Constructing Separators and Adjustment Sets in Ancestral Graphs

Benito van der Zander, Maciej Liśkiewicz
Theoretical Computer Science
University of Lübeck, Germany
{benito,liskiewi}@tcs.uni-luebeck.de

Johannes Textor
Theoretical Biology & Bioinformatics
Utrecht University, The Netherlands
johannes.textor@gmx.de

Abstract

Ancestral graphs (AGs) are graphical causal models that can represent uncertainty about the presence of latent confounders, and can be inferred from data. Here, we present an algorithmic framework for efficiently testing, constructing, and enumerating m-separators in AGs. Moreover, we present a new constructive criterion for covariate adjustment in directed acyclic graphs (DAGs) and maximal ancestral graphs (MAGs) that characterizes adjustment sets as m-separators in a subgraph. Jointly, these results allow to find all adjustment sets that can identify a desired causal effect with multivariate exposures and outcomes in the presence of latent confounding. Our results generalize and improve upon several existing solutions for special cases of these problems.

1 INTRODUCTION

Graphical causal models endow researchers with a language to codify assumptions about a data generating process (Pearl, 2009; Elwert, 2013). Using graphical criteria, one can asses whether the assumptions encoded in such a model allow estimation of a causal effect from observational data, which is a key issue in Epidemiology (Rothman et al., 2008), the Social Sciences (Elwert, 2013) and other fields where controlled experimentation is typically impossible. Specifically, the famous back-door criterion by Pearl (2009) can identify cases where causal effect identification is possible by standard covariate adjustment, and other methods like the front-door criterion or do-calculus can even permit identification even if the back-door criterion fails (Pearl, 2009). In current practice, however, covariate adjustment is highly preferred to such alternatives because its statistical properties are well understood, giving access to useful methodology like robust estimators and confidence intervals. In contrast, knowledge about the statistical properties of e.g. front-door estimation is still considerably lacking (VanderWeele, 2009; Glynn and Kashin, 2013)[1]. Unfortunately, the back-door criterion is not complete, i.e., it does not find all possible options for covariate adjustment that are allowed by a given graphical causal model.

In this paper, we aim to efficiently find a definitive answer for the following question: Given a causal graph $\mathcal{G}$, which covariates $\mathbf{Z}$ do we need to adjust for to estimate the causal effect of the exposures $\mathbf{X}$ on the outcomes $\mathbf{Y}$? To our knowledge, no efficient algorithm has been shown to answer this question, not even when $\mathcal{G}$ is a directed acyclic graph (DAG), though constructive solutions do exist for special cases like singleton $\mathbf{X} = \{X\}$ (Pearl, 2009), and a subclass of DAGs (Textor and Liśkiewicz, 2011). Here, we provide algorithms for adjustment sets in DAGs as well as in maximal ancestral graphs (MAGs), which extend DAGs allowing to account for unspecified latent variables. Our algorithms are guaranteed to find all valid adjustment sets for a given DAG or MAG with polynomial delay, and we also provide variants to list only those sets that minimize a user-supplied cost function or to quickly construct a simple adjustment set if one exists. Modelling multiple, possibly interrelated exposures $\mathbf{X}$ is important e.g. in case-control studies that screen several putative causes of a disease (Greenland, 1994). Likewise, the presence of unspecified latent variables often cannot be excluded in real-world settings, and the causal structure between the observed variables may not be completely known. We hope that the ability to quickly deduce from a given DAG or MAG whether and how covariate adjustment can render a causal effect identifiable will benefit researchers in such areas.

We have two main contributions. First, in Section 3, we present algorithms for verifying, constructing, and listing m-separating sets in AGs. This subsumes a number of earlier solutions for special cases of these problems, e.g.

[1]Quoting VanderWeele (2009), "Time will perhaps tell whether results like Pearl's front-door path adjustment theorem and its generalizations are actually useful for epidemiologic research or whether the results are simply of theoretical interest."

the Bayes-Ball algorithm for verification of d-separating sets (Shachter, 1998), the use of network flow calculations to find minimal d-separating sets in DAGs (Tian et al., 1998; Acid and de Campos, 2003), and an algorithm to list minimal adjustment sets for a certain subclass of DAGs (Textor and Liśkiewicz, 2011). Our verification and construction algorithms for single separators are asymptotically runtime-optimal. Although we apply our algorithms only to adjustment set construction, they are likely useful in other settings as separating sets are involved in most graphical criteria for causal effect identification. Moreover, the separators themselves constitute statistically testable implications of the causal assumptions encoded in the graph.

Second, we give a graphical criterion that characterizes adjustment sets in terms of separating sets, and is sound and complete for DAGs and MAGs without selection variables. This generalizes the sound and complete criterion for DAGs by Shpitser et al. (2010), and the sound but incomplete adjustment criterion for MAGs without selection variables by Maathuis and Colombo (2013). Our criterion exhaustively addresses adjustment set construction in the presence of latent covariates and with incomplete knowledge of causal structure if at least a MAG can be specified. We give the criterion separately for DAGs (Section 4) and MAGs (Section 5) because the same graph usually admits more adjustment options if viewed as a DAG than if viewed as a MAG.

2 PRELIMINARIES

We denote sets by bold upper case letters ($\mathbf{S}$), and sometimes abbreviate singleton sets as $\{S\} = S$. Graphs are written calligraphically ($\mathcal{G}$), and variables in upper-case (X).

Mixed graphs and paths. We consider mixed graphs $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ with nodes (vertices, variables) $\mathbf{V}$ and directed ($A \to B$), undirected ($A - B$), and bidirected ($A \leftrightarrow B$) edges $\mathbf{E}$. Nodes linked by an edge are *adjacent*. A *walk* of length n is a node sequence $V_1, \dots, V_{n+1}$ such that there exists an edge sequence $E_1, E_2, \dots, E_n$ for which every edge E_i connects V_i, V_{i+1}. Then V_1 is called the *start node* and V_{n+1} the *end node* of the walk. A *path* is a walk in which no node occurs more than once. Given a node set $\mathbf{X}$ and a node set $\mathbf{Y}$, a walk from $X \in \mathbf{X}$ to $Y \in \mathbf{Y}$ is called *proper* if only its start node is in $\mathbf{X}$. Given a graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and a node set $\mathbf{V}'$, the *induced subgraph* $\mathcal{G}_{\mathbf{V}'} = (\mathbf{V}', \mathbf{E}')$ contains the edges $\mathbf{E}'$ from $\mathcal{G}$ that are adjacent only to nodes in $\mathbf{V}'$.

Ancestry. A walk of the form $V_1 \to \dots \to V_n$ is *directed*, or *causal*. If there is a directed walk from U to V, then U is called an *ancestor* of V and V a *descendant* of U. A graph is *acyclic* if no directed walk from a node to itself is longer than 0. All directed walks in an acyclic graph are paths. A walk is *anterior* if it were directed after replacing all edges $U - V$ by $U \to V$. If there is an anterior path from U to V, then U is called an *anterior* of V. All ancestors of V are anteriors of V. Every node is its own ancestor, descendant, and anterior. For a node set $\mathbf{X}$, the set of all of its ancestors is written as $An(\mathbf{X})$. The descendant and anterior sets $De(\mathbf{X}), Ant(\mathbf{X})$ are analogously defined. Also, we denote by $Pa(\mathbf{X})$, ($Ch(\mathbf{X})$), the set of parents (children) of $\mathbf{X}$.

***m*-Separation.** A node V on a walk w is called a *collider* if two arrowheads of w meet at V, e.g. if w contains $U \leftrightarrow V \leftarrow Q$. There can be no collider if w is shorter than 2. Two nodes U, V are called *collider connected* if there is a path between them on which all nodes except U and V are colliders. Adjacent vertices are collider connected. Two nodes U, V are called *m-connected* by a set $\mathbf{Z}$ if there is a path π between them on which every node that is a collider is in $An(\mathbf{Z})$ and every node that is not a collider is not in $\mathbf{Z}$. Then π is called an m-connecting path. The same definition can be stated simpler using walks: U, V are called m-connected by $\mathbf{Z}$ if there is a walk between them on which all colliders and only colliders are in $\mathbf{Z}$. If U, V are m-connected by the empty set, we simply say they are m-connected. If U, V are not m-connected by $\mathbf{Z}$, we say that $\mathbf{Z}$ *m-separates* them or *blocks* all paths between them. Two node sets $\mathbf{X}, \mathbf{Y}$ are m-separated by $\mathbf{Z}$ if all their nodes are pairwise m-separated by $\mathbf{Z}$. In DAGs, m-separation is equivalent to the well-known d-separation criterion (Pearl, 2009).

Ancestral graphs and DAGs. A mixed graph $\mathcal{G} = (V, E)$ is called an *ancestral graph* (AG) if the following two conditions hold: (1) For each edge $A \leftarrow B$ or $A \leftrightarrow B$, A is not an ancestor of B. (2) For each edge $A - B$, there are no edges $A \leftarrow C$, $A \leftrightarrow C$, $B \leftarrow C$ or $B \leftrightarrow C$. There can be at most one edge between two nodes in an AG (Richardson and Spirtes, 2002). Syntactically, all DAGs are AGs and all AGs containing only directed edges are DAGs. An AG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ is a *maximal ancestral graph* (MAG) if every non-adjacent pair of nodes U, V can be m-separated by some $\mathbf{Z} \subseteq \mathbf{V} \setminus \{U, V\}$. Every AG $\mathcal{G}$ can be turned into a MAG $\mathcal{M}$ by adding bidirected edges between node pairs that cannot be m-separated (Richardson and Spirtes, 2002).

3 ALGORITHMS FOR m-SEPARATION

In this section, we compile an algorithmic framework for solving a host of problems related to verification, construction, and enumeration of m-separating sets in AGs. The problems are defined in Fig. 1, which also shows the asymptotic runtime of their solutions. Throughout, n stands for the number of nodes and m for the number of edges in a graph. All of these problems except LISTSEP can be solved by rather straightforward modifications of existing algorithms (Acid and Campos, 1996; Shachter, 1998; Tian et al., 1998; Textor and Liśkiewicz, 2011). We therefore refrain in this paper from presenting them in detail, Pseudocodes of these algorithms are shown for reference and implementation in the online version of this paper[2]. The online version also contains proof details that had to be omitted from this paper for space reasons.

An important tool for solving similar problems for d-separation is *moralization*, by which d-separation can be reduced to a vertex cut in an undirected graph. This reduction allows to solve problems like FINDMINSEP using standard network flow algorithms (Acid and Campos, 1996). Moralization can be generalized to AGs in the following manner.

Definition 3.1 (Moralization of AGs (Richardson and Spirtes, 2002))**.** *Given an AG $\mathcal{G}$, the augmented graph $(\mathcal{G})^a$ is an undirected graph with the same node set as $\mathcal{G}$ such that $X - Y$ is an edge in $(\mathcal{G})^a$ if and only if X and Y are collider connected in $\mathcal{G}$.*

Theorem 3.2 (Reduction of m-Separation to vertex cuts (Richardson and Spirtes, 2002))**.** *Given an AG $\mathcal{G}$ and three node sets $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$, $\mathbf{Z}$ m-separates $\mathbf{X}$ and $\mathbf{Y}$ if and only if $\mathbf{Z}$ is an $\mathbf{X}$-$\mathbf{Y}$ node cut in $(\mathcal{G}_{Ant(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z})})^a$.*

A direct implementation of Definition 3.1 would lead to a suboptimal algorithm. Therefore, we first give an asymptotically optimal (linear time in output size) moralization algorithm for AGs. We then solve TESTMINSEP, FINDMINSEP, FINDMINCOSTSEP and LISTMINSEP by generalizing existing correctness proofs of the moralization approach for d-separation (Tian et al., 1998).

Not all our solutions are based on moralization, however. Moralization takes time $O(n^2)$, and TESTSEP and FINDSEP can be solved faster, i.e. in asymptotically optimal time $O(n + m)$.

Lemma 3.3 (Efficient AG moralization)**.** *Given an AG $\mathcal{G}$, the augmented graph $(\mathcal{G})^a$ can be computed in time $O(n^2)$.*

Proof. The algorithm proceeds in four steps. (1) Start by setting $(\mathcal{G})^a$ to $\mathcal{G}$ replacing all edges by undirected ones. (2) Identify all connected components in $\mathcal{G}$ with respect to bidirected edges (two nodes are in the same such component if they are connected by a path consisting only of bidirected edges). Nodes without adjacent bidirected edges form singleton components. (3) For each pair U, V of nodes from the same component, add the edge $U - V$ to $(\mathcal{G})^a$ if it did not exist already. (4) For each component, identify all its parents (nodes U with an edge $U \to V$ where U is in the component) and link them all by undirected edges in $(\mathcal{G})^a$. Now two nodes are adjacent in $(\mathcal{G})^a$ if and only if they are collider connected in $\mathcal{G}$. All four steps can be performed in time $O(n^2)$. □

Lemma 3.4. *Let $\mathbf{X}, \mathbf{Y}, \mathbf{I}, \mathbf{R}$ be sets of nodes with $\mathbf{I} \subseteq \mathbf{R}$, $\mathbf{R} \cap (\mathbf{X} \cup \mathbf{Y}) = \emptyset$. If there exists an m-separator $\mathbf{Z}_0$, with $\mathbf{I} \subseteq \mathbf{Z}_0 \subseteq \mathbf{R}$ then $\mathbf{Z} = Ant(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{I}) \cap \mathbf{R}$ is an m-separator.*

[2]URL: `theory.bio.uu.nl/textor/uai14.pdf`

Corollary 3.5 (Ancestry of minimal separators)**.** *Given an AG $\mathcal{G}$, and three sets $\mathbf{X}, \mathbf{Y}, \mathbf{I}$, every minimal set $\mathbf{Z}$ over all m-separators containing $\mathbf{I}$ is a subset of $Ant(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{I})$.*

Proof. Assume there is a minimal separator $\mathbf{Z}$ with $\mathbf{Z} \not\subseteq Ant(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{I})$. According to Lemma 3.4 we have that $\mathbf{Z}' = Ant(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{I}) \cap \mathbf{Z}$ is a separator with $\mathbf{I} \subseteq \mathbf{Z}'$. But $\mathbf{Z}' \subseteq Ant(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{I})$ and $\mathbf{Z}' \subseteq \mathbf{Z}$, so $\mathbf{Z} \neq \mathbf{Z}'$ and $\mathbf{Z}$ is not a minimal separator. □

Corollary 3.5 applies to minimum-cost separators as well because every minimum-cost separator must be minimal. Now we can solve FINDMINCOSTSEP and FINDMINSIZESEP by using weighted min-cut, which takes time $O(n^3)$ using practical algorithms, and LISTMINSEP by using Takata's algorithm to enumerate minimal vertex cuts with delay $O(n^3)$ (Takata, 2010).

However, for FINDMINSEP and TESTMINSEP, we can do better than using standard vertex cuts.

Proposition 3.6. *The task* FINDMINSEP *can be solved in time $O(n^2)$.*

Proof. Two algorithms are given in the online appendix, one with runtime $O(n^2)$ and one with runtime $O(nm)$. □

Corollary 3.7. *The task* TESTMINSEP *can be solved in time $O(n^2)$.*

Proof. First verify whether $\mathbf{Z}$ is an m-separator using moralization. If not, return "no". Otherwise, set $S = \mathbf{Z}$ and solve FINDMINSEP. Return "yes" if the output is $\mathbf{Z}$ and "no", otherwise. □

Moralization can in the worst case quadratically increase the size of a graph. Therefore, in some cases, it may be preferable to avoid moralization if the task at hand is rather simple, as are the two tasks considered below.

Proposition 3.8. *The task* FINDSEP *can be solved in time $O(n + m)$.*

Proof. This follows directly from Lemma 3.4, and the fact that the set $Ant(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{I}) \cap \mathbf{R}$ can be found in linear time from the MAG without moralization. Note that unlike in DAGs, two non-adjacent nodes cannot always be m-separated in ancestral graphs. □

By modifying the Bayes-Ball algorithm (Shachter, 1998) appropriately, we get the following.

Proposition 3.9. *The task* TESTSEP *can be solved in time $O(n + m)$.*

Lastly, we consider the problem of listing *all* m-separators. Here is an algorithm to solve that problem with polynomial delay.

Verification: For given $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ decide if ...		
TESTSEP	$\mathbf{Z}$ m-separates $\mathbf{X}, \mathbf{Y}$	$O(n+m)$
TESTMINSEP	$\mathbf{Z}$ m-separates $\mathbf{X}, \mathbf{Y}$ but no $\mathbf{Z}' \subsetneq \mathbf{Z}$ does	$O(n^2)$
Construction: For given $\mathbf{X}, \mathbf{Y}$ and auxiliary $\mathbf{I}, \mathbf{R}$, output ...		
FINDSEP	an m-separator $\mathbf{Z}$ with $\mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{R}$	$O(n+m)$
FINDMINSEP	a minimal m-separator $\mathbf{Z}$ with $\mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{R}$	$O(n^2)$
FINDMINCOSTSEP	a minimum-cost m-separator $\mathbf{Z}$ with $\mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{R}$	$O(n^3)$
Enumeration: For given $\mathbf{X}, \mathbf{Y}, \mathbf{I}, \mathbf{R}$ enumerate all ...		
LISTSEP	m-separators $\mathbf{Z}$ with $\mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{R}$	$O(n(n+m))$ delay
LISTMINSEP	minimal m-separators $\mathbf{Z}$ with $\mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{R}$	$O(n^3)$ delay

Table 1: Definitions of algorithmic tasks related to m-separation. Throughout, $\mathbf{X}, \mathbf{Y}, \mathbf{R}$ are pairwise disjoint node sets, $\mathbf{Z}$ is disjoint with $\mathbf{X}, \mathbf{Y}$ which are nonempty, and $\mathbf{I}, \mathbf{R}, \mathbf{Z}$ can be empty. By a minimal m-separator $\mathbf{Z}$, with $\mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{R}$, we mean a set such that no proper subset $\mathbf{Z}'$ of $\mathbf{Z}$, with $\mathbf{I} \subseteq \mathbf{Z}'$, m-separates the pair $\mathbf{X}$ and $\mathbf{Y}$. Analogously, we define a minimal and a minimum-cost m-separator. The construction algorithms will output $\perp$ if no set fulfilling the listed condition exists. Delay complexity for e.g. LISTMINSEP refers to the time needed to output one solution when there can be exponentially many solutions (see Takata (2010)).

```
function ListSep(G, X, Y, I, R)
    if FindSep(G, X, Y, I, R) ≠ ⊥ then
        if I = R then Output I
        else
            V ← an arbitrary node of R \ I
            ListSep(G, X, Y, I ∪ {V}, R)
            ListSep(G, X, Y, I, R \ {V})
```

Figure 1: ListSep

Proposition 3.10. *The task* LISTSEP *can be solved with polynomial delay* $O(n(n+m))$.

Proof. Algorithm LISTSEP performs backtracking to enumerate all $\mathbf{Z}$ with $\mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{R}$ aborting branches that will not find a valid separator. Since every leaf will output a separator, the tree height is at most n and the existence check needs $O(n+m)$, the delay time is $O(n(n+m))$. The algorithm generates every separator exactly once: if initially $\mathbf{I} \subsetneq \mathbf{R}$, with $V \in \mathbf{R} \setminus \mathbf{I}$, then the first recursive call returns all separators $\mathbf{Z}$ with $V \in \mathbf{Z}$ and the second call returns all $\mathbf{Z}'$ with $V \notin \mathbf{Z}'$. Thus the generated separators are pairwise disjoint. This is a modification of the enumeration algorithm for minimal vertex separators (Takata, 2010). □

4 ADJUSTMENT IN DAGS

In this section, we leverage the algorithmic framework of the last section together with a new constructive, sound and complete criterion for covariate adjustment in DAGs to solve all problems listed in Table 1 for adjustment sets instead of m-separators in the same asymptotic time. First, however, we need to introduce some more notation pertaining to the causal interpretation DAGs.

Do-operator and adjustment sets. A DAG $\mathcal{G}$ encodes the factorization of joint distribution π for the set of variables $\mathbf{V} = \{X_1, \dots, X_n\}$ as $p(\mathbf{v}) = \prod_{j=1}^n p(x_j|pa_j)$, where pa_j denotes a particular realization of the parent variables of X_j in $\mathcal{G}$. When interpreted causally, an edge $X_i \to X_j$ is taken to represent a direct causal effect of X_i on X_j. For disjoint $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$, the *(total) causal effect* of $\mathbf{X}$ on $\mathbf{Y}$ is $p(\mathbf{y}|do(\mathbf{x}))$ where $do(\mathbf{x})$ represents an intervention that sets $\mathbf{X} = \mathbf{x}$. In a DAG, this intervention corresponds to removing all edges into $\mathbf{X}$, disconnecting $\mathbf{X}$ from its parents. We denote the resulting graph as $\mathcal{G}_{\overline{\mathbf{X}}}$. Given DAG $\mathcal{G}$ and a joint probability density π for $\mathbf{V}$ the post-intervention distribution can be expressed in a truncated factorization formula:

$$p(\mathbf{v}|do(\mathbf{x})) = \begin{cases} \prod\limits_{X_j \in \mathbf{V} \setminus \mathbf{X}} p(x_j|pa_j) & \text{for } \mathbf{V} \text{ consistent with } \mathbf{x} \\ 0 & \text{otherwise.} \end{cases}$$

Definition 4.1 (Adjustment (Pearl, 2009))**.** *Given a DAG* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ *and pairwise disjoint* $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, $\mathbf{Z}$ *is called* covariate adjustment for estimating the causal effect of $\mathbf{X}$ on $\mathbf{Y}$, *or simply* adjustment, *if for every distribution p consistent with $\mathcal{G}$ we have* $p(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} p(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) p(\mathbf{z})$.

Definition 4.2 (Adjustment criterion (Shpitser et al., 2010; Shpitser, 2012)[3])**.** *Let* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ *be a DAG, and* $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$ *be pairwise disjoint subsets of variables. The set* $\mathbf{Z}$ *satisfies the adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *if*

(*a*) *no element in* $\mathbf{Z}$ *is a descendant in* $\mathcal{G}$ *of any* $W \in \mathbf{V} \setminus \mathbf{X}$ *which lies on a proper causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *and*

(*b*) *all proper non-causal paths in* $\mathcal{G}$ *from* $\mathbf{X}$ *to* $\mathbf{Y}$ *are blocked by* $\mathbf{Z}$.

Analogously to $\mathcal{G}_{\overline{\mathbf{X}}}$, by $\mathcal{G}_{\underline{\mathbf{X}}}$ we denote a DAG obtained from $\mathcal{G}$ by removing all edges leaving $\mathbf{X}$.

[3] In Shpitser et al. (2010), the criterion is stated using $\mathcal{G}_{\overline{\mathbf{X}}}$ instead of $\mathcal{G}$. However, one can easily prove that both criteria are equivalent.

4.1 CONSTRUCTIVE BACK-DOOR CRITERION

Definition 4.3 (Proper back-door graph). *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG, and $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$ be pairwise disjoint subsets of variables. The* proper back-door graph*, denoted as $\mathcal{G}^{pbd}_{\mathbf{XY}}$, is obtained from $\mathcal{G}$ by removing the first edge of every proper causal path form $\mathbf{X}$ to $\mathbf{Y}$.*

Note the difference between the back-door graph $\mathcal{G}_{\underline{\mathbf{X}}}$ and the proper back-door graph $\mathcal{G}^{pbd}_{\mathbf{XY}}$: in $\mathcal{G}_{\underline{\mathbf{X}}}$ *all* edges leaving $\mathbf{X}$ are removed while in $\mathcal{G}^{pbd}_{\mathbf{XY}}$ only those that lie on a proper causal path. However, to construct $\mathcal{G}^{pbd}_{\mathbf{XY}}$ still only elementary operations are sufficient. Indeed, we remove all edges $X \to D$ in $\mathbf{E}$ such that $X \in \mathbf{X}$ and D is in the subset, which we call $PCP(\mathbf{X}, \mathbf{Y})$, obtained as follows:

$$PCP(\mathbf{X}, \mathbf{Y}) = (De_{\overline{\mathbf{X}}}(\mathbf{X}) \setminus \mathbf{X}) \cap An_{\underline{\mathbf{X}}}(\mathbf{Y}) \tag{1}$$

where $De_{\overline{\mathbf{X}}}(\mathbf{W})$ denotes descendants of $\mathbf{W}$ in $\mathcal{G}_{\overline{\mathbf{X}}}$. $An_{\underline{\mathbf{X}}}(\mathbf{W})$ is defined analogously for $\mathcal{G}_{\underline{\mathbf{X}}}$. Hence, the proper back-door graph can be constructed from $\mathcal{G}$ in linear time $O(m + n)$.

Now we propose the following adjustment criterion. For short, we will denote the set $De(PCP(\mathbf{X}, \mathbf{Y}))$ as $Dpcp(\mathbf{X}, \mathbf{Y})$.

Definition 4.4 (Constructive back-door criterion (CBC)). *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG, and let $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$ be pairwise disjoint subsets of variables. The set $\mathbf{Z}$ satisfies the* constructive back-door criterion *relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{G}$ if*

(*a*) $\mathbf{Z} \subseteq \mathbf{V} \setminus Dpcp(\mathbf{X}, \mathbf{Y})$ *and*

(*b*) $\mathbf{Z}$ *d-separates $\mathbf{X}$ and $\mathbf{Y}$ in the proper back-door graph $\mathcal{G}^{pbd}_{\mathbf{XY}}$.*

Theorem 4.5. *The constructive back-door criterion is equivalent to the adjustment criterion.*

Proof. First observe that the conditions (*a*) of both criteria are identical. Assume conditions (*a*) and (*b*) of the adjustment criterion hold. We show that (*b*) of the constructive back-door criterion follows. Let π be any proper path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}^{pbd}_{\mathbf{XY}}$. Because $\mathcal{G}^{pbd}_{\mathbf{XY}}$ does not contain causal paths from $\mathbf{X}$ to $\mathbf{Y}$, π is not causal and has to be blocked by $\mathbf{Z}$ in $\mathcal{G}$ by the assumption. Since removing edges cannot open paths, π is blocked by $\mathbf{Z}$ in $\mathcal{G}^{pbd}_{\mathbf{XY}}$ as well.

Now we show that (*a*) and (*b*) of the constructive back-door criterion together imply (*b*) of the adjustment criterion. If that were not the case, then there could exist a proper non-causal path π from $\mathbf{X}$ to $\mathbf{Y}$ that is blocked in $\mathcal{G}^{pbd}_{\mathbf{XY}}$ but open in $\mathcal{G}$. There can be two reasons why π is blocked in $\mathcal{G}^{pbd}_{\mathbf{XY}}$: (1) The path starts with an edge $X \to D$ that does not exist in $\mathcal{G}^{pbd}_{\mathbf{XY}}$. Then we have $D \in PCP(\mathbf{X}, \mathbf{Y})$. For π to be non-causal, it would have to contain a collider $C \in An(\mathbf{Z}) \cap De(D) \subseteq An(\mathbf{Z}) \cap Dpcp(\mathbf{X}, \mathbf{Y})$. But because of (*a*), $An(\mathbf{Z}) \cap Dpcp(\mathbf{X}, \mathbf{Y})$ is empty. (2) A collider C on π is an ancestor of $\mathbf{Z}$ in $\mathcal{G}$, but not in $\mathcal{G}^{pbd}_{\mathbf{XY}}$. Then there must be a directed path from C to $\mathbf{Z}$ via an edge $X \to D$ with $D \in An(\mathbf{Z}) \cap PCP(\mathbf{X}, \mathbf{Y})$, contradicting (a). □

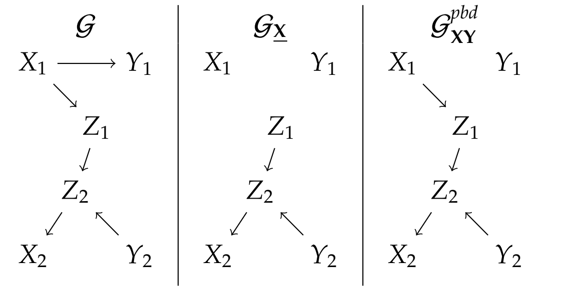

Figure 2: A DAG where for $\mathbf{X} = \{X_1, X_2\}$ and $\mathbf{Y} = \{Y_1, Y_2\}$, $\mathbf{Z} = \{Z_1, Z_2\}$ is a valid and minimal adjustment, but no set fulfills the back-door criterion (Pearl, 2009), and the parents of $\mathbf{X}$ are not a valid adjustment set either.

4.2 ADJUSTING FOR MULTIPLE EXPOSURES

For a singleton set $\mathbf{X} = \{X\}$ of exposures we know that if a set of variables $\mathbf{Y}$ is disjoint from $\{X\} \cup Pa(X)$ then one obtains easily an adjustment set with respect to X and $\mathbf{Y}$ as $\mathbf{Z} = Pa(X)$ (Pearl, 2009, Theorem 3.2.2). The situation changes drastically if the effect of multiple exposures is estimated. Theorem 3.2.5 in Pearl (2009) claims that the expression for $P(\mathbf{y} \mid do(\mathbf{x}))$ is obtained by adjusting for $Pa(\mathbf{X})$ if $\mathbf{Y}$ is disjoint from $\mathbf{X} \cup Pa(\mathbf{X})$, but, as the DAG in Fig. 2 shows, this is not true: the set $\mathbf{Z} = Pa(X_1, X_2) = \{Z_2\}$ is not an adjustment set according to $\{X_1, X_2\}$ and Y. In this case one can identify the causal effect by adjusting for $\mathbf{Z} = \{Z_1, Z_2\}$ only. Indeed, for more than one exposure, no adjustment set may exist at all even without latent covariates and even though $\mathbf{Y} \cap (\mathbf{X} \cup Pa(\mathbf{X})) = \emptyset$, e.g. in the DAG

$X_1 \quad X_2 \leftrightarrows Z \longleftarrow Y$.

Using our criterion, we can construct a simple adjustment set explicitly if one exists. For a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ we define the set

$$Adj(\mathbf{X}, \mathbf{Y}) = An(\mathbf{X}, \mathbf{Y}) \setminus (\mathbf{X} \cup \mathbf{Y} \cup Dpcp(\mathbf{X}, \mathbf{Y})).$$

Theorem 4.6. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG and let $\mathbf{X}, \mathbf{Y} \subseteq V$ be distinct node sets. Then the following statements are equivalent:*

1. *There exists an adjustment in $\mathcal{G}$ w.r.t. $\mathbf{X}$ and $\mathbf{Y}$.*
2. *$Adj(\mathbf{X}, \mathbf{Y})$ is an adjustment w.r.t. $\mathbf{X}$ and $\mathbf{Y}$.*
3. *$Adj(\mathbf{X}, \mathbf{Y})$ d-separates $\mathbf{X}$ and $\mathbf{Y}$ in the proper back-door graph $\mathcal{G}^{pbd}_{\mathbf{XY}}$.*

Proof. The implication (3) ⇒ (2) follows directly from the criterion Def. 4.4 and the definition of $Adj(\mathbf{X}, \mathbf{Y})$. Since the implication (2) $\Rightarrow$ (1) is obvious, it remains to prove (1) $\Rightarrow$ (3).

Assume there exists an adjustment set $\mathbf{Z}_0$ w.r.t. $\mathbf{X}$ and $\mathbf{Y}$. From Theorem 4.5 we know that $\mathbf{Z}_0 \cap Dpcp(\mathbf{X}, \mathbf{Y}) = \emptyset$ and that $\mathbf{Z}_0$ d-separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}^{pbd}_{\mathbf{XY}}$. Our task is to show that $Adj(\mathbf{X}, \mathbf{Y})$ d-separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}^{pbd}_{\mathbf{XY}}$. This follows from Lemma 3.4 used for the proper back-door graph $\mathcal{G}^{pbd}_{\mathbf{XY}}$ if we take $\mathbf{I} = \emptyset$, $\mathbf{R} = \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y} \cup Dpcp(\mathbf{X}, \mathbf{Y}))$. $\square$

From Equation 1 and the definition $Dpcp(\mathbf{X}, \mathbf{Y}) = De(PCP(\mathbf{X}, \mathbf{Y}))$ we then obtain immediately:

Corollary 4.7. *Given two distinct sets* $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$, $Adj(\mathbf{X}, \mathbf{Y})$ *can be found in* $\mathcal{O}(n + m)$ *time.*

4.3 TESTING, COMPUTING, AND ENUMERATING ADJUSTMENT SETS

Using our criterion, every algorithm for m-separating sets $\mathbf{Z}$ between $\mathbf{X}$ and $\mathbf{Y}$ can be used for adjustment sets with respect to $\mathbf{X}$ and $\mathbf{Y}$, by requiring that $\mathbf{Z}$ not contain any node in $Dpcp(\mathbf{X}, \mathbf{Y})$. This allows solving all problems listed in Table 1 for adjustment sets in DAGs instead of m-separators. Below, we name those problems analogously as for m-separation, e.g. the problem to decide whether $\mathbf{Z}$ is an adjustment set w.r.t. $\mathbf{X}, \mathbf{Y}$ is named TESTADJ in analogy to TESTSEP.

TESTADJ can be solved by testing if $\mathbf{Z} \cap Dpcp(\mathbf{X}, \mathbf{Y}) = \emptyset$ and $\mathbf{Z}$ is a d-separator in the proper back-door graph $\mathcal{G}^{pbd}_{\mathbf{XY}}$. Since $\mathcal{G}^{pbd}_{\mathbf{XY}}$ can be constructed from $\mathcal{G}$ in linear time, the total time complexity of this algorithm is $\mathcal{O}(n + m)$.

TESTMINADJ can be solved with an algorithm that iteratively removes nodes from $\mathbf{Z}$ and tests if the resulting set remains an adjustment set w.r.t. $\mathbf{X}$ and $\mathbf{Y}$. This can be done in time $\mathcal{O}(n(n + m))$. Alternatively, one can construct the proper back-door graph $\mathcal{G}^{pbd}_{\mathbf{XY}}$ from $\mathcal{G}$ and test if $\mathbf{Z}$ is a minimal d-separator, with $\mathbf{Z} \subseteq \mathbf{V} \setminus Dpcp(\mathbf{X}, \mathbf{Y})$ between $\mathbf{X}$ and $\mathbf{Y}$. This can be computed in time $\mathcal{O}(n^2)$. The correctness of these algorithms follows from the proposition below, which is a generalization of the result in Tian et al. (1998).

Proposition 4.8. *If no single node Z can be removed from an adjustment set* $\mathbf{Z}$ *such that the resulting set* $\mathbf{Z}' = \mathbf{Z} \setminus Z$ *is no longer an adjustment set, then* $\mathbf{Z}$ *is minimal.*

The remaining problems like FINDADJ, FINDMINADJ etc. can be solved using corresponding algorithms for finding, resp. listing m-separations applied for proper back-door graphs. Since the proper back-door graph can be constructed in linear time the time complexities to solve the problems above are as listed in Table 1.

5 ADJUSTMENT IN MAGS

We now generalize the results from the previous section to MAGs. Two examples may illustrate why this generalization is not trivial. First, take $\mathcal{G} = X \to Y$. If $\mathcal{G}$ is interpreted as a DAG, then the empty set is valid for adjustment. If $\mathcal{G}$ is however taken as a MAG, then there exists no adjustment set as $\mathcal{G}$ represents among others the DAG $U \to X \to Y$ where U is an unobserved confounder. Second, take $\mathcal{G} = A \to X \to Y$. In that case, the empty set is an adjustment set regardless of whether $\mathcal{G}$ is interpreted as a DAG or a MAG. The reasons will become clear as we move on. First, let us recall the semantics of a MAG. The following definition can easily be given for AGs in general, but we do not need this generality for our purpose.

Definition 5.1 (DAG representation by MAGs (Richardson and Spirtes, 2002))**.** *Let* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ *be a DAG, and let* $\mathbf{S}, \mathbf{L} \subseteq \mathbf{V}$. *The MAG* $\mathcal{M} = \mathcal{G}[^{\mathbf{L}}_{\mathbf{S}}$ *is a graph with nodes* $\mathbf{V} \setminus (\mathbf{S} \cup \mathbf{L})$ *and defined as follows. (1) Two nodes U and V are adjacent in* $\mathcal{G}[^{\mathbf{L}}_{\mathbf{S}}$ *if they cannot be m-separated by any* $\mathbf{Z}$ *with* $\mathbf{S} \subseteq \mathbf{Z} \subseteq \mathbf{V} \setminus \mathbf{L}$ *in* $\mathcal{G}$. *(2) The edge between U and V is*

$U - V$ *if* $U \in An(\mathbf{S} \cup V)$ *and* $V \in An(\mathbf{S} \cup U)$;

$U \to V$ *if* $U \in An(\mathbf{S} \cup V)$ *and* $V \notin An(\mathbf{S} \cup U)$;

$U \leftrightarrow V$ *if* $U \notin An(\mathbf{S} \cup V)$ *and* $V \notin An(\mathbf{S} \cup U)$.

We call $\mathbf{L}$ latent *variables and* $\mathbf{S}$ selection *variables. We say there is* selection bias *if* $\mathbf{S} \neq \emptyset$.

Hence, every MAG represents an infinite set of underlying DAGs that all share the same ancestral relationships. For a given MAG $\mathcal{M}$, we can construct a represented DAG $\mathcal{G}$ by replacing every edge $X - Y$ by a path $X \to S \leftarrow Y$, and every edge $X \leftrightarrow Y$ by $X \leftarrow L \to Y$, where S and L are new nodes; then $\mathcal{M} = \mathcal{G}[^{\mathbf{L}}_{\mathbf{S}}$ where $\mathbf{S}$ and $\mathbf{L}$ are all new nodes. $\mathcal{G}$ is called the *canonical DAG* of $\mathcal{M}$ (Richardson and Spirtes, 2002), which we write as $\mathcal{C}(\mathcal{M})$.

Lemma 5.2 (Preservation of separating sets (Richardson and Spirtes, 2002))**.** $\mathbf{Z}$ *m-separates* $\mathbf{X}, \mathbf{Y}$ *in* $\mathcal{G}[^{\mathbf{L}}_{\mathbf{S}}$ *if and only if* $\mathbf{Z} \cup \mathbf{S}$ *m-separates* $\mathbf{X}, \mathbf{Y}$ *in* $\mathcal{G}$.

We now extend the concept of adjustment to MAGs in the usual way (Maathuis and Colombo, 2013).

Definition 5.3 (Adjustment in MAGs)**.** *Given a MAG* $\mathcal{M} = (\mathbf{V}, \mathbf{E})$ *and two variable sets* $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$, $\mathbf{Z} \subseteq \mathbf{V}$ *is an adjustment set for* $\mathbf{X}, \mathbf{Y}$ *in* $\mathcal{M}$ *if for every probability distribution* $p(\mathbf{v}')$ *consistent with a DAG* $\mathcal{G} = (\mathbf{V}', \mathbf{E}')$ *for which* $\mathcal{G}[^{\mathbf{L}}_{\mathbf{S}} = \mathcal{M}$ *for some* $\mathbf{S}, \mathbf{L} \subseteq \mathbf{V}' \setminus \mathbf{V}$, *we have*

$$p(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} p(\mathbf{y} \mid \mathbf{x}, \mathbf{z}, \mathbf{s}) p(\mathbf{z} \mid \mathbf{s}) \,. \qquad (2)$$

Selection bias (i.e., $\mathbf{S} \neq \emptyset$) substantially complicates adjustment, and in fact nonparametric causal inference in general (Zhang, 2008)[4]. Due to these limitations, we restrict ourselves to the case $\mathbf{S} = \emptyset$ in the rest of this section. Note however that recovery from selection bias is sometimes possible with additional population data, and graphical conditions exist to identify such cases (Barenboim et al., 2014).

5.1 ADJUSTMENT AMENABILITY

In this section we first identify a class of MAGs in which adjustment is impossible because of causal ambiguities – e.g., the simple MAG $X \to Y$ falls into this class, but the larger MAG $A \to X \to Y$ does not.

Definition 5.4 (Visible edge (Zhang, 2008))**.** *Given a MAG $\mathcal{M} = (\mathbf{V}, \mathbf{E})$, an edge $X \to Y \in \mathbf{E}$ is called* visible *if in all DAGs $\mathcal{G} = (\mathbf{V}', \mathbf{E}')$ with $\mathcal{G}[^{\mathbf{L}}_{\mathbf{S}} = \mathcal{M}$ for some $\mathbf{S}, \mathbf{L} \subseteq \mathbf{V}'$, all d-connected walks between X and Y in $\mathcal{G}$ that contain only nodes of $\mathbf{S} \cup \mathbf{L} \cup X \cup Y$ are directed paths.*

Intuitively, an invisible directed edge $X \to Y$ means that there may still hidden confounding factors between X and Y, which is guaranteed not to be the case if the edge is visible.

Lemma 5.5 (Graphical conditions for edge visibility (Zhang, 2008))**.** *In a MAG $\mathcal{M} = (\mathbf{V}, \mathbf{E})$, an edge $X \to Y$ is visible if and only if there is a node A not adjacent to Y where (1) $A \to X \in \mathbf{E}$ or $A \leftrightarrow X \in \mathbf{E}$, or (2) there is a collider path $A \leftrightarrow V_1 \leftrightarrow \ldots \leftrightarrow V_n \leftrightarrow X$ or $A \to V_1 \leftrightarrow \ldots \leftrightarrow V_n \leftrightarrow X$ where all V_i are parents of Y.*

Definition 5.6. *We call a MAG $\mathcal{M} = (\mathbf{V}, \mathbf{E})$* adjustment amenable *w.r.t. $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$ if all proper causal paths from $\mathbf{X}$ to $\mathbf{Y}$ start with a visible directed edge.*

Lemma 5.7. *If a MAG $\mathcal{M} = (\mathbf{V}, \mathbf{E})$ is not adjustment amenable w.r.t. $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$ then there exists no adjustment set $\mathbf{W}$ for $\mathbf{X}, \mathbf{Y}$ in $\mathcal{M}$.*

Proof. If the first edge $X \to D$ on some causal path to $\mathbf{Y}$ in $\mathcal{M}$ is not visible, then there exists a consistent DAG $\mathcal{G}$ where there is a non-causal path between X and $\mathbf{Y}$ via V that could only be blocked in $\mathcal{M}$ by conditioning on D or some of its descendants. But such conditioning would violate the adjustment criterion in $\mathcal{G}$. □

5.2 ADJUSTMENT CRITERION FOR MAGS

We now show that DAG adjustment criterion generalizes to adjustment amenable MAGs. The adjustment criterion and

[4] A counterexample is the graph $A \leftarrow X \to Y$, where we can safely assume that A is the ancestor of a selection variable. A sufficient and necessary condition for adjustment under selection bias is $\mathbf{Y} \perp\!\!\!\perp \mathbf{S} \mid \mathbf{X}$ (Barenboim et al., 2014), which is so restrictive that most statisticians would probably not even speak of "selection bias" anymore in such a case.

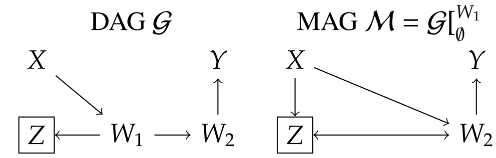

Figure 3: Illustration of the case in the proof of Theorem 5.8 where Z descends from W_1 which in a DAG $\mathcal{G}$ is on a proper causal path from X to Y, but is not a descendant of a node on a proper causal path from X to Y in the MAG $\mathcal{M}$ after marginalizing W_1. In such cases, conditioning on Z will m-connect X and Y in $\mathcal{M}$ via a proper non-causal path.

the constructive back-door criterion are defined like their DAG counterparts (Definitions 4.2 and 4.3), replacing d-with m-separation for the latter.

Theorem 5.8. *Given an adjustment amenable MAG $\mathcal{M} = (\mathbf{V}, \mathbf{E})$ and three disjoint node sets $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, the following statements are equivalent:*

(i) $\mathbf{Z}$ *is an adjustment relative to* $\mathbf{X}, \mathbf{Y}$ *in* $\mathcal{M}$.

(ii) $\mathbf{Z}$ *fulfills the adjustment criterion (AC) w.r.t.* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{M}$.

(iii) $\mathbf{Z}$ *fulfills the constructive backdoor criterion (CBC) w.r.t.* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{M}$.

Proof. The equivalence of (ii) and (iii) is established by observing that the proof of Theorem 4.5 generalizes to m-separation. Below we establish equivalence of (i) and (ii).

$\neg(ii) \Rightarrow \neg(i)$: If $\mathbf{Z}$ violates the adjustment criterion in $\mathcal{M}$, it does so in the canonical DAG $\mathcal{C}(\mathcal{M})$, and thus is not an adjustment in $\mathcal{M}$.

$\neg(i) \Rightarrow \neg(ii)$: Let $\mathcal{G}$ be a DAG with $\mathcal{G}[^{\mathbf{L}}_{\emptyset} = \mathcal{M}$ in which $\mathbf{Z}$ violates the AC. We show that (a) if $\mathbf{Z} \cap Dpcp(\mathbf{X}, \mathbf{Y}) \neq \emptyset$ in $\mathcal{G}$ then $\mathbf{Z} \cap Dpcp(\mathbf{X}, \mathbf{Y}) \neq \emptyset$ in $\mathcal{M}$ as well, or there exists a proper non-causal path in $\mathcal{M}$ that cannot be m-separated; and (b) if $\mathbf{Z} \cap Dpcp(\mathbf{X}, \mathbf{Y}) = \emptyset$ in $\mathcal{G}$ and $\mathbf{Z}$ d-connects a proper non-causal walk in $\mathcal{G}$, then it m-connects a proper non-causal walk in $\mathcal{M}$.

(a) Suppose that in $\mathcal{G}$, $\mathbf{Z}$ contains a node Z in $Dpcp(\mathbf{X}, \mathbf{Y})$, and let $\mathbf{W} = \mathrm{PCP}(\mathbf{X}, \mathbf{Y}) \cap An(Z)$. If $\mathcal{M}$ still contains at least one node $W_1 \in \mathbf{W}$, then W_1 lies on a proper causal path in $\mathcal{M}$ and Z is a descendant of W_1 in $\mathcal{M}$. Otherwise, $\mathcal{M}$ must contain a node $W_2 \in \mathrm{PCP}_{\mathcal{G}}(\mathbf{X}, \mathbf{Y}) \setminus An(Z)$ (possibly $W_2 \in \mathbf{Y}$) such that $W_2 \leftrightarrow A$, $X \to W_2$, and $X \to A$ are edges in $\mathcal{M}$, where $A \in An(Z)$ (possibly $A = Z$; see Fig. 3). Then $\mathcal{M}$ contains an m-connected proper non-causal path $X \to A \leftrightarrow W \to W_2 \to \ldots \to Y$.

(b) Suppose that in $\mathcal{G}$, $\mathbf{Z} \cap Dpcp(\mathbf{X}, \mathbf{Y}) = \emptyset$, and there exists an open proper non-causal path from $\mathbf{X}$ to $\mathbf{Y}$. Then there

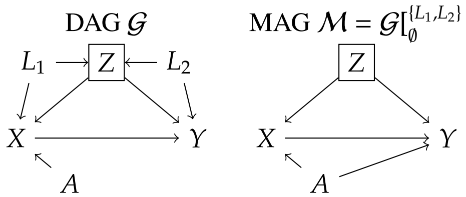

Figure 4: Case (b) in the proof of Theorem 5.8: A proper non-causal path $w_{\mathcal{G}} = X \leftarrow L_1 \rightarrow Z \leftarrow L_s \rightarrow Y$ in a DAG is d-connected by $\mathbf{Z}$, but the corresponding proper non-casual path $w_{\mathcal{M}} = X \leftarrow Z \rightarrow Y$ is not m-connected in the MAG, and its m-connected subpath $w'_{\mathcal{M}} = X \rightarrow Y$ is proper causal. However, this also renders the edge $X \rightarrow Y$ invisible, because otherwise A could be m-separated from Y by $\mathbf{U} = \{X, Z\}$ in $\mathcal{M}$ but not in $\mathcal{G}$.

must then also be a proper non-causal *walk* $w_{\mathcal{G}}$ from some $X \in \mathbf{X}$ to some $Y \in \mathbf{Y}$ (Lemma 7.1), which is d-connected by $\mathbf{Z}$ in $\mathcal{G}$. Let $w_{\mathcal{M}}$ denote the subsequence of $w_{\mathcal{G}}$ formed by nodes in $\mathcal{M}$, which includes all colliders on $w_{\mathcal{G}}$. The sequence $w_{\mathcal{M}}$ is a path in $\mathcal{M}$, but is not necessarily m-connected by $\mathbf{Z}$; all colliders on $w_{\mathcal{M}}$ are in $\mathbf{Z}$ because every non-$\mathbf{Z}$ must be a parent of at least one of its neighbours, but there can subsequences $U, Z_1, \ldots, Z_k, V$ on $w_{\mathcal{M}}$ where all $Z_i \in \mathbf{Z}$ but some of the Z_i are not colliders on $w_{\mathcal{M}}$. However, then we can form from $w_{\mathcal{M}}$ an m-connected walk by bypassing some sequences of $\mathbf{Z}$-nodes (Lemma 7.2). Let $w'_{\mathcal{M}}$ be the resulting walk.

If $w'_{\mathcal{M}}$ is a proper non-causal walk, then there must also exist a proper non-causal path in $\mathcal{M}$ (Lemma 7.1), violating the AC. It therefore remains to show that $w'_{\mathcal{M}}$ is not a proper causal path. This must be the case if $w_{\mathcal{G}}$ does not contain colliders, because then the first edge of $w_{\mathcal{M}} = w'_{\mathcal{M}}$ cannot be a visible directed edge out of X. Otherwise, the only way for $w'_{\mathcal{M}}$ to be proper causal is if all $\mathbf{Z}$-nodes in $w_{\mathcal{M}}$ have been bypassed in $w'_{\mathcal{M}}$ by edges pointing away from $\mathbf{X}$. In that case, one can show by several case distinctions that the first edge $X \rightarrow D$ of $w'_{\mathcal{M}}$, where $D \notin \mathbf{Z}$, cannot be visible (see Figure 4 for an example of such a case). □

5.3 ADJUSTMENT SET CONSTRUCTION

In the previous section, we have already shown that the CBC is equivalent to the AC for MAGs as well; hence, adjustment sets for a given MAG $\mathcal{M}$ can be found by forming the proper back-door graph $\mathcal{M}^{pbd}_{\mathbf{XY}}$ and then applying the algorithms from the previous section. In principle, care must be taken when removing edges from MAGs as the result might not be a MAG; however, this is not the case when removing only directed edges.

Lemma 5.9 (Closure of maximality under removal of directed edges)**.** *Given a MAG $\mathcal{M}$, every graph $\mathcal{M}'$ formed by removing only directed edges from $\mathcal{M}$ is also a MAG.*

Proof. Suppose the converse, i.e. $\mathcal{M}$ is no longer a MAG after removal of some edge $X \rightarrow D$. Then X and D cannot be m-separated even after the edge is removed because X and D are collider connected via a path whose nodes are all ancestors of X or D (Richardson and Spirtes, 2002). The last edge on this path must be $C \leftrightarrow D$ or $C \leftarrow D$, hence $C \notin An(D)$, and thus we must have $C \in An(X)$. But then we get $C \in An(D)$ in $\mathcal{M}$ via the edge $X \rightarrow V$, a contradiction. □

Corollary 5.10. *For every MAG $\mathcal{M}$, the proper back-door graph $\mathcal{M}^{pbd}_{\mathbf{XY}}$ is also a MAG.*

For MAGs that are not adjustment amenable, the CBC might falsely indicate that an adjustment set exists even though that set may not be valid for some represented graph. Fortunately, adjustment amenability is easily tested using the graphical criteria of Lemma 5.5. For each child D of $\mathbf{X}$ in $Dpcp(\mathbf{X}, \mathbf{Y})$, we can test the visibility of all edges $\mathbf{X} \rightarrow D$ simultaneously using depth first search. This means that we can check all potentially problematic edges in time $O(n + m)$. If all tests pass, we are licensed to apply the CBC, as shown above. Hence, we can solve all algorithmic tasks in Table 1 for MAGs in the same way as for DAGs after an $O(k(n + m))$ check of adjustment amenability, where $k \leq |Ch(\mathbf{X})|$.

6 DISCUSSION

We have compiled efficient algorithms for solving several tasks related to m-separators in ancestral graphs, and applied those together with a new, constructive adjustment criterion to provide a complete and informative answer to the question when, and how, a desired causal effect can be estimated by covariate adjustment. Our results fully generalize to MAGs in the absence of selection bias. One may argue that the MAG result is more useful for exploratory applications (inferring a graph from data) than confirmatory ones (drawing a graph based on theory), as researchers will prefer drawing DAGs instead of MAGs due to the easier causal interpretation of the former. Nevertheless, in such settings the results can provide a means to construct more "robust" adjustment sets: If there are several options for covariate adjustment in a DAG, then one can by interpreting the same graph as a MAG possibly generate an adjustment set that is provably valid for a much larger class of DAGs. This might partially address the typical criticism that complete knowledge of the causal structure is unrealistic.

Our adjustment criterion generalizes the work of Shpitser et al. (2010) to MAGs and therefore now completely characterizes when causal effects are estimable by covariate adjustment in the presence of unmeasured confounders with multivariate exposures and outcomes. This also generalizes recent work by Maathuis and Colombo (2013) who provide a criterion which, for DAGs and MAGs without selection bias, is stronger than the back-door criterion but weaker than ours. They moreover show their criterion to hold also for CPDAGs and PAGs, which represent equivalence classes of DAGs and MAGs as they are constructed by causal discovery algorithms. It is possible that the constructive back-door criterion could be generalized further to those cases, which we leave for future work.

7 APPENDIX

In this appendix, we prove Lemma 3.4 and two auxiliary Lemmas that are used in the proof of Theorem 5.8.

Proof of Lemma 3.4. Let us consider a proper walk $w = X, V_1, \ldots, V_n, Y$ with $X \in \mathbf{X}, Y \in \mathbf{Y}$. If w does not contain a collider, all nodes V_i are in $Ant(\mathbf{X} \cup \mathbf{Y})$ and the walk is blocked by $\mathbf{Z}$, unless $\{V_1, \ldots, V_n\} \cap \mathbf{R} = \emptyset$ in which case the walk is not blocked by $\mathbf{Z}_0$ either. If the walk contains colliders $\mathbf{C}$, it is blocked, unless $\mathbf{C} \subseteq \mathbf{Z} \subseteq \mathbf{R}$. Then all nodes V_i are in $Ant(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{I})$ and the walk is blocked, unless $\{V_1, \ldots, V_n\} \cap \mathbf{R} = \mathbf{C}$. Since $\mathbf{C} \subseteq \mathbf{Z}$ is a set of anteriors, there exists a shortest (possible containing 0 edges) path $\pi_j = V_j \to \ldots \to W_j$ for each $V_j \in \mathbf{C}$ with $W_j \in \mathbf{X} \cup \mathbf{Y} \cup \mathbf{I}$ (it cannot contain an undirected edge, since there is an arrow pointing to V_j). Let $\pi'_j = V_j \to \ldots \to W'_j$ be the shortest subpath of π_j that is not blocked by $\mathbf{Z}_0$. Let w' be the walk w after replacing each V_j by the walk $V_j \to \ldots \to W'_j \leftarrow \ldots \leftarrow V_j$. If any of the W_j is in $\mathbf{X} \cup \mathbf{Y}$ we truncate the walk, such that we get the shortest walk between nodes of $\mathbf{X}$ and $\mathbf{Y}$. Since π'_j is not blocked, w' contains no colliders except w'_j and all other nodes of w' are not in $\mathbf{R}$, w' is not blocked and $\mathbf{Z}_0$ is not a separator. □

Lemma 7.1. *Given a DAG $\mathcal{G}$ and sets $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$ satisfying $\mathbf{Z} \cap Dpcp(\mathbf{X}, \mathbf{Y}) = \emptyset$, $\mathbf{Z}$ m-connects a proper non-causal path between $\mathbf{X}$ and $\mathbf{Y}$ if and only if it m-connects a proper non-causal walk between $\mathbf{X}$ and $\mathbf{Y}$.*

Proof. $\Leftarrow$: Let w be the m-connected proper non-causal walk. It can be transformed to an m-connected path π by removing loops of nodes that are visited multiple times. Since no nodes have been added, π remains proper, and the first edges of π and w are the same. So if w does not start with a $\to$ edge, π is non-causal. If w starts with an edge $X \to D$, there exists a collider with a descendant in $\mathbf{Z}$ which is in $De(D)$. So π has to be non-causal, or it would contradict $\mathbf{Z} \cap Dpcp(\mathbf{X}, \mathbf{Y}) = \emptyset$.

$\Rightarrow$: Let π be an m-connected proper non-causal path. It can be changed to an m-connected walk w by inserting $C_i \to \ldots \to Z_i \leftarrow \ldots \leftarrow C_i$ for every collider C_i on π and a corresponding $Z_i \in \mathbf{Z}$. Since no edges are removed from π, w is non-causal, but not necessarily proper, since the inserted walks might contain nodes of $\mathbf{X}$. However, in that case, w can be truncated to a proper walk w' starting at the last node of $\mathbf{X}$ on w. Then w' is non-causal, since it contains the subpath $\mathbf{X} \leftarrow \ldots \leftarrow C_i$. □

Lemma 7.2. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG and let $w_{\mathcal{G}}$ be a walk from $X \in \mathbf{V}$ to $Y \in \mathbf{V}$ that is d-connected by $\mathbf{Z} \subseteq \mathbf{V}$. Let $\mathcal{M} = \mathcal{G}[^{\mathbf{L}}_{\emptyset}$, where $\mathbf{L} \subseteq \mathbf{V} \setminus (\mathbf{Z} \cup X \cup Y)$. Let $w_{\mathcal{M}} = V_1, \ldots, V_{n+1}$ be the subsequence of $w_{\mathcal{G}}$ consisting only of the nodes in $\mathcal{M}$. Then $\mathbf{Z}$ m-connects X and Y in $\mathcal{M}$ via a path along a subsequence $w'_{\mathcal{G}}$ formed from $w_{\mathcal{G}}$ by removing some nodes of $\mathbf{Z}$ (possibly $w'_{\mathcal{G}} = w_{\mathcal{G}}$).*

Proof. (Sketch) The subsequences removed from $w_{\mathcal{M}}$ correspond to maximally long inducing walks in $w_{\mathcal{G}}$ with respect to $\mathbf{L}$. An inducing walk is a collider connected path on which all nodes are ancestors of one of the endpoints, and all non-colliders are in $\mathbf{L}$. The endpoints of inducing walks with respect to $\mathbf{L}$ must be adjacent to each other in $\mathcal{M}$ (similar to Richardson and Spirtes, 2002, for inducing paths); It is easy to see that all $\mathbf{Z}$-nodes which are not colliders on $w_{\mathcal{M}}$ can be removed in this way, e.g. $X \leftarrow Z_1 \leftrightarrow Z_2 \to Y$ can be truncated to X, Y because there must have been an inducing walk in $\mathcal{G}$ via Z_1, Z_2. Additionally, it can be necessary to remove nodes that are colliders on $w_{\mathcal{M}}$, e.g. if $w_{\mathcal{M}} = X \leftarrow Z_1 \to Z_2 \leftarrow Y$ and $Z_2 \in An(X)$, then $w_{\mathcal{G}}$ must have been an inducing walk, and $w'_{\mathcal{M}}$ contains only X and Y even though Z_2 is a collider. To obtain the Lemma, it remains to be shown that no new colliders are created when bypassing nodes in this way. This is done by case distinctions; e.g., in the example $w_{\mathcal{M}} = X \leftarrow Z_1 \to Z_2 \leftarrow Y$ and $Z_2 \in An(X)$, we also have $Y \in An(X)$ and hence $w'_{\mathcal{M}}$ cannot be $X \to Y$ or $x \leftrightarrow y$. □

References

Silvia Acid and Luis M. De Campos. An algorithm for finding minimum d-separating sets in belief networks. In *Proceedings of UAI 1996*, pages 3–10, 1996.

Silvia Acid and Luis M. de Campos. Searching for bayesian network structures in the space of restricted acyclic partially directed graphs. *J. Artif. Intell. Res. (JAIR)*, 18:445–490, 2003.

Elias Barenboim, Jin Tian, and Judea Pearl. Recovering from selection bias in causal and statistical inference. In *Proceedings of AAAI-14*, 2014.

Felix Elwert. *Graphical Causal Models*, pages 245–273. Handbooks of Sociology and Social Research. Springer, 2013.

Adam Glynn and Konstantin Kashin. Front-door versus back-door adjustment with unmeasured confounding: Bias formulas for front-door and hybrid adjustments. Technical report, Harvard University, 2013.

Sander Greenland. Hierarchical regression for epidemiologic analyses of multiple exposures. *Environ. Health Perspect.*, 102 Suppl 8:33–39, Nov 1994.

Marloes H. Maathuis and Diego Colombo. A generalized backdoor criterion. arXiv:1307.5636, 2013.

Judea Pearl. *Causality*. Cambridge University Press, 2009. ISBN 0-521-77362-8.

Thomas Richardson and Peter Spirtes. Ancestral graph markov models. *Annals of Statistics*, 30:927–1223, 2002.

Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash. *Modern Epidemiology*. Wolters Kluwer, 2008. ISBN 0781755646.

Ross D. Shachter. Bayes-ball: The rational pastime. In *Proceedings of UAI 1998*, pages 480–487, 1998.

Ilya Shpitser. Appendum to on the validity of covariate adjustment for estimating causal effects, 2012. unpublished manuscript.

Ilya Shpitser, Tyler VanderWeele, and James Robins. On the validity of covariate adjustment for estimating causal effects. In *Proceedings of UAI 2010*, pages 527–536. AUAI Press, 2010.

Ken Takata. Space-optimal, backtracking algorithms to list the minimal vertex separators of a graph. *Discrete Applied Mathematics*, 158:1660–1667, 2010.

Johannes Textor and Maciej Liśkiewicz. Adjustment criteria in causal diagrams: An algorithmic perspective. In *Proceedings of UAI*, pages 681–688, 2011.

Jin Tian, Azaria Paz, and Judea Pearl. Finding minimal d-separators. Technical Report R-254, University of California, Los Angeles, 1998. URL `ftp.cs.ucla.edu/pub/stat_ser/r254.pdf`.

Tyler J. VanderWeele. On the relative nature of overadjustment and unnecessary adjustment. *Epidemiology*, 20(4): 496–499, Jul 2009.

Jiji Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474, 2008.