On Causal Identification under Markov Equivalence*

Amin Jaber\textsuperscript{1}, Jiji Zhang\textsuperscript{2} and Elias Bareinboim\textsuperscript{1}
\textsuperscript{1}Department of Computer Science, Purdue University
\textsuperscript{2}Department of Philosophy, Lingnan University
\{jaber0, eb\}@purdue.edu, jijizhang@ln.edu.hk

Abstract

In this work, we investigate the problem of computing an experimental distribution from a combination of the observational distribution and a partial qualitative description of the causal structure of the domain under investigation. This description is given by a partial ancestral graph (PAG) that represents a Markov equivalence class of causal diagrams, i.e., diagrams that entail the same conditional independence model over observed variables, and is learnable from the observational data. Accordingly, we develop a complete algorithm to compute the causal effect of an arbitrary set of intervention variables on an arbitrary outcome set.

1 Introduction

A prominent approach to infer causal effects leverages a combination of substantive knowledge about the domain under investigation, usually encoded in the form of a causal diagram, with observational data [Pearl, 2000; Bareinboim and Pearl, 2016]. A sample diagram is shown in Fig. 1a, where the nodes represent variables, directed edges represent direct causal relations from tails to heads, and bi-directed arcs represent the presence of unobserved (latent) variables that generate spurious association between the variables. The decision problem of whether an interventional distribution can be computed from a combination of observational and experimental data together with the causal diagram is known as the problem of identification of causal effects (identification, for short). For instance, a possible task is to identify the effect of $do(X=x)$ on $V_4=v_4$, i.e. $P_x(v_4)$, given the diagram in Fig. 1a and data from the distribution $P(X, V_1, ..., V_4)$.

This problem has been extensively studied in the literature with a number of criteria established [Pearl, 1993; Galles and Pearl, 1995; Kuroki and Miyakawa, 1999; Tian and Pearl, 2002; Huang and Valtorta, 2006; Shpitser and Pearl, 2006; Bareinboim and Pearl, 2012], which include the back-door criterion and the do-calculus [Pearl, 1995]. Despite their power, these techniques require a fully specified causal diagram, which is not always available in practical settings.

*This is an abridged version of [Jaber et al., 2018a] which won the best student paper award at UAI-18 and [Jaber et al., 2019].

Figure 1: A causal diagram (left) and the inferred PAG (right).

One may try to learn the causal diagram from data. However, it is common that only an equivalence class of causal diagrams can be consistently inferred from observational data [Verma, 1993; Spirtes et al., 2001]. A useful representation of such an equivalence class is a partial ancestral graph (PAG) [Zhang, 2008b]. Fig. 1b shows the PAG that can be inferred from observational data generated by Fig. 1a. In PAGs, directed edges signify (possibly indirect) causal relations and circle marks indicate structural uncertainty.

In this work, we analyze the marriage of these two lines of investigation. Identification from an equivalence class is considerably more challenging than from a single diagram due to the structural uncertainty regarding both the causal relations among the variables and the presence of hidden variables. Zhang [2007] extended the do-calculus to PAGs. In practice, however, it is computationally hard to decide whether there exists a sequence of applications of the rules of the generalized calculus to identify an effect. Perković et al. [2015] generalized the back-door criterion to PAGs, and provided a complete algorithm to find a back-door admissible set, should such a set exist. However, in practice, no adjustment set exists for many identifiable effects. In this extended abstract, we summarize the approach taken in [Jaber et al., 2018b; Jaber et al., 2018b; Jaber et al., 2019] that culminated in a complete algorithm to identify causal effects given PAGs. Specifically, our contributions are as follows:

1. We revisit the original identification algorithm given causal diagrams and introduce a new formulation which is more amenable under structural uncertainties.
2. We derive crucial PAG properties including a novel graph-decomposition strategy that breaks a target causal distribution into an equal product of more tractable ones.
3. We develop a complete algorithm to compute the effect of an arbitrary set of intervention variables on an arbitrary outcome set from a PAG and observational data.
2 Preliminaries

In this section, we introduce the basic notation and machinery used throughout the paper. Bold capital letters denote sets of variables, while bold lowercase letters stand for particular value assignments to those variables.

2.1 Structural Causal Models

We use the language of Structural Causal Models (SCM) [Pearl, 2000, pp. 204-207] as our basic semantic framework. Formally, an SCM $M$ is a 4-tuple $(U, V, F, P(U))$, where $U$ is a set of exogenous (latent) variables and $V$ is a set of endogenous (measured) variables. $F$ represents a collection of functions $\{f_i\}$ such that each endogenous variable $V_i \in V$ is determined by a function $f_i \in F$, where $f_i$ is a mapping from the respective domain of $U_i \cup Pa_i$ to $V_i$, $U_i \subseteq U$, $Pa_i \subseteq V \setminus \{V_i\}$. The uncertainty is encoded through a probability distribution over the exogenous variables, $P(U)$, and the marginal distribution induced over the endogenous variables $P(V)$ is called observational. Every SCM is associated with a causal diagram where every variable $V_i \in V$ is a node, and there exists a directed edge from every node in $Pa_i$ to $V_i$. Also, for every pair $V_i, V_j \in V$ such that $U_i \cap U_j \neq \emptyset$, there exists a bi-directed edge between $V_i$ and $V_j$. We restrict our study to recursive systems, which means that the corresponding diagram will be acyclic.

Within the structural semantics, performing an action $X = x$ is represented through the do-operator, $do(X = x)$, which encodes the operation of replacing the original equation for $X$ by the constant $x$ and induces a submodel $M_x$. The resulting distribution is denoted by $P_x$, which is the main target for identification in this paper. For further details on structural models, we refer readers to [Pearl, 2000].

2.2 Ancestral Graphs

We now introduce a graphical representation of equivalence classes of causal diagrams. A mixed graph can contain directed and bi-directed edges. A is an ancestor of $B$ if they share a directed path out of $A$. $A$ is a spouse of $B$ if $A \leftrightarrow B$ is present. An almost directed cycle happens when $A$ is both a spouse and an ancestor of $B$. An inducing path is a path on which every node (except for the endpoints) is a collider on the path (i.e., both edges incident to $X$ are into $X$) and every collider is an ancestor of an endpoint of the path. A mixed graph is ancestrally directed if it doesn’t contain a directed or almost directed cycle. It is maximal if there is no inducing path between any two non-adjacent nodes. A Maximal Ancestral Graph (MAG) is a graph that is both ancestral and maximal [Richardson and Spirtes, 2002].

In short, a MAG represents a set of causal diagrams that entail the same independence and ancestral relations among the observed variables. Different MAGs may be Markov equivalent in that they entail the same exact independence model. A partial ancestral graph (PAG) represents an equivalence class of MAGs $[M]$, which shares the same adjacencies as every MAG in $[M]$ and displays all and only the invariant edge marks. A circle indicates an uncommon edge mark. A PAG is learnable from the conditional independence relations among the observed variables and the FCI algorithm is a standard method to learn such an object [Zhang, 2008b].

2.3 Graphical Notions

Given a causal diagram, a MAG, or a PAG, a path between $X$ and $Y$ is potentially directed (causal) from $X$ to $Y$ if there is no arrowhead on the path pointing towards $X$. $X$ is a possible ancestor of $Y$, i.e., $X \in \text{An}(Y)$, if there is a potentially directed path from $X$ to $Y$. By stipulation, $X \in \text{An}(X)$. $Y$ is called a possible child of $X$, i.e., $Y \in \text{Ch}(X)$, if they are adjacent and the edge is not into $X$. For a set of nodes $X$, we have $\text{Ch}(X) = \bigcup_{x \in X} \text{Ch}(X)$. If the edge marks on a path between $X$ and $Y$ are all circles, we call the path a circle path. We refer to the closure of nodes connected with circle paths as a bucket. Obviously, given a PAG, nodes are partitioned into a unique set of buckets.

A directed edge $X \rightarrow Y$ in a MAG or PAG is visible if there exists no causal diagram in the corresponding equivalence class where there is an inducing path between $X$ and $Y$ that is into $X$. This implies that a visible edge is not confounded ($X \not\rightarrow \rightarrow Y$ doesn’t exist). Which directed edges are visible is easily decidable by a graphical condition [Zhang, 2008a, Def. 8], so we simply mark visible edges by $v$. For brevity, we refer to any edge that is not a visible directed edge as invisible.

3 Revisit Identification in Causal Diagrams

Tian and Pearl [2002] presented an identification algorithm based on a decomposition strategy of the causal diagram into a set of so-called c-components (confounded components).

Definition 1 (C-Component). In a causal diagram, two nodes are said to be in the same c-component if and only if they are connected by a bi-directed path, i.e., a path composed solely of bi-directed edges.

For any set $C \subseteq V$, the quantity $Q[C]$ is defined to denote the post-intervention distribution of $C$ under an intervention on $V \setminus C$, i.e., $P_{V \setminus C}(\cdot)$. Given a diagram $\mathcal{D}$, $Q[C]$ decomposes into a product of sub-queries over the c-components in $\mathcal{D}_C$, where $\mathcal{D}_C$ denotes the (induced) subgraph of $\mathcal{D}$ over $C$. Hence, we get the following decomposition, where $C_i$ is a c-component in $\mathcal{D}_C$:

$$Q[C] = \prod_i Q[C_i]$$

(1)

The significance of c-components and their decomposition is evident from [Tian, 2002, Lemmas 10, 11], which are the basis of Tian’s identification algorithm. Our goal is to reformulate the procedure with a more local, atomic criterion shown in Lem. 1. The step-wise algorithm is shown in Alg. 1.

Lemma 1. Given a causal diagram $\mathcal{D}$ over $V$, $X \in T \subseteq V$, and $P_{V \setminus T}$, i.e., an expression for $Q[T]$. If $X$ is not in the same c-component with a child in $\mathcal{D}_T$, then $Q[T \setminus \{X\}]$ is identifiable and given by

$$Q[T \setminus \{X\}] = \frac{P_{V \setminus T}}{Q[S^X]} \times \sum_x Q[S^X]$$

(2)

where $S^X$ is the c-component of $X$ in $\mathcal{D}_T$ and $Q[S^X]$ is computable from $P_{V \setminus T}$ by [Tian, 2002, Lemma 11].
The revised algorithm requires checking an atomic criterion at every instance of the recursive routine `IDENTIFY`. This might not be crucial when the precise causal diagram is known and the induced subgraphs preserve complete information about the c-components and the ancestral relations between the nodes. However, it becomes significant when the domain description is an equivalence class represented by a PAG, in which structural information is partial.

4 PAG Properties and Q-Decomposition

Evidently, induced subgraphs of the original causal diagram play a critical role in identification (cf Alg. 1). It is natural to expect that in the generalized setting we study here, induced subgraphs of the given PAG will also play an important role. An immediate challenge, however, is that a subgraph of a PAG over \( V \) is in general, not a PAG that represents a full Markov equivalence class. In particular, if \( D \) is a diagram in the equivalence class represented by \( \mathcal{P} \), \( \mathcal{P}_A \) is in general not the PAG that represents the equivalence class of \( D_A \). For example, let \( D \) and \( \mathcal{P} \) denote the diagram and the corresponding PAG in Figure 1, respectively, and let \( A = \{V_2, V_3, X, V_4\} \). Then, \( V_4 \) is disconnected in \( D_A \) while \( X \) is a child of \( V \) in \( \mathcal{P}_A \). Despite this subtlety, we establish a few facts showing that for any \( A \subseteq V \) and any diagram \( D \) in the equivalence class represented by \( \mathcal{P} \), some information about \( D_A \), which is particularly relevant to identification, can still be read off from \( \mathcal{P}_A \). In [Jaber et al., 2018a], we discuss a key property related to c-components.

Definition 2 (PC-Component). In a MAG, a PAG, or any induced subgraph thereof, two nodes are in the same possible c-component (pc-component) if there is a path between them such that (1) all non-endpoint nodes along the path are colliders, and (2) none of the edges is visible.

As mentioned earlier, a c-component in a causal diagram plays a central role in identification. The following proposition establishes a graphical condition in an induced subgraph \( \mathcal{P}_A \) that is necessary for two nodes being in the same c-component in \( D_A \) for some diagram \( D \) represented by \( \mathcal{P} \).

Proposition 1. Let \( \mathcal{P} \) be a PAG over \( V \), and \( D \) be any causal diagram in the equivalence class represented by \( \mathcal{P} \). For any \( X, Y \in A \subseteq V \), if \( X \) and \( Y \) are in the same c-component in \( D_A \), then \( X \) and \( Y \) are in the same pc-component in \( \mathcal{P}_A \).

This result provides a sufficient condition for not belonging to the same c-component in any of the induced causal diagrams. In the aforementioned \( \mathcal{P}_A \) of Fig. 1b, for example, \( V_3 \) and \( V_1 \) or \( X \) and \( V_4 \) are not in the same pc-component, which implies by Prop. 1 that they are not in the same c-component in \( D_A \) for any \( D \) in the equivalence class represented by the PAG in Fig. 1b. As a special case of Def. 2, we define the following notion, which will prove useful later on.

Definition 3 (DC-Component). In a MAG, a PAG, or any induced subgraph thereof, two nodes are in the same definite c-component (dc-component) if they are connected with a bi-directed path, i.e. a path composed of bi-directed edges.

Next, we use the pc-component property to devise a decomposition for \( Q[C] \) in PAGs akin to that for causal diagrams presented in Eq. 1. However, the decomposition in PAGs is challenging due to the structural uncertainties; most relevant the presence of latent confounders. For instance, given the query \( Q[Y] \) over the PAG in Fig. 2, the sequence of nodes \( \{Y_2, Y_3, Y_4, Y_5, Y_1\} \) is connected with invisible edges, which are possibly confounded. Hence, any naive decomposition of \( Q[Y] \) into a product of sub-queries over subsets of \( Y \) is invalid since we can construct a diagram in the equivalence class which violates this decomposition.

To develop a valid decomposition, we start by introducing the notion of a region. In short, a region is the pc-component of a set \( A \) appended with the corresponding buckets of the nodes. We append the pc-component set with the corresponding buckets of the nodes to avoid non-identifiability of the sub-queries since no sufficient causal information is present within a bucket. Consider the PAG in Figure 2, and let \( C = Y \) and \( A = \{Y_3\} \). Then, \( R^C_A = \{Y_3, Y_2, Y_4, Y_5\} \) since \( Y_2 \) and \( Y_4 \) are in the pc-component of \( Y_3 \) and \( Y_5 \) is in the same bucket as \( Y_4 \). For simplicity, we often drop \( C \), i.e. \( R^C_A \), whenever it is clear from the context. Using this construction, we derive the decomposition in Theorem 1.

Definition 4 (Region \( R^C_A \)). Given a PAG or a MAG \( \mathcal{G} \) over \( V \), and \( A \subseteq C \subseteq V \). Let the region of \( A \) with respect to \( C \), denoted \( R^C_A \), be the union of the buckets that contain nodes in the pc-component of \( A \) in the induced subgraph \( \mathcal{G}_C \).

Theorem 1. Given a PAG \( \mathcal{P} \) over \( V \) and set \( C \subseteq V \), \( Q[C] \) decomposes as follows, where \( A \subset C \) and \( R^{(\cdot)} = R^C_{(\cdot)} \).

\[
Q[C] = \frac{Q[R^A_C]Q[R^C_{C \setminus R^C_A}]}{Q[R^A_C]Q[R^C_{C \setminus R^C_A}]}
\]

6183
Back to the query \( Q[Y] \) over the PAG in Fig. 2, it can be decomposed as follows with \( A = \{ Y_3 \} \):

\[
Q[Y] = \frac{Q[Y \setminus \{ Y_3 \}] Q[\{ Y_1, Y_4, Y_5 \}]}{Q[\{ Y_4, Y_5 \}]} \tag{3}
\]

5 Identification in PAGs

We start by formally defining the notion of identification given a PAG, which generalizes the model-specific notion \cite{Pearl:2000:AAA:358533}. Let the query be 

\[
\begin{align*}
Q[Y] = & Q[Y \setminus fY_3g] \cdot Q[\{ Y_1, Y_4, Y_5 \}] \\
= & \frac{P(Y)}{P(Y_1)P(Y_2)} \cdot P(Y_3 | Y_4, Y_5)
\end{align*}
\]

for \( P(x) \) is a result of Eqs. 5, 6, and 7 as follows.

\[
P(x) = P(Y_1)P(Y_2) \cdot P(Y_3 | Y_4, Y_5)
\]


\[
P(x) = P(Y_1)P(Y_2) \cdot P(Y_3 | Y_4, Y_5)
\]


Algorithm 2 IDP(x,y) given PAG \( P \)

\begin{itemize}
\item \textbf{Input:} two disjoint sets \( X, Y \subset V \)
\item \textbf{Output:} Expression for \( P_x(y) \) or \text{FAIL}
\end{itemize}

\begin{algorithmic}
1: Let \( D = \{ V \} \times \emptyset \) and \text{FAIL}
2: \( P_x(y) = \sum_{d \in D} \text{IDP}(D, V, P) \)
3: \textbf{function} \text{IDP}(C, T, Q; Q = Q[T])
4: \text{if} \( C = \emptyset \) then return 1
5: \text{if} \( C = T \) then return \( Q \)
6: \text{if} \( \exists B \in T \setminus C \) such that \( C^B \cap \text{ch}(B) \subseteq B \) then
7: \quad Compute \( Q[T \setminus B] \) from \( Q \) via Thm. 2
8: \quad return \text{IDP}(C, T \setminus B, Q[T \setminus B])
9: \text{else} \text{if} \( \exists B \in C \) such that \( R_B \neq C \) then
10: \quad return \text{IDP}(R_B \setminus T, Q) \cdot \text{IDP}(R_C \setminus R_B \setminus T, Q)
11: \text{else}
12: \quad throw \text{FAIL}
\end{algorithmic}

6 Identification in PAGs

5 Identification in PAGs

We start by formally defining the notion of identification given a PAG, which generalizes the model-specific notion \cite{Pearl:2000:AAA:358533}. Let the query be

\[
\begin{align*}
Q[Y] = & Q[Y \setminus fY_3g] \cdot Q[\{ Y_1, Y_4, Y_5 \}] \\
= & \frac{P(Y)}{P(Y_1)P(Y_2)} \cdot P(Y_3 | Y_4, Y_5)
\end{align*}
\]

for \( P(x) \) is a result of Eqs. 5, 6, and 7 as follows.

\[
P(x) = P(Y_1)P(Y_2) \cdot P(Y_3 | Y_4, Y_5)
\]


Algorithm 2 IDP(x,y) given PAG \( P \)

\begin{itemize}
\item \textbf{Input:} two disjoint sets \( X, Y \subset V \)
\item \textbf{Output:} Expression for \( P_x(y) \) or \text{FAIL}
\end{itemize}

\begin{algorithmic}
1: Let \( D = \{ V \} \times \emptyset \) and \text{FAIL}
2: \( P_x(y) = \sum_{d \in D} \text{IDP}(D, V, P) \)
3: \textbf{function} \text{IDP}(C, T, Q; Q = Q[T])
4: \text{if} \( C = \emptyset \) then return 1
5: \text{if} \( C = T \) then return \( Q \)
6: \text{if} \( \exists B \in T \setminus C \) such that \( C^B \cap \text{ch}(B) \subseteq B \) then
7: \quad Compute \( Q[T \setminus B] \) from \( Q \) via Thm. 2
8: \quad return \text{IDP}(C, T \setminus B, Q[T \setminus B])
9: \text{else} \text{if} \( \exists B \in C \) such that \( R_B \neq C \) then
10: \quad return \text{IDP}(R_B \setminus T, Q) \cdot \text{IDP}(R_C \setminus R_B \setminus T, Q)
11: \text{else}
12: \quad throw \text{FAIL}
\end{algorithmic}

6 Identification in PAGs

5 Identification in PAGs

We start by formally defining the notion of identification given a PAG, which generalizes the model-specific notion \cite{Pearl:2000:AAA:358533}. Let the query be

\[
\begin{align*}
Q[Y] = & Q[Y \setminus fY_3g] \cdot Q[\{ Y_1, Y_4, Y_5 \}] \\
= & \frac{P(Y)}{P(Y_1)P(Y_2)} \cdot P(Y_3 | Y_4, Y_5)
\end{align*}
\]

for \( P(x) \) is a result of Eqs. 5, 6, and 7 as follows.

\[
P(x) = P(Y_1)P(Y_2) \cdot P(Y_3 | Y_4, Y_5)
\]


Algorithm 2 IDP(x,y) given PAG \( P \)

\begin{itemize}
\item \textbf{Input:} two disjoint sets \( X, Y \subset V \)
\item \textbf{Output:} Expression for \( P_x(y) \) or \text{FAIL}
\end{itemize}

\begin{algorithmic}
1: Let \( D = \{ V \} \times \emptyset \) and \text{FAIL}
2: \( P_x(y) = \sum_{d \in D} \text{IDP}(D, V, P) \)
3: \textbf{function} \text{IDP}(C, T, Q; Q = Q[T])
4: \text{if} \( C = \emptyset \) then return 1
5: \text{if} \( C = T \) then return \( Q \)
6: \text{if} \( \exists B \in T \setminus C \) such that \( C^B \cap \text{ch}(B) \subseteq B \) then
7: \quad Compute \( Q[T \setminus B] \) from \( Q \) via Thm. 2
8: \quad return \text{IDP}(C, T \setminus B, Q[T \setminus B])
9: \text{else} \text{if} \( \exists B \in C \) such that \( R_B \neq C \) then
10: \quad return \text{IDP}(R_B \setminus T, Q) \cdot \text{IDP}(R_C \setminus R_B \setminus T, Q)
11: \text{else}
12: \quad throw \text{FAIL}
\end{algorithmic}
We can reformulate Eq. 8 using independence relations to,
\[ P_X(y) = P(y_2; y_3|x_2; y_4) \cdot P(y_1|x_1; y_5) \cdot P(y_4; y_5) \]

6 Conclusion
In this work, we investigated the problem of identification of interventional distributions in Markov equivalence classes represented by PAGs. We first revisited the identification algorithm given a specific causal diagram and reformulated it using an atomic criterion. We then derived new graphical properties for induced subgraphs of PAGs over an arbitrary subset of nodes, including a novel decomposition strategy. Finally, building on these results, we proposed an identification criterion in PAGs and employed it in a complete identification algorithm. We believe the proposed approach enables empirical researchers to perform causal reasoning while lacking substantive background knowledge. This is considered more “data-driven”, and more aligned with the zeitgeist in machine learning today.

Acknowledgements
Bareinboim and Jaber are supported in parts by grants from NSF IIS-1704352, IIS1750807 (CAREER), IBM Research, and Adobe Research. Zhang’s research was supported in part by the Research Grants Council of Hong Kong under the General Research Fund LU13602818.

References


