Interpreting and using CPDAGs with background knowledge

Emilija Perković
Seminar for Statistics
ETH Zurich, Switzerland
perkovic@stat.math.ethz.ch

Markus Kalisch
Seminar for Statistics
ETH Zurich, Switzerland
kalisch@stat.math.ethz.ch

Marloes H. Maathuis
Seminar for Statistics
ETH Zurich, Switzerland
maathuis@stat.math.ethz.ch

Abstract

We develop terminology and methods for working with maximally oriented partially directed acyclic graphs (maximal PDAGs). Maximal PDAGs arise from imposing restrictions on a Markov equivalence class of directed acyclic graphs, or equivalently on its graphical representation as a completed partially directed acyclic graph (CPDAG), for example when adding background knowledge about certain edge orientations. Although maximal PDAGs often arise in practice, causal methods have been mostly developed for CPDAGs. In this paper, we extend such methodology to maximal PDAGs. In particular, we develop methodology to read off possible ancestral relationships, we introduce a graphical criterion for covariate adjustment to estimate total causal effects, and we adapt the IDA and joint-IDA frameworks to estimate multi-sets of possible causal effects. We also present a simulation study that illustrates the gain in identifiability of total causal effects as the background knowledge increases. All methods are implemented in the `R` package `pcalg`.

1 INTRODUCTION

Directed acyclic graphs (DAGs) are used for causal reasoning (e.g., Pearl, 2009), where directed edges represent direct causal effects. In general, it is impossible to learn a DAG from (observational) data. Instead, one can learn a completed partially directed acyclic graph (CPDAG) (e.g., Spirtes et al., 2000; Chickering, 2002), which usually contains some undirected edges and represents a Markov equivalence class of DAGs (see Section 2 for definitions).

Maximally oriented partially directed acyclic graphs (maximal PDAGs) generally contain fewer undirected edges than their corresponding CPDAGs and thus represent fewer DAGs. They arise in various scenarios, for example when adding background knowledge about edge orientations to a CPDAG (Meek, 1995), when imposing a partial ordering (tiers) of variables before conducting causal structure learning (Scheines et al., 1998), in causal structure learning from both observational and interventional data (Hauser and Bühlmann, 2012; Wang et al., 2017), or in structure learning under certain model restrictions (Hoyer et al., 2008; Ernest et al., 2016; Eigenmann et al., 2017).

Maximal PDAGs appear in the literature under various names, such as interventional essential graphs (Hauser and Bühlmann, 2012; Wang et al., 2017), distribution equivalence patterns (Hoyer et al., 2008), aggregated PDAGs (Eigenmann et al., 2017) or simply CPDAGs with background knowledge (Meek, 1995). We refer to them as maximal PDAGs in this paper.

Maximal PDAGs can represent more information about causal relationships than CPDAGs. However, this additional information has not been fully exploited in practice, since causal methods that are applicable to CPDAGs are not directly applicable to general maximal PDAGs.

We now illustrate the difficulties and differences in working with maximal PDAGs, as opposed to CPDAGs. Consider the CPDAG $\mathcal{C}$ in Figure 1a. All DAGs represented by $\mathcal{C}$ (see Figure 1b) have the same adjacencies and unshielded colliders as $\mathcal{C}$. The undirected edge $A - B$ in $\mathcal{C}$ means that there exists a DAG represented by $\mathcal{C}$ that contains $A \to B$, as well as a DAG that contains $A \leftarrow B$. Conversely, if $A \to B$ was in $\mathcal{C}$, then $A \to B$ would be in all DAGs represented by $\mathcal{C}$.

The paths $B - D$, $B - C - D$ and $B - A - D$ are possibly directed paths in $\mathcal{C}$, and hence B is a possible ancestor of D in $\mathcal{C}$. In Figure 1b, we see that there indeed

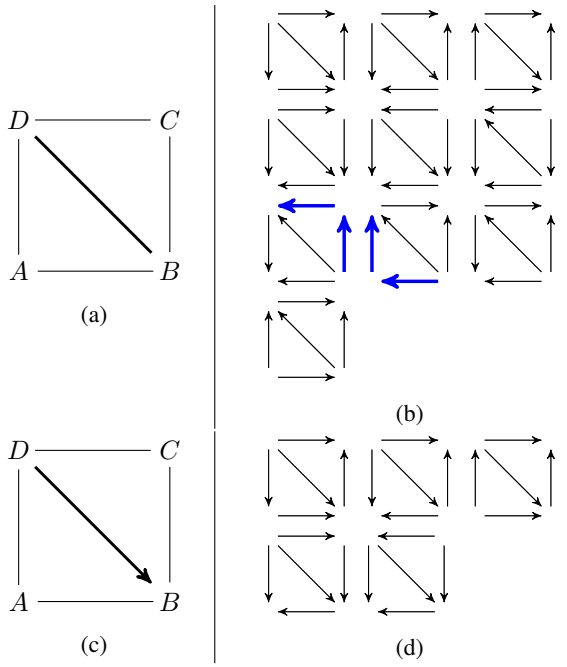

Figure 1: (a) CPDAG $\mathcal{C}$, (b) all DAGs represented by $\mathcal{C}$, (c) maximally oriented PDAG $\mathcal{G}$, (d) all DAGs represented by $\mathcal{G}$.

exist DAGs represented by $\mathcal{C}$ that contain directed paths $B \to D$, $B \to C \to D$, or $B \to A \to D$, and B is an ancestor of D in half of the DAGs represented by $\mathcal{C}$.

Now suppose that we know from background knowledge that $D \to B$. Adding $D \to B$ to $\mathcal{C}$ results in the maximal PDAG $\mathcal{G}$ in Figure 1c. The DAGs represented by $\mathcal{G}$, shown in Figure 1d, are exactly the first five DAGs of Figure 1b. Again, all DAGs represented by $\mathcal{G}$ have the same adjacencies and unshielded colliders as $\mathcal{G}$. Moreover, the interpretation of directed and undirected edges is the same as in the CPDAG.

In Figure 1c, paths $B - C - D$ and $B - A - D$ look like possibly directed paths from B to D, so that one could think that B is a possible ancestor of D. However, due to $D \to B$ and the acyclicity of DAGs, B is not an ancestor of D in any DAG represented by $\mathcal{G}$. Hence, we need a new way of reading off possible ancestral relationships.

An important difference between maximal PDAGs and CPDAGs, is that the former can contain partially directed cycles while the latter cannot. In Figure 1c, $D \to B - C - D$ and $D \to B - A - D$ are examples of such cycles. As a consequence, the important property of CPDAGs which states that if $D \to B - C$ is in a CPDAG $\mathcal{C}$, then $D \to C$ is also in $\mathcal{C}$ (Lemma 1 from Meek, 1995), does not hold in maximal PDAGs.

Furthermore, let $\mathcal{C}_{undir}$ denote the subgraph of a CPDAG $\mathcal{C}$ that contains only the undirected edges from $\mathcal{C}$. It is well known that $\mathcal{C}_{undir}$ is triangulated (chordal) and that the connected components of $\mathcal{C}_{undir}$ can be oriented into DAGs without unshielded colliders, independently of each other and the other edges in $\mathcal{C}$, to form all DAGs represented by $\mathcal{C}$ (see the proof of Theorem 4 in Meek, 1995). This result is not valid in maximal PDAGs. To see this, consider the undirected component $\mathcal{G}_{undir}$ of the maximal PDAG $\mathcal{G}$ in Figure 1c. Then $\mathcal{G}_{undir}$ is $A - B - C - D - A$ and is not triangulated. Orienting $A-B-C-D-A$ into a DAG always leads to at least one unshielded collider. This unshielded collider in $\mathcal{G}_{undir}$ must occur at A or C, so that it is shielded in $\mathcal{G}$ by the edge $D \to B$. Hence, one can no longer orient $\mathcal{G}_{undir}$ independently of the directed edges.

Previous work on maximal PDAGs deals with these issues in two ways, by considering special cases of maximal PDAGs that do not contain partially directed cycles and that retain the same path interpretations as CPDAGs (van der Zander and Liśkiewicz, 2016), or by exhaustively searching the space of DAGs represented by the maximal PDAG to find a solution consistent with all these DAGs (Hyttinen et al., 2015).

In this paper, we develop methodology to work with general maximal PDAGs directly. In **Section 2**, we introduce terminology and definitions. In **Section 3**, we define possible causal relationships in maximal PDAGs and construct a method to easily read these off from a given maximal PDAG.

In **Section 4**, we consider the problem of estimating (possible) total causal effects in maximal PDAGs. In **Section 4.1**, we give a necessary and sufficient graphical criterion for computing total causal effects in maximal PDAGs via covariate adjustment. Our criterion is called the b-adjustment criterion ('b' for background), and builds on Shpitser et al. (2010); Shpitser (2012) and Perković et al. (2015, 2017). We also construct fast algorithms for finding all adjustment sets in maximal PDAGs, using results from van der Zander et al. (2014). In **Section 4.2**, we no longer focus on total causal effects that are identifiable via covariate adjustment, but rather consider computing all possible total causal effects of a node set $\mathbf{X}$ on a node Y in a maximal PDAG $\mathcal{G}$, and collect all these values in a multi-set. Such an approach already exists for CPDAGs under the name (joint-)IDA (Maathuis et al., 2009, 2010; Nandy et al., 2017). We develop an efficient semi-local version of (joint-)IDA that is applicable to maximal PDAGs.

In **Section 5**, we discuss our implementation and present a simulation study where we apply our b-adjustment criterion and semi-local IDA algorithm to CPDAGs with various amounts of added background knowledge. We

demonstrate that background knowledge can help considerably in identifying total causal effects via covariate adjustment and in reducing the number of unique values in estimated multi-sets of possible total causal effects. We close with a discussion in **Section 6**. All proofs can be found in the supplementary material.

2 PRELIMINARIES

Nodes, edges and subgraphs. A graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ consists of a set of nodes (variables) $\mathbf{V} = \{X_1, \ldots, X_p\}$ and a set of edges $\mathbf{E}$. There is at most one edge between any two nodes, and two nodes are *adjacent* if an edge connects them. We call $\rightarrow$ a directed and $-$ a undirected edge. An *induced subgraph* $\mathcal{G}' = (\mathbf{V}', \mathbf{E}')$ of $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ consists of a subset of nodes $\mathbf{V}' \subseteq \mathbf{V}$ and edges $\mathbf{E}' \subseteq \mathbf{E}$ where $\mathbf{E}'$ are all edges in $\mathbf{E}$ between nodes in $\mathbf{V}'$.

Paths. A *path* p from X to Y in $\mathcal{G}$ is a sequence of distinct nodes $\langle X, \ldots, Y \rangle$ in which every pair of successive nodes is adjacent. A node V *lies on a path* p if V occurs in the sequence of nodes. If $p = \langle X_1, X_2, \ldots, X_k, \rangle, k \geq 2$, then X_1 and X_k are *endpoints* of p, and any other node $X_i, 1 < i < k$, is a *non-endpoint* node on p. A *directed path* or *causal path* from X to Y is a path from X to Y in which all edges are directed towards Y, that is $X \rightarrow \cdots \rightarrow Y$. A *possibly directed path* or *possibly causal path* from X to Y is a path from X to Y that does not contain an edge directed towards X. A *non-causal* path from X to Y contains at least one edge directed towards X. Throughout, we will refer to a path p from X to Y as non-causal (causal) if and only if it is non-causal (causal) from X to Y. For two disjoint subsets $\mathbf{X}$ and $\mathbf{Y}$ of $\mathbf{V}$, a path from $\mathbf{X}$ to $\mathbf{Y}$ is a path from some $X \in \mathbf{X}$ to some $Y \in \mathbf{Y}$. A path from $\mathbf{X}$ to $\mathbf{Y}$ is *proper* (w.r.t. $\mathbf{X}$) if only its first node is in $\mathbf{X}$. If $\mathcal{G}$ and $\mathcal{G}^*$ are two graphs with identical adjacencies and p is a path in $\mathcal{G}$, then the *corresponding path* p^* in $\mathcal{G}^*$ consists of the same node sequence as p.

Partially directed and directed cycles. A directed path from X to Y, together with $Y \rightarrow X$, forms a *directed cycle*. A *partially directed cycle* is formed by a possibly directed path from X to Y, together with $Y \rightarrow X$.

Subsequences and subpaths. A *subsequence* of a path p is obtained by deleting some nodes from p without changing the order of the remaining nodes. A subsequence of a path is not necessarily a path. For a path $p = \langle X_1, X_2, \ldots, X_m \rangle$, the *subpath* from X_i to X_k ($1 \leq i \leq k \leq m$) is the path $p(X_i, X_k) = \langle X_i, X_{i+1}, \ldots, X_k \rangle$.

Ancestral relationships. If $X \rightarrow Y$, then X is a *parent* of Y. If $X - Y$, then X is a *sibling* of Y. If there is a causal path from X to Y, then X is an *ancestor* of Y, and Y is a *descendant* of X. We also use the convention that every node is a descendant and an ancestor of itself. The sets of parents, siblings, ancestors and descendants of X in $\mathcal{G}$ are denoted by $\mathrm{Pa}(X, \mathcal{G})$, $\mathrm{Sib}(X, \mathcal{G})$, $\mathrm{An}(X, \mathcal{G})$ and $\mathrm{De}(X, \mathcal{G})$ respectively. For a set of nodes $\mathbf{X} \subseteq \mathbf{V}$, we let $\mathrm{Pa}(\mathbf{X}, \mathcal{G}) = \cup_{X \in \mathbf{X}} \mathrm{Pa}(X, \mathcal{G})$, with analogous definitions for $\mathrm{Sib}(\mathbf{X}, \mathcal{G})$, $\mathrm{An}(\mathbf{X}, \mathcal{G})$ and $\mathrm{De}(\mathbf{X}, \mathcal{G})$.

Colliders, shields and definite status paths. If a path p contains $X_i \rightarrow X_j \leftarrow X_k$ as a subpath, then X_j is a *collider* on p. A path $\langle X_i, X_j, X_k \rangle$ is an *(un)shielded triple* if X_i and X_k are (not) adjacent. A path is *unshielded* if all successive triples on the path are unshielded. A node X_j is a *definite non-collider* on a path p if there is at least one edge out of X_j on p, or if $X_i - X_j - X_k$ is a subpath of p and $\langle X_i, X_j, X_k \rangle$ is an unshielded triple. A node is of *definite status* on a path if it is a collider, a definite non-collider or an endpoint on the path. A path p is of definite status if every node on p is of definite status.

D-connection and blocking. A definite status path p from X to Y is *d-connecting* given a node set $\mathbf{Z}$ ($X, Y \notin \mathbf{Z}$) if every definite non-collider on p is not in $\mathbf{Z}$, and every collider on p has a descendant in $\mathbf{Z}$. Otherwise, $\mathbf{Z}$ blocks p.

DAGs, PDAGs and CPDAGs. A *directed graph* contains only directed edges. A *partially directed graph* may contain both directed and undirected edges. A directed graph without directed cycles is a *directed acyclic graph* (DAG). A *partially directed acyclic graph* (PDAG) is a partially directed graph without directed cycles.

Two disjoint node sets $\mathbf{X}$ and $\mathbf{Y}$ in a DAG $\mathcal{D}$ are *d-separated* given a node set $\mathbf{Z}$ (pairwise disjoint with $\mathbf{X}$ and $\mathbf{Y}$) if and only if every path from $\mathbf{X}$ to $\mathbf{Y}$ is blocked by $\mathbf{Z}$. Several DAGs can encode the same d-separation relationships. Such DAGs form a *Markov equivalence class* which is uniquely represented by a *completed partially directed acyclic graph* (CPDAG) (Meek, 1995). A directed edge $X \rightarrow Y$ in a CPDAG $\mathcal{C}$ corresponds to $X \rightarrow Y$ in every DAG in the Markov equivalence class represented by $\mathcal{C}$. For any undirected edge $X - Y$ in a CPDAG $\mathcal{C}$, the Markov equivalence class represented by $\mathcal{C}$ contains a DAG with $X \rightarrow Y$ and a DAG with $X \leftarrow Y$.

A PDAG $\mathcal{G}'$ is *represented* by another PDAG $\mathcal{G}$ (equivalently $\mathcal{G}$ represents $\mathcal{G}'$) if $\mathcal{G}'$ and $\mathcal{G}$ have the same adjacencies and unshielded colliders and every directed edge $X \rightarrow Y$ in $\mathcal{G}$ is also in $\mathcal{G}'$.

Maximal PDAGs. A PDAG $\mathcal{G}$ is a *maximally oriented* PDAG if and only if the edge orientations in $\mathcal{G}$ are closed under the orientation rules in Figure 2. Throughout, we will refer to maximally oriented PDAGs as *maximal PDAGs*.

Algorithm 1: ConstructMaxPDAG

Data: maximal PDAG $\mathcal{G}$, background knowledge $\mathbf{R}$
Result: maximal PDAG $\mathcal{G}'$ or FAIL
Let $\mathcal{G}' = \mathcal{G}$;
while $\mathbf{R} \neq \emptyset$ **do**
  Choose an edge $\{X \to Y\}$ in $\mathbf{R}$;
  $\mathbf{R} = \mathbf{R} \setminus \{X \to Y\}$;
  if $\{X - Y\}$ *or* $\{X \to Y\}$ *is in* $\mathcal{G}'$ **then**
    Orient $\{X \to Y\}$ in $\mathcal{G}'$;
    Close the edge orientations under the rules in Figure 2 in $\mathcal{G}'$;
  else
    FAIL;
  end
end

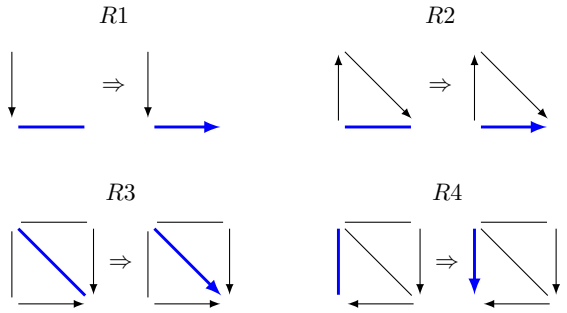

Figure 2: The orientation rules from Meek (1995). If the graph on the left-hand side of a rule is an induced subgraph of a PDAG $\mathcal{G}$, then *orient* the blue undirected edge ($-$) as shown on the right-hand side of the rule.

Let $\mathbf{R}$ be a set of required directed edges representing background knowledge. Algorithm 1 of Meek (1995) describes how to incorporate background knowledge $\mathbf{R}$ in a maximal PDAG $\mathcal{G}$. If Algorithm 1 does not return a FAIL, then it returns a new maximal PDAG $\mathcal{G}'$ that is represented by $\mathcal{G}$. Background knowledge $\mathbf{R}$ is *consistent* with maximal PDAG $\mathcal{G}$ if and only if Algorithm 1 does not return a FAIL (Meek, 1995).

$\mathcal{G}$ and $[\mathcal{G}]$. If $\mathcal{G}$ is a maximal PDAG, then $[\mathcal{G}]$ denotes every maximal PDAG represented by $\mathcal{G}$. Thus, $[\mathcal{G}]$ contains all DAGs represented by $\mathcal{G}$, but also all PDAGs (with the same adjacencies and unshielded colliders) that contain more orientations than $\mathcal{G}$.

Causal DAGs and PDAGs. A density f of $\mathbf{V} = \{X_1, \ldots, X_p\}$ is *consistent* with a DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ if it factorizes as $f(\mathbf{V}) = \prod_{i=1}^{p} f(X_i|Pa(X_i, \mathcal{G}))$ (Pearl, 2009). A DAG is *causal* if every edge $X_i \to X_j$ in $\mathcal{D}$ represents a direct causal effect of X_i on X_j (wrt $\mathbf{V}$). A PDAG is *causal* if it represents a causal DAG.

We consider interventions $do(\mathbf{X} = \mathbf{x})$ (for $\mathbf{X} \subseteq \mathbf{V}$) or $do(\mathbf{x})$ for shorthand, which represent outside interventions that set $\mathbf{X}$ to $\mathbf{x}$ (Pearl, 2009). A density f of $\mathbf{V} = \{X_1, \ldots, X_p\}$ is *consistent with a causal DAG* $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ if all post-intervention densities $f(\mathbf{v}|do(\mathbf{x}))$ factorize as:

$$f(\mathbf{v}|do(\mathbf{x})) = \\ = \begin{cases} \prod_{X_i \in \mathbf{V} \setminus \mathbf{X}} f(x_i|\operatorname{Pa}(x_i, \mathcal{D})), & \text{if } \mathbf{X} = \mathbf{x}, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Equation (1) is known as the truncated factorization formula (Pearl, 2009), manipulated density formula (Spirtes et al., 2000) or the g-formula (Robins, 1986). A density f is *consistent with a causal PDAG* $\mathcal{G}$ if it is consistent with a causal DAG in $[\mathcal{G}]$.

3 POSSIBLY CAUSAL RELATIONSHIPS IN MAXIMAL PDAGs

A basic task for causal reasoning based on a maximal PDAG $\mathcal{G}$ is reading off possible causal relationships between nodes directly from $\mathcal{G}$. The alternative is of course to list all DAGs in $[\mathcal{G}]$ and to read off the causal relationship in each DAG. If $\mathcal{G}$ has many undirected edges, this task quickly becomes a computational burden. We show that possible causal relationships can be read off from the maximal PDAG directly if we modify the definitions of possibly causal and non-causal paths.

In Definition 3.1, we define a *b-possibly causal* path and its complement the *b-non-causal* path for maximal PDAGs. Throughout, we use the prefix *b-* (for background) to distinguish from the analogous terms for CPDAGs. Lemma 3.2 demonstrates the soundness of Definition 3.1.

Definition 3.1. ***(b-possibly causal path, b-non-causal path)*** *Let $p = \langle X = V_0, \ldots, V_k = Y \rangle$, $k \geq 1$ be a path from node X to node Y in a maximal* PDAG *$\mathcal{G}$. We say that p is b-possibly causal in $\mathcal{G}$ if and only if no edge $V_i \leftarrow V_j, 0 \leq i < j \leq k$ is in $\mathcal{G}$. Otherwise, we say that p is b-non-causal path in $\mathcal{G}$.*

Remark. Note that Definition 3.1 involves edges that are not on p.

Lemma 3.2. *Let p^* be a path from X to Y in a maximal* PDAG *$\mathcal{G}$. If p^* is b-non-causal in $\mathcal{G}$, then for every* DAG *$\mathcal{D}$ in $[\mathcal{G}]$ the corresponding path in $\mathcal{D}$ is non-causal. By contraposition, if p is a causal path in at least one* DAG *$\mathcal{D}$ in $[\mathcal{G}]$, then the corresponding path in $\mathcal{G}$ is b-possibly causal.*

Note that since every CPDAG is also a maximal PDAG, all results in this paper subsume existing results for

CPDAGs. We now define *b-possible descendants* (*b-possible ancestors*) in a maximal PDAG.

Definition 3.3. ($\text{b-PossDe}(\mathbf{X}, \mathcal{G})$ *and* $\text{b-PossAn}(\mathbf{X}, \mathcal{G})$) *Let* $\mathbf{X}$ *be a node set in a maximal* PDAG $\mathcal{G}$*. Then* W *is a b-possible descendant (b-possible ancestor) of* $\mathbf{X}$ *in* $\mathcal{G}$*, and we write* $W \in \text{b-PossDe}(\mathbf{X}, \mathcal{G})$ ($W \in \text{b-PossAn}(\mathbf{X}, \mathcal{G})$) *if and only if* $W \in \mathbf{X}$ *or there is a node* $X \in \mathbf{X}$ *and a b-possibly causal path from* X *to* W (W *to* X) *in* $\mathcal{G}$*.*

Example 3.4. *We use this example to illustrate b-possibly causal paths and b-possible descendants. Consider* CPDAG $\mathcal{C}$ *in Figure 1a and maximal* PDAG $\mathcal{G}$ *in Figure 1c. We see that* $B - C - D$ *is a b-possibly causal path in* $\mathcal{C}$*, and a b-non-causal path in* $\mathcal{G}$ *due to* $D \to B$ *in* $\mathcal{G}$*. Conversely,* $D - C - B$ *is a b-possibly causal path in both* $\mathcal{C}$ *and* $\mathcal{G}$*.*

Furthermore, $\text{b-PossDe}(B, \mathcal{C}) = \{A, B, C, D\}$ *and* $\text{b-PossDe}(B, \mathcal{G}) = \{A, B, C\}$*. The b-possible descendants of* A, C *and* D *nodes are the same in* $\mathcal{G}$ *and* $\mathcal{C}$*.*

3.1 EFFICIENTLY FINDING ALL B-POSSIBLE DESCENDANTS/ANCESTORS

The b-possibly causal paths present an elegant extension of the notion of possibly causal paths from CPDAGs to maximal PDAGs. Nevertheless, finding all b-possible descendants by checking the b-possibly causal status of a path is non trivial, since it involves considering many edges not on the path. This is cumbersome if the graphs we are dealing with are large and/or dense.

To solve this issue, we use Lemma 3.6 (analogous to Lemma B.1 in Zhang, 2008) to only consider paths which are unshielded and hence, of definite status. We then only need consider the edges which are on the definite status paths (Lemma 3.5). Thus, the task of finding all b-possible descendants (ancestors) of $\mathbf{X}$ in a maximal PDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ can be done using a depth first search algorithm with computational complexity $O(|\mathbf{V}| + |\mathbf{E}|)$.

Lemma 3.5. *Let* $p^* = \langle V_1, \ldots, V_k \rangle$ *be a definite status path in a maximal* PDAG $\mathcal{G}$*. Then* p^* *is b-possibly causal if and only if there is no* $V_i \leftarrow V_{i+1}$*, for* $i \in \{1, \ldots, k-1\}$ *in* $\mathcal{G}$*.*

Lemma 3.6. *Let* X *and* Y *be distinct nodes in a maximal* PDAG $\mathcal{G}$*. If* p *is s a b-possibly causal path from* X *to* Y *in* $\mathcal{G}$*, then a subsequence* p^* *of* p *forms a b-possibly causal unshielded path from* X *to* Y *in* $\mathcal{G}$*.*

4 ESTIMATING TOTAL CAUSAL EFFECTS WITH MAXIMAL PDAGS

After inferring the existence of a possibly non-zero total causal effect of $\mathbf{X}$ on $\mathbf{Y}$ in a maximal PDAG $\mathcal{G}$ by finding a b-possible causal path from X to Y, the natural question to ask is, how big is this effect? How do we calculate this effect using observational data? Throughout, let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ represent a causal maximal PDAG, and let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ be pairwise disjoint subsets of $\mathbf{V}$. We are interested in the total causal effect of $\mathbf{X}$ on $\mathbf{Y}$.

4.1 ADJUSTMENT IN MAXIMAL PDAGs

The most commonly used tool for estimating total causal effects from observational data is covariate adjustment. We first define the concept of a adjustment sets for maximal PDAGs.

Definition 4.1. ***(Adjustment set)*** *Let* $\mathbf{X}, \mathbf{Y}$ *and* $\mathbf{Z}$ *be pairwise disjoint node sets in a causal maximal* PDAG $\mathcal{G}$*. Then* $\mathbf{Z}$ *is an adjustment set relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *if for any density* f *consistent with* $\mathcal{G}$ *we have*

$$f(\mathbf{y}|do(\mathbf{x})) = \begin{cases} f(\mathbf{y}|\mathbf{x}) & \text{if } \mathbf{Z} = \emptyset, \\ \int_{\mathbf{z}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}) f(\mathbf{z}) d\mathbf{z} & \text{otherwise.} \end{cases}$$

Thus, if $\mathbf{Z}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{G}$, we do not need to find the true underlying causal DAG in order to find the post-intervention density $f(\mathbf{y}|do(\mathbf{x}))$.

Under the assumption that the density consistent with the underlying causal DAG can be generated by a linear structural equation model (SEM) with additive noise (see Ch.5 in Pearl, 2009, for definition of SEMs), covariate adjustment allows the researcher to estimate the total causal effect by performing one multiple linear regression (see Nandy et al., 2017, for the non-Gaussian noise result). In this setting, if $\mathbf{Z}$ is an adjustment set relative to some nodes X, Y in $\mathcal{G}$, then the coefficient of X in the linear regression of Y on X and $\mathbf{Z}$ is the total causal effect of X on Y.

There has been recent work on finding graphical criteria for sets that satisfy Definition 4.1 relative to some $(\mathbf{X}, \mathbf{Y})$. The first such criterion was the back-door criterion (Pearl, 2009), which is sound but not complete for adjustment (wrt Definition 4.1). The sound and complete adjustment criterion for DAGs was introduced in Shpitser et al. (2010); Shpitser (2012) and extended to more general graph classes in Perković et al. (2015, 2017). Furthermore, van der Zander and Liśkiewicz (2016) presented an adjustment criterion that can be used in maximal PDAGs that do not contain partially directed cycles. None of these criteria is applicable to general maximal PDAGs.

We present our b-adjustment criterion in Definition 4.3, building on the results of Shpitser et al. (2010); Shpitser (2012) and Perković et al. (2015, 2017). Our criterion, as well as other results presented in this section, is phrased in the same way as the results in Perković et al. (2015, 2017). The only difference is the use of b-possibly causal paths as opposed to possibly causal paths. This similarity is intentional, as it makes our results easier to follow and further demonstrates that the results for DAGs and CPDAGs can be leveraged for maximal PDAGs. The proofs do not follow from previous results and require special consideration, especially due to the partially directed cycles that can occur in maximal PDAGs. In DAGs and CPDAGs our b-adjustment criterion reduces to the adjustment criteria from Shpitser et al. (2010); Shpitser (2012) and Perković et al. (2015, 2017). For consistency however, we will refer to the b-adjustment criterion for all graph types.

To define our b-adjustment criterion, we first introduce the concept of the b-forbidden set for maximal PDAGs. The b-forbidden set contains all nodes that cannot be used for adjustment.

Definition 4.2. $(\text{b-Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}))$ *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be disjoint node sets in a maximal* PDAG $\mathcal{G}$*. We define the b-forbidden set relative to* $(\mathbf{X}, \mathbf{Y})$ *as:*

$$\text{b-Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \{W' \in \mathbf{V} : W' \in \text{b-PossDe}(W, \mathcal{G}), \text{ for some } W \notin \mathbf{X} \text{ which lies on a proper b-possibly causal path from } \mathbf{X} \text{ to } \mathbf{Y} \text{ in } \mathcal{G}\}.$$

Definition 4.3. ***(b-adjustment criterion)*** *Let* $\mathbf{X}, \mathbf{Y}$ *and* $\mathbf{Z}$ *be pairwise disjoint node sets in a maximal* PDAG $\mathcal{G}$*. Then* $\mathbf{Z}$ *satisfies the b-adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *if:*

- ***(b-amenability)*** *all proper b-possibly causal paths from* $\mathbf{X}$ *to* $\mathbf{Y}$ *start with a directed edge out of* $\mathbf{X}$ *in* $\mathcal{G}$*,*
- ***(b-forbidden set)*** $\mathbf{Z} \cap \text{b-Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$*,*
- ***(b-blocking)*** *all proper b-non-causal definite status paths from* $\mathbf{X}$ *to* $\mathbf{Y}$ *are blocked by* $\mathbf{Z}$ *in* $\mathcal{G}$*.*

In Theorem 4.4 we show that the b-adjustment criterion is sound and complete for adjustment. Furthermore, in Theorem 4.6 we show that there exists an adjustment set if and only if the specific set $\text{b-Adjust}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ (Definition 4.5) is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{G}$.

Theorem 4.4. *Let* $\mathbf{X}, \mathbf{Y}$ *and* $\mathbf{Z}$ *be pairwise disjoint node sets in a causal maximal* PDAG $\mathcal{G}$*. Then* $\mathbf{Z}$ *is an adjustment set (Definition 4.1) relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *if and only if* $\mathbf{Z}$ *satisfies the b-adjustment criterion (Definition 4.3) relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$*.*

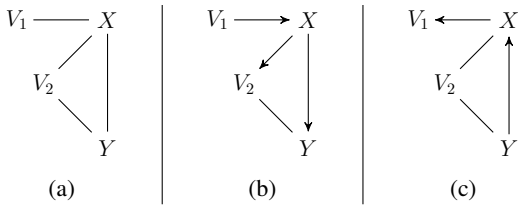

Figure 3: (a) CPDAG $\mathcal{C}$, (b) maximal PDAG $\mathcal{G}_1$, (c) maximal PDAG $\mathcal{G}_2$ used in Example 4.7.

Definition 4.5. $(\text{b-Adjust}(\mathbf{X}, \mathbf{Y}, \mathcal{G}))$ *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be disjoint node sets in a maximal* PDAG $\mathcal{G}$*. We define*

$$\text{b-Adjust}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \text{b-PossAn}(\mathbf{X} \cup \mathbf{Y}, \mathcal{G}) \setminus (\mathbf{X} \cup \mathbf{Y} \cup \text{b-Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})).$$

Theorem 4.6. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be disjoint node sets in a maximal* PDAG $\mathcal{G}$*. There exists a set that satisfies the b-adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *if and only if* $\text{b-Adjust}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ *satisfies the b-adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$*.*

Example 4.7. *We use this example to illustrate the b-adjustment criterion. Consider Figure 3 with a* CPDAG $\mathcal{C}$ *in (a), and two maximal* PDAG*s* $\mathcal{G}_1$ *and* $\mathcal{G}_2$ *in* $[\mathcal{G}]$ *in (b) and (c). Maximal* PDAG $\mathcal{G}_1$ *(*$\mathcal{G}_2$*) can be obtained from* $\mathcal{C}$ *by adding* $V_1 \to X$ *(*$Y \to X$*) as background knowledge and completing the orientation rules from Figure 2.*

$\mathcal{C}$ is not b-amenable relative to (X, Y) *due to* $X - Y$ *and* $X - V_2 - Y$*. Hence, there is no adjustment set (no set satisfies Definition 4.3) relative to* (X, Y) *in* $\mathcal{G}$*.*

On the other hand, $\mathcal{G}_1$ *is b-amenable relative to* (X, Y) *and* $\text{b-Forb}(X, Y, \mathcal{G}_1) = \{V_2, Y\}$*. Since there are no b-non-causal paths from* X *to* Y *in* $\mathcal{G}_1$ *any set of nodes disjoint with* $\text{b-Forb}(X, Y, \mathcal{G}_1) \cup \{X, Y\}$ *satisfies the b-adjustment criterion relative to* (X, Y)*. Hence, all valid adjustment sets relative to* (X, Y) *in* $\mathcal{G}_1$ *are* $\emptyset$ *and* $\{V_1\}$*.*

Maximal PDAG $\mathcal{G}_2$ *is also b-amenable relative to* (X, Y) *(since* $X - V_2 - Y$ *is a b-non-causal path). Since* $X \leftarrow Y$ *is in* $\mathcal{G}_2$*,* $\text{b-Forb}(X, Y, \mathcal{G}_2) = \emptyset$*. However, since* $X \leftarrow Y$ *is a proper b-non-causal definite status path from* X *to* Y *that cannot be blocked by any set of nodes, there is no adjustment set (no set satisfies Definition 4.3) relative to* (X, Y) *in* $\mathcal{G}_2$*. Nevertheless, since* $Y \notin \text{b-PossDe}(X, \mathcal{G}_2)$*, we can conclude that the total causal effect of* X *on* Y *in* $\mathcal{G}_2$ *is zero.*

4.1.1 Constructing adjustment sets

Checking whether $\mathbf{Z}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{G}$ requires checking the three conditions in Definition 4.3: b-amenability, b-forbidden set and b-blocking. Checking the b-amenability condition or the

b-forbidden set condition is computationally straightforward and depends only on constructing the set of all b-possible descendants of $\mathbf{X}$. Naively checking the b-blocking condition, however, requires keeping track of all paths between $\mathbf{X}$ and $\mathbf{Y}$. This scales very poorly with the size of the graph. To deal with this issue we rely on Lemma C.3 in the supplement, which is analogous to Lemma 10 in Perković et al., 2017.

If $\mathcal{G}$ is b-amenable and $\mathbf{Z}$ satisfies the b-forbidden set condition relative to $(\mathbf{X}, \mathbf{Y})$, then in order to verify that $\mathbf{Z}$ satisfies the b-blocking condition in $\mathcal{G}$ it is enough to verify that $\mathbf{Z}$ satisfies the b-blocking condition in one DAG in $[\mathcal{G}]$ ((iii) in Lemma C.3 in the supplement). In van der Zander et al. (2014), the authors propose fast algorithms to verify the b-blocking condition and construct adjustment sets in DAGs. Using the above mentioned result we can leverage these for use in maximal PDAGs.

The computational complexity of finding one DAG $\mathcal{D}$ in $[\mathcal{G}]$ is $O(|\mathbf{V}||\mathbf{E}|)$, as shown in Dor and Tarsi (1992). The computational complexity of verifying whether a set satisfies the b-adjustment criterion in $\mathcal{D}$ is $O(|\mathbf{V}| + |\mathbf{E}|)$, as shown in van der Zander et al. (2014). Furthermore, the computational complexity of listing all or all minimal adjustment sets in $\mathcal{D}$, is at most polynomial in $|\mathbf{V}|$ and $|\mathbf{E}|$ per output set, as shown in van der Zander et al. (2014). The complexity of verifying the b-blocking condition in $\mathcal{G}$ when exploiting the results of Lemma C.3, Dor and Tarsi (1992) and van der Zander et al. (2014) is polynomial in $|\mathbf{V}|$ and $|\mathbf{E}|$ and listing all or all minimal adjustment sets in $\mathcal{G}$ is polynomial in $|\mathbf{V}|$ and $|\mathbf{E}|$ per output set.

4.2 IDA AND JOINT-IDA IN MAXIMAL PDAGs

It is not always possible to find an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in a maximal PDAG $\mathcal{G}$. For example, if the total causal effect of $\mathbf{X}$ on $\mathbf{Y}$ differs in some distinct DAGs in $[\mathcal{G}]$, then this effect is not identifiable in $\mathcal{G}$, and is certainly not identifiable via adjustment.

The IDA algorithm for CPDAGs from Maathuis et al. (2009) was developed with precisely this issue in mind. In order to estimate the possible total causal effects of a node X on a node Y based on a CPDAG $\mathcal{C}$ one can consider listing all DAGs in $[\mathcal{C}]$ and estimating the total causal effect of X on Y in each. Since this effect may differ between different DAGs in $\mathcal{G}$, the output of such an algorithm is a multi-set of possible total causal effects of X on Y. The joint-IDA algorithm for CPDAGs from Nandy et al. (2017) employs the same idea to estimate the possible total joint causal effect of a node set $\mathbf{X}$ on a node Y.

The IDA and joint-IDA algorithms use all possible (joint) parent sets of $\mathbf{X}$ in the DAGs in $[\mathcal{G}]$. If Y is not a parent of X in a causal DAG $\mathcal{D}$, it is well known that $\mathrm{Pa}(X, \mathcal{D})$ is an adjustment set relative to (X, Y) in $\mathcal{D}$ (Pearl, 2009).

Algorithm 2: Semi-locally find all joint parent sets of $\mathbf{X} = (X_1, \ldots, X_k), k \geq 1$, in a maximal PDAG $\mathcal{G}$.

Data: maximal PDAG $\mathcal{G}$, $\mathbf{X} = \{X_1, \ldots, X_k\}, k \geq 1$
Result: Multi-set $\mathbf{PossPa}$ of all joint parent sets of $\mathbf{X}$
$\mathbf{PossPa} = \emptyset$;
$\mathbf{Sib_1} = \{A : A - X_1 \text{ in } \mathcal{G}\}$;
if $k > 1$ **then**
  $\mathbf{Sib}_i = \{A : A - X_i \text{ in } \mathcal{G}\} \setminus \{X_1, \ldots, X_{i-1}\}$;
end
forall $\mathbf{S}_i \subseteq \mathbf{Sib}_i$, $i = 1, \ldots, k$, **do**
  $\mathbf{LocalBg} = \cup_{i=1}^{k} \{A \to X_i : A \in \mathbf{S}_i\} \cup \cup_{i=1}^{k} \{X_i \to A : A \in \mathbf{Sib}_i \setminus \mathbf{S}_i\}$;
  if $\mathrm{ConstructMaxPDAG}(\mathcal{G}, \mathbf{LocalBg}) \neq$ *FAIL* **then**
    add $(\mathrm{Pa}(X_1, \mathcal{G}) \cup \mathbf{S}_1, \ldots, \mathrm{Pa}(X_k, \mathcal{G}) \cup \mathbf{S}_k)$ to $\mathbf{PossPa}$;
  end
end

The (joint-)IDA algorithm has different algorithmic variants for finding (joint) parent sets: global, local and semi-local. The global variant applies the above mentioned idea of listing all DAGs represented by a CPDAG $\mathcal{C}$ in order to find all possible (joint) parent sets of $\mathbf{X}$. This method can be applied directly to a maximal PDAG $\mathcal{G}$, but scales poorly with the number of undirected edges in $\mathcal{G}$. The local variant of IDA from Maathuis et al. (2009) and the semi-local variant of joint-IDA from Nandy et al. (2017) dramatically reduce the computational complexity of the algorithm. However, they are not applicable to maximal PDAGs, as we demonstrate in Examples 4.8 and 4.9.

In Algorithm 2, we present our semi-local variant for finding possible (joint) parent sets in maximal PDAGs. This algorithm exploits the soundness and completeness of orientations in the maximal PDAG (Meek, 1995). To find a possible parent set of X, Algorithm 2 simply considers all $\mathbf{S} \subseteq \mathrm{Sib}(X, \mathcal{G})$ and imposes $S \to X$ and $X \to \bar{S}$, for all $S \in \mathbf{S}$, and $\bar{S} \in \mathrm{Sib}(X, \mathcal{G}) \setminus \mathbf{S}$ as background knowledge called $\mathbf{LocalBg}$. If this background knowledge is consistent with $\mathcal{G}$, then $\mathbf{S} \cup \mathrm{Pa}(X, \mathcal{G})$ is a possible parent set of X. In case of joint interventions ($|\mathbf{X}| = k > 1$), Algorithm 2 does this for every $X_i \in \mathbf{X}, i \in \{1, \ldots, k\}$ and $(\mathrm{Pa}(X_1, \mathcal{G}) \cup \mathbf{S}_1, \ldots, \mathrm{Pa}(X_k, \mathcal{G}) \cup \mathbf{S}_k)$ is the possible joint parent set of $\mathbf{X}$.

Since the local IDA and semi-local joint-IDA are not applicable to maximal PDAGs, we can only compare the computational complexity of our Algorithm 2 with the local IDA and semi-local joint-IDA by applying them all

to a CPDAG. In this case, Algorithm 2 will in general be slower than local IDA, as it requires closing the orientation rules of Meek (1995), but also in general somewhat faster than semi-local joint IDA, since it does not require orienting an entire undirected component of the CPDAG. We compare the runtimes of local IDA and our semi-local IDA in an empirical study in Section D of the supplement.

4.2.1 Examples

Example 4.8. *Consider the maximal* PDAG $\mathcal{G}$ *in Figure 1c. Suppose we want to estimate the total causal effect of* C *on* A *in* $\mathcal{G}$*. All possible parent sets of* C *according to Algorithm 2 are:* $\emptyset, \{D\}, \{B, D\}$*. These are also all possible parent sets of* C *according to Figure 1d.*

We now show that local IDA cannot be applied. For every $\mathbf{S} \subseteq \mathrm{Sib}(C, \mathcal{G})$*, local IDA from Maathuis et al. (2009) orients* $S - C$ *as* $S \to C$ *for every* $S \in \mathbf{S}$*. If this does not introduce a new unshielded collider* $\to C \leftarrow$ *it returns* $\mathrm{Pa}(C, \mathcal{G}) \cup \mathbf{S}$ *as a valid parent set. Local IDA returns the following sets as the possible parent sets of* C*:* $\emptyset, \{B\}, \{D\}, \{B, D\}$*. However, parent set* $\{B\}$ *means adding* $B \to C$ *and* $C \to D$ *and this introduces a cycle due to* $D \to B$*. Hence,* $\{B\}$ *will never be a parent set of* C *in* $[\mathcal{C}]$ *and local IDA is not valid for maximal* PDAG*s.*

Example 4.9. *Consider again the maximal* PDAG $\mathcal{G}$ *in Figure 1c. Suppose we want to estimate the total causal effect of* (C, D) *on* A *in* $\mathcal{G}$*. All possible joint parent sets of* (C, D) *according to Algorithm 2 are:* $(\emptyset, \{C\}), (\{D\}, \emptyset), (\{B, D\}, \emptyset), (\{B, D\}, \{A\})$*. These are also all possible joint parent sets of* (C, D) *according to Figure 1d.*

Semi-local joint-IDA from Nandy et al. (2017) would attempt to learn all possible joint parent sets of (C, D) *by orienting the undirected component* $A - B - C - D - A$ *on* $\mathcal{G}$ *into all possible* DAG*s without unshielded colliders. However, it is not possible to orient* $A - B - C - D - A$ *into a* DAG *without creating an unshielded collider. Hence, semi-local joint-IDA is not valid for maximal* PDAG*s.*

5 IMPLEMENTATION AND SIMULATION STUDY

We investigated the effect of including background knowledge into a CPDAG in a simulation study using `R (3.3.3)` and the R-package `pcalg (2.4-6)` (Kalisch et al., 2012) where the following functions were added or modified: `isValidGraph()`, `addBgKnowledge()`, `adjustment()`, `gac()`, `ida()`, `jointIda()`. Details about the simulation can be found in the package vignette.

We first sampled 1000 settings that were used to generate graphs later on. The following settings were drawn uniformly at random: the number of nodes $p \in \{20, 30, ..., 100\}$, and the expected neighborhood size $E[N] \in \{3, 4, ..., 10\}$.

For each of these 1000 settings, 20 DAGs were randomly generated and then transformed into the corresponding CPDAG. This resulted in 20 000 CPDAGs. For each graph, we randomly chose a node X and then randomly chose a node Y that was connected to X but was not a parent of X in the true underlying DAG. Furthermore, for each DAG we generated a data set with sample size $n = 200$.

For each of the 20 000 CPDAGs, we then generated additional maximal PDAGs by replacing a fraction of randomly chosen undirected edges by the true directed edges of the underlying DAG and applying the orientation rules in Figure 2 to further orient edges if possible. The fraction of background knowledge varied through all values in the set $\{0, 0.1, \ldots, 0.9, 1\}$, resulting in eleven maximal PDAGs. Note that a fraction of 0 corresponds to the CPDAG and a fraction of 1 corresponds to the true underlying DAG.

For each of the 220 000 maximal PDAGs $\mathcal{G}$ we analyzed two questions:

Q1: Is there a set that satisfies the b-adjustment criterion relative to (X, Y) in $\mathcal{G}$?

Q2: What is the multi-set of possible total causal effects of X on Y given $\mathcal{G}$ and the sampled data of the corresponding DAG? In particular, what is the number of unique estimates in the multi-set?

The results for question (**Q1**) are shown in Figure 4. We see that without any background knowledge around 90% of all total causal effects could be identified via adjustment, while around 70% of the non-zero total causal effects could be identified via adjustment. When the proportion of background knowledge increases, the fraction of total causal effects that we identify via adjustment increases, both for all total causal effects and for the non-zero total causal effects. With 100% background knowledge, the maximal PDAG is identical to the true underlying DAG and identification of the total causal effect of X on Y via covariate adjustment is always possible, since X is a single intervention and Y is not a parent of X in the DAG (Pearl, 2009).

The results for questions (**Q2**) are shown in Figure 5, restricting ourselves to the 4315 CPDAGs that had more

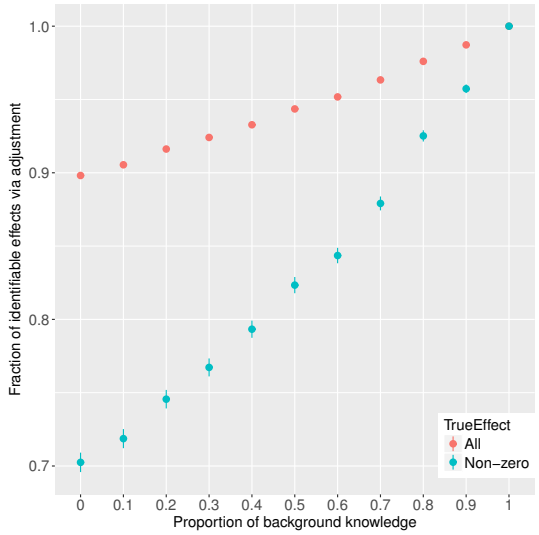

Figure 4: Fraction of total causal effects that are identifiable with our adjustment criterion, with respect to the proportion of background knowledge. Red: all simulations; blue: those simulations where the true total causal effect is non-zero. Error bars indicate standard errors and are partly too small to be seen.

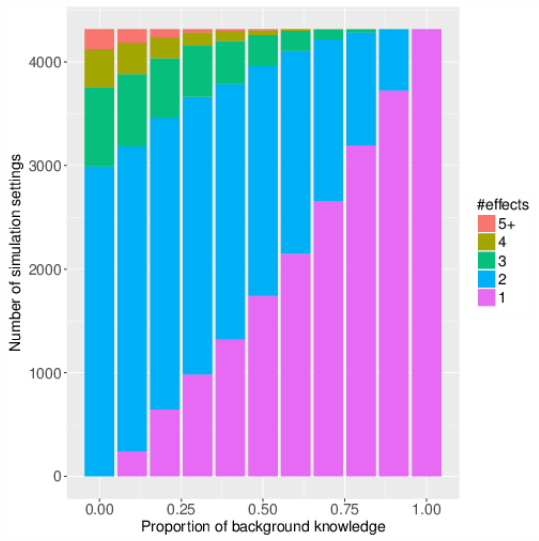

Figure 5: Stacked barplot of the number of unique possible total causal effects in the multi-sets of our semi-local IDA, with respect to the proportion of background knowledge. The results are restricted to the 4315 CPDAGs (out of 20 000) that have more than one unique estimated possible total causal effect in the multi-set.

than one unique estimate in the multi-set. When including background knowledge, the fraction of multi-sets with exactly one unique estimate gradually increases. Finally, with 100% background knowledge, the maximal PDAG is identical to the true underlying DAG and the multi-set always contains one unique element.

6 DISCUSSION

Although maximal PDAGs typically contain more orientation information than CPDAGs, this additional information has not been fully exploited in practice, due to a lack of understanding and methodology for maximal PDAGs. Our paper aims to make an important step in bridging this gap and in opening the way for the use of maximal PDAGs in practice.

This paper introduces various tools for working with maximal PDAGs. In particular, we are now able to read off possible ancestral relationships directly from a maximal PDAG and to estimate (possible) total causal effects when a maximal PDAG is given. Since CPDAGs and DAGs are special cases of maximal PDAGs, our b-adjustment criterion and semi-local (joint-)IDA methods for maximal PDAGs generalize existing results for CPDAGs and DAGs (Maathuis et al., 2009; Shpitser et al., 2010; Perković et al., 2015, 2017; Nandy et al., 2017). All methods are implemented in the `R` package `pcalg`.

The examples and the simulation study in our paper involve maximal PDAGs generated by adding background knowledge to a CPDAG. Nevertheless, we emphasize that maximal PDAGs can arise in many different ways, e.g., by adding background knowledge before structure learning (Scheines et al., 1998), by structure learning from a combination of observational and interventional data (Hauser and Bühlmann, 2012; Wang et al., 2017), or by structure learning for certain restricted model classes (Hoyer et al., 2008; Ernest et al., 2016; Eigenmann et al., 2017).

It would be interesting to extend our methods to settings with hidden variables, e.g., considering partial ancestral graphs (PAGs; Richardson and Spirtes, 2002; Ali et al., 2005) with background knowledge. An important missing link for such an extension is a clear understanding of PAGs with background knowledge. In particular, one would need to develop complete orientation rules for PAGs with background knowledge, analogous to the work of Meek (1995). Once this is in place, it seems feasible to generalize graphical criteria for covariate adjustment in PAGs (Perković et al., 2015, 2017) and an IDA type method for PAGs, called LV-IDA (Malinsky and Spirtes, 2017).

Acknowledgements

This work was supported in part by Swiss NSF Grant 200021_172603.

References

Ali, A. R., Richardson, T. S., Spirtes, P. L., and Zhang, J. (2005). Towards characterizing Markov equivalence classes for directed acyclic graphs with latent variables. In *Proceedings of UAI 2005*.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *J. Mach. Learn. Res.*, 3:507–554.

Dor, D. and Tarsi, M. (1992). A simple algorithm to construct a consistent extension of a partially oriented graph. *Technicial Report R-185, Cognitive Systems Laboratory, UCLA*.

Eigenmann, M., Nandy, P., and Maathuis, M. H. (2017). Structure learning of linear Gaussian structural equation models with weak edges. In *Proceedings of UAI 2017*.

Ernest, J., Rothenhäusler, D., and Bühlmann, P. (2016). Causal inference in partially linear structural equation models: identifiability and estimation. arXiv:1607.05980.

Hauser, A. and Bühlmann, P. (2012). Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *J. Mach. Learn. Res.*, 13:2409–2464.

Hoyer, P. O., Hyvarinen, A., Scheines, R., Spirtes, P. L., Ramsey, J., Lacerda, G., and Shimizu, S. (2008). Causal discovery of linear acyclic models with arbitrary distributions. In *Proceedings of UAI 2008*, pages 282–289.

Hyttinen, A., Eberhardt, F., and Järvisalo, M. (2015). Do-calculus when the true graph is unknown. In *Proceedings of UAI 2015*, pages 395–404.

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

Maathuis, M. H., Colombo, D., Kalisch, M., and Bühlmann, P. (2010). Predicting causal effects in large-scale systems from observational data. *Nat. Methods*, 7:247–248.

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

Malinsky, D. and Spirtes, P. (2017). Estimating bounds on causal effects in high-dimensional and possibly confounded systems. *Int. J. of Approx. Reason.*

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In *Proceedings of UAI 1995*, pages 403–410.

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

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, second edition.

Perković, E., Textor, J., Kalisch, M., and Maathuis, M. H. (2015). A complete generalized adjustment criterion. In *Proceedings of UAI 2015*, pages 682–691.

Perković, E., Textor, J., Kalisch, M., and Maathuis, M. H. (2017). Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. *J. Mach. Learn. Res.* to appear.

Richardson, T. S. and Spirtes, P. (2002). Ancestral graph Markov models. *Ann. Stat.*, 30:962–1030.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Math. Mod.*, 7:1393–1512.

Scheines, R., Spirtes, P., Glymour, C., Meek, C., and Richardson, T. (1998). The TETRAD project: constraint based aids to causal model specification. *Multivar. Behav. Res.*, 33(1):65–117.

Shpitser, I. (2012). Appendum to "On the validity of covariate adjustment for estimating causal effects". Personal communication.

Shpitser, I., VanderWeele, T., and Robins, J. M. (2010). On the validity of covariate adjustment for estimating causal effects. In *Proceedings of UAI 2010*, pages 527–536.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, second edition.

van der Zander, B. and Liśkiewicz, M. (2016). Separators and adjustment sets in Markov equivalent DAGs. In *Proceedings of AAAI 2016*, pages 3315–3321.

van der Zander, B., Liśkiewicz, M., and Textor, J. (2014). Constructing separators and adjustment sets in ancestral graphs. In *Proceedings of UAI 2014*, pages 907–916.

Wang, Y., Solus, L., Yang, K. D., and Uhler, C. (2017). Permutation-based causal inference algorithms with interventions. arXiv:1705.10220.

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artif. Intell.*, 172:1873–1896.