Testing Edges by Truncations

Ilya Shpitser
Department of Epidemiology
Harvard School of Public Health
ishpitse@hsph.harvard.edu

Thomas S. Richardson
Department of Statistics
University of Washington
tsr@stat.washington.edu

James M. Robins
Department of Epidemiology
Harvard School of Public Health
robins@hsph.harvard.edu

Abstract

We consider the problem of testing whether two variables should be adjacent (either due to a direct effect between them, or due to a hidden common cause) given an observational distribution, and a set of causal assumptions encoded as a causal diagram. In other words, given a set of edges in the diagram known to be true, we are interested in testing whether another edge ought to be in the diagram. In fully observable faithful models this problem can be easily solved with conditional independence tests. Latent variables make the problem significantly harder since they can imply certain non-adjacent variable pairs, namely those connected by so called inducing paths, are not independent conditioned on any set of variables. We characterize which variable pairs can be determined to be non-adjacent by a class of constraints due to dormant independence, that is conditional independence in identifiable interventional distributions. Furthermore, we show that particular operations on joint distributions, which we call truncations are sufficient for exhibiting these non-adjacencies. This suggests a causal discovery procedure taking advantage of these constraints in the latent variable case can restrict itself to truncations.

1 Introduction

Causal discovery, that is the problem of learning causal theories from observations, is central to empirical science. In graphical models, which is a popular formalism for representing causal assumptions in the presence of uncertainty, there is a large literature on this problem [Cooper and Dietterich, 1992], [Spirtes *et al.*, 1993], [Suzuki, 1993]. In such models, causal assumptions are represented by means of a directed acyclic graph (dag) called a causal diagram, where nodes are variables of interest, and arrows represent, informally, direct causal influences.

In this setting, causal discovery amounts to learning aspects of the causal diagram from observations summarized as an observable joint probability distribution $P(\mathbf{v})$. The simplest causal discovery problem assumes causal sufficiency, e.g. it assumes whenever two observed variables share a common cause, that cause is itself observed. There are two approaches to causal discovery in this setting. The score-based approach [Suzuki, 1993], assigns to each possible graph a score consisting of two terms – the likelihood term which measures how well the graph fits the data, and the model complexity term, which penalizes large graphs. A search is then performed for high scoring graphs. The second, so called constraint-based approach [Spirtes *et al.*, 1993] rules out graphs which are not compatible with constraints observed in the data. Both approaches to causal discovery in the presence of causal sufficiency rely on faithfulness, which is a property stating that a notion of path-separation known as d-separation [Pearl, 1988] in the graph precisely characterizes conditional independence inherent in the distribution which the graph represents.

In the absence of causal sufficiency, the causal discovery problem is significantly harder, especially if, as in this paper, no parametric assumptions are made. One problem is that in this setting graphs may entail constraints which cannot be expressed as conditional independence constraints.

Previous work showed that certain non-independence constraints, which we refer to as Verma constraints, may be represented by an identified dormant independence (i.e., a conditional independence in an interventional distribution [Robins, 1986], [Spirtes *et al.*, 1993], [Pearl, 2000] that can be used to test for the presence of an edge in the causal diagram [Robins, 1999], [Tian and Pearl, 2002b], [Shpitser and Pearl, 2008]. We show that there exist identified dormant independencies that cannot be used to test for the presence of an edge and thus do not represent a Verma constraint.

Using the notion of edge testing, explored in [Shpitser and Pearl, 2008], we characterize which dormant independences do give rise to Verma constraints. Finally, we prove that every such dormant independence can be obtained from the joint distribution by a sequence of simple operations we call truncations, and construct an algorithm which does this. This implies that an algorithm trying to recover correct causal structure from data which uses Verma constraints can restrict itself to searching over possible truncations.

2 An Example of Edge Testing

Consider the causal graph in Fig. 1 (a), where a bidirected arc corresponds to an unobserved cause of W and Y. Any model

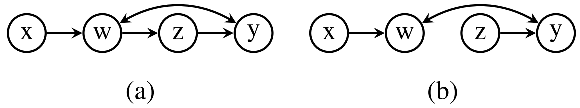

Figure 1: (a) The Verma graph. (b) The graph of the submodel M_z derived from the Verma graph.

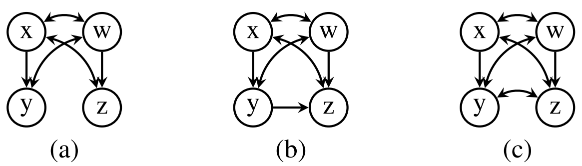

Figure 2: (a) A graph where Y and Z are independent given $do(x, w)$, yet there is no Verma constraint. Illustration of how Theorem 3 fails for identifying a directed edge from Y to Z (b) and bidirected edge from Y to Z (c).

compatible with this graph imposes certain constraints on its observable distribution $P(x, w, z, y)$, here assumed positive. Some of these constraints are in the form of conditional independences. For instance, in any such model X is independent of Z given W, which means $P(x|w) = P(x|w, z)$.

However, there is an additional constraint implied by this graph which cannot be expressed in terms of conditional independence in the observable distribution. This constraint, noted in [Verma and Pearl, 1990], [Robins, 1999], states that the expression $\sum_w P(y|z, w, x)P(w|x)$ is a function of y and z, but not of x. In fact, constraints of this type do emanate from conditional independences, albeit not in the original observable distribution, but in a distribution resulting from an intervention.

An intervention, written $do(\mathbf{x})$ [Pearl, 2000], is an operation which forces variables $\mathbf{X}$ to attain values $\mathbf{x}$ regardless of their usual behavior in a causal model. The result of applying an intervention $do(\mathbf{x})$ on a model M with a set of observable variables $\mathbf{V}$ is a *submodel* $M_{\mathbf{x}}$, with stochastic behavior of $\mathbf{V}$ described by an *interventional distribution* written as $P_{\mathbf{x}}(\mathbf{v})$ or $P(\mathbf{v}|do(\mathbf{x}))$. The graph induced by $M_{\mathbf{x}}$ is almost the same as the graph induced by M, except it is missing all arrows incoming to $\mathbf{X}$, to represent the fact that an intervention sets the values of $\mathbf{X}$ independently of its usual causal influences, represented by such arrows. We will denote such a graph as $G_{\overline{\mathbf{x}}}$. We will also consider the so called stochastic interventions, where values of $\mathbf{X}$ are set according to a new distribution $P^*(\mathbf{x})$. We denote this intervention as $do(\mathbf{x} \sim P^*)$, and its result by $P(\mathbf{v}|do(\mathbf{x} \sim P^*))$ or $P_{\mathbf{x} \sim P^*}(\mathbf{v})$. The graphical representation of stochastic interventions is the same as that of ordinary interventions.

A key idea in causal inference is that in certain causal models, some interventional distributions can be *identified* from the observational distribution. Consider for instance a model M inducing the graph in Fig. 1 (a). If we intervene on Z in M, we obtain the submodel M_z inducing the graph in Fig. 1 (b). The distribution of the unfixed observables in this submodel, $P_z(x, w, y)$, is identifiable from $P(x, w, z, y)$ and equals to $P(y|z, w, x)P(w|x)P(x)$ [Pearl and Robins, 1995], [Tian and Pearl, 2002a]. Moreover, by d-separation [Pearl, 1988], the graph in Fig. 1 (b) implies that X is independent of Y in $P_z(x, w, y)$, or $P_z(y|x) = P_z(y)$. But it's not hard to show that $P_z(y|x)$ is equal to $\sum_w P(y|z, w, x)P(w|x)$, which means this expression depends only on z and y. What we have just shown is that independence in $P_z(x, w, y)$, which is an identifiable distribution, leads to a constraint on observational distributions in the original, unmutilated model M. The same reasoning applies to stochastic interventions, since all identification results carry over without change. Conditional independences in interventional distributions are called dormant [Shpitser and Pearl, 2008]. Given an appropriate notion of faithfulness, we can conclude from this constraint that there should be no edge between X and Y in Fig. 1 (a). In the remainder of the paper, we explore this sort of constraint-based testing in more detail.

2.1 A Vacuous Dormant Independence

It turns out that not every dormant independence leads to a Verma constraint. Consider the graph in Fig. 2 (a). In this graph, variables Y and Z cannot be d-separated by any conditioning set. However, it's not difficult to see that Y and Z are independent if we fix X and W, in other words Y is independent of Z in $P(y, z|do(x, w))$. Moreover, $P(y, z|do(x, w))$ is identifiable in this graph, and equal to $P(y|x)P(z|w)$. In other words, there exists an (identifiable) dormant independence between Y and Z. However, if we translate what this independence asserts about the joint distribution $P(x, y, z, w)$ we get that Y and Z are independent in the distribution $P(y|x)P(z|w)$ where x, w are held constant and y, z are allowed to vary. But this independence trivially holds by construction regardless of what $P(\mathbf{v})$ we choose. In other words, this dormant independence does not constrain $P(x, y, z, w)$ in any way.

Which dormant independences do give rise to Verma constraints? For intuition we consider the easier case of conditional independence. In faithful models the presence of conditional independence between X and Y implies the lack of an edge connecting these variables. Similarly, we would expect (identifiable) dormant independences between two variables which lead to constraints to imply the lack of an edge between those variables. However, unlike the conditional independence case, identifying dormant independences relies on the absence of certain edges (since the process of identifying interventional distributions relies on causal assumptions embodied by missing edges in the causal diagram). The key point is that identifying the dormant independence used to test a particular edge e cannot rely on the absence of that edge in the process of identification!

In particular, the distribution $P(y, z|do(x, w))$ in Fig. 2 (a) is only identifiable if we assume there is no edge between Y and Z (the very thing we would be trying to test by a dormant independence between Y and Z). In particular, if we add a directed edge from Y to Z as in Fig. 2 (b), $P(y, z|do(x, w))$ becomes non-identifiable since the graph becomes a Z-rooted C-tree [Shpitser and Pearl, 2006b]. Similarly, $P(y, z|do(x, w))$ becomes non-identifiable if we add a bidirected edge from Y to Z as in Fig. 2 (c).

3 Preliminaries

One way of formalizing causal inference is with probabilistic causal models (PCMs). Such models consist of two sets of variables, the observable set $\mathbf{V}$ representing the domain of interest, and the unobservable set $\mathbf{U}$ representing the background to the model that we are ignorant of. Associated with each observable variable V_i in $\mathbf{V}$ is a function f_i which determines the value of V_i in terms of values of other variables in $\mathbf{V} \cup \mathbf{U}$. Finally, there is a joint probability distribution $P(\mathbf{u})$ over the unobservable variables, signifying our ignorance of the background conditions of the model.

The causal relationships in a PCM are represented by the functions f_i, each function causally determines the corresponding V_i in terms of its inputs. Causal relationships entailed by a given PCM have an intuitive visual representation using a graph called a causal diagram. Causal diagrams contain two kinds of edges. Directed edges are drawn from a variable X to a variable V_i if X appears as an input of f_i. Directed edges from the same unobservable U_i to two observables V_j, V_k can be replaced by a bidirected edge between V_j to V_k. We will consider models which induce acyclic graphs where $P(\mathbf{u}) = \prod_i P(u_i)$, and each U_i has at most two observable children. A graph obtained in this way from a model is said to be induced by said model.

The importance of causal diagrams stems from the fact that conditional independences between observable variables correspond to graphical features in the diagram. Since the rest of the paper will rely heavily on this correspondence, we introduce probabilistic and graphical notions we will need to make use of it. A set $\mathbf{X}$ is independent of $\mathbf{Y}$ conditional on $\mathbf{Z}$ (written as $\mathbf{X} \perp\!\!\!\perp_P \mathbf{Y}|\mathbf{Z}$) if $P(\mathbf{x}|\mathbf{y},\mathbf{z}) = P(\mathbf{x}|\mathbf{z})$. We will use the following graph-theoretic notation. $An(.)_G, De(.)_G, Pa(.)_G$ stand for the set of ancestors, descendants and parents of a given variable set in G. The sets $An(.)_G$ and $De(.)_G$ will be inclusive, in other words, for every $An(X)_G, De(X)_G$, $X \in An(X)_G$ and $X \in De(X)_G$. The set $C(X)_G$ stands for the C-component of X [Tian and Pearl, 2002a], that is the maximal set of nodes containing X where any two nodes are pairwise connected by a bidirected path contained in the set. A graph $G_{\mathbf{x}}$ is the subgraph of G containing only nodes in $\mathbf{X}$ and edges between them.

It's possible to show that whenever edges in a causal diagram are drawn according to the above rules, the distribution $P(\mathbf{u},\mathbf{v})$ induced by $P(\mathbf{u})$ and the f_i functions factorizes as $\prod_{X_i \in \mathbf{V} \cup \mathbf{U}} P(x_i|Pa(x_i))$. This Markov factorization implies that conditional independences in $P(\mathbf{u},\mathbf{v})$ are mirrored by a well-known graphical notion of "path blocking" known as d-separation [Pearl, 1988], which we will not reproduce here.[1]

Two sets $\mathbf{X}, \mathbf{Y}$ are said to be d-separated given $\mathbf{Z}$ (written $\mathbf{X} \perp_G \mathbf{Y}|\mathbf{Z}$) in G if all paths from $\mathbf{X}$ to $\mathbf{Y}$ in G are d-separated by $\mathbf{Z}$. Paths or sets which are not d-separated are said to be d-connected. The relationship between d-separation and conditional independence is provided by the following theorem.

Theorem 1 (Pearl) *Let G be a causal diagram. Then in any model M inducing G and P, if $\mathbf{X} \perp_G \mathbf{Y}|\mathbf{Z}$, then $\mathbf{X} \perp\!\!\!\perp_P \mathbf{Y}|\mathbf{Z}$.*

[1] In fact d-separation is defined for dags, although a natural generalization exists for mixed graphs with bidirected arcs [Richardson and Spirtes, 2002].

Using d-separation as a guide, we can look for a conditioning set $\mathbf{Z}$ which renders given sets $\mathbf{X}$ and $\mathbf{Y}$ independent by only examining the causal diagram, without having to inspect the probability distribution $P(\mathbf{v})$.

A causal dag with an arbitrary set of hidden variables has a graphical representation called the latent projection [Pearl, 2000], such that the projection only contains nodes corresponding to observable variables, it contains only directed and bidirected arcs, is acyclic, and preserves the set of d-separation statements over observed nodes in the dag. In this sense the mixed graphs we consider include causal models with arbitrary latent variables.

We examine probabilistic independences in distributions resulting from not only conditioning but a second, powerful operation of intervention, defined in the previous section. An intervention is a more powerful operation than conditioning, for the purposes of determining probabilistic independence. This is because conditioning on a variable can d-separate certain paths, but also d-connect certain paths (due to the presence of the so called collider triples). On the other hand, interventions can only block paths, since incoming arrows are cut by interventions, destroying all colliders involving the intervened variable. Some interventions can be computed from $P(\mathbf{v})$ and the graph G, due to a general notion called identifiability, defined as follows.

Definition 1 (identifiability) *Consider a class of models $\boldsymbol{M}$ with a description T, and two objects ϕ and θ computable from each model. We say that ϕ is θ-identified in T if all models in $\boldsymbol{M}$ which agree on θ also agree on ϕ.*

If ϕ is θ-identifiable in T, we write $T, \theta \vdash_{id} \phi$. Otherwise, we write $T, \theta \not\vdash_{id} \phi$. Often, the model class T corresponds to a causal graph, θ is the observational distribution $P(\mathbf{v})$, and ϕ is the causal effect of interest. For example, in Fig. 1 (a), $G, P(\mathbf{v}) \vdash_{id} P_z(x, w, y)$. Conditional independences in interventional distributions are called *dormant* in [Shpitser and Pearl, 2008].

Definition 2 (dormant independence) *A dormant independence exists between variable sets $\boldsymbol{X}, \boldsymbol{Y}$ in $P(\boldsymbol{v})$ obtained from the causal graph G if there exist variable sets $\boldsymbol{Z}, \boldsymbol{W}$ (with $\boldsymbol{W}$ possibly intersecting $\boldsymbol{X} \cup \boldsymbol{Y}$) such that $P_{\boldsymbol{w} \sim P^*}(\boldsymbol{y}|\boldsymbol{z}) = P_{\boldsymbol{w} \sim P^*}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z})$. If $P(\boldsymbol{v}), G \vdash_{id} P_{\boldsymbol{w} \sim P^*}(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{z})$, the dormant independence is identifiable; we denote this as $\boldsymbol{X} \perp\!\!\!\perp_{\boldsymbol{w},G,P} \boldsymbol{Y}|\boldsymbol{Z}$.*[2]

A natural graphical analogue of d-separation exists for dormant independence.

Definition 3 (d^*-separation) *Let G be a causal diagram. Variable sets $\boldsymbol{X}, \boldsymbol{Y}$ are d^*-separated in G given $\boldsymbol{Z}, \boldsymbol{W}$ (written $\boldsymbol{X} \perp_{\boldsymbol{w},G} \boldsymbol{Y}|\boldsymbol{Z}$), if we can find sets $\boldsymbol{Z}, \boldsymbol{W}$, such that $\boldsymbol{X} \perp_{G_{\overline{\boldsymbol{w}}}} \boldsymbol{Y}|\boldsymbol{Z}$, and $P(\boldsymbol{v}), G \vdash_{id} P_{\boldsymbol{w} \sim P^*}(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{z})$.*

Note that despite the presence of probability notation in the definition, this is a purely graphical notion, since identification can be determined using only the graph [Shpitser and

[2] Earlier work [Shpitser and Pearl, 2008] used $\perp\!\!\!\perp_{\mathbf{w}}$. We use this modified notation to emphasize the fact that dormant independence unlike conditional independence is both graph and distribution dependent.

Pearl, 2006a]. A theorem analogous to Theorem 1 links d^*-separation and identifiable dormant independence [Shpitser and Pearl, 2008]. [3]

Theorem 2 *Let G be a causal diagram. Then in any model M inducing G, if $\mathbf{X} \perp_{\mathbf{w},G} \mathbf{Y}|\mathbf{Z}$, then $\mathbf{X} \perp\!\!\!\perp_{\mathbf{w},G,P} \mathbf{Y}|\mathbf{Z}$.* [4]

In the next section we consider the question of edge identification, namely determining which dormant independences can test non-adjacency of variable pairs.

4 Edge Identification

Typically, questions of identifiability are posed about distributions obtained from interventions, such as causal effects and counterfactuals. In such cases, a given distribution P is identifiable in a model class $\mathbf{M}$ if every model in the class agrees on P. This can be extended to any aspect of the model, such as an edge. An edge e is said to be θ-identifiable in $\mathbf{M}$ (written $\theta \vdash_{id} e$) if every model in $\mathbf{M}$ which agrees on θ agrees on the presence of e.

For faithful, causally sufficient models, $P(\mathbf{v})$-identified edges are those edges on which the Markov equivalence class of models [Pearl, 1988] consistent with $P(\mathbf{v})$ agree.

We are interested in exploring which edges can be identified if, in addition to $P(\mathbf{v})$, we assume partial causal knowledge in the form of a graph, where every edge is either correctly specified, or extraneous. We call such graphs *valid*. As we saw in an earlier example, such problems may be solved using constraints implied by dormant independence, given an appropriate notion of faithfulness. It turns out this notion is identifiable faithfulness.[5] A model is *identifiable faithful* if the subgraph corresponding to every identifiable $P_{\mathbf{x}\sim P^*}(\mathbf{y})$ is faithful to $P_{\mathbf{x}\sim P^*}(\mathbf{y})$. [6] We can give a general theorem for testing edges in valid graphs of such models.

Theorem 3 *Let M be an identifiable faithful model with observable distribution $P(\mathbf{v})$, G a graph valid for M, and e a (possibly extraneous) edge between X and Y in G. Then $G, P(\mathbf{v}) \vdash_{id} e$ if there exist variable sets $\mathbf{W}, \mathbf{Z}$, such that $G, P(\mathbf{v}) \vdash_{id} P_{\mathbf{w}}(x, y|\mathbf{z})$, $X \perp_{G_{\overline{\mathbf{w}}}\setminus\{e\}} Y|\mathbf{Z}$, and $G_{\overline{\mathbf{w}}} \neq G_{\overline{\mathbf{w}}} \setminus \{e\}$. $G_{\overline{\mathbf{w}}} \setminus \{e\}$ is obtained by removing incoming arrows to $\mathbf{W}$, and then removing e from G.*

Proof: Assume such sets $\mathbf{W}, \mathbf{Z}$ exist. If X and Y are independent in $P_{\mathbf{w}\sim P^*}(x, y|\mathbf{z})$, then due to results in [Shpitser and Pearl, 2006a], there exists a set $\mathbf{Z}'$ such that $G, P(\mathbf{v}) \vdash_{id} P_{\mathbf{w},\mathbf{z}'}(x, y, \mathbf{z} \setminus \mathbf{z}')$ iff $G, P(\mathbf{v}) \vdash_{id} P_{\mathbf{w}}(x, y|\mathbf{z})$. By identifiable faithfulness of M, the true graph G' corresponding to $P_{\mathbf{w},\mathbf{z}'}(x, y, \mathbf{z} \setminus \mathbf{z}')$ is faithful to $P_{\mathbf{w},\mathbf{z}'}(x, y, \mathbf{z} \setminus \mathbf{z}')$, so it cannot contain e. Since $\mathbf{Z}$ cannot intersect $\{X, Y\}$, and we assumed $G_{\overline{\mathbf{w}}} \neq G_{\overline{\mathbf{w}}} \setminus \{e\}$, $do(\mathbf{w} \cup \mathbf{z}')$ does not cut e. We conclude that e is extraneous in G.

If X and Y are dependent in $P_{\mathbf{w}}(x, y|\mathbf{z})$, then by identifiable faithfulness of M, the true graph of M must contain e since, $G_{\overline{\mathbf{w}}} \setminus \{e\}$ d-separates X and Y given $\mathbf{Z}$. Thus the presence of e is a function of G and $P_{\mathbf{w}}(x, y|\mathbf{z})$, and the latter is a function of $P(\mathbf{v})$. $\square$

We use existing graphical conditions for identification of conditional causal effects [Shpitser and Pearl, 2006a] to characterize exactly when edges can be identified in a valid graph via Theorem 3, although we make no claims about the existence of other methods by which edges may be identified. In the remainder of the paper we use "identified" and "identified by Theorem 3" as synonyms.

We start by characterizing directed edge testing. First, we define some graphical terminology. A *C-forest* is a graph consisting of a single C-component where every node has at most one child. A C-forest with a single childless node is called *C-tree*. The set of childless nodes $\mathbf{R}$ of a C-forest is called its root set. The unique maximum set of nodes that forms a Y-rooted C-tree in G is called the *maximum ancestral confounding set (MACS)* of Y, and is denoted by T_y^G [Shpitser and Pearl, 2008].

Theorem 4 *Let e be a directed edge $X \rightarrow Y$ in G. Then $P(\mathbf{v}), G \vdash_{id} e$ (via Theorem 3) if and only if $X \notin T_y^G$, and the only directed edge from X to T_y^G is e.*

Proof: Let $G' = G \setminus \{e\}$. By Theorem 3, $Y \notin \mathbf{W}$ for any separating $\mathbf{W}$. Assume $X \in T_y^G$. This implies there is an inducing path from X to Y in G, and moreover no element of T_y^G can be fixed, since $G, P(\mathbf{v}) \not\vdash_{id} P_{\mathbf{t}}(y)$, for $\mathbf{T} \subseteq T_y^G$. If we remove e from T_y^G, nodes in T_y^G either form an $\{X, Y\}$-rooted C-forest or stay a C-tree in G'. In both cases the inducing path remains. If there is a directed arrow from X to T_y^G other than e, and $X \notin T_y^G$ then there is an inducing path from X to Y in G, and no element in it can be (identifiably) fixed. Moreover, fixing X, even if identifiable, does not cut this path. Finally, if we remove e from G, the inducing path remains.

Assume $X \notin T_y^G$, and the only directed edge from X to T_y^G is e. This implies $T_y^G = T_y^{G'}$, and by definition, $T_x^G = T_x^{G'}$. Since G is acyclic, there cannot be a directed arc from Y to $T_x^{G'}$. What remains is to show there is no bidirected arc from $T_x^{G'}$ to $T_y^{G'}$. But if such an arc does exist, then $T_x^{G'} \cup T_y^{G'}$ would form a Y-rooted C-tree in G, which means $T_x^{G'} \cup T_y^{G'} \subseteq T_y^G$. But we assumed $X \notin T_y^G$. By results in [Shpitser and Pearl, 2008], X, Y are d^*-separable in G' by $\mathbf{Z}, \mathbf{W}$, where $\mathbf{Z} = T_x^{G'} \cup T_y^{G'} \setminus \{X, Y\}$, $\mathbf{W} = (Pa(T_x^{G'}) \cup Pa(T_y^{G'})) \setminus (T_x^{G'} \cup T_y^{G'})$. If $G, P(\mathbf{v}) \not\vdash_{id} P_{\mathbf{w}}(x, y|\mathbf{z})$, then there is a hedge [Shpitser and Pearl, 2006b] ancestral to $\{X, Y\}$ in G for this effect, and it must contain e (or it would exist in G' which we ruled out). But the existence of this hedge implies $T_y^G \neq T_y^{G'}$, which is a contradiction. $\square$

There are simple examples where an edge e between X and Y is not identifiable if it is bidirected, but identifiable if

[3] The reference did not consider stochastic interventions, although the theorem extends to this setting.

[4] This theorem also holds if we drop the requirement of identifiability from both sides of the implication.

[5] In fact, it is possible to relax this assumption somewhat by restricting faithfulness only to those identifiable distributions which can lead to edge identification. However, characterizing all such distributions is outside the scope of this paper (though we give one such distribution for each identifiable edge), so we used a slightly stronger assumption that was easier to state.

[6] This graph can be obtained by applying graphical rules for marginalization found in [Richardson and Spirtes, 2002] to remove variables in $\mathbf{V} \setminus (\mathbf{Y} \cup \mathbf{X})$, and removing all arrows incoming to $\mathbf{X}$.

it is directed. For instance, if we add an edge $X \leftrightarrow Z$ to Fig. 1 (a), the directed edge from X to Y is identifiable using the same reasoning as in our example, while the bidirected edge from X to Y is not identifiable (no sets $\mathbf{Z}, \mathbf{W}$ both separate X and Y and result in an identifiable distribution). The next natural question is the converse, can there be a graph where a bidirected arc can be identified but a directed arc cannot be. It turns out the answer is no. In fact, the same is true if we reverse the direction of a directed arc.

Theorem 5 *Let G be a graph where a directed arc e from X to Y is not identified. Let G' be an acyclic graph obtained from G by replacing e by a bidirected (or directed but reversing direction if this is possible) edge e'. Then e' is not identified in G'.*

Proof: If e between X, Y is not identified, then for any sets $\mathbf{Z}, \mathbf{W}$ either $G, P(\mathbf{v}) \not\vdash_{id} P(x, y|do(\mathbf{w}), \mathbf{z})$ or X, Y are not d^*-separated by conditioning on $\mathbf{Z}$, and fixing $\mathbf{W}$ in the graph $G \setminus \{e\}$, or $do(\mathbf{w})$ cuts e directly. In particular, if the second case is true, then it is also true in the graph $G' \setminus \{e'\}$.

If the first case is true, assume without loss of generality $G, P(\mathbf{v}) \not\vdash_{id} P(x, y, \mathbf{z}|do(\mathbf{w}))$. Results in [Shpitser and Pearl, 2006b] imply there is a hedge for this effect in G. Results in [Shpitser and Pearl, 2008] imply it is sufficient to restrict our attention to $\mathbf{Z} \in An(\{X, Y\})$. Since there is a hedge for $P(x, y, \mathbf{z}|do(\mathbf{w}))$ (e.g. ancestral to $\{X, Y\} \cup \mathbf{Z}$), then there is one ancestral to $\{X, Y\}$. If this hedge does not involve the edge e, then it remains in G' and remains ancestral to $\{X, Y\}$ in G', so our conclusion follows. If the hedge does involve e, then at least one of the X, Y nodes which e connects must be in F (the smaller of the two C-forests of the hedge). Removing e will then preserve the $\{X, Y\}$-rooted status of both C-forests in the hedge, and our conclusion follows.

If the third case is true, then if e becomes bidirected, it stays true. If e points the other way, then due to Theorem 4 either X is an ancestor of Y, in which case we cannot reverse e without introducing a cycle into G', or reversing e creates $T_x^{G'}$ which contains Y. □

Due to Theorem 5, there is no need to give conditions for testing for the absence of bidirected edges, although it is not hard to rephrase Theorem 4 for this purpose (and this may be useful in cases where background knowledge forbids direct effects, but confounding is still desirable to test for).

We also note that, assuming all constraints of the type that appear in Fig. 1 are due to missing edges that are identifiable, Theorem 4 gives a characterization of *dense inducing paths*, that is inducing paths [Verma and Pearl, 1990] which prevent separation of X and Y not only by conditioning but by identifiable interventions.

5 Sufficiency of Truncations

In the previous section, we characterized identification of edges in valid graphs of identifiable faithful models – a notion of identification which corresponds to observable constraints implied by dormant independence. In this section, we give an algorithm for identifying interventional distributions where such constraints appear, show this algorithm complete for this problem, and use this fact to conclude that a particular operation which we call truncation is sufficient for showing such constraints.

The algorithm, shown in Fig. 3, consists of four functions. The function **Test-Edge** is a top level function, and determines the candidate sets $\mathbf{W}, \mathbf{Z}$ to fix and condition respectively in order to d^*-separate X and Y. In order to do so it uses the notion of a maximum ancestral confounded set (MACS), which was shown to be sufficient for this purpose [Shpitser and Pearl, 2008]. The function **Find-MACS** finds the MACS (such sets are unique) for a singleton node argument in a given graph. The function **Truncate-IDC** identifies a conditional interventional distribution by rephrasing the query to be without conditioning, and calling **Truncate-ID**. It was shown that such rephrasing is without loss of generality for conditional effects [Shpitser and Pearl, 2006a]. Finally, the function **Truncate-ID** identifies the resulting distribution by means of two operations: marginalization and truncation. The latter operation consists of dividing by a conditional distribution term $P(x|nd(x))$, where $Nd(X)$ is the set of nondescendants of X.

The key theorem about **Test-Edge** is that it can succeed on every edge identifiable by Theorem 3. Before proving this result, we prove a utility lemma.

Lemma 1 *$P_x(\mathbf{v} \setminus x) = P(\mathbf{v})/P(x|nd(x))$ in G if there is no bidirected path from X to $De(X)$.*

Proof: $P_x(\mathbf{v} \setminus x) = P_x(de(x)|nd(x))P_x(nd(x))$ (by chain rule). $P_x(nd(x)) = P(nd(x))$ by rule 3 of do-calculus [Pearl, 2000]. We claim $P_x(de(x)|nd(x)) = P(de(x)|nd(x), x)$ by rule 2 of do-calculus. We must show there are no d-connected back-door paths from X to $De(x)$ conditioned on $Nd(x)$. Such a path must start either with an directed arrow pointing to X or a bidirected arrow. In the former case, such an arrow will have a tail pointing to an element in $Nd(X)$ which is conditioned on, implying no d-connected path. In the latter case, as long as the path consists entirely of bidirected arcs, it must stay within $Nd(X)$ by assumption. However, the first directed arc on the path will render it d-separated by the above reasoning if the arrow points to the path fragment, and by definition if the arrow points away. □

Theorem 6 *Let e be an edge in a valid graph G of an identifiable faithful model M. Then $P(\mathbf{v}), G \vdash_{id} e$ due to Theorem 3 if and only if **Test-Edge** succeeds. Moreover, the successful output of **Test-Edge** gives an identifiable distribution witnessing identification of e.*

Proof: Soundness of **Find-MACS** is shown in [Shpitser and Pearl, 2008]. Soundness of **Truncate-ID** follows by Lemma 1 and the soundness proof of the **ID** algorithm is found in [Shpitser and Pearl, 2006b]. Soundness of **Truncate-IDC** follows by soundness of do-calculus, and the fact that the input distribution $P^*(\mathbf{v})$ at the point of failure of **Truncate-ID** was soundly identified from $P(\mathbf{v})$. Soundness of **Test-Edge** follows from the soundness of other functions, and results in [Shpitser and Pearl, 2008].

To show completeness, we must show that if **Test-Edges** returns **FAIL**, the preconditions of Theorem 3 fail. If **Test-Edges** returns **FAIL** itself, then by known results X and Y are not d^*-separable [Shpitser and Pearl, 2008]. If

function **Test-Edge**$(P(\mathbf{v}), G, e)$.
INPUT: $P(\mathbf{v})$ a probability distribution, G a causal diagram, e an edge from X to Y.
OUTPUT: Expression for $P_{\mathbf{w}}(x, y|\mathbf{z})$ that tests e in G or **FAIL**.

1 $T^G_x =$ **Find-MACS**(G, X), $T^G_y =$ **Find-MACS**(G, Y).

2 If X is a parent of T^G_y or Y is a parent of T^G_x or there is a bidirected arc between T^G_x and T^G_y, return **FAIL**.

3 Let $\mathbf{T} = T^G_x \cup T^G_y$, $\mathbf{Z} = \mathbf{T} \setminus \{X, Y\}$, $\mathbf{W} = Pa(\mathbf{T}) \setminus \mathbf{T}$.

4 return **Truncate-IDC**$(P(\mathbf{v}), G, \{X, Y\}, \mathbf{W}, \mathbf{Z}, e)$.

function **Find-MACS**(G, Y)
INPUT: G, a causal diagram, Y a node in G.
OUTPUT: T^G_y, the maximum ancestral confounded set for Y in G.

1 If $(\exists X \notin An(Y)_G)$, return **Find-MACS**$(G_{An(Y)}, Y)$.

2 If $(\exists X \notin C(Y)_G)$, return **Find-MACS**$(G_{C(Y)}, Y)$.

3 Otherwise, return G.

function **Truncate-IDC**$(P(\mathbf{v}), G, \mathbf{Y}, \mathbf{X}, \mathbf{Z}, e)$
INPUT: $P(\mathbf{v})$ a probability distribution, G a causal diagram, **X**,**Y**,**Z** variable sets, e an edge between X and Y.
OUTPUT: Expression for $P_{\mathbf{w}}(x, y|\mathbf{z})$ such that $X \perp\!\!\!\perp_{\mathbf{w},G,P} Y|\mathbf{Z}$ or **FAIL**.

1 If $(\exists Z \in \mathbf{Z})(\mathbf{Y} \perp\!\!\!\perp_P Z|\mathbf{X}, \mathbf{Z} \setminus \{Z\})_{G_{\overline{\mathbf{x}},\underline{z}}}$,
return **Truncate-IDC**$(P(\mathbf{v}), G, \mathbf{y}, \mathbf{x} \cup \{z\}, \mathbf{z} \setminus \{z\}, e)$.

2 Let $P' =$ **Truncate-ID**$(P(\mathbf{v}), G \cup \{e\}, \mathbf{y} \cup \mathbf{z}, \mathbf{x})$.

4 Otherwise, return $P'/\sum_{\mathbf{y}} P'$.

function **Truncate-ID**$(P(\mathbf{v}), G, \mathbf{Y}, \mathbf{X})$.
INPUT: $P(\mathbf{v})$ a probability distribution, G a causal diagram, **X**, **Y**, variable sets.
OUTPUT: An expression for $P(\mathbf{y}|do(\mathbf{x}))$ or **FAIL**.

1 If **X** is empty, return $\sum_{\mathbf{v} \setminus \mathbf{y}} P(\mathbf{v})$.

2 If $Na(\mathbf{Y}) = \mathbf{V} \setminus An(\mathbf{Y})$ is not empty, return **Truncate-ID**$(\sum_{na(\mathbf{y})} P(\mathbf{v}), G_{an(\mathbf{y})}, \mathbf{Y}, \mathbf{X} \setminus Na(\mathbf{Y}))$.

3 If $\mathbf{W} = (\mathbf{V} \setminus \mathbf{X}) \setminus An(\mathbf{Y})_{G_{\overline{\mathbf{x}}}}$ is not empty, return **Truncate-ID**$(P(\mathbf{v}), G, \mathbf{Y}, \mathbf{X} \cup \mathbf{W})$.

4 If there is a node $X \in \mathbf{X}$ with no bidirected paths to $De(X)$ in G,
return **Truncate-ID**$(\frac{P(\mathbf{v}) * P^*(x)}{P(x|nd(x))}, G_{\overline{x}}, P(\mathbf{y}|do(\mathbf{x} \setminus \{x\}))$.

5 If $C(X)_G \neq C(Y)_G$, return **Test-Edge**(
Truncate-ID$(P(\mathbf{v}), G, C(Y)_G, C(X)_G), G_{\overline{c(x)_g}}, e)$.

6 Otherwise, return **FAIL**.

Figure 3: An identification algorithm for interventional distributions which lead to edge identification. $G_{C(Y)}$ is the subgraph of G containing the C-component of Y. $G_{\overline{\mathbf{x}},\underline{\mathbf{z}}}$ is the graph obtained from G by removing incoming arrows to **X** and outgoing arrows from **Z**.

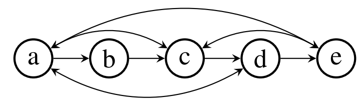

Figure 4: A causal diagram with a Verma constraint between B and E.

Truncate-ID returns **FAIL**, then by construction of the algorithm, the only remaining nodes other than X and Y are nodes Z such that Z is an ancestor of X, Y, Z has a bidirected path to a descendant of Z, and Z is being fixed. But this implies (by induction) that every remaining node has a bidirected path to $\{X, Y\}$.

Since e can be assumed directed, then all remaining nodes are ancestors of either X or Y, say Y. Then these nodes (including X) either form a Y-rooted C-tree, in which case Theorem 4 applies, or there are two C-components, one containing X and another containing Y. In this case, $P(c(y)|do(c(x)))$ is identifiable by truncations, and line 5 of **Truncate-ID** uses this fact to continue the recursion. Either failure case after this step implies either Theorem 4 holds, or X and Y are not d^*-separable – in the original graph.

Finally, though complete identification results in [Shpitser and Pearl, 2006b], [Shpitser and Pearl, 2006a] were for the set of all causal models, it is simple to extend them to also hold in all faithful models. □

We illustrate the operation of the algorithm by testing an edge between B and E in Fig. 4. In this graph, $T^G_b = \{B\}$, and $T^G_e = \{E\}$. Thus, the algorithm tries to fix the parents of $\{B, E\}$, namely $\{A, D\}$. However, both of these nodes have bidirected paths to their descendants (an arc from A to C, and a path $D \leftrightarrow A \leftrightarrow E$), so these interventions cannot be identified by truncations. Instead, since the C-component containing B (which is just B itself) and the C-component containing E are disjoint, the algorithm recurses and truncates out B, e.g. it considers $\frac{P^*(b)}{P(b|a)}P(a, b, c, d, e)$. This distribution is equal to $P_{b \sim P^*}(a, b, c, d, e)$, and corresponds to the graph where the arrow from A to B is cut. The algorithm then proceeds to marginalize A and truncate D, resulting in an expression $\frac{P^{**}(d)}{P(d|b,c)} \sum_a \frac{P^*(b)}{P(b|a)} P(a, b, c, d, e)$ which corresponds to $P_{b \sim P^*, d \sim P^{**}}(b, c, d, e)$, where B and E are independent.

Note that in this example in order to find the independence between B and E we had to fix B itself. Since typically interventions hold the variable constant, an intervention $do(b)$ would make B and E trivially independent. By considering stochastic interventions like $do(b \sim P^*)$, we are able to show that B and E remain independent after appropriate truncations even if B is allowed to vary (although in a way that no longer depends on B's causal ancestors).

The distributions returned by **Test-Edges** can be thought of as resulting from applying a sequence of nested operations, where some operations are marginalizations and others are truncations. What we are going to show is that all truncations can be applied before every marginalization, in which case marginalizations can be dispensed with as they do not affect conditional independence.

Theorem 7 *Let $P(\mathbf{v})$ be a probability distribution, and $P^* = f_1(f_2(...f_k(P(\mathbf{v}))...))$, where $f_i(p)$ is either $\sum_{x_i} p$, for some $X_i \in \mathbf{V}$, or $\frac{p}{P(x_i|z_i)}$, and such that $X_i \neq X_j$ for $i \neq j$, $\mathbf{Z}_i$ is some set not containing either X_i or any X_j mentioned in $f_{i+1}, ..., f_k$ ($\mathbf{Z}_i$ and $\mathbf{Z}_j$ can intersect for $j > i$). Then there exists an ordering $j_1, ..., j_k$ of f_i such that $P^* = f_{j_1}(f_{j_2}(...f_{j_k}(P(\mathbf{v}))...))$, and for any f_{j_i} that is a marginalization, $f_{j_1}, ..., f_{j_{i-1}}$ are also marginalizations.*

Proof: By definition of $\mathbf{Z}_i$, every truncation f_i commutes with every marginalization f_j where $j > i$, while marginalizations commute with each other due to axioms of probability. ▫

Corollary 1 *Let M be an identifiable faithful model with a valid graph G and observable distribution $P(\mathbf{v})$. Then if an edge e is identifiable by Theorem 3, there exists a set of truncations $f_1, ..., f_k$ and a set $\mathbf{Z}$ such that $X \perp\!\!\!\perp_P Y|\mathbf{Z}$ in $f_1(...f_k(P(\mathbf{v}))...)$.*

Proof: This follows from Theorems 6 and 7, the fact that conditional independence $X \perp\!\!\!\perp_P Y|\mathbf{Z}$ holds in $P(\mathbf{v})$ if and only if it holds in $\sum_{\mathbf{w}} P(\mathbf{v})$ for any $\mathbf{W}$ which does not intersect $\{X, Y\} \cup \mathbf{Z}$, and the fact that **Truncate-ID** never mentions nodes it removes on each step in subsequent steps. ▫

Corollary 1 suggests that in order to unearth dormant independence constraints which are due to edge absences in the graph, it is sufficient to consider truncation operations on a probability distribution. For instance, the independence between B and E in the expression $\frac{P^{**}(d)}{P(d|b,c)} \sum_a \frac{P^*(b)}{P(b|a)} P(a, b, c, d, e)$ obtained by running **Test-Edge** on Fig. 4, also holds in the expression $\frac{P^{**}(d)}{P(d|b,c)} \frac{P^*(b)}{P(b|a)} P(a, b, c, d, e)$.

We hasten to add that since Theorem 7 "pushes truncations inward," and Corollary 1 ignores marginalizations, the independences we observe in the distributions obtained by truncations alone using these two results can lose their causal interpretability as independences in identifiable interventional distributions, since such distributions rely on truncations and marginalizations performed in a certain order. In particular, the expression above with two nested truncations does not correspond in an obvious way to any identifiable interventional distribution. Losing this interpretability is the price we pay for restricting the set of operations we consider.

What remains is to characterize all the constraints between two variables which are discoverable via truncation operations, and show, as we conjecture, that under reasonable faithfulness assumptions such constraints can only arise due to edge absences. The proof of this conjecture would allow the use of dormant independence in causal discovery, although unfortunately it is outside the scope of this paper.

Conclusion

In this paper we consider probabilistic constraints due to dormant independence which can be used to test the presence of edges in a causal diagram. We characterized when these constraints arise, and give an algorithm which finds them, given a graph where every edge is either correct or extraneous, i.e. the graph is a (possibly non-proper) supergraph of the true graph. Furthermore, we showed that applying an operation we call truncation in sequence to an observable probability distribution is sufficient to unearth all constraints of this type.

What remains to show before dormant independence constraints can be used in causal discovery algorithms is that sequences of truncation operations only lead to types of constraints we characterize and no others, given an appropriate notion of faithfulness. Proving this conjecture, and developing a causal discovery algorithm based on these ideas are the next steps in our investigation.

References

[Cooper and Dietterich, 1992] Gregory F. Cooper and Tom Dietterich. A Bayesian method for the induction of probabilistic networks from data. In *Machine Learning*, pages 309–347, 1992.

[Pearl and Robins, 1995] Judea Pearl and J. M. Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Uncertainty in Artificial Intelligence*, volume 11, pages 444–453, 1995.

[Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan and Kaufmann, San Mateo, 1988.

[Pearl, 2000] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[Richardson and Spirtes, 2002] Thomas Richardson and Peter Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30:962–1030, 2002.

[Robins, 1986] James M. Robins. A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512, 1986.

[Robins, 1999] James M. Robins. Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models. *Computation, Causation, and Discovery*, pages 349–405, 1999.

[Shpitser and Pearl, 2006a] Ilya Shpitser and Judea Pearl. Identification of conditional interventional distributions. In *Uncertainty in Artificial Intelligence*, volume 22, 2006.

[Shpitser and Pearl, 2006b] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Twenty-First National Conference on Artificial Intelligence*, 2006.

[Shpitser and Pearl, 2008] Ilya Shpitser and Judea Pearl. Dormant independence. In *Proceedings of the Twenty-Third Conference on Artificial Intelligence*, pages 1081–1087, 2008.

[Spirtes *et al.*, 1993] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. Springer Verlag, New York, 1993.

[Suzuki, 1993] J. Suzuki. A construction of Bayesian networks from databases based on an mdl scheme. In *UAI 93*, pages 266–273, 1993.

[Tian and Pearl, 2002a] Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Eighteenth National Conference on Artificial Intelligence*, pages 567–573, 2002.

[Tian and Pearl, 2002b] Jin Tian and Judea Pearl. On the testable implications of causal models with hidden variables. In *Proceedings of UAI-02*, pages 519–527, 2002.

[Verma and Pearl, 1990] T. S. Verma and Judea Pearl. Equivalence and synthesis of causal models. Technical Report R-150, Department of Computer Science, University of California, Los Angeles, 1990.