On Identifiability of Conditional Causal Effects

Yaroslav Kivva[1] **Jalal Etesami**[1, 3] **Negar Kiyavash**[1,2]

[1]School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland
[2]College of Management of Technology, EPFL, Lausanne, Switzerland
[3]TUM School of Computation, Information and Technology, Technical University of Munich

Abstract

We address the problem of identifiability of an arbitrary *conditional causal effect* given both the causal graph and a set of any observational and/or interventional distributions of the form $Q[S] := P(S|do(V\backslash S))$, where V denotes the set of all observed variables and $S \subseteq V$. We call this problem conditional generalized identifiability (**c-gID** in short) and prove the completeness of Pearl's *do*-calculus for the c-gID problem by providing sound and complete algorithm for the c-gID problem. This work revisited the c-gID problem in Lee et al. [2020], Correa et al. [2021] by adding explicitly the positivity assumption which is crucial for identifiability. It extends the results of [Lee et al., 2019, Kivva et al., 2022] on general identifiability (gID) which studied the problem for *unconditional* causal effects and Shpitser and Pearl [2006b] on identifiability of conditional causal effects given *merely* the observational distribution $P(\mathbf{V})$ as our algorithm generalizes the algorithms proposed in [Kivva et al., 2022] and [Shpitser and Pearl, 2006b].

1 INTRODUCTION

This paper addresses the problem of identification of a conditional post-interventional distribution from the combination of observational and/or interventional distributions. Formally, the relationships between the variables of interest are established by a directed acyclic graph (DAG) Pearl [1995]. Each node in the causal graph represents some random variable that may simulate real-life measurements, and each directed edge encodes a possible causal relationship between the variables. In general, a subset of the nodes in DAG are observed and others may be hidden. The hidden nodes could result in spurious correlations between observed variables and complicate the question of identifiability. On the other hand, when all the variables in the system are observable and the distribution over them is known then any conditional causal effect is identifiable.

The question of identification of the causal effect has been one of the central focus of research in causal inference literature. The classical setting of the problem asks whether the causal effect $P_{\mathbf{x}}(\mathbf{y})$[1] is identifiable in a given graph $\mathcal{G}$ from observational distribution $P(\mathbf{V})$ ($\mathbf{V}$ is a set of all observed nodes in the graph $\mathcal{G}$). The problem was solved in Shpitser and Pearl [2006a], Huang and Valtorta [2006] and later Shpitser and Pearl [2006b] extended the result by answering the question when a conditional causal effect $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ is identifiable in a given graph $\mathcal{G}$. The work of Bareinboim and Pearl [2012], Lee et al. [2019], Kivva et al. [2022] solved a generalization of the classical identifiability problem, namely identifiability of unconditional causal effect $P_{\mathbf{x}}(\mathbf{y})$ from a specific mix of observational and interventional distributions. It is noteworthy that all aforementioned works proved that the rules of do-calculus are sound and complete for the identification of the causal effect in their settings. Furthermore, the work of Tikka et al. [2021], Mokhtarian et al. [2022], Bareinboim and Pearl [2014], Bareinboim and Tian [2015] considers the problem of identifiability in a presence of additional information to observational/interventional distributions and the causal graph $\mathcal{G}$. More specifically, Mokhtarian et al. [2022] considers the identifiability problem in the presence of additional knowledge in the form of context-specific independence for some variables. Tikka et al. [2021] assumes that they have access to multiple incomplete data sources and Bareinboim and Tian [2015] studies the identifiability problem under a selection bias.

In this paper, we extend both the general identifiability (gID) result of Kivva et al. [2022] and the conditional identifiability result of Shpitser and Pearl [2006b]. More specifically,

[1]This notation indicates causal effect on $\mathbf{y}$ after intervention $do(\mathbf{X} = \mathbf{x})$, That is, $P(\mathbf{y}|do(\mathbf{X} = \mathbf{x}))$ shortened to $P_{\mathbf{x}}(\mathbf{y})$.

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

our work answers the question of identifiability of an *arbitrary* conditional causal effect $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ under the same set of assumptions as in gID problem. We call this problem *conditional general identifiability*, for short **c-gID**. This problem has been studied in Lee et al. [2020], Correa et al. [2021]. The authors of Lee et al. [2020] generalizes the problem of c-gID by assuming that observable data is available from multiple domains and Correa et al. [2021] considers the c-gID problem as an identifiability problem of counterfactual quantities. However, both of the aforementioned works are based on causal models that violate the positivity assumption (See Appendix B) which is crucial for identification as it is discussed in Kivva et al. [2022]. Since they did not discuss whether their proposed models can be modified such that the positivity assumption holds and it is not straightforward whether such modifications exist, herein we present an alternative proof for the c-gID problem including its soundness and completeness. The causal models developed here for proving the completeness of our algorithm are novel and satisfy the positivity assumption.

2 PRELIMINARIES

2.1 NOTATION AND DEFINITIONS

We denote random variables by capital letters and their realization by their lower-case version. Similarly, a set of random variables and their realizations are denoted by bold capital and bold lower-case letters, respectively. For two integers $a \leqslant b$, we define $[a:b] := \{a, a+1, \dots, b\}$. For any random variable X, we denote its domain set by $\mathfrak{X}(X)$ and for any set of random variables $\mathbf{X}$, we denote by $\mathfrak{X}(\mathbf{X})$, the Cartesian product of the domains of the variables in $\mathbf{X}$. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are arbitrary sets of random variables, then we say that realizations $\mathbf{x}$ and $\mathbf{y}$ are *consistent*, if the values of $\mathbf{X} \cap \mathbf{Y}$ in $\mathbf{x}$ and $\mathbf{y}$ are the same. Also, we use $\mathfrak{X}_{\mathbf{y}}(\mathbf{X})$ to denote a set of realizations of $\mathbf{X}$ that are consistent with $\mathbf{y}$. Suppose that $\mathbf{X}' \subseteq \mathbf{X}$ and $\mathbf{x}$ to be a realization of $\mathbf{X}$. Then, we use $\mathbf{x}[\mathbf{X}']$ to denote a realization of $\mathbf{X}'$ that is consistent with $\mathbf{x}$. When it is clear from the context, we write $\mathbf{x}'$ instead of $\mathbf{x}[\mathbf{X}']$.

Causal Graph: Consider a directed graph $\mathcal{G} := (\mathbf{V} \cup \mathbf{U}, \mathbf{E})$ over node $\mathbf{V} \cup \mathbf{U}$ in which $\mathbf{V}$ and $\mathbf{U}$ denote the set of observed and hidden variables, respectively and $\mathbf{E} \subseteq (\mathbf{V} \cup \mathbf{U}) \times (\mathbf{V} \cup \mathbf{U})$ denotes the set of directed edges. A causal graph $\mathcal{G}$ is a directed acyclic[2] graph (DAG). We say that node X is a parent of another node Y (subsequently, Y is a child of X) if and only if there exists a direct edge from X to Y in $\mathcal{G}$, e.g. $(X, Y) \in \mathbf{E}$. Similarly, X is said to be an ancestor of Y (subsequently, Y is a descendant of X) if and only if there is a directed path from X to Y in $\mathcal{G}$. We denote the set of parents, children, ancestors, and descendants of X by $Pa_{\mathcal{G}}(X)$, $Ch_{\mathcal{G}}(X)$, $Anc_{\mathcal{G}}(X), De_{\mathcal{G}}(X)$ respectively.

[2]It contains no directed cycle.

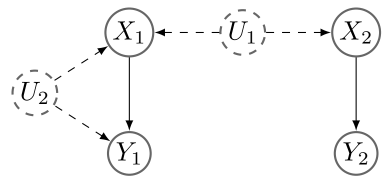

Figure 1: A semi-Markovian DAG over the set of observed variable $\mathbf{V} = \{X_1, X_2, Y_1, Y_2\}$ and the set of hidden variables $\mathbf{U} = \{U_1, U_2\}$.

We assume that X belongs to all the aforementioned sets. Additionally, for a subset of nodes $\mathbf{X}$, we define $Pa_{\mathcal{G}}(\mathbf{X}) := \bigcup_{X \in \mathbf{X}} Pa_{\mathcal{G}}(X)$ and analogously, define $Ch_{\mathcal{G}}(\mathbf{X})$, $Anc_{\mathcal{G}}(\mathbf{X})$ and $De_{\mathcal{G}}(\mathbf{X})$.

A causal graph $\mathcal{G}$ is called a semi-Markovian, if any node from $\mathbf{U}$ has exactly two children without any parents. Suppose that $\mathcal{G}$ is a semi-Morkovain graph and $\mathbf{X} \subseteq \mathbf{V}$. In this case, we use $\mathcal{G}[\mathbf{X}]$ to denote the induced subgraph of $\mathcal{G}$ over variables in $\mathbf{X}$ including all unobserved variables that have both children in $\mathbf{X}$. We also use $\widehat{\mathcal{G}}[\mathbf{X}]$ to denote the dual graph of $\mathcal{G}[\mathbf{X}]$ that is a mixed[3] graph and it is constructed from $\mathcal{G}[\mathbf{X}]$ by replacing unobserved variables and their outgoing arrows with bidirected edges. By the abuse of notation, we use $\mathcal{G}[\mathbf{X}]$ and $\widehat{\mathcal{G}}[\mathbf{X}]$ interchangeably.

Definition 1 (c-component)**.** *C-components of a subset* $\mathbf{X}$ *in* $\mathcal{G}$ *are the connected components in* $\widehat{\mathcal{G}}[\mathbf{X}]$ *after removing all directed edges, i.e., nodes in each c-component are connected via bidirected edges.* $\mathbf{X}$ *is called a single c-component if* $\mathbf{X}$ *has only one c-component, i.e.,* $\widehat{\mathcal{G}}[\mathbf{X}]$ *is a connected graph after removing all directed edges.*

For instance in Figure 1, c-components of $\{X_1, X_2, Y_2\}$ are $\{X_1, X_2\}$ and $\{Y_2\}$. In this DAG, $\{X_1, X_2, Y_1\}$ and $\{Y_2\}$ are each single c-components.

Definition 2 (c-forest)**.** *Let* $\mathcal{F}$ *be a subgraph of* $\mathcal{G}$ *over a set of nodes* $\mathbf{X}$*. The maximal subset of* $\mathbf{X}$ *with no children in* $\mathcal{F}$ *is called the root set and denoted by* $\mathbf{R}$*.* $\mathcal{F}$ *is a* $\mathbf{R}$*-rooted c-forest if* $\mathbf{X}$ *is a single c-component with root set* $\mathbf{R}$*, and all observable nodes in* $\mathbf{X}$ *have at most one child in* $\mathcal{F}$*.*

In Figure 1, $\mathcal{G}[\{X_1, X_2, Y_1\}]$ is a $\{Y_1, X_2\}$-rooted c-forest.

Causal Model: A causal model $\mathcal{M}$ is defined over a set of random variables $\mathbf{V} \cup \mathbf{U}$ via Structural Equation Model (SEM) [Pearl, 2009] with a causal graph $\mathcal{G}$. In a SEM with a causal graph $\mathcal{G}$, each variable $X \in \mathbf{V} \cup \mathbf{U}$ is determined by its parents and an exogenous variable ϵ_X, i.e. $X = f_X(Pa_{\mathcal{G}}(X), \epsilon_X)$. It is assumed that the set of exogenous variables, $\{\epsilon_X | X \in \mathbf{V} \cup \mathbf{U}\}$, are mutually independent. If graph $\mathcal{G}$ is a semi-Markovian, then $\mathcal{M}$ is said to be a semi-Markovian causal model. Because, the problem

[3]It contains both directed and bidirected edges.

of identifiability in a DAG is equivalent to a relative identifiability problem in a semi-Markovian DAG [Huang and Valtorta, 2006], in this work, we only consider the problem of identifiability in semi-Markovian models.

In a semi-Markovian causal model, by Markov property [Pearl, 2009], the induced joint distribution can be factorized as follows

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

where the summation is over latent variables in $\mathbf{U}$. We use $\mathbb{M}(\mathcal{G})$ to denote the set of causal models with graph $\mathcal{G}$ such that for any $\mathcal{M} \in \mathbb{M}(\mathcal{G})$ and any realization $\mathbf{v} \in \mathfrak{X}(\mathbf{V})$, $P^{\mathcal{M}}(\mathbf{v}) > 0$. In the remainder of this work, we assume that all causal models belong to $\mathbb{M}$. This is known as the positivity assumption in the causality literature and as it is discussed in Kivva et al. [2022], it is crucial for developing sound identification algorithm.

In a causal model $\mathcal{M}$, post-interventional distribution is defined using do-operation. An intervention $do(X = x)$ modifies the corresponding SEM by replacing the equation of $X = f_X(Pa_{\mathcal{G}}(X), \epsilon_X)$ by $X = x$. The conditional post-interventional distribution of $\mathbf{y}$ given $\mathbf{s}$ after intervening on $do(X = x)$ is denoted by $P(\mathbf{y}|do(\mathbf{X} = \mathbf{x}), \mathbf{s}) := P_{\mathbf{x}}(\mathbf{y}|\mathbf{s})$.

Suppose that $\mathbf{S} = \mathbf{S}' \cup \mathbf{S}''$, where $\mathbf{S}'$ and $\mathbf{S}''$ are two disjoint subsets of observed variables $\mathbf{V}$. We define Q-notations, $Q[\mathbf{S}](\cdot)$ and $Q[\mathbf{S}'|\mathbf{S}''](\cdot)$ as follows:

$$Q[\mathbf{S}](\mathbf{v}) := P_{\mathbf{v} \backslash \mathbf{s}}(\mathbf{s}), \tag{1}$$

$$Q[\mathbf{S}'|\mathbf{S}''](\mathbf{v}) := P_{\mathbf{v} \backslash \mathbf{s}}(\mathbf{s}'|\mathbf{s}''), \tag{2}$$

where $\mathbf{s} = \mathbf{v}[\mathbf{S}], \mathbf{v} \backslash \mathbf{s} = \mathbf{v}[\mathbf{V} \backslash \mathbf{S}], \mathbf{s}' = \mathbf{v}[\mathbf{S}']$, and $\mathbf{s}'' = \mathbf{v}[\mathbf{S}'']$. Note that $Q[\mathbf{V}](\mathbf{v}) = P(\mathbf{V} = \mathbf{v})$. By Markov property and basic probabilistic manipulation, we have

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

$$Q[\mathbf{S}'|\mathbf{S}''](\mathbf{v}) = \frac{Q[\mathbf{S}](\mathbf{v})}{\sum_{\mathbf{v}' \in \mathfrak{X}_{\mathbf{v} \backslash \mathbf{s}'}(\mathbf{V})} Q[\mathbf{S}](\mathbf{v}')} \tag{4}$$

Definition 3 (Blocked path)**.** *A path in $\mathcal{G}$ is a non-repeated sequence of connected nodes. A path p in $\mathcal{G}$ is said to be blocked by a set of nodes in $\mathbf{Z}$ if and only if*

- *p contains a chain $X \to W \to Y$ or fork $X \leftarrow W \to Y$, such that $W \in \mathbf{Z}$, or*
- *p contains a collider $X \to W \leftarrow Y$ (node W is called a collider), such that $\mathbf{Z} \cap De_{\mathcal{G}}(W) = \varnothing$.*

Two disjoint sets of nodes $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in $\mathcal{G}$ if any path between $\mathbf{X}$ and $\mathbf{Y}$ are blocked by $\mathbf{Z}$ and denote it by $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_{\mathcal{G}}$. Using d-separation, we introduce rules of do-calculus [Pearl, 2009] as the main tools for causal effect identification.

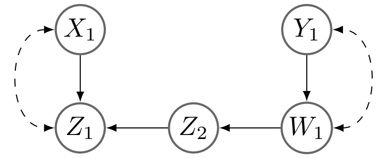

Figure 2: A semi-Markovian DAG over $\mathbf{V} = \{X_1, Y_1, Z_1, Z_2, W_1\}$

- Rule 1: $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}, \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y}|\mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}$.
- Rule 2: $P_{\mathbf{x},\mathbf{z}}(\mathbf{y}|\mathbf{w}) = P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}, \mathbf{w})$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y}|\mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}},\underline{\mathbf{Z}}}}$.
- Rule 3: $P_{\mathbf{x},\mathbf{z}}(\mathbf{y}|\mathbf{w}) = P_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y}|\mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}},\overline{\mathbf{Z}_W}}}$,

where $\mathcal{G}_{\overline{\mathbf{X}},\underline{\mathbf{Y}}}$ denotes an edge subgraph of $\mathcal{G}$ where all incoming arrows into $\mathbf{X}$ and all outgoing arrows from $\mathbf{Y}$ are deleted and $\mathbf{Z}_W := \mathbf{Z} \backslash Anc_{\mathcal{G}_{\overline{\mathbf{X}}}}(\mathbf{W})$.

2.2 CLASSICAL IDENTIFIABILITY (ID)

Classical identifiability problem refers to computing a causal effect $P_{\mathbf{x}}(\mathbf{y})$ from a given joint distribution $P(\mathbf{V})$ in a causal graph $\mathcal{G}$. This problem was solved independently by Shpitser and Pearl [2006a] and Huang and Valtorta [2006]. Shpitser and Pearl [2006b] extended this result to identifiability of a conditional causal effect, i.e., $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$.

Definition 4 (conditional ID)**.** *Suppose $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are three disjoint subsets of $\mathbf{V}$. The causal effect $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ is said to be conditional ID in $\mathcal{G}$ if for any $\mathbf{x} \in \mathfrak{X}(\mathbf{X})$, $\mathbf{y} \in \mathfrak{X}(\mathbf{Y})$, and $\mathbf{z} \in \mathfrak{X}(\mathbf{Z})$, $P^{\mathcal{M}}_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ is uniquely computable from $P^{\mathcal{M}}(\mathbf{V})$ in any causal model $\mathcal{M} \in \mathbb{M}(\mathcal{G})$.*

Knowing $P_{\mathbf{x}}(\mathbf{y}, \mathbf{z})$, it is straightforward to uniquely compute $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ from $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x}}(\mathbf{y}, \mathbf{z}) / \sum_{\mathbf{Y}'} P_{\mathbf{x}}(\mathbf{y}', \mathbf{z})$. On the other hand, Tian [2004] showed that $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ might be identifiable in $\mathcal{G}$ even if $P_{\mathbf{x}}(\mathbf{y}, \mathbf{z})$ is not identifiable. This happens when the "non-identifiable parts" of $P_{\mathbf{x}}(\mathbf{y}, \mathbf{z})$ in the nominator cancel out with the non-identifiable parts of $P_{\mathbf{x}}(\mathbf{z})$ in the denominator. Next example demonstrates such a scenario.

Example: Consider the causal graph $\mathcal{G}$ as on Figure 2. Assume that one wants to compute the causal effect $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$, where $\mathbf{X} = \{X_1\}$, $\mathbf{Y} = \{Y_1\}$ and $\mathbf{Z} = \{Z_1, Z_2\}$. Then,

$$P_{x_1}(y_1|z_1, z_2) = \frac{P_{x_1}(y_1, z_1, z_2)}{P_{x_1}(z_1, z_2)}.$$

where

$$P_{x_1}(y_1, z_1, z_2) = \sum_{w_1 \in \mathfrak{X}(W_1)} P_{x_1}(y_1, w_1, z_1, z_2),$$

Table 1: Different types of identifiability problems.

Problem	Target	Input	Solved
Causal effect identifiability (ID) Shpitser and Pearl [2006a] Huang and Valtorta [2006]	$P_{\mathbf{x}}(\mathbf{y})$	$\mathcal{G}, P(\mathbf{V})$	✓
Conditional causal effect identifiability (c-ID) Shpitser and Pearl [2006b]	$P_{\mathbf{x}}(\mathbf{y}\vert\mathbf{z})$	$\mathcal{G}, P(\mathbf{V})$	✓
z-identifiability (zID) Bareinboim and Pearl [2012]	$P_{\mathbf{x}}(\mathbf{y})$	$\mathcal{G}, P(\mathbf{V}), \{P_{\mathbf{V}\backslash\mathbf{A}'}(\mathbf{A}')\vert\forall\mathbf{A}' \subset \mathbf{A}\}$	✓
g-identifiability (gID) Lee et al. [2019], Kivva et al. [2022]	$P_{\mathbf{x}}(\mathbf{y})$	$\mathcal{G}, \{P(\mathbf{A}_i\vert do(\mathbf{V}\backslash\mathbf{A}))\}_{i=0}^m$	✓
Conditional general identifiability (c-gID) Lee et al. [2020], Correa et al. [2021]	$P_{\mathbf{x}}(\mathbf{y}\vert\mathbf{z})$	$\mathcal{G}, \{P(\mathbf{A}_i\vert do(\mathbf{V}\backslash\mathbf{A}))\}_{i=0}^m$	✓our work
Generalized identifiability	$P_{\mathbf{x}}(\mathbf{y}\vert\mathbf{z})$	$\mathcal{G}, \{P(\mathbf{A}_i\vert do(\mathbf{B}_i), \mathbf{C}_i)\}_{i=0}^m$	?

and

$$P_{x_1}(z_1, z_2) = \sum_{\substack{w_1 \in \mathfrak{X}(W_1) \\ y_1 \in \mathfrak{X}(Y_1)}} P_{x_1}(y_1, w_1, z_1, z_2).$$

In terms of Q-notation, we have

$$\begin{aligned} P_{x_1}(y_1, z_1, z_2) &= \sum_{W_1} Q[Y_1, W_1, Z_1, Z_2], \\ P_{x_1}(z_1, z_2) &= \sum_{W_1, Y_1} Q[Y_1, W_1, Z_1, Z_2]. \end{aligned}$$

Using results of Huang and Valtorta [2006], the above equations can be simplified as follows,

$$\begin{aligned} P_{x_1}(y_1, z_1, z_2) &= Q[Z_1] \sum_{W_1} Q[Y_1, W_1, Z_2], \\ P_{x_1}(z_1, z_2) &= Q[Z_1] \sum_{W_1, Y_1} Q[Y_1, W_1, Z_2]. \end{aligned}$$

Results of Huang and Valtorta [2006], Shpitser and Pearl [2006a] imply that $Q[Z_1]$ is not ID from $\mathcal{G}$, however $Q[Y_1, W_1, Z_2]$ is ID in $\mathcal{G}$. Therefore, both causal effects $P_{\mathbf{x}}(\mathbf{y}, \mathbf{z})$ and $P_{\mathbf{x}}(\mathbf{z})$ are not ID in $\mathcal{G}$, but clearly

$$P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = \frac{\sum_{W_1} Q[Y_1, W_1, Z_2]}{\sum_{W_1, Y_1} Q[Y_1, W_1, Z_2]}$$

is identifiable in $\mathcal{G}$.

2.3 GENERALIZED IDENTIFIABILITY (GID)

In this problem, the goal is to identify a causal effect in a given graph $\mathcal{G}$ from a set of observational and/or interventional distributions instead of only observational distribution $P(\mathbf{V})$. This problem, to the best of our knowledge, remains open when the set of given distributions are arbitrary. In the special case, when the set of given distributions are in the form of Q-notations, the problem is called generalized identifiability (gID) (See below for a formal definition) and was solved by [Lee et al., 2019, Kivva et al., 2022]. See Table 1 for a summary of solved and unsolved problems in the causal identifiability context.

Definition 5 (gID)**.** *Suppose $\mathbf{X}$ and $\mathbf{Y}$ are two disjoint subsets of $\mathbf{V}$ and $\mathbb{A} := \{\mathbf{A}_i\}_{i=0}^m$ is a collection of subsets of $\mathbf{V}$, i.e., $\mathbf{A}_i \subseteq \mathbf{V}$ for all $i \in [0:m]$. The causal effect $P_{\mathbf{x}}(\mathbf{y})$ is said to be gID from $(\mathbb{A}, \mathcal{G})$ if for any $\mathbf{x} \in \mathfrak{X}(\mathbf{X})$ and $\mathbf{y} \in \mathfrak{X}(\mathbf{Y})$ if $P_{\mathbf{x}}^{\mathcal{M}}(\mathbf{y})$ is uniquely computable from $\{Q^{\mathcal{M}}[\mathbf{A}_i]\}_{i=0}^m$ in any causal model $\mathcal{M} \in \mathbb{M}(\mathcal{G})$.*

Note that the classical ID problem is a special case of the gID problem when $\mathbb{A} = \{\mathbf{V}\}$. More than a decade after Shpitser and Pearl [2006b] proposed a sound and complete algorithm for ID, Kivva et al. [2022] solved the gID problem by showing that gID problem can be reduced to a series of separated ID problems. Formally, they showed the following result.

Theorem 1 (Kivva et al. [2022])**.** *Suppose that $\mathbf{S} \subseteq \mathbf{V}$ is a single c-component in $\mathcal{G}$. Then, $Q[\mathbf{S}]$ is gID 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 $\mathcal{G}[\mathbf{A}]$.*

2.4 CONDITIONAL GENERALIZED IDENTIFIABILITY (C-GID)

In this work, we address an extension of both conditional ID and g-ID problem in which the goal is to identify a conditional causal effect from a set of observational and/or interventional distributions.

Definition 6. *(c-gID) Suppose $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are three disjoint subsets of $\mathbf{V}$ and $\mathbb{A} := \{\mathbf{A}_i\}_{i=0}^m$ is a collection of subsets of $\mathbf{V}$, i.e., $\mathbf{A}_i \subseteq \mathbf{V}$ for all $i \in [0:m]$. The causal* *effect* $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ *is said to be c-gID from* $(\mathbb{A}, \mathcal{G})$ *if for any* $\mathbf{x} \in \mathfrak{X}(\mathbf{X})$*,* $\mathbf{y} \in \mathfrak{X}(\mathbf{Y})$*, and* $\mathbf{z} \in \mathfrak{X}(\mathbf{Z})$*,* $P^{\mathcal{M}}_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ *is uniquely computable from* $\{Q^{\mathcal{M}}[\mathbf{A}_i]\}_{i=0}^{m}$ *in any causal model* $\mathcal{M} \in \mathbb{M}(\mathcal{G})$*.*

From this definition, it is clear that c-gID covers both conditional ID and gID. Namely, when $\mathbf{Z} = \varnothing$, then c-gID reduces to the gID problem, studied by Lee et al. [2019], Kivva et al. [2022]. When $\mathbb{A} = \{\mathbf{V}\}$, c-gID becomes the conditional ID problem studied by Shpitser and Pearl [2006b]. Both Lee et al. [2020] and Correa et al. [2021] proposed algorithms for identification problems that can also be used for solving c-gID problem. However, the completeness of their algorithms rely on causal models that violate the positivity assumption. For more details see Appendix B. Additionally, they miss discussions on whether this issue in their proofs can be resolved.

Next we propose an alternative solution for the c-gID problem under the positivity assumption. The soundness and completeness of our solution are based on novel techniques that we believe they are important for further generalizations of identifiability problems.

3 MAIN RESULT

The main idea presented in this work for solving the c-gID problem is to construct an equivalent gID problem and then use the results of [Lee et al., 2019, Kivva et al., 2022] to solve the equivalent gID problem.

Suppose $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ are three disjoint subsets of $\mathbf{V}$ and $\mathbb{A}$ is a collection of subsets of $\mathbf{V}$. We are interested in identifying $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ from $(\mathcal{G}, \mathbb{A})$. To this end, we define $\mathbf{W}$ to be the maximal subset of $\mathbf{Z}$, such that $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x},\mathbf{w}}(\mathbf{y}|\mathbf{z}\backslash\mathbf{w})$. Shpitser and Pearl [2006b] proved that such a maximal set is unique and it is given by

$$\mathbf{W} = \bigcup_{W' \in \mathbf{Z}} \{W' |\, P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x},w'}(\mathbf{y}|\mathbf{z}\backslash\{w'\})\}. \quad (5)$$

More precisely, they showed the following result.

Theorem 2 (Shpitser and Pearl [2006b])**.** *For a given graph* $\mathcal{G}$ *and any conditional effect* $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$*, there exists a unique maximal set* $\mathbf{W} = \{W \in \mathbf{Z} | P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x},w}(\mathbf{y}|\mathbf{z}\backslash\{w\})\}$ *such that rule 2 of do-calculus applies to* $\mathbf{W}$ *in* $\mathcal{G}$ *for* $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$*.*

In a special case when $\mathbf{W} = \mathbf{Z}$, it is trivial that the equivalent gID problem boils down to identifying $P_{\mathbf{x},\mathbf{z}}(\mathbf{y})$ from $(\mathcal{G}, \mathbb{A})$. In the next result, we present the form of an equivalent gID problem for a general c-gID problem.

Theorem 3. *Let* $\mathbf{W}$ *be the maximal subset of* $\mathbf{Z}$*, such that* $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x},\mathbf{w}}(\mathbf{y}|\mathbf{z}\backslash\mathbf{w})$*. Then,* $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ *is c-gID from* $(\mathbb{A}, \mathcal{G})$ *if and only if* $P_{\mathbf{x},\mathbf{w}}(\mathbf{y}, \mathbf{z}\backslash\mathbf{w})$ *is gID from* $(\mathbb{A}, \mathcal{G})$*.*

Algorithm 1 c-gID

Function C-GID$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbb{A} = \{\mathbf{A}_i\}_{i=0}^{m}, \mathcal{G})$
Output: True, if $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ is c-gID from $(\mathbb{A}, \mathcal{G})$.
$\mathbf{W} \leftarrow$ **MaxBI**$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G})$
Return GID$(\mathbf{X} \cup \mathbf{W}, \mathbf{Y} \cup (\mathbf{Z}\backslash\mathbf{W}), \mathbb{A}, \mathcal{G})$

Function MaxBI$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G})$
Output: set $\mathbf{W}$
$\mathbf{W} \leftarrow \varnothing$
for Z in $\mathbf{Z}$ **do**
    if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}, \mathbf{Z}\backslash\{Z\})_{\mathcal{G}_{\overline{\mathbf{X}},\underline{\mathbf{Z}\backslash\{Z\}}}}$ **then**
        $\mathbf{W} \leftarrow \mathbf{W} \cup \{Z\}$
    end if
end for
Return W

A sketch of the proof of this Theorem appears in Section 4. This result extends the result of Shpitser and Pearl [2006b] for conditional ID to c-gID. Furthermore, Theorem 3 allows us to develop an algorithm for solving the c-gID problem. Algorithm 1 summarizes the steps of the proposed algorithm. The algorithm consists of two main steps:
1. Find the maximal set $\mathbf{W} \subseteq \mathbf{Z}$ in $\mathcal{G}$, such that $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x},\mathbf{w}}(\mathbf{y}|\mathbf{z}\backslash\mathbf{w})$. For this part, we propose function **MaxBI** presented in Algorithm 1 that is based on Equation (5).
2. Run any sound and complete gID algorithm (e.g., the proposed algorithm by Kivva et al. [2022]) for checking the gID of $P_{\mathbf{x},\mathbf{w}}(\mathbf{y}, \mathbf{z}\backslash\mathbf{w})$ from $(\mathbb{A}, \mathcal{G})$.

Theorem 4. *Algorithm 1 is sound and complete.*

Proof. The result immediately follows from Theorem 3 since the gID algorithm is sound and complete. □

Corollary 1. *Rules of do-calculus are sound and complete for the c-gID problems.*

Remark 1. *Algorithm 1 is polynomial time in the input size.*

In subroutine **MaxBI**, a conditional independence test is performed for each variable in $\mathbf{Z}$. Subsequently, the problem is reduced to the gID problem, which can be solved in polynomial number of steps by using any of the algorithms proposed in Lee et al. [2019], Kivva et al. [2022].

4 PROOF OF THE THEOREM 3

In this section, we present the main steps of proof of Theorem 3. Further details can be found in Appendix A. Before going into the details and purely for simpler representation, we define the following notations, $\mathbf{X}' := \mathbf{X} \cup \mathbf{W}$, $\mathbf{Y}' := \mathbf{Y}$, and $\mathbf{Z}' := \mathbf{Z}\backslash\mathbf{W}$. Note that by the definition of $\mathbf{W}$ and Theorem 2, we have $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$.

The proof consists of two main parts: sufficiency and necessity. In the sufficiency part, which is more straightforward, we show that if $P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}')$ is gID from $(\mathbb{A},\mathcal{G})$, then $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ is c-gID. For the reverse, which is much more involved, we use a proof by contradiction. That is we show if $P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}')$ is not gID from $(\mathbb{A},\mathcal{G})$, then $P_x(\mathbf{y}|\mathbf{z})$ is also not c-gID.

Sufficiency: Suppose $P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}')$ is gID from $(\mathbb{A},\mathcal{G})$, then the result follows immediately from the Bayes rule and the fact that $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}') = P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$, i.e.,

$$P_x(\mathbf{y}|\mathbf{z}) = \frac{P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}')}{\sum_{\mathbf{y}''\in\mathfrak{X}(\mathbf{Y})} P_{\mathbf{x}'}(\mathbf{y}'',\mathbf{z}')}. \tag{6}$$

Necessity: Suppose that $P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}')$ is not gID from $(\mathbb{A},\mathcal{G})$. To show the non-identifiability of $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$ from $(\mathbb{A},\mathcal{G})$, we construct two causal models $\mathcal{M}_1$ and $\mathcal{M}_2$ from $\mathbb{M}(\mathcal{G})$, such that for each $i \in [0:m]$ and any $\mathbf{v} \in \mathfrak{X}(\mathbf{V})$,

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

but there exists a triple $(\mathbf{x}',\mathbf{y}',\mathbf{z}') \in \mathfrak{X}(\mathbf{X}') \times \mathfrak{X}(\mathbf{Y}') \times \mathfrak{X}(\mathbf{Z}')$, such that $P^{\mathcal{M}_1}_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}') \neq P^{\mathcal{M}_2}_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$.

Huang and Valtorta [2006] showed that that $P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}')$ can be written as follows

$$P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}') = \sum_{\mathbf{S}\setminus(\mathbf{Y}'\cup\mathbf{Z}')} Q[\mathbf{S}](\mathbf{v}),$$

where $\mathbf{S} := Anc_{\mathcal{G}[\mathbf{V}\setminus\mathbf{X}']}(\mathbf{Y}'\cup\mathbf{Z}')$ and the marginalization is over all variables in set $\mathbf{S}\setminus(\mathbf{Y}'\cup\mathbf{Z}')$. Suppose that $\mathbf{S}_1,\mathbf{S}_2,\dots,\mathbf{S}_n$ are the c-components of $\mathbf{S}$ in a graph $\mathcal{G}[\mathbf{S}]$. It is known by Huang and Valtorta [2006] that

$$Q[\mathbf{S}](\mathbf{v}) = \prod_{i=1}^{n} Q[\mathbf{S}_i](\mathbf{v}).$$

Since $P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}')$ is not gID from $(\mathbb{A},\mathcal{G})$, using Proposition 4 and Theorem 1 in Kivva et al. [2022], we conclude that there exists $i \in [1:n]$, such that for any $j \in [0:m]$, the causal effect $Q[\mathbf{S}_i]$ is not ID from $\mathcal{G}[\mathbf{A}_j]$.

Analogously, let $\mathbf{S}' := Anc_{\mathcal{G}[\mathbf{V}\setminus\mathbf{X}']}(\mathbf{Z}')$ and assume $\mathbf{S}'_1,\mathbf{S}'_2,\dots,\mathbf{S}'_{n'}$ are the c-components of $\mathbf{S}'$ in graph $\mathcal{G}[\mathbf{S}']$. Then, we have

$$P_{\mathbf{x}'}(\mathbf{z}') = \sum_{\mathbf{S}'\setminus\mathbf{Z}'} \prod_{i=1}^{n'} Q[\mathbf{S}'_i](\mathbf{v}). \tag{7}$$

Consequently, we obtain the following expression

$$P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}') = \frac{\sum_{\mathbf{S}\setminus(\mathbf{Y}'\cup\mathbf{Z}')}\prod_{i=1}^{n} Q[\mathbf{S}_i](\mathbf{v})}{\sum_{\mathbf{S}'\setminus\mathbf{Z}'}\prod_{i=1}^{n'} Q[\mathbf{S}'_i](\mathbf{v})}.$$

Note that $\mathbf{S}' \subseteq \mathbf{S}$ and for any $i \in [1:n]$ and $j \in [1:n']$ either $\mathbf{S}'_j$ and $\mathbf{S}_i$ are disjoint or $\mathbf{S}'_j \subseteq \mathbf{S}_i$.

Depending on the relationships between $\{Q[\mathbf{S}_i]\}_{i=1}^{n}$ and $\{Q[\mathbf{S}'_j]\}_{j=1}^{n'}$ and which parts are gID, in the remainder, we consider two different cases and study each one separately.

4.1 FIRST CASE

In this case, we assume that there exists an index $i \in [1:n]$, such that both $Q[\mathbf{S}_i]$ is not gID from $(\mathbb{A},\mathcal{G})$ and $\mathbf{S}_i \neq \mathbf{S}'_j$ for all $j \in [1:n']$.

If we show that $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$ remain not c-gID even after adding additional knowledge about the distributions $\{Q[\mathbf{S}'_j]\}_{j=1}^{n'}$ to $\{Q[\mathbf{A}_k]\}_{k=0}^{m}$, then, we can conclude that $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$ is also not c-gID from $(\mathbb{A},\mathcal{G})$. To do so, let $\mathbb{A}' := \mathbb{A} \cup (\bigcup_{j=1}^{n'}\{\mathbf{S}'_i\})$.

Clearly, $P_{\mathbf{x}'}(\mathbf{z}')$ is c-gID from $(\mathbb{A}',\mathcal{G})$ as all the terms in (7) are given in $\mathbb{A}'$. On the other hand, $Q[\mathbf{S}_i]$ is not gID from $(\mathbb{A}',\mathcal{G})$. This is due to the assumptions of this setting, that are $Q[\mathbf{S}_i]$ is not gID from $(\mathbb{A},\mathcal{G})$ and $\mathbf{S}_i \not\subseteq \mathbf{S}'_j$ for all $j \in [1:n']$. The latter assumption implies that none of the additional distributions $\{Q[\mathbf{S}'_j]\}_{j=1}^{n'}$ can be used to identify $Q[\mathbf{S}_i]$. Since, we have established that $Q[\mathbf{S}_i]$ and consequently $P_{\mathbf{x}'}(\mathbf{y}',\mathbf{z}')$ are not gID from $(\mathbb{A}',\mathcal{G})$, there exists two models $\mathcal{M}_1,\mathcal{M}_2 \in \mathbb{M}(\mathcal{G})$, such that for any $\mathbf{v} \in \mathfrak{X}(\mathbf{V})$,

$$Q^{\mathcal{M}_1}[\mathbf{A}_j](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}_j](\mathbf{v}), \quad j \in [0:m],$$
$$Q^{\mathcal{M}_1}[\mathbf{S}_{j'}](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{S}_{j'}](\mathbf{v}), \quad j' \in [1:n'],$$

and there exists $(\hat{\mathbf{x}}',\hat{\mathbf{y}}',\hat{\mathbf{z}}') \in \mathfrak{X}(\mathbf{X}')\times\mathfrak{X}(\mathbf{Y}')\times\mathfrak{X}(\mathbf{Z}')$, such that

$$P^{\mathcal{M}_1}_{\hat{\mathbf{x}}'}(\hat{\mathbf{y}}',\hat{\mathbf{z}}') \neq P^{\mathcal{M}_2}_{\hat{\mathbf{x}}'}(\hat{\mathbf{y}}',\hat{\mathbf{z}}').$$

Because $P_{\mathbf{x}'}(\mathbf{z}')$ is gID from $(\mathbb{A}',\mathcal{G})$ and from (6), we have

$$P^{\mathcal{M}_1}_{\widehat{\mathbf{x}'}}(\widehat{\mathbf{y}'}|\widehat{\mathbf{z}'}) \neq P^{\mathcal{M}_2}_{\widehat{\mathbf{x}'}}(\widehat{\mathbf{y}'}|\widehat{\mathbf{z}'}).$$

This implies that $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$ is not c-gID from $(\mathbb{A}',\mathcal{G})$.

4.2 SECOND CASE

Suppose that there is no $i \in [1:n]$, such that both $Q[\mathbf{S}_i]$ is not gID from $(\mathbb{A},\mathcal{G})$ and $\mathbf{S}_i \neq \mathbf{S}'_j$ for all $j \in [1:n']$.

Without loss of generality, suppose that for some $k \leqslant n$, all $Q[\mathbf{S}_1], Q[\mathbf{S}_2],\dots,Q[\mathbf{S}_k]$ are not gID from $(\mathbb{A},\mathcal{G})$ and the remaining $Q[\mathbf{S}_{k+1}],\dots,Q[\mathbf{S}_n]$ are gID from $(\mathbb{A},\mathcal{G})$. By the assumption of this case, for each $i \in [1:k]$, there exists $j_i \in [1:n']$ such that $\mathbf{S}_i = \mathbf{S}'_{j_i}$. Without loss generality, suppose that $j_i = i$ for all $i \in [1:k]$, i.e., $\mathbf{S}_1 = \mathbf{S}'_1$, ..., $\mathbf{S}_k = \mathbf{S}'_k$. Therefore, $\mathbf{S}_i \subset \mathbf{S}' = Anc_{\mathcal{G}[\mathbf{V}\setminus\mathbf{X}']}(\mathbf{Z}')$, for all $i \in [1:k]$.

To establish the result, we further consider three different sub-cases:
1: $\mathbf{Y}'\cap\mathbf{S}_1 \neq \varnothing$, 2: $\mathbf{S}_1 \subseteq \mathbf{Z}'$, and 3: $\mathbf{S}_1\setminus(\mathbf{Z}'\cup\mathbf{Y}') \neq \varnothing$.

Remark 2. *Although, the above sub-cases may have non-empty intersection, it is easy to see that their union covers all possible scenarios of the second case.*

4.2.1 Sub-case 1: $\mathbf{Y}' \cap \mathbf{S}_1 \neq \varnothing$

Let Y denotes a random variable in $\mathbf{Y}' \cap \mathbf{S}_1$. Since Y belongs to $\mathbf{S}_1 = \mathbf{S}'_1$, Y is an ancestor of a variable in $\mathbf{Z}'$ in a graph $\mathcal{G}[\mathbf{V}\backslash\mathbf{X}']$, i.e. $Y \in Anc_{\mathcal{G}[\mathbf{V}\backslash\mathbf{X}']}(\mathbf{Z}') = \mathbf{S}'$. This implies that

$$P_{\mathbf{x}'}(y|\mathbf{z}') = \frac{P_{\mathbf{x}'}(y,\mathbf{z}')}{P_{\mathbf{x}'}(\mathbf{z}')} = \frac{\sum_{\mathbf{S}'\backslash(\mathbf{Z}'\cup\{Y\})}\prod_{i=1}^{n'} Q[\mathbf{S}'_i](\mathbf{v})}{\sum_{\mathbf{S}'\backslash\mathbf{Z}'}\prod_{i=1}^{n'} Q[\mathbf{S}'_i](\mathbf{v})} \quad (8)$$

We prove this sub-case by showing that $P_{\mathbf{x}'}(y|\mathbf{z}')$ is not c-gID from $(\mathbb{A},\mathcal{G})$ and subsequently $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$ is not c-gID from $(\mathbb{A},\mathcal{G})$. To this end, first, we prove **I:** $Q[\{Y\}|\mathbf{S}'_1\backslash\{Y\}]$ is not c-gID from $(\mathbb{A},\mathcal{G}_{\underline{\{Y\}}})$, and then use it to show **II:** $Q[\{Y\}|\mathbf{S}'_1\backslash\{Y\}]$ is not c-gID from $(\mathbb{A},\mathcal{G})$. Finally, we show **III:** $P_{\mathbf{x}'}(y|\mathbf{z}')$ is not c-gID from $(\mathbb{A},\mathcal{G})$.

I: In graph $\mathcal{G}_{\underline{\{Y\}}}$ and using (8), we obtain

$$\begin{aligned} Q\big[\{Y\}|\mathbf{S}'\backslash\{Y\}\big] &= \frac{\prod_{i=1}^{n'} Q[\mathbf{S}'_i]}{\sum_Y \prod_{i=1}^{n'} Q[\mathbf{S}'_i]} \\ &= \frac{Q[\mathbf{S}_1]}{\sum_Y Q[\mathbf{S}_1]} = Q\big[\{Y\}|\mathbf{S}_1\backslash\{Y\}\big]. \end{aligned}$$

Recall that $\mathbf{S}_1 = \mathbf{S}'_1$. Next result shows that $Q[\{Y\}|\mathbf{S}_1\backslash\{Y\}]$ is not c-gID from $(\mathbb{A},\mathcal{G}_{\underline{\{Y\}}})$ because $Q[\mathbf{S}_1]$ is not gID from $(\mathbb{A},\mathcal{G}_{\underline{\{Y\}}})$. A proof is presented in Appendix A.

Lemma 1. *Suppose $\mathbf{L} \subseteq \mathbf{V}$ is a single c-component, such that $\mathbf{L} = \mathbf{L}' \cup \mathbf{L}''$ for some disjoint sets $\mathbf{L}'$ and $\mathbf{L}''$. $Q[\mathbf{L}'|\mathbf{L}'']$ is c-gID from $(\mathbb{A},\mathcal{G})$ if and only if $Q[\mathbf{L}' \cup \mathbf{L}'']$ is gID from $(\mathbb{A},\mathcal{G})$.*

II: Shpitser and Pearl [2006a] showed the following result for a non-identifiable causal effect.

Lemma 2 (Shpitser and Pearl [2006a])**.** *Suppose $\mathbf{L} \subseteq \mathbf{A} \subseteq \mathbf{V}$. $Q[\mathbf{L}]$ is not identifiable from $\mathcal{G}[\mathbf{A}]$ if and only if there exists at least one $\mathbf{L}$-rooted c-forest $\mathcal{F}$ with the set of observed variables $\mathbf{B}$ such that $\mathbf{L} \subsetneq \mathbf{B} \subseteq \mathbf{A}$, the bidirected edges of $\widehat{\mathcal{F}}[\mathbf{B}]$ form a spanning tree, and $\widehat{\mathcal{F}}[\mathbf{L}]$ is a connected graph with respect to the bidirected edges.*

On the other hand, because $Q[\mathbf{S}_1]$ is not gID from $(\mathbb{A},\mathcal{G})$, by the results of Kivva et al. [2022], $Q[\mathbf{S}_1]$ is not ID from $\mathcal{G}[\mathbf{A}_i]$ for all $i \in [0:m]$. Lemma 2 implies that adding or removing outgoing edges from $Y \in \mathbf{S}_1$ will not affect the non-identifiability of $Q[\mathbf{S}_1]$ from $\mathcal{G}[\mathbf{A}_i]$ for all $i \in [0:m]$. Thus, we have $Q[\mathbf{S}_1]$ is not gID from $(\mathbb{A},\mathcal{G}_{\underline{\{Y\}}})$. This means that exists two causal models $\mathcal{M}_1$ and $\mathcal{M}_2$ from $\mathbb{M}(\mathcal{G}_{\underline{\{Y\}}})$ which are consistent with all known distributions but disagree on the causal effect $Q[\mathbf{S}_1]$, i.e., there exists $\tilde{\mathbf{v}} \in \mathfrak{X}(\mathbf{V})$ such that

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

Note that $\mathbb{M}(\mathcal{G}_{\underline{\{Y\}}}) \subset \mathbb{M}(\mathcal{G})$ which in combination with the above result yield that $Q[\{Y\}|\mathbf{S}\backslash\{Y\}]$ is not c-gID from $(\mathbb{A},\mathcal{G})$.

III: To prove this part, we first present the following result. A proof is provided in Appendix A.

Lemma 3. *Suppose that $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are disjoint subsets of $\mathbf{V}$ in graph $\mathcal{G}$ and variables $Z_1 \in \mathbf{Z}$, $Z_2 \in \mathbf{Y} \cup \mathbf{Z}$, such that there is a directed edge from Z_1 to Z_2 in $\mathcal{G}$. If the causal effect $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ is not c-gID from $(\mathbb{A},\mathcal{G})$, then the causal effect $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}\backslash\{z_1\})$ is also not c-gID from $(\mathbb{A},\mathcal{G})$.*

Note that $P_{\mathbf{x}'}(\mathbf{s}') = Q[\mathbf{S}']$ since $\mathbf{S}' = Anc_{\mathcal{G}[\mathbf{V}\backslash\mathbf{X}']}(\mathbf{S}')$. Therefore, by the definition of Q-notation, we have

$$Q\big[\{Y\}|\mathbf{S}'\backslash\{Y\}\big] = P_{\mathbf{x}'}(y|\mathbf{s}'\backslash\{y\}),$$

which is shown to be not c-gID from $(\mathbb{A},\mathcal{G})$ in part **II**. In the remainder of this part of our proof, we introduce a set of nodes in $\mathbf{S}'$ that satisfy the condition in Lemma 3 and thus, can be eliminated without affecting the non-identifiability. Bellow, we show that the nodes in $\mathbf{S}'\backslash(\mathbf{Z}' \cup \{Y\})$ satisfy Lemma 3's condition and by deleting them, we conclude that $P_{\mathbf{x}'}(y|\mathbf{z}')$ is not c-gID from $(\mathbb{A},\mathcal{G})$.

Recall that $\mathbf{S}' = Anc_{\mathcal{G}[\mathbf{V}\backslash\mathbf{X}']}(\mathbf{Z}')$ which means that from any node in $\mathbf{S}'\backslash(\mathbf{Z}' \cup \{Y\})$, there exists a directed path to a node in $\mathbf{Z}'$ in graph $\mathcal{G}[\mathbf{V}\backslash\mathbf{X}']$. We assign a real number to each node in $\mathbf{S}'\backslash(\mathbf{Z}' \cup \{Y\})$, namely, the length of its shortest path to set $\mathbf{Z}$. Let $(W_1, W_2, \ldots, W_\eta)$ denote the nodes in $\mathbf{S}'\backslash(\mathbf{Z}' \cup \{Y\})$ sorted in a descending order using their assigned numbers. Observe that for any $i \in [1:\eta]$, there is a direct edge from W_i to a node in $\{Y\} \cup \mathbf{Z}' \bigcup_{j=i+1}^{\eta}\{W_j\}$. In other words, Lemma 3 allows us to delete W_i from $\mathbf{S}'\backslash\big(\{Y\} \bigcup_{j=1}^{i-1}\{W_j\}\big)$ without violating the non-identifiability.

4.2.2 Sub-case 2: $\mathbf{S}_1 \subseteq \mathbf{Z}'$

In this sub-case, we prove non-identifiability of $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$ from $(\mathbb{A},\mathcal{G})$ in two steps: **I:** we introduce a conditional causal effect that is not c-gID from $(\mathbb{A},\mathcal{G})$. **II:** Analogous to the previous sub-case, we apply Lemma 3 to prune this causal effect and conclude the result.

I: Let Z' be a node in $\mathbf{S}_1$. Recall that $\mathbf{W}$ is the maximal set such that $P_{\mathbf{x},\mathbf{w}}(\mathbf{y}|\mathbf{z}\backslash\mathbf{w}) = P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$, which means that we can not apply the second rule of do-calculus to Z' in $\mathcal{G}$ for $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$, i.e.,

$$(\mathbf{Y}' \not\perp\!\!\!\perp Z'|\mathbf{X}', \mathbf{Z}'\backslash\{Z'\})_{\mathcal{G}_{\overline{\mathbf{X}'},\underline{\{Z'\}}}}.$$

This implies that there exists at least a unblocked backdoor path from Z' to $\mathbf{Y}'$ given $\mathbf{X}' \cup \mathbf{Z}'\backslash\{Z'\}$. We use p to denote an unblocked path from Z' to $\mathbf{Y}'$ with the least number of colliders. Path p satisfies the following properties:

1. If path p contains a chain $W' \to W \to W''$ or a fork $W' \leftarrow W \to W''$, then node W does not belong to any of the sets $\mathbf{X}'$, $\mathbf{Z}'$ or $\mathbf{Y}'$.
2. If path p contains a collider $W' \to W \leftarrow W''$, then there is a directed path p_W from W to a node in $\mathbf{Z}'$. Moreover, none of the intermediate nodes in the path p_W belong to the set $\mathbf{X}' \cup \mathbf{Z}' \cup \mathbf{Y}'$.
3. Path p does not contain any node from the set $\mathbf{X}'$.

Proofs of the above statements are provided in Appendix A. Suppose $\mathbf{F}$ is a set of all colliders on the path p. We use $\mathcal{P}$ to denote a collection of paths $\{p\} \cup \{p_W | W \in \mathbf{F}\}$ and use $\mathbf{D}$ to denote the set of all nodes on the paths in $\mathcal{P}$ excluding the ones in $\mathbf{Z}'$. Given the above definitions, we are ready to introduce the non-identifiable conditional causal effect in the next result.

Lemma 4. *Let* $\mathbf{S} := Anc_{\mathcal{G}[\mathbf{V}\backslash\mathbf{X}']}(\mathbf{Y}' \cup \mathbf{Z}')$ *and* $\mathbf{D}$ *denote the set defined above. Then,*

$$P_{\mathbf{x}'}(\mathbf{d}|\mathbf{s}\backslash\mathbf{d}) = \frac{Q[\mathbf{S}]}{\sum_{\mathbf{D}} Q[\mathbf{S}]} = Q[\mathbf{D}|\mathbf{S}\backslash\mathbf{D}] \tag{9}$$

is not c-gID from $(\mathbb{A}, \mathcal{G})$.

Proof of this lemma is presented in Appendix A.

II: In order to complete the proof of this part, besides Lemma 3, we require the following technical lemmas.

Lemma 5. *Suppose that* $\mathbf{X}$*,* $\mathbf{Y}$ *and* $\mathbf{Z}$ *are disjoint subsets of* $\mathbf{V}$ *and* $Z \in \mathbf{Z}$*. If the conditional causal effect* $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ *is not c-gID from* $(\mathbb{A}, \mathcal{G})$*, the conditional causal effect* $P_{\mathbf{x}}(\mathbf{y}, z|\mathbf{z}\backslash\{z\})$ *is not c-gID from* $(\mathbb{A}, \mathcal{G})$ *as well.*

Proof. Proof is by contradiction. Suppose that $P_{\mathbf{x}}(\mathbf{y}, z|\mathbf{z}\backslash\{z\})$ is c-gID from $(\mathbb{A}, \mathcal{G})$. This implies that $P_{\mathbf{x}}(z|\mathbf{z}\backslash\{z\})$ is also c-gID from $(\mathbb{A}, \mathcal{G})$. Applying Bayes rule yields

$$P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = \frac{P_{\mathbf{x}}(\mathbf{y}, z|\mathbf{z}\backslash\{z\})}{P_{\mathbf{x}}(z|\mathbf{z}\backslash\{z\})},$$

which results in c-gID of $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ from $(\mathbb{A}, \mathcal{G})$. This contradicts the non-identifiability assumption on $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$. □

Lemma 6. *Suppose that* $\mathbf{X}$*,* $\mathbf{Y}$ *and* $\mathbf{Z}$ *are disjoint subsets of* $\mathbf{V}$ *in graph* $\mathcal{G}$ *and variables* $Y_1 \in \mathbf{Y}$*,* $Y_2 \in \mathbf{Y} \cup \mathbf{Z}$*, such that there is a directed edge from* Y_1 *to* Y_2 *in* $\mathcal{G}$*. If the causal effect* $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ *is not c-gID from* $(\mathbb{A}, \mathcal{G})$*, then the causal effect* $P_{\mathbf{x}}(\mathbf{y}\backslash\{y_1\}|\mathbf{z})$ *is also not c-gID from* $(\mathbb{A}, \mathcal{G})$*.*

Proof of this lemma is presented in Appendix A.

Recall that the goal is to prune the conditional causal effect in (9) to get $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$. We do this in two pruning steps: first using Lemma 5 and then via Lemmas 3, 6. Let $\mathbf{Y}'' := \mathbf{Y}'\backslash\mathbf{D}$. Recall that $\mathbf{S} = Anc_{\mathcal{G}[\mathbf{V}\backslash\mathbf{X}']}(\mathbf{Y}', \mathbf{Z}')$. It is easy to see that $\mathbf{Y}''$ is a subset of $\mathbf{S}\backslash\mathbf{D}$ and thus we can apply Lemma 5 to the causal effect $P_{\mathbf{x}'}(\mathbf{d}|\mathbf{s}\backslash\mathbf{d})$ and conclude that $P_{\mathbf{x}'}(\mathbf{d} \cup \mathbf{y}'|\mathbf{s}\backslash(\mathbf{d} \cup \mathbf{y}'))$ is not c-gID from $(\mathbb{A}, \mathcal{G})$.

To use Lemmas 3, 6 for the second pruning steps, we use similar type of argument as in the first sub-case. More precisely, using the fact that there exists a direct path for each node in $\mathbf{S}\backslash(\mathbf{Z}' \cup \mathbf{Y}')$ to a node in $\mathbf{Z}' \cup \mathbf{Y}'$, we sort the nodes in

$$\mathbf{W}' := \mathbf{S}\backslash(\mathbf{Z}' \cup \mathbf{Y}')$$

in a descending order based on the length of their corresponding shortest direct path to the set $\mathbf{Z}' \cup \mathbf{Y}'$. We denote these sorted nodes by $(W'_1, W'_2, \ldots, W'_{\eta'})$. Note that for any $i \in [1 : \eta']$, there exists a direct edge from W'_i to a node in $\mathbf{Y}' \cup \mathbf{Z}' \cup \{W'_j\}_{j=i+1}^{\eta'}$.

Since $\mathbf{W}'$ is a subset of $\mathbf{S}\backslash(\mathbf{Z}' \cup \mathbf{Y}')$, similar to the second sub-case, we apply Lemmas 3, 6 to the causal effect $P_{\mathbf{x}'}(\mathbf{d} \cup \mathbf{y}'|\mathbf{s}\backslash(\mathbf{d} \cup \mathbf{y}'))$ and remove variables $(W'_1, \ldots, W'_{\eta'})$ one by one from the $P_{\mathbf{x}'}(\mathbf{d} \cup \mathbf{y}'|\mathbf{s}\backslash(\mathbf{d} \cup \mathbf{y}'))$. From definitions of $\mathbf{D}$ and $\mathbf{Z}'$, we have $\mathbf{D} \cap \mathbf{Z}' = \varnothing$, which means

$$\mathbf{S}\backslash(\mathbf{W}' \cup \mathbf{Y}' \cup \mathbf{D}) = \mathbf{Z}'.$$

Therefore, after removing all nodes of $\mathbf{W}'$ from the set $\mathbf{S}\backslash(\mathbf{D} \cup \mathbf{Y}')$ without affecting the non-identifiability of $P_{\mathbf{x}'}(\mathbf{d} \cup \mathbf{y}'|\mathbf{s}\backslash(\mathbf{d} \cup \mathbf{y}'))$, we can claim that $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$ is not c-gID from $(\mathbb{A}, \mathcal{G})$.

4.2.3 Sub-case 3: $\mathbf{S}_1\backslash(\mathbf{Z}' \cup \mathbf{Y}') \neq \varnothing$

The proof of this sub-case is quite similar to the second sub-case with a few twists. Let T be an arbitrary node in $\mathbf{S}_1\backslash(\mathbf{Z}' \cup \mathbf{Y}')$. Since $\mathbf{S}_1$ is a subset of the ancestors of $\mathbf{Z}'$, then there exists a directed path from T to the set $\mathbf{Z}'$. Let p_T denote the shortest directed path from node T to a node Z' in the set $\mathbf{Z}'$. Analogous to the second sub-case, we define $\widetilde{p}$ to be an unblocked backdoor path from Z' to $\mathbf{Y}'$ given $\mathbf{X}', \mathbf{Z}'\backslash\{Z'\}$ with the least number of colliders. Path $\widetilde{p}$ satisfies the following properties:
1. Assume that path $\widetilde{p}$ contains a chain $W' \to W \to W''$ or a fork $W' \leftarrow W \to W''$, then W does not belong to any of the sets $\mathbf{X}'$, $\mathbf{Z}'$ or $\mathbf{Y}'$.
2. Assume that path $\widetilde{p}$ contains an inverted fork $W' \to W \leftarrow W''$, then there is a directed path p_W from the node W to a node in the set $\mathbf{Z}'$. Moreover, none of the intermediate nodes on this path p_W belong to set $\mathbf{X}' \cup \mathbf{Z}' \cup \mathbf{Y}'$.
3. Path $\widetilde{p}$ does not contain any node from the set $\mathbf{X}'$

Proofs of the above statements are provided in Appendix A. Let $\widetilde{\mathbf{F}}$ be the set of all colliders on the path $\widetilde{p}$. Define $\widetilde{\mathcal{P}} := \{\widetilde{p}\} \cup \{p_T\} \cup \{\widetilde{p}_W | W \in \widetilde{\mathbf{F}}\}$ and $\widetilde{\mathbf{D}}$ to be a set containing all nodes on the paths from $\widetilde{\mathcal{P}}$ excluding the nodes in $\mathbf{Z}'$.

Lemma 7. *Let* $\mathbf{S} := Anc_{\mathcal{G}[\mathbf{V}\backslash\mathbf{X}']}(\mathbf{Y}', \mathbf{Z}')$ *and* $\widetilde{\mathbf{D}}$ *denote the* *set defined above. Then,*

$$P_{\mathbf{x}'}(\widetilde{\mathbf{d}}|\mathbf{s}\backslash\widetilde{\mathbf{d}}) = \frac{Q[\mathbf{S}]}{\sum_{\widetilde{\mathbf{D}}} Q[\mathbf{S}]} = Q[\widetilde{\mathbf{D}}|\mathbf{S}\backslash\widetilde{\mathbf{D}}]$$

is not c-gID from $(\mathbb{A}, \mathcal{G})$.

A proof for this lemma is presented in Appendix A. The remainder of the proof of this sub-case is identical to the proof of the second sub-case.

In both cases considered in Sections 4.1-4.2, we proved that $P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$ is not c-gID from $(\mathbb{A}, \mathcal{G})$. Recall that $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = P_{\mathbf{x}'}(\mathbf{y}'|\mathbf{z}')$. This concludes the proof of the necessity part of Theorem 3.

Summing up: Recall that the necessity part required us to show when $P_{\mathbf{x}'}(\mathbf{y}', \mathbf{z}')$ is not gID from $(\mathbf{A}, \mathcal{G})$, $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ is not c-gID from $(\mathbf{A}, \mathcal{G})$. In the sufficiency part, had to show that $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ is c-gID from $(\mathbf{A}, \mathcal{G})$ whenever $P_{\mathbf{x}'}(\mathbf{y}', \mathbf{z}')$ is gID from $(\mathbf{A}, \mathcal{G})$. These two results together conclude the proof of Theorem 3.

5 CONCLUSION

We considered the problem of identifying a conditional causal effect from a causal graph $\mathcal{G}$ and a particular set of known observational/interventional distributions in the form of Q-notations. We called this problem c-gID and showed that any c-gID problem has an equivalent g-ID problem. Using this equivalency, we proposed the first sound and complete algorithm for solving c-gID problem.

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.

Juan Correa, Sanghack Lee, and Elias Bareinboim. Nested counterfactual identification from arbitrary surrogate experiments. *Advances in Neural Information Processing Systems*, 34:6856–6867, 2021.

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

Yaroslav Kivva, Ehsan Mokhtarian, Jalal Etesami, and Negar Kiyavash. Revisiting the general identifiability problem. In *The 38th Conference on Uncertainty in Artificial Intelligence*, 2022.

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

Sanghack Lee, Juan Correa, and Elias Bareinboim. General transportability–synthesizing observations and experiments from heterogeneous domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10210–10217, 2020.

Ehsan Mokhtarian, Fateme Jamshidi, Jalal Etesami, and Negar Kiyavash. Causal effect identification with context-specific independence relations of control variables. In *International Conference on Artificial Intelligence and Statistics*, pages 11237–11246. PMLR, 2022.

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. Identifying conditional causal effects. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 561–568, 2004.

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.