
Dependency in DAG models with Hidden Variables
(Supplementary Material)

Robin J. Evans[1]

[1]Department of Statistics, University of Oxford, Oxford, UK

B GRAPHS

The first concept we will need is an extension to ADMGs in which we allow some vertices to be 'fixed'. We define the *siblings* of a vertex to be its neighbours via bidirected edges:

$$\text{sib}_{\mathcal{G}}(v) \equiv \{w : v \leftrightarrow w \text{ in } \mathcal{G}\}.$$

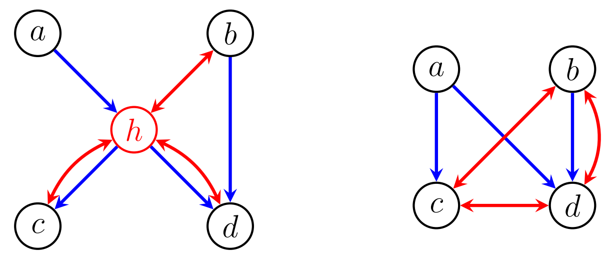
A CADMG $\mathcal{G}(V, W)$ is an ADMG with a set of *random* vertices V and *fixed* vertices W, with the property that $\text{sib}_{\mathcal{G}}(w) \cup \text{pa}_{\mathcal{G}}(w) = \emptyset$ for every $w \in W$. An example can be found in Figure 10(b); note that we depict fixed vertices with rectangular nodes, and random vertices with round nodes. Random vertices in a CADMG correspond to random variables, as in standard graphical models, while fixed vertices correspond to variables that were fixed to a specific value by some operation, such as conditioning or causal interventions. The genealogical relations in Section 2 generalize in a straightforward way to CADMGs by ignoring the distinction between V and W; the only exception is that districts are only defined for random vertices, so that the districts in the graph partition only V, rather than $V \cup W$.

B.1 LATENT PROJECTION

The *latent projection* of a CADMG $\mathcal{G}(V \dot\cup L, W)$ to another graph $\mathcal{G}'(V, W)$ is given by following the rules:

- if there is a directed path from $a \in V \cup W$ to $b \in V$, and any interior vertices are in L, then add $a \to b$;
- if there is a path between $a, b \in V$ without any adjacent arrowheads, and any interior vertices are in L, that starts and ends with an arrow at a and b, then add $a \leftrightarrow b$.

As an example, consider the ADMG in Figure 8(a), with variable h designated as latent. Then the projection of this is given by the ADMG in Figure 8(b).



(a) (b)

Figure 8: (a) An ADMG in which h is latent; (b) its latent projection over $\{a, b, c, d\}$.

B.2 ARID PROJECTION

Example B.1. The *maximal arid projection* of the ADMG $\mathcal{G}$ in Figure 9(a) is given in 9(b). In the graph (a) we have $\langle d \rangle_{\mathcal{G}} = \{b, d\}$, so $\text{pa}_{\mathcal{G}}(\langle d \rangle) = \{a, b, c\}$. As a result, in (b) all these vertices are parents of d. In addition, $\langle \{d, e\} \rangle_{\mathcal{G}} = \{b, c, d, e\}$ which is bidirected connected, so we add the edge $d \leftrightarrow e$ into (b). All other adjacencies are as in (a).

C THE NESTED MARKOV MODEL

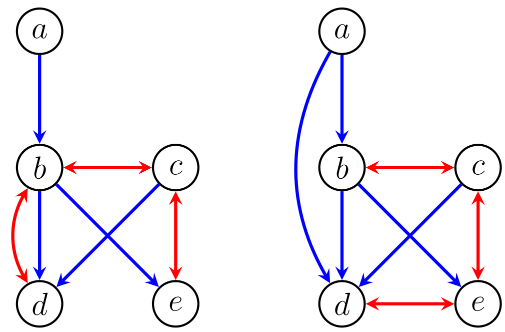
C.1 FIXING

A vertex $r \in V$ is said to be *fixable* in a CADMG $\mathcal{G}(V, W)$ if $\text{dis}_{\mathcal{G}}(r) \cap \text{de}_{\mathcal{G}}(r) = \emptyset$. For instance, the vertices a, c and d are all fixable in the graph in Figure 10(a), but b is not because d is both its descendant and its sibling.

For any $v \in V$, such that $\text{ch}_{\mathcal{G}}(v) = \emptyset$, the *Markov blanket* of v in a CADMG $\mathcal{G}$ is defined as

$$\text{mb}_{\mathcal{G}}(v) \equiv (\text{dis}_{\mathcal{G}}(v) \cup \text{pa}_{\mathcal{G}}(\text{dis}_{\mathcal{G}}(v))) \setminus \{v\};$$

that is, the set of vertices that are connected to v by paths with an arrow at v and two arrowheads at each internal



(a) (b)

Figure 9: (a) An ADMG $\mathcal{G}$ which is neither maximal nor arid; (b) its maximal arid projection.

vertex. We can generalize this definition to any vertex that is childless within its own district.

Given a CADMG $\mathcal{G}(V, W)$, and a fixable $r \in V$, the fixing operation $\phi_r(\mathcal{G})$ yields a new CADMG $\widetilde{\mathcal{G}}(V \setminus \{r\}, W \cup \{r\})$ obtained from $\mathcal{G}(V, W)$ by removing all edges of the form $\to r$ and $\leftrightarrow r$, and keeping all other edges. Given a kernel $q_V(x_V \mid x_W)$ associated with a CADMG $\mathcal{G}(V, W)$, and a fixable $r \in V$, the fixing operation $\phi_r(q_V; \mathcal{G})$ yields a new kernel

$$\widetilde{q}_{V\setminus\{r\}}(x_{V\setminus\{r\}} \mid x_W, x_r) \equiv \frac{q_V(x_V \mid x_W)}{q_V(x_r \mid x_{\mathrm{mb}_{\mathcal{G}}(r)})}.$$

A result in Richardson et al. [2017] allows us to unambiguously define

$$\phi_R(\mathcal{G}) \equiv \phi_{r_k}(\ldots \phi_{r_2}(\phi_{r_1}(\mathcal{G}))\ldots),$$

and similarly the kernel $\phi_R(p; \mathcal{G})$ for distributions that are nested Markov with respect to $\mathcal{G}$ (defined below). Consequently, we just use sets to index fixings from now on.

If a fixing sequence exists for a set $R \subseteq V$ in $\mathcal{G}(V, W)$, we say $V \setminus R$ is a *reachable set*. Such a set is called *intrinsic* if the vertices in $V \setminus R$ are bidirected-connected (so that $\phi_{V\setminus R}(\mathcal{G})$ has only a single district); this definition is equivalent to the definition in the main paper. We denote the collections of reachable and intrinsic sets in $\mathcal{G}$ respectively by $\mathcal{R}(\mathcal{G})$ and $\mathcal{I}(\mathcal{G})$.

For a CADMG $\mathcal{G}(V, W)$, a (reachable) subset $C \subseteq V$ is called a *reachable closure* for $S \subseteq C$ if the set of fixable vertices in $\phi_{V\setminus C}(\mathcal{G})$ is a subset of S. Every set S in $\mathcal{G}$ has a unique reachable closure, which we denote $\langle S \rangle_{\mathcal{G}}$ [Shpitser et al., 2018]. Note that this set is generally a subset of what we earlier called the 'closure'.

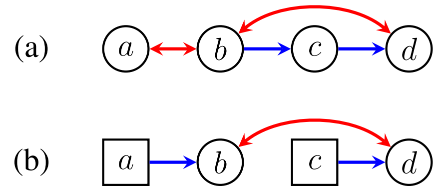


Figure 10: (a) An ADMG $\mathcal{G}$ that is not ancestral; (b) a CADMG obtained from $\mathcal{G}$ in (a) by fixing a and c.

C.2 NESTED MARKOV MODEL

We are now ready to define the nested Markov model $\mathcal{M}_n(\mathcal{G})$. Given an ADMG $\mathcal{G}$, we say that a distribution p obeys the *nested Markov property* with respect to $\mathcal{G}$ if for any reachable set R, we have that $\phi_{V\setminus R}(p; \mathcal{G})$ factorizes into kernels as

$$\phi_{V\setminus R}(p; \mathcal{G}) = \prod_{D \in \mathcal{D}(\phi_{V\setminus R}(\mathcal{G}))} \phi_{V\setminus D}(p; \mathcal{G}).$$

In other words, for any reachable graph, the associated kernel factorizes into a product of the districts in that graph conditional on the parents of those districts.

Note that this also means that $\phi_{V\setminus R}(p; \mathcal{G})$ will be Markov with respect to the CADMG $\phi_{V\setminus R}(\mathcal{G})$ for each reachable set R; see Richardson [2003] for more details on this.

Example C.1. Consider the ADMG in Figure 10(a). The vertices a, c and d all satisfy the condition of being fixable, but b does not since d is both a descendant of, and in the same district as, b. The CADMG $\mathcal{G}(\{b, d\}, \{a, c\})$ obtained after fixing a and c is shown in Figure 10(b). Notice that fixing c removes the edge $b \to c$, but that the edge $c \to d$ is preserved. Applying m-separation to the graph shown in Figure 10(b), yields

$$X_d \perp\!\!\!\perp X_a \mid X_c \quad [\phi_{ac}(p(x_{abcd}); \mathcal{G})].$$

In addition, one can see easily that if an edge $a \to d$ had been present in the original graph, then we would not have obtained this m-separation.

C.3 DENSELY CONNECTED VERTICES

Here we give a couple of slightly more detailed examples than in the main text.

Example C.2. The vertices a and c in Figure 1(b) are 'densely connected', because they cannot be separated by any combination of conditioning or fixing, except by fixing c (which just amounts to marginalizing it from the graph). Separately, for 'gadget' graph in Figure 2(b) the vertices c and d are also 'densely connected'. Naturally, any pair of vertices joined by an edge is also densely connected.

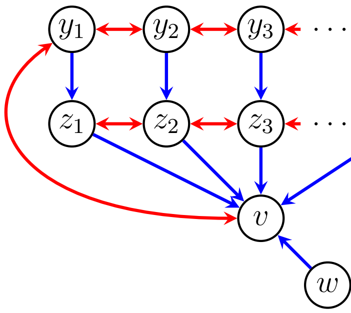


Figure 11: An ADMG for which a search for a set to satisfy Proposition 5.3 is computationally difficult.

D ALGORITHMS

D.1 SPANNING TREE

Given a set C and its subset of childless nodes B (in our case this will be either $\{v\}$ or $\{v, w\}$), pick a topological order on the vertices that places all elements of B at the end. Then, the last vertex before B must be a parent of some element of B; pick the largest such element under the topological order.

We then move backwards in the topological order, and each time a vertex has more than one child, we join it to the vertex which has the shortest path to an element of B; if there is a tie, then we pick the largest element under the topological order. This ensures that each vertex is joined to B by the shortest possible directed path.

D.2 DIFFICULT GRAPHS

Consider the graph shown in Figure 11. This can clearly be reduced to the graph $w \to v$, but the application of Proposition 5.3 is computationally difficult. Note that no subset will work apart from $\{z_1, \ldots, z_k\}$, and there are $3^k - 1$ possible sets to choose.

Algorithm 1 (with complexity proven to be $O(|V|)$) can be applied instead and will immediately return the graph $w \to v$.

References

T. S. Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavial Journal of Statistics*, 30(1): 145–157, 2003.

T. S. Richardson, R. J. Evans, J. M. Robins, and I. Shpitser. Nested Markov properties for acyclic directed mixed graphs. arXiv:1701.06686, 2017.

I. Shpitser, R. J. Evans, and T. S. Richardson. Acyclic linear SEMs obey the nested Markov property. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI-18)*, pages 735–745, 2018.