Collapsible IDA: Collapsing Parental Sets for Locally Estimating Possible Causal Effects

Yue Liu, Zhuangyan Fang, Yangbo He*, Zhi Geng
School of Mathematical Sciences
Peking University
Beijing, China

Abstract

It is clear that some causal effects cannot be identified from observational data when the causal directed acyclic graph is absent. In such cases, IDA is a useful framework which estimates all possible causal effects by adjusting for all possible parental sets. In this paper, we combine the adjustment set selection procedure with the original IDA framework. Our goal is to find a common set that can be subtracted from all possible parental sets without influencing the back-door adjustment. To this end, we first introduce graphical conditions to decide whether a treatment's neighbor or parent in a completed partially directed acyclic graph (CPDAG) can be subtracted and then provide a procedure to construct a subtractable set from those subtractable vertices. We next combine the procedure with the IDA framework and provide a fully local modification of IDA. Experimental results show that, with our modification, both the number of possible parental sets and the size of each possible parental set enumerated by the modified IDA decrease, making it possible to estimate all possible causal effects more efficiently.

1 INTRODUCTION

Causal directed acyclic graphs are often used to give interpretable and compact representations of causal relations and the generative mechanisms of observational data (Pearl, 1995; Spirtes et al., 2000; Geng et al., 2019). If the underlying causal DAG is provided (He & Geng, 2008; Hauser & Bühlmann, 2012), the causal effect of a treatment on a target can be estimated from observational data via the back-door adjustment (Pearl, 2009), or more generally, via the covariate adjustment (Shpitser et al., 2010; Perković et al., 2015, 2017; Perković et al., 2018). However, in some situations, based on observational data one can only obtain a set of statistically equivalent DAGs, forming a Markov equivalence class represented by a completed partially directed acyclic graph (CPDAG) (Meek, 1995; Andersson et al., 1997; Spirtes et al., 2000; Chickering, 2002). Since the causal effects of a treatment on a target may vary in Markov equivalent DAGs, it is challenging to estimate causal effects with a CPDAG only.

The recent progress in causal inference shows that some causal effects can be uniquely identified from observational data without a fully specified causal DAG (Perković et al., 2017; Perković et al., 2018; Jaber et al., 2019; Perković, 2019). Despite these criteria, there are still many causal effects that cannot be identified. To deal with this problem, Maathuis et al. (2009) proposed an alternative framework called intervention do-calculus when the DAG is absent (IDA), which enumerates all possible causal effects of a treatment on a target. Since knowing the parental set of a treatment is enough for the back-door adjustment, IDA and its generalizations only enumerate possible parental sets instead of all equivalent DAGs (Maathuis et al., 2009; Nandy et al., 2017; Perković et al., 2017), making them suitable for sparse DAGs.

Though enumerating possible parental sets is sufficient for estimating all possible causal effects, with limited samples adjusting for possible parental sets may still be challenging, since the total number of possible parental sets and the size of each possible parental set could be very large. There are many papers studying the adjustment set selection problem (Kuroki & Cai, 2004; VanderWeele & Shpitser, 2011; Henckel et al., 2019; Andrea & Ezequiel, 2019), but it is difficult to combine them with the IDA framework directly without breaking the local nature of IDA, since most of them focus on the case where the causal effect can be uniquely identified.

*Correspondence to: heyb@pku.edu.cn.

Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), PMLR volume 124, 2020.

In this paper, we consider the problem of finding a common set that can be subtracted from all possible parental sets simultaneously without affecting the back-door adjustment. This problem is similar to the one considered in VanderWeele & Shpitser (2011); Henckel et al. (2019); Andrea & Ezequiel (2019), but the major difference is that their work focuses on pruning each possible parental set separately. We first introduce a new concept called uniform collapsibility, which basically states that a set of possible parental sets is uniformly collapsible over a set $\mathbf{Z}$, if after subtracting $\mathbf{Z}$ from all of those possible parental sets, the remaining parts are still back-door adjustment sets. Next, given a CPDAG, we provide graphical conditions to decide which subset of treatment's neighbors and parents in the CPDAG can be subtracted. Based on these results, we provide a modification of the original IDA which includes the adjustment set selection procedure while keeps the local nature of IDA. Experimental results show that, with our modification, both the number of possible parental sets and the size of each possible parental set enumerated by the modified IDA decrease, making it possible to estimate all possible causal effects more efficiently.

2 PRELIMINARIES

In this section, we review some basic concepts.

2.1 NOTATION AND DEFINITIONS

Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ denote a graph. Two vertices are adjacent if there is an edge between them. If $X_i \to X_j$, then X_i is a *parent* of X_j and X_j is a *child* of X_i. If $X_i - X_j$ then they are *neighbors* of each other. A graph is called *directed* (*undirected*, or *partially directed*) if the edges in the graph are directed (undirected, or a mixture of directed and undirected). We agree that directed and undirected graphs are also partially directed. The *skeleton* of a (partially) directed graph is the undirected graph obtained by replacing all directed edges with undirected ones. If $U \to W \leftarrow V$ and U, V are not adjacent in $\mathcal{G}$, then (U, W, V) forms a *v-structure* collided on W. A *path* in a graph is a sequence of distinct vertices such that any two consecutive vertices are adjacent in the graph. Let $\pi = (X_0, X_1, ..., X_n)$ be a path in $\mathcal{G}$, if $X_{i-1} \to X_i$ or $X_{i-1} - X_i$ for all $i = 1, 2, ..., n$, then π is a *partially directed* path from X_0 to X_n. If all edges on a partially directed path are directed (undirected), then the path is *directed* (*undirected*). If there is a directed path from X_i to X_j or $X_i = X_j$, then X_i is an *ancestor* of X_j and X_j is a *descendant* of X_i. Given a graph $\mathcal{G}$, the parents, children, neighbors, ancestors and descendants of a set $\mathbf{X}$ are the union of those of each $X \in \mathbf{X}$ in $\mathcal{G}$, and are denoted by $pa(\mathbf{X}, \mathcal{G})$, $ch(\mathbf{X}, \mathcal{G})$, $ne(\mathbf{X}, \mathcal{G})$, $an(\mathbf{X}, \mathcal{G})$ and $de(\mathbf{X}, \mathcal{G})$, respectively. $\mathcal{G}$ will be omitted from these notations if the context is clear, . A (directed) *cycle* is a (directed) path that starts and ends with the same vertex. A directed graph without directed cycle is called a *directed acyclic graph* (DAG).

2.2 CAUSAL GRAPHICAL MODELS

Based on the notion of d-separation (Pearl, 2009, or see, e.g. Appendix A.1), a DAG encodes a set of conditional independence relationships. DAGs encoding the same conditional independencies are called *Markov equivalent* and form a *Markov equivalence class*. Two equivalent DAGs have the same skeleton and the same v-structures (Pearl et al., 1989). A Markov equivalence class can be uniquely represented by a *completed partially directed acyclic graph* (CPDAG) $\mathcal{G}^*$. As proved by Andersson et al. (1997), $\mathcal{G}^*$ is a *chain graph*. We use $chcomp(X, \mathcal{G}^*)$ to denote the *chain component* containing X in $\mathcal{G}^*$, and use $[\mathcal{G}^*]$ or $[\mathcal{G}]$ to denote the Markov equivalence class represented by $\mathcal{G}^*$ or containing $\mathcal{G}$, respectively. It can be shown that the skeleton of a CPDAG $\mathcal{G}^*$ is the same as the skeleton of every DAG in $[\mathcal{G}^*]$, and an edge is directed in a CPDAG if and only if it is directed in every DAG in $[\mathcal{G}^*]$ (Pearl et al., 1989).

Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG and f be a distribution over $\mathbf{V}$. We use $\{X \perp\!\!\!\perp Y \mid \mathbf{Z}\}_{\mathcal{G}}$ to denote that X and Y are d-separated by $\mathbf{Z}$ in $\mathcal{G}$, and use $\{X \perp\!\!\!\perp Y \mid \mathbf{Z}\}_f$ to denote that X and Y are independent conditioning on $\mathbf{Z}$ w.r.t. f. We say that f is *Markovian* to $\mathcal{G}$ if $\{X \perp\!\!\!\perp Y \mid \mathbf{Z}\}_{\mathcal{G}}$ implies $\{X \perp\!\!\!\perp Y \mid \mathbf{Z}\}_f$. Any distribution f Markovian to a DAG $\mathcal{G}$ can be factorized as,

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i | pa(x_i, \mathcal{G})).$$

A causal graphical model consists of a DAG and a distribution Markovian to that DAG.

2.3 INTERVENTION CALCULUS

In order to obtain the effect of an intervention on a target variable, Pearl (2009) employed the *do-operator* to formulate the post-intervention distribution as follows:

$$\begin{aligned} &f(x_1, ..., x_n | do(X_j = x_j')) \\ &= \begin{cases} \prod\limits_{i=1, i\neq j}^{n} f(x_i | pa(x_i))|_{x_j = x_j'}, & \text{if } x_j = x_j', \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{1}$$

Here, $f(x_1, ..., x_n | do(X_j = x_j'))$ is the post-intervention distribution over $\mathbf{V} = \{X_1, ..., X_n\}$ after intervening on X_j, by forcing X_j to equal x_j'. The post-intervention distribution $f(x_i | do(X_j = x_j))$ is defined by integrating out all variables other than x_i in $f(x_1, ..., x_n|do(X_j = x_j))$. Given a treatment X and a target Y, if there exists an $x \neq x'$ such that $f(y|do(X = x)) \neq f(y|do(X = x'))$, then X has causal effect on Y (Pearl, 2009). It is common to summarize the distribution generated by an intervention by its mean (Pearl, 2009; Maathuis et al., 2009), i.e., the mean of Y w.r.t. $f(y|do(X = x))$, which is denoted by $E(Y|do(X = x))$. $E(Y|do(X = x))$ is a function of x. If X is continuous, or more precisely, $E(Y|do(X = x))$ is differentiable w.r.t. x, then we can define the average causal effect (ACE) of $do(X = x)$ on Y, i.e., $ACE(Y|do(X = x))$, by

$$ACE(Y|do(X = x)) = \frac{\partial E(Y|do(X = x))}{\partial x}.$$

If X is discrete or $E(Y|do(X = x))$ is not differentiable w.r.t. x, we can set a reference value x_0 and define

$$\begin{aligned} & ACE(Y|do(X = x)) \\ & \qquad = E(Y|do(X = x)) - E(Y|do(X = x_0)). \end{aligned}$$

In general, the post-intervention distribution $f(y|do(X = x))$ is not identical to the conditional distribution $f(y|X = x)$, meaning that we can not estimate $f(y|do(X = x))$ by $f(y|X = x)$. Fortunately, Pearl (2009) showed that, if f is Markovian to $\mathcal{G}$, then for any $Y \notin pa(X, \mathcal{G})$, we have $f(y|do(X = x), pa(x, \mathcal{G})) = f(y|X = x, pa(x, \mathcal{G}))$. Therefore,

$$\begin{aligned} & f(y|do(X = x)) \\ & = \int f(y|do(X = x), pa(x)) f\left(pa(x)\right) d\left(pa(x)\right) \\ & = \int f(y|X = x, pa(x)) f\left(pa(x)\right) d\left(pa(x)\right). \end{aligned} \tag{2}$$

Here, $pa(x)$ is an abbreviation for $pa(x, \mathcal{G})$. Equation (2) shows that, if we known $pa(X, \mathcal{G})$, then we can estimate $f(y|do(X = x))$ from observational data. In fact, Equation (2) is a special case of so-called *back-door adjustment*(Pearl, 2009), and $pa(X, \mathcal{G})$ is a *back-door adjustment set*. The general definition of back-door adjustment set is given as follows (Pearl, 2009, Definition 3.3.1):

Definition 1 (Back-Door Adjustment Set) *Let* $\mathbf{W}$ *be a variable set and* $X, Y \notin \mathbf{W}$ *be two distinct variables in a DAG* $\mathcal{G}$*. Then we say that* $\mathbf{W}$ *is a back-door adjustment set for* (X, Y) *w.r.t.* $\mathcal{G}$ *if:*

(1) no node in $\mathbf{W}$ *is a descendant of* X*; and*

(2) $\mathbf{W}$ *blocks every path between* X *and* Y *that contains an arrow into* X*.*

Pearl (2009) showed that, if $\mathbf{W}$ is a back-door adjustment set, then $f(y|do(X = x), \mathbf{w}) = f(y|X = x, \mathbf{w})$ and

$$f(\mathbf{y}|do(X = x)) = \int f(\mathbf{y}|\mathbf{w}, x) f(\mathbf{w}) d\mathbf{w}.$$

Algorithm 1 The IDA algorithm

Require: A CPDAG $\mathcal{G}^*$, a variable X and a target Y in $\mathcal{G}^*$,
Ensure: A multi-set Θ which stores all possible causal effects of X on Y.
Initialize $\Theta = \emptyset$,
for each $\mathbf{S} \subseteq ne(X, \mathcal{G}^*)$ such that $\mathbf{S}$ is a clique **do**
  estimate the causal effect θ of X on Y by adjusting for $\mathbf{S} \cup pa(X, \mathcal{G}^*)$, and add θ to Θ,
end for
return Θ.

Given a DAG $\mathcal{G}$ and a variable X in $\mathcal{G}$, we define the *manipulated graph* $\mathcal{G}_{\bar{X}}$ as the subgraph of $\mathcal{G}$ by deleting all directed edges pointing at X (Spirtes et al., 2000; Pearl, 2009). Manipulated graphs are important in causal inference, as one can see that if f is Markovian to $\mathcal{G}$, then $f(\cdot|do(X = x))$ is Markovian to $\mathcal{G}_{\bar{X}}$.

2.4 THE IDA FRAMEWORK

The back-door adjustment provides an efficient way to compute post-intervention distributions. However, a causal DAG must be prespecified. In general, due to the existence of Markov equivalent DAGs, it is possible that one can only obtain a CPDAG from observational data instead of a DAG. Much research has been devoted to estimating post-intervention distributions and causal effects when the DAG is absent (Maathuis & Colombo, 2015; Perković et al., 2015, 2017). However, in some cases, the causal effect of a treatment on a target may not be identifiable. For example, if the causal effects of a treatment on a target vary in different equivalent DAGs, then it is impossible to estimate the true causal effect without knowing the underlying causal DAG.

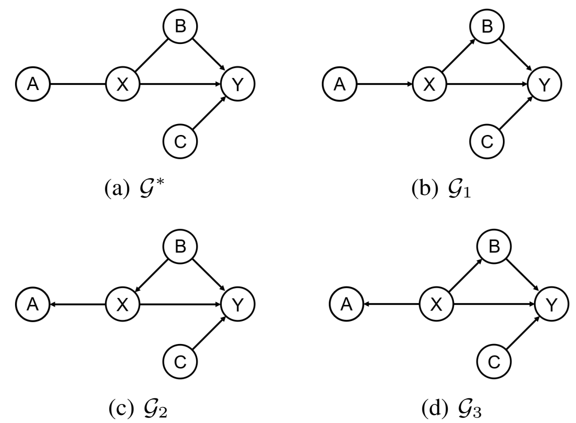
To deal with the unidentifiable cases, Maathuis et al. (2009) proposed an alternative framework called intervention do-calculus when the DAG is absent (IDA) (see Algorithm1 for the details). For a treatment and a target, IDA estimates all possible causal effects of the treatment on the target, by using Equation (2) to compute the causal effect w.r.t. each of the equivalent DAGs.[1] Since Equation (2) only requires the parental set of X in each DAG, to avoid enumerating equivalent DAGs, IDA enumerates all possible parental sets by using the following lemma.

Lemma 1 (Maathuis et al., 2009, Lemma 3.1) *Let* $\mathcal{G}^*$ *be a CPDAG,* X *be a vertex of* $\mathcal{G}^*$*, and* $\mathbf{S} \subset ne(X, \mathcal{G}^*)$.

[1] We note that, although Maathuis et al. (2009) assumed a linear Gaussian model and used ACE to summarize the causal effects, but IDA can be easily extended beyond those assumptions. Similarly, the results in our paper do not need such assumptions either.

Then there is a DAG $\mathcal{G} \in [\mathcal{G}^*]$ such that $pa(X, \mathcal{G}) = pa(X, \mathcal{G}^*) \cup \mathbf{S}$ if and only if orienting $S \to X$ for every $S \in \mathbf{S}$ in $\mathcal{G}^*$ does not introduce any new v-structure.

Meek (1995, Lemma 1) proved that if $Y \in pa(X, \mathcal{G}^*)$, then $Y \in pa(X', \mathcal{G}^*)$ for every $X' \in ne(X, \mathcal{G}^*)$. From this we can prove that the condition in Lemma 1 holds if and only if $\mathbf{S}$ is a clique, i.e., $\mathbf{S}$ is either an empty set, or a singleton set, or for any two distinct vertices $S, S' \in \mathbf{S}$, S and S' are adjacent in $\mathcal{G}^*$. Clearly, enumerating possible parental sets is more efficient than enumerating DAGs (He et al., 2015). However, when the size of $ne(X, \mathcal{G}^*)$ is large, it may take a long time to finish enumeration. Moreover, if the sample size is small, the estimation of $f(Y|do(X = x))$ may have a large variance. In the following, we will provide a method to reduce both the number of possible parental sets and the size of each possible parental set.



(a) $\mathcal{G}^*$ (b) $\mathcal{G}_1$

(c) $\mathcal{G}_2$ (d) $\mathcal{G}_3$

Figure 1: This example shows how to collapsing $pa(X, \mathcal{G})$ for estimating all possible causal effects under the IDA framework.

3 UNIFORM COLLAPSIBILITY FOR POSSIBLE PARENTAL SETS

As discussed in Section 2.4, our goal is to reduce both the number of possible parental sets and the size of each possible parental set when estimating all possible causal effects. The start point is Equation (2). For a DAG $\mathcal{G}$, if we can find a subset $\mathbf{Z}$ of $pa(x, \mathcal{G})$, such that

$$f(y|X = x, pa(x, \mathcal{G})) = f(y|X = x, pa(x, \mathcal{G}) \setminus \mathbf{Z}),$$

then we can estimate $f(y|do(X = x))$ by adjusting for $pa(x, \mathcal{G}) \setminus \mathbf{Z}$,

$$\begin{aligned} & f(y|do(X = x)) \\ & = \int f(y|do(X = x), pa(x)) f\left(pa(x)\right) d\left(pa(x)\right) \\ & = \int f(y|X = x, pa(x)) f\left(pa(x)\right) d\left(pa(x)\right) \\ & = \int f(y|X = x, pa(x) \setminus \mathbf{Z}) f\left(pa(x) \setminus \mathbf{Z}\right) d\left(pa(x) \setminus \mathbf{Z}\right). \end{aligned}$$

Since $pa(x, \mathcal{G}) \setminus \mathbf{Z}$ contains less variables, adjusting for $pa(x, \mathcal{G}) \setminus \mathbf{Z}$ may lead to a more accurate estimation (Henckel et al., 2019; Andrea & Ezequiel, 2019).

3.1 COLLAPSIBILITY

To formulate the idea given at the beginning, we introduce the following concept.

Definition 2 (Collapsibility) *Let X, Y be distinct vertices in a DAG $\mathcal{G}$ such that $Y \notin pa(X, \mathcal{G})$, and $\mathbf{W}$ is a back-door adjustment set for (X, Y) w.r.t. $\mathcal{G}$. We say that $\mathbf{W}$ is collapsible over $\mathbf{Z} \subseteq \mathbf{W}$ (or onto $\mathbf{W} \setminus \mathbf{Z}$) w.r.t. $\mathcal{G}$ and (X, Y), if either $\mathbf{W} = \emptyset$, or $\mathbf{Z} \neq \emptyset$ and $\mathbf{W} \setminus \mathbf{Z}$ is a back-door adjustment set for (X, Y) w.r.t. $\mathcal{G}$.*

In Definition 2, if $\mathbf{Z} = \{Z\}$ is a singleton set, we simply say that $\mathbf{W}$ is collapsible over Z w.r.t. $\mathcal{G}$ and (X, Y). Moreover, if $\mathbf{W}$ is collapsible over $\mathbf{Z}$, then $\mathbf{Z}$ is called *subtractable* from $\mathbf{W}$.[2] Back to the IDA framework, if $pa(X, \mathcal{G})$ is collapsible over $\mathbf{Z}(\mathcal{G})$, then estimating $f(y|do(X = x))$ by adjusting for $pa(X, \mathcal{G}) \setminus \mathbf{Z}(\mathcal{G})$ may improve the efficiency and accuracy of the estimation (Henckel et al., 2019; Andrea & Ezequiel, 2019).

Example 1 *Figure 1 shows how to collapse $pa(X, \mathcal{G})$ for estimating all possible causal effects under the IDA framework. The CPDAG $\mathcal{G}^*$ is shown in Figure 1(a), and Figures 1(b)-1(d) enumerate all equivalent DAGs. Since $ne(X, \mathcal{G}^*) = \{A, B\}$ and $pa(X, \mathcal{G}^*) = \emptyset$, all possible parental sets of X are $\{A\}$, $\{B\}$ and $\emptyset$, which correspond to Figures 1(b)-1(d) respectively. However, in Figure 1(b), $pa(X, \mathcal{G}_1)$ is collapsible over A. Therefore, $f(y|do(X = x)) = \int f(y|X = x, a) f(a)\, da = f(y|X = x)$, meaning that the post-intervention distribution is reduced to the conditional distribution. On the other hand, since neither $\{B\}$ in $\mathcal{G}_2$ nor $\emptyset$ in $\mathcal{G}_3$ is collapsible, the final possible back-door adjustment sets are $\{B\}$ and $\emptyset$.*

Example 1 shows that collapsing $pa(X, \mathcal{G})$ can indeed reduce both the number of possible parental sets and the size of each parental set when estimating all possible causal effects. However, as shown in Example 1, for different $\mathcal{G}$'s, $pa(X, \mathcal{G})$'s may be collapsible over different

[2]The terminology of 'collapsibility' is borrowed from statistics (see, e.g. Xie & Geng, 2009). In statistics, collapsibility means that the same statistical result of interest can be obtained before and after marginalization over some variables.

$\mathbf{Z}(\mathcal{G})$'s. Thus, we need a simple rule to check whether a set can be subtracted from $pa(X, \mathcal{G})$.

Proposition 1 *Suppose that X and $Y \notin pa(X, \mathcal{G})$ are distinct vertices in a DAG $\mathcal{G}$, and $\mathbf{Z}(\mathcal{G})$ is a subset of $pa(X, \mathcal{G})$. Then $pa(X, \mathcal{G})$ is collapsible over $\mathbf{Z}(\mathcal{G})$ w.r.t. $\mathcal{G}$ and (X, Y) if and only if $\{\mathbf{Z}(\mathcal{G}) \perp\!\!\!\perp Y \mid X \cup pa(X, \mathcal{G}) \setminus \mathbf{Z}(\mathcal{G})\}_{\mathcal{G}}$.*

All detailed proofs of the theoretical results in this paper are present in Appendix A. The sufficiency of Proposition 1 follows from Henckel et al. (2019, Lemma D.1). In fact, we can also prove that,

Proposition 2 *With the assumptions in Proposition 1, $pa(X, \mathcal{G})$ is collapsible over $\mathbf{Z}(\mathcal{G})$ w.r.t. $\mathcal{G}$ and (X, Y) if and only if $\{\mathbf{Z}(\mathcal{G}) \perp\!\!\!\perp Y \mid X \cup pa(X, \mathcal{G}) \setminus \mathbf{Z}(\mathcal{G})\}_{\mathcal{G}_{\bar{X}}}$.*

Henckel et al. (2019, Algorithm 1) also provided an algorithm to construct a subtractable set. However, combining this algorithm with IDA locally is still challenging. In fact, it may take much more effort to find $\mathbf{Z}(\mathcal{G})$ than simply adjusting for $pa(X, \mathcal{G})$. Therefore, in this paper, we focus on another strategy. We would like to find a fixed variable set $\mathbf{Z}$ which can be subtracted from all possible parental sets.

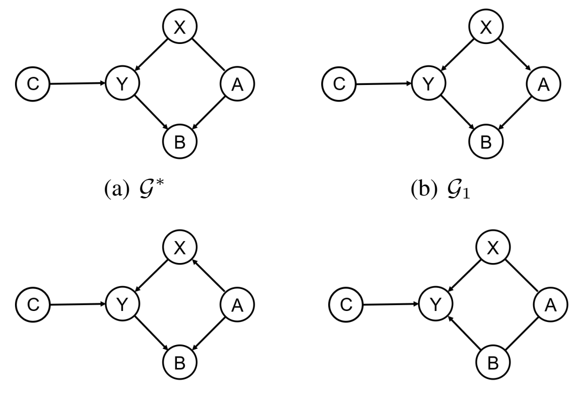
3.2 UNIFORM COLLAPSIBILITY

In this section, we introduce a new concept called uniform collapsibility for a set of back-door adjustment sets.

Definition 3 (Uniform Collapsibility) *Let $\mathbf{Z}$ be a variable set, and $X, Y \notin \mathbf{Z}$ are two distinct vertices in a CPDAG $\mathcal{G}^*$ and $Y \notin pa(X, \mathcal{G}^*)$. Given a set of back-door adjustment sets $\mathcal{W} = \{\mathbf{W}(\mathcal{G}) \mid \mathcal{G} \in [\mathcal{G}^*]$ and $Y \notin pa(X, \mathcal{G})\}$, where $\mathbf{W}(\mathcal{G})$ is a back-door adjustment set for (X, Y) w.r.t. $\mathcal{G}$, we say that $\mathcal{W}$ is uniformly collapsible over $\mathbf{Z}$ w.r.t. $\mathcal{G}^*$ and (X, Y), if $\mathbf{W}(\mathcal{G})$ is collapsible over $\mathbf{W}(\mathcal{G}) \cap \mathbf{Z}$ w.r.t. $\mathcal{G}$ and (X, Y) for every $\mathbf{W}(\mathcal{G}) \in \mathcal{W}$.*

If $\mathcal{W}$ is uniformly collapsible over $\mathbf{Z}$, then $\mathbf{Z}$ is called *uniformly subtractable* from $\mathcal{W}$. Clearly, $\mathcal{W}$ is uniformly collapsible over any $\mathbf{Z}$ such that $\mathbf{W}(\mathcal{G}) \cap \mathbf{Z} = \emptyset$ for every $\mathcal{G} \in [\mathcal{G}^*]$. We call such $\mathbf{Z}$ *trivial*. Conversely, if there exists a non-trivial set which is uniformly subtractable from $\mathcal{W}$, then the size of at least one back-door adjustment set in $\mathcal{W}$ can be reduced. Next example shows that non-trivial sets do exist for some CPDAGs.

Example 2 *Consider the CPDAG $\mathcal{G}^*$ shown in Figure 2(a) and two equivalent DAGs in Figures 2(b) and 2(c). Let $\mathcal{W} = \{\emptyset, \{A\}\}$. Clearly, $\emptyset$ and $\{A\}$ are back-door adjustment sets for (X, Y) w.r.t. $\mathcal{G}_1$ and $\mathcal{G}_2$ respectively. Since $\{A\}$ is collapsible over $\{A\}$ in $\mathcal{G}_2$ based on Proposition 1, and $\emptyset$ is collapsible over $\emptyset \cap \{A\}$ in $\mathcal{G}_1$, $\mathcal{W}$ is uniformly collapsible over $\{A\}$. After collapsing $\mathcal{W}$, we only need adjust for $\emptyset$ in the IDA framework.*



(a) $\mathcal{G}^*$ (b) $\mathcal{G}_1$ (c) $\mathcal{G}_2$ (d) $\mathcal{G}^{**}$

Figure 2: An example to show that non-trivial sets exist for some CPDAGs.

Conversely, for some CPDAGs, such non-trivial subtractable sets may not exist. For example, if we consider the CPDAG in Figure 2(d), then $\mathcal{W} = \{\emptyset, \{A\}\}$ is not uniformly collapsible over A w.r.t. $\mathcal{G}^{**}$ and (X, Y).

3.3 CHARACTERIZATIONS AND CONSTRUCTIONS

Based on the IDA framework, our goal is to characterize and construct a set $\mathbf{Z}$ which is uniformly subtractable from $\mathcal{W} = \{pa(X, \mathcal{G}) \mid \mathcal{G} \in [\mathcal{G}^*]$ and $Y \notin pa(X, \mathcal{G})\}$. The road map is as follows: we first discuss when a single vertex can be uniformly subtracted from $\mathcal{W}$ (Theorems 1 and 2), then we consider how to construct a larger subtractable set from those singleton sets (Theorems 3).

The first result, which is given in Theorem 1, provides a sufficient and necessary condition under which a single vertex in $ne(X, \mathcal{G})$ is uniformly subtractable from $\mathcal{W} = \{pa(X, \mathcal{G}) \mid \mathcal{G} \in [\mathcal{G}^*]$ and $Y \notin pa(X, \mathcal{G})\}$.

Theorem 1 *Suppose that $\mathcal{G}^*$ is a CPDAG, and X, Y, Z are three distinct vertices in $\mathcal{G}^*$ such that $Y \notin pa(X, \mathcal{G}^*)$ and $Z \in ne(X, \mathcal{G}^*)$. Let $\mathcal{W} = \{pa(X, \mathcal{G}) \mid \mathcal{G} \in [\mathcal{G}^*]$ and $Y \notin pa(X, \mathcal{G})\}$, then the following statements are equivalent.*

(1) $\mathcal{W}$ is uniformly collapsible over Z w.r.t. $\mathcal{G}^*$ and (X, Y),

(2) $\{Z \perp\!\!\!\perp Y \mid X \cup pa(X, \mathcal{G}) \setminus Z\}_{\mathcal{G}}$ holds for every $pa(X, \mathcal{G}) \in \mathcal{W}$,

(3) $\{Z \perp\!\!\!\perp Y \mid X \cup pa(X, \mathcal{G}) \setminus Z\}_{\mathcal{G}_{\bar{X}}}$ holds for every $pa(X, \mathcal{G}) \in \mathcal{W}$, and

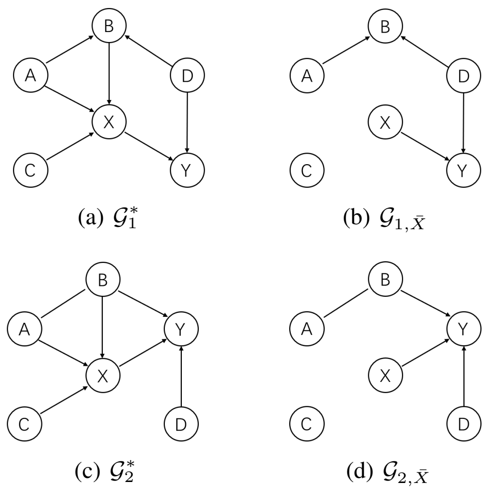


Figure 3: This example shows the results in Theorem 1 no longer hold when $Z \in pa(X, \mathcal{G}^*)$.

(4) **(graphical criterion)** *all partially directed paths from Z to Y, if any, passes X.*

The fourth statement in Theorem 1 gives a necessary and sufficient graphical criterion to decide whether a singleton subset of $ne(X, \mathcal{G}^*)$ is uniformly subtractable from $\mathcal{W}$. Note that, if none of the paths from Z to Y is partially directed, then $\mathcal{W}$ is also uniformly collapsible over Z. We also note that, the graphical criterion only holds for $Z \in ne(X, \mathcal{G}^*)$. If $Z \in pa(X, \mathcal{G}^*)$, the criterion is neither sufficient nor necessary. Below we give an example.

Example 3 *Figure 3(a) shows a CPDAG $\mathcal{G}_1^*$ containing directed edges only. Figure 3(c) shows another CPDAG $\mathcal{G}_2^*$ in which only A and B are connected by an undirected edge. Since $\mathcal{G}_1^*$ has no undirected edge, the only DAG in the Markov equivalence class represented by $\mathcal{G}_1^*$ is itself. Thus, the corresponding manipulated graph is $\mathcal{G}_{1,\bar{X}}$, as shown in 3(b). Similarly, the corresponding manipulated graphs of the DAGs in $[\mathcal{G}_2^*]$ are shown in 3(d), where $A - B$ in $\mathcal{G}_{2,\bar{X}}$ can be oriented as $A \to B$ or $A \leftarrow B$.*

We first show that statement (4) is not sufficient. As shown in Figure 3(a), all partially directed paths from A to Y pass through X. However, by proposition 1 A is not subtractable from $\{A, B, C\}$, as $A \to B \leftarrow D \to Y$ is a d-connected path given B, C and X in both $\mathcal{G}_1^*$ and $\mathcal{G}_{1,\bar{X}}$. To show that statement (4) is not necessary either, let us consider Figures 3(c) and 3(d). Although $A - B \to Y$ is a partially directed path from X to Y which bypasses X, A is d-separated from Y given B, C, X in both $\mathcal{G}_2^*$ and $\mathcal{G}_{2,\bar{X}}$. Thus, A is subtractable.

Next, we consider the singleton subsets of $pa(X, \mathcal{G}^*)$. It can be shown that,

Proposition 3 *Let $\mathcal{G}^*$ be a CPDAG, X, Y be two distinct vertices in $\mathcal{G}^*$ such that $Y \notin pa(X, \mathcal{G}^*)$. Suppose there exists a $Z \in ne(X, \mathcal{G}^*)$ such that $Z \neq Y$ and $\mathcal{W} = \{pa(X, \mathcal{G}) \mid \mathcal{G} \in [\mathcal{G}^*] \text{ and } Y \notin pa(X, \mathcal{G})\}$ is not uniformly collapsible over Z, then $\mathcal{W}$ is not uniformly collapsible over any subset of $pa(X, \mathcal{G}^*)$.*

Proposition 3 is a necessary condition for the collapsibility of $pa(X, \mathcal{G}^*)$. It shows that, if some neighbors of X are not subtractable, then we do not bother to collapse $pa(X, \mathcal{G}^*)$. Conversely, if the entire set $ne(X, \mathcal{G}^*)$ can be subtracted, then the causal effect of X on Y is identifiable. In this case, many criteria are useful for selecting and constructing an adjustment set (see, e.g. Henckel et al., 2019). For the sake of completeness, we also provide a sufficient condition for collapsing $pa(X, \mathcal{G}^*)$.

Theorem 2 *Suppose that $\mathcal{G}^*$ is a CPDAG, and X, Y, Z are three distinct vertices in $\mathcal{G}^*$ such that $Y \notin pa(X, \mathcal{G}^*)$ and $Z \in pa(X, \mathcal{G}^*)$. Let $\mathcal{W} = \{pa(X, \mathcal{G}) \mid \mathcal{G} \in [\mathcal{G}^*] \text{ and } Y \notin pa(X, \mathcal{G})\}$, then $\mathcal{W}$ is uniformly collapsible over Z w.r.t. $\mathcal{G}^*$ and (X, Y) if every path from Z to Y, if any, passes X.*

Clearly, Theorem 2 is still valid if we replace Z by a subset $\mathbf{Z} \subset pa(X, \mathcal{G})$. Notice that, unlike Theorem 1, Theorem 2 only provides a sufficient condition. To see why Theorem 2 is not necessary, let us consider the following Example 4.

Example 4 *In this example, we show that the condition in Theorem 2 is not necessary. As shown in Figures 3(c) and 3(d), A is a parent of X in $\mathcal{G}^*$, and $A - B \to Y$ is a path from A to Y bypassing X. However, as discussed in Example 3, $\{A, B, C\}$ is uniformly collapsible over A.*

Since $pa(X, \mathcal{G}) \subset pa(X, \mathcal{G}^*) \cup ne(X, \mathcal{G}^*)$ for any $\mathcal{G} \in [\mathcal{G}^*]$, we do not have to consider $Z \in ch(X, \mathcal{G}^*)$. Thus, the remaining problem is how to construct a subtractable set containing more than just one vertex. The following Theorem 3 provides an answer.

Theorem 3 *Suppose that $\mathcal{G}^*$ is a CPDAG, X and Y are two distinct vertices in $\mathcal{G}^*$ and $Y \notin pa(X, \mathcal{G}^*)$, and $\mathbf{Z}_1, \mathbf{Z}_2$ are two subsets of variables such that at least one of them is a subset of $ne(X, \mathcal{G}^*)$. Let $\mathcal{W} = \{pa(X, \mathcal{G}) \mid \mathcal{G} \in [\mathcal{G}^*] \text{ and } Y \notin pa(X, \mathcal{G})\}$, if $\mathcal{W}$ is uniformly collapsible over both $\mathbf{Z}_1$ and $\mathbf{Z}_2$ w.r.t. $\mathcal{G}^*$ and (X, Y), then $\mathcal{W}$ is uniformly collapsible over $\mathbf{Z}_1 \cup \mathbf{Z}_2$ w.r.t. $\mathcal{G}^*$ and (X, Y).*

Based on the above theorems, we have,

Corollary 1 *Suppose that $\mathcal{G}^*$ is a CPDAG, X and Y are two distinct vertices in $\mathcal{G}^*$ and $Y \notin pa(X, \mathcal{G}^*)$. Let $\mathbf{Z}_{ne} \subset ne(X, \mathcal{G}^*)$ and $\mathbf{Z}_{pa} \subset pa(X, \mathcal{G}^*)$ be the sets*

Algorithm 2 The collapsible IDA algorithm

Require: A CPDAG $\mathcal{G}^*$, a variable X and a target Y in $\mathcal{G}^*$,

Ensure: A multi-set Θ which stores all possible causal effects of X on Y.

1: Initialize $\Theta = \emptyset$,
2: find all vertices in $ne(X, \mathcal{G}^*)$ from which there is no partially directed path to Y in $\mathcal{G}^*$ that bypasses X, and denote them by $\mathbf{Z}_{ne}$,
3: **if** $\mathbf{Z}_{ne}$ is not identical to $ne(X, \mathcal{G}^*)$, **then**
4:     set $\mathbf{Z}_{pa} = \emptyset$,
5: **else**
6:     find all vertices in $pa(X, \mathcal{G}^*)$ from which every path to Y passes through X, and denote them by $\mathbf{Z}_{pa}$,
7: **end if**
8: **for** each $\mathbf{S} \subseteq ne(X, \mathcal{G}^*) \setminus \mathbf{Z}_{ne}$ such that $\mathbf{S}$ is a clique, **do**
9:     estimate the causal effect θ of X on Y by adjusting for $\mathbf{S} \cup pa(X, \mathcal{G}^*) \setminus \mathbf{Z}_{pa}$, and add θ to Θ,
10: **end for**
11: **return** Θ.

of vertices satisfying the graphical criteria in Theorems 1 and 2, respectively. Then $\mathcal{W} = \{pa(X, \mathcal{G}) \mid \mathcal{G} \in [\mathcal{G}^*]$ *and* $Y \notin pa(X, \mathcal{G})\}$ *is uniformly collapsible over* $\mathbf{Z}_1 \cup \mathbf{Z}_2$ *w.r.t.* $\mathcal{G}^*$ *and* (X, Y).

Hence, with Corollary 1, we can separately find all singleton sets satisfying the graphical criteria in Theorems 1 and 2, respectively, then the union of these singleton sets is a non-trivial set which $\mathcal{W}$ is uniformly collapsible over.

4 ALGORITHM

In this section, we apply the theoretical results given in the last section to modifying IDA. The proposed algorithm, which is called collapsible IDA, is shown in Algorithm 2.

In Algorithm 2, we first use the graphical criterion provided in Theorem 1 to find $\mathbf{Z}_{ne}$. If $\mathbf{Z}_{ne}$ is not identical to $ne(X, \mathcal{G}^*)$, then based on Proposition 3, $pa(X, \mathcal{G}^*)$ is not collapsible. Thus, we simply let $\mathbf{Z}_{pa} = \emptyset$. On the other hand, if $\mathbf{Z}_{ne} = ne(X, \mathcal{G}^*)$, we construct $\mathbf{Z}_{pa}$ based on Theorem 2. Notice that, other criteria in Henckel et al. (2019, Section 3.2) can also be applied to this case. Finally, we enumerate all subsets of $ne(X, \mathcal{G}^*) \setminus \mathbf{Z}_{ne}$ in order to find all cliques, and for each clique $\mathbf{S}$, we estimate one possible causal effect by adjusting for $\mathbf{S} \cup pa(X, \mathcal{G}^*) \setminus \mathbf{Z}_{pa}$.

To avoid enumerating all partially directed paths from Z to Y when building $\mathbf{Z}_{ne}$, we can use the following proposition to further reduce the complexity.

Proposition 4 *Given a CPDAG* $\mathcal{G}^*$ *and three distinct vertices* X, Y *and* $Z \in ne(X, \mathcal{G}^*)$. *Then, every partially directed path from* Z *to* Y *passes through* X *in* $\mathcal{G}^*$ *if and only if for any* $U \in chcomp(X, \mathcal{G}^*) \cap an(Y, \mathcal{G}^*)$, *every partially directed path from* Z *to* U *passes through* X *in* $\mathcal{G}^*$.

Note that any vertex is an ancestor of itself, thus for any $U \in ne(X, \mathcal{G}^*) \cap an(Y, \mathcal{G}^*)$, $\mathcal{W}$ is not uniformly collapsible over U since there is a zero-length path from U to itself, which definitely bypasses X. The next example shows the usefulness of Proposition 4.

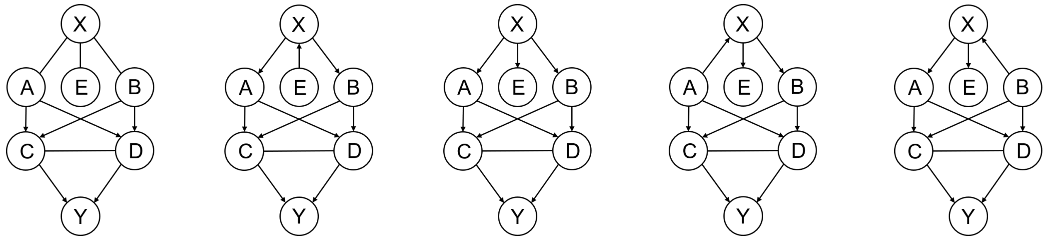
Example 5 *Figure 4 shows an example of finding* $\mathbf{Z}_{ne}$ *with Proposition 4. The CPDAG* $\mathcal{G}^*$ *is shown in Figure 4(a). Since* $ne(X, \mathcal{G}^*) = \{A, B, E\}$ *and* $pa(X, \mathcal{G}^*) = \emptyset$, *all possible parental sets of* X *are* $\{A\}$, $\{B\}$, $\{E\}$ *and* $\emptyset$. *The partially directed graphs given in Figures 4(b) to 4(e) enumerate all the cases. Note that* $chcomp(X) \cap an(Y) = \{A, B\}$, *thus* $\mathcal{W}$ *is not uniformly collapsible over* A, B. *On the other hand, all undirected paths from* E *to* A, B *pass through* X, *hence,* $\mathcal{W}$ *is uniformly collapsible over* E. *In fact,* $X \leftarrow A \rightarrow C \rightarrow Y$ *is a back-door path in* $\mathcal{G}_3^*$, *and* $X \leftarrow B \rightarrow D \rightarrow Y$ *is a back-door path in* $\mathcal{G}_4^*$. *Therefore, both* A *and* B *are needed in some back-door adjustment sets.*

The major difference between the collapsible IDA and the original IDA is that, the collapsible IDA only enumerates the subsets of $ne(X, \mathcal{G}^*) \setminus \mathbf{Z}_{ne}$, while IDA enumerates the subsets of $ne(X, \mathcal{G}^*)$. This modification can reduce both the number of possible parental sets and the size of each possible parental set.

When implementing Algorithm 2, one may use a simple trick to combine the collapsible IDA and the original IDA together. In fact, after line 7 in Algorithm 2, we can remove all edges between $\mathbf{Z}_{ne} \cup \mathbf{Z}_{pa}$ and X, and the resulting graph is a partially directed graph denoted by $\mathcal{H}$. Next, we can simply call IDA with the input graph $\mathcal{H}$, input treatment X, and input target Y. Although $\mathcal{H}$ is not a CPDAG as required by IDA, from the construction given above, it is straightforward to verify that the resulting multi-set is identical to the one returned by Algorithm 2.

5 SIMULATIONS

In this section, we use simulated data to compare our method with IDA (Maathuis et al., 2009). The input CPDAG is either the true CPDAG (Perković et al., 2017), or the one learned from data using the PC algorithm (Maathuis et al., 2009). All experiments were implemented with `R` and run on a computer with 2.50GHz CPU and 8 GB of memory. PC and IDA were called from `pcalg R-package` (Kalisch et al., 2012). All



(a) $\mathcal{G}^*$ (b) $\mathcal{G}_1^*$ (c) $\mathcal{G}_2^*$ (d) $\mathcal{G}_3^*$ (e) $\mathcal{G}_4^*$

Figure 4: An example of finding $\mathbf{Z}_{ne}$ with Proposition 4.

statistical independence tests were performed under the significance level $\alpha = 0.001$.

The data was generated as follows. We first sampled a random DAG $\mathcal{G}$ with 50 vertices and expected degree $d \in \{1, 2, 3, 4, 5\}$ based on a Erdös-Rényi random graph model. Then we generated a joint Gaussian distribution Markovian to this DAG as follows. For each directed edge $X_i \to X_j$, we first independently drawn an edge weight β_{ij} from a Uniform($[0.5, 2]$) or a Uniform($[-2, -0.5] \cup [0.5, 2]$). The DAG $\mathcal{G}$ together with these edge weights $\{\beta_{ij}\}$ gives a distribution over the variable set through the following equations:

$$X_j = \sum_{X_i \in pa(X_j)} \beta_{ij} X_i + \epsilon_j, \quad j = 1, ..., n,$$

where $\epsilon_1, ..., \epsilon_n \ i.i.d. \sim \mathcal{N}(0, 1)$. After obtaining the distribution, we randomly generated two data sets with sample size $N_1 = 1000$ and $N_2 \in \{20, 50\}$, respectively. The first data set was used to learn the CPDAG and the second was used to estimate all possible causal effects. Finally, we sampled an X and used the original IDA and the collapsible IDA ('CIDA' for short) to estimate all possible effects of X on all other variables. The input CPDAG was set to be the true one representing the Markov equivalence class containing $\mathcal{G}$, or the one learned by the PC algorithm. All experiments were repeated 100 times.

We use the following metrics to assess the results. After estimating all possible causal effects of X on other variables, for each method, we computed the total number of possible causal effects of X on all other variables (denoted by N_{IDA} or N_{CIDA}), the maximum size of possible parental sets (denoted by M_{IDA} or M_{CIDA}), and the total estimation bias (denoted by B_{IDA} or B_{CIDA}). The total estimation bias is defined as,

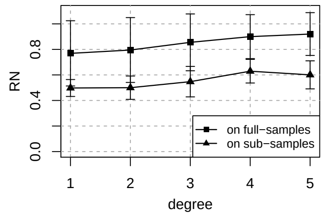
$$B_{\text{method}} = \sqrt{\sum_Y \sum_i (PE_{X \to Y, i} - TE_{X \to Y})^2},$$

where $PE_{X \to Y, i}$ is the i-th possible effect of X on Y estimated with the input CPDAG and $TE_{X \to Y}$ is the true effect of X on Y estimated with the underlying DAG. For ease of comparison, we only stored $RN = N_{\text{CIDA}}/N_{\text{IDA}}$, $RM = M_{\text{CIDA}}/M_{\text{IDA}}$ and $RB = B_{\text{CIDA}}/B_{\text{IDA}}$.

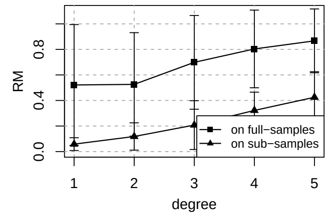
Due to page limits, we only show the results for mixed edge weights in Figure 5. Additional results are given in Appendix B. As discussed in Section 4, if the treatment X does not have any neighbors, the number of possible parental sets cannot be reduced. Thus, in Figure 5, we not only report the average quantities over 100 times repetitions (full-samples), but also report the average quantities over the cases where X has neighbors (sub-samples).

From Figure 5 we can draw the following conclusions. (1) For an arbitrary treatment, the total number of possible effects can be reduced by 10%-20% if we use the collapsible IDA, and for a treatment with neighbors, the total number of possible effects can be approximately reduced by 40%-50%. (2) Using the collapsible IDA can significantly reduce the maximum size of parental sets. For an arbitrary treatment, M_{CIDA} is reduced by 10%-50%, and for a treatment with neighbors, M_{CIDA} is reduced by 60%-90% compared with M_{IDA}. (3) The estimation bias B_{CIDA} is reduced by 5%-20% and 10%-30% for an arbitrary treatment and a treatment with neighbors, respectively, compared with B_{IDA} (4) As the graph becomes dense, the results of the collapsible IDA become less significant.

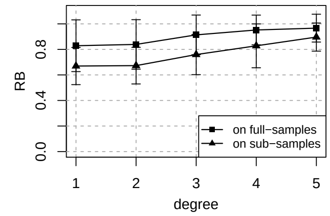
To explain the above results, we recall that collapsing treatments' parents in a CPDAG does not reduce the number of possible effects but only reduce the maximum size of a possible parental set, while collapsing treatments' neighbors can reduce both the number of possible effects and the maximum size of a possible parental set. Therefore, RM is generally lower than RN in the same setting. On the other hand, when the underlying DAG is sparse, the corresponding CPDAG contains many undirected edges. However, when the graph becomes dense, the number of v-structures increases, and thus less undirected edges are in the CPDAG. Therefore, RN, RM and RB increase when we increase the expected degree. Finally, when the input CPDAG is learned from data, we



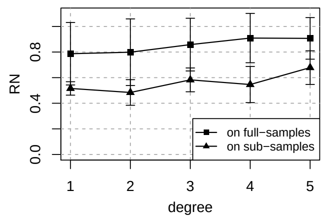
(a) True CPDAG, $N_2 = 20$



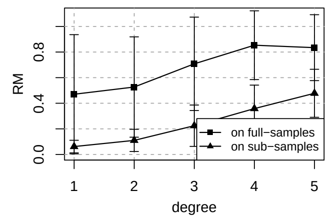
(b) True CPDAG, $N_2 = 20$



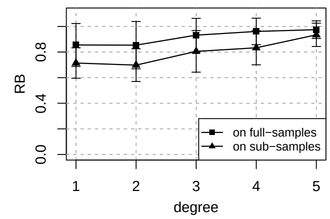
(c) True CPDAG, $N_2 = 20$



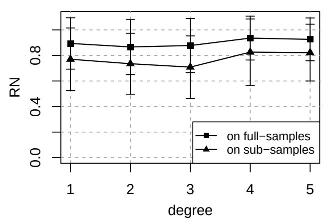
(d) True CPDAG, $N_2 = 50$



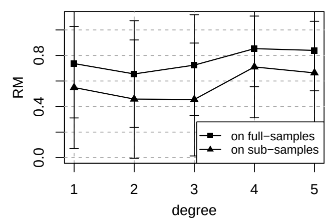
(e) True CPDAG, $N_2 = 50$



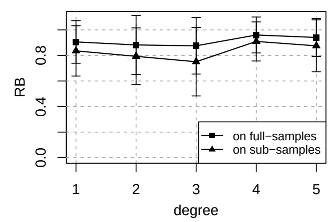
(f) True CPDAG, $N_2 = 50$



(g) Learned CPDAG, $N_2 = 50$



(h) Learned CPDAG, $N_2 = 50$



(i) Learned CPDAG, $N_2 = 50$

Figure 5: Experimental results. The first two rows report the results for using the true CPDAGs as inputs, while the third row reports the results for using the learned CPDAGs. The first row corresponds to $N_2 = 20$, while the second and the third row correspond to $N_2 = 50$. The edge weights were sampled from Uniform($[-2, -0.5] \cup [0.5, 2]$).

empirically find that there are many falsely discovered v-structures in the learned graphs. Consequently, the difference between the collapsible IDA and the original IDA, as well as the distance between two lines in each figure is narrowed.

6 CONCLUDING REMARKS

IDA is a general framework for estimating all possible causal effects of a treatment on a target when the true effect is not identifiable. In this paper, we combine the adjustment set selection procedure with the IDA framework, by providing a method to subtract a common set from all possible parental sets without influencing the back-door adjustment and estimating possible causal effects. With our modification, both the number of possible parental sets and the size of each possible parental set enumerated by IDA decrease, while the local nature of IDA remains unchanged.

There are many possible future directions. For example, how to extend our work to more generalized graphs such as maximal PDAGs is interesting (Perković et al., 2017). Besides, as discussed in Henckel et al. (2019); Andrea & Ezequiel (2019), some additional covariates are beneficial for efficiency, and thus should be included. Therefore, how to extend our work to identify those variables locally and apply them to the IDA framework is also useful.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive suggestions. We would also like to thank Assoc. Prof. Jiji Zhang from Lingnan University for many helpful comments. This work was supported by National Key R&D Program of China (2018YFB1004300), NSFC (11671020, 11771028).

References

Andersson, S. A., Madigan, D., and Perlman, M. D. A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2):505–541, 04 1997.

Andrea, R. and Ezequiel, S. Efficient adjustment sets for population average treatment effect estimation in non-parametric causal graphical models. *arXiv e-prints*, art. arXiv:1912.00306, Dec 2019.

Chickering, D. M. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3(Nov):507–554, 2002.

Geng, Z., Liu, Y., Liu, C., and Miao, W. Evaluation of causal effects and local structure learning of causal networks. *Annual Review of Statistics and Its Application*, 6(1), 2019.

Hauser, A. and Bühlmann, P. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 13(Aug):2409–2464, 2012.

He, Y. and Geng, Z. Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9(Nov):2523–2547, 2008.

He, Y., Jia, J., and Yu, B. Counting and exploring sizes of Markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 16:2589–2609, 2015.

Henckel, L., Perković, E., and Maathuis, M. H. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *arXiv e-prints*, art. arXiv:1907.02435, Jul 2019.

Jaber, A., Zhang, J., and Bareinboim, E. Causal identification under Markov equivalence: Completeness results. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2981–2989. PMLR, 09–15 Jun 2019.

Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., and Bühlmann, P. Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, 47(11):1–26, 2012.

Kuroki, M. and Cai, Z. Selection of identifiability criteria for total effects by using path diagrams. In *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence*, pp. 333–340, 2004.

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

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

Meek, C. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 403–410. Morgan Kaufmann Publishers Inc., 1995.

Nandy, P., Maathuis, M. H., and Richardson, T. S. Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *The Annals of Statistics*, 45(2):647–674, 04 2017.

Pearl, J. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 12 1995.

Pearl, J. *Causality*. Cambridge University Press, 2009.

Pearl, J., Geiger, D., and Verma, T. Conditional independence and its representations. *Kybernetika*, 25(7):33–44, 1989.

Perković, E. Identifying causal effects in maximally oriented partially directed acyclic graphs. *arXiv e-prints*, art. arXiv:1910.02997, Oct 2019.

Perković, E., Textor, J., Kalisch, M., and Maathuis, M. H. A complete generalized adjustment criterion. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pp. 682–691. AUAI Press, 2015.

Perković, E., Kalisch, M., and Maathuis, M. H. Interpreting and using CPDAGs with background knowledge. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence*. AUAI press, 2017.

Perković, E., Textor, J., Kalisch, M., and Maathuis, M. H. Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18(220):1–62, 2018.

Shpitser, I., Vanderweele, T. J., and Robins, J. M. On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pp. 527–536. AUAI press, 2010.

Spirtes, P., Glymour, C. N., and Scheines, R. *Causation, Prediction, and Search*. MIT Press, second edition, 2000.

VanderWeele, T. J. and Shpitser, I. A new criterion for confounder selection. *Biometrics*, 67(4):1406–1413, 2011.

Xie, X. and Geng, Z. Collapsibility for directed acyclic graphs. *Scandinavian Journal of Statistics*, pp. 185–203, 2009.