Causal Effect Identification in Alternative Acyclic Directed Mixed Graphs

Jose M. Peña JOSE.M.PENA@LIU.SE
Linköping University
Linköping (Sweden)

Abstract

Alternative acyclic directed mixed graphs (ADMGs) are graphs that may allow causal effect identification in scenarios where Pearl's original ADMGs may not, and vice versa. Therefore, they complement each other. In this paper, we introduce a sound algorithm for identifying arbitrary causal effects from alternative ADMGs. Moreover, we show that the algorithm is complete for identifying the causal effect of a single random variable on the rest. We also show that the algorithm follows from a calculus similar to Pearl's *do*-calculus.

Keywords: Causality; causal effect identification; acyclic directed mixed graphs.

1. Introduction

Undirected graphs (UGs), bidirected graphs (BGs), and directed and acyclic graphs (DAGs) have extensively been studied as representations of independence models. DAGs have also been studied as representation of causal models, because they can model asymmetric relationships between random variables. DAGs and UGs (respectively BGs) have been extended into chain graphs (CGs), which are graphs with directed and undirected (respectively bidirected) edges but without semidirected cycles. Therefore, CGs can model both symmetric and asymmetric relationships between random variables. CGs with directed and undirected edges may represent a different independence model depending on whether the Lauritzen-Wermuth-Frydenberg (LWF) or the Andersson-Madigan-Perlman (AMP) interpretation is considered (Lauritzen, 1996; Andersson et al., 2001). CGs with directed and bidirected edges have a unique interpretation, the so-called multivariate regression (MVR) interpretation (Cox and Wermuth, 1996). MVR CGs have been extended by (i) relaxing the semidirected acyclity constraint so that only directed cycles are forbidden, and (ii) allowing up to two edges between any pair of nodes. The resulting models are called acyclic directed mixed graphs (ADMGs) (Richardson, 2003). AMP CGs have also been extended similarly (Peña, 2016). The resulting models are called alternative acyclic directed mixed graphs (aADMGs). It is worth mentioning that neither the original ADMGs nor any other family of mixed graphical models that we know of (e.g. summary graphs (Cox and Wermuth, 1996), ancestral graphs (Richardson and Spirtes, 2002), MC graphs (Koster, 2002) or loopless mixed graphs (Sadeghi and Lauritzen, 2014)) subsume AMP CGs and hence aADMGs. To see it, we refer the reader to the works by Richardson and Spirtes (2002, p. 1025) and Sadeghi and Lauritzen (2014, Section 4.1).

In addition to represent independence models, some of the graphical models mentioned above have been used for causal effect identification, i.e. to determine if the causal effect of

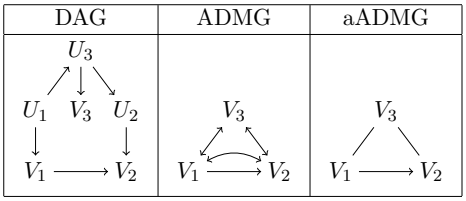

Figure 1: Example where $f(v_2|\widehat{v_1})$ is identifiable from the aADMG but not from the ADMG.

an intervention is identifiable from observational quantities. For instance, Pearl's approach to causal effect identification makes use of ADMGs to represent causal models over the observed variables (Pearl, 2009). The directed edges represent causal relationships, whereas the bidirected edges represent confounding, i.e. a latent common cause. A key feature of Pearl's approach is that no assumption is made about the functional form of the causal relationships. That is, each variable A is an unconstrained function of its observed causes $Pa(A)$ and its unobserved causes U_A, i.e. $A = g(Pa(A), U_A)$. Without loss of generality, we can consider U_A as being unidimensional (Mooij et al., 2016, Proposition 4). We do so. This U_A is sometimes called noise or error. In this paper, we study causal effect identification under the assumption that $A = g(Pa(A)) + U_A$, also called additive noise model. This is a rather common assumption in causal discovery (Bühlmann et al., 2014; Mooij et al., 2016; Peters et al., 2014). Note also that linear structural equation models, which have extensively been studied for causal effect identification (Pearl, 2009, Chapter 5), are additive noise models. As argued by Peña (2016), aADMGs are suitable for representing causal models with additive noise. The main difference between ADMGs and aADMGs is that an edge $A - B$ in an aADMG represents dependence between U_A and U_B given the unobserved causes of the rest of the observed nodes, as opposed to a bidirected edge in an ADMG which represents marginal dependence due to confounding. The reason for studying aADMGs for causal effect identification is that they may allow identifying causal effects that ADMGs may not. We illustrate this with the example in Figure 1, which is borrowed from Peña (2016). Peña and Bendtsen (2017) assign fictitious meanings to the variables in the example and provide additional examples. The ADMG and aADMG represent the causal model over the observed variables represented by the DAG. The ADMG is derived from the DAG by keeping the directed edges between observed variables, and adding a bidirected edge between two observed variables if and only if they have a confounder (Tian and Pearl, 2002b, Section 5). The aADMG is derived from the DAG by keeping the directed edges between observed variables, and adding an undirected edge between two observed variables if and only if their unobserved causes are not separated in the DAG given the unobserved causes of the rest of the observed variables. Clearly, the causal effect on V_2 of intervening on V_1, denoted as the density function $f(v_2|\widehat{v_1})$, is not identifiable from the ADMG (Tian and Pearl, 2002a, Theorem 4). However, $f(v_2|\widehat{v_1})$ is identifiable from the aADMG and is given by

$$f(v_2|\widehat{v_1}) = \int f(v_2|v_1, v_3) f(v_3)\, dv_3. \tag{1}$$

To see it, recall that we assumed that V_3 determines U_3, which blocks the path $V_1 \leftarrow U_1 \rightarrow U_3 \rightarrow U_2 \rightarrow V_2$ in the DAG. This can also be seen directly in the aADMG, as V_3 blocks the path $V_1 - V_3 - V_2$. Therefore, we can identify the desired causal effect by just adjusting for V_3, since V_3 blocks all non-causal paths from V_1 to V_2.[1] It is worth mentioning that Peña (2016) also provides an example where the ADMG allows for causal effect identification whereas the aADMG does not: Simply reverse the edge $U_3 \rightarrow U_2$ in Figure 1. Therefore, ADMGs and aADMGs are more complementary than competing causal models.

As mentioned, aADGMs were proposed by Peña (2016), who mainly studied them as representation of statistical independence models. In particular, their global, local and pairwise Markov properties were studied. Later, Peña and Bendtsen (2017) considered aADGMs for causal effect identification. Specifically, they presented a calculus similar to Pearl's *do*-calculus (Pearl, 2009; Shpitser and Pearl, 2006), and a decomposition of the density function represented by an aADMG that is similar to the Q-decomposition by Tian and Pearl (2002a,b). In this paper, we extend the decomposition to identify further causal effects. The result is a sound algorithm for causal effect identification in aADMGs. Although the algorithm is not complete for arbitrary causal effects, we show that it is complete for the identification of the causal effect of a single variable on the rest of the variables. We also show that the algorithm follows from the calculus of interventions in Peña and Bendtsen (2017). The rest of the paper is organized as follows. Section 2 introduces some preliminaries, including a detailed account of aADMGs for causal modeling. Section 3 presents our novel algorithm for causal effect identification, and proves its soundness for arbitrary causal effects and completeness for restricted causal effects. It also proves that the algorithm follows from a calculus of interventions. Section 4 closes the paper with some lines of future research.

2. Preliminaries

Unless otherwise stated, all the graphs and density functions in this paper are defined over a finite set of continuous random variables V. The elements of V are not distinguished from singletons. An aADMG G is a simple graph with possibly directed and undirected edges but without directed cycles. There may be up to two edges between any pair of nodes, but in that case the edges must be different and one of them must be undirected to avoid directed cycles. Edges between a node and itself are not allowed.

Given an aADMG G, the parents of a set $X \subseteq V$ in G are $Pa_G(X) = \{A | A \rightarrow B$ is in G with $B \in X\}$. The children of X in G are $Ch_G(X) = \{A | A \leftarrow B$ is in G with $B \in X\}$. The neighbours of X in G are $Ne_G(X) = \{A | A - B$ is in G with $B \in X\}$. The ancestors of X in G are $An_G(X) = \{A | A \rightarrow \ldots \rightarrow B$ is in G with $B \in X$ or $A \in X\}$. Moreover, X is called an ancestral set if $X = An_G(X)$. The descendants of X in G are $De_G(X) = \{A | A \leftarrow \ldots \leftarrow B$ is in G with $B \in X$ or $A \in X\}$. A route between a node V_1 and a node V_n on G is a sequence of (not necessarily distinct) nodes $V_1, \ldots, V_n$ such that V_i and V_{i+1} are adjacent in G for all $1 \leq i < n$. We do not distinguish between the sequences $V_1, \ldots, V_n$ and $V_n, \ldots, V_1$,

1. True that there are ADMGs that represent the correct independence model over the observed variables and allow for identification via Equation 1, specifically the ADMGs $G_1 = \{V_1 \rightarrow V_2, V_1 \leftarrow V_3, V_3 \rightarrow V_2\}$, $G_2 = \{V_1 \rightarrow V_2, V_1 \leftrightarrow V_3, V_3 \rightarrow V_2\}$ and $G_3 = \{V_1 \rightarrow V_2, V_1 \leftarrow V_3, V_3 \leftrightarrow V_2\}$. However, these ADMGs do not represent the correct causal model, i.e. they contain false causal relationships such as $V_1 \leftarrow V_3$ or $V_3 \rightarrow V_2$. This leads to wrong expressions for other causal effects such as $f(v_1|\widehat{v_3}) = f(v_1|v_3)$ in G_1 and G_3, and $f(v_2|\widehat{v_3}, v_1) = f(v_2|v_3, v_1)$ in G_2.

	Input: An aADMG G.
	Output: The magnified aADMG G'.
1	Set $G' = G$
2	For each node A in G
3	Add the node U_A and the edge $U_A \rightarrow A$ to G'
4	For each edge $A - B$ in G
5	Replace $A - B$ with the edge $U_A - U_B$ in G'
6	Return G'

Table 1: Algorithm for magnifying an aADMG.

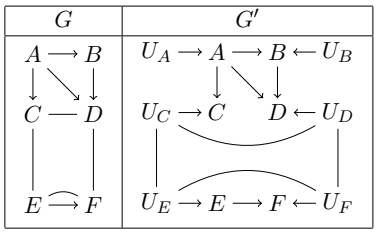

Figure 2: Example of the magnification of an aADMG.

i.e. they represent the same route. A node C on a route in G is said to be a collider on the route if $A \rightarrow C \leftarrow B$ or $A \rightarrow C - B$ is a subroute. Note that maybe $A = B$. Moreover, the route is said to be connecting given $Z \subseteq V$ when every collider on the route is in Z, and every non-collider on the route is outside Z. Let X, Y and Z denote three disjoint subsets of V. When there is no route in G connecting a node in X and a node in Y given Z, we say that X is separated from Y given Z in G and denote it as $X \perp_G Y|Z$. We say that a density function $f(v)$ satisfies the global Markov property with respect to G if every separation in G implies an independence in $f(v)$.

Given an aADMG G and a set $W \subseteq V$, let G_W denote the subgraph of G induced by W. Similarly, let G^W denote the aADMG over W constructed as follows: $A \rightarrow B$ is in G^W if and only if $A \rightarrow B$ is in G, whereas $A - B$ is in G^W if and only if $A - B$ is in G or $A - V_1 - \ldots - V_n - B$ is in G with $V_1, \ldots, V_n \notin W$.

2.1 Causal Interpretation

Let us assume that V is normally distributed. In this section, we show that an aADMG G can be interpreted as a system of structural equations with correlated errors. Specifically, the system includes an equation for each $A \in V$, which is of the form

$$A = \beta_A Pa_G(A) + U_A \tag{2}$$

where U_A denotes the noise or error term. The error terms are represented implicitly in G. They can be represented explicitly by magnifying G into the aADMG G' as shown in Table 1. The magnification basically consists in adding the error nodes U_A to G and connect them appropriately. Figure 2 shows an example. Note that Equation 2 implies that A is determined by $Pa_G(A) \cup U_A$ and U_A is determined by $A \cup Pa_G(A)$. Let U denote all the

Input: An aADMG G and a set $X \subseteq V$.
Output: aADMG after intervening on X in G.

Delete from G all the edges $A \rightarrow B$ with $B \in X$
For each path $A - V_1 - \ldots - V_n - B$ in G with $A, B \notin X$ and $V_1, \ldots, V_n \in X$
  Add the edge $A - B$ to G
Delete from G all the edges $A - B$ with $B \in X$
Return G

Table 2: Algorithm for intervening on an aADMG.

error nodes in G'. Formally, we say that $A \in V \cup U$ is determined by $Z \subseteq V \cup U$ when $A \in Z$ or A is a function of Z. We use $Dt(Z)$ to denote all the nodes that are determined by Z. From the point of view of the separations, that a node outside the conditioning set of a separation is determined by the conditioning set has the same effect as if the node were actually in the conditioning set. Bearing this in mind, it can be proven that, as desired, G and G' represent the same separations over V (Peña, 2016, Theorem 9).

Finally, let $U \sim \mathcal{N}(0, \Lambda)$ such that $(\Lambda^{-1})_{U_A, U_B} = 0$ if $U_A - U_B$ is not in G'. Then, G can be interpreted as a system of structural equations with correlated errors as follows. For any $A \in V$

$$A = \sum_{B \in Pa_G(A)} \beta_{AB} B + U_A \tag{3}$$

and for any other $B \in V$

$$covariance(U_A, U_B) = \Lambda_{U_A, U_B}. \tag{4}$$

It can be proven that this causal interpretation of aADMGs works as intended: Every density function $f(v)$ specified by Equations 3 and 4 is Gaussian, and it satisfies the global Markov property with respect to G (Peña, 2016, Theorems 10 and 11).

A less formal but more intuitive interpretation of aADMGs is as follows. We can interpret the parents of each node in an aADMG as its observed causes. Its unobserved causes are summarized by an error node that is represented implicitly in the aADMG. We can interpret the undirected edges in the aADMG as the correlation relationships between the different error nodes. The causal structure is constrained to be a DAG, but the correlation structure can be any UG. This causal interpretation of aADMGs parallels that of the original ADMGs. There are however two main differences. First, the noise in the ADMGs is not necessarily additive normal. Second, the correlation structure of the error nodes in the ADMGs is represented by a covariance or bidirected graph. Therefore, whereas a missing edge between two error nodes in ADMGs represents marginal independence, in aADMGs it represents conditional independence given the rest of the error nodes. This means that ADMGs and aADMGs represent complementary causal models. Consequently, there may be causal effects that can be identified with one but not with the other. An example was provided in Section 1.

Given the above causal interpretation of an aADMG G, intervening on a set $X \subseteq V$ so as to change the natural causal mechanism of X amounts to modifying the right-hand side of the equations for the random variables in X. For simplicity, we only consider interventions that set variables to fixed values. Graphically, an intervention amounts to modifying G as

shown in Table 2. Line 1 is shared with an intervention on an original ADMG. Lines 2-4 are best understood in terms of the magnified aADMG G': They correspond to marginalizing the error nodes associated with the nodes in X out of G'_U, the UG that represents the correlation structure of the error nodes. In other words, lines 2-4 replace G'_U with $(G'_U)^{U \setminus U_X}$, the marginal graph of G'_U over $U \setminus U_X$. This makes sense since U_X is no longer associated with X due to the intervention and, thus, we may want to marginalize it out because it is unobserved. This is exactly what lines 2-4 imply. Note that the aADMG after the intervention and the magnified aADMG after the intervention represent the same separations over V (Peña, 2016, Theorem 9). It can be proven that this definition of intervention works as intended: If $f(v)$ is specified by Equations 3 and 4, then $f(v \setminus x|\widehat{x})$ satisfies the global Markov property with respect to the aADMG resulting from intervening on X in G (Peña and Bendtsen, 2017, Corollary 5).

It is worth mentioning that Equations 3 and 4 specify each node as a linear function of its parents with additive normal noise. The equations can be generalized to nonlinear or nonparametric functions as long as the noise remains additive normal. That is, $A = g(Pa_G(A)) + U_A$ for all $A \in V$, with $U \sim \mathcal{N}(0, \Lambda)$ such that $(\Lambda^{-1})_{U_A,U_B} = 0$ if $U_A - U_B$ is not in G'. That the noise is additive normal ensures that U_A is determined by $A \cup Pa_G(A)$, which is needed for Theorems 9 and 11 by Peña (2016) and Corollary 5 by Peña and Bendtsen (2017) to remain valid.

Finally, an assumption-free exact algorithm for learning aADMGs from observations and interventions via answer set programming exists (Peña, 2016).

3. Causal Effect Identification

In this section, we present a novel sound algorithm for identifying arbitrary causal effects from aADMGs. We show that the algorithm is also complete for the identification of the causal effect of a single variable on the rest of the variables. The algorithm is based on a decomposition of $f(v)$. We also show that the algorithm follows from a calculus of interventions.

3.1 Identification by Decomposition

Note that the system of structural equations corresponding to the causal model represented by an aADMG G induces a density function over $V = \{V_1, \ldots, V_n\}$, namely

$$f(v) = \int [\prod_i f(v_i|pa_G(V_i), u_i)] f(u)\, du. \tag{5}$$

Moreover, we say that two nodes belong to the same component if and only if they are connected by an undirected path in G. Assume that V is partitioned into components $S_1, \ldots, S_k$. Then, $f(u)$ factorizes as

$$f(u) = \prod_j f(u_{S_j}) \tag{6}$$

because, as mentioned above, $f(u)$ satisfies the global Markov property with respect to G'_U. When a density function can be written as in Equations 5 and 6, we say that it factorizes according to G.

The density function induced by the post-interventional system of structural equations can be obtained from Equation 5 by simply removing the terms for the variables intervened upon, that is

$$f(v \setminus x|\widehat{x}) = \int [\prod_{V_i \in V \setminus X} f(v_i|pa_G(V_i), u_i)] f(u)\, du$$
$$= \int [\prod_{V_i \in V \setminus X} f(v_i|pa_G(V_i), u_i)] f(u_{V \setminus X})\, du_{V \setminus X}. \quad (7)$$

Moreover, we define the factor $q(c)$ with $C \subseteq V$ as follows:

$$q(c) = \int [\prod_{V_i \in C} f(v_i|pa_G(V_i), u_i)] f(u)\, du = \int [\prod_{V_i \in C} f(v_i|pa_G(V_i), u_i)] f(u_C)\, du_C.$$

Note from the two previous equations that $q(c) = f(c|\widehat{v \setminus c})$. Note also that $q(c)$ factorizes according to G^C. The next lemma shows that $q(c)$ is identifiable if C is a component in G.[2]

Lemma 1 *Given an aADMG G, assume that V is partitioned into components $S_1, \ldots, S_k$. Then,*

$$f(v) = \prod_j q(s_j)$$

and

$$q(s_j) = \prod_{V_i \in S_j} f(v_i|v^{(i-1)})$$

where $V_1 < \ldots < V_n$ is a topological order of V with respect to G, and $V^{(i)} = \{V_1, \ldots, V_i\}$.

The following two lemmas show how certain factors are related. They will be instrumental later.

Lemma 2 *Given an aADMG G and two sets $E \subseteq C \subseteq V$ such that E is an ancestral set in G^C, then*

$$q(e) = \int q(c)\, d(c \setminus e).$$

Lemma 3 *Given an aADMG G, assume that a set $C \subseteq V$ is partitioned into components $C_1, \ldots, C_k$ in G^C. Then,*

$$q(c) = \prod_j q(c_j)$$

and

$$q(c_j) = \prod_{V_i \in C_j} \frac{q(c^{(i)})}{q(c^{(i-1)})}$$

where $V_1 < \ldots < V_n$ is a topological order of C with respect to G^C, and $C^{(i)} = \{V_1, \ldots, V_i\}$. Moreover,

$$q(c^{(i)}) = \int q(c)\, d(c \setminus c^{(i)}).$$

2. The proofs of the lemmas and theorems in the paper can be found in the supplementary material. Some proofs are adaptations of those by Tian and Pearl (2002a,b). We provide them for completeness.

The previous lemmas can be generalized as follows. Let $A \subseteq V$ be an ancestral set in G, and let $B = V \setminus A$. Given $C \subseteq B$, we define the factor $q(c|a)$ as follows:

$$q(c|a) = \int [\prod_{V_i \in C} f(v_i|pa_G(V_i), u_i)] f(u_B|u_A)\, du_B = \int [\prod_{V_i \in C} f(v_i|pa_G(V_i), u_i)] f(u_C|u_A)\, du_C.$$

Note that that A is an ancestral set in G implies that it determines U_A in the expression above. We show now that $q(c|a) = f(c|\widehat{b \setminus c}, a)$. Note that Equation 7 implies that

$$f(a, c|\widehat{v \setminus \{a, c\}}) = [\int [\prod_{V_i \in C} f(v_i|pa_G(V_i), u_i)] f(u_C|u_A)\, du_C] \prod_{V_i \in A} f(v_i|pa_G(V_i), u_i) f(u_A)$$

and thus

$$f(a|\widehat{v \setminus \{a, c\}}) = \prod_{V_i \in A} f(v_i|pa_G(V_i), u_i) f(u_A)$$

by marginalization in the previous equation and recalling that A is an ancestral set in G which implies that no node in A has a parent in C. Then, combining the two previous equations implies that

$$f(c|\widehat{v \setminus \{a, c\}}, a) = \frac{f(a, c|\widehat{v \setminus \{a, c\}})}{f(a|\widehat{v \setminus \{a, c\}})} = \int [\prod_{V_i \in C} f(v_i|pa_G(V_i), u_i)] f(u_C|u_A)\, du_C = q(c|a).$$

Note that $f(u_B|u_A)$ factorizes according to G'_{U_B} and, thus, $q(b|a)$ factorizes according to $H = G_B$. To see it, set $C = B$ in the previous equation. Then, $q(c|a)$ factorizes according to H^C.

Lemma 4 *Given an aADMG G and two disjoint sets $A, C \subseteq V$, then*

$$q(c|a) = \frac{q(a, c)}{\int q(a, c)\, dc}.$$

Moreover, if A is an ancestral set in $G^{A \cup C}$, then

$$q(c|a) = \frac{q(a, c)}{q(a)}.$$

The following three lemmas can be proven in much the same way as Lemmas 1-3.

Lemma 5 *Given an aADMG G and an ancestral set A in G, assume that $B = V \setminus A$ is partitioned into components $S_1, \ldots, S_k$ in $H = G_B$. Then,*

$$f(b|a) = \prod_j q(s_j|a)$$

and

$$q(s_j|a) = \prod_{V_i \in S_j} f(v_i|v^{(i-1)}, a)$$

where $V_1 < \ldots < V_n$ is a topological order of B with respect to H, and $V^{(i)} = \{V_1, \ldots, V_i\}$.

Lemma 6 *Given an aADMG G, an ancestral set A in G, and two sets $E \subseteq C \subseteq V \setminus A$ such that E is an ancestral set in $(G_{V\setminus A})^C$, then*

$$q(e|a) = \int q(c|a)\, d(c \setminus e).$$

Lemma 7 *Given an aADMG G and an ancestral set A in G, let $B = V \setminus A$, $H = G_B$ and assume that a set $C \subseteq B$ is partitioned into components $C_1, \ldots, C_k$ in H^C. Then,*

$$q(c|a) = \prod_j q(c_j|a)$$

and

$$q(c_j|a) = \prod_{V_i \in C_j} \frac{q(c^{(i)}|a)}{q(c^{(i-1)}|a)}$$

where $V_1 < \ldots < V_n$ is a topological order of C with respect to H^C, and $C^{(i)} = \{V_1, \ldots, V_i\}$. Moreover,

$$q(c^{(i)}|a) = \int q(c|a)\, d(c \setminus c^{(i)}).$$

We are now in the position to introduce our sound algorithm for identifying an arbitrary causal effect $f(y|\widehat{x})$ from an aADMG G. Let X' be a maximal subset of X such that, for any $V_1 \in X'$, there is a path $V_1 \to \ldots \to V_n$ in G such that $V_n \in Y$ and $V_2, \ldots, V_n \notin X'$. Note that $f(y|\widehat{x}) = f(y|\widehat{x'})$. Hereinafter, we assume without loss of generality that $X' = X$. Let $B = De_G(X)$ and $A = V \setminus B$. Note that A is an ancestral set in G. Let $Y_1 = Y \cap A$ and $Y_2 = Y \cap B$. Then,

$$\begin{aligned} f(y|\widehat{x}) = \int f(y_2, a|\widehat{x})\, d(a \setminus y_1) &= \int f(y_2|\widehat{x}, a) f(a|\widehat{x})\, d(a \setminus y_1) \\ &= \int f(y_2|\widehat{x}, a) f(a)\, d(a \setminus y_1) \end{aligned} \tag{8}$$

where the third equality follows from the fact that $A \cap De_G(X) = \emptyset$. Moreover,

$$f(y_2|\widehat{x}, a) = \int f(b \setminus x|\widehat{x}, a)\, d(b \setminus \{x, y_2\}) = \int q(b \setminus x|a)\, d(b \setminus \{x, y_2\}).$$

Let $C = An_{(G_B)^{B\setminus X}}(Y_2)$. Then by Lemma 6,

$$f(y_2|\widehat{x}, a) = \int \int q(b \setminus x|a)\, d(b \setminus \{x, c\})\, d(c \setminus y_2) = \int q(c|a)\, d(c \setminus y_2).$$

Assume that C is partitioned into components $C_1, \ldots, C_l$ in $(G_B)^C$. Then by Lemma 7,

$$f(y_2|\widehat{x}, a) = \int \prod_j q(c_j|a)\, d(c \setminus y_2). \tag{9}$$

Consequently, Equations 8 and 9 imply that $f(y|\widehat{x})$ is identifiable if $q(c_j|a)$ is identifiable for all j. Assume that B is partitioned into components $S_1, \ldots, S_k$ in G_B. Note that $C_j \subseteq S_i$ for some i, and recall that $q(s_i|a)$ is identifiable by Lemma 5, which implies that $q(c_j|a)$ is identifiable by Lemma 6 if C_j is an ancestral set in $(G_B)^{S_i}$. Table 3 summarizes the just described steps and the following theorem summarizes their correctness.

	Input: An aADMG G and two disjoint sets $X, Y \subseteq V$.
	Output: An expression to compute $f(y\|\widehat{x})$ from $f(v)$ or FAIL.
1	Let $B = De_G(X)$ and $A = V \setminus B$
2	Let $Y_1 = Y \cap A$ and $Y_2 = Y \cap B$
3	Assume that B is partitioned into components $S_1, \ldots, S_k$ in G_B
4	Let $C = An_{(G_B)^{B \setminus X}}(Y_2)$
5	Assume that C is partitioned into components $C_1, \ldots, C_l$ in $(G_B)^C$
6	For each C_j such that $C_j \subseteq S_i$ do
7	Compute $q(s_i\|a)$ by Lemma 5
8	If C_j is an ancestral set in $(G_B)^{S_i}$ then
9	Compute $q(c_j\|a)$ from $q(s_i\|a)$ by Lemma 6
10	Else return FAIL
11	Return $\int [\int \prod_j q(c_j\|a)\, d(c \setminus y_2)] f(a)\, d(a \setminus y_1)$ by Lemma 7

Table 3: Algorithm for causal effect identification from aADMGs.

Theorem 8 *Given an aADMG G and two disjoint sets $X, Y \subseteq V$, if the algorithm in Table 3 returns an expression for $f(y|\widehat{x})$, then it is correct.*

As an example, we run the algorithm in Table 3 to identify $f(v_2|\widehat{v_1})$ from the aADMG in Figure 1. Then, $X = V_1$ and $Y = V_2$. Thus, $B = \{V_1, V_2\}$ and $A = V_3$ in line 1, and $Y_1 = \emptyset$ and $Y_2 = V_2$ in line 2. Then, $S_1 = V_1$ and $S_2 = V_2$ in line 3. Then, $C = V_2$ in line 4 and, thus, $C_1 = V_2$ in line 5. Note that $C_1 \subseteq S_2$ and, thus, $q(v_2|v_3) = f(v_2|v_1, v_3)$ by lines 6-9. Therefore, the algorithm returns $\int f(v_2|v_1, v_3) f(v_3)\, dv_3$ which is the correct answer.

The algorithm in Table 3 is not only sound but also complete for identifying the causal effect of a single random variable on the rest, i.e. the algorithm identifies every such causal effect that can be computed uniquely from $f(v)$. The following theorem proves this result.

Theorem 9 *Given an aADMG G and an element $X \in V$, if the algorithm in Table 3 fails to return an expression for $f(v \setminus x|\widehat{x})$, then the expression does not exist.*

3.2 Identification by Calculus

An alternative to the algorithm in Table 3 consists in repeatedly applying the rules below which, together with standard probability manipulations, aim to transform the causal effect of interest into an expression that only involves observational quantities. The rules are sound (Peña and Bendtsen, 2017, Theorem 7). Given an aADMG G, let X, Y, Z and W be disjoint subsets of V. The rules are as follows:

- Rule 1 (insertion/deletion of observations):

$$f(y|\widehat{x}, z, w) = f(y|\widehat{x}, w) \text{ if } Y \perp_{G_{\overrightarrow{\overline{X}}}} Z|W$$

  where $G_{\overrightarrow{\overline{X}}}$ denotes the graph obtained from G by deleting all directed edges in and out of X.

- Rule 2 (intervention/observation exchange):

$$f(y|\widehat{x}, \widehat{z}, w) = f(y|\widehat{x}, z, w) \text{ if } Y \perp_{G_{\overrightarrow{\overline{X}}\overrightarrow{\underline{Z}}}} Z|W$$

where $G_{\overrightarrow{X}\overrightarrow{Z}}$ denotes the graph obtained from G by deleting all directed edges in and out of X and out of Z.

- Rule 3 (insertion/deletion of interventions):

$$f(y|\widehat{x},\widehat{z},w) = f(y|\widehat{x},w) \text{ if } Y \perp_{G_{\overrightarrow{X}\overrightarrow{\overline{Z(W)}}}} Z|W$$

where $Z(W)$ denotes the nodes in Z that are not ancestors of W in $G_{\overrightarrow{X}}$, and $G_{\overrightarrow{X}\overrightarrow{\overline{Z(W)}}}$ denotes the graph obtained from G by deleting all directed edges in and out of X and all undirected and directed edges into $Z(W)$.

We prove below that the algorithm in Table 3 actually follows from rules 1-3 and standard probability manipulations. To see it, note that all the steps in the algorithm involve standard probability manipulations except the application of Lemmas 5-7, which involve interventions. We prove below that these lemmas follow from rules 1-3. First, we prove that rule 1 is not really needed.

Lemma 10 *Rule 1 follows from rules 2 and 3.*

Lemma 11 *Lemmas 2 and 6 follow from rule 3.*

Lemma 12 *Lemmas 1, 3, 5 and 7 follow from rules 2 and 3.*

The following theorem summarizes the lemmas above.

Theorem 13 *Given an aADMG G and two disjoint sets $X, Y \subseteq V$, if the algorithm in Table 3 returns an expression for $f(y|\widehat{x})$, then it is correct. Moreover, the expression can also be obtained by repeated application of rules 2 and 3.*

4. Conclusions

Alternative ADMGs are mixed graphs that may allow causal effect identification in scenarios where Pearl's original ADMGs may not, and vice versa. Therefore, they complement each other. In this paper, we have shown that, as for the original ADMGs, it is possible to develop a sound algorithm for identifying arbitrary causal effects from alternative ADMGs. We have also shown that the algorithm is complete for identifying the causal effect of a single random variable on the rest. We have also shown that the algorithm follows from a calculus similar to Pearl's *do*-calculus.

In the future, we would like to extend the algorithm in this paper so that it becomes complete for identifying arbitrary causal effects. We would also like to evaluate how accurate the output of our algorithm is when the additive noise assumption does not hold. Finally, we would like to combine the original and alternative ADMGs into a family of mixed graphs with three types of edges, and develop a sound and complete algorithm for causal effect identification from them.

References

S. A. Andersson, D. Madigan, and M. D. Perlman. Alternative Markov Properties for Chain Graphs. *Scandinavian Journal of Statistics*, 28:33–85, 2001.

P. Bühlmann, J. Peters, and J. Ernest. CAM: Causal Additive Models, High-Dimensional Order Search and Penalized Regression. *Annals of Statistics*, 42:2526–2556, 2014.

D. R. Cox and N. Wermuth. *Multivariate Dependencies - Models, Analysis and Interpretation*. Chapman & Hall, 1996.

J. T. A. Koster. Marginalizing and Conditioning in Graphical Models. *Bernoulli*, 8:817–840, 2002.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

J. M. Mooij, J. Peters, D. Janzing, J. Zscheischler, and B. Schölkopf. Distinguishing Cause from Effect Using Observational Data: Methods and Benchmarks. *Journal of Machine Learning Research*, 17:1–102, 2016.

J. M. Peña. Alternative Markov and Causal Properties for Acyclic Directed Mixed Graphs. In *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence*, pages 577–586, 2016.

J. M. Peña and M. Bendtsen. Causal Effect Identification in Acyclic Directed Mixed Graphs and Gated Models. *International Journal of Approximate Reasoning*, 2017.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf. Causal Discovery with Continuous Additive Noise Models. *Journal of Machine Learning Research*, 15:2009–2053, 2014.

T. Richardson. Markov Properties for Acyclic Directed Mixed Graphs. *Scandinavian Journal of Statistics*, 30:145–157, 2003.

T. Richardson and P. Spirtes. Ancestral Graph Markov Models. *The Annals of Statistics*, 30:962–1030, 2002.

K. Sadeghi and S. L. Lauritzen. Markov Properties for Mixed Graphs. *Bernoulli*, 20: 676–696, 2014.

I. Shpitser and J. Pearl. Identification of Conditional Interventional Distributions. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, pages 437–444, 2006.

J. Tian and J. Pearl. A General Identification Condition for Causal Effects. In *Proceedings of the 18th National Conference on Artificial Intelligence*, pages 567–573, 2002a.

J. Tian and J. Pearl. On the Identification of Causal Effects. *Technical report R-290-L, Department of Computer Science, University of California, Los Angeles*, 2002b.