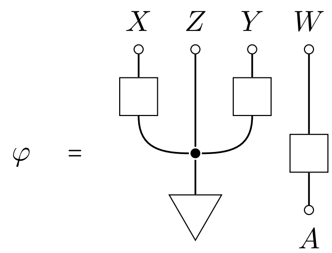


which still contains an undirected path $X - Y$. The same reasoning applies when $\mathcal{Z} = \{W\}$ or $\mathcal{Z} = \{W, Z\}$. However, if $\mathcal{Z} = \emptyset$, then

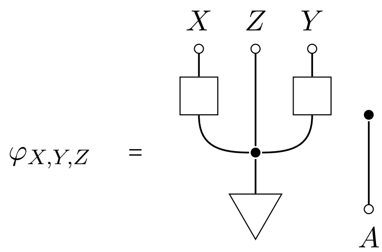


which has disconnected X and Y. Therefore $\emptyset$ d-separates X and Y.

(iv) Consider the morphism



The normalized marginal $\varphi_{X,Y,Z}$ is given by



which again shows that Z d-separates X from Y, since cutting Z makes X and Y disconnected. △

In order to define the classical notion of d-separation, we note that every gs-monoidal string diagram has an underlying DAG, given by using wires as nodes and taking the edges to be → as in Definition 16. We use the term *DAG path* to refer to an undirected path in this DAG, i.e. to a sequence of wires connected by boxes from input to output or vice versa (but *not* from input to input or output to output). We also define the ancestor wires of a given set of wires as

$$\mathsf{An}(\mathcal{X}) = \{U \in W(G) : \exists X \in \mathcal{X} \text{ such that } U \twoheadrightarrow X\}.$$

and the set of descendant wires as

$$\mathsf{Dec}(\mathcal{X}) = \{U \in W(G) : \exists X \in \mathcal{X} \text{ such that } X \twoheadrightarrow U\}.$$

Note that $\mathcal{X} \subseteq \mathsf{An}(\mathcal{X}), \mathsf{Dec}(\mathcal{X})$. To state the following classical definition [18, Definition 1.2.3] in our language, we restrict further to those generalized causal models that are determined by their underlying DAGs. In a causal structure as represented by a DAG, it is (implicitly) assumed that every node or variable has its own causal mechanism associated with it; in our framework, this means that every box has exactly one output. Moreover, DAGs have no global inputs which implies in our framework that $\mathsf{in}(\varphi) = \emptyset$.

Definition 19 (Classical d-separation)**.** *Let φ be a pure bloom causal model with $\mathsf{in}(\varphi) = \emptyset$ and such that every box has exactly one output. Then:*

(*a*) *A DAG path p in φ is called d-*separated *by a set of wires $\mathcal{Z} \subseteq \mathsf{out}(\varphi)$ if:*

(i) *p contains a chain $W \to Z \to U$ or a fork $W \leftarrow Z \to U$ for some $Z \in \mathcal{Z}$.*
(ii) *p contains a collider $W \to M \leftarrow U$ where $M \notin \mathsf{An}(\mathcal{Z})$.*

(b) *$\mathcal{X}$ is d-separated from $\mathcal{Y}$ by $\mathcal{Z}$ if every DAG path between every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ is d-separated by $\mathcal{Z}$.*

We will now prove the equivalence of categorical d-separation with classical d-separation for the class of causal models on which the latter is defined. This requires some preparation.

Lemma 20. *Let φ be a pure bloom causal model, $b \in B(G)$ a box in φ and $\mathcal{W} \subseteq \mathsf{out}(\varphi)$ a subset of its wires. The following statements are equivalent:*

(i) $\mathsf{out}(b) \cap \mathsf{An}(\mathcal{W}) = \emptyset$.
(ii) *b gets discarded in $\varphi_{\mathcal{W}} = \mathsf{norm}(\mathsf{del}_{\mathcal{W}^c} \circ \varphi)$.*

Proof. (ii) $\Longrightarrow$ (i): To prove the contrapositive, assume $\exists A \in \mathsf{out}(b)$ such that $A \in \mathsf{An}(\mathcal{W})$. Then there is a path $A \twoheadrightarrow W$ with $W \in \mathcal{W}$. Since W is still an overall output that does not get discarded, this path is still valid in $\mathsf{del}_{\mathcal{W}^c} \circ \varphi$. Therefore b remains in $\mathsf{norm}(\mathsf{del}_{\mathcal{W}^c} \circ \varphi)$.

(i) $\Longrightarrow$ (ii): Consider the set $\mathsf{Dec}(\mathsf{out}(b))$. By assumption, we have that $\mathsf{Dec}(\mathsf{out}(b)) \cap \mathsf{An}(\mathcal{W}) = \emptyset$. We show that the box b gets discarded in $\mathsf{norm}(\mathsf{del}_{\mathsf{Dec}(\mathsf{out}(b))} \circ \varphi)$, which is enough because of $\mathcal{W}^c \supseteq \mathsf{Dec}(\mathsf{out}(b))$. By definition of $\mathsf{Dec}(\mathsf{out}(b))$, there exists a final box[8] $\hat{b}$ such that $\mathsf{out}(\hat{b}) \subseteq \mathsf{Dec}(\mathsf{out}(b))$. This shows that $\hat{b}$ gets discarded in $\mathsf{norm}(\mathsf{del}_{\mathsf{Dec}(\mathsf{out}(b))} \circ \varphi)$.

Define $\widetilde{\varphi} \coloneqq \mathsf{norm}(\mathsf{del}_{\mathsf{Dec}(\mathsf{out}(\hat{b}))} \circ \varphi)$. Repeating the above procedure with $\widetilde{\varphi}$, we arrive after finite number of steps at b itself being a final box. Since it is then eliminable after composing with $\mathsf{del}_{\mathsf{Dec}(\mathsf{out}(b))}$, it indeed no longer appears in the normalization. □

We can now show the promised equivalence result between categorical d-separation and classical d-separation in the cases where φ represents a causal structure given by a DAG, i.e. $\mathsf{in}(\varphi) = \emptyset$ and every box has a single output.

Proposition 21. *Both concepts of d-separation coincide on pure bloom causal models φ with $\mathsf{in}(\varphi) = \emptyset$ and in which every box has exactly one output.*

Proof. To make the proof more intuitive, we introduce the term *d-connected* as the negation of d-separated (in either version).

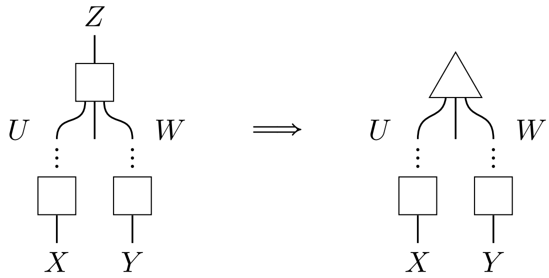
We start by showing that classical d-connectedness implies categorical d-connectedness. Let p be a DAG path between some $X \in \mathcal{X}$ and some $Y \in \mathcal{Y}$ which witnesses that $\mathcal{Z}$ makes $\mathcal{X}$ and $\mathcal{Y}$ be d-connected in the classical sense, which means that the following hold:

(i) For every chain $W \to M \to U$ or fork $W \leftarrow M \to U$ in p, we have $M \notin \mathcal{Z}$.
(ii) For every collider $W \to M \leftarrow U$ in p, we have $M \in \mathsf{An}(\mathcal{Z})$.

For simplicity, we also assume without loss of generality that p contains only one wire twice from $\mathcal{X}$ and $\mathcal{Y}$ each, say X and Y respectively. Then this p can also be interpreted as an undirected path in φ, but generally not in $\varphi_{\mathsf{cut}} \coloneqq \mathsf{Cut}_{\mathcal{Z}}(\varphi_{\mathcal{X},\mathcal{Y},\mathcal{Z}})$, since it may traverse wires that are not in φ_{cut}. However, we now show that there still is an undirected path p' between X and Y in φ_{cut}. By the above assumption (i), if p contains a wire $Z \in \mathcal{Z}$, then it has to arise from a collider $U \to Z \leftarrow W$ in p. Removing wire Z from p still defines a valid

[8] As defined in [10], a final box is one whose outputs are global outputs of φ without further copy or discard. Such a box always exists since φ is pure bloom and normalized (compare [10, Lemma 4.6]).

undirected path between X and Y, pictorially:



We prove that the path p' obtained by removing all wires in $\mathcal{Z}$ from p like this is an undirected path in φ_{cut}, which implies categorical d-connectedness. To this end, it only remains to show that each wire in p' is an existing wire in φ_{cut}, which we do as follows:

(i) X and Y themselves are still in φ_{cut}.
(ii) Every $Z \in \mathcal{Z}$ in p is part of a collider $U \to Z \leftarrow W$ as above, so that $U, W \in \mathsf{An}(\mathcal{Z})$. This implies that U and W survive in $\varphi_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$ by Lemma 20.
(iii) Since U and W are themselves either the middle node in a chain or fork or the start or end of p, we have $U, W \notin \mathcal{Z}$. This implies $U, W \in \mathsf{An}(\mathcal{Z}) \setminus \mathcal{Z}$, and therefore U and W survive also in φ_{cut}.
(iv) For every chain $W \to M \to U$ in p, if U survives in φ_{cut}, then so does M (since it survives in $\varphi_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$ and $M \notin \mathcal{Z}$).
(v) For every fork $W \leftarrow M \to U$ in p, if U or W survives in φ_{cut}, then so does M (since it survives in $\varphi_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$ and $M \notin \mathcal{Z}$).

Since the wires in p' are exactly those of p minus some of the colliders, we can start with the first two observations and then apply the latter two repeatedly on any segment bounded by colliders or the starting node X or the final node Y in order to conclude that all wires in p' are present in φ_{cut}. This concludes one direction of the proof.

The converse direction of showing that categorical d-connectedness implies classical d-connectedness works similarly. Let p be an undirected path between $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ in φ_{cut}. We assume without loss of generality that all wires in p are distinct and that X and Y are the only elements of $\mathcal{X}$ and $\mathcal{Y}$ in p. Furthermore, we also assume without loss of generality that p is of the form

$$X \twoheadleftarrow A - B \twoheadrightarrow Y, \tag{14}$$

where no wire that is in between A and B is in $\mathsf{An}(\mathcal{X})$ or $\mathsf{An}(\mathcal{Y})$, or equivalently that every wire in p that is in $\mathsf{An}(\mathcal{X})$ is directly reached from X by output-to-input traversals in p, and similarly for all wires in $\mathsf{An}(\mathcal{Y})$. This property can be achieved by taking every wire in p in $\mathsf{An}(\mathcal{X})$ for which this is not the case and replacing the path from X to it by a sequence of output-to-input traversals, and similarly for every wire in $\mathsf{An}(\mathcal{Y})$. Note that this replacement may involve changing the starting and ending wires X and Y as well. In order to turn p into a DAG path p' that witnesses classical d-separation, we need to remove all direct input-to-input traversals of a box in p; direct output-to-output traversals cannot occur due to the assumption that every box has exactly one output. We can hence simply add to p the unique output wire of every box that has an input-to-input traversal in p, and we obtain a valid DAG path p'.

It remains to verify the conditions on chains, forks and colliders. Clearly p' does not contain any chain $W \to Z \to U$ or fork $W \leftarrow Z \to U$ with $Z \in \mathcal{Z}$, since such a configuration cannot occur in p to begin with. For a collider $W \to M \leftarrow U$, the unique box which outputs M must be contained in $\varphi_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$, and therefore be in $\mathsf{An}(\mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z})$ by Lemma 20. However,

M being in $\mathsf{An}(\mathcal{X})$ or $\mathsf{An}(\mathcal{Y})$ violates the assumption that p is of the form (14). Therefore M has to be in $\mathsf{An}(\mathcal{Z})$, showing the collider condition (ii). □

We record one more observation on categorical d-separation for further use below.

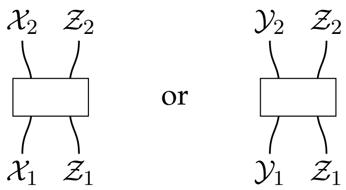
Lemma 22. *Let φ be a pure bloom causal model and $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathsf{out}(\varphi)$ a partition of all output wires such that $\mathcal{Z}$ categorically d-separates $\mathcal{X}$ and $\mathcal{Y}$. Then every box $b \in B(G)$ in φ satisfies at least one of the cases:*

(*i*) $\mathsf{in}(b), \mathsf{out}(b) \subseteq \mathcal{X} \cup \mathcal{Z}$.
(*ii*) $\mathsf{in}(b), \mathsf{out}(b) \subseteq \mathcal{Y} \cup \mathcal{Z}$.

Proof. If there exist $Y \in \mathcal{Y} \cap \mathsf{out}(b)$ and $X \in \mathcal{X} \cap \mathsf{out}(b)$, then these wires are still in the output of b in φ_{cut}, and this contradicts the assumed disconnectedness of φ_{cut} with respect to $\mathcal{X}$ and $\mathcal{Y}$. Since $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ form a partition, this shows that either $\mathsf{out}(b) \subseteq \mathcal{X} \cup \mathcal{Z}$ or $\mathsf{out}(b) \subseteq \mathcal{Y} \cup \mathcal{Z}$.

Proving $\mathsf{out}(b) \subseteq \mathcal{X} \cup \mathcal{Z} \Rightarrow \mathsf{in}(b) \subseteq \mathcal{X} \cup \mathcal{Z}$ and $\mathsf{out}(b) \subseteq \mathcal{Y} \cup \mathcal{Z} \Rightarrow \mathsf{in}(b) \subseteq \mathcal{Y} \cup \mathcal{Z}$ works similarly, and this then proves the statement. □

Pictorially, Lemma 22 shows that if $\mathcal{Z}$ d-separates $\mathcal{X}$ and $\mathcal{Y}$, then every box b in φ is of the form



where $\mathcal{X}_i \subseteq \mathcal{X}$, $\mathcal{Y}_i \subseteq \mathcal{Y}$, $\mathcal{Z}_i \subseteq \mathcal{Z}$.

6.2. **Causal compatibility.** In the following, we show that d-separation implies conditional independence for any generalized causal model. We first prove this result for a partition of wires in a pure bloom causal model in Lemma 23. We then refine it to any disjoint collection of wires in Corollary 25 in any generalized causal model. Finally, we show in Theorem 27 that d-separation fully characterizes causal compatibility for pure bloom causal models in all Markov categories with conditionals.

Throughout, we also use the following convenient notation: If a morphism f in $\mathbf{C}$ is compatible with a causal model φ in the sense of Definition 10, then we refer to the wires of φ to indicate conditional independence instead of the objects in the tensor factorization of f. In other words, instead of writing $\mathcal{X}' \perp \mathcal{Y}' \mid \mathcal{Z}'$, we simply write $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$. Here, each $W' = F(\mathsf{type}(W))$ is the object in $\mathbf{C}$ associated to the wire W by the causal model functor F (see Definition 10).

Lemma 23. *Let $\mathbf{C}$ be a strict Markov category with conditionals, and let φ be a pure bloom causal model. Further, let $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathsf{out}(\varphi)$ be a partition of wires in φ such that $\mathsf{in}(\varphi) \subseteq \mathcal{X} \cup \mathcal{Z}$ and $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$.*

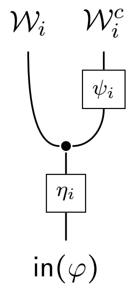
If a morphism f in $\mathbf{C}$ is compatible with φ, then $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ (as in Definition 14).

Proof. Choose a total ordering of all boxes $b_1, \ldots, b_{k-1} \in B(G)$ and a chain of sets of wires in $\mathsf{out}(\varphi)$,

$$\mathsf{in}(\varphi) = \mathcal{W}_1 \subseteq \ldots \subseteq \mathcal{W}_k = \mathsf{out}(\varphi),$$

such that $\mathcal{W}_{i+1} = \mathsf{out}(b_i) \cup \mathcal{W}_i$ and $\mathsf{An}(\mathcal{W}_i) = \mathcal{W}_i$. Note that there is a factorization

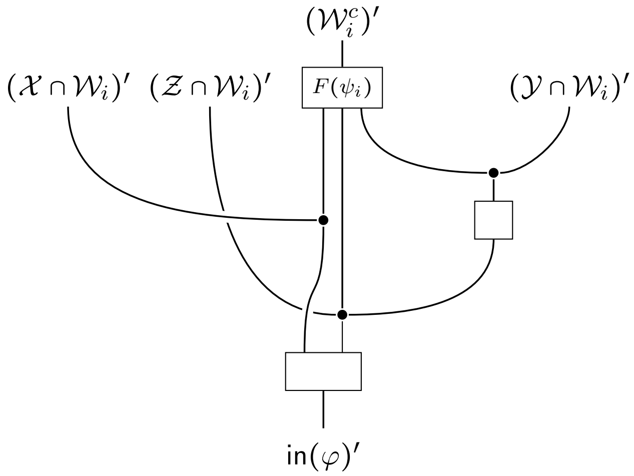
φ =



in $\mathbf{FreeMarkov}_\Sigma$, where η_i is again a pure bloom and η_1 is an identity morphism. The existence of such a chain of sets follows easily by induction on the number of boxes based on the existence of a final box.[9]

Then for every $i \in \{1, \dots, k\}$, we show the existence of a decomposition

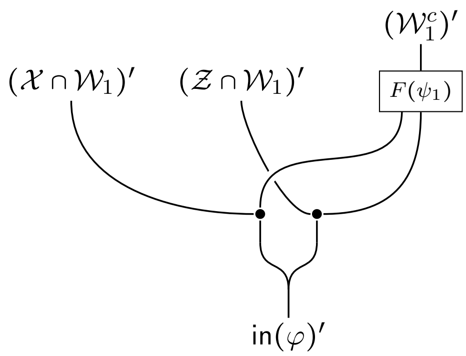
f = (15)



Since $\mathcal{W}_n^c = \emptyset$, setting $i = k$ proves the desired statement.

We prove this stronger claim by induction on i. The start of the induction at $i = 1$ is trivial since η_1 is the identity and therefore

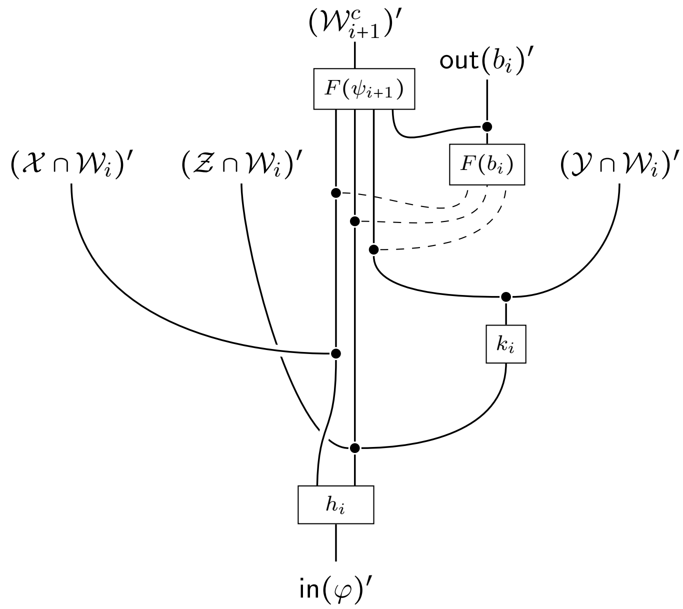
f =



since $\mathcal{Y} \cap \mathcal{W}_1 = \mathcal{Y} \cap \mathsf{in}(\varphi) = \emptyset$. For the induction step, we prove the statement at $i+1$. Since φ is pure bloom and since $\mathsf{An}(\mathcal{W}_{i+1}) = \mathcal{W}_{i+1}$, we can peel off the box b_i with $\mathcal{W}_{i+1} = \mathsf{out}(b_i) \cup \mathcal{W}_i$

[9] It is also worth noting that for causal models which correspond do DAGs, this statement amounts to the standard fact that every DAG can be refined to a total order.

from ψ_i, so as to achieve the decomposition

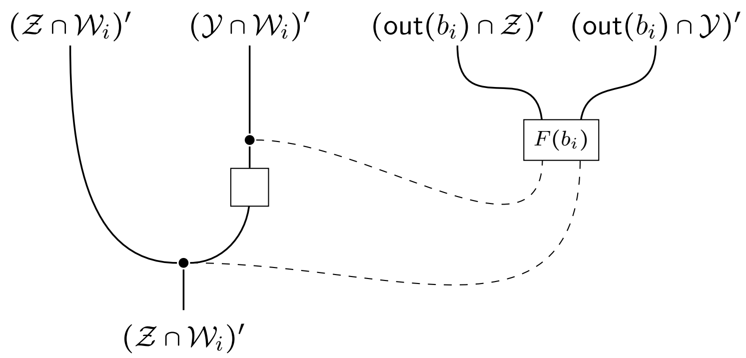
f =  (16)

where we have used the induction assumption to obtain a decomposition as in the lower half, and the dashed wires indicate that only some of them may be present, since the inputs of b_i are an unspecified subset of $\mathcal{W}_i$.

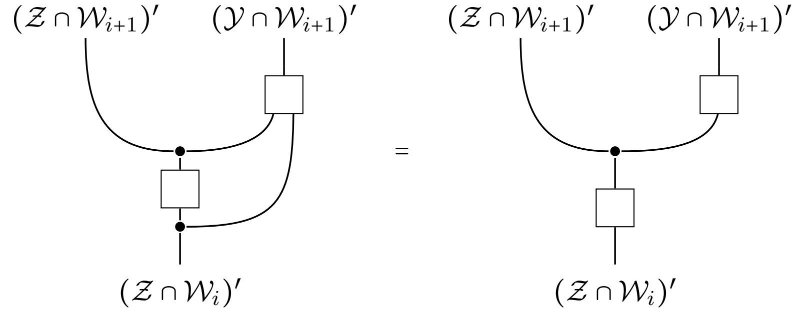
By Lemma 22 we have to distinguish two cases:

(i) $\mathsf{in}(b_i), \mathsf{out}(b_i) \subseteq \mathcal{X} \cup \mathcal{Z}$. Then, the third dashed wire in the above decomposition of f is not needed, and we can merge $F(b_i)$ with h_i, which shows the statement.

(ii) $\mathsf{in}(b_i), \mathsf{out}(b_i) \subseteq \mathcal{Y} \cup \mathcal{Z}$. Then, the first dashed wire in the above decomposition of f is not needed, and we consider the morphism

g := 

which is part of that decomposition. By the existence of conditionals, we can rewrite g in the form

g = 

where both lower boxes can be refined with internal structure consisting of carrying $(\mathcal{Z} \cap \mathcal{W}_i)'$ forward on a separate wire, but this internal structure is not relevant for the remainder of the proof. Substituting this form of g into Equation (16), i.e. replacing the morphism k_i there with the right box here and merging the lower box here with h_i there, proves the induction step. □

We now aim at generalizing Lemma 23 to all generalized causal models and to arbitrary disjoint sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ which do not necessarily partition the set of all wires.

Lemma 24. *Let φ be a generalized causal model and $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathsf{out}(\varphi)$ a tripartition of output wires in φ such that $\mathsf{in}(\varphi) \subseteq \mathcal{X} \cup \mathcal{Z}$ and such that $\mathcal{Z}$ categorically d-separates $\mathcal{X}$ and $\mathcal{Y}$. Then there exists a tripartition of wires $\widetilde{\mathcal{X}} \supseteq \mathcal{X}$, $\widetilde{\mathcal{Y}} \supseteq \mathcal{Y}$, $\mathcal{Z}$ in the pure bloom version $\varphi_{pure\text{-}bloom}$ of φ*[10] *such that*

$$\mathcal{Z} \text{ } d\text{-separates } \widetilde{\mathcal{X}} \text{ and } \widetilde{\mathcal{Y}} \qquad \text{in} \qquad \varphi_{pure\text{-}bloom}$$

Proof. With $\varphi_{\text{cut}} \coloneqq \mathsf{Cut}_{\mathcal{Z}}(\varphi_{\text{pure-bloom}})$, define

$$\widetilde{\mathcal{X}} \coloneqq \{U \in \mathsf{out}(\varphi_{\text{cut}}) : \exists X \in \mathcal{X} : X - U \text{ in } \varphi_{\text{cut}}\} \supseteq \mathcal{X}$$

to be the connected component of $\mathcal{X}$ in φ_{cut}, and

$$\widetilde{\mathcal{Y}} \coloneqq \mathsf{out}(\varphi_{\text{pure-bloom}}) \setminus \left(\widetilde{\mathcal{X}} \cup \mathcal{Z}\right) \supseteq \mathcal{Y}.$$

By definition, $\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}, \mathcal{Z}$ form a tripartition of wires in $\varphi_{\text{pure-bloom}}$. Moreover, $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$ are categorically d-separated by $\mathcal{Z}$ since any path in $\mathsf{Cut}_{\mathcal{Z}}(\varphi_{\text{pure-bloom}})$ is a valid path in $\mathsf{Cut}_{\mathcal{Z}}(\varphi)$ and vice versa. □

Corollary 25. *Let $\mathbf{C}$ be a strict Markov category with conditionals, and let φ be a generalized causal model.*[11] *Further, let $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathsf{out}(\varphi)$ be disjoint sets of output wires in φ such that $\mathsf{in}(\varphi) \subseteq \mathcal{X} \cup \mathcal{Z}$ and $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$. If f is compatible with φ, then $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$.*

In this following statement, we use another standard convention: when the disjoint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ do not partition the set of wires of φ, then the conditional independence $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ is to be understood as Definition 14 applied to the corresponding marginal $f_{\mathcal{X}',\mathcal{Y}',\mathcal{Z}'}$ rather than to f itself.

Proof. We prove this statement by reducing it to the case of pure bloom causal models treated in Lemma 23.

Consider the restricted causal model $\psi \coloneqq \varphi_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$ and its compatible morphism $g \coloneqq F(\psi)$ obtained from the compatibility of f with φ, which is a marginal of f. By the definition of categorical d-separation, $\mathcal{Z}$ d-separates $\mathcal{X}$ and $\mathcal{Y}$ also in ψ. Let $\psi_{\text{pure-bloom}}$ be the pure bloom version of ψ. Since g is compatible with ψ, we can extend g to a pure bloom version

$$g_{\text{pure-bloom}} \coloneqq F(\psi_{\text{pure-bloom}})$$

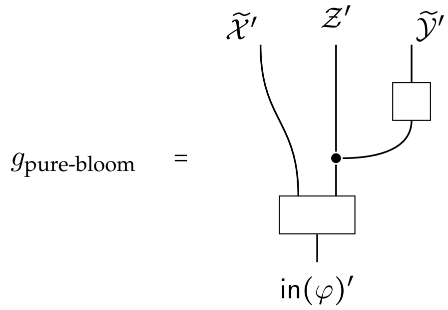
of which g is a marginal.

By Lemma 24, for $\psi_{\text{pure-bloom}}$ there is a tripartition of output wires $\widetilde{\mathcal{X}} \supseteq \mathcal{X}, \widetilde{\mathcal{Y}} \supseteq \mathcal{Y}$, $\mathcal{Z}$ such that $\mathcal{Z}$ d-separates $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$. Since $\mathbf{C}$ has conditionals, Lemma 23 provides us with a

[10] The pure-bloom version $\varphi_{\text{pure-bloom}}$ is obtained by copying each wire so to make it into an output. It is part of the bloom-circuitry factorization of [10].

[11] In this situation φ does not need to be pure bloom.

decomposition of the form



By marginalizing over $\widetilde{\mathcal{X}}' \setminus \mathcal{X}'$, $\widetilde{\mathcal{Y}}' \setminus \mathcal{Y}'$ in $g_{\text{pure-bloom}}$, we obtain the desired conditional independence for the marginal $f_{\mathcal{X}',\mathcal{Y}',\mathcal{Z}'}$. □

Note that this result includes the soundness of the classical d-separation criterion in the classical case of discrete random variables in Bayesian networks[12], which is obtained upon restricting to pure bloom causal models with $\mathsf{in}(\varphi) = \emptyset$, the Markov category **FinStoch**, and every box having precisely one output, since then conditional independence reduces to Definition 12 by Remark 15.

Definition 26. *Let φ be a generalized causal model and f a morphism in a strict Markov category* **C**. *Then we say that f satisfies:*

(i) *the* global Markov property *with respect to φ if for every three disjoint sets of outputs $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathsf{out}(\varphi)$ with $\mathsf{in}(\varphi) \subseteq \mathcal{X} \cup \mathcal{Z}$:*

$$\mathcal{X} \text{ and } \mathcal{Y} \text{ are categorically } d\text{-separated by } \mathcal{Z} \text{ in } \varphi \quad \Longrightarrow \quad \mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z} \text{ in } f.$$

(ii) *the* local Markov property *with respect to φ if for every box b in φ, we have*

$$\mathsf{Dec}(\mathsf{out}(b))^c \setminus \mathsf{in}(b) \perp \mathsf{out}(b) \mid \mathsf{in}(b) \ \text{ in } f.$$

Theorem 27. *Suppose that we are given the following:*

- **C** *is a strict Markov category with conditionals.*
- *φ is a pure bloom causal model over a hypergraph Σ such that the boxes in φ have distinct types in Σ.*
- *$f : \bigotimes_{i=1}^{n} W_i' \to \bigotimes_{j=1}^{m} V_j'$ is a morphism in* **C**.

Then the following statements are equivalent:

(i) *f is compatible with the causal model φ.*
(ii) *f satisfies the global Markov property.*
(iii) *f satisfies the local Markov property.*

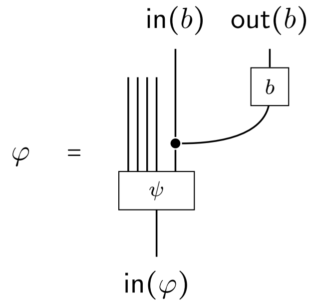
Proof. $(i) \Longrightarrow (ii)$: The global Markov property is precisely the statement of Corollary 25.

$(ii) \Longrightarrow (iii)$: This follows from the fact that $\mathsf{Dec}(\mathsf{out}(b))^c$ and $\mathsf{out}(b)$ are d-separated by $\mathsf{in}(b)$, and $\mathsf{in}(\varphi) \subseteq \mathsf{Dec}(\mathsf{out}(b))^c$, which makes the global Markov property specialize to the local one.

$(iii) \Longrightarrow (i)$: We prove this statement by induction over the number of boxes $k \coloneqq |B(G)|$. The case $k = 1$ is trivial. For the step from k to $k+1$, let b be a final box in φ, which means

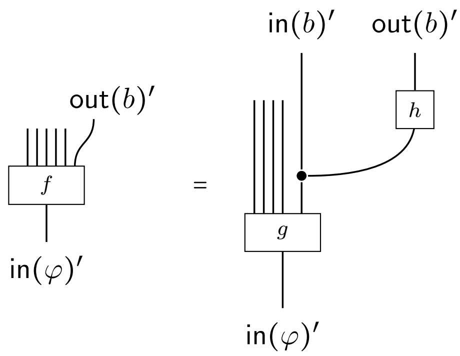
[12] See [21] for the original proof and [18, Theorem 1.2.5(i)] for a textbook account.

that $\mathsf{Dec}(\mathsf{out}(b)) = \mathsf{out}(b)$. Then, φ factorizes as



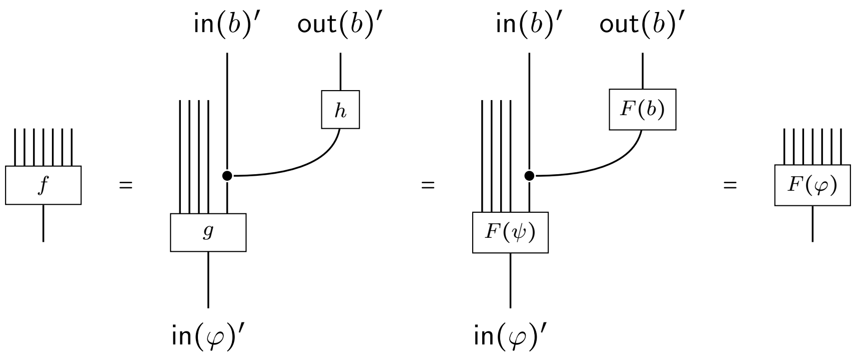
where ψ is another causal model satisfying all of our assumptions, and no box in ψ has the same type in Σ as b does.

In order to construct a functor F as in Definition 10, note first that it must satisfy (11), which already lets us write the domain of f as $\mathsf{in}(\varphi)'$, and similarly for the codomain. Since f satisfies the local Markov property with respect to b, we can decompose f by Definition 14 as



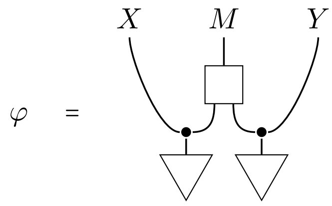
(17)

By induction hypothesis, we have that g is compatible with ψ since g satisfies the local Markov properties specified by ψ. Since the box b appears only once in φ, we can freely define the action of the functor F on b as $F(b) \coloneqq h$. Then, we obtain

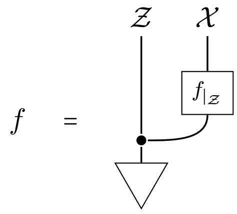


where we use in the first step Equation (17) and in the last the fact that F is a Markov functor. □

Remark 28. (i) Note that we have used the assumption that **C** has conditionals only for the implication (i) $\implies$ (ii). Therefore, for an arbitrary strict Markov category, the global as well as the local Markov property is a sufficient condition for the compatibility of a morphism with a generalized causal model (satisfying our assumptions). However, these Markov properties require implicitly the existence of certain conditionals. Consider for example the generalized causal model



where all boxes are of distinct types. Choosing $\mathcal{X} = \{M\}, \mathcal{Y} = \emptyset$ and $\mathcal{Z} = \{X, Y\}$, a morphism f satisfying the global Markov property displays in particular the conditional independence $\{M\} \perp \emptyset \mid \mathcal{Z}$, pictorially:



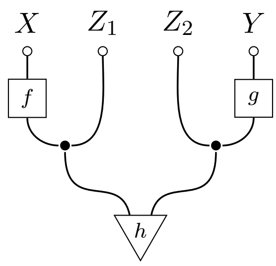
This shows that the conditional $f_{|\mathcal{Z}}$ exists, and this recovers the box that outputs M (up to almost sure equality).

(ii) Theorem 27 shows that d-separation correctly detects causal compatibility for the Markov categories **FinStoch**, **Gauss** or **BorelStoch**. For the Markov category **Stoch**, which does not have conditionals, the global and local Markov properties are at least sufficient for the compatibility, since our proof of these implications has not used conditionals.

(iii) Note that Theorem 27 only applies to causal models where each box appears at most once in the model (which in particular implies that φ has no nontrivial symmetries). However, the implication $(i) \implies (ii)$ applies to arbitrary pure bloom causal models as proven in Lemma 23. △

Example 29. We now present two examples which go beyond the classical d-separation criterion. In (i) we will study a causal structure which does not arise from a DAG, while in (ii) we study a DAG causal structure with continuous variables.

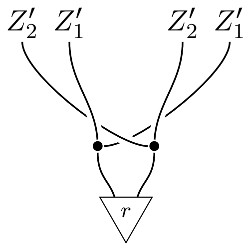
(i) Let φ be the causal structure



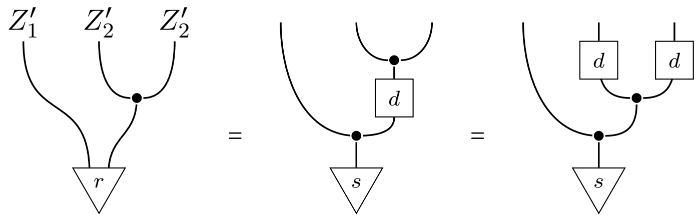
and let **C** be a strict Markov category with conditionals. By Theorem 27, a morphism $t : I \to X' \otimes Z'_1 \otimes Z'_2 \otimes Y'$ in **C** is compatible with this structure if and only if it satisfies

$$X \perp \{Y, Z_2\} \mid Z_1 \quad \text{and} \quad Y \perp \{X, Z_1\} \mid Z_2$$

For a general class of examples, consider $X' = Z_2'$ and $Y' = Z_1'$ and any morphism in $\mathbf{C}$ of the form



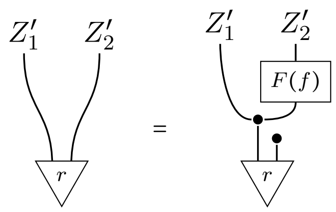
We claim that such a distribution is compatible with φ if and only if there exist morphisms d and d' such that



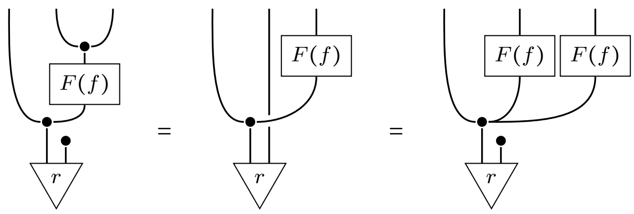
(18)

where s is the first marginal of r, and similarly $d' : Z_2' \to Z_1'$ satisfies the same equations the other way around. Here, the second equation states that the morphism d is s-a.s. deterministic [9, Definition 13.11], and similarly for d'.

Indeed, assuming compatibility we have that

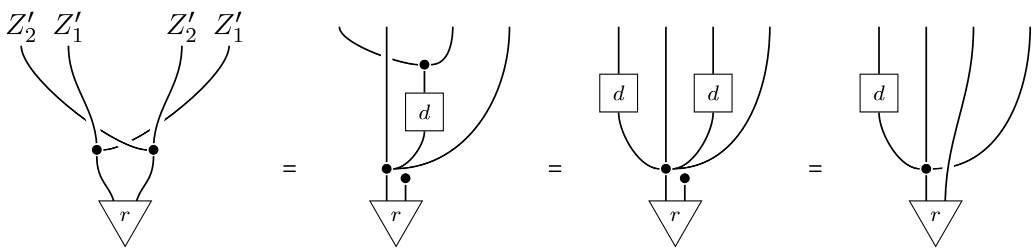


which shows the first equality in Equation (18). For the second equality we have that



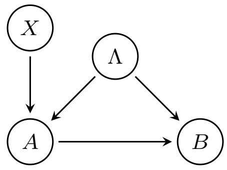
Proving the existence of d' works analogously by interchanging the roles of X and Y as well as Z_1 and Z_2.

Conversely, we have

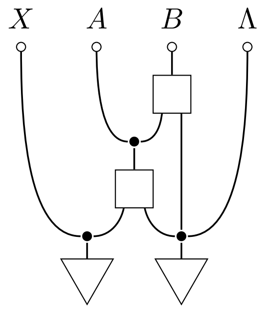


where we have used the assumption that d is s-a.s. deterministic in the second equation. Repeating this calculation interchanging the roles of Z_1 and Z_2 as well as X and Y shows the statement.

(ii) Consider the instrumental scenario, given by the DAG



This has been previously studied mainly in the context of DAGs with latent variables [17, 2]. For our analysis we assume each variable to be observed, which means that the causal structure reads string-diagrammatically as

φ =  (19)

There are two non-trivial d-separations:
(a) Between X and B by $\{A, \Lambda\}$,
(b) Between X and Λ.

Therefore, Theorem 27 implies that a distribution P on a four-fold tensor product object in a Markov category with conditionals is compatible with φ if and only if $X \perp B \mid A, \Lambda$ and $X \perp \Lambda$. In **BorelStoch**, this means that P is compatible with φ if and only if

$$\begin{aligned} &P(X \in E_1, A \in E_2, B \in E_3, \Lambda \in E_4) \\ &= \int_{E_2}\int_{E_4} P_{X|A,\Lambda}(X \in E_1|a,\lambda)\, P_{B|A,\Lambda}(B \in E_3|a,\lambda)\, P_{A,\Lambda}(\mathrm{d}a, \mathrm{d}\lambda) \end{aligned}$$

and

$$P(X \in E_1, \Lambda \in E_4) = P(X \in E_1) \cdot P(\Lambda \in E_4)$$

where E_i are measurable sets in the Borel σ-algebras of the spaces X', A', B' and Λ'.

For simplicity, assume that all random variables take values in $\mathbb{R}$ and are absolutely continuous, i.e. there exists a density $f : X' \times A' \times B \times \Lambda' \to [0, \infty)$ such that

$$P(X \in E_1, A \in E_2, B \in E_3, \Lambda \in E_4) = \int_{E_1\times E_2\times E_3\times E_4} f(x, a, b, \lambda)\, \mathrm{d}x\, \mathrm{d}a\, \mathrm{d}b\, \mathrm{d}\lambda$$

The causal compatibility now amounts to the following two conditions:
(a) $X \perp \Lambda$, i.e.

$$f_{X,\Lambda}(x, \lambda) = f_X(x) \cdot f_\Lambda(\lambda) \tag{20}$$

where a.e. means almost everywhere with respect to the Lebesgue measure on $\mathbb{R}$.
(b) $X \perp B \mid A, \Lambda$, i.e.

$$f(x, a, b, \lambda) = f_{X|A,\Lambda}(x, a, \lambda) \cdot f_{A,\Lambda}(a, \lambda) \cdot f_{B|A,\Lambda}(b, a, \lambda), \tag{21}$$

where the conditional densities are defined implicitly by

$$f_{A|X\Lambda}(a|x,\lambda)\cdot f_{X,\Lambda}(x,\lambda) = f_{X,A,\Lambda}(x,a,\lambda) \quad \text{a.e.}$$

Combining Eq. (20) and Eq. (21) results in

$$f(x,a,b,\lambda) = f_{\Lambda}(\lambda)\cdot f_X(x)\cdot f_{A|X\Lambda}(a,x,\lambda)\cdot f_{B|A,\Lambda}(b,a,\lambda) \quad \text{a.e.}$$

which is the usual factorization condition for compatibility with the causal structure in (19). △

Question 30. *Can Theorem 27 be extended to more general causal models? In particular, what about allowing the same box to appear several times in φ?*

References

[1] F. Bonchi, F. Gadducci, A. Kissinger, P. Sobociński, and F. Zanasi. Rewriting modulo symmetric monoidal structure. *Proceedings - Symposium on Logic in Computer Science*, pages 710–719, 2016. https://doi.org/10.1145/2933575.2935316, https://arxiv.org/abs/1602.06771.

[2] B. Bonet. Instrumentality tests revisited. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 48–55. Morgan Kaufmann Publishers Inc., 2001. https://dl.acm.org/doi/10.5555/2074022.2074029.

[3] K. Cho and B. Jacobs. Disintegration and Bayesian inversion via string diagrams. *Mathematical Structures in Computer Science*, 29:938–971, 2019. https://doi.org/10.1017/S0960129518000488, https://arxiv.org/abs/1709.00322.

[4] B. Coecke and R. W. Spekkens. Picturing classical and quantum Bayesian inference. *Synthese*, 186:651, 2011. http://dx.doi.org/10.1007/s11229-011-9917-5, http://arxiv.org/abs/1102.2368.

[5] A. Corradini and F. Gadducci. An algebraic presentation of term graphs, via gs-monoidal categories. *Applied Categorical Structures*, 7:299–331, 1999. http://dx.doi.org10.1023/A:1008647417502.

[6] B. Fong. Causal theories: A categorical perspective on Bayesian networks. Master's thesis, University of Oxford, 2013.

[7] P. Forré. Transitional conditional independence, 2021. http://arxiv.org/abs/2104.11547.

[8] P. Forré and J. M. Mooij. Markov properties for graphical models with cycles and latent variables, 2017. http://arxiv.org/abs/1710.08775.

[9] T. Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:1–98, 2020. https://doi.org/10.1016/j.aim.2020.107239, https://arxiv.org/abs/1908.07021.

[10] T. Fritz and W. Liang. Free gs-monoidal categories and free Markov categories, 2022. https://arxiv.org/abs/2204.02284.

[11] F. Gadducci. *On The Algebraic Approach To Concurrent Term Rewriting*. PhD thesis, University of Pisa, 1996. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.5651.

[12] D. Gao. Functorial causal models: Towards an algebraic approach to causal inference. Master's thesis, Brown University, 2022. https://www.brown.edu/academics/math/sites/math/files/Gao%2C%20Dichuan_0.pdf.

[13] B. Jacobs, A. Kissinger, and F. Zanasi. Causal inference by string diagram surgery. *Lecture Notes in Computer Science*, 11425 LNCS:313–329, 2019. https://doi.org/10.1007/978-3-030-17127-8_18.

[14] S. L. Lauritzen. *Graphical Models*, volume 18 of *Oxford Statistical Science Series*. Oxford University Press, 1996.

[15] S. L. Lauritzen, A. P. Dawid, B. N. Larsen, and H.-G. Leimer. Independence properties of directed markov fields. *Networks*, 20:491–505, 1990. https://doi.org/10.1002/net.3230200503.

[16] A. Milosavljevic and F. Zanasi. String diagram rewriting modulo commutative monoid structure, 2022. http://arxiv.org/abs/2204.04274.

[17] J. Pearl. On the testability of causal models with latent and instrumental variables. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 435–443. Morgan Kaufmann Publishers Inc., 1995. https://dl.acm.org/doi/10.5555/2074158.2074208.

[18] J. Pearl. *Causality: Models, reasoning, and inference*. Cambridge University Press, 2009. Second edition.

[19] E. F. Rischel. The category theory of causal models. Master's thesis, University of Copenhagen, 2020. https://www.erischel.com/documents/mscthesis.pdf.

[20] E. F. Rischel and S. Weichwald. Compositional abstraction error and a category of causal models. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161, pages 1013–1023. MLResearchPress, 2021. https://proceedings.mlr.press/v161/rischel21a.html.

[21] T. Verma and J. Pearl. Causal networks: Semantics and expressiveness. In *Proceedings of the Fourth Conference on Uncertainty in Artificial Intelligence*, volume 9, pages 69–76. Morgan Kaufmann Publishers, 1990. https://doi.org/10.1016/B978-0-444-88650-7.50011-1.