Supplement to Identifying causal effects in maximally oriented partially directed acyclic graphs

Emilija Perković
Department of Statistics
University of Washington
Seattle, WA 98195-4322

A PRELIMINARIES

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. For a path $p = \langle X_1, X_2, \dots, X_m \rangle$, the *subpath* from X_i to X_k $(1 \le i \le k \le m)$ is the path $p(X_i, X_k) = \langle X_i, X_{i+1}, \dots, X_k \rangle$.

Concatenation. We denote concatenation of paths by $\oplus$, so that for a path $p = \langle X_1, X_2, \dots, X_m \rangle$, $p = p(X_1, X_r) \oplus p(X_r, X_m)$, for $1 \le r \le m$.

D-separation. If $\mathbf{X}$ and $\mathbf{Y}$ are d-separated given $\mathbf{Z}$ in a DAG $\mathcal{D}$, we write $\mathbf{X} \perp_{\mathcal{D}} \mathbf{Y} | \mathbf{Z}$.

Possible Descendants. If there is a possibly causal path from X to Y, then Y is a *possible descendant* of X. We use the convention that every node is a possible descendant of itself. The set of possible descendants of X in $\mathcal{G}$ is $\mathrm{PossDe}(X, \mathcal{G})$. For a set of nodes $\mathbf{X} \subseteq \mathbf{V}$, we let $\mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \cup_{X \in \mathbf{X}}) \,\mathrm{PossDe}(X, \mathcal{G})$.

Bayesian And Causal Bayesian Networks. If a density f over $\mathbf{V}$ is consistent with DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$, then $(\mathcal{D}, f)$ form a *Bayesian network*. Let $\mathbf{F}$ be a set of density functions made up of all interventional densities $f(\mathbf{v}'|do(\mathbf{x}))$ for any $\mathbf{X} \subset \mathbf{V}$ and $\mathbf{V}' = \mathbf{V} \setminus \mathbf{X}$ that are consistent with $\mathcal{D}$ ($\mathbf{F}$ also includes all observational densities consistent with $\mathcal{D}$), then $(\mathcal{D}, \mathbf{F})$ form a *causal Bayesian network*.

Rules Of The Do-calculus (Pearl, 2009). Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ and $\mathbf{W}$ be pairwise disjoint (possibly empty) sets of nodes in a DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ Let $\mathcal{D}_{\overline{\mathbf{X}}}$ denote the graph obtained by deleting all edges into $\mathbf{X}$ from $\mathcal{D}$. Similarly, let $\mathcal{D}_{\underline{\mathbf{X}}}$ denote the graph obtained by deleting all edges out of $\mathbf{X}$ in $\mathcal{D}$ and let $\mathcal{D}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ denote the graph obtained by deleting all edges into $\mathbf{X}$ and all edges out of $\mathbf{Z}$ in $\mathcal{D}$. Let $(\mathcal{D}, \mathbf{F})$ be a causal Bayesian network, the following rules hold for densities in $\mathbf{F}$.

Rule 1 (Insertion/deletion of observations). If $\mathbf{Y} \perp_{\mathcal{D}_{\overline{\mathbf{X}}}} \mathbf{Z} | \mathbf{X} \cup \mathbf{W}$, then

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{w}) = f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}). \tag{1}$$

Rule 2. If $\mathbf{Y} \perp_{\mathcal{D}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}} \mathbf{Z} | \mathbf{X} \cup \mathbf{W}$, then

$$f(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}). \tag{2}$$

Rule 3. If $\mathbf{Y} \perp_{\mathcal{D}_{\overline{\mathbf{X}\mathbf{Z}(\mathbf{W})}}} \mathbf{Z} | \mathbf{X} \cup \mathbf{W}$, then

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{w}) = f(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}), \tag{3}$$

where $\mathbf{Z}(\mathbf{W}) = \mathbf{Z} \setminus \mathrm{An}(\mathbf{W}, \mathcal{D}_{\overline{\mathbf{X}}})$.

A.1 EXISTING RESULTS

Theorem A.1 (Wright's rule Wright, 1921). *Let $\mathbf{X} = \mathbf{A}\mathbf{X} + \epsilon$, where $\mathbf{A} \in \mathbb{R}^{k \times k}$, $\mathbf{X} = (X_1, \dots, X_k)^T$ and $\epsilon = (\epsilon_1, \dots, \epsilon_k)^T$ is a vector of mutually independent errors with means zero. Moreover, let $\mathrm{Var}(\mathbf{X}) = \mathbf{I}$. Let $\mathcal{D} = (\mathbf{X}, \mathbf{E})$, be the corresponding DAG such that $X_i \to X_j$ in $\mathcal{D}$ if and only if $A_{ji} \neq 0$. A nonzero entry A_{ji} is called the edge coefficient of $X_i \to X_j$. For two distinct nodes $X_i, X_j \in \mathbf{X}$, let $p_1, \dots, p_r$ be all paths between X_i and X_j in $\mathcal{D}$ that do not contain a collider. Then $\mathrm{Cov}(X_i, X_j) = \sum_{s=1}^{r} \pi_s$, where π_s is the product of all edge coefficients along path p_s, $s \in \{1, \dots, r\}$.*

Theorem A.2 (c.f. Theorem 3.2.4 Mardia et al., 1980)**.** *Let $\mathbf{X} = (\mathbf{X_1}^T, \mathbf{X_2}^T)^T$ be a p-dimensional multivariate Gaussian random vector with mean vector $\mu = (\mu_\mathbf{1}^T, \mu_\mathbf{2}^T)^T$ and covariance matrix $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma_{11}} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma_{21}} & \mathbf{\Sigma}_{22} \end{bmatrix}$, so that $\mathbf{X_1}$ is a q-dimensional multivariate Gaussian random vector with mean vector $\mu_\mathbf{1}$ and covariance matrix $\mathbf{\Sigma}_{11}$ and $\mathbf{X_2}$ is a $(p-q)$-dimensional multivariate Gaussian random vector with mean vector $\mu_\mathbf{2}$ and covariance matrix $\mathbf{\Sigma}_{22}$. Then $E[\mathbf{X_2}|\mathbf{X_1} = \mathbf{x_1}] = \mu_\mathbf{2} + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{x_1} - \mu_\mathbf{1})$.*

Algorithm 2: PTO algorithm (Jaber et al., 2018)

input : DAG or CPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$.
output : An ordered list $\mathbf{B} = (\mathbf{B_1}, \dots, \mathbf{B_k}), k \geq 1$ of the bucket decomposition of $\mathbf{V}$ in $\mathcal{G}$.

Let $\mathbf{ConComp}$ be the bucket decomposition of $\mathbf{V}$ in $\mathcal{G}$;
Let $\mathbf{B}$ be an empty list;
while $\mathbf{ConComp} \neq \emptyset$ **do**
  Let $\mathbf{C} \in \mathbf{ConComp}$;
  Let $\overline{\mathbf{C}}$ be the set of nodes in $\mathbf{ConComp}$ that are not in $\mathbf{C}$;
  if *all edges between* $\mathbf{C}$ *and* $\overline{\mathbf{C}}$ *are into* $\mathbf{C}$ *in* $\mathcal{G}$ **then**
    Add $\mathbf{C}$ to the beginning of $\mathbf{B}$;
  end
end
return $\mathbf{B}$;

Lemma A.3 (c.f. Lemma C.1 of Perković et al., 2017, Lemma 8 of Perković et al., 2018)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in a* MPDAG $\mathcal{G}$. *Suppose that there is a proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ that starts with an undirected edge in $\mathcal{G}$, then there is one such path $q = \langle X, V_1, \dots, Y\rangle$, $X \in \mathbf{X}$, $Y \in \mathbf{Y}$ in $\mathcal{G}$ and DAGs $\mathcal{D}^1, \mathcal{D}^2$ in $[\mathcal{G}]$ such that the path in $\mathcal{D}^1$ consisting of the same sequence of nodes as q is of the form $X \to V_1 \to \cdots \to Y$ and in $\mathcal{D}^2$ the path consisting of the same sequence of nodes as q is of the form $X \leftarrow V_1 \to \cdots \to Y$.*

Lemma A.4 (Lemma 3.2 of Perković et al., 2017)**.** *Let p^* be a path from X to Y in a* MPDAG $\mathcal{G}$. *If p^* is non-causal in $\mathcal{G}$, then for every* DAG $\mathcal{D}$ *in $[\mathcal{G}]$ the corresponding path to p^* in $\mathcal{D}$ is non-causal. Conversely, if p is a causal path in at least one* DAG $\mathcal{D}$ *in $[\mathcal{G}]$, then the corresponding path to p in $\mathcal{G}$ is possibly causal.*

Lemma A.5 (Lemma 3.5 of Perković et al., 2017)**.** *Let $p = \langle V_1, \dots, V_k\rangle$ be a definite status path in a* MPDAG $\mathcal{G}$. *Then p is possibly causal if and only if there is no $V_i \leftarrow V_{i+1}$, for $i \in \{1, \dots, k-1\}$ in $\mathcal{G}$.*

Lemma A.6 (Lemma 3.6 of Perković et al., 2017)**.** *Let X and Y be distinct nodes in a* MPDAG $\mathcal{G}$. *If p is a possibly causal path from X to Y in $\mathcal{G}$, then a subsequence p^* of p forms a possibly causal unshielded path from X to Y in $\mathcal{G}$.*

Lemma A.7 (c.f. Lemma 1 of Jaber et al., 2018)**.** *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a CPDAG or DAG and let $\mathbf{B} = (\mathbf{B_1}, \dots, \mathbf{B_k})$, $k \geq 1$, be the output of $\mathtt{PTO}(\mathcal{G})$ (Algorithm 2). Then for each $i, j \in \{1, \dots k\}$, $\mathbf{B_i}$ and $\mathbf{B_j}$ are buckets in $\mathbf{V}$ and if $i < j$, then $\mathbf{B_i} < \mathbf{B_j}$.*

Lemma A.8 (c.f. Lemma E.6 of Henckel et al., 2019)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in an MPDAG $\mathcal{G}$ and suppose that there is no proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ that starts with an undirected edge in $\mathcal{G}$. Let $\mathcal{D}$ be a DAG in $[\mathcal{G}]$. Then $\mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \subseteq \mathrm{De}(\mathbf{X}, \mathcal{G})$.*

B PROOFS FOR SECTION 3.1 OF THE MAIN TEXT

Proof of Proposition 3.2. This proof follows a similar reasoning as the proof of Theorem 2 of Shpitser and Pearl (2006) and proof of Theorem 57 of Perković et al. (2018).

By Lemma A.3, there is a proper possibly causal path $q = \langle X, V_1, \dots, Y\rangle$, $k \geq 1$, $X \in \mathbf{X}$, $Y \in \mathbf{Y}$ in $\mathcal{G}$ and DAGs $\mathcal{D}^1$ and $\mathcal{D}^2$ in $[\mathcal{G}]$ such that $X \to V_1 \to \cdots \to Y$ is in $\mathcal{D}^1$ and $X \leftarrow V_1 \to \cdots \to Y$ is in $\mathcal{D}^2$ (the special case when $k = 1$ is $X \leftarrow Y$).

Consider a multivariate Gaussian density over $\mathbf{V}$ with mean vector zero, constructed using a linear structural causal model (SCM) with Gaussian noise. In particular, each random variable $A \in \mathbf{V}$ is a linear combination of its parents in $\mathcal{D}^1$ and a designated Gaussian noise variable ϵ_A with zero mean and a fixed variance. The Gaussian noise variables $\{\epsilon_A : A \in \mathbf{V}\}$, are mutually independent.

We define the SCM such that all edge coefficients except for the ones on q_1 are 0, and all edge coefficients on q_1 are in $(0, 1)$ and small enough so that we can choose the residual variances so that the variance of every random variable in $\mathbf{V}$ is 1.

The density f of $\mathbf{V}$ generated in this way is consistent with $\mathcal{D}^1$ and thus, f is also consistent with $\mathcal{G}$ and $\mathcal{D}^2$ (Lauritzen et al., 1990). Moreover, f is consistent with DAG $\mathcal{D}^{11}$ that is obtained from $\mathcal{D}^1$ by removing all edges except for the ones on q_1. Analogously, f is also consistent with DAG $\mathcal{D}^{21}$ that is obtained from $\mathcal{D}^2$ by removing all edges except for the ones on q_2. Hence, let $f_1(\mathbf{v}) = f(\mathbf{v})$ and let $f_2(\mathbf{v}) = f(\mathbf{v})$.

Let $f_1(\mathbf{v}'|do(\mathbf{x}))$ be an interventional density consistent with $\mathcal{D}^{11}$. Similarly let $f_2(\mathbf{v}'|do(\mathbf{x}))$ be an interventional density consistent with $\mathcal{D}^{21}$. Then $f_1(\mathbf{v}'|do(\mathbf{x}))$ and $f_1(\mathbf{v}'|do(\mathbf{x}))$ are also interventional densities consistent with $\mathcal{D}^1$ and $\mathcal{D}^2$, respectively. Now, $f_1(\mathbf{y}|do(\mathbf{x}))$ is a marginal interventional density of $\mathbf{Y}$ that can be calculated from the density $f_1(\mathbf{v}'|do(\mathbf{x}))$ and the analagous is true for $f_2(\mathbf{y}|do(\mathbf{x}))$ and $f_2(\mathbf{v}'|do(\mathbf{x}))$.

In order to show that $f_1(\mathbf{y}|do(\mathbf{x})) \neq f_2(\mathbf{y}|do(\mathbf{x}))$, it suffices to show that $f_1(y|do(\mathbf{x} = 1)) \neq f_2(y|do(\mathbf{x} = \mathbf{1}))$ for at least one $Y \in \mathbf{Y}$ when all $\mathbf{X}$ variables are set to 1 by a do-intervention. In order for $f_1(y|do(\mathbf{x} = 1)) \neq$

$f_2(y|do(\mathbf{x}=\mathbf{1}))$ to hold, it is enough to show that the expectation of Y is not the same under these two densities. Hence, let $E_1[Y \mid do(\mathbf{X}=\mathbf{1})]$ denote the expectation of Y, under $f_1(y|do(\mathbf{X}=\mathbf{1}))$ and let $E_2[Y \mid do(\mathbf{X}=\mathbf{1})]$ denote the expectation of $\mathbf{Y}$, under $f_2(y|do(\mathbf{X}=\mathbf{1}))$.

Since Y is d-separated from $\mathbf{X}$ in $\mathcal{D}^{21}_{\overline{\mathbf{X}}}$ we can use Rule 3 of the do-calculus (see equation (3)) to conclude that $E_2[Y \mid do(\mathbf{X}=\mathbf{1})] = E[Y] = 0$. Similarly, since Y is d-separated from $\mathbf{X}$ in $\mathcal{D}^{11}_{\underline{\mathbf{X}}}$, we can use Rule 2 of the do-calculus (see equation (2)) to conclude that $E_1[Y \mid do(\mathbf{X}=\mathbf{1})] = E[Y|X=1]$. By Theorems A.2 and A.1, $E[Y \mid X=1] = \mathrm{Cov}(X,Y) = a$, where a is the product of all edge coefficients on q_1. Since $a \neq 0$, $E_1[Y \mid do(\mathbf{X}=\mathbf{1})] \neq E_2[Y \mid do(\mathbf{X}=\mathbf{1})]$. $\square$

C PROOFS FOR SECTION 3.2 OF THE MAIN TEXT

Lemma C.1. *Let $\mathbf{D}$ be any subset of $\mathbf{V}$ in* MPDAG $\mathcal{G}=(\mathbf{V},\mathbf{E})$. *Then the call to algorithm $PCO(\mathbf{D},\mathcal{G})$ will complete. Meaning that, at each iteration of the while loop in $PCO(\mathbf{D},\mathcal{G})$ (Algorithm 1), there is a bucket $\mathbf{C}$ among the remaining buckets in $\mathbf{ConComp}$ (the bucket decomposition of $\mathbf{V}$) such that all edges between $\mathbf{C}$ and $\mathbf{ConComp} \setminus \mathbf{C}$ are into $\mathbf{C}$ in $\mathcal{G}$.*

Proof of Lemma C.1. Let $\mathbf{C_1},\dots,\mathbf{C_k}$ be the buckets in $\mathbf{ConComp}$ at some iteration of the while loop in the call to PCO$(\mathbf{D},\mathcal{G})$. Suppose for contradiction that there is no bucket $\mathbf{C_i}$, $i \in \{1,\dots,k\}$ such that all edges between $\mathbf{C_i}$ and $\cup_{j=1}^{k}\mathbf{C_j} \setminus \mathbf{C_i}$ are into $\mathbf{C_i}$. We will show that this leads to the conclusion that $\mathcal{G}$ is not acyclic (a contradiction).

Consider a directed graph $\mathcal{G}_1$ constructed so that each bucket in $\mathbf{ConComp}$ represents one node in $\mathcal{G}_1$. Meaning, a bucket $\mathbf{C_i}$, $i \in \{1,\dots,k\}$ is represented by a node C_i in $\mathcal{G}_1$. Also, let $C_i \to C_j$, $i,j \in \{1,\dots,k\}$, be in $\mathcal{G}_1$ if $A \to B$ is in $\mathcal{G}$ and $A \in C_i$, $B \in C_j$.

Since there is no bucket $\mathbf{C_i}$ in $\mathbf{ConComp}$ such that all edges between $\mathbf{C_i}$ and $\cup_{j=1}^{k}\mathbf{C_j} \setminus \mathbf{C_i}$ are into $\mathbf{C_i}$, there is either a directed cycle in $\mathcal{G}_1$, or $C_l \to C_r$ and $C_r \to C_l$ is in $\mathcal{G}_1$ for some $l,r \in \{1,...,k\}$. For simplicity, we will refer to both previously mentioned cases as directed cycles.

Let us choose one such directed cycle in $\mathcal{G}_1$, that is, let $C_{r_1} \to \dots \to C_{r_m} \to C_{r_1}$, $2 \leq m \leq k$, $r_1,\dots,r_m \in \{1,\dots,k\}$, be in $\mathcal{G}_1$. Let $A_i \in \mathbf{C_{r_i}}$ and $B_{i+1} \in \mathbf{C_{r_{i+1}}}$, for all $i \in \{1,\dots,m-1\}$, such that $A_i \to B_{i+1}$ is in $\mathcal{G}$. Additionally, let $A_m \in \mathbf{C_{r_m}}$, and $B_1 \in \mathbf{C_{r_1}}$ such that $A_m \to B_1$ is in $\mathcal{G}$.

Since $A_1 \to B_2$ is in $\mathcal{G}$ and B_2 and A_2 are in the same bucket $\mathbf{C_{r_2}}$ in $\mathcal{G}$, by Lemma C.2, $A_1 \to A_2$. The same reasoning can be applied to conclude that $A_i \to A_{i+1}$, for all $i \in \{1,...,m-1\}$ and also that $A_m \to A_1$ is in $\mathcal{G}$. Thus, $A_1 \to A_2 \to \dots \to A_m \to A_1$, a directed cycle is in $\mathcal{G}$, a contradiction. $\square$

Proof of Lemma 3.5. Lemma C.2 and Lemma A.7 together imply that Algorithm 2 can be applied to a MPDAG $\mathcal{G}$ and also that the output of $\texttt{PTO}(\mathcal{G})$ is the same as that of $\texttt{PCO}(\mathbf{V},\mathcal{G})$. Furthermore, $\texttt{PTO}(\mathcal{G})$ = $\texttt{PCO}(\mathbf{V},\mathcal{G})$ = $(\overline{\mathbf{B_1}},\dots,\overline{\mathbf{B_r}})$ $r \geq k$, where for all $i,j \in \{1,\dots,r\}$, $\overline{\mathbf{B_i}}$ and $\overline{\mathbf{B_j}}$ are buckets in $\mathbf{V}$ in $\mathcal{G}$, and if $i<j$, then $\overline{\mathbf{B_i}} < \overline{\mathbf{B_j}}$ with respect to $\mathcal{G}$.

The statement of the lemma then follows directly from the definition of buckets (Definition 3.3) and Corollary 3.4, since for each $l \in \{1,\dots,k\}$, there exists $s \in \{1,\dots,r\}$ such that $\mathbf{B_l} = \mathbf{D} \cap \overline{\mathbf{B_s}}$ and $(\mathbf{B_1},\dots,\mathbf{B_k})$ is exactly the output of $\texttt{PCO}(\mathbf{V},\mathcal{G})$. $\square$

Lemma C.2. *Let $\mathbf{B}$ be a bucket in $\mathbf{V}$ in* MPDAG $\mathcal{G}=(\mathbf{V},\mathbf{E})$ *and let $X \in \mathbf{V}$, $X \notin \mathbf{B}$. If there is a causal path from X to $\mathbf{B}$ in $\mathcal{G}$, then for every node $B \in \mathbf{B}$ there is a causal path from X to B in $\mathcal{G}$.*

Proof of Lemma C.2. Let p be a shortest causal path from X to $\mathbf{B}$ in $\mathcal{G}$. Then p is of the form $X \to \dots A \to B$, possibly $X=A$ and $A \notin \mathbf{B}$.

Let $B' \in \mathbf{B}$, $B' \neq B$ and let $q = \langle B=W_1,\dots,W_r=B' \rangle$, $r>1$ be a shortest undirected path from B to B' in $\mathcal{G}$. It is enough to show that there is an edge $A \to B'$ is in $\mathcal{G}$.

Since $A \to B - W_2$, by the properties of MPDAGs (Meek, 1995, see Figure 2 in the main text), $A \to W_2$ or $A - W_2$ is in $\mathcal{G}$. Since $A \notin \mathbf{B}$, $A \to W_2$ is in $\mathcal{G}$. If $r=2$, we are done. Otherwise, $A \to W_2 - W_3 - \dots - W_k$ is in $\mathcal{G}$ and and we can apply the same reasoning as above iteratively until we obtain $A \to W_k$ is in $\mathcal{G}$. $\square$

D PROOFS FOR SECTION 3.3 OF THE MAIN TEXT

The proof of Theorem 3.6 is given in the main text. Here we provide proofs for the supporting results.

Lemma D.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in $\mathbf{V}$ in* MPDAG $\mathcal{G}=(\mathbf{V},\mathbf{E})$ *and suppose that there is no proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ that starts with an undirected edge in $\mathcal{G}$. Further, let $(\mathbf{B_1},\dots\mathbf{B_k}) =$ $\texttt{PCO}(\mathrm{An}(\mathbf{Y},\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}),\mathcal{G})$, $k \geq 1$.*

(i) For $i \in \{1,\dots,k\}$, there is no proper possibly causal path from $\mathbf{X}$ to $\mathbf{B_i}$ that starts with an undirected edge in $\mathcal{G}$.

(ii) *For $i \in \{2,\dots,k\}$, let $\mathbf{P_i} = (\cup_{j=1}^{i-1}\mathbf{B_i}) \cap \mathrm{Pa}(\mathbf{B_i},\mathcal{G})$. Then for every DAG $\mathcal{D}$ in $[\mathcal{G}]$ and every interventional density f consistent with $\mathcal{D}$ we have*

$$f(\mathbf{b_i}|\mathbf{b_{i-1}},\dots,\mathbf{b_1}, do(\mathbf{x})) = f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x})).$$

(iii) *For $i \in \{2,\dots,k\}$, let $\mathbf{P_i} = (\cup_{j=1}^{i-1}\mathbf{B_i}) \cap \mathrm{Pa}(\mathbf{B_i},\mathcal{G})$. For $i \in \{1,\dots,k\}$, let $\mathbf{X_{p_i}} = \mathbf{X} \cap \mathrm{Pa}(\mathbf{B_i},\mathcal{G})$. Then for every DAG $\mathcal{D}$ in $[\mathcal{G}]$ and every interventional density f consistent with $\mathcal{D}$ we have*

$$f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x})) = f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x_{p_i}})).$$

Additionally, $f(\mathbf{b_1}|do(\mathbf{x})) = f(\mathbf{b_1}|do(\mathbf{x_{p_1}}))$.

(iv) *For $i \in \{2,\dots,k\}$, let $\mathbf{P_i} = (\cup_{j=1}^{i-1}\mathbf{B_i}) \cap \mathrm{Pa}(\mathbf{B_i},\mathcal{G})$. For $i \in \{1,\dots,k\}$, let $\mathbf{X_{p_i}} = \mathbf{X} \cap \mathrm{Pa}(\mathbf{B_i},\mathcal{G})$. Then for every DAG $\mathcal{D}$ in $[\mathcal{G}]$ and every interventional density f consistent with $\mathcal{D}$ we have*

$$f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x_{p_i}})) = f(\mathbf{b_i}|\,\mathrm{pa}(\mathbf{b_i},\mathcal{G})),$$

for values $\mathrm{pa}(\mathbf{b_i},\mathcal{G})$ of $\mathrm{Pa}(\mathbf{b_i},\mathcal{G})$ that are in agreement with $\mathbf{x}$.

Proof of Lemma D.1. **(i):** Suppose for a contradiction that there is a proper possibly causal path from $\mathbf{X}$ to $\mathbf{B_i}$ that starts with an undirected edge in $\mathcal{G}$. Let $p = \langle X,\dots,B\rangle$, $X \in \mathbf{X}$, $B \in \mathbf{B_i}$, be a shortest such path in $\mathcal{G}$. Then p is unshielded in $\mathcal{G}$ (Lemma A.6).

Since $B \in \mathrm{An}(\mathbf{Y}, \mathcal{G}_{\mathbf{V}\setminus\mathbf{X}})$ there is a causal path q from B to $\mathbf{Y}$ in $\mathcal{G}$ that does not contain a node in $\mathbf{X}$. No node other than B is both on q and p (otherwise, by definition p is not possibly causal from X to B). Hence, by Lemma D.2, $p \oplus q$ is a proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ that starts with an undirected edge in $\mathcal{G}$, which is a contradiction.

(ii): Let $\mathbf{N_i} = (\cup_{j=1}^{i-1}\mathbf{B_j}) \setminus \mathrm{Pa}(\mathbf{B_i},\mathcal{G})$. If $\mathbf{B_i} \perp_{\mathcal{D}_{\overline{\mathbf{X}}}} \mathbf{N_i} \mid (\mathbf{X} \cup \mathbf{P_i})$, then by Rule 1 of the do calculus: $f(\mathbf{b_i}|\mathbf{b_{i-1}},\dots,\mathbf{b_1}, do(\mathbf{x})) = f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x}))$ (see equation (1)).

Suppose for a contradiction that there is a path from $\mathbf{B_i}$ to $\mathbf{N_i}$ that is d-connecting given $\mathbf{X} \cup \mathbf{P_i}$ in $\mathcal{D}_{\overline{\mathbf{X}}}$. Let $p = \langle B_i,\dots,N\rangle$, $B_i \in \mathbf{B_i}$, $N \in \mathbf{N_i}$ be a shortest such path. Let p^* be the path in $\mathcal{G}$ that consists of the same sequence of nodes as p in $\mathcal{D}_{\overline{\mathbf{X}}}$.

First suppose that p is of the form $B_i \to \dots N$. Since $B_i \in \mathbf{B_i}$ and $\mathbf{N_i} \subseteq (\cup_{j=1}^{i-1}\mathbf{B_j})$, p is not causal from B_i to N (Lemma 3.5). Hence, let C be the closest collider to B_i on p, that is, p has the form $B_i \to \dots \to C \leftarrow \dots N$. Since p is d-connecting given $\mathbf{X} \cup \mathbf{P_i}$ in $\mathcal{D}_{\overline{\mathbf{X}}}$, C must be an ancestor of $\mathbf{P_i}$ in $\mathcal{D}_{\overline{\mathbf{X}}}$. However, then there is a causal path from $B_i \in \mathbf{B_i}$ to $\mathbf{P_i} \subseteq (\cup_{j=1}^{i-1}\mathbf{B_j})$ which contradicts Lemma 3.5.

Next, suppose that p is of the form $B_i \leftarrow A \dots N$, $A \notin \mathbf{B_i}$. Since $\mathrm{Pa}(\mathbf{B_i},\mathcal{G}) \subseteq (\mathbf{X} \cup \mathbf{P_i})$ and since p is d-connecting given $(\mathbf{X} \cup \mathbf{P_i})$, $B_i - A$ is in $\mathcal{G}$ and $A \notin (\mathbf{X} \cup \mathbf{P_i})$.

Note that p^* cannot be undirected, since that would imply that $N \in \mathbf{B_i}$ and contradict Lemma 3.5. Hence, let B be the closest node to B_i on p^* such that $p^*(B,N)$ starts with a directed edge (possibly $B = A$). Then p^* is either of the form $B_i - A - \dots - L - B \to R \dots N$ or of the form $B_i - A - \dots - L - B \leftarrow R \dots N$.

Suppose first that p^* is of the form $B_i - A - \dots - L - B \to R \dots N$. Then $B \notin (\mathbf{X} \cup \mathbf{P_i} \cup \mathbf{B_i})$ otherwise, p is either blocked by $\mathbf{X} \cup \mathbf{P_i}$, or a shorter path could have been chosen.

Let $(\mathbf{B_1'},\dots\mathbf{B_r'}) = \mathrm{PCO}(\mathbf{V},\mathcal{G})$, $r \geq k$. Let $l \in \{i,\dots,r\}$ such that $\mathbf{B_l'} \cap \mathbf{B_i} \neq \emptyset$, then $B_i, B \in \mathbf{B_l'}$ and $N \in (\cup_{j=1}^{l-1}\mathbf{B_j'})$. Now consider subpath $p(B,N)$. By Lemma 3.5, $p(B,N)$ cannot be causal from B to N. Hence, there is a collider on $p(B,N)$ and we can derive the contradiction using the same reasoning as above.

Suppose next that p^* is of the form $B_i - A - \dots - L - B \leftarrow R \dots N$. Then either $R \to L$ or $R - L$ is in $\mathcal{G}$ (Meek, 1995, see Figure 4 in the main text). Then $\langle L,R\rangle$ is also an edge in $\mathcal{D}_{\overline{\mathbf{X}}}$ otherwise, L or R is in $\mathbf{X}$ and a non-collider on p, so p would be blocked by $\mathbf{X} \cup \mathbf{P_i}$.

Hence, $q = p(B_i,L) \oplus \langle L,R\rangle \oplus p(R,N)$ is a shorter path than p in $\mathcal{D}_{\overline{\mathbf{X}}}$. If L and R have the same collider/non-collider status on q on p, then q is also d-connecting given $\mathbf{X} \cup \mathbf{P_i}$, which would contradict our choice of p. Hence, the collider/non-collider status of L or R, is different on p and q. We now discuss the cases for the change of collider/non-collider status of L and R and derive a contradiction in each.

Suppose that L is a collider on q, and a non-collider on p. This implies that $W \to L \to B \leftarrow R$ is a subpath of p and $L \leftarrow R$ is in $\mathcal{D}_{\overline{\mathbf{X}}}$. Even though L is not a collider on p, B is a collider on p and $L \in \mathrm{An}(B, \mathcal{D}_{\overline{\mathbf{X}}})$. Since p is d-connecting given $\mathbf{X} \cup \mathbf{P_i}$, $\mathrm{De}(B,\mathcal{D}_{\overline{\mathbf{X}}}) \cap (\mathbf{X} \cup \mathbf{P_i}) \neq \emptyset$. However, then also $\mathrm{De}(L,\mathcal{D}_{\overline{\mathbf{X}}}) \cap (\mathbf{X} \cup \mathbf{P_i}) \neq \emptyset$ and q is also d-connecting given $\mathbf{X} \cup \mathbf{P_i}$ and a shorter path between $\mathbf{B_i}$ and $\mathbf{N_i}$ than p, which is a contradiction.

The contradiction can be derived in exactly the same way as above in the case when R is a collider on q, and a non-collider on p. Since $B \leftarrow R$ is in $\mathcal{D}_{\overline{\mathbf{X}}}$, R cannot be anything but a non-collider on q, so the only case left to consider is if L is a non-collider on q and a collider on p.

For L to be a non-collider on q and a collider on p, $W \to L \leftarrow B \leftarrow R$ must be a subpath of p and $L \to R$ should be in $\mathcal{D}_{\overline{\mathbf{X}}}$. But then there is a cycle in $\mathcal{D}_{\overline{\mathbf{X}}}$, which is a contradiction.

(iii): We will show that $f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x})) = f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x_{p_i}}))$. The simpler case, $f(\mathbf{b_1}|do(\mathbf{x})) = f(\mathbf{b_1}|(\mathbf{x_{p_1}})$ follows from the same proof, when $\mathbf{B_i}$ is replaced by $\mathbf{B_1}$ and $\mathbf{P_i}$ is removed.

Let $\mathbf{X_{n_i}} = \mathbf{X} \setminus \mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$ and let $\mathbf{X}'_{\mathbf{n_i}} = \mathbf{X_{n_i}} \setminus \mathrm{An}(\mathbf{P_i}, \mathcal{D}_{\overline{\mathbf{X_{p_i}}}})$. That is $X \in \mathbf{X}'_{\mathbf{n_i}}$ if $X \in \mathbf{X_{n_i}}$ and if there is no causal path from X to $\mathbf{P_i}$ in $\mathcal{D}$ that does not contain a node in $\mathbf{X_{p_i}}$.

Note that $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G}) = \mathbf{X_{p_i}} \cup \mathbf{P_i}$. By Rule 3 of the do-calculus, for $f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x})) = f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x_{p_i}}))$ to hold, it is enough to show that $\mathbf{B_i} \perp_{\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}} \mathbf{X_{n_i}} | \mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$ (see equation (3)).

Suppose for a contradiction that there is a d-connecting path from $\mathbf{B_i}$ to $\mathbf{X_{n_i}}$ in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. Let $p = \langle B_i, \dots, X \rangle$, $B_i \in \mathbf{B_i}$, $X \in \mathbf{X_{n_i}}$, be a shortest such path in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. Let p^* be the path in $\mathcal{G}$ that consists of the same sequence of nodes as p in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. This proof follows a very similar line of reasoning to the proof of (ii) above.

Let $(\mathbf{B}'_\mathbf{1}, \dots \mathbf{B}'_\mathbf{r}) = \mathrm{PCO}(\mathbf{V}, \mathcal{G})$, $r \geq k$. Let $l \in \{i, \dots, r\}$ such that $\mathbf{B}'_\mathbf{l} \cap \mathbf{B_i} \neq \emptyset$, then $B_i \in \mathbf{B}'_\mathbf{l}$ and $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G}) \subseteq (\cup_{j=1}^{i-1} \mathbf{B_j})$.

Suppose that p is of the form $B_i \to \dots X$. If $X \in \mathbf{X}'_{\mathbf{n_i}}$, then p is not a causal path since p is a path in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. Otherwise, $X \in \mathrm{An}(\mathbf{P_i}, \mathcal{D}_{\overline{\mathbf{X_{p_i}}}})$ and so any causal path from B_i to X would need to contain a node in $\mathbf{X_{p_i}}$ and hence, would be blocked by $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$. Thus, p is not a causal path from B_i to X.

Hence, let C be the closest collider to B_i on p, that is, p has the form $B_i \to \dots \to C \leftarrow \dots X$. Since p is d-connecting given $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$, C is be an ancestor of $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$ in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. However, this would imply that there is a causal path from $B_i \in \mathbf{B}'_\mathbf{l}$ to $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G}) \subseteq (\cup_{j=1}^{i-1} \mathbf{B_j})$ in $\mathcal{D}_{\underline{\mathbf{X_{p_i}}}}$, which contradicts Lemma 3.5.

Next, suppose that p is of the form $B_i \leftarrow A \dots X$, $A \notin \mathbf{B_i}$. Since p is d-connecting given $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$, $A \notin \mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$. Hence, $B_i - A$ is in $\mathcal{G}$.

Then $A \in \mathbf{B}'_\mathbf{l}$. Note that by (i) above, $\mathbf{X} \cap \mathbf{B}'_\mathbf{l} = \emptyset$, so p^* is not an undirected path in $\mathcal{G}$. Hence, let B be the closest node to B_i on p^* such that $p^*(B, X)$ starts with a directed edge (possibly $B = A$). Then p^* is either of the form $B_i - A - \dots - L - B \to R \dots X$ or of the form $B_i - A - \dots - L - B \leftarrow R \dots X$.

Suppose first that p^* is of $B_i - A - \dots - L - B \to R \dots X$. Then $B \in \mathbf{B}'_\mathbf{l}$ and so $B \notin \mathbf{X}$. Since p is d-connecting given $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$, $B \notin \mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$ and additionally, $B \notin \mathbf{B_i}$ otherwise, a shorter path could have been chosen.

Now consider subpath $p(B, X)$. There is at least one collider on $p(B, X)$. Since $B, B_i \in \mathbf{B}'_\mathbf{l}$, the same reasoning as above can be used to derive a contradiction in this case.

Suppose next that p^* is of the form $B_i - A - \dots - L - B \leftarrow R \dots X$. Then either $R \to L$ or $R - L$ is in $\mathcal{G}$ (Meek, 1995, see Figure 4 in the main text). We first show that in either case, edge $\langle L, R \rangle$ is also in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$.

Since $L \in \mathbf{B}'_\mathbf{l}$ and since $\mathbf{X} \cap \mathbf{B}'_\mathbf{l} = \emptyset$, $L \notin \mathbf{X}$. Hence, if $R \to L$ is in $\mathcal{G}$, $R \to L$ is in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. If $R - L$ is in $\mathcal{G}$, then $R \in \mathbf{B}'_\mathbf{l}$ and since $\mathbf{X} \cap \mathbf{B}'_\mathbf{l} = \emptyset$, $R \notin \mathbf{X}$, so $\langle L, R \rangle$ is in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$.

Hence, $q = p(B_i, L) \oplus \langle L, R \rangle \oplus p(R, X)$ is a shorter path than p in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. If L and R have the same collider/non-collider status on q on p, then q is also d-connecting given $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$, which would contradict our choice of p. Hence, the collider/non-collider status of L or R, is different on p and q. We now discuss the cases for the change of collider/non-collider status of L and R and derive a contradiction in each.

Suppose that L is a collider on q, and a non-collider on p. This implies that $W \to L \to B \leftarrow R$ is a subpath of p and $L \leftarrow R$ are in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. Even though L is not a collider on p, B is a collider on p and $L \in \mathrm{An}(B, \mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}})$. Since p is d-connecting given $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$, $\mathrm{De}(B, \mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}) \cap \mathrm{Pa}(\mathbf{B_i}, \mathcal{G}) \neq \emptyset$. However, then also $\mathrm{De}(L, \mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}) \cap \mathrm{Pa}(\mathbf{B_i}, \mathcal{G}) \neq \emptyset$ and q is also d-connecting given $\mathrm{Pa}(\mathbf{B_i}, \mathcal{G})$ and a shorter path between $\mathbf{B_i}$ and $\mathbf{X_{n_i}}$ than p, which is a contradiction.

The contradiction can be derived in exactly the same way as above in the case when R is a collider on q, and a non-collider on p. Since $B \leftarrow R$ is in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$, R cannot be anything but a non-collider on q, so the only case left to consider is if L is a non-collider on q and a collider on p.

For L to be a non-collider on q and a collider on p, $W \to L \leftarrow B \leftarrow R$ must be a subpath of p and $L \to R$ should be in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$. But then there is a cycle in $\mathcal{D}_{\overline{\mathbf{X_{p_i}}\mathbf{X}'_{\mathbf{n_i}}}}$, which is a contradiction.

(iv):. If $\mathbf{B_i} \perp_{\mathcal{D}_{\underline{\mathbf{X_{p_i}}}}} \mathbf{X_{p_i}} | \mathbf{P_i}$, then $f(\mathbf{b_i}|\mathbf{p_i}, do(\mathbf{x_{p_i}})) = f(\mathbf{b_i}|pa(\mathbf{b_i}, \mathcal{G}))$ by Rule 2 of the do calculus (equation (2)).

Suppose for a contradiction that there is a d-connecting path from $\mathbf{B_i}$ to $\mathbf{X_{p_i}}$ in $\mathcal{D}_{\underline{\mathbf{X_{p_i}}}}$. Let $p = \langle B_i, \dots, X \rangle$, $B_i \in \mathbf{B_i}$, $X \in \mathbf{X_{p_i}}$, be a shortest such path in $\mathcal{D}_{\underline{\mathbf{X_{p_i}}}}$. Let

p^* be the path in $\mathcal{G}$ that consists of the same sequence of nodes as p in $\mathcal{D}_{\overline{\mathbf{X}}}$. This proof follows a very similar line of reasoning to the proof of (ii) above.

Let $(\mathbf{B_1'}, \dots \mathbf{B_r'}) = \texttt{PCO}(\mathbf{V}, \mathcal{G})$, $r \geq k$. Let $l \in \{i, \dots, r\}$ such that $\mathbf{B_l'} \cap \mathbf{B_i} \neq \emptyset$, then $B_i \in \mathbf{B_l'}$ and by (i) above, $\mathbf{X_{P_i}} \subseteq (\cup_{j=1}^{l-1} \mathbf{B_j'})$.

Suppose that p is of the form $B_i \to \dots X$. Since $B_i \in \mathbf{B_l'}$ and $\mathbf{X_{P_i}} \subseteq (\cup_{j=1}^{l-1} \mathbf{B_j'})$, by Lemma 3.5, there is at least one collider on p. Hence, let C be the closest collider to B_i on p, that is, p has the form $B_i \to \dots \to C \leftarrow \dots X$. Since p is d-connecting given $\mathbf{P_i}$ in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$, C is be an ancestor of $\mathbf{P_i}$ in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$. However, this would imply that there is a causal path from $B_i \in \mathbf{B_i}$ to $\mathbf{P_i} \subseteq (\cup_{j=1}^{i-1} \mathbf{B_j})$ in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$, which contradicts Lemma 3.5.

Next, suppose that p is of the form $B_i \leftarrow A \dots X$, $A \notin \mathbf{B_i}$. Since p is a path in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$, $A \notin \mathbf{X_{P_i}}$. Additionally, since p is d-connecting given $\mathbf{P_i}$, $A \notin \mathbf{P_i}$. Hence, $B_i - A$ is in $\mathcal{G}$.

Then $A \in \mathbf{B_l'}$ and since $X \in (\cup_{j=1}^{l-1} \mathbf{B_j'})$, $p^*(A, X)$ is not an undirected path in $\mathcal{G}$. Hence, let B be the closest node to B_i on p^* such that $p^*(B, X)$ starts with a directed edge (possibly $B = A$). Then p^* is either of the form $B_i - A - \dots - L - B \to R \dots X$ or of the form $B_i - A - \dots - L - B \leftarrow R \dots X$.

Suppose first that p^* is of $B_i - A - \dots - L - B \to R \dots X$. Then $B \in \mathbf{B_l'}$ and since $\mathbf{X_{P_i}} \subseteq (\cup_{j=1}^{l-1} \mathbf{B_j'})$, $B \notin \mathbf{X_{P_i}}$. Since p is d-connecting given $\mathbf{P_i}$, $B \notin \mathbf{P_i}$ and additionally, $B \notin \mathbf{B_i}$ otherwise, a shorter path could have been chosen.

Now consider subpath $p(B, X)$. Since $B, B_i \in \mathbf{B_l'}$, the same reasoning as above can be used to derive a contradiction in this case.

Suppose next that p^* is of the form $B_i - A - \dots - L - B \leftarrow R \dots X$. Then either $R \to L$ or $R - L$ is in $\mathcal{G}$ (Meek, 1995, see Figure 4 in the main text). Since $R \to B$ is in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$, $R \notin \mathbf{X_{P_i}}$. Since $L \in \mathbf{B_l'}$, $L \notin \mathbf{X_{P_i}}$, so $\langle L, R \rangle$ is also in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$.

Hence, $q = p(B_i, L) \oplus \langle L, R \rangle \oplus p(R, X)$ is a shorter path than p in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$. If L and R have the same collider/non-collider status on q on p, then q is also d-connecting given $\mathbf{P_i}$, which would contradict our choice of p. Hence, the collider/non-collider status of L or R, is different on p and q. We now discuss the cases for the change of collider/non-collider status of L and R and derive a contradiction in each.

Suppose that L is a collider on q, and a non-collider on p. This implies that $W \to L \to B \leftarrow R$ is a subpath of p and $L \leftarrow R$ are in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$. Even though, L is not a collider on p, B is a collider on p and $L \in \mathrm{An}(B, \mathcal{D}_{\underline{\mathbf{X_{P_i}}}})$. Since p is d-connecting given $\mathbf{P_i}$, $\mathrm{De}(B, \mathcal{D}_{\underline{\mathbf{X_{P_i}}}}) \cap \overline{\mathbf{P_i}} \neq \emptyset$. However, then also $\mathrm{De}(L, \mathcal{D}_{\underline{\mathbf{X_{P_i}}}}) \cap \mathbf{P_i} \neq \emptyset$ and q is also d-connecting given $\mathbf{P_i}$ and a shorter path between $\mathbf{B_i}$ and $\mathbf{X_{P_i}}$ than p, which is a contradiction.

The contradiction can be derived in exactly the same way as above in the case when R is a collider on q, and a non-collider on p. Since $B \leftarrow R$ is in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$, R cannot be anything but a non-collider on q, so the only case left to consider is if L is a non-collider on q and a collider on p.

For L to be a non-collider on q and a collider on p, $W \to L \leftarrow B \leftarrow R$ must be a subpath of p and $L \to R$ should be in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$. But then there is a cycle in $\mathcal{D}_{\underline{\mathbf{X_{P_i}}}}$, which is a contradiction. $\square$

Lemma D.2. *Let* X, Y *and* Z *be distinct nodes in* MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. *Suppose that there is an unshielded possibly causal path* p *from* X *to* Y *and a causal path* q *from* Y *to* Z *in* $\mathcal{G}$ *such that the only node that* p *and* q *have in common is* Y. *Then* $p \oplus q$ *is a possibly causal path from* X *to* Z.

Proof of Lemma D.2. Suppose for a contradiction that there is an edge $V_q \to V_p$, where V_q is a node on q and V_p is a node on p (additionally, $V_p \neq Y \neq V_q$). Then $p(V_p, Y)$ cannot be a causal path from V_p to Y since otherwise there is a cycle in $\mathcal{G}$. So $p(V_p, Y)$ takes the form $V_p - V_{p+1} \dots Y$.

Let $\mathcal{D}$ be a DAG in $[\mathcal{G}]$, that contains $V_p \to V_{p+1}$. Since $p(V_p, Y)$ is an unshielded possibly causal path in $\mathcal{G}$, it corresponds to $V_p \to \dots \to Y$ in $\mathcal{D}$. Then $V_q \to V_p \to \dots \to Y$ and $q(Y, V_q)$ form a cycle in $\mathcal{D}$, a contradiction. $\square$

Proof of Corollary 3.7. The first statement in Corollary 3.7 follows from the proof of Theorem 3.6 when replacing $\mathbf{Y}$ with $\mathbf{V}$ and $\mathbf{X}$ with empty set.

For the second statement in Corollary 3.7, note that since there are no undirected edges $X - V$ in $\mathcal{G}$, where $X \in \mathbf{X}$ and $V \in \mathbf{V}'$, some of the buckets $\mathbf{V_i}$, $i \in \{1, \dots, k\}$ in the bucket decomposition of $\mathbf{V}$ will contain only nodes in $\mathbf{X}$. Hence, obtaining the bucket decomposition of $\mathbf{V}' = \mathbf{V} \setminus \mathbf{X}$ is the same as leaving out buckets $\mathbf{V_i}$ that contain only nodes in $\mathbf{X}$ from $\mathbf{V_1}, \dots, \mathbf{V_k}$. The statement then follows from Theorem 3.6 when taking $\mathbf{Y} = \mathbf{V}'$. $\square$

E PROOFS FOR SECTION 4 OF THE MAIN TEXT

Proof of Proposition 4.2. If the causal effect of X on Y is not identifiable in $\mathcal{G}$, by Theorem 3.6, there is a proper possibly causal path from X to Y that starts with an undirected edge in $\mathcal{G}$. Then by Theorem 4.1, there is no adjustment set relative to (X, Y) in $\mathcal{G}$.

Hence, suppose that there is no proper possibly causal path from X to Y that starts with an undirected edge in $\mathcal{G}$ and consider $\mathrm{Pa}(X, \mathcal{G})$. By Theorem 4.1, it is enough to show that $\mathrm{Pa}(X, \mathcal{G})$ satisfies the generalized adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$.

If $\mathcal{G}$ is a DAG, $\mathrm{Pa}(X, \mathcal{G})$ is an adjustment set relative to (X, Y) by Theorem 3.3.2 of Pearl (2009). Hence, suppose that $\mathcal{G}$ is not a DAG.

Since $\mathcal{G}$ is acyclic, $\mathrm{Pa}(X, \mathcal{G}) \cap \mathrm{De}(X, \mathcal{G}) = \emptyset$. Additionally, by Lemma A.8, $\mathrm{Forb}(X, Y, \mathcal{G}) \subseteq \mathrm{De}(X, \mathcal{G})$. Hence, $\mathrm{Pa}(X, \mathcal{G})$ satisfies $\mathrm{Pa}(X, \mathcal{G}) \cap \mathrm{Forb}(X, Y, \mathcal{G}) = \emptyset$, that is, condition 2 in Theorem 4.1 relative to (X, Y) in $\mathcal{G}$.

Consider a non-causal definite status path p from X to Y. If p is of the form $X \leftarrow \dots Y$ in $\mathcal{G}$, then p is blocked by $\mathrm{Pa}(X, \mathcal{G})$. If p is of the form $X \rightarrow \dots Y$, then p contains at least one collider $C \in \mathrm{De}(X, \mathcal{G})$ and since $\mathrm{Pa}(X, \mathcal{G}) \cap \mathrm{De}(X, \mathcal{G}) = \emptyset$, p is blocked by $\mathrm{Pa}(X, \mathcal{G})$.

Lastly, suppose that p is of the form $X - \dots Y$. Since p is a non-causal path from X to Y and since p is of definite status in $\mathcal{G}$, by Lemma A.5, there is at least one edge pointing towards X on p. Let D be the closest node to X on p such that $p(D, Y)$ is of the form $D \leftarrow \dots Y$ in $\mathcal{G}$. Then by Lemma A.5, $p(X, D)$ is a possibly causal path from X to D so let p' be an unshielded subsequence of $p(X, D)$ that forms a possibly causal path from X to D in $\mathcal{G}$ (Lemma A.6). Additionally, p is of definite status, so D must be a collider on p.

In order for p to be blocked by $\mathrm{Pa}(X, \mathcal{G})$ it is enough to show that $\mathrm{De}(D, \mathcal{G}) \cap \mathrm{Pa}(X, \mathcal{G}) = \emptyset$. Suppose for a contradiction that $E \in \mathrm{De}(D, \mathcal{G}) \cap \mathrm{Pa}(X, \mathcal{G})$. Let $q$be a directed path from D to E in $\mathcal{G}$. Then p' and q satisfy Lemma D.2 in $\mathcal{G}$, so $p' \oplus q$ is a possibly causal path from X to E. By definition of a possibly causal path in MPDAGs, this contradicts that $E \in \mathrm{Pa}(X, \mathcal{G})$. □

Lemma E.1. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be disjoint node sets in an* MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. *If there is no possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$, *then for any observational density* f *consistent with* $\mathcal{G}$ *we have*

$$f(\mathbf{y}|do(\mathbf{x})) = f(\mathbf{y}).$$

Proof of Lemma E.1. Lemma E.1 follows from Lemma A.4 and Rule 3 of the do-calculus of Pearl (2009) (see equation (3)). □

References

Henckel, L., Perković, E., and Maathuis, M. H. (2019). Graphical criteria for efficient total effect estimation via adjustment in causal linear models. arXiv:1907.02435.

Jaber, A., Zhang, J., and Bareinboim, E. (2018). A graphical criterion for effect identification in equivalence classes of causal diagrams. In *Proceedings of IJCAI 2018*, pages 5024–5030.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, 20(5):491–505.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1980). *Multivariate analysis (probability and mathematical statistics)*. Academic Press London.

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

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

Perković, E., Kalisch, M., and Maathuis, M. H. (2017). Interpreting and using CPDAGs with background knowledge. In *Proceedings of UAI 2017*.

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

Shpitser, I. and Pearl, J. (2006). Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of AAAI 2006*, pages 1219–1226.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20(7):557–585.