Revisiting the General Identifiability Problem

Abstract

We revisit the problem of general identifiability originally introduced in [Lee et al., 2019] for causal inference and note that it is necessary to add positivity assumption of observational distribution to the original definition of the problem. We show that without such an assumption the rules of do-calculus and consequently the proposed algorithm in [Lee et al., 2019] are not sound. Moreover, adding the assumption will cause the completeness proof in [Lee et al., 2019] to fail. Under positivity assumption, we present a new algorithm that is provably both sound and complete. A nice property of this new algorithm is that it establishes a connection between general identifiability and classical identifiability by Pearl [1995] through decomposing the general identifiability problem into a series of classical identifiability sub-problems.

1 INTRODUCTION

Causal effect identification (or ID for short) problem, a central concern in causal inference, pertains to whether given a causal graph an interventional distribution can be uniquely computed from observational distribution [Pearl, 2009]. When all the variables in the system are observable, Pearl's do-calculus (a collection of three rules), allows to determine whether a causal effect is identifiable [Pearl, 1995]. Moreover, it was shown that Pearl's do-calculus is both sound and complete for ID problem [Shpitser and Pearl, 2006a, Huang and Valtorta, 2008].

There are two main assumptions in the classical setting of ID problem: I) both the causal graph and the observational distribution, denoted by $P(\mathbf{V})$ ($\mathbf{V}$ is the set of observed variables in the causal graph) are given, and II) no interventional distribution are available. Several recent work in the literature[1] relax these assumptions such as [Shpitser and Pearl, 2006b, Bareinboim and Tian, 2015, Bareinboim and Pearl, 2014]. Before, discussing these results, let us introduce a notion. We denote by $P_{\mathbf{x}}(\mathbf{Y})$, the distribution of set of variables $\mathbf{Y}$ resulting from intervening on another set of variables $\mathbf{X}$. Bareinboim and Pearl [2012] introduced the z-identification problem (or zID for short) in which for a fixed set $\mathbf{Z} \subseteq \mathbf{V}$, given a set of interventional distributions of the form $\{P_{\mathbf{z}'}(\mathbf{V} \setminus \mathbf{Z}') : \forall \mathbf{Z}' \subseteq \mathbf{Z}\}$, we ask whether $P_{\mathbf{x}}(\mathbf{Y})$ is identifiable. Note that the observational distribution $P(\mathbf{V})$ always belong to the set of available distributions. Lee et al. [2019] generalized zID and proposed so-called general identifiability problem (or gID for short). In the gID, observational distribution is not necessarily given but instead we have access to $\{P_{\mathbf{z}_i}(\mathbf{V} \setminus \mathbf{Z}_i)\}_{i=0}^m$ for some subsets $\{\mathbf{Z}_i\}_{i=0}^m$ of observed variables. When one of Z_is is an empty set, then we have access to $P(\mathbf{V})$.

We give formal definitions of identifiability (Definition 3) and general identifiability (Definition 5) in Section 2. An important contribution of this paper is to add an assumption on the positivity of the observational distribution in these definitions, i.e., $P(\mathbf{v}) > 0$ for all the realizations of observed variables. As we shall discuss in detail in Section 3, this assumption is crucial. More specifically, do-calculus-based methods are no longer sound for the ID problem if we ignore the positivity assumption. In other words, there exist causal graphs with non-positive distribution $P(\mathbf{V})$ such that do-calculus would claim a causal effect is identifiable while it cannot be uniquely computed from mere observational distribution. An even more important reason for including the positivity assumption is that without it the proposed algorithm in the original gID in [Lee et al., 2019] is not sound and complete. As we shall discuss in Section 3, the proof of completeness in [Lee et al., 2019] relies on building two models that have zero probabilities for certain realizations of observed variables. Therefore, unfortunately, simply adding the positivity assumption to the definition of general identifiability (g-identifiability) will result in the failure of

[1] See [Tikka et al., 2021] for an overview.

Submitted to the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022). **To be used for reviewing only**.

the proof technique in [Lee et al., 2019]. On the other hand, ignoring the positivity assumption makes the soundness of their algorithm incorrect.

In summary, our main contributions are as fallows. We redefine the g-identifiability by adding the positivity assumption of observational distribution (Definition 5). We show in Section 3 that this assumption is essential for the gID problem. We then provide a sound and complete algorithm for the gID problem (Algorithm 2). A nice property of our algorithm is that it establishes a connection between gID and classical ID problem by showing that gID can be reduced to solving a series of ID problems (Theorem 1).

2 PRELIMINARIES

2.1 TERMINOLOGY

Throughout the paper, we denote random variables by capital letters (e.g. X), their realizations by small letters (e.g. x), and sets by bold letters (e.g. $\mathbf{X}$ or $\mathbf{x}$). We use $\mathfrak{X}_X$ to denote the domain of random variable X and $\mathfrak{X}_{\mathbf{X}}$ to denote the Cartesian product of the domains of all the variables in set $\mathbf{X}$, i.e., $\prod_{X\in\mathbf{X}} \mathfrak{X}_X$. For integer numbers $a \leq b$, we use $[a:b]$ to denote $\{a, a+1, \cdots, b\}$.

Suppose $\mathcal{G} = (\mathbf{V}\cup\mathbf{U}, \mathbf{E})$ is a directed acyclic graph (DAG) over vertex set $\mathbf{V}\cup\mathbf{U}$, where $\mathbf{V}$ and $\mathbf{U}$ represent the set of observed and unobserved variables, respectively. For each edge $(X, Y) \in \mathbf{E}$, X is called a parent of Y and Y is called a child of X. Vertex X is an ancestor of Y in $\mathcal{G}$ if there exists a directed path from X to Y in $\mathcal{G}$. Note that X is an ancestor of itself. $Pa_{\mathcal{G}}(X)$, $Ch_{\mathcal{G}}(X)$, and $Anc_{\mathcal{G}}(X)$ denote the set of parents, children, and ancestors of X in $\mathcal{G}$, respectively. These notations are also used for a set of vertices. In this case, they refer to the union over the elements of the set. For instance, $Pa_{\mathcal{G}}(\mathbf{X}) = \bigcup_{X\in\mathbf{X}} Pa_{\mathcal{G}}(X)$. We assume $\mathcal{G}$ is semi-Markovian, that is for each $U \in \mathbf{U}$, $Pa_{\mathcal{G}}(U) = \varnothing$ and $|Ch_{\mathcal{G}}(U)| = 2$. Note that this is not a restrictive assumption as there exists an equivalency for identifiability in DAGs and semi-Markovian DAGs [Huang and Valtorta, 2006].

Structural Equation Models (SEMs) are used to model causal systems [Pearl, 2009]. $\mathcal{G}$ is a causal graph for SEM $\mathcal{M}$ if each $X \in \mathbf{V}\cup\mathbf{U}$ is generated as $f_X(Pa_{\mathcal{G}}(X), \epsilon_X)$, where $\{\epsilon_X : X \in \mathbf{V}\}$ is a set of mutually independent exogenous random variables. We denote by $P^{\mathcal{M}}(\cdot)$ the joint distribution of the variables in $\mathcal{M}$ and drop the superscript $\mathcal{M}$ when it is clear from the context. Markov factorization property implies that $P^{\mathcal{M}}(\cdot)$ can get factorized as

$$P(\mathbf{v}) = \sum_{\mathbf{U}} \prod_{X\in\mathbf{V}} P(x|Pa_{\mathcal{G}}(X)) \prod_{U\in\mathbf{U}} P(u), \qquad (1)$$

where $\sum_{\mathbf{U}}$ denotes the marginalization over $\mathbf{U}$.

Definition 1. *$\mathbb{M}(\mathcal{G})$ denotes the set of SEMs with causal graph $\mathcal{G}$. $\mathbb{M}^+(\mathcal{G})$ denotes the set of SEMs $\mathcal{M} \in \mathbb{M}(\mathcal{G})$ such that $P^{\mathcal{M}}(\mathbf{v}) > 0$ for each $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}}$.*

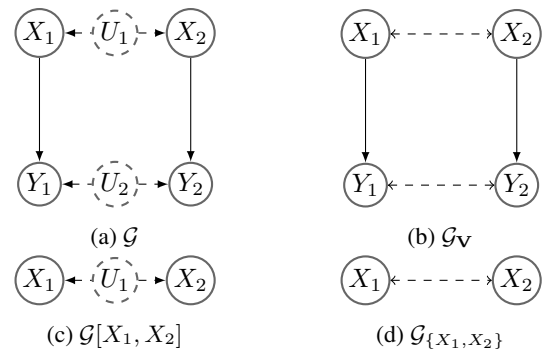


Figure 1: An example for a causal DAG $\mathcal{G}$ over observed variables $\mathbf{V} = \{X_1, X_2, Y_1, Y_2\}$ and unobserved variables $\mathbf{U} = \{U_1, U_2\}$.

For $\mathbf{X} \subseteq \mathbf{V}$ and $\mathbf{x} \in \mathfrak{X}_{\mathbf{X}}$, the intervention $do(\mathbf{X} = \mathbf{x})$ converts $\mathcal{M}$ to a new SEM where the equations of $\mathbf{X}$ in $\mathcal{M}$ are replaced by the constants in $\mathbf{x}$. We denote by $P_{\mathbf{x}}(\cdot)$ the corresponding post interventional distribution.

Remark 1. *For three disjoint subsets $\mathbf{X}, \mathbf{Y}, \mathbf{W}$ of $\mathbf{V}$, if $\mathcal{M} \in \mathbb{M}^+(\mathcal{G})$, then $P^{\mathcal{M}}_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w}) > 0$ for any $\mathbf{x} \in \mathfrak{X}_{\mathbf{X}}$, $\mathbf{y} \in \mathfrak{X}_{\mathbf{Y}}$, and $\mathbf{w} \in \mathfrak{X}_{\mathbf{W}}$.*

For $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}}$ and $\mathbf{S} \subseteq \mathbf{V}$, we define $Q[\mathbf{S}](\cdot)$ by

$$Q[\mathbf{S}](\mathbf{v}) := P_{\mathbf{v}\setminus\mathbf{s}}(\mathbf{s}). \qquad (2)$$

Similar to Equation (1), $Q[\mathbf{S}]$ can get factorized as

$$Q[\mathbf{S}](\mathbf{v}) = \sum_{\mathbf{U}} \prod_{S\in\mathbf{S}} P(s|Pa_{\mathcal{G}}(S)) \prod_{U\in\mathbf{U}} P(u). \qquad (3)$$

For $\mathbf{X} \subseteq \mathbf{V}$, $\mathcal{G}[\mathbf{X}]$ denotes the inducing subgraph of $\mathcal{G}$ over $\mathbf{X}$ and the unobserved variables with both children in $\mathbf{X}$. Note that $\mathcal{G}$ is semi-Markovian. Furthermore, we denote by $\mathcal{G}_{\mathbf{X}}$ the partially directed graph over $\mathbf{X}$ obtained by removing unobserved variables of $\mathcal{G}[\mathbf{X}]$ and replacing them by bidirected edges.

Definition 2 (**c-component, c-forest**). *For $\mathbf{X} \subseteq \mathbf{V}$, confounded components or c-components of $\mathbf{X}$ are the connected components of the graph obtained by only the bidirected edges of $\mathcal{G}_{\mathbf{X}}$. Also, a subgraph of $\mathcal{G}_{\mathbf{V}}$ is called a single c-component if its bidirected edges form a connected graph. Suppose $\mathcal{H}$ is a subgraph of $\mathcal{G}$ over observed vertices $\mathbf{X}$. The root set of $\mathcal{H}$ is the maximal subset of $\mathbf{X}$ with no children in $\mathcal{H}$. $\mathcal{H}$ is called $\mathbf{R}$-rooted c-forest if $\mathbf{R}$ is the root set of $\mathcal{H}$, $\mathcal{H}_{\mathbf{X}}$ is a single c-component, and each node in $\mathbf{X}$ has at most one child in $\mathcal{H}$.*

Example 1: Consider the causal graph $\mathcal{G}$ in Figure 1a, where $\mathbf{V} = \{X_1, X_2, Y_1, Y_2\}$ and $\mathbf{U} = \{U_1, U_2\}$. $\mathcal{G}_{\mathbf{V}}$ is depicted

in Figure 1b. The c-components of $\mathbf{V}$ are $\{X_1, X_2\}$ and $\{Y_1, Y_2\}$. Figure 1c depicts the inducing subgraph of $\mathcal{G}$ over $\{X_1, X_2\}$ and $\mathcal{G}_{\{X_1,X_2\}}$ is depicted in Figure 1d. Herein, $\mathcal{G}[\{X_1, X_2\}]$ is $\{X_1, X_2\}$-rooted c-forest since $\mathcal{G}_{\{X_1,X_2\}}$ is single c-component.

2.2 IDENTIFIABILITY

The goal in the *identifiability* problem is to understand whether a post-interventional distribution can be uniquely computed from observational distribution $P(\mathbf{V})$, given the causal graph [Pearl, 2009].

Definition 3 (**identifiability**)**.** *Suppose* $\mathbf{X}$ *and* $\mathbf{Y}$ *are two disjoint subsets of* $\mathbf{V}$*. The causal effect of* $\mathbf{X}$ *on* $\mathbf{Y}$ *is said to be identifiable from* $\mathcal{G}$ *if for any* $\mathbf{x} \in \mathfrak{X}_{\mathbf{X}}$ *and* $\mathbf{y} \in \mathfrak{X}_{\mathbf{Y}}$*,* $P^{\mathcal{M}}_{\mathbf{x}}(\mathbf{y})$ *is uniquely computable from* $P^{\mathcal{M}}(\mathbf{V})$ *in any SEM* $\mathcal{M} \in \mathbb{M}^{+}(\mathcal{G})$*. Also,* $Q[\mathbf{Y}]$ *is said to be identifiable from* $\mathcal{G}$ *if the causal effect of* $\mathbf{V} \setminus \mathbf{Y}$ *on* $\mathbf{Y}$ *is identifiable from* $\mathcal{G}$*.*

Huang and Valtorta [2008] showed that identifiability of a causal effect is equivalent to identifiability of a specific $Q[\cdot]$.

Proposition 1 (Huang and Valtorta [2008])**.** *The causal effect of* $\mathbf{X}$ *on* $\mathbf{Y}$ *is identifiable from* $\mathcal{G}$ *if and only* $Q[Anc_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}(\mathbf{Y})]$ *is identifiable from* $\mathcal{G}$*.*

For a subset $\mathbf{S}$ of observed nodes, [Tian and Pearl, 2003] showed that identifiability of a $Q[\mathbf{S}]$ is equivalent to identifiability of all its c-components.

Proposition 2 (Tian and Pearl [2003])**.** *Suppose* $\mathbf{S}_1, \cdots, \mathbf{S}_l$ *are the c-components of* $\mathbf{S} \subseteq \mathbf{V}$*.* $Q[\mathbf{S}]$ *is identifiable from* $\mathcal{G}$ *if and only if* $Q[\mathbf{S}_i]$ *is identifiable from* $\mathcal{G}$ *for all* $i \in [1 : l]$*.*

Based on Propositions 1 and 2, [Tian and Pearl, 2003] proposed an algorithm that for two disjoint subsets $\mathbf{X}$ and $\mathbf{Y}$ checks the identifiability of the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ from observational distribution given the causal graph $\mathcal{G}$. As we will use their algorithm as a subroutine in our algorithm for g-identifiability, we present their method in Algorithm 1. In this algorithm, function **ID_Single** determines whether $Q[\mathbf{S}]$ is identifiable from $\mathcal{G}$ when $\mathbf{S}$ is a single c-component. More precisely, this function starts from $\mathbf{Y} = \mathbf{V}$ and at each step, it decreases $\mathbf{Y}$ such that both $Q[\mathbf{Y}]$ remains identifiable from $\mathcal{G}$ and $\mathbf{S} \subseteq \mathbf{Y}$. If this procedure can reduce $\mathbf{Y}$ to $\mathbf{S}$, then $Q[\mathbf{S}]$ is identifiable otherwise, $Q[\mathbf{S}]$ is not identifiable. This algorithm is both sound and complete [Shpitser and Pearl, 2006a, Huang and Valtorta, 2008].

Algorithm 1: Identifiability

Function ID$(\mathbf{X}, \mathbf{Y}, \mathcal{G})$
Output: True, if the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is identifiable from $\mathcal{G}$.
$\mathbf{S} \leftarrow Anc_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}(\mathbf{Y})$
$\{\mathbf{S}_1, \ldots, \mathbf{S}_l\} \leftarrow$ c-components of $\mathbf{S}$
for i from 1 to l **do**
  if ID_Single$(\mathbf{S}_i, \mathcal{G})$ = False **then**
    Return False
Return True

Function ID_Single$(\mathbf{S}, \mathcal{G})$
Output: True, if $Q[\mathbf{S}]$ is identifiable from $\mathcal{G}$, where $\mathbf{S}$ is a single c-component.
$\mathbf{Y} \leftarrow \mathbf{V}$
while $\mathbf{Y} \neq \mathbf{S}$ **do**
  $\mathbf{A} \leftarrow Anc_{\mathcal{G}_{\mathbf{Y}}}(\mathbf{S})$
  $\mathbf{Y}_{new} \leftarrow$ The c-component of $\mathbf{A}$ that contains $\mathbf{S}$
  if $\mathbf{Y}_{new} = \mathbf{Y}$ **then**
    Return False
  else
    $\mathbf{Y} \leftarrow \mathbf{Y}_{new}$
Return True

2.3 GENERAL IDENTIFIABILITY

In the previous section, we explained the classical identifiability problem which determines whether a causal effect is identifiable from observational distribution given the causal graph. As we discussed earlier, in many problems of interest, the goal is to identify a causal effect from a set of both observational and interventional distributions given a causal graph. A variant of the this problem was defined by Lee et al. [2019] under the name g-identifiability.

Definition 4 (g-identifiability in [Lee et al., 2019])**.** *Let* $\mathbf{X}, \mathbf{Y}$ *be disjoint subsets of* $\mathbf{V}$*,* $\mathbb{Z} = \{\mathbf{Z}_i\}_{i=0}^{m}$ *be a collection of subsets of* $\mathbf{V}$*, and let* $\mathcal{G}$ *be a causal diagram.* $P_{\mathbf{x}}(\mathbf{y})$ *is said to be g-identifiable from* $\mathbb{Z}$ *in* $\mathcal{G}$*, if* $P_{\mathbf{x}}(\mathbf{y})$ *is uniquely computable from distributions* $\{P(\mathbf{V}|do(\mathbf{z}))\}_{\mathbf{Z}\in\mathbb{Z},\mathbf{z}\in\mathfrak{X}_{\mathbf{Z}}}$ *in any causal model which induces* $\mathcal{G}$*.*

Note that the causal model in this definition belongs to $\mathbb{M}(\mathcal{G})$. However, as we shall discuss in Section 3, it is crucial to assume that the causal model is positive, i.e., it belongs to $\mathbb{M}^{+}(\mathcal{G})$. Therefore, we modify the above definition as follows.

Definition 5 (g-identifiability)**.** *Suppose* $\mathbb{A} = \{\mathbf{A}_i\}_{i=0}^{m}$ *is a collection of subsets of* $\mathbf{V}$ *and* $\mathbf{X}, \mathbf{Y}$ *are two disjoint subsets of* $\mathbf{V}$*. The causal effect of* $\mathbf{X}$ *on* $\mathbf{Y}$ *is said to be g-identifiable from* $(\mathbb{A}, \mathcal{G})$ *if for any* $\mathbf{x} \in \mathfrak{X}_{\mathbf{X}}$ *and* $\mathbf{y} \in \mathfrak{X}_{\mathbf{Y}}$*,* $P^{\mathcal{M}}_{\mathbf{x}}(\mathbf{y})$ *is uniquely computable from the set of distributions* $\{Q[\mathbf{A}_i]\}_{i=0}^{m}$ *in any SEM* $\mathcal{M} \in \mathbb{M}^{+}(\mathcal{G})$*. Also,* $Q[\mathbf{Y}]$ *is said to be g-identifiable from* $(\mathbb{A}, \mathcal{G})$ *if the causal effect of* $\mathbf{V} \setminus \mathbf{Y}$ *on* $\mathbf{Y}$ *is g-identifiable from* $(\mathbb{A}, \mathcal{G})$*.*

Note that knowing $P(\mathbf{V}|do(\mathbf{Z}))$ for some subset $\mathbf{Z} \subseteq \mathbf{V}$ is equivalent to knowing $Q[\mathbf{V} \setminus \mathbf{Z}]$, and therefore, by setting $\mathbf{A}_i = \mathbf{V} \setminus \mathbf{Z}_i$, the two aforementioned definitions are the

same except for the positivity assumption. For the remainder of this paper, we use Definition 5 for g-identifiability.

3 ON THE POSITIVITY ASSUMPTION IN g-IDENTIFIABILITY

In our definition of g-identifiability (Definitions 5) and the classical definition of identifiability (Definition 3), only SEMs that belong to $\mathbb{M}^+(\mathcal{G})$ instead of $\mathbb{M}(\mathcal{G})$ are considered [Huang and Valtorta, 2008, Shpitser and Pearl, 2006a]. That is, SEMs with positive probabilities for any realization $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}}$. In this section, we discuss why this assumption is crucial by showing that ignoring positivity leads to wrong conclusions. As a consequence, since Lee et al. [2019] presented the soundness and completeness of their algorithm for g-identifiability ignoring the positivity assumption, we discuss how after imposing the assumption, their results are no longer valid. After this discussion, we present a new algorithm for g-identifiability and prove its soundness and completeness under the positivity assumption.

Next example shows that do-calculus-based methods (e.g., Algorithm 1) are no longer sound for the ID problem ignoring the positivity assumption.

Example 2: Consider again the causal graph in Figure 1. Herein, do-calculus-based methods (e.g., Algorithm 1) would report that the causal effect of $\mathbf{X} = \{X_1, X_2\}$ on $\mathbf{Y} = \{Y_1, Y_2\}$ is identifiable given $\mathcal{G}$. However, by ignoring the positivity assumption, we can introduce two SEMs $\mathcal{M}_1$ and $\mathcal{M}_2$ in $\mathbb{M}(\mathcal{G})$ that have the same observational distribution but result in two different post-interventional distributions after intervening on $\{X_1, X_2\}$. This clearly contradicts with the identifiability of $P_{x_1,x_2}(y_1, y_2)$.

All variables in both models are binary. Also, for both models and $i \in \{1, 2\}$, we define $P(U_i = 0) = P(U_i = 1) = 0.5$ and $X_i = U_1$. In model $\mathcal{M}_1$, we define Y_1, Y_2 to have the following conditional distributions:

$$P^{\mathcal{M}_1}(y_1 \mid u_2, x_1) = \frac{1}{3}\mathbb{1}_{y_1=u_2} + \frac{2}{3}\mathbb{1}_{y_1\neq u_2},$$
$$P^{\mathcal{M}_1}(y_2 \mid u_2, x_2) = \frac{1}{3}\mathbb{1}_{y_2=(u_2\oplus x_2)} + \frac{2}{3}\mathbb{1}_{y_2\neq(u_2\oplus x_2)},$$

where $\mathbb{1}_A$ is the indicator function which is one whenever the statement in A is true and is zero otherwise. For model $\mathcal{M}_2$, we define the conditional distributions of Y_1, Y_2 as

$$P^{\mathcal{M}_2}(y_1 \mid u_2, x_1) = \frac{2}{3}\mathbb{1}_{y_1=(u_2\oplus x_1)} + \frac{1}{3}\mathbb{1}_{y_1\neq(u_2\oplus x_1)},$$
$$P^{\mathcal{M}_2}(y_2|u_2, x_2) = \frac{2}{3}\mathbb{1}_{y_2=u_2} + \frac{1}{3}\mathbb{1}_{y_2\neq u_2}.$$

It is straightforward to see that for any realizations $(x_1, x_2, y_1, y_2) \in \mathfrak{X}_{\mathbf{V}}$, we have

$$P^{\mathcal{M}_1}(x_1, x_2, y_1, y_2) = P^{\mathcal{M}_2}(x_1, x_2, y_1, y_2).$$

However,

$$\frac{4}{9} = P^{\mathcal{M}_1}_{x_1=0,x_2=1}(Y_1 = 0, Y_2 = 0)$$
$$\neq P^{\mathcal{M}_2}_{x_1=0,x_2=1}(Y_1 = 0, Y_2 = 0) = \frac{5}{9}.$$

Note that $\mathcal{M}_1$ and $\mathcal{M}_2$ do not belong to $\mathbb{M}^+(\mathcal{G})$, since $P(x_1 = 0, x_2 = 1, y_1, y_2) = 0$ for any $y_1 \in \mathfrak{X}_{Y_1}$ and $y_2 \in \mathfrak{X}_{Y_2}$. This example shows that if we use $\mathbb{M}(\mathcal{G})$ instead of $\mathbb{M}^+(\mathcal{G})$ in Definition 3, the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is not identifiable from $\mathcal{G}$, and therefore, do-calculus-based methods such as the proposed algorithm in Lee et al. [2019] are not sound. Specifically, the proposed algorithm in Lee et al. [2019] suggests the causal effect in this example is g-identifiable and returns the following expression:

$$P_{x_1,x_2}(y_1, , y_2) = P(y_1|x_1, x_2)P(y_2|y_1, x_2, x_1).$$

This expression is not well-defined for all realizations ignoring the positivity assumption because for some realizations $P(x_1, x_2)$ is zero which means the conditional distribution $P(y_1|x_1, x_2)$ is not well-defined. Thus, the algorithm in Lee et al. [2019] *is not sound*.

Next, we discuss the g-identifiability in Lee et al. [2019] and show that the completeness result provided in that work relies on two models in $\mathbb{M}(\mathcal{G})$ that violate the positivity assumption.

[Lee et al., 2019] presented necessary and sufficient conditions to determine if a causal effect $P_{\mathbf{x}}(\mathbf{y})$ is g-identifiable w.r.t. the Definition 4. To prove that their proposed conditions are necessary for g-identifiability, they construct two models $\mathcal{M}_1$ and $\mathcal{M}_2$ such that the available distributions in the definition of the problem are the same for both models yet $P^{\mathcal{M}_1}_{\mathbf{x}}(\mathbf{y}) \neq P^{\mathcal{M}_2}_{\mathbf{x}}(\mathbf{y})$. The issue here is that ignoring the positivity assumption, they constructed their models allowing for zero probability for some realizations. In fact, having zero probabilities in their model is essential for the proof. For instance, Lemma 3 in Lee et al. [2019] states that under certain conditions, there is an observed variable $R \in \mathbf{V}$ such that it takes value zero in both their models with probability one. In other words, the probability of R not being zero is zero (see Appendix A.2 for more details.)

It is noteworthy to mention that an alternative positivity assumption is introduced by Shpitser and Pearl [2006a]. Below, we describe this assumption and discuss that the models introduced in Lee et al. [2019] also violate this assumption.

Shpitser and Pearl [2006a] claim that in the ID problem of a causal effect $P_{\mathbf{x}}(\mathbf{y})$, one can relax the positivity constraint $P(\mathbf{V}) > 0$ to $P(\mathbf{X}|\,(Pa_{\mathcal{G}}(\mathbf{X}) \cap \mathbf{V}) \setminus \mathbf{X}) > 0$. However, as we mentioned, even the relaxed constraint does not hold for the constructed models in Lee et al. [2019]. More precisely, consider the causal graph $\mathcal{G}$ in Figure 2 which is brought here from Lee et al. [2019]. Assume that we are interested

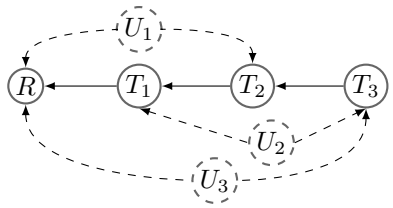


Figure 2: A causal graph of [Lee et al., 2019] that shows the violation of relaxed positivity assumption in constructed models of [Lee et al., 2019].

in g-identifying the causal effect $Q[R]$ from $\mathbb{Z} = \{\varnothing\}$, i.e., from mere observational distribution $P(\mathbf{V})$, w.r.t. Definition 4. In this case, $\mathbf{X} = \{T_1, T_2, T_3\}$ and therefore: $P(\mathbf{X}|\,(Pa_{\mathcal{G}}(\mathbf{X}) \cap \mathbf{V}) \setminus \mathbf{X}) = P(T_1, T_2, T_3)$.

The result in Lee et al. [2019] implies that the causal effect $Q[R]$ is not g-identifiable given the causal graph $\mathcal{G}$ in Figure 2. To prove the non g-identifiability, Lee et al. [2019] constructed two models $\mathcal{M}_1$ and $\mathcal{M}_2$ that impose similar observational distributions, i.e., $P^{\mathcal{M}_1}(\mathbf{V}) = P^{\mathcal{M}_2}(\mathbf{V})$, while the causal effect $Q[R]$ under these two models are not the same for at least one realization. Next, we present these two models and show that they violate the positivity assumption claimed in Shpitser and Pearl [2006a], i.e., $P(T_1, T_2, T_3)$ is zero for certain realizations of $\{T_1, T_2, T_3\}$.

By the construction in Lee et al. [2019], variables T_3, U_1, U_2, U_3 are binary variables and T_1, T_2 are binary vectors of length two. For both models, all unobserved variables are defined to be binary with uniform distribution and the observed variables T_1, T_2, T_3 are defined as follows,

$$\begin{aligned} &T_3 = U_2 \oplus U_3, \\ &T_{2,1} = T_3, \quad T_{2,2} = U_1, \\ &T_{1,1} = T_{2,1} \oplus U_2, \quad T_{1,2} = T_{2,2}. \end{aligned}$$

In model $\mathcal{M}_1$, variable R is defined as

$$R = \mathbb{1}_{T_{1,1}=0} \wedge \mathbb{1}_{T_{1,2}=0} \wedge \mathbb{1}_{U_3=1} \wedge \mathbb{1}_{U_1=1},$$

and in model $\mathcal{M}_2$, it is defined to be zero, i.e., $R = 0$.

Given the above models, it is clear that the probability $P(t_1, t_2, t_3)$ is equal to zero whenever $t_{2,1} \neq t_3$, and therefore, the relaxed positivity constraint $P(T_1, T_2, T_3) > 0$ does not hold for the models in Lee et al. [2019]. See Appendix A for more details.

To summarize, in this section, our goal was to prove the importance of positivity assumption in both classical ID and its generalization gID. We did so, by showing that the rules of *do-calculus* and consequently the proposed algorithm in Lee et al. [2019] are not sound without the positivity assumption. Moreover, we discussed that the completeness proof in Lee et al. [2019] only holds when there is no positivity assumption. This motivates our work to revisit the gID problem by including the positivity assumption in the definition of gID, and presenting a new algorithm which is provably sound and complete.

Algorithm 2: g-identifiability

Function GID($\mathbf{X}, \mathbf{Y}, \mathbb{A} = \{\mathbf{A}_i\}_{i=0}^m, \mathcal{G}$)
Output: True, if the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.
$\mathbf{S} \leftarrow Anc_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}(\mathbf{Y})$
$\{\mathbf{S}_1, \ldots, \mathbf{S}_l\} \leftarrow$ c-components of $\mathbf{S}$
for i from 1 to l **do**
  if GID_Single($\mathbf{S}_i, \mathbb{A} = \{\mathbf{A}_i\}_{i=0}^m, \mathcal{G}$) = False **then**
    Return False
Return True

Function GID_Single($\mathbf{S}, \mathbb{A} = \{\mathbf{A}_i\}_{i=0}^m, \mathcal{G}$)
Output: True, if $Q[\mathbf{S}]$ is identifiable from $(\mathbb{A}, \mathcal{G})$, where $\mathbf{S}$ is a single c-component.
for i from 0 to m **do**
  if $\mathbf{S} \subseteq \mathbf{A}_i$ and **ID_Single**($\mathbf{S}, \mathcal{G}[\mathbf{A}_i]$) = True **then**
    Return True
Return False

4 AN ALGORITHM FOR gID

In this section, we propose an algorithm for gID from $(\mathbb{A}, \mathcal{G})$, where $\mathbb{A} = \{\mathbf{A}_i\}_{i=0}^m$ is a collection of subsets of $\mathbf{V}$. To this end, we first extend Propositions 1 and 2 from identifiability to g-identifiability.

Proposition 3. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two disjoint subsets of $\mathbf{V}$. The causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[Anc_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}(\mathbf{Y})]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.*

Proposition 4. *Suppose $\mathbf{S}_1, \cdots, \mathbf{S}_l$ are the c-components of $\mathbf{S} \subseteq \mathbf{V}$. $Q[\mathbf{S}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{S}_i]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ for all $i \in [1 : l]$.*

Proofs are provided in Appendix B. Proposition 3 allows us to solve the gID problem for $P_{\mathbf{x}}(\mathbf{Y})$ by solving an equivalent problem for $Q[\mathbf{S}]$, where $\mathbf{S}$ is given in the same proposition. Proposition 4 shows that the g-identifiability of $Q[\mathbf{S}]$ from $(\mathbb{A}, \mathcal{G})$ is equivalent to g-identifiability of its single c-components. Next result provide a method for solving g-identifiability of $Q[\mathbf{S}]$, when $\mathbf{S}$ is a single c-component.

Theorem 1. *Suppose $\mathbf{S} \subseteq \mathbf{V}$ is a single c-component. $Q[\mathbf{S}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if there exists $\mathbf{A} \in \mathbb{A}$ such that $\mathbf{S} \subseteq \mathbf{A}$ and $Q[\mathbf{S}]$ is identifiable from $\mathcal{G}[\mathbf{A}]$.*

A proof for Theorem 1 is provided in Section 5. Note that the equivalent condition provided in Theorem 1 is identifiability of a $Q[\cdot]$. This can be checked by function **ID_Single** in Algorithm 1. Therefore, when $\mathbf{S}$ is a single c-component, in order to check whether $Q[\mathbf{S}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$,

we need to check the identifiability of $Q[\mathbf{S}]$ from $\mathcal{G}[\mathbf{A}]$ for all $\mathbf{A} \in \mathbb{A}$ that $\mathbf{S} \subseteq \mathbf{A}$. Algorithm 2 summarizes the steps for solving g-identifiability of a causal effect given $(\mathbb{A}, \mathcal{G})$.

Theorem 2. *Algorithm 2 is sound and complete.*

Proof. It directly follows from Propositions 3 and 4 and Theorem 1. □

Suppose Algorithm 2 determines that the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$. Analogous to the method in Tian and Pearl [2003], we can derive a formula for $P_{\mathbf{x}}(\mathbf{Y})$ as follows:
First, for each $\mathbf{S}_i \in \{\mathbf{S}_1, \cdots, \mathbf{S}_l\}$, we can derive a formula for $Q[\mathbf{S}_i]$ using **ID_Single** function in line 4 of **GID_Single**. This allows us to compute $Q[\mathbf{S}]$ using

$$Q[\mathbf{S}] = \prod_{i=1}^{l} Q[\mathbf{S}_i].$$

Finally, the expression for $P_{\mathbf{x}}(\mathbf{Y})$ will be $\sum_{\mathfrak{X}_{\mathbf{S}\setminus\mathbf{Y}}} Q[\mathbf{S}]$.

5 MAIN RESULT: THEOREM 1

In this section, we present the main steps of the proof of Theorem 1. The technical lemmas in this section are proved in Appendix B.

Sufficient part: This part is straightforward: if $Q[\mathbf{S}]$ is identifiable from $\mathcal{G}[\mathbf{A}]$ for some $\mathbf{A} \in \mathbb{A}$ such that $\mathbf{S} \subseteq \mathbf{A}$, then $Q[\mathbf{S}]$ is uniquely computable from $Q[\mathbf{A}]$, and therefore, $Q[\mathbf{S}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.

Necessary part: Suppose $\mathbf{S}$ is a single c-component and $Q[\mathbf{S}]$ is not identifiable from $\mathcal{G}[\mathbf{A}_i]$ for all $\mathbf{A}_i \in \mathbb{A}$ such that $\mathbf{S} \subseteq \mathbf{A}_i$. We need to show that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. Recall that $\mathbb{A} = \{\mathbf{A}_i\}_{i=0}^m$. To this end, we will introduce two SEMs $\mathcal{M}_1$ and $\mathcal{M}_2$ in $\mathbb{M}^+(\mathcal{G})$ such that for each $i \in [0:m]$ and any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}}$,

$$Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}), \tag{4}$$

but there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}}$ such that

$$Q^{\mathcal{M}_1}[\mathbf{S}](\mathbf{v}_0) \neq Q^{\mathcal{M}_2}[\mathbf{S}](\mathbf{v}_0). \tag{5}$$

This shows that $Q[\mathbf{S}]$ cannot be uniquely computed from $\{Q[\mathbf{A}_i]\}_{i=0}^m$.

For sake of space, we assume that there exists at least one $i \in [0, m]$ such that $\mathbf{S} \subset \mathbf{A}_i$. In this case, without loss of generality, we assume that there exists $k \in [0:m]$ such that $\mathbf{S} \subset \mathbf{A}_i$ for $i \in [0:k]$ and $\mathbf{S} \nsubseteq \mathbf{A}_i$ for $i \in [k+1:m]$. A proof for the case in which $\mathbf{S}$ is not a subset of $\mathbf{A}_i$ for all $i \in [0, m]$ is provided in Appendix C.

We first modify $\mathcal{G}$ by deleting some nodes and edges and show that it is enough to prove Theorem 1 for the modified graph. Then, we provide our method for constructing $\mathcal{M}_2$ from $\mathcal{M}_1$ by introducing a system of linear equations.

Graph modification: Since $\mathbf{S}$ is single c-component, the bidirected edges in $\mathcal{G}_{\mathbf{S}}$ form a connected graph over $\mathbf{S}$. Let $\mathcal{F}^{\mathbf{S}}$ be a minimal (in terms of edges) spanning subgraph of $\mathcal{G}[\mathbf{S}]$ such that $\mathcal{F}^{\mathbf{S}}_{\mathbf{S}}$ is single c-component. Thus, $\mathcal{F}^{\mathbf{S}}_{\mathbf{S}}$ has no directed edges and its bidirected edges form a spanning tree.

Lemma 1 (Shpitser and Pearl [2006a])**.** *Suppose $\mathbf{S} \subseteq \mathbf{A} \subseteq \mathbf{V}$. $Q[\mathbf{S}]$ is not identifiable from $\mathcal{G}[\mathbf{A}]$ if and only if there exists at least one $\mathbf{S}$-rooted c-forest $\mathcal{F}$ with the set of observed variables $\mathbf{B}$ such that $\mathbf{S} \subsetneq \mathbf{B} \subseteq \mathbf{A}$, the bidirected edges of $\mathcal{F}_{\mathbf{B}}$ form a spanning tree, and the induced subgraph of $\mathcal{F}$ over $\mathbf{S}$ is $\mathcal{F}^{\mathbf{S}}$, i.e., $\mathcal{F}^{\mathbf{S}} = \mathcal{F}[\mathbf{S}]$.*

Recall that for each $i \in [0:k]$, $\mathbf{S} \subset \mathbf{A}_i$ and $Q[\mathbf{S}]$ is not identifiable from $Q[\mathbf{A}_i]$. Hence, Lemma 1 implies that for each $i \in [0:k]$, there exists a $\mathbf{S}$-rooted c-forest $\mathcal{F}_i$ over a set of observed variables $\mathbf{B}_i$ such that $\mathbf{S} \subsetneq \mathbf{B}_i \subseteq \mathbf{A}_i$, the bidirected edges of $(\mathcal{F}_i)_{\mathbf{S}}$ form a spanning tree, and $\mathcal{F}^{\mathbf{S}} = \mathcal{F}_i[\mathbf{S}]$. Next, we use $\{\mathcal{F}_i\}_{i=0}^k$ to modify $\mathcal{G}$.

We define $\mathcal{G}'$ to be the union of all the subgraphs in $\{\mathcal{F}_i\}_{i=0}^k$ with the observed variables $\mathbf{V}' := \bigcup_{i=0}^k \mathbf{B}_i$ and unobserved variables $\mathbf{U}'$. Furthermore, let $\mathbb{A}' := \{\mathbf{A}'_i := \mathbf{A}_i \cap \mathbf{V}'\}_{i=0}^m$. Because for each $i \in [0:k]$, $\mathcal{F}_i$ is a $\mathbf{S}$-rooted c-forest in $\mathcal{G}'$, Lemma 1 implies that $Q[\mathbf{S}]$ is not identifiable from $\mathcal{G}'[\mathbf{A}'_i]$.

Next result establishes the connection between non g-identifiability of $Q[\mathbf{S}]$ from $(\mathbb{A}, \mathcal{G})$ and non g-identifiability of $Q[\mathbf{S}]$ from $(\mathbb{A}', \mathcal{G}')$.

Lemma 2. *If $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$, then $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$.*

To complete the proof using Lemma 2, it is enough to show that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$.

From g-identifiability to a system of linear equations: To show that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$, we introduce two models in $\mathbb{M}^+(\mathcal{G}')$ such that equations (4) and (5) are satisfied. That is, $Q[\mathbf{S}]$ cannot be uniquely computed from $\{Q[\mathbf{A}'_i]\}_{i=0}^m$.

Note that to define a SEM $\mathcal{M}$ over a causal graph $\mathcal{G}'$, it suffices to define the domains $\mathfrak{X}_X$ and either the conditional distributions $P^{\mathcal{M}}(X|Pa_{\mathcal{G}'}(X))$ or the corresponding equation in the SEM for all $X \in \mathbf{V}' \cup \mathbf{U}'$, where $\mathbf{V}'$ and $\mathbf{U}'$ denote the observed and unobserved variables in causal graph $\mathcal{G}'$. We define the domains of all variables to be finite, i.e., $|\mathfrak{X}_X| < \infty$ for all $X \in \mathbf{V}' \cup \mathbf{U}'$. Let $U_0 \in \mathbf{U}'$ be a fixed unobserved variable (we will discuss later how to select U_0) with domain $\mathfrak{X}_{U_0} := \{\gamma_1, \cdots, \gamma_d\}$. We define both models $\mathcal{M}_1$ and $\mathcal{M}_2$ to have similar distributions over all variables

except variable U_0 (We will specify these distributions in Section 5.) More specifically, for all $V \in \mathbf{V}'$,

$$P^{\mathcal{M}_1}(V \mid Pa_{\mathcal{G}'}(V)) = P^{\mathcal{M}_2}(V \mid Pa_{\mathcal{G}'}(V)), \quad (6)$$

and for all $U \in \mathbf{U}' \setminus \{U_0\}$,

$$P^{\mathcal{M}_1}(U) = P^{\mathcal{M}_2}(U) = \frac{1}{|\mathfrak{X}_U|}. \quad (7)$$

As the distributions in Equations (6) and (7) are the same for both models, for the sake of brevity, we drop the superscripts $\mathcal{M}_1$ and $\mathcal{M}_2$ from here on. For $j \in [1:d]$, We define $P^{\mathcal{M}_1}(U_0 = \gamma_j) = 1/d$ and $P^{\mathcal{M}_2}(U_0 = \gamma_j) = p_j$, where we will specify $\{p_j\}_{j=1}^d$ later such that $\mathcal{M}_2 \in \mathbb{M}^+(\mathcal{G}')$ and both Equations (4) and (5) hold.

For $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$, $i \in [0:m]$, and $j \in [1:d]$, we define

$$\theta_{i,j}(\mathbf{v}) := \sum_{\mathbf{U}' \setminus \{U_0\}} \prod_{X \in \mathbf{A}'_i} P(x \mid Pa_{\mathcal{G}'}(X)) \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u),$$

$$\eta_j(\mathbf{v}) := \sum_{\mathbf{U}' \setminus \{U_0\}} \prod_{X \in \mathbf{S}} P(x \mid Pa_{\mathcal{G}'}(X)) \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u),$$

where the index j indicates that $U_0 = \gamma_j$ in the factorizations. Using these definitions, we can write $\{Q[\mathbf{A}'_i]\}_{i=0}^m$ and $Q[\mathbf{S}]$ for both models $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows:

$$\begin{aligned} Q^{\mathcal{M}_1}[\mathbf{A}'_i](\mathbf{v}) &= \sum_{j=1}^d \frac{1}{d}\theta_{i,j}(\mathbf{v}), \\ Q^{\mathcal{M}_2}[\mathbf{A}'_i](\mathbf{v}) &= \sum_{j=1}^d p_j\theta_{i,j}(\mathbf{v}), \end{aligned} \quad (8)$$

and

$$\begin{aligned} Q^{\mathcal{M}_1}[\mathbf{S}](\mathbf{v}) &= \sum_{j=1}^d \frac{1}{d}\eta_j(\mathbf{v}), \\ Q^{\mathcal{M}_2}[\mathbf{S}](\mathbf{v}) &= \sum_{j=1}^d p_j\eta_j(\mathbf{v}). \end{aligned} \quad (9)$$

As we mentioned, we need to define $\{p_j\}_{j=1}^d$ such that $\mathcal{M}_2 \in \mathbb{M}^+(\mathcal{G}')$ and both Equations (4) and (5) hold. Substituting Equations (8) and (9) into (4) and (5) yield the following set of equations.

$$\begin{aligned} &\sum_{j=1}^d (p_j - \frac{1}{d})\theta_{i,j}(\mathbf{v}) = 0, \quad \forall \mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}, i \in [0,m], \\ &\sum_{j=1}^d (p_j - \frac{1}{d})\eta_j(\mathbf{v}_0) \neq 0, \quad \exists \mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}, \\ &\sum_{j=1}^d p_j = 1, \\ &0 < p_j < 1, \quad \forall j \in [1:d]. \end{aligned} \quad (10)$$

Note that the last inequalities ensure that $\mathcal{M}_2 \in \mathbb{M}^+(\mathcal{G}')$. The system of linear equations in (10) is solvable with respect to $\{p_j\}_{j=1}^d$ if and only if the following system of linear equations is solvable with respect to $\{\beta_j\}_{j=1}^d$.

$$\begin{aligned} &\sum_{j=1}^d \beta_j\theta_{i,j}(\mathbf{v}) = 0, \quad \forall \mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}, i \in [0:m] \\ &\sum_{j=1}^d \beta_j\eta_j(\mathbf{v}_0) \neq 0, \quad \exists \mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'} \\ &\sum_{j=1}^d \beta_j = 0. \end{aligned} \quad (11)$$

Remark 2. *If $\{\beta_j^*\}_{j=1}^d$ is a solution for* (11)*, then*

$$p_j^* := \frac{1}{d} + \frac{\beta_j^*}{2hd}, \quad (12)$$

is a solution for (10)*, where $h = \max_{j \in [1:d]} |\beta_j^*|$. Note that the division by $2h$ in Equation* (12) *ensures that $0 < p_j^* < 1$ for each $j \in [1:d]$.*

A solution to the system of linear equations in (11) will specify the distribution of U_0 in model $\mathcal{M}_2$. Clearly, existence of a solution to (11) depends on the choices of $\{\theta_{i,j}(\mathbf{v})\}$ and $\{\eta_j(\mathbf{v})\}$. Next result presents a necessary condition under which (11) admits a solution.

For $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$ and $i \in [0:m]$, let $\theta_i(\mathbf{v})$ and $\eta(\mathbf{v})$ denote the vectors $(\theta_{i,1}(\mathbf{v}), ..., \theta_{i,d}(\mathbf{v}))$ and $(\eta_1(\mathbf{v}), ..., \eta_d(\mathbf{v}))$ in $\mathbb{R}^d$, respectively.

Lemma 3. *Consider the following set of vectors in $\mathbb{R}^d$*

$$\mathbf{\Omega} := \{\theta_i(\mathbf{v}) : \ i \in [0:m], \mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}\} \cup \{\mathbb{1}_d\}, \quad (13)$$

where $\mathbb{1}_d$ denotes the all-ones vector in $\mathbb{R}^d$. If there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ such that $\eta(\mathbf{v}_0)$ is linearly independent from all the vectors in $\mathbf{\Omega}$, then the system of linear equations in (11) *admits a solution.*

To summarize, so far we have introduced two models for proving the necessary part of Theorem 1. In order to complete the proof, it remains to specify the conditional distributions in (6) for all observed variables which consequently specify the vectors in $\mathbf{\Omega}$ in Equation (13) and to find a realization $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ such that $\eta(\mathbf{v}_0)$ is linearly independent from the set of the vectors in $\mathbf{\Omega}$.

Constructing the conditional distributions: In order to specify the conditional distributions in (6), we first introduce the following definitions and notations.

Let U_0 be an unobserved variable in subgraph $\mathcal{F}_0$ that has one child in $\mathbf{S}$ and one child in $\mathbf{T} := \mathbf{V}' \setminus \mathbf{S}$. Note that such variable exists because $\mathbf{B}_0$ is a single c-component and

$\mathbf{S} \subsetneq \mathbf{B}_0 \subseteq \mathbf{V}'$. We denote the set of unobserved variables in $\mathcal{G}[\mathbf{S}]$ by $\mathbf{U}^{\mathbf{S}}$ and define $\mathbf{U}^{\mathbf{T}} := \mathbf{U}' \setminus (\mathbf{U}^{\mathbf{S}} \cup \{U_0\})$. For $X \in \mathbf{V}' \cup \mathbf{U}'$, we define $\alpha(X)$ to denote the number of graphs in $\{\mathcal{F}_j\}_{j=0}^k$ that contains X.

For each $i \in [0:k]$, let T_i denotes a node in $\mathbf{B}_i \setminus \mathbf{S}$ such that $Ch_{\mathcal{F}_i}(T_i) \cap \mathbf{S} \neq \varnothing$. Note that such variables exist because $\mathcal{F}_i$s are $\mathbf{S}$-rooted c-forest.

Now, we are ready to introduce the domains of all variables in $\mathbf{V}' \cup \mathbf{U}'$. Recall that $\mathbf{V}' = \mathbf{S} \cup \mathbf{T}$ and $\mathbf{U}' = \mathbf{U}^{\mathbf{S}} \cup \mathbf{U}^{\mathbf{T}} \cup \{U_0\}$.

$$\begin{aligned}
&\mathfrak{X}_X := [0:\kappa], \quad \forall X \in \mathbf{S},\\
&\mathfrak{X}_X := \{0,1\}^{\alpha(T)}, \quad \forall X \in \mathbf{T},\\
&\mathfrak{X}_X := [0:\kappa], \quad \forall X \in \mathbf{U}^{\mathbf{S}},\\
&\mathfrak{X}_X := \{0,1\}^{\alpha(U)}, \quad \forall X \in \mathbf{U}^{\mathbf{T}},\\
&\mathfrak{X}_{U_0} := [0:\kappa] \times \{0,1\}^{\alpha(U_0)-1}.
\end{aligned}$$

In the above definition, κ is an arbitrary odd integer greater than 4. Note that the number of elements in $\mathfrak{X}_{U_0}$ is $d = (\kappa+1)2^{\alpha(U_0)-1}$.

According to the above definitions, for each $X \in \mathbf{T} \cup \mathbf{U}^{\mathbf{T}} \cup \{U_0\}$, its domain $\mathfrak{X}_X$ is a subset of $\mathbb{R}^{\alpha(X)}$ and it belongs to exactly $\alpha(X)$ number of subgraphs in $\{\mathcal{F}_i\}_{i=0}^k$. Suppose X belongs to $\mathcal{F}_{i_1}, \cdots, \mathcal{F}_{i_{\alpha(X)}}$, where $i_1 < \cdots < i_{\alpha(X)}$. Thus, we denote X by a vector $(X[i_1], \cdots, X[i_{\alpha(X)}])$ of length $\alpha(X)$. Next, we construct the conditional distributions of the observed variables by specifying their functional dependencies to their parents.

When $X \in \mathbf{T}$, we define the entries of its corresponding vector as

$$X[i_j] \equiv \left(\sum_{Y \in Pa_{\mathcal{F}_{i_j}}(X)} Y[i_j] \right) \pmod 2,$$

where $j \in [1:\alpha(X)]$.

We now construct the variables in $\mathbf{S}$. Recall that U_0 has one child in $\mathbf{S}$ which we denote it by S_0. For each $S \in \mathbf{S} \setminus \{S_0\}$ and any realization of $Pa_{\mathcal{G}'}(S)$, we define $\mathbb{I}(S)$ to be one if there exists $i \in [0:k]$ such that

1. $T_i \in Pa_{\mathcal{G}'}(S)$ and $T_i[i] = 0$, or
2. there exists $X \in Pa_{\mathcal{G}'}(S) \setminus (\mathbf{U}^{\mathbf{S}} \cup \{T_i\})$ such that $\mathcal{F}_i$ contains X and $X[i] = 1$,

and zero, otherwise. Note that according to the definition of T_i, it belongs to $\mathcal{F}_i$ and therefore, $T_i[i]$ exists. Analogously, we define $\mathbb{I}(S_0)$ to be one if there exists $i \in [0:k]$ such that

1. $T_i \in Pa_{\mathcal{G}'}(S)$ and $T_i[i] = 0$, or
2. $i \neq 0$, $\mathcal{F}_i$ contains U_0, and $U_0[i] = 1$, or
3. there exists $X \in Pa_{\mathcal{G}'}(S) \setminus (\mathbf{U}^{\mathbf{S}} \cup \{T_i, U_0\})$ such that $\mathcal{F}_i$ contains X and $X[i] = 1$.

Now, for each $S \in \mathbf{S}$ and $s \in [0:\kappa]$, we define $P(S = s \mid Pa_{\mathcal{G}'}(S))$ as

$$\begin{cases}
\frac{1}{\kappa+1} & \text{if } \mathbb{I}(S) = 1\\
1 - \kappa\epsilon & \text{if } \mathbb{I}(S) = 0 \text{ and } s \equiv M(S) \pmod{\kappa+1},\\
\epsilon & \text{if } \mathbb{I}(S) = 0 \text{ and } s \not\equiv M(S) \pmod{\kappa+1},
\end{cases}$$

where $0 < \epsilon < \frac{1}{\kappa}$ and

$$M(S) := \begin{cases}
\sum_{x \in Pa_{\mathcal{G}'[\mathbf{S}]}(S)} x & \text{, if } S \in \mathbf{S} \setminus \{S_0\},\\
u_0[0] + \sum_{x \in Pa_{\mathcal{G}'[\mathbf{S}]}(S)} x & \text{, if } S = S_0 \,.
\end{cases}$$

Note that $M(S)$ is an integer number. This is because $Pa_{\mathcal{G}'[\mathbf{S}]}(S) \subseteq \mathbf{U}^{\mathbf{S}}$ and thus all terms in the above definition belong to $[0:\kappa]$.

Lemma 4. *The SEM constructed above belongs to* $\mathbb{M}^+(\mathcal{G}')$.

Existence of realization $\mathbf{v}_0$: Herein, we show that for the aforementioned conditional distributions, there exists a realization $\mathbf{v}_0$ such that $\eta(\mathbf{v}_0)$ is linearly independent from the set of the vectors in $\mathbf{\Omega}$ (in Equation (13)). Consider the following subset of $\mathfrak{X}_{U_0} = \{\gamma_1, ..., \gamma_d\}$ with $\frac{\kappa+1}{2}$ elements:

$$\mathbf{\Gamma} := \Big\{ (2x, 0, \cdots, 0) \colon\ x \in [0 : \frac{\kappa-1}{2}] \Big\}.$$

Recall that for $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$ and $i \in [0:m]$, $\theta_i(\mathbf{v})$ and $\eta(\mathbf{v})$ are two vectors in $\mathbb{R}^d$ with j-th entry corresponds to $U_0 = \gamma_j$. Suppose that $\mathbf{\Gamma} = \{\gamma_{j_1}, ..., \gamma_{j_{\frac{\kappa+1}{2}}}\}$. Next result shows that in our constructed models, all entries of $\theta_i(\mathbf{v})$ with indices in $\{j_1, ..., j_{\frac{\kappa+1}{2}}\}$ are equal.

Lemma 5. *For any* $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$ *and* $i \in [0:m]$,

$$\theta_{i,j_1}(\mathbf{v}) = \theta_{i,j_2}(\mathbf{v}) = \cdots = \theta_{i,j_{\frac{\kappa+1}{2}}}(\mathbf{v}).$$

An immediate consequence of this result is that any linear combination of the vectors in $\mathbf{\Omega}$ will have equal entries at the indices in $\{j_1, ..., j_{\frac{\kappa+1}{2}}\}$. Next, we show there exists a realization $\mathbf{v}_0$ for which $\eta(\mathbf{v}_0)$ does not follow this pattern and thus it is linearly independent of all vectors in $\mathbf{\Omega}$.

Lemma 6. *There exists* $0 < \epsilon < \frac{1}{\kappa}$ *such that there exists* $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ *and* $1 \leq r < t \leq \frac{\kappa+1}{2}$ *such that*

$$\eta_{j_r}(\mathbf{v}_0) \neq \eta_{j_t}(\mathbf{v}_0).$$

6 CONCLUSION

We revisited the problem of general identifiability and showed that positivity assumption of observational distributions is crucial for the soundness of do-calculus rules. This assumption was ignored in previous work. We presented a novel algorithm for g-identifiability which is provably sound and complete considering the positivity assumption.

References

Elias Bareinboim and Judea Pearl. Causal inference by surrogate experiments: Z-identifiability. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, page 113–120, Arlington, Virginia, USA, 2012. AUAI Press.

Elias Bareinboim and Judea Pearl. Transportability from multiple environments with limited experiments: Completeness results. *Advances in neural information processing systems*, 27, 2014.

Elias Bareinboim and Jin Tian. Recovering causal effects from selection bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

Yimin Huang and Marco Valtorta. Identifiability in causal bayesian networks: A sound and complete algorithm. In *AAAI*, pages 1149–1154, 2006.

Yimin Huang and Marco Valtorta. On the completeness of an identifiability algorithm for semi-markovian models. *Annals of Mathematics and Artificial Intelligence*, 54(4): 363–408, 2008.

Sanghack Lee, Juan D Correa, and Elias Bareinboim. General identifiability with arbitrary surrogate experiments. In *Uncertainty in Artificial Intelligence*, pages 389–398. PMLR, 2019.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

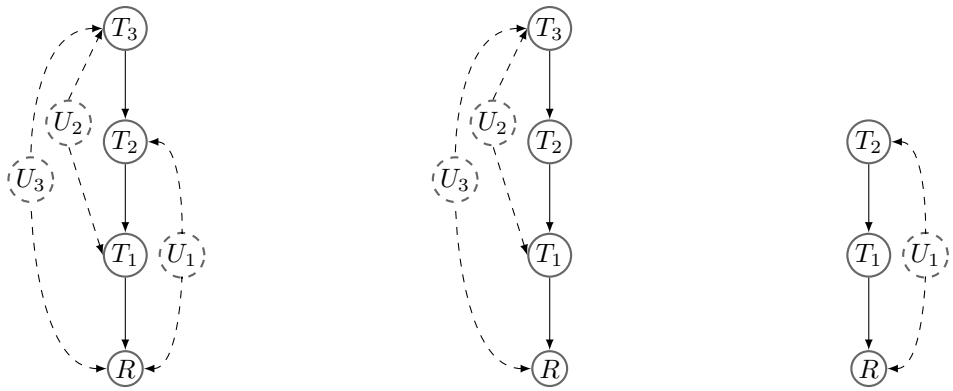
Judea Pearl. *Causality*. Cambridge university press, 2009.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 1219, 2006a.

Ilya Shpitser and Judea Pearl. Identification of conditional interventional distributions. *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, 2006b.

Jin Tian and Judea Pearl. On the identification of causal effects. Technical report, Department of Computer Science, University of California, 2003.

Santtu Tikka, Antti Hyttinen, and Juha Karvanen. Causal effect identification from multiple incomplete data sources: A general search-based approach. *Journal of Statistical Software*, 2021.



(a) Thicket $\mathcal{J}$ (b) Hedgelet $\mathcal{H}_1$ (c) Hedgelet $\mathcal{H}_2$

Figure 3: (a) Thicket is formed for the causal effect of $\{T_1, T_2, T_3\}$ on $\{R\}$ in Example 2; (b) and (c) are the hedgelets formed by the thicket $\mathcal{J}$

APPENDIX

A ON THE POSITIVITY ASSUMPTION

We first present some definitions and notations from [Lee et al., 2019] including their illustrations using the causal graph $\mathcal{G}$ from Example 2 (Figure 2).

A.1 NOTATION

Definition 6 ([Lee et al., 2019])**.** *Assume that* $\mathbf{R}$ *is a subset of observed variables* $\mathbf{V}$*. A hedge is pair of* $\mathbf{R}$*-rooted c-forests* $\langle \mathcal{F}, \mathcal{F}' \rangle$ *such that* $\mathcal{F}'$ *is a subgraph of* $\mathcal{F}$*.*

In Figure 2: Subgraphs $\mathcal{F} = \mathcal{G}[\{R, T_1, T_2, T_3\}]$ and $\mathcal{F}' = \mathcal{G}[\{R\}]$ form a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$.

Denoted by $\mathcal{C}(\mathcal{G}) = \{\mathbf{W}_i\}_{i=1}^k$, the set of c-components that partitions observed variables in $\mathcal{G}$ such that each $\mathbf{W}_i$ is a maximal c-component. Maximal in the sense of number of nodes that is there is no $\mathbf{W} \in \mathbf{V}$ such that $\mathbf{W}_i \subsetneq \mathbf{W}$ and $\mathbf{W}$ is a c-component in $\mathcal{G}$. Assume that $\mathbf{T}$ is set of all observed variables in $\mathcal{F}$ but not in $\mathcal{F}'$. We define $\mathcal{F}'' := \mathcal{F}[\mathbf{T}]$.

In Figure 2: $\mathcal{C}(\mathcal{G}[\{T_1, T_2, T_3\}]) = \{\{T_1, T_3\}, \{T_2\}\}$. Additionally, $\mathcal{F}'' = \mathcal{G}[\{T_1, T_2, T_3\}]$ for the hedge constructed before.

Definition 7 (Lee et al. [2019])**.** *Given a hedge* $\langle \mathcal{F}, \mathcal{F}' \rangle$*. Denote by* $\mathbf{V}'$ *a set of all observed variables of* $\mathcal{F}'$*. The hedgelet decomposition of a hedge* $\langle \mathcal{F}, \mathcal{F}' \rangle$ *is a collection of hedgelets* $\{\mathcal{F}(\mathbf{W})\}_{\mathbf{W} \in \mathcal{C}(\mathcal{F}'')}$ *where each hedgelet* $\mathcal{F}(\mathbf{W})$ *is a subgraph of* $\mathcal{F}$ *made of (i)* $\mathcal{F}[\mathbf{W} \cup \mathbf{V}']$ *and (ii)* $\mathcal{F}[De_{\mathcal{F}}(\mathbf{W})]$ *without bidirected edges, that is all observed descendants of* $\mathbf{W}$ *and all directed edges between them. Let* $\mathbb{H}_{\mathcal{F}} := \{\mathcal{F}(\mathbf{W})\}_{\mathbf{W} \in \mathcal{C}(\mathcal{F}'')}$ *be the set of hedgelets of* $\langle \mathcal{F}, \mathcal{F}' \rangle$*.*

In Figure 2: For the hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$, where $\mathcal{F} = \mathcal{G}[\{R, T_1, T_2, T_3\}]$ and $\mathcal{F}' = \mathcal{G}[\{R\}]$, there are two hedgelets $\mathcal{H}_1, \mathcal{H}_2$ displayed in Figures (3b)-(3c). Moreover, we have $\mathbb{H}_{\mathcal{F}} = \{\mathcal{H}_1, \mathcal{H}_2\}$.

Definition 8 ([Lee et al., 2019])**.** *Let* $\mathbf{R}$ *be a non-empty set of variables and* $\mathbb{Z}$ *be a collection of sets of variables in* $\mathcal{G}$*. A thicket* $\mathcal{J}$ *is a subgraph of* $\mathcal{G}$ *which is an* $\mathbf{R}$*-rooted c-component consisting of a minimal c-component over* $\mathbf{R}$ *and hedges*

$$\mathbb{F}_{\mathcal{J}} := \{\langle \mathcal{F}_{\mathbf{Z}}, \mathcal{J}[\mathbf{R}] \rangle \mid \mathcal{F}_{\mathbf{Z}} \subseteq \mathcal{G}[\mathbf{V} \setminus \mathbf{Z}], \mathbf{Z} \cap \mathbf{R} = \varnothing\}_{\mathbf{Z} \in \mathbb{Z}}.$$

Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint sets of observed variables in $\mathcal{G}$. A thicket $\mathcal{J}$ is said to be formed for $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$ with respect to $\mathbb{Z}$ if $\mathbf{R} \subseteq Anc_{\mathcal{G}[\mathbf{V} \setminus \mathbf{X}]}(\mathbf{Y})$ and every hedgelet of each hedge $\langle \mathcal{F}_{\mathbf{Z}}, \mathcal{J}[\mathbf{R}] \rangle$ intersects with $\mathbf{X}$.

In Figure 2: This graph is a thicket, also displayed in Figure 3a. Let $\mathbb{F}_{\mathcal{J}}$ be

$$\mathbb{F}_{\mathcal{J}} = \{\langle \mathcal{F}, \mathcal{F}' \rangle\},$$

where $\mathcal{F} = \mathcal{G}[\{R, T_1, T_2, T_3\}]$ and $\mathcal{F}' = \mathcal{G}[\{R\}]$. One can observe that thicket $\mathcal{J}$ is formed for the causal effect $\mathbf{X} = \{T_1, T_2, T_3\}$ on $\mathbf{Y} = \{R\}$.

Denote by $\mathbf{T}$ all observed variables in thicket $\mathcal{J}$ outside of subgraph $\mathcal{J}[\mathbf{R}]$. Let $\mathbb{H} = \bigcup_{\{\langle \mathcal{F}, \mathcal{F}' \rangle\} \in \mathbb{F}_{\mathcal{J}}} \mathbb{H}_{\mathcal{F}}$, that is, a collection of all hedgelets induced by the hedges of $\mathcal{J}$.

In Figure 2: $\mathbf{T} = \{T_1, T_2, T_3\}$ and $\mathbb{H} = \{\mathcal{H}_1, \mathcal{H}_2\}$.

A.2 ON THE POSITIVITY ASSUMPTION $P(\mathbf{V}) > 0$

Given the above definitions, we can state Lemma 3 in [Lee et al., 2019].

Lemma. *Let $\mathbf{T}' \subsetneq \mathbf{T}$ such that there exists a hedgelet $\mathcal{H} \in \mathbb{H} \setminus \mathbb{H}(\mathbf{T}')$, where $\mathbb{H}(\mathbf{T}')$ is a set of hedgelets from $\mathbb{H}$ which contain at least one variable from $\boldsymbol{T}'$. Then, under the intervention $do(\mathbf{t}')$, there exists $R \in \mathbf{R}$, for any instantiation of $\mathbf{U}$, such that $r = 0$ in both models.*

Note that by the construction in [Lee et al., 2019], R in the above Lemma is a binary random variable. In the above Lemma, let $\mathbf{T}' = \varnothing$. Based on this Lemma, for any instantiation of unobserved variables $\mathbf{U}$, $P(\mathbf{V} = \mathbf{v}) = 0$, where $\mathbf{v}$ is a realization for observed variables in which $r = 1$. This clearly shows that the constructed models in [Lee et al., 2019] violates the positivity assumption.

A.3 ON THE POSITIVITY ASSUMPTION $P(\mathbf{X} \mid \boldsymbol{Pa}_{\mathcal{G}}(\mathbf{X}) \setminus \mathbf{X}) > 0$

Herein, we study Figure 2 in more details and show that the models in [Lee et al., 2019] violate the relaxed positivity assumption. To this end, we present the models $\mathcal{M}_1$ and $\mathcal{M}_2$ constructed in [Lee et al., 2019] for the thicket $\mathcal{J}$ which is defined for this case in Appendix A.1. By the construction, each variable from $\{U_1, U_2, U_3, T_3\}$ is a binary number, i.e., $\{0, 1\}$ and each variable from $\{T_1, T_2\}$ is a vector of length two, because each variable from $\{U_1, U_2, U_3, T_3\}$ appears in only one hedgelet and each variables in $\{T_1, T_2\}$ appears in exactly two different hedgelets. Thus, $T_1 = (T_{1,1}, T_{1,2})$ and $T_2 = (T_{2,1}, T_{2,2})$, where $T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}$ are binary numbers. The first coordinate captures some properties of the hedgelet $\mathcal{H}_1$ while the second coordinate captures some properties of the hedgelet $\mathcal{H}_2$. Lee et al. [2019] define both models $\mathcal{M}_1, \mathcal{M}_2$ for the hedgelet $\mathcal{H}_1$ as

$$T_3 = U_2 \oplus U_3, \quad T_{2,1} = T_3, \quad T_{1,1} = T_{2,1} \oplus U_2,$$

and for the hedgelet $\mathcal{H}_2$ as

$$T_{2,2} = U_1, \quad T_{1,2} = T_{2,2}, \quad T_{2,2} = U_1.$$

Additionally, in model $\mathcal{M}_1$, variable R is defined by

$$R = \mathbb{1}_{T_{1,1}=0} \wedge \mathbb{1}_{T_{1,2}=0} \wedge \mathbb{1}_{U_3=1} \wedge \mathbb{1}_{U_1=1},$$

and in model $\mathcal{M}_2$, it is defined to be zero, i.e., $R = 0$.

B TECHNICAL PROOFS

In this section, we first present some technical lemmas which we use throughout our proofs. The proofs of the lemmas and propositions within the main text are provided in Subsections B.2 and B.3.

The logical order of our proofs is depicted in Figure 4. For instance, we use Theorem 1 to prove Lemma 8. Also note that the proof of Theorem 1 is provided in the main text using Lemmas 2, 3, 4, 5, and 6.

Definition 9 (Ancestral)**.** *We say a subset $\mathbf{X}$ of observed variables $\mathbf{V}$ is ancestral in $\mathcal{G}$, if $\mathbf{X} = Anc_{\mathcal{G}_{\mathbf{V}}}(\mathbf{X})$.*

B.1 TECHNICAL LEMMAS

Lemma 7 ([Tian and Pearl, 2003])**.** *Let $\mathbf{W} \subseteq \mathbf{C} \subseteq \mathbf{V}$, $\mathbf{T} = \mathbf{C} \setminus \mathbf{W}$, $\mathbf{S} = \mathbf{V} \setminus \mathbf{T}$. If $\mathbf{W}$ is an ancestral set in $\mathcal{G}[\mathbf{C}]$, then:*

$$Q[\mathbf{W}] = \sum_{\mathbf{C} \setminus \mathbf{W}} Q[\mathbf{C}].$$

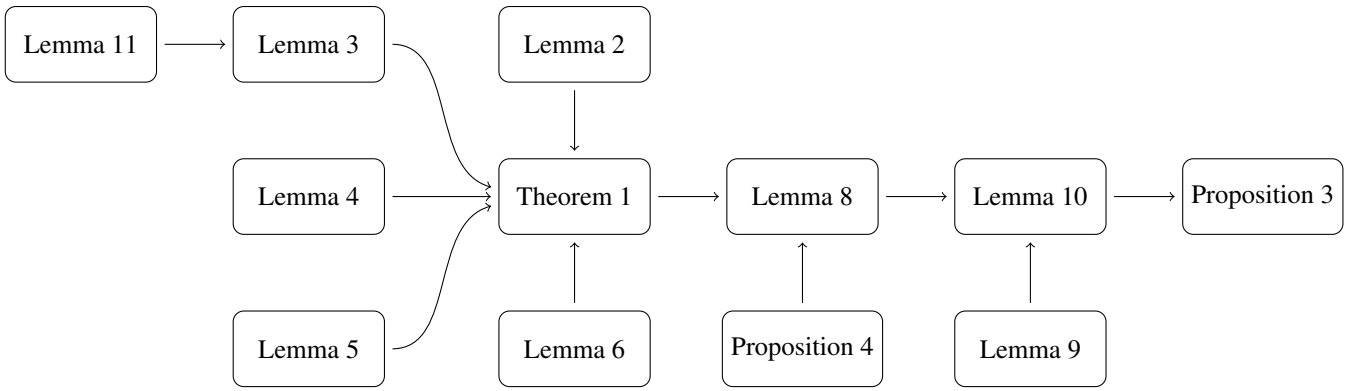


Figure 4: Logical order of our proofs.

Lemma 8. *Consider a causal graph $\mathcal{G}$ with observed variables $\mathbf{V}$. Suppose $\mathbf{X} \subseteq \mathbf{V}$ and $e := (X_1, Z)$ is a directed edge such that $X_1 \in \mathbf{X}$. $Q[\mathbf{X}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{X}]$ is g-identifiable from $(\mathbb{A}, \mathcal{H})$, where $\mathcal{H}$ is the graph obtained by deleting e from $\mathcal{G}$.*

Proof. $\mathbf{X}$ has the same c-components in $\mathcal{G}$ and $\mathcal{H}$ since $\mathcal{G}_{\mathbf{V}}$ and $\mathcal{H}_{\mathbf{V}}$ have the same undirected edges. Let $\mathbf{X}_1, \cdots, \mathbf{X}_l$ be the c-components of $\mathbf{X}$. For any $i \in [1 : l]$ and $\mathbf{A} \in \mathbb{A}$ such that $\mathbf{X}_i \subseteq \mathbf{A}$, Huang and Valtorta [2008] showed that $Q[\mathbf{X}_i]$ is identifiable from $\mathcal{G}[\mathbf{A}]$ if and only if $Q[\mathbf{X}_i]$ is identifiable from $\mathcal{H}[\mathbf{A}]$. Hence, Theorem 1 implies that $Q[\mathbf{X}_i]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{X}_i]$ is g-identifiable from $(\mathbb{A}, \mathcal{H})$. In this case, Proposition 4 implies that $Q[\mathbf{X}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{X}]$ is g-identifiable from $(\mathbb{A}, \mathcal{H})$.

□

Lemma 9. *Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are disjoint subsets of $\mathbf{V}$. Let (Y_1, Y_2) (i.e., $Y_1 \to Y_2$) denotes a directed edge in $\mathcal{G}$, where $Y_1, Y_2 \in \mathbf{Y}$. Let $\mathcal{G}'$ denotes the resulting graph after removing edge (Y_1, Y_2) from $\mathcal{G}$. If the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}')$, then the causal effect of $\mathbf{X}$ on $\mathbf{Y} \setminus \{Y_1\}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$.*

Proof. Herein, we provide a proof that is similar to one of the proofs in [Huang and Valtorta, 2008].

Using Markov factorization property in graph $\mathcal{G}'$, $P_{\mathbf{x}}(\mathbf{y})$ is given by

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y})} \sum_{\mathbf{U}} \prod_{W \in \mathbf{V} \setminus \mathbf{X}} P(w \mid Pa_{\mathcal{G}'}(W)) \prod_{U \in \mathbf{U}} P(u).$$

Similarly, in graph $\mathcal{G}$ we have

$$P_{\mathbf{x}}(\mathbf{y} \setminus \{Y_1\}) = \sum_{\{Y_1\} \cup (\mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y}))} \sum_{\mathbf{U}} \prod_{W \in \mathbf{V} \setminus \mathbf{X}} P(w \mid Pa_{\mathcal{G}}(W)) \prod_{U \in \mathbf{U}} P(u).$$

Since the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}')$, there exists $\mathcal{M}_1$ and $\mathcal{M}_2$ in $\mathbb{M}^+(\mathcal{G}')$ such that:

$$Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}), \ \forall \mathbf{v} \in \mathfrak{X}_{\mathbf{V}}, \ \forall i \in [0 : m],$$

$$P_{\mathbf{x}}^{\mathcal{M}_1}(\mathbf{y}) \neq P_{\mathbf{x}}^{\mathcal{M}_2}(\mathbf{y}), \ \exists \mathbf{x} \in \mathfrak{X}_{\mathbf{X}}, \ \exists \mathbf{y} \in \mathfrak{X}_{\mathbf{Y}}.$$

Using $\mathcal{M}_1$ and $\mathcal{M}_2$, we construct two SEMs $\mathcal{M}_1'$ and $\mathcal{M}_2'$ in $\mathbb{M}^+(\mathcal{G})$. Define a surjective function $F\colon \ \mathfrak{X}_{Y_1} \to \{0, 1\}$ and a function $\Psi\colon \ \{0, 1\} \times \mathfrak{X}_{Y_1} \to (0, 1)$ such that $\Psi(0, y_1) + \Psi(1, y_1) = 1$ for each $y_1 \in \mathfrak{X}_{Y_1}$. We will later assume some constraints for these functions, but for now lets assume they are arbitrary.

For any node S which is either unobserved or in $\mathbf{V} \setminus (\{Y_2\} \cup Ch_{\mathcal{G}}(Y_2))$, we define

$$P^{\mathcal{M}_i'}(s | Pa_{\mathcal{G}}(S)) = P^{\mathcal{M}_i}(s | Pa_{\mathcal{G}'}(S)),$$

where $i \in \{1, 2\}$. The domain of Y_2 in $\mathcal{M}_i'$ is defined as $\mathfrak{X}_{Y_2}^{\mathcal{M}} \times \{0, 1\}$, where $\mathfrak{X}_{Y_2}^{\mathcal{M}}$ is the domain of Y_2 in $\mathcal{M}_i$. For $y_2 \in \mathfrak{X}_{Y_2}^{\mathcal{M}}$, $i \in \{0, 1\}$, and $k \in \{0, 1\}$ we define:

$$P^{\mathcal{M}_i'}((y_2, k) \mid Pa_{\mathcal{G}'}(Y_2), y_1) = P^{\mathcal{M}_i}(y_2 \mid Pa_{\mathcal{G}'}(Y_2)) \Psi(F(y_1) \oplus k, y_1).$$

Note that $Pa_{\mathcal{G}'}(Y_2) \cup \{Y_1\} = Pa_{\mathcal{G}}(Y_2)$. Moreover, for a fixed realization $(Pa_{\mathcal{G}'}(Y_2), y_1)$, we have:

$$\sum_{k \in \{0,1\}} \sum_{y_2 \in \mathfrak{X}^{\mathcal{M}}_{Y_2}} P^{\mathcal{M}'_i}((y_2, k) | pa(Y_2), y_1) = 1.$$

For each $S \in Ch_{\mathcal{G}}(Y_2)$, we define:

$$P^{\mathcal{M}'_i}(s \mid Pa_{\mathcal{G}}(S) \setminus \{Y_2\}, (y_2, k)) = P^{\mathcal{M}_i}(s \mid Pa_{\mathcal{G}}(S) \setminus \{Y_2\}, y_2).$$

Next, we show that $Q^{\mathcal{M}'_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}'_2}[\mathbf{A}_i](\mathbf{v})$ for each $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}}$ and $i \in [0:m]$. Suppose $\mathbf{v}$ is a realization of $\mathbf{V}$ in $\mathcal{M}'_1$ with realizations y_1 and (y_2, k) for Y_1 and Y_2, respectively. Consider two cases:

- If $Y_2 \notin \mathbf{A}_i$:

$$\begin{aligned}
Q^{\mathcal{M}'_1}[\mathbf{A}_i](\mathbf{v}) &= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}'_1}(a \mid Pa_{\mathcal{G}}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}'_1}(u) \\
&= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_1}(a \mid Pa_{\mathcal{G}'}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_1}(u) = Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}) \\
&= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_2}(a \mid Pa_{\mathcal{G}'}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_2}(u) \\
&= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}'_2}(a \mid Pa_{\mathcal{G}}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}'_2}(u) \\
&= Q^{\mathcal{M}'_2}[\mathbf{A}_i](\mathbf{v}).
\end{aligned}$$

- If $Y_2 \in \mathbf{A}_i$:

$$\begin{aligned}
Q^{\mathcal{M}'_1}[\mathbf{A}_i](\mathbf{v}) &= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}'_1}(a \mid Pa_{\mathcal{G}}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}'_1}(u) \\
&= \Psi\left(F(y_1) \oplus k, y_1\right) \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_1}(a \mid Pa_{\mathcal{G}'}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_1}(u) \\
&= \Psi(F(y_1) \oplus k, y_1) Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = \Psi(F(y_1) \oplus k, y_1) Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}) \\
&= \Psi(F(y_1) \oplus k, y_1) \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{M_2}(a \mid Pa_{\mathcal{G}'}(A)) \prod_{U \in \mathbf{U}} P^{M_2}(u) \\
&= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}'_2}(a \mid Pa_{\mathcal{G}}(A))) \prod_{U \in \mathbf{U}} P^{\mathcal{M}'_2}(u) \\
&= Q^{\mathcal{M}'_2}[\mathbf{A}_i](\mathbf{v}).
\end{aligned}$$

Therefore, $Q^{\mathcal{M}'_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}'_2}[\mathbf{A}_i](\mathbf{v})$ for each $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}}$ and $i \in [0:m]$.

We know that there exists $\hat{\mathbf{x}} \in \mathfrak{X}^{\mathcal{M}}_{\mathbf{X}}$ and $\hat{\mathbf{y}} \in \mathfrak{X}^{\mathcal{M}}_{\mathbf{Y}}$ such that $P^{\mathcal{M}_1}_{\hat{\mathbf{x}}}(\hat{\mathbf{y}}) \neq P^{\mathcal{M}_2}_{\hat{\mathbf{x}}}(\hat{\mathbf{y}})$. Denote by $\hat{y}_1$ and $\hat{y}_2$ the realizations of Y_1 and Y_2 in the realization $\hat{\mathbf{y}}$, respectively. Assume that $P^{\mathcal{M}_1}_{\hat{\mathbf{x}}}(\hat{\mathbf{y}}) = d_1 > P^{\mathcal{M}_2}_{\hat{\mathbf{x}}}(\hat{\mathbf{y}}) = d_2$. Assume that $\Psi(F(\hat{y}_1) \oplus 0, \hat{y}_1) = 0.5$ and $\Psi(F(y) \oplus 0, y) = \frac{d_1 - d_2}{4}$ for all $y \in \mathfrak{X}_{Y_1} \setminus \{\hat{y}_1\}$. Then we have:

$$\begin{aligned}
P^{\mathcal{M}'_1}_{\hat{\mathbf{x}}}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\}) &= \sum_{y_1 \in \mathfrak{X}_{Y_1}} \sum_{\mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V} \setminus \mathbf{X}} P^{\mathcal{M}'_1}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&> \sum_{y_1 = \hat{y}_1} \sum_{\mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V} \setminus \mathbf{X}} P^{\mathcal{M}'_1}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&= P^{\mathcal{M}_1}_{\hat{\mathbf{x}}}(\hat{\mathbf{y}}) \Psi(F(\hat{y}_1) \oplus 0, \hat{y}_1) = 0.5 d_1.
\end{aligned}$$

but,

$$
\begin{aligned}
P_{\hat{\mathbf{x}}}^{\mathcal{M}_2'}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\}) &= \sum_{y_1 \in \mathfrak{X}_{Y_1}} \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V}\setminus\mathbf{X}} P^{\mathcal{M}_1'}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&= \sum_{y_1 = \hat{y}_1} \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V}\setminus\mathbf{X}} P^{\mathcal{M}_1'}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&+ \sum_{y_1 \in \mathfrak{X}_{Y_1}\setminus\{\hat{y}_1\}} \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V}\setminus\mathbf{X}} P^{M_1'}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&\leq P_{\hat{\mathbf{x}}}^{\mathcal{M}_2}(\hat{\mathbf{y}})\Psi(F(\hat{y}_1) \oplus 0, \hat{y}_1) + P_{\hat{\mathbf{x}}}^{\mathcal{M}_2}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\})\Psi(F(Y_1 \neq y_1) \oplus 0, Y_1 \neq y_1) \\
&= 0.5d_2 + \frac{d_1 - d_2}{4} < 0.5d_1.
\end{aligned}
$$

This implies that $P_{\hat{\mathbf{x}}}^{\mathcal{M}_1'}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\}) \neq P_{\hat{\mathbf{x}}}^{\mathcal{M}_2'}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\})$ which concludes the proof. □

Lemma 10. *Assume $\mathbf{Y} \subset \mathbf{W} \subset \mathbf{V}$ such that for each $W \in \mathbf{W} \setminus \mathbf{Y}$, there exists a directed path in $\mathcal{G}[\mathbf{W}]$ from W to a variable in $\mathbf{Y}$. Then, the causal effect of $\mathbf{V} \setminus \mathbf{W}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{W}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.*

Proof. Let $\mathbf{X} := \mathbf{V} \setminus \mathbf{W}$.

Sufficient part: Suppose $Q[\mathbf{W}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$. Since $Q[\mathbf{W}] = P_{\mathbf{x}}(\mathbf{W})$, we have

$$
P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{W}\setminus\mathbf{Y}} Q[\mathbf{W}].
$$

Hence, $P_{\mathbf{x}}(\mathbf{y})$ is uniquely computed and the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.

Necessary part: Suppose $Q[\mathbf{W}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$, we will show that $P_{\mathbf{x}}(\mathbf{y})$ is also not g-identifiable. To this end, first, we order the nodes in $\mathbf{W} \setminus \mathbf{Y}$, say $(W_1, W_2, \cdots, W_n)$, such that for each $1 \leq i \leq n$, W_i is a parent of at least one node in $\mathbf{Y} \cup \{W_1, W_2, \ldots, W_{i-1}\}$. Assume that e_i is the directed edge from W_i to its child in $\mathbf{Y} \cup \{W_1, W_2, \ldots, W_{i-1}\}$. We also define $\mathcal{G}'$ to be the graph obtained by deleting all the edges $\{e_i\}_{i=1}^n$ from $\mathcal{G}$. Applying Lemma 8 repeatedly n times imply that $Q[\mathbf{W}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}')$.

Let $\mathcal{G}_n := \mathcal{G}$ and for $0 \leq i \leq n-1$, we define $\mathcal{G}_i$ to be the graph obtained by removing e_{i+1} from $\mathcal{G}_{i+1}$. From Lemma 9, we know that if $Q[\mathbf{W}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}')$, then adding edge e_1 will make the causal effect of $\mathbf{X}$ on $\mathbf{W} \setminus \{W_1\}$ not g-identifiable from $(\mathbb{A}, \mathcal{G}_1)$. Note that $\mathcal{G}_1$ is obtained from $\mathcal{G}'$ by adding edge e_1. Using this lemma again implies that the causal effect of $\mathbf{X}$ on $\mathbf{W} \setminus \{W_1, W_2\}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}_2)$. Repeating this procedure yields that the causal effect of $\mathbf{X}$ on $\mathbf{W} \setminus \{W_1, \ldots W_n\} = \mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}_n)$. Since $\mathcal{G}_n = \mathcal{G}$, the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. □

Lemma 11. *Consider a set of vectors $\{c_i\}_{i=1}^n$, where $c_i \in \mathbb{R}^d$. Assume $c \in \mathbb{R}^d$ is a vector that is linearly independent of $\{c_i\}_{i=1}^n$, then there is a vector $b \in \mathbb{R}^d$ such that*

$$
\begin{aligned}
&\langle c_i, b \rangle = 0, \quad \forall i \in [1:n], \\
&\langle c, b \rangle \neq 0.
\end{aligned}
$$

Proof. Denote by $\{\phi_i\}_{i=1}^l$ a subset of $\{c_i\}_{i=1}^n$ which forms a basis for the vectors in $\{c_i\}_{i=1}^n$. Clearly, we have $l < d$. Now, consider the following system of linear equations with respect to b:

$$
\begin{aligned}
&\langle \phi_i, b \rangle = 0, \quad \forall i \in [1:l], \\
&\langle c, b \rangle = 13 \neq 0.
\end{aligned} \tag{14}
$$

By the assumption, vectors in $\{\phi_i\}_{i=1}^l \cup \{c\}$ are linearly independent, thus there exists a solution to (14). □

B.2 PROOFS OF SECTION 4

Proposition 3. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two disjoint subsets of $\mathbf{V}$. The causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[Anc_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}(\mathbf{Y})]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.*

Proof. Let $\mathbf{W} := Anc_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}(\mathbf{Y})$. Since $Q[\mathbf{V} \setminus \mathbf{X}] = P_{\mathbf{x}}(\mathbf{V} \setminus \mathbf{X})$, using marginalization, we obtain

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} Q[\mathbf{V} \setminus \mathbf{X}] = \sum_{\mathbf{W}\setminus\mathbf{Y}} \sum_{\mathbf{V}\setminus(\mathbf{W}\cup\mathbf{X})} Q[\mathbf{V} \setminus \mathbf{X}]. \tag{15}$$

Since $\mathbf{W}$ is an ancestral set in $\mathcal{G}[\mathbf{V} \setminus \mathbf{X}]$, Lemma 7 implies

$$\sum_{\mathbf{V}\setminus(\mathbf{W}\cup\mathbf{X})} Q[\mathbf{V} \setminus \mathbf{X}] = Q[\mathbf{W}].$$

Substituting the above equation into (15) implies

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{W}\setminus\mathbf{Y}} Q[\mathbf{W}] = P_{\mathbf{v}\setminus\mathbf{w}}(\mathbf{y}). \tag{16}$$

Sufficient part: Suppose $Q[\mathbf{W}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$. Equation (16) implies that $P_{\mathbf{x}}(\mathbf{y})$ is uniquely computable from $Q[\mathbf{W}]$, and therefore, the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.

Necessary part: Suppose $Q[\mathbf{W}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. For each $W \in \mathbf{W} \setminus \mathbf{Y}$, there exists a directed path in $\mathcal{G}[\mathbf{W}]$ from W to a variable in $\mathbf{Y}$. Hence, Lemma 10 implies that the causal effect of $\mathbf{V} \setminus \mathbf{W}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. Hence, Equation (16) implies that $P_{\mathbf{x}}(\mathbf{y})$ cannot be uniquely computed and the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$.

□

Proposition 4. *Suppose $\mathbf{S} \subseteq \mathbf{V}$ and $\mathbf{S}_1, \cdots, \mathbf{S}_l$ are the c-components of $\mathbf{S}$. $Q[\mathbf{S}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{S}_i]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ for each $i \in [1 : l]$.*

Proof. Sufficient part: Suppose $Q[\mathbf{S}_i]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ for each $i \in [1 : l]$. Tian and Pearl [2003] showed that

$$Q[\mathbf{S}] = \prod_{i=1}^{l} Q[\mathbf{S}_i].$$

Hence, $Q[\mathbf{S}]$ is uniquely computable and therefore, g-identifiable from $(\mathbb{A}, \mathcal{G})$.

Necessary part: Suppose $Q[\mathbf{S}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$. For $i \in [1 : l]$, Tian and Pearl [2003] provided a formula for computing $Q[\mathbf{S}_i]$ from $Q[\mathbf{S}]$ (Lemma 4, Equations (71) and (72) in [Tian and Pearl, 2003]). Hence, for each $i \in [1 : l]$, $Q[\mathbf{S}]$ is uniquely computable and therefore, g-identifiable from $(\mathbb{A}, \mathcal{G})$.

□

B.3 PROOFS OF SECTION 5

Lemma 2. *If $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$, then $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$.*

Proof. If $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$, then there exists two models $\mathcal{M}'_1$ and $\mathcal{M}'_2$ in $\mathbb{M}^+(\mathcal{G}')$ such that for each $i \in [0 : m]$ and any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$,

$$Q^{\mathcal{M}'_1}[\mathbf{A}'_i](\mathbf{v}) = Q^{\mathcal{M}'_2}[\mathbf{A}'_i](\mathbf{v}),$$

and there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ such that

$$Q^{\mathcal{M}'_1}[\mathbf{S}](\mathbf{v}_0) \neq Q^{\mathcal{M}'_2}[\mathbf{S}](\mathbf{v}_0).$$

Next, we will construct two models $\mathcal{M}_1$ and $\mathcal{M}_2$ in $\mathbb{M}^+(\mathcal{G})$ to prove that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. We define the domains of variables in $\mathbf{V}'$ in the model $\mathcal{M}_i$ similar to model $\mathcal{M}'_i$, for $i \in \{1, 2\}$. Since for each node $V \in \mathbf{V}'$, we have $Pa_{\mathcal{G}'}(V) \subseteq Pa_{\mathcal{G}}(V)$, then for all $V \in \mathbf{V}'$ and $i \in \{1, 2\}$, we can define:

$$P^{M_i}(V|Pa_{\mathcal{G}}(V)) := P^{M'_i}(V \mid Pa_{\mathcal{G}'}(V)).$$

And for $V \in \mathbf{V} \setminus \mathbf{V}'$, we define:

$$\mathfrak{X}_V = \{0\}, \quad P(V = 0) = 1.$$

Because variable $V \in \mathbf{V} \setminus \mathbf{V}'$ can only take value 0 with probability one, then $Q^{\mathcal{M}_j}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}'_j}[\mathbf{A}'_i](\mathbf{v})$ for all i and $Q^{\mathcal{M}_j}[\mathbf{S}](\mathbf{v}_0) = Q^{\mathcal{M}'_j}[\mathbf{S}](\mathbf{v}_0)$ for $j \in \{1, 2\}$. Thus, we have

$$\begin{aligned} &Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}), \quad i \in [0 : m], \\ &Q^{\mathcal{M}_1}[\mathbf{S}](\mathbf{v}_0) \neq Q^{\mathcal{M}_2}[\mathbf{S}](\mathbf{v}_0). \end{aligned}$$

This shows that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. □

Lemma 3. *Consider the following set of vectors in $\mathbb{R}^d$*

$$\mathbf{\Omega} := \{\theta_i(\mathbf{v}) : \ i \in [0 : m], \mathbf{v} \in \mathfrak{X}_{\mathbf{V}}\} \cup \mathbb{1}_d, \tag{17}$$

where $\mathbb{1}_d$ denotes the all-ones vector in $\mathbb{R}^d$. If there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}}$ such that $\eta(\mathbf{v}_0)$ is linearly independent from all the vectors in $\mathbf{\Omega}$, then the system of linear equations in (11) *admits a solution.*

Proof. This is a direct consequence of Lemma 11 with $\{c_i\}$ to be $\mathbf{\Omega}$ and c to be $\eta(\mathbf{v}_0)$. □

Lemma 4. *The SEM constructed above belongs to $\mathbb{M}^+(\mathcal{G}')$.*

Proof. By the construction, it is clear that the model belongs to $\mathbb{M}(\mathcal{G}')$. Hence, we need to show that $P(\mathbf{v}) > 0$ for any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$. To this end, it is enough to show that for any realization $\mathbf{v} \in \mathfrak{X}'_{\mathbf{V}}$, there exists a realization $\hat{\mathbf{u}} \in \mathfrak{X}_{\mathbf{U}'}$ such that $P(\mathbf{v}, \hat{\mathbf{u}}) > 0$, because in this case we have

$$P(\mathbf{v}) = \sum_{\mathbf{u} = \mathfrak{X}_{\mathbf{U}'}} P(\mathbf{v}, \mathbf{u}) \geq P(\mathbf{v}, \hat{\mathbf{u}}) > 0.$$

Let $\mathbf{v}$ be a fixed realization in $\mathfrak{X}_{\mathbf{V}'}$. For the rest of the proof, we assume all the realizations for $\mathbf{V}'$ are consistent with $\mathbf{v}$.

By Markov factorization property, for any $\mathbf{u} \in \mathfrak{X}_{\mathbf{U}'}$ we have

$$P(\mathbf{v}, \mathbf{u}) = \prod_{V \in \mathbf{V}'} P(v \mid Pa_{\mathcal{G}'}(V)) \prod_{U \in \mathbf{U}'} P(u). \tag{18}$$

By the construction of our model, we have $P(u) > 0$ for any $U \in \mathbf{U}'$ and $u \in \mathfrak{X}_U$. Moreover, for any $X \in \mathbf{S}$ and any realization for $Pa_{\mathcal{G}'}(X) \cap \mathbf{U}'$ we have $P(x \mid Pa_{\mathcal{G}'}(X)) > 0$. Hence, it is enough to show that there exists $\hat{\mathbf{u}} \in \mathfrak{X}_{\mathbf{U}'}$ such that $P(x \mid Pa_{\mathcal{G}'}(X)) > 0$ for each $X \in \mathbf{T}$.

Recall that for each $X \in \mathbf{T}$, we have $X = (X[i_1], \cdots, X[i_{\alpha(X)}])$, where X belongs to $\mathcal{F}_{i_1}, \cdots, \mathcal{F}_{i_{\alpha(X)}}$ and

$$X[i_j] \equiv \left(\sum_{Y \in Pa_{\mathcal{F}_{i_j}}(X)} Y[i_j] \right) \pmod 2.$$

By the construction, we define the entries corresponding to each $\mathcal{F}_i$ separately. For each $i \in [0 : k]$, let $\mathbf{U}_i$ to be the set of unobserved variables in $\mathbf{U}^{\mathbf{T}}$ that are in $\mathcal{F}_i$.

Let us fix an $i \in [0 : k]$. To finish the prove, we will introduce a method to determine $\hat{u}[i]$ for each $U \in \mathbf{U}_i$ such that

$$x[i] \equiv \left(\sum_{Y \in Pa_{\mathcal{F}_i}(X)} y[i] \right) \pmod 2, \tag{19}$$

for each $X \in \mathbf{T} \cap \mathbf{B}_i$.

Lets start with an arbitrary set of values for $\{\hat{u}[i]\colon\ U \in \mathbf{U}_i\}$ which are either 0 or 1. Suppose $X \in \mathbf{T} \cap \mathbf{B}_i$. We introduce a trick such that $x[i]$ will be replaced by $1 - x[i]$ while for all $Y \in \mathbf{T} \cap \mathbf{B}_i$, $y[i]$ remains the same:
By the construction of $\mathcal{F}_i$, there exists a path $(X = X_1, U_1, X_2, \cdots, X_l, U_l, Z = X_{l+1})$ from X to a variable $Z \in \mathbf{S}$ such that $\{U_1, \cdots, U_l\} \subseteq \mathbf{U}_i$, $\{X_1, \cdots, X_l\} \subseteq \mathbf{B}_i \cap \mathbf{T}$, and $Ch_{\mathcal{F}_i}(U_j) = \{X_j, X_{j+1}\}$ for each $j \in [1 : l]$. Now for each $j \in [1 : l]$, we replace $\hat{u}_j[i]$ by $1 - \hat{u}_j[i]$. Since Equation (19) is in mod 2, the value of $x_j[i]$ will be the same for each $j \in [2 : l]$ while $x[i]$ will be replaced by $1 - x[i]$. Note that $X_{l+1} = Z \notin \mathbf{T}$.

With the trick that we described above, we can construct any realization for the i-th bit of the variables in $\mathbf{T} \cap \mathbf{B}_i$. Hence, we can construct $\hat{\mathbf{u}} \in \mathfrak{X}_{\mathbf{U}'}$ such that $P(x \mid Pa_{\mathcal{G}'}(X)) > 0$ for each $X \in \mathbf{T}$.

□

Lemma 5. *For any* $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$ *and* $i \in [0 : m]$,

$$\theta_{i,j_1}(\mathbf{v}) = \theta_{i,j_2}(\mathbf{v}) = \cdots = \theta_{i,j_{\frac{\kappa+1}{2}}}(\mathbf{v}).$$

Proof. Lets fix a realization $\mathbf{v}$ for the observed variables $\mathbf{V}'$. Suppose that l_1 and l_2 are two integers such that

$$\begin{aligned}\gamma_{l_1} &= (2x, 0, \ldots, 0)),\\ \gamma_{l_2} &= (2x + 2 \pmod{\kappa + 1}, 0, \ldots, 0),\end{aligned}$$

where x is any fixed integer in $[0 : \frac{\kappa-1}{2}]$. To show the result, we will prove that $\theta_{i,l_1}(\mathbf{v}) = \theta_{i,l_2}(\mathbf{v})$. Let

$$\begin{aligned}f_{i,j}(\mathbf{v}, \mathbf{u}^{\mathbf{T}}) &:= \sum_{\mathbf{u} \in \mathbf{U}^{\mathbf{S}}} \prod_{V \in \mathbf{A}'_i} P(v \mid Pa_{\mathcal{G}'}(V)) \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u)\\ &= \prod_{V \in \mathbf{A}'_i \setminus \mathbf{B}_i} P(v \mid Pa_{\mathcal{G}'}(V)) \prod_{V \in \mathbf{B}_i \setminus \mathbf{S}} P(v \mid Pa_{\mathcal{G}'}(V)) \sum_{\mathbf{u} \in \mathbf{U}^{\mathbf{S}}} \prod_{V \in \mathbf{S}} P(v \mid Pa_{\mathcal{G}'}(V)) \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u).\end{aligned}$$

where index j indicates $U_0 = \gamma_j$. Note that variable U_0 may appear in the parent set of some observed variables. Using the above definition, we have

$$\theta_{i,j}(\mathbf{v}) = \sum_{\mathbf{u}^{\mathbf{T}} \in \mathbf{U}^{\mathbf{T}}} f_{i,j}(\mathbf{v}, \mathbf{u}^{\mathbf{T}}).$$

Hence, if we show $f_{i,l_1}(\mathbf{v}, \mathbf{u}^{\mathbf{T}}) = f_{i,l_2}(\mathbf{v}, \mathbf{u}^{\mathbf{T}})$ for any fixed realization $\mathbf{u}^{\mathbf{T}}$, the above equation implies $\theta_{i,l_1}(\mathbf{v}) = \theta_{i,l_2}(\mathbf{v})$.

When $T \in \mathbf{A}'_i \setminus \mathbf{B}_i$, then for fixed realizations of $\mathbf{u}^{\mathbf{T}}$, $P(t|Pa_{\mathcal{G}'}(T))$ is the same for both realizations γ_{l_1} and γ_{l_2} since $\gamma_{l_1} \equiv \gamma_{l_2} \bmod 2$.

When $T \in \mathbf{B}_i \setminus \mathbf{S}$, unobserved variables in $Pa_{\mathcal{G}'}(T)$ are a subset of $\mathbf{U}^{\mathbf{T}} \cup \{U_0\}$. Note that in the definition of $f_{i,j}(\mathbf{v}, \mathbf{u}^{\mathbf{T}})$, all such unobserved variables are fixed. Thus, if there exists $T \in \mathbf{B}_i \setminus \mathbf{S}$, such that $P(t|Pa_{\mathcal{G}'}(T)) = 0$, then

$$f_{i,l_1}(\mathbf{v}, \mathbf{u}^{\mathbf{T}}) = f_{i,l_2}(\mathbf{v}, \mathbf{u}^{\mathbf{T}}) = 0.$$

When $P(t|Pa_{\mathcal{G}'}(T)) = 1$ for all $T \in \mathbf{B}_i \setminus \mathbf{S}$, to prove $f_{i,l_1}(\mathbf{v}, \mathbf{u}^{\mathbf{T}}) = f_{i,l_2}(\mathbf{v}, \mathbf{u}^{\mathbf{S}})$, we show that for any realization $(\mathbf{u}_1, \gamma_{l_1})$ of $(\mathbf{U}^{\mathbf{S}}, U_0)$, there is a realization $(\mathbf{u}_2, \gamma_{l_2})$ of $(\mathbf{U}^{\mathbf{S}}, U_0)$ such that

$$\prod_{V \in \mathbf{S}} P(v \mid Pa_{\mathcal{G}'}(V))\Big|_{(\mathbf{U}^{\mathbf{S}}, U_0) = (\mathbf{u}_1, \gamma_{l_1})} = \prod_{V \in \mathbf{S}} P(v \mid Pa_{\mathcal{G}'}(V))\Big|_{(\mathbf{U}^{\mathbf{S}}, U_0) = (\mathbf{u}_2, \gamma_{l_2})},$$

where $P(v \mid Pa_{\mathcal{G}'}(V))\Big|_{(\mathbf{U}^{\mathbf{S}}, U_0) = (\mathbf{u}_1, \gamma_{l_1})}$ denotes the conditional probability of v given its parents in which the unobserved variables $(\mathbf{U}^{\mathbf{S}}, U_0)$ are fixed to be $(\mathbf{u}_1, \gamma_{l_1})$. To this end, we consider two cases depending on i.

First case, when $i \in [0 : k]$**:** In this case, we have

$$t[i] = \left(\sum_{Y \in Pa_{\mathcal{F}_i}(T)} y[i] \right) \pmod 2. \tag{20}$$

Consider the set $\mathbf{\Lambda} := Pa_{\mathcal{F}_i}(\mathbf{S}) \setminus Pa_{\mathcal{F}_i[\mathbf{S}]}(\mathbf{S})$, that is the set of all parents of nodes in $\mathbf{S}$ that are outside of $\mathbf{S}$. By the construction of our models, summation of the values of the observed and unobserved nodes in $\mathbf{\Lambda}$ are the same, i.e.,

$$\sum_{W \in \mathbf{\Lambda} \cap \mathbf{B}_i} w[i] \equiv \sum_{W \in \mathbf{\Lambda} \cap \mathbf{U}'} w[i] \pmod 2,$$

or equivalently

$$\sum_{W \in \mathbf{\Lambda}} w[i] \equiv 0 \pmod 2. \tag{21}$$

This is because in graph $\mathcal{F}_i$, each observed variable outside of $\mathbf{S}$ has at most one child outside of $\mathbf{S}$ and each unobserved node has either one or two children outside of $\mathbf{S}$. According to (20), those unobserved nodes with two children outside of $\mathbf{S}$ do not belong to $\mathbf{\Lambda} \cap \mathbf{U}'$. Such unobserved nodes have exactly two observed descendants in $\mathbf{\Lambda} \cap \mathbf{B}_i$ and because both descendants appear in (21), their summation is zero mod 2. On the other hand, the unobserved nodes with only one child outside of $\mathbf{S}$ belong to $\mathbf{\Lambda} \cap \mathbf{U}'$ and have exactly one observed descendant in $\mathbf{\Lambda} \cap \mathbf{B}_i$. Thus, the summation of such unobserved variables and their observed descendant is again zero mod 2 in (21).

If $\mathbb{I}(S) = 0$ for all $S \in \mathbf{S}$, then by our model construction, for any variable $W \in \mathbf{\Lambda} \setminus \{T_i\}$, $w[i]$ is an even number but T_i takes value 1 with probability one. Hence, the summation in (21) cannot be an even number. Therefore, there exists at least a variable $S \in \mathbf{S}$ such that $\mathbb{I}(S) = 1$. In this case, the value of $P(S|Pa_{\mathcal{G}'}(S))$ does not depend on the realizations of variables in $\mathbf{U}^{\mathbf{S}}$. Next, we show that for any realization $\mathbf{u}_1$ of $\mathbf{U}^{\mathbf{S}}$, there is a realization $\mathbf{u}_2$ such that

$$P(s|Pa_{\mathcal{G}'}(S))\Big|_{(\mathbf{U}^{\mathbf{S}},U_0)=(\mathbf{u}_1,\gamma_{l_1})} = P(s|Pa_{\mathcal{G}'}(S))\Big|_{(\mathbf{U}^{\mathbf{S}},U_0)=(\mathbf{u}_2,\gamma_{l_2})}. \tag{22}$$

Since $\mathcal{G}'_{\mathbf{S}}$ is a c-component, there exists a sequence of variables $U_0, \hat{S}_1, \hat{U}_1, \hat{S}_2, \hat{U}_2, \ldots, \hat{U}_l, S$, such that U_0 is a parent of $\hat{S}_1$, S is a children of $\hat{U}_l$ and $\hat{U}_j$ is a parent of variables $\hat{S}_j$ and $\hat{S}_{j+1}$ for $j \in [1 : l-1]$. Let $\hat{\mathbf{U}} := \{\hat{U}_1, \ldots, \hat{U}_l\}$. For realization $\mathbf{u}_1$, we define $\mathbf{u}_2$ by

$$\begin{aligned} &u_{2,\hat{U}_j} := u_{1,\hat{U}_j} + 2(-1)^j \pmod{\kappa+1}, \quad j \in [1:l], \\ &u_{2,U} := u_{1,U}, \quad \forall U \in \mathbf{U}' \setminus (\hat{\mathbf{U}} \cup \{U_0\}), \end{aligned} \tag{23}$$

where $u_{2,U}$ denotes the realization for variable U in $\mathbf{u}_2$. It is straightforward to see that this mapping is a bijection between $\mathbf{u}_1$ and $\mathbf{u}_2$ and (22) holds.

Second case, when $i \in [k+1 : m]$: In this case, $\mathbf{S} \setminus \mathbf{A}'_i \neq \varnothing$. Since $\mathcal{G}'_{\mathbf{S}}$ is a c-component, there exists a sequence of variables $U_0, \hat{S}_1, \hat{U}_1, \hat{S}_2, \hat{U}_2, \ldots, \hat{U}_l, S$, such that U_0 is a parent of $\hat{S}_1$, $S \in \mathbf{S} \setminus \mathbf{A}'_i$ is a children of $\hat{U}_l$ and $\hat{U}_j$ is a parent of variables $\hat{S}_j$ and $\hat{S}_{j+1}$ for $j \in [1 : l-1]$. Let $\hat{\mathbf{U}} := \{\hat{U}_1, \ldots, \hat{U}_l\}$. Similar to the previous case, for a given realization $\mathbf{u}_1$ of $\mathbf{U}^{\mathbf{S}}$, we define $\mathbf{u}_2 \in \mathfrak{X}_{\mathbf{U}^{\mathbf{S}}}$ by

$$\begin{aligned} &u_{2,\hat{U}_j} := u_{1,\hat{U}_j} + 2(-1)^j \pmod{\kappa+1}, \quad j \in [1:l], \\ &u_{2,U} := u_{1,U}, \quad \forall U \in \mathbf{U}' \setminus (\hat{\mathbf{U}} \cup \{U_0\}), \end{aligned} \tag{24}$$

where $u_{2,\hat{U}}$ denotes the realization for variable U in $\mathbf{u}_2$. Analogous to the previous setting, we have (22).

Herein, we proved that $\theta_{i,l_1}(\mathbf{v}) = \theta_{i,l_2}(\mathbf{v})$. By varying x within $[0 : \frac{\kappa-1}{2}]$ in the definition of γ_{l_1} and γ_{l_2}, we conclude the lemma. □

Lemma 6. *There exists $0 < \epsilon < \frac{1}{\kappa}$ such that there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ and $1 \leq r < t \leq \frac{\kappa+1}{2}$ such that*

$$\eta_{j_r}(\mathbf{v}_0) \neq \eta_{j_t}(\mathbf{v}_0).$$

Proof. Lets consider r and t such that $\gamma_r = (0, 0, \ldots, 0)$ and $\gamma_t = (2, 0, \ldots, 0)$. Recall that:

$$\eta_r(\mathbf{v}) := \sum_{\mathbf{U}' \setminus \{U_0\}} \prod_{X \in \mathbf{S}} P(x \mid Pa_{\mathcal{G}'}(X))\Big|_{U_0=\gamma_r} \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u), \tag{25}$$

$$\eta_t(\mathbf{v}) := \sum_{\mathbf{U}' \setminus \{U_0\}} \prod_{X \in \mathbf{S}} P(x \mid Pa_{\mathcal{G}'}(X))\Big|_{U_0=\gamma_t} \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u). \tag{26}$$

We choose $\mathbf{v}_0$ as follows: set all variables in $\mathbf{S}$ to be zero and select a realization for variables in $\mathbf{V}' \setminus \mathbf{S}$ such that $\mathbb{I}(S) = 0$ for all $S \in \mathbf{S}$. Denote by S_0 a child of U_0 in $\mathbf{S}$.

Note that there is a term in the summation of the right side of equation (25) that is $(1 - \kappa\epsilon)^{|\mathbf{S}|}$. For instance, this occurs when all realizations of unobserved variables in $\mathbf{U}^{\mathbf{S}}$ are zero.

Next, we prove that there is no realization of unobserved variables $\mathbf{U}^{\mathbf{S}}$ such that $P(S|Pa_{\mathcal{G}'}(S)) = 1 - \epsilon\kappa$ for all $S \in \mathbf{S}$ and $U_0 = \gamma_t$. In other words, each term in the summation of (26) has at least a term ϵ. To do so, it suffices to show that there is no realization of $\mathbf{U}^{\mathbf{S}}$ such that:

$$s = \sum_{W \in Pa_{\mathcal{G}'[\mathbf{S}]}(S)} w, \quad S \in \mathbf{S} \setminus \{S_0\},$$
$$s_0 = u_0[0] + \sum_{W \in Pa_{\mathcal{G}'[\mathbf{S}]}(S)} w.$$

Suppose, there is a realization of $\mathbf{U}^{\mathbf{S}}$ such that the above equations hold. In this case, since $\mathcal{G}'_{\mathbf{S}}$ is a tree, we can color its nodes by two colors, red and black, such that connected nodes by biderected edges have different colors. Suppose that $\mathbf{S}_1$ is the set of black variables and $\mathbf{S}_2$ is the set of red variables which (without loss of generality) contains $S_0 \in \mathbf{S}_1$. Then:

$$\sum_{W \in \mathbf{S}_1} w \equiv u_0[0] + \sum_{U \in \mathbf{U}^{\mathbf{S}}} u \pmod{\kappa + 1},$$
$$\sum_{W \in \mathbf{S}_2} w \equiv \sum_{U \in \mathbf{U}^{\mathbf{S}}} u \pmod{\kappa + 1}.$$

Left hand sides of both above equations are zero because of our choice of $\mathbf{v}_0$. However, the right hand sides cannot be the same since $u_0[0] = 2$. Hence, in Equation (26), there exists a term in the summation with probability ϵ. Therefore, in extreme case, when $\epsilon = 0$, $\eta_t(\mathbf{v}') = 0$. However, $\eta_r(\mathbf{v}') \geq (1 - \kappa\epsilon)^{|\mathbf{S}|} \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u) > 0$. Since $\eta_r(v)$ and $\eta_t(v)$ are polynomial functions of ϵ and they are not equal at $\epsilon = 0$, then there exists a small enough $0 < \epsilon < \frac{1}{\kappa}$ such that $\eta_r(\mathbf{v}') \neq \eta_t(\mathbf{v}')$. □

C A SPECIAL CASE IN THE PROOF OF THEOREM 1

In this section, we provide our proof for the necessary part of Theorem 1 when $\mathbf{S} \nsubseteq \mathbf{A}'_i$ for all $i \in [0 : m]$.

We define $\mathcal{F}^{\mathbf{S}}$ to be a minimal (in terms of edges) spanning subgraph of $\mathcal{G}[\mathbf{S}]$ such that $\mathcal{F}^{\mathbf{S}}_{\mathbf{S}}$ is a single c-component. In this case, we can assume $\mathbf{V}' = \mathbf{S}$, $\mathcal{G}'$ is $\mathcal{F}^{\mathbf{S}}$, and $\mathbb{A}' = \{\mathbf{A}'_i := \mathbf{A}_i \cap \mathbf{V}'\}_{i=0}^m$. For each $i \in [0 : m]$, we have $\mathbf{A}'_i \subsetneq \mathbf{V}'$. Note that Lemma 2 holds for this case. Hence, it is enough to show that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$.

Recall that our assumptions and goal in this section are as follows:
$\mathcal{G}'$ is a DAG with observed variables $\mathbf{V}'$ and unobserved variables $\mathbf{U}'$ such that $\mathcal{G}'_{\mathbf{V}'}$ has no directed edges and its bidirected edges form a spanning tree over $\mathbf{V}'$. $\mathbb{A}' = \{\mathbf{A}'_i\}_{i=0}^m$ is a collection of subsets such that $\mathbf{A}'_i \subsetneq \mathbf{V}'$. The goal is to show that $Q[\mathbf{V}']$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$.

For this case we will define two model $\mathcal{M}_1$ and $\mathcal{M}_2$ such that for each $i \in [0 : m]$ and any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$,

$$Q^{\mathcal{M}_1}[\mathbf{A}'_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}'_i](\mathbf{v}),$$

but there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ such that

$$Q^{\mathcal{M}_1}[\mathbf{S}](\mathbf{v}_0) \neq Q^{\mathcal{M}_2}[\mathbf{S}](\mathbf{v}_0).$$

For both models $\mathcal{M}_1$ and $\mathcal{M}_2$ we define each observed and unobserved variable to be binary, i.e $\mathfrak{X}_W = \{0, 1\}$ for all $W \in \mathbf{V}' \cup \mathbf{U}'$. Next, we define the equation of the variables in each model.

Model 1: For $V \in \mathbf{V}'$:

$$V = \begin{cases} \bigoplus Pa_{\mathcal{G}'}(V), & \text{with probability } 1 - \epsilon, \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{27}$$

and for $U \in \mathbf{U}'$:

$$P(U=0) = P(U=1) = 0.5.$$

Model 2: Suppose V_1 is a fixed observed variable in $\mathbf{V}'$. Then, for all V in $\mathbf{V}' \setminus \{V_1\}$ we define:

$$V = \begin{cases} \bigoplus Pa_{\mathcal{G}'}(V), & \text{with probability } 1-\epsilon \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{28}$$

and for V_1:

$$V_1 = \begin{cases} \neg \bigoplus Pa_{\mathcal{G}'}(V_1), & \text{with probability } 1-\epsilon \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{29}$$

where $\neg$ denotes the logical not. Similar to the first mode, for each unobserved variables $U \in \mathbf{U}'$,

$$P(U=0) = P(U=1) = 0.5.$$

Lemma 17. *Let $i \in [0, m]$ and denote the cardinality of $\mathbf{A}'_i$ by n, i.e. $|\mathbf{A}'_i| = n$. Then for any realization $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$:*

$$Q^{\mathcal{M}_1}[\mathbf{A}'_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}'_i](\mathbf{v}) = \frac{1}{2^n}.$$

Proof. Suppose $\mathbf{A}'_i := \{A_1, A_2, \ldots, A_n\}$. Since $\mathbf{A}'_i \subsetneq \mathbf{V}'$, there are distinct unobserved variables $U_1, U_2, \ldots, U_n$, such that U_j is a parent of the A_j for $j \in [1:n]$. Denote by $\mathcal{M}$ any of the model $\mathcal{M}_1$ or $\mathcal{M}_2$.

Assume that for some realization of observed and unobserved variables, exactly $t \in [0, n]$ variables in $\mathbf{A}'_i$ are defined by the XOR or $\neg XOR$ of their parents. Without loss of generality, assume that these variables are $\{A_1, A_2, \ldots, A_t\}$. If we know all unobserved variables $\mathbf{U}'$ except $\{U_1, U_2, \ldots, U_t\}$, then we can determine uniquely the values of $\{U_1, U_2, \ldots, U_t\}$ from the following equations:

$$A_i = \bigotimes Pa_{\mathcal{G}'}(A_i), \quad i \in [1:t],$$

where $\bigotimes$ denotes the corresponding equation, either XOR or $\neg XOR$, for variable A_i in model $\mathcal{M}$. Thus, by considering all possible realizations of unobserved variables that lead to a realization $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$, we obtain

$$Q[\mathbf{A}'_i] = \sum_{j=0}^{n} C_n^j (1-\epsilon)^j \left(\frac{\epsilon}{2}\right)^{n-j} \left(\frac{1}{2}\right)^j = \left(\frac{1}{2}\right)^n,$$

where C_n^j is the number of different ways to choose j variables out of n, such that with probability $(1-\epsilon)$ their values are determined by either XOR or $\neg XOR$ equation. All other $n-j$ variables are equal to either 0 or 1 with probability $\frac{\epsilon}{2}$. $\square$

Lemma 18. *Let $\mathbf{v} = \mathbf{0}$ be the realization of $\mathbf{V}'$ such that all observed variables are equal to 0. Then $Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{v}) \neq Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{v})$.*

Proof. Define $n = |\mathbf{V}'|$ and $\mathbf{V}' = \{V_1, V_2, \ldots, V_n\}$. Firstly, we will prove that for any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$, the value of $Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{v})$ does not depend on the position of V_1 in graph $\mathcal{G}'$. Denote by V_2 an observed variable which is connected to the V_1 by a bidirected edge in $\mathcal{G}'_{\mathbf{V}'}$. Let U denotes the unobserved variable (corresponding to the bidirected edge) which is a parent of V_1 and V_2. Next, we define a new model $\mathcal{M}'_2$ in which all variables in $\mathbf{V}'$ are defined similarly as they are defined in model $\mathcal{M}_2$ except for variables V_1 and V_2. In $\mathcal{M}'_2$, we define V_2 in the same way as V_1 is defined in $\mathcal{M}_2$. We also define V_1 in $\mathcal{M}'_2$ in the same way as V_2 is defined in $\mathcal{M}_2$. Then, we have

$$\prod_{i=1}^{n} P^{\mathcal{M}_2}(v_i | Pa_{\mathcal{G}'}(V_i)) = P^{\mathcal{M}_2}(v_1 | Pa_{\mathcal{G}'}(V_1)) P^{\mathcal{M}_2}(v_2 | Pa_{\mathcal{G}'}(V_2)) \prod_{i=3}^{n} P(v_i | Pa_{\mathcal{G}'}(V_i))$$

$$= P^{\mathcal{M}'_2}(v_1 | Pa_{\mathcal{G}'}(V_1) \setminus \{U\}, u \oplus 1) P^{\mathcal{M}'_2}(v_2 | Pa_{\mathcal{G}'}(V_2) \setminus \{U\}, u \oplus 1) \prod_{i=3}^{n} P(v_i | Pa_{\mathcal{G}'}(V_i)).$$

This implies that substituting V_1 by V_2 does not change the value of $Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{v})$.

Without loss of generality, suppose that V_1 is a leaf in $\mathcal{G}'$ and U_1 is a parent of V_1. Note that there are exactly $n-1$ unobserved variables in graph $\mathcal{G}'$. This is because $\mathcal{G}'_{\mathbf{V}'}$ is a tree with bidirected edges over $\mathbf{V}'$. Therefore, we have

$$2^{n-1}Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{0}) = P^{\mathcal{M}_1}(V_1=0|U_1=0)\sum_{\mathbf{U}'\setminus\{U_1\}}\prod_{j>1}P(v_j|Pa_{\mathcal{G}'}(V_j)) + P^{\mathcal{M}_1}(V_1=0|U_1=1)\sum_{\mathbf{U}'\setminus\{U_1\}}\prod_{j>1}P(v_j|Pa_{\mathcal{G}'}(V_j)),$$
$$2^{n-1}Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{0}) = P^{\mathcal{M}_2}(V_1=0|U_1=0)\sum_{\mathbf{U}'\setminus\{U_1\}}\prod_{j>1}P(v_j|Pa_{\mathcal{G}'}(V_j)) + P^{\mathcal{M}_2}(V_1=0|U_1=1)\sum_{\mathbf{U}'\setminus\{U_1\}}\prod_{j>1}P(v_j|Pa_{\mathcal{G}'}(V_j)).$$

Note that:

$$\begin{aligned}
P^{\mathcal{M}_1}(V_1=0|U_1=0) &= 1-\frac{\epsilon}{2}\\
P^{\mathcal{M}_1}(V_1=0|U_1=1) &= \frac{\epsilon}{2}\\
P^{\mathcal{M}_2}(V_1=0|U_1=0) &= \frac{\epsilon}{2}\\
P^{\mathcal{M}_2}(V_1=0|U_1=1) &= 1-\frac{\epsilon}{2}
\end{aligned}$$

More over, we have

$$\sum_{U_1=0,\mathbf{U}'\setminus\{U_1\}}\prod_{j>1}P(v_j|Pa_{\mathcal{G}'}(V_j)) + \sum_{U_1=1,\mathbf{U}'\setminus\{U_1\}}\prod_{j>1}P(v_j|Pa_{\mathcal{G}'}(V_j)) = Q[\mathbf{V}'\setminus\{V_1\}] = \left(\frac{1}{2}\right)^{n-1}$$

This yields

$$\begin{aligned}
2^{n-1}Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{0}) &= \left(1-\frac{\epsilon}{2}\right)a + \frac{\epsilon}{2}b,\\
2^{n-1}Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{0}) &= \left(1-\frac{\epsilon}{2}\right)b + \frac{\epsilon}{2}a,
\end{aligned}$$

where

$$\begin{aligned}
a &= \sum_{U_1=0,\mathbf{U}'\setminus\{U_1\}}\prod_{j>1}P(V_j=0|Pa_{\mathcal{G}'}(V_j)),\\
b &= \sum_{U_1=1,\mathbf{U}'\setminus\{U_1\}}\prod_{j>1}P(V_j=0|Pa_{\mathcal{G}'}(V_j)).
\end{aligned}$$

To prove that $Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{0}) \neq Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{0})$, it is enough to show that $a \neq b$.

Denote by S_n an observed variable connected to the V_1 by a bidirect edge in $\mathcal{G}'_{\mathbf{V}'}$. We define $\mathbf{V}'_{n-1} := \mathbf{V}'\setminus\{V_1\}$, $\mathbf{U}'_{n-1} := \mathbf{U}'\setminus\{U_1\}$ and $\mathcal{G}_{n-1} := \mathcal{G}'[\mathbf{V}'\setminus\{V_1\}]$. We also define models $\mathcal{M}_1^{(n-1)}$ and $\mathcal{M}_2^{(n-1)}$ as follows:

New model $\mathcal{M}_1^{(n-1)}$: For $V \in \mathbf{V}'_{n-1}$:

$$V = \begin{cases} \bigoplus Pa_{\mathcal{G}_{n-1}}(V), & \text{with probability } 1-\epsilon,\\ 1, & \text{with probability } \frac{\epsilon}{2},\\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{30}$$

and for $U \in \mathbf{U}'_{n-1}$:

$$P(U=0) = P(U=1) = 0.5.$$

Model $\mathcal{M}_2^{(n-1)}$: For all V in $\mathbf{V}'_{n-1}\setminus\{S_n\}$:

$$V = \begin{cases} \bigoplus Pa_{\mathcal{G}_{n-1}}(V), & \text{with probability } 1-\epsilon\\ 1, & \text{with probability } \frac{\epsilon}{2},\\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{31}$$

and for S_n:

$$S_n = \begin{cases} \neg \bigoplus Pa_{\mathcal{G}_{n-1}}(S_n), & \text{with probability } 1-\epsilon \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}. \end{cases} \tag{32}$$

Similar to the first model, for each unobserved variables $U \in \mathbf{U}'_{n-1}$, we define

$$P(U=0) = P(U=1) = 0.5.$$

Note that:

$$\left(\frac{1}{2}\right)^{n-2} \sum_{U_1=0, \mathbf{U}'_{n-1}} \prod_{j>1} P(V_j | Pa_{\mathcal{G}'}(V_j)) = \left(\frac{1}{2}\right)^{n-2} a = Q^{\mathcal{M}_1^{(n-1)}}[\mathbf{V}'_{n-1}](\mathbf{0}),$$
$$\left(\frac{1}{2}\right)^{n-2} \sum_{U_1=1, \mathbf{U}'_{n-1}} \prod_{j>1} P(V_j | Pa_{\mathcal{G}'}(V_j)) = \left(\frac{1}{2}\right)^{n-2} b = Q^{\mathcal{M}_2^{(n-1)}}[\mathbf{V}'_{n-1}](\mathbf{0}).$$

It remains to show $Q^{\mathcal{M}_1^{(n-1)}}[\mathbf{V}'_{n-1}](\mathbf{0}) \neq Q^{\mathcal{M}_2^{(n-1)}}[\mathbf{V}'_{n-1}](\mathbf{0})$. Note that if this holds, then by our construction, $Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{0}) \neq Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{0})$. In other words, we could reduce the size of the graph while keeping the same problem. Thus, by continuing this procedure, we eventually reach graph $\mathcal{G}_2$ that consists of only two observed nodes and showing $Q^{\mathcal{M}_1^{(2)}}[\mathbf{V}'_2](\mathbf{0}) \neq Q^{\mathcal{M}_2^{(2)}}[\mathbf{V}'_2](\mathbf{0})$ in that graph will conclude the result. For graph $\mathcal{G}_2$, we have

$$Q^{\mathcal{M}_1^{(2)}}[\mathbf{V}'_2](\mathbf{0}) = \left(\frac{\epsilon}{2}\right)^2 + 2\frac{\epsilon}{2}(1-\epsilon)\frac{1}{2} + (1-\epsilon)^2\frac{1}{2},$$
$$Q^{\mathcal{M}_2^{(2)}}[\mathbf{V}'_2](\mathbf{0}) = \left(\frac{\epsilon}{2}\right)^2 + 2\frac{\epsilon}{2}(1-\epsilon)\frac{1}{2}.$$

This clearly shows that $Q^{\mathcal{M}_1^{(2)}}[\mathbf{V}'_2](\mathbf{0}) \neq Q^{\mathcal{M}_2^{(2)}}[\mathbf{V}'_2](\mathbf{0})$. □