Estimating Bounds on Causal Effects Considering Unmeasured Common Causes

Sebastián Bejos SEBASTIAN.BEJOS@INAOEP.MX
Luis Enrique Sucar ESUCAR@INAOEP.MX
Eduardo F. Morales EMORALES@INAOEP.MX
Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México

Editors: J.H.P. Kwisthout & S. Renooij

Abstract

Maximal ancestral graphs (MAGs) can represent causal relationships in systems that include unmeasured common direct causes. Constraint-based causal discovery methods are able to find solely the Markov Equivalence Class (MEC) of the causal structure given a set of observational data. To bound the total effect estimation between a pair of variables, when the MEC of the causal structure is known, the causal effect on each member in the MEC are computed, while keeping the minimum and maximum values as the lower and upper bounds for the total causal effect. However, when the modeling is done using MAGs, i.e., the MEC is encoded as a Partial Ancestral Graph (PAG), it is not always possible to find an adjustment set over some pairs of variables for the computation of the causal effect by covariance adjustment. In such cases, the LV-IDA algorithm returns missing values on the causal effects computation for some, and occasionally all, of the MAGs in the PAG. We present an extension of the LV-IDA algorithm, which we call the LV-IDA+ algorithm, that can compute approximated bounds of causal effects between every pair of the variables on a PAG. To achieve this, we propose a way to approximate the causal effect estimations when it is not possible to find adjustment sets for some pairs of variables on the MAGs in a PAG. We evaluate the performance of LV-IDA+ using simulated data generated by a canonical DAGs and compare with the LV-IDA algorithm. The results suggest the approximations of causal effects computed by LV-IDA+, are better than the missing values (simple NAs) returned by the LV-IDA algorithm, at least for the case of observational data generated by a canonical DAGs with latent variables.
Keywords: Causal Graphical Models, Causal Effect Estimation, Causally Insufficient Systems, Maximal Ancestral Graphs, Gaussian Linear Models, Unmeasured Common Causes.

1. Introduction

The causal effect between a pair of variables (X, Y) it is a measure of how much the variable Y is modified when manipulating the variable X. The prevailing mean of estimating such effects is using randomized controlled experiments, in which the treatment variable (the cause) is randomized while the outcome variable (the effect) is passively observed. Nevertheless, experimental data are not always available given that randomized controlled experiments can be unethical, infeasible, time consuming, or expensive. On the other hand, observational data, i.e., data associated with processes that cannot be reproduced and are therefore not appropriate for conducting controlled experiments, are often abundant. In this research, we consider the problem of estimating the causal effects between pairs of variables given only observational data. Causal effect estimation is notoriously hard to calculate from this kind of data, but recently there has been much interest to this problem (see Montero-Hernandez et al. (2018); Perković et al. (2018); van der Zander et al. (2019)). However, most works consider the assumption of causal sufficiency. This assumption stipulates that no variables which are common direct causes of at least two measured variables are unmeasured. In this paper we do not consider this assumption, which is often not realistic in applied contexts.

The general approach we use to solve this problem is the covariate adjustment method in the context of causal graphical models. This method can be divided in three parts: (i) First, a causal structure represented by a graph is estimated, where the vertices are the variables in the system and the edges indicate direct causal relationships between them. (ii) Then, the structure of the causal graph is used to find a set of variables, called the adjustment set, which is a sufficient set, along with the treatment variable, to compute the needed interventions. (iii) Lastly, using an adjustment set, the interventions required to estimate total causal effect can be calculated (see Section 3 for details).

Our contribution deals with the second and third part of the covariant adjustment method. Concerning the first part, we consider that the causal structure is estimated by a constraint-based algorithm. These algorithms consider conditional dependencies on the observational distribution, to infer the causal directed acyclic graph (DAG) that generated the data. Unfortunately, multiple DAGs can encode the same set of conditional independence relationships so these methods can exclusively find an equivalence class, called the Markov Equivalence Class (MEC), of the underlying causal structure given a set of observational data (see Glymour et al. (2019)). If the assumption of causal sufficiency is relaxed, then the causal structure is represented as a Maximal Ancestral Graph (MAG). Under this scenario, a constraint-based algorithms, such as the Fast Causal Inference algorithm (FCI), will estimate a Partial Ancestral Graph (PAG), which encodes a MEC of MAGs (see Zhang (2008b); Ogarrio et al. (2016)).

Regarding the search for adjustment sets, in Maathuis and Colombo (2015); Perković et al. (2018) and van der Zander et al. (2019) are defined graphical criterion over DAGs and MAGs for finding sets of variables that can be used in order to estimate the total causal effect from observational data. In particular, in Perković et al. (2018) and van der Zander et al. (2019), sound and complete algorithms are shown for constructing sets that satisfy their criterion for MAGs. These results play a fundamental role for the causal effects estimations in the algorithm we propose.

Since by means of a constraint-based algorithm it is not possible to specify a single causal structure $\mathcal{M}$ but a summary structure $\mathcal{P}$ that represents the MEC to which $\mathcal{M}$ belongs. It is possible to unfold $\mathcal{P}$ to list all the structures represented by $\mathcal{P}$ and calculate the causal effect among a variables pair (X, Y) in each of the causal structures in $\mathcal{P}$, performing the second and third part of the covariante adjustment method. Under this framework, we can bound the real causal effect between a pair of variables (X, Y) as the minimum and maximum of the effects calculated in each of the models in $\mathcal{P}$. This framework to bound total causal effects in causal sufficient system, i.e., considering that $\mathcal{M}$ is given as a DAG and $\mathcal{P}$ as a Completed Partially Directed Acyclic Graph, was proposed in Maathuis et al. (2009) as the IDA ("Intervention when the DAG is Absent") algorithm. The IDA algorithm was extended for insufficient system in Malinsky and Spirtes (2017), where the LV-IDA (Latent Variables

IDA) algorithm was proposed. This algorithm works with MAGs as the underlying causal structures, $\mathcal{M}$, and PAGs as their summary structure, $\mathcal{P}$, allowing the possibility of latent confounders in the modeling.

In this paper, we propose an algorithm based on the IDA and LV-IDA algorithms to calculate the bounds of the causal effects between any pair of variables (X, Y) in a insufficient system, i.e., considering non-measured direct causes. As in IDA and LV-IDA, we consider the case that the system $\boldsymbol{V} = \{X_1, \ldots, X_p\}$ is jointly Gaussian, and as in LV-IDA, we use MAGs and PAGs, to represent the causal structure and the MEC, respectively, of the system. The main difference of the LV-IDA algorithm with respect to the proposed algorithm, LV-IDA+, is that LV-IDA+ always guarantees the calculation of the causal effect between any pair of variables in the system. LV-IDA cannot always calculate this effect for some pairs of variables and then occasionally returns missing values as output. This is due to the fact that in these cases there is no adjustment set for the pair of variables (X, Y) in some, and occasionally all, MAGs in the PAG. Our main contribution proposes a way to approximate the causal effect when these degenerate cases are presented. The LV-IDA+ algorithm uses the adjustment sets of the canonical DAGs associated with the MAGs in the PAG to compute the bounds on the causal effects in these special cases. Experimental results over synthetic models show a higher accuracy in the estimation of the bounds on causal effects using the proposed approach compared to LV-IDA, in particular when LV-IDA returns NAs.

After discussing the preliminaries on MAGs and its relation with causal DAGs in the next section; in Section 3, we review the needed concepts on interventions calculus and covariate adjustment over Causal Bayesian Networks in the context of the IDA framework, upon which we base this research. Following this, we give the description of the proposed method in Section 4, and present the experimental setup and the simulations results in Section 5.

2. Preliminaries

A directed mixed graph $\mathcal{M}$ is a graph $(\boldsymbol{V}, \boldsymbol{E})$ that may contain two kinds of edges: directed edges ($\rightarrow$) and bi-directed edges ($\leftrightarrow$).[1] Between any two vertices there is at most one edge. The two ends of an edge are called marks and there are two kinds of marks: arrowhead ($>$) and tail ($-$). Two vertices are said to be adjacent in a graph if there is an edge between them. If $X \rightarrow Y$ in $\mathcal{M}$ then X is a parent of Y and $X \in \boldsymbol{pa}(Y)$. A vertex X is said to be an ancestor of a vertex Y, $X \in \boldsymbol{an}(Y)$, if either there is a directed path $X \rightarrow \cdots \rightarrow Y$ from X to Y, or $X = Y$. A mixed directed graph is an ancestral graph if: (a) there are no directed cycles (b) whenever there is an edge $X \leftrightarrow Y$, then there is no directed path from X to Y, or from Y to X. In an ancestral graph, a nonendpoint vertex X on a path is said to be a collider if two arrowheads meet at X. All other nonendpoint vertices on a path are noncolliders. A path along which every nonendpoint is a collider is called a collider path. An inducing path π relative to a set $\boldsymbol{L}$, between vertices X and Y in an ancestral graph $\mathcal{G}$, is a path on which every nonendpoint vertex not in $\boldsymbol{L}$ is both a collider on π and an ancestor of at least one of the endpoints, X and Y. An ancestral graph is called a Maximal Ancestral Graph (MAG)

1. We use the notation convention that all sets in the article appear in bold and reserve calligraphic font for graphics, e.g., $\mathcal{G}$ is a graph and $\boldsymbol{E}$ is a set of edges.

if there is no inducing path between any two non-adjacent vertices (see Zhang (2008a)). As a probabilistic model, the vertices of a MAG $\mathcal{M}$ represent random variables, and the structure of $\mathcal{M}$ is interpreted as encoding a set of conditional independence relations by the m-separation graphical criterion (see Zhang (2008a,b)), which generalizes the well-known d-separation criterion for Directed Acyclic Graphs (DAGs).

A causal model is a pair $(\mathcal{M}, \boldsymbol{\Phi_{\mathcal{M}}})$ where $\mathcal{M}$ is a graph that is called the causal structure and $\boldsymbol{\Phi}_{\mathcal{M}}$ is the set of parameter of the model. The set of parameter $\boldsymbol{\Phi}$ is given as set of functional relationship among the variables which are known as Structural Equations (SE). This set of SEs $\boldsymbol{\Phi_{\mathcal{M}}}$ assign a function $X_i = f_i(\boldsymbol{pa}(X_i), \epsilon_i)$ to each $X_i \in \boldsymbol{V}$ and a probability distribution $P(\epsilon_i)$ to each ϵ_i, where ϵ_i is a random disturbance distributed according to $P(\epsilon_i)$, independently of all other ϵ_j with $i \neq j$ (see Pearl (2009)). A special case, important for this work, is when the f_i functions are linear and the errors ϵ_i are Gaussian distributed random variables.

It is common to assume that the set of variables $\boldsymbol{V}$, also referred to as the system, is causally sufficient and therefore, the causal structure of the system is represented by a DAG $\mathcal{D}$, i.e., an ancestral graph $(\boldsymbol{V}, \boldsymbol{E})$ with no bi-directed edges, where the vertices in $\boldsymbol{V}$ corresponds to a set of random variables $\boldsymbol{V} = \{X_1, \ldots, X_p\}$ and each edge in $\boldsymbol{E}$ represents a direct functional relationship among the corresponding variables, which is expressed by saying that X_i is the direct cause of X_j for an edge $(X_i, X_j) \in \boldsymbol{E}$. The system $\boldsymbol{V}$ is said to be causally sufficient if there are no variables in $\boldsymbol{V}$ which are common direct causes of at least two measured variables that are unmeasured. If some of the variables in the set $\boldsymbol{V}$ are unmeasured, $\boldsymbol{V}$ can be partitioned as $\boldsymbol{V} = \boldsymbol{O} \cup \boldsymbol{L}$, where $\boldsymbol{O}$ are the observed (measured) variables and $\boldsymbol{L}$ are the latents variables (unmeasured). MAGs can represent conditional independence information and causal relationships in DAGs that include unmeasured (hidden or latent) variables. A MAG represents a DAG after all latent variables have been marginalized out, and it preserves all entailed conditional independence relations among the measured variables (see Zhang (2008a)). This is, given any DAG $\mathcal{D}$ over $\boldsymbol{V} = \boldsymbol{O} \cup \boldsymbol{L}$ there is a MAG $\mathcal{M}$ over $\boldsymbol{O}$ alone, such that for any disjoint sets $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{O}$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated by $\boldsymbol{Z}$ in $\mathcal{D}$ (See Pearl (2009)) if and only if they are m-separated by $\boldsymbol{Z}$ in the MAG $\mathcal{M}$. The following construction gives us such a MAG: (i) For each pair of variables $X, Y \in \boldsymbol{O}$, X and Y are adjacent in $\mathcal{M}$ if and only if there is an inducing path between them relative to $\boldsymbol{L}$ in $\mathcal{D}$. (ii) For each pair of adjacent variables X, Y in $\mathcal{M}$: (a) orient the edge as $X \rightarrow Y$ in $\mathcal{M}$ if X is an ancestor of Y in $\mathcal{D}$; (b) orient it as $X \leftarrow Y$ in $\mathcal{M}$ if Y is an ancestor of X in $\mathcal{D}$; (c) orient it as $X \leftrightarrow Y$ in $\mathcal{M}$, otherwise. (see Figure 1a for an example of this construction).

On the other hand, if $\mathcal{M}$ is a MAG with vertex set $\boldsymbol{V}$, then we define the canonical DAG, $\mathcal{D}(\mathcal{M})$ associated with $\mathcal{M}$ as follows (see Richardson and Spirtes (2002) and Figure 1b): (i) Let $\boldsymbol{L}_{\mathcal{D}(\mathcal{M})} = \{\lambda_{XY} \mid X \leftrightarrow Y \text{ in } \mathcal{M}\}$. (ii) DAG $\mathcal{D}(\mathcal{M})$ has vertex set $\boldsymbol{V} \cup \boldsymbol{L}_{\mathcal{D}(\mathcal{M})}$ and edge set defined as:

$$\text{If } \left\{ \begin{array}{c} X \rightarrow Y \\ X \leftrightarrow Y \end{array} \right\} \text{in } \mathcal{M} \text{ then } \left\{ \begin{array}{l} X \rightarrow Y \\ X \leftarrow \lambda_{XY} \rightarrow Y \end{array} \right\} \text{in } \mathcal{D}(\mathcal{M}).$$

Several MAGs can also encode the same conditional independencies via m-separation. Such MAGs form a Markov Equivalence Class (MEC) which can be described uniquely by a partial ancestral graph (PAG). Let $[\mathcal{M}]$ be the MEC of an arbitrary MAG $\mathcal{M}$. The PAG

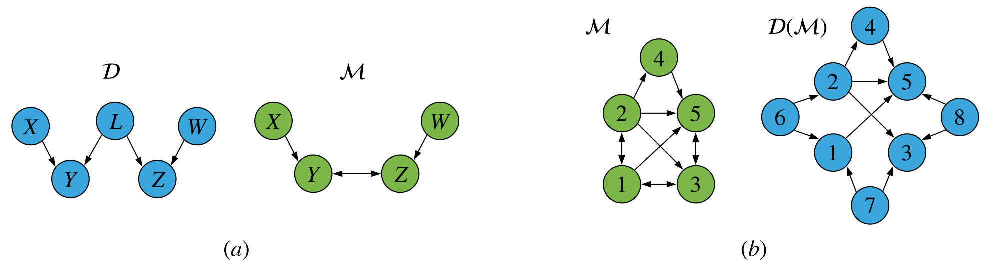

Figure 1: (a) The construction of a MAG $\mathcal{M}$ (green) over $\boldsymbol{O} = \{X, Y, Z, W\}$, from a DAG $\mathcal{D}$ (blue) over the set of variables $\boldsymbol{O} \cup \boldsymbol{L}$, with $\boldsymbol{L} = \{L\}$. (b) The canonical DAG $\mathcal{D}(\mathcal{M})$ (blue) associated with the MAG $\mathcal{M}$ (green), with $\boldsymbol{V} = \{1, 2, 3, 4, 5\}$ and $\boldsymbol{L}_{\mathcal{D}(\mathcal{M})}]\{\lambda_{12} = 6, \lambda_{13} = 7, \lambda_{35} = 8\}$. (Best seen in color.)

for $[\mathcal{M}]$, $\mathcal{P}_{[\mathcal{M}]}$, is a partial mixed graph with possibly three kinds of mark: arrowhead ($>$), tail ($-$) or a circle ($\circ$), such that (i) $\mathcal{P}_{[\mathcal{M}]}$ has the same adjacencies as $\mathcal{M}$ (and any member of $[\mathcal{M}]$) does; (ii) Every non-circle mark in $\mathcal{P}$ is an invariant mark in $[\mathcal{M}]$, i.e., a mark of arrowhead is in $\mathcal{P}_{[\mathcal{M}]}$ if and only if it is shared by all MAGs in $[\mathcal{M}]$; a mark of tail is in $\mathcal{P}_{[\mathcal{M}]}$ if and only if it is shared by all MAGs in $[\mathcal{M}]$; and a mark of circle is in $\mathcal{P}_{[\mathcal{M}]}$, otherwise. In Figure 2 we show a PAG and the MEC of MAGs it encodes.

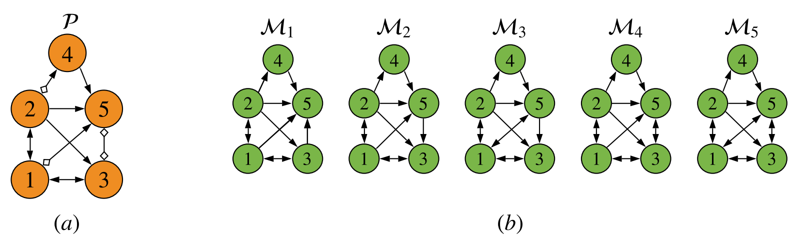

Figure 2: (a) A PAG $\mathcal{P}$ (orange) of five variables with four circle marks representing the MEC $[\mathcal{M}] = \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5\}$. (b) The five MAG $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5 \in [\mathcal{M}]$.

3. Covariate Adjustment and the IDA Framework

A causal graphical model over a set of variables $\boldsymbol{V} = \{X_1, \ldots, X_p\}$ can be represented as a causal Bayesian network (CBN), i.e., as a pair $(\mathcal{G}, f)$, where $\mathcal{G}$ is the causal structure and f is the joint density for $\boldsymbol{V}$ that factorizes as $f(\boldsymbol{V}) = \prod_{i=1}^{p} f(X_i \mid \boldsymbol{pa}(X_i))$, where these factors act as the set of parameters $\Phi_{\mathcal{G}}$ of the causal model. A density f is consistent with a DAG $\mathcal{D}$ if the pair $(\mathcal{D}, f)$ forms a CBN and f is consistent with a MAG $\mathcal{M}$ if there exists a CBN $(\mathcal{D}, g)$ such that $\mathcal{M}$ represents $\mathcal{D}$ and f is the observed marginal density of g. Given a CBN, it is possible to derive post-intervention densities. In particular, we are interested in interventions that set $\boldsymbol{X}$ to $\boldsymbol{x}$ uniformly in the population, which are denoted using the do-calculus as $do(\boldsymbol{X} = \boldsymbol{x})$ or shorthand $do(\boldsymbol{x})$, with $\boldsymbol{X} \subset \boldsymbol{V}$ (see Pearl (2009)). For this kind of interventions the post-intervention densities are given by the so called truncated factorization formula:

$$f(\boldsymbol{v} \mid do(\boldsymbol{x})) = \begin{cases} \prod_{\{i \mid X_i \in \boldsymbol{V} \setminus \boldsymbol{X}\}} f(x_i \mid \boldsymbol{pa}(x_i)), & \text{if } \boldsymbol{v} \text{ is conssitent with } \boldsymbol{x}, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathcal{G} = (\boldsymbol{V}, \boldsymbol{E})$ be a DAG or a MAG, and let $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{Z}$ be pairwise disjoint subsets of the set of continuous random variables $\boldsymbol{V}$, with $\boldsymbol{X} \neq \emptyset$ and $\boldsymbol{Y} \neq \emptyset$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ represents the manipulated and outcome variables, respectively. The set of variables $\boldsymbol{Z}$ is an adjustment set relative to $(\boldsymbol{X}, \boldsymbol{Y})$ in $\mathcal{G}$ if for any density f consistent with G we have that:

$$f(\boldsymbol{y} \mid do(\boldsymbol{x})) = \begin{cases} f(\boldsymbol{y} \mid \boldsymbol{x}), & \text{if } \boldsymbol{Z} = \emptyset \\ \int_{\boldsymbol{Z}} f(\boldsymbol{y} \mid \boldsymbol{xz}) f(\boldsymbol{z}) d\boldsymbol{z}, & \text{otherwise.} \end{cases} \quad (1)$$

Observe that adjustment sets allow post-intervention densities involving the do-operator to be identified, i.e., expressed only as specific functions of conditional densities which can be estimated from observational data, so the search for adjustment sets is fundamental for the computation of causal effects. The use of Equation 1 for computing total causal effects over a CBN is what is called the covariate adjustment method. For the particular case of computing the covariate adjustment of pairs of single variables X and Y over multivariate Gaussian densities, (case in which we are interested in this research) we can use the fact that this kind of densities are fully defined by expectations, and for expressing conditional independencies $P(Y \mid x, \boldsymbol{z}) = P(Y \mid \boldsymbol{z})$ we can use equivalently conditional expectations $E(Y \mid x, \boldsymbol{z}) = E(Y \mid \boldsymbol{z})$. Furthermore, since conditional expectations are linear in a multivariate Gaussian distribution, the substitution of probabilities by expectations, allows to use regression for estimate $E(Y \mid x, \boldsymbol{z})$ as $E(Y \mid x, \boldsymbol{z}) = \alpha + \beta x + \gamma^T \boldsymbol{z}$, for some $\alpha, \beta \in \mathbb{R}$ and $\gamma \in \mathbb{R}^{|\boldsymbol{z}|}$. The total causal effect of X on Y for continuous random variables setting is defined as $\frac{\partial}{\partial x} E(Y \mid do(x))$ (see Maathuis et al. (2009)), so the total causal effect of X on Y in this setting is β, that is, the coefficient of X in the regression of Y on X and the adjustment set $\boldsymbol{Z}$, since by the adjustment set definition, we have that

$$E(Y \mid do(x)) = \int_{\boldsymbol{z}} E(Y \mid x, \boldsymbol{z}) f(\boldsymbol{z}) d\boldsymbol{z} = \alpha + \beta x + \gamma^T E(\boldsymbol{Z}).$$

Causal structure learning methods for CBNs consider conditional independencies in the observational distribution, to reconstruct the causal structure that generated the data. Unfortunately, this does not solve completely the problem of learning the causal structure, because the same set of conditional independencies relationships can be encoded by several DAGs of MAGs (see Zhang (2008b)). The majority of causal structure learning methods estimate graphical representations of such Markov Equivalence Class (MEC). If we want to estimate the total causal effect of X on a response variable Y, when only a representation of the MEC of the causal structure is known, as is the case after a constraint-based causal structure learning methods are powered with only observational data, Maathuis et al. (2009) proposed the "Intervention-calculus when the DAG is Absent" (IDA) algorithm. The general idea of the IDA algorithm is as follows: After estimating a representation of the MEC given as a Completed Partial DAG (CPDAG), list all DAGs in the MEC and then apply covariate adjustment for each DAG, yielding an estimated total causal effect of X

on Y for each possible DAG. All these total causal effects, one for each DAG in the MEC, are collected in a multiset $\hat{\mathbf{\Theta}}$, and the minimum and maximum value in $\hat{\mathbf{\Theta}}$ are returned as bounds estimators for the true causal effect.

The conceptual framework of the IDA algorihm was extended to insufficient causal systems by Malinsky and Spirtes in (Malinsky and Spirtes, 2017). They named their algorithm LV-IDA (Latent Variables IDA), where they worked with PAGs as representations of MECs, and used the results in Maathuis and Colombo (2015) to found adjustment set in MAGs for the covariate adjustment computations. Listing all the MAGs represented by a PAG is more complicated than listing all the DAGs represented by a CPDAG. The naive approach would be to exhaustively try every combination for substitute the circle marks and then checks by the m-separation criteria if the resulting graph is Markov equivalent to the starting graph. An important contribution from the LV-IDA algorithm that we will use in our proposed algorithm, is the procedure for listing all the MAGs. This procedure is based on the transformational characterization of equivalence between MAGs introduced in Zhang and Spirtes (2005). We have been talking about the global versions of IDA and LV-IDA. Nonetheless, it is worth mentioning that local versions of these algorithms were also defined in Maathuis et al. (2009) and Malinsky and Spirtes (2017), respectively. Local versions work on the principle that only certain information about DAGs or MAGs in the MEC needs to be known, and this is enough to determine adjustment sets for the desired pairs of variables. When we mention LV-IDA we refer to its global version. The main motivation of this paper is to approximate the estimates of causal effects when it is not possible to find adjustment sets on a MAG for a pair of variables. Finally, we specify that both, IDA and LV-IDA algorithms, assume a Gaussian linear parameterization of the causal model, as we will also assume.

4. The LV-IDA+ Algorithm

A mayor drawback of the LV-IDA algorithm is that it is not always possible to find an adjustment set for some pairs of nodes in some MAGs to perform covariate adjustment. In such cases, the estimated multisets of total causal effects $\hat{\mathbf{\Theta}}$ contains missing values and the fundamental idea of using the minimum and maximum value in $\hat{\mathbf{\Theta}}$ to bound the true causal effect, it is not longer valid. We refer to this kind of missing values as simple NAs. More over, in some cases there are only missing values in $\hat{\mathbf{\Theta}}$ for some pairs of variables. We call these more problematic types of missing values, extreme NAs. Following the same principles of the IDA and LV-IDA algorithms, the proposed LV-IDA+ algorithm presents a way to approximate the calculation of the total causal effect when it is not possible to find an adjustment sets on the MAGs that belong to the MEC for some pairs of variables.

A schematic representation of the operation of the LV-IDA+ algorithm is shown in Figure 3. The LV-IDA+ algorithm, shown in Algorithm 1 computes a matrix of estimated causal effect intervals: $CE^* = \left([\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij}\right)_{p\times p}$, where the interval $[\hat{\theta}_{min}, \hat{\theta}_{max}]$ is calculated for each pair (i, j) of the p variables in the system, i.e., $i, j = 1 \dots p$, and $\hat{\theta}_{min}$, $\hat{\theta}_{max}$ are the lower, upper bounds, respectively, for the estimated causal effect. After listing all MAGs in the MEC encoded as a PAG (Line 1 in the algorithm), at first, the LV-IDA+ algorithm computes a matrix $CE_{\mathcal{M}_w} = (\hat{\theta}_{ij}^{\mathcal{M}_w})_{p\times p}$ for each of the k MAGs in the MEC (Lines 7, 8, 9, 11), where each total casual effect $\hat{\theta}_{ij}^{\mathcal{M}_w}$ is estimated by covariate adjustment

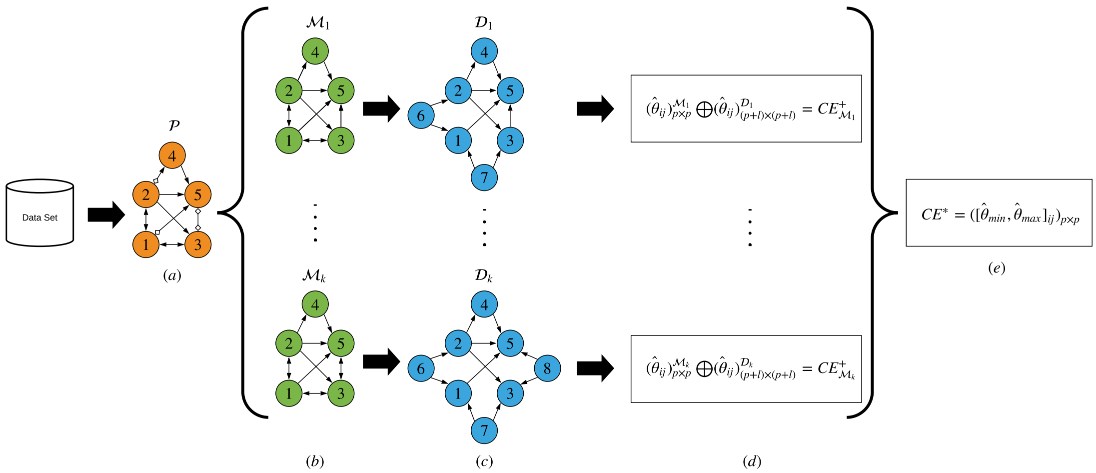

Figure 3: The LV-IDA+ framework. (a) A PAG $\mathcal{P}$ (orange) learned from data. (b) The set of MAGs (green) in the Markov equivalences class $[\mathcal{M}]$ encoded by $\mathcal{P}$. (c) The set of associated canonical DAGs (blue) from each of the MAGs in $[\mathcal{M}]$. (d) $CE_{\mathcal{M}_w} = (\hat{\theta}_{ij})^{\mathcal{M}_w}_{p\times p}$ and $CE_{\mathcal{D}_w} = (\hat{\theta}_{ij})^{\mathcal{D}_w}_{(p+l_w)\times(p+l_w)}$ are the square matrices of causal effects estimated from the MAG $\mathcal{M}_w$, and from the canonical DAG $\mathcal{D}_w$, respectively, where $w = 1 \dots k$, p is the number of variables and l_w is the number of latent variables in the $\mathcal{D}_w$ DAG. Note that the $CE_{\mathcal{M}_w}$ matrices may have missing values but the $CE_{\mathcal{D}_w}$ matrices not. We denote by $\oplus$, the operation of substitute all the NA values in a $CE_{\mathcal{M}_w}$ matrix by the corresponding values in the $CE_{\mathcal{D}_w}$ matrix. (e) $CE^* = ([\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij})_{p\times p}$ denote the square matrix of interval of causal effect, where the extremes in the intervals $[\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij}$ are obtained by getting the minimum $\hat{\theta}_{min}$ and maximum $\hat{\theta}_{max}$, from the $\hat{\theta}_{ij}$ estimations in the $CE^{+}_{\mathcal{M}_w} = (\hat{\theta}_{ij})_{p\times p}$ matrices. (Best seen in color.)

with adjustment sets found using the sound and complete algorithms in van der Zander et al. (2019) and Perković et al. (2018). These algorithms guarantee finding adjustment sets for the variables of interest on MAGs when they exist. However, when this is not the case, the matrices $CE_{\mathcal{M}_w} = (\hat{\theta}^{\mathcal{M}_w}_{ij})_{p\times p}$ obtain NA values. To approximate these missing values, LV-IDA+ generates the canonical DAGs associated with each of the k MAGs in the MEC (Lines 3, 4). Then, it learns the parameters for each of these canonical DAGs and calculate the matrix of causal effects $CE_{\mathcal{D}_w} = (\hat{\theta}^{\mathcal{D}_w}_{ij})_{(p+l_w)\times(p+l_w)}$ for each of them, where l_w stands for the number of latent variables in the canonical DAG $\mathcal{D}_w$ (Lines 13, 14). It is always possible to find an adjustment set in a DAG, i.e., calculate a value for the estimation of the causal effect between a pair of variables. However it is important to note that even though some of the parameters can be learned from the original data, in the construction of the canonical DAG $\mathcal{D}_w$ for each MAG $\mathcal{M}_w$, hidden variables are introduced for which there are no data. This issue is solved by Expectation Maximization (EM) techniques to learn the parameters in Gaussians DAGs given the structure of the model in the presence of hidden variables (see Sucar (2015)). With matrices $CE_{\mathcal{M}_w}$ and $CE_{\mathcal{D}_w}$ computed, the LV-IDA+ algorithm obtains a matrix $CE^{+}_{\mathcal{M}_w}$ of size $p \times p$ (the same dimension as matrix $CE_{\mathcal{M}_w}$) for each MAG in the PAG by combining the matrices $CE_{\mathcal{M}_w}$ and $CE_{\mathcal{D}_w}$, in such a way that if $\hat{\theta}_{ij}^{\mathcal{M}_w}$ is NA in $CE_{\mathcal{M}_w}$ it is replaced by the value of $\hat{\theta}_{ij}^{\mathcal{D}_w}$ in $CE_{\mathcal{D}_w}$, and the $\hat{\theta}_{ij}^{\mathcal{M}_w}$ values of the matrix $CE_{\mathcal{M}_w}$ are kept in $CE_{\mathcal{M}_w}^{+}$ in any other case (Line 17). Finally, the matrix of estimated causal effect intervals $CE^* = ([\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij})_{p \times p}$, i.e., the output of the LV-IDA+ algoritm, is computed by recording the minimum $(\hat{\theta}_{min})_{ij}$ and maximum $(\hat{\theta}_{max})_{ij}$ values for each of the k MAGs in the matrices $CE_{\mathcal{M}_w}^{+}$ (Line 19 of the algorithm).

The rationale behind the LV-IDA+ algorithm is that it is reasonable to calculate the causal effect on one of the DAGs that belongs to the set of DAGs represented by the MAG as an approximation to the causal effect on a pair of variables when this cannot be calculated directly on the MAG in question. The canonical DAG has the advantage that it can be efficiently built from a MAG and this type of DAGs encompasses the most significant aspects of the causal structure of an infinity of DAGs represented by a MAG (see Appendix B for more details).

Algorithm 1: LV-IDA+

Input: A PAG $\mathcal{P}$, and a set of observational data D, for p variables
Output: A matrix of interval of causal effects $CE^* = \left([\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij}\right)_{p \times p}$

1 Determine all MAGs $\mathcal{M}_1, \dots \mathcal{M}_k$ in the MEC $[\mathcal{M}]$ encoded as $\mathcal{P}$
2 **for** $w = 1, \dots, k$ **do**
3   Compute de canonical DAG $\mathcal{D}_w$ from $\mathcal{M}_w$
4   Learn the parameters of $\mathcal{D}_w$ from data D using EM for the l_w hidden variables in $\mathcal{D}_w$
5   **for** $i = 1, \dots, p$ **do**
6   **for** $j = 1, \dots, p$ **do**
7   Compute an adjustment set $\boldsymbol{Z}_{(i,j)}^{\mathcal{M}_w}$ for the pair of variables (i, j) in $\mathcal{M}_w$
8   **if** an adjustment set $\boldsymbol{Z}_{(i,j)}^{\mathcal{M}_w}$ exists **then**
9                 $\hat{\theta}_{ij}^{\mathcal{M}_w} \leftarrow \beta$, where β is the coefficient of i in the regression $j = \alpha + \beta i + \gamma^T E(\boldsymbol{Z}_{(i,j)}^{\mathcal{M}_w})$
10   **else** $\hat{\theta}_{ij}^{\mathcal{M}_w} \leftarrow$ NA
11
12   **end**
13   Compute an adjustment set $\boldsymbol{Z}_{(i,j)}^{\mathcal{D}_w}$ for the pair of variables (i, j) in $\mathcal{D}_w$
14            $\hat{\theta}_{ij}^{\mathcal{D}_w} \leftarrow \beta$, where β is the coefficient of i in the regression $j = \alpha + \beta i + \gamma^T E(\boldsymbol{Z}_{(i,j)}^{\mathcal{D}_w})$
15   **end**
16   **end**
17    $CE_{\mathcal{M}_w}^{+} \leftarrow CE_{\mathcal{M}_w} \oplus CE_{\mathcal{D}_w}$
18 **end**
19 **return** $CE^* = \left([\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij}\right)_{p \times p}$

5. Experiments and Results

Evaluation Metric. To evaluate the causal effect interval matrix returned by LV-IDA+ and compare it against the computed by LV-IDA, we use the Average Interval Mean Square Error (AIntMSE), base on the Interval Mean Square Error metric (IntMSE) (see Malinsky and Spirtes (2017)). The AIntMSE is the average IntMSE error calculated on the estimates of the causal effect between each ordered pair of variables (X, Y) of a model with p variables.

Let $CEReal = (\theta_{ij})$ be the real total causal effects matrix, where θ_{ij} is the true total causal effect from the pair of variables (i, j) for $i, j = 1, \ldots, p$. The IntMSE between the true total effect θ_{ij} from $CEReal$, and the estimate interval of causal effect $[\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij}$ from the CE^* matrix, returned by the LV-IDA+ and the LV-IDA algorithms, is defined as:

$$\text{IntMSE} = \begin{cases} 0 & \text{if } \theta_{ij} \in [\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij} \\ \min\{|\theta_{ij} - (\hat{\theta}_{min})_{ij}|, |\theta_{ij} - (\hat{\theta}_{max})_{ij}|\} & \text{otherwise,} \end{cases}$$

Let $IEM = (\epsilon_{ij})$ be the Interval Error Matrix where the ϵ_{ij} are defined as the IntMSE between the true total effect θ_{ij} in $CEReal$, and the estimate interval of causal effect $[\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij}$ in CE^*. Let Σ be the sum of the elements ϵ_{ij} of the Interval Error Matrix $IEM = (\epsilon_{ij})$. The Average Interval Mean Square Error (AIntMSE), is defined as:

$$\text{AIntMSE} = \frac{\Sigma}{p^2 - (p + \text{eNA})},$$

where p is the number of variables and eNA is the number of extreme NAs. Note that the denominator corresponds to the total number of elements of the matrices CE^*, minus the elements of the diagonal (which always has value $[1, 1]$), and the number of extreme NAs, which are the returned NA values on the matrices CE^* when it is not possible to estimate the interval of causal effects by LV-IDA. This corresponds to the number of estimated interval actually computed by the LV-IDA+ or LV-IDA algorithms, so this evaluation gives us an average of the error in the calculation of the estimated intervals of causal effect in the CE^* matrices. We include an example of the AIntMSE evaluation metric computation for a real case in the experimentation in Appendix A.

Data Generation Process. We evaluate the LV-IDA+ and LV-IDA algorithms using data simulated from synthetic models. We describe the data generation process for insufficient systems next. First, a random PAG $\mathcal{P}$ is generated. Later, unfolding $\mathcal{P}$ the set of MAGs in the MEC $[\mathcal{M}]$ are listed, and we selected randomly a MAG $\mathcal{M}^*$ from $[\mathcal{M}]$. Then, the canonical DAG $\mathcal{D}^*$ is obtained from $\mathcal{M}^*$. Finally, a data set D' is generated simulating the DAG $\mathcal{D}^*$ (see Figure 4). We parameterized the DAG $\mathcal{D}^*$ with linear Gaussian structural equations, where the coefficients are distributed as $\pm$Uniform($[0.5, 1.5]$), and the random disturbances according to a normal distribution with mean zero and standard deviations taken from a Uniform($[1, 3]$)).

Experimentation. From the data generation process we keep the PAG $\mathcal{P}$, and remove the columns of the variables that were added in the construction of the canonical DAG $\mathcal{D}^*$ over the data set D', to build the data set D. This PAG $\mathcal{P}$ and the data set D are used as the input for the LV-IDA+ and LV-IDA algorithms. We preserve the parameterized DAG $\mathcal{D}^*$ and the complete data set D' to calculate the true causal effects in the $CEReal$ matrix. We generate 80 synthetic causal models by varying the number of variables p using

$p \in \{5, 8, 11, 14\}$, 20 different synthetic models for each value of p and their respective data sets, generated with samples of size 1000. For a total of 7,360 causal effect estimations, for each of the two algorithms: LV-IDA+ and LV-IDA. We use the LV-IDA implementation on the R programming language, given in Malinsky and Spirtes (2017), and we implement the LV-IDA+ algorithm also in the R language. Both implementations of the algorithms were used to calculate the causal effect on all pairs of variables in each of the synthetic models. All experiments were run on a computer with an Intel Core i5 processor running at 2.0 GHz and 8 GB of RAM.

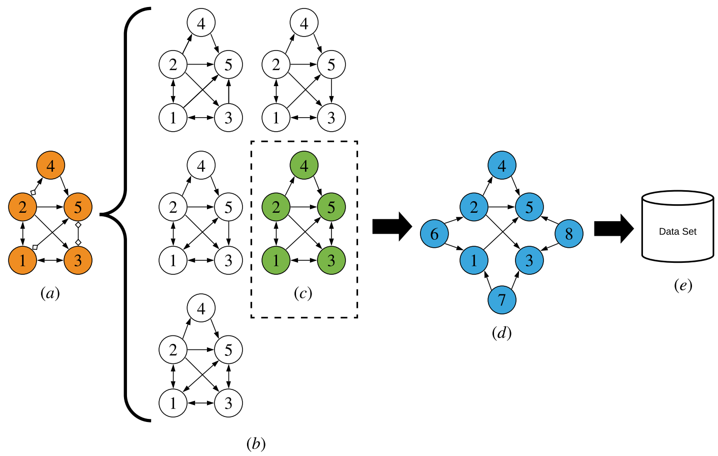

Figure 4: The data generation process. (a) A random generated PAG $\mathcal{P}$ (orange). (b) The set of MAGs in the Markov equivalences class $[\mathcal{M}]$. (c) The random selected MAG $\mathcal{M}^*$ (green) from $[\mathcal{M}]$. (d) The canonical DAG $\mathcal{D}^*$ obtained from $\mathcal{M}^*$ (blue). (e) The data set is generated simulating the DAG $\mathcal{D}^*$. (Best seen in color.)

	AIntMSE	**SimpleNAs**	**ExtremeNAs**	**Time (sec)**
$\boldsymbol{p = 5}$				
LV-IDA+	**0.0223 ± 0.0182**	**0**	**0**	6.6 ± 2.4
LV-IDA	0.0357 ± 0.0413	28.88 ± 16.04	3.50 ± 2.27	**1.4 ± 0.7**
$\boldsymbol{p = 8}$				
LV-IDA+	**0.0441 ± 0.0333**	**0**	**0**	14.4 ± 7.5
LV-IDA	0.0710 ± 0.0611	107.8 ± 28.32	20.3 ± 4.44	**4.6 ± 1.6**
$\boldsymbol{p = 11}$				
LV-IDA+	**0.1040 ± 0.1130**	**0**	**0**	17.0 ± 11.1
LV-IDA	0.1797 ± 0.2133	177.3 ± 78.2	49.4 ± 26.6	**16.4 ± 7.1**
$\boldsymbol{p = 14}$				
LV-IDA+	**0.2353 ± 0.2028**	**0**	**0**	18.4 ± 18.2
LV-IDA	0.4101 ± 0.3605	340.3 ± 98.69	74.1 ± 8.37	**12.2 ± 21.6**

Table 1: Performance comparisons between LV-IDA+ and LV-IDA (mean ± standard deviation) over AIntMSE, for $p = 5, 8, 11$ and 14 where p is the number of variables. The best results for each evaluation are highlighted in bold type. The mean and standard deviation of simple and extreme NAs, and the running time in seconds for each algorithm are shown.

Results. Table 1 summarizes the estimation results for the 80 different synthetic causal models. One can observe the growth of the number of simple and extreme NAs for the LV-IDA algorithm as the number of variables in the models increases. The count for extremes NA values in the results give us an idea of how many more estimation bounds on causal effects we can calculate (or at least approximate) for pairs of variables in a model with the LV-IDA+ algorithm. The results suggest that ignoring missing values in the multisets of causal effects (simple NAs) in the LV-IDA algorithm produce misleading bounds in causal effects, and according to the AIntMSE evaluation metric, using the approximations of causal effects computed by LV-IDA+ algorithm, is better than the missing values (simple NAs) returned by the LV-IDA algorithm, at least for the case of observational data generated by a canonical DAGs with latent variables. We are aware that anti-canonical DAGs, could be problematic for the approximation used in the LV-IDA algorithm (see Appendix B). Therefore, to explore further the generality of our algorithm, it is necessary to extend the spectrum of synthetic models for the data generation in the experimentation.

6. Conclusions and Future Work

The main contribution of this work is a proposal for approximating causal effects when it is not possible to find adjustment sets for a pair of variables on a causal insufficient system represented by a MAG. With this, the proposed LV-IDA+ algorithm can estimate the causal effects among all pairs of variables (or at least approximate the estimate in some cases) in each of the MAGs in a Markov equivalence class and thus it can obtain approximate lower and upper bounds of the real value of the causal effect for each pair of system variables. The proposed method is in principle an efficient baseline that can motivate the search for better approximations for causal effects bounds. Since, to our knowledge, no other way for approximating causal effects has been proposed when they cannot be calculated by covariate adjustments directly in a MAG, for causal insufficient systems.

The experimental results for the LV-IDA+ algorithm are promising but it would be important to extend the experimentation. An analysis of why canonical DAGs are good representatives for a large sub-collection of DAGs with latent variables represented by a MAG, for the purpose of estimating causal effects is presented in Appendix B. However, it is necessary to carry out more experimentation with these structures, to see if indeed a canonical DAG is a good representation for them. In the same analysis, a collection of DAGs with latent variables for which we conjecture they are not well represented by canonical DAGs for the same purpose, which we call anti-canonical DAGs, was also described but also it is necessary to evaluate with more experiments how much a representation through a canonical DAG fits models with anti-canonical patterns.

Extending the work in Montero-Hernandez et al. (2018) to insufficient systems is an interesting application of this research that we are interested in exploring. In the latter, intervals of causal effects are used to single out a unique model from the Markov equivalence class in a causal structure learning process. For the aforementioned, it is essential to be able to compute intervals of causal effects for any pairs of variables in the causal system. Therefore, the proposed LV-IDA+ algorithm is fundamental to continue the ideas of in Montero-Hernandez et al. (2018) and bring them to the domain of insufficient causal systems.

References

C. Glymour, K. Zhang, and P. Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10:524, 2019.

M. H. Maathuis and D. Colombo. A generalized back-door criterion. *Ann. Statist.*, 43(3): 1060–1088, 06 2015.

M. H. Maathuis, M. Kalisch, and P. Bühlmann. Estimating high-dimensional intervention effects from observational data. *Ann. Statist.*, 37(6A):3133–3164, 12 2009.

D. Malinsky and P. Spirtes. Estimating bounds on causal effects in high-dimensional and possibly confounded systems. *International Journal of Approximate Reasoning*, 88:371–384, 2017.

S. Montero-Hernandez, F. Orihuela-Espina, and L. E. Sucar. Intervals of causal effects for learning causal graphical models. In *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*, volume 72 of *Proceedings of Machine Learning Research*, pages 296–307, Prague, Czech Republic, 11–14 Sep 2018. PMLR.

J. M. Ogarrio, P. Spirtes, and J. Ramsey. A Hybrid Causal Search Algorithm for Latent Variable Models. *JMLR workshop and conference proceedings*, 52:368–379, 2016.

J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009.

E. Perković, J. Textor, M. Kalisch, and M. H. Maathuis. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18:1–62, 2018.

T. Richardson and P. Spirtes. Ancestral graph markov models. *Ann. Statist.*, 30(4):962–1030, 08 2002.

L. E. Sucar. *Probabilistic Graphical Models: Principles and Applications*. Springer Publishing Company, Incorporated, 2015.

B. van der Zander, M. Liśkiewicz, and J. Textor. Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artificial Intelligence*, 270: 1–40, 2019.

J. Zhang. Causal reasoning with ancestral graphs. *J. Mach. Learn. Res.*, 9:1437–1474, June 2008a. ISSN 1532-4435.

J. Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17):1873–1896, 2008b.

J. Zhang and P. Spirtes. A transformational characterization of markov equivalence for directed acyclic graphs with latent variables. In *Proceedings of the Twenty-First Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pages 667–674, Arlington, Virginia, 2005. AUAI Press.

Appendix A. AIntMSE Evaluation Metric Example

To further clarify how the AIntMSE evaluation metric is computed, an example for a real case in the experimentation is shown next. Below, we show the real total causal effect matrix $CEReal(\theta_{ij})$ computed from a synthetic $\mathcal{G}$ model with $p = 5$ variables and present the interval matrices $CE^*_{LV-IDA+}$ and CE^*_{LV-IDA} returned by the LV-IDA+ and LV-IDA algorithms, respectively, for the synthetic $\mathcal{G}$ model. Then, the interval error matrices $IEM_{LV-IDA+}$ for $CE^*_{LV-IDA+}$ and IEM_{LV-IDA} for CE^*_{LV-IDA} are computed. Finally, we show the computation of the AIntMSE for each of these CE^* matrices, $CE^*_{LV-IDA+}$ and CE^*_{LV-IDA}.

Equation 2 shows the $CEReal$ matrix that is calculated from a synthetic $\mathcal{G}$ model with $p = 5$ variables.

$$CEReal = (\theta_{ij}) = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.67 \\ 0.00 & 1.00 & 0.79 & 0.58 & 1.74 \\ 0.00 & 0.30 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.53 & 0.00 & 1.00 & 0.62 \\ 0.40 & 0.31 & 0.00 & 0.25 & 1.00 \end{bmatrix} \tag{2}$$

Equation 3 shows the $CE^*_{LV-IDA+}$ matrix returned by the LV-IDA+ algorithm for the $\mathcal{G}$ model.

$$CE^*_{LV-IDA+} = \begin{bmatrix} [1.00, 1.00] & [0.00, 0.00] & [0.00, 0.16] & [0.01, 0.01] & [0.00, 0.67] \\ [0.00, 0.00] & [1.00, 1.00] & [0.68, 0.80] & [0.55, 0.55] & [1.69, 2.18] \\ [0.00, 0.00] & [0.43, 0.48] & [1.00, 1.00] & [-0.04, 0.00] & [0.00, 1.02] \\ [0.05, 0.05] & [0.42, 0.42] & [0.02, 0.15] & [1.00, 1.00] & [0.61, 0.64] \\ [0.00, 0.14] & [0.34, 0.36] & [0.00, 0.32] & [0.29, 0.33] & [1.00, 1.00] \end{bmatrix} \tag{3}$$

Equation 4 shows the CE^*_{LV-IDA} matrix returned by the LV-IDA algorithm for the $\mathcal{G}$ model.

$$CE^*_{LV-IDA} = \begin{bmatrix} [1.00, 1.00] & [0.00, 0.00] & [0.00, 0.00] & [0.00, 0.00] & [0.00, 0.00] \\ [0.00, 0.00] & [1.00, 1.00] & \text{NA} & [0.55, 0.55] & \text{NA} \\ [0.00, 0.00] & [0.00, 0.00] & [1.00, 1.00] & [0.00, 0.00] & [0.00, 0.00] \\ [0.00, 0.00] & [0.00, 0.00] & [0.00, 0.00] & [1.00, 1.00] & \text{NA} \\ [0.00, 0.00] & [0.00, 0.00] & [0.00, 0.27] & [0.00, 0.00] & [1.00, 1.00] \end{bmatrix} \tag{4}$$

Equation 5 shows the $IEM_{LV-IDA+}$ matrix calculated using the $CEReal$ matrix and the $CE^*_{LV-IDA+}$ matrix returned by the LV-IDA+ algorithm for the $\mathcal{G}$ model. Based on this, the AIntMSE evaluation metric is computed below.

$$IEM_{LV-IDA+} = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.03 & 0.00 \\ 0.00 & 0.13 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.11 & 0.02 & 0.00 & 0.00 \\ 0.26 & 0.03 & 0.00 & 0.04 & 0.00 \end{bmatrix} \tag{5}$$

$$\text{AIntMSE} = \frac{\Sigma}{p^2 - (p + \text{eNA})} = \frac{0.68}{25 - (5+0)} = \frac{.68}{20} = 0.034$$

Equation 6 shows the IEM_{LV-IDA} matrix calculated using the $CEReal$ matrix and the CE^*_{LV-IDA} matrix returned by the LV-IDA algorithm for the $\mathcal{G}$ model. Note that since the CE^*_{LV-IDA} matrix contains NAs (produced by extreme NAs), the IEM_{LV-IDA} matrix also contains NAs. Based on this IEM_{LV-IDA} matriz, the AIntMSE evaluation metric is computed below.

$$IEM_{LV-IDA} = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.67 \\ 0.00 & 0.00 & \text{NA} & 0.03 & \text{NA} \\ 0.00 & 0.30 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.53 & 0.00 & 0.00 & \text{NA} \\ 0.40 & 0.31 & 0 & 0.25 & 0.00 \end{bmatrix} \tag{6}$$

$$\text{AIntMSE} = \frac{\Sigma}{p^2 - (p + \text{eNA})} = \frac{2.49}{25 - (5+3)} = \frac{2.49}{17} = 0.146$$

Appendix B. Anti-canonical DAGs

The rationale behind our method is that it is reasonable to calculate the causal effect on one of the DAGs that belongs to the set of DAGs represented by the MAG as an approximation to the causal effect on a pair of variables when this cannot be calculated directly on the MAG in question. With this in mind, we address the main question that motivated this work, about what would be a simple approximation to estimate the effect on a pair of variables when there are no adjustment sets on a MAG, and avoid to answer that such effect cannot be computed. An infinite number of DAGs with latent variables are transformed into the same MAG. However, we can at least rebuild some of the most representative graphs DAGs represented by a MAG. The canonical DAG has the advantage that it can be efficiently built from a MAG and this type of DAGs encompasses the most significant aspects of the causal structure of an infinity subset of DAGs represented by a MAG. We will now go into more detail on the latter below.

From the process of marginalizing the latent variables of a DAG, $\mathcal{D}$, to construct a MAG, $\mathcal{M}$, and then construct the canonical DAG, $\mathcal{D}^*$, associated with this MAG, we were able to identify two types of substructure patterns in which although a canonical DAG, $\mathcal{D}^*$, is not an exact portrayal of the original DAG, $\mathcal{D}$, a canonical DAG offers a good representation for estimating causal effects (see Figures 5 and 6). The first type of substructure pattern is one in which the original DAG contains a complex substructure of a set $\boldsymbol{L}$ of several unmeasured variables and one of the latent variables L^* is a common cause of a pair of observed variables, say O_1 and O_2. When the latent variables in this substructure are marginalized, a MAG, $\mathcal{M}$, is created where none of the latent variables of the substructure appear and it contains a bidirectional edge between O_1 and O_2. Then, when constructing the canonical DAG, $\mathcal{D}^*$, for this MAG, a single variable L is added as a common cause for the observed variables O_1 and O_2 (see Figure 5 for an instance of this substructure pattern). For this first substructure pattern, it can be seen that the weight of all latent variables in

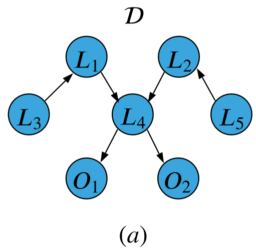

Figure 5: (a) A DAG $\mathcal{D}$ with a complex substructure of several unmeasured variables, $\boldsymbol{L} = \{L_1, L_2, L_3, L_4, L_5\}$ and $L^* = L_4$. (b) The MAG $\mathcal{M}$ obtained from $\mathcal{D}$ by marginalizing the latent variables L_1, L_2, L_3, L_4 and L_5. (c) The canonical DAG $\mathcal{D}^*$ associated to MAG $\mathcal{M}$.

the substructure can be absorbed by a single variable, L, just as in the reconstruction of the original DAG, $\mathcal{D}$, by mean of a canonical DAG, $\mathcal{D}^*$. Therefore, a canonical DAG is a good representation for estimating the causal effect on the class of DAGs that present this type of substructure pattern.

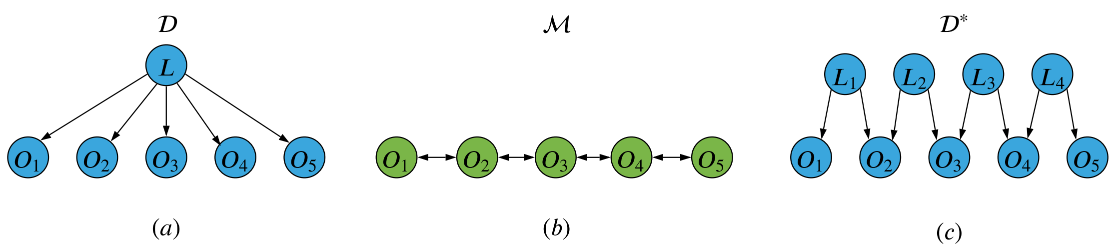

Figure 6: (a) A DAG $\mathcal{D}$ with a latent variable L that is a common cause for several observed variables $\{O_1, O_2, O_3, O_4, O_5\}$. (b) The MAG $\mathcal{M}$ obtained from $\mathcal{D}$ by marginalizing the latent variable L. (c) The canonical DAG $\mathcal{D}^*$ associated to MAG $\mathcal{M}$.

The second type of structural pattern occurs when there is a latent variable, L, that is a common cause of not only a couple of measured variables but several more, this translates into several bi-directed edges in the MAG, $\mathcal{M}$, when this variable is marginalized. When the canonical DAG associated to this MAG is built, a new variable is added for each bi-directed edge, although in reality it is a single unmeasured variable which is the common cause for several variables on the bi-directed edges (see Figure 6 for an instance of this substructure pattern). This structural pattern is not problematic for the calculation of the causal effect with LV-IDA+ as it seems, since what happens is that this unmeasured common cause is simply repeated for each pair of variables measured in the canonical DAG. For the LV-IDA+ algorithm, the essential is that an unmeasured common cause has been identified at each bidirectional edge.

On the other hand, a class of DAGs that we identify are not well represented by a canonical DAG are those we call *anti-canonical* DAGs. We call anti-canonical DAG to a

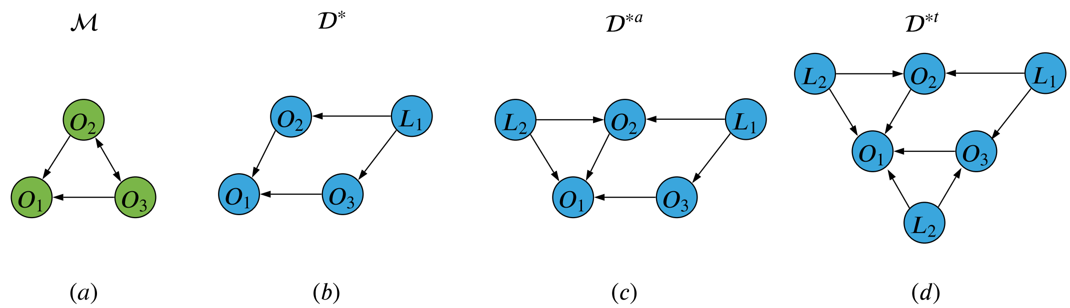

Figure 7: (a) A MAG $\mathcal{M}$ with two directed edges and one bi-directed edge.(b) The canonical DAG associated to the MAG $\mathcal{M}$. (c) An anti-canonical DAG $\mathcal{D}^{*a}$ associated to the MAG $\mathcal{M}$ with an anti-canonical pattern over the directed edge from O_1 to O_2. (d) The total anti-canonical DAG $\mathcal{D}^{*t}$ associated to the MAG $\mathcal{M}$.

DAG that is built from a MAG, $\mathcal{M}$, in which at least one directed edge from a variable O_1 to a variable O_2 in the MAG is transformed into the directed edge from O_1 to O_2 and augmenting a latent variable L as the common cause of O_1 and O_2 in the DAG (see Figure 7 (c)). We call to this transformation from a directed edge in a MAG to a directed edge with a latent common cause in the DAG an anti-canonical pattern. If all the directed edges in the MAG are transformed as anti-canonical patterns we call the resulting DAG, a total anti-canonical DAG (see Figure 7 (d)).

Anti-canonical DAGs are problematic because, if in the original DAG, $\mathcal{D}$, that generates the data there is an unmeasured common cause, L, for two variables O_1 and O_2, and in addition, O_1 is the direct cause of O_2. When forming a MAG $\mathcal{M}$ by marginalizing L from the DAG, $\mathcal{D}$, the MAG has the directed edge from O_1 to O_2 (not a bi-directed edge indicating the latent variable L) and then the canonical DAG, $\mathcal{D}^*$, associated with $\mathcal{M}$ would have simply the same directed edge from O_1 to O_2. This means that if we construct the canonical DAG, $\mathcal{D}^*$, associated to $\mathcal{M}$ we are ruling out the existence of the unmeasured common cause L. When a direct common cause is not contemplated in a system, this causes greater values than what they actually are on the estimations of some causal effects among the variables in the system. So we identify this case as a possible limitation on the proposal to approximate the calculation of the causal effect using the canonical DAG when it is not possible to calculate it for a couple of variables directly on the MAG.