Acyclic Directed Mixed Graphs

Matrix Algebra, Statistical Models, Confounder Selection

Qingyuan Zhao

Statistical Laboratory, University of Cambridge

September 26, 2024 @ TU Munich, Germany

Acknowledgement

- Engineering and Physical Sciences Research Council (EPSRC grant EP/V049968/1).

Introduction

Acyclic directed mixed graphs (ADMGs)

- ADMGs have directed edges ($\longrightarrow$), bidirected edges ($\longleftrightarrow$), and no directed cycles.
- They were first used by Sewall Wright a century ago for genetics. Stayed popular in economics (e.g. instrumental variable methods) and social science (e.g. LISREL).
- They play a critical role in modern theory for causal modeling and identification (work by Pearl, Verma, Spirtes, Richardson, Tian, Shpitser, Robins, ...).

This talk

- Share some useful things I found while learning this theory.
- Main message:

Use **ADMGs**, not **DAGs** (for skeptical causal reasoning).

Outline

1. **Matrix algebra**: Understand and describe the "semantics" of ADMGs.
   - Qingyuan Zhao (2024a). *A Matrix Algebra for Graphical Statistical Models*. arXiv: 2407.15744 [math, stat].
2. **Statistical models**: Discuss various interpretations of ADMGs and highlight one of them.
   - Qingyuan Zhao (2024b). *On Statistical Models Associated with Acyclic Directed Mixed Graphs*. (working draft available upon request).
3. **Confounder selection**: A new interactive algorithm via iterative graph expansion.
   - F. Richard Guo and Qingyuan Zhao (2023). *Confounder Selection via Iterative Graph Expansion*. arXiv: 2309.06053 [math, stat].

Outline

Matrix algebra

Statistical models

Confounder selection

Outline

Matrix algebra

Statistical models

Confounder selection

Where it starts

Folklore in the community

Many results for nonparametric graphical models have their origins in (Gaussian) linear SEMs.

- Examples: (global) identifiability (Drton 2018); nested Markov property; proximal causal inference; sufficient and efficient adjustment sets.
- This is also explored in Spirtes et al. (1993) and also Judea Pearl (May 2013). "Linear Models: A Useful "Microscope" for Causal Analysis". In: *Journal of Causal Inference* 1.1, pp. 155–170.

This work

- I realized 2 years ago that the single most useful observation in linear SEMs is

$$[(I - A)^{-1}]_{jk} = [I + A + A^2 + \dots]_{jk} = \delta_{jk} + \# \{\text{directed walks from } j \text{ to } k\},$$

  where A is the adjacency matrix of a directed graph.
- Matrix abstractly: **Matrix multiplication is just composition of relations**.
- Goal: Develop a matrix algebra for ADMGs using the matrix algebra for linear SEMs.

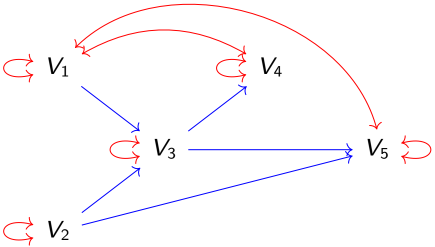
Gaussian linear systems

- Real-valued random vector $V = (V_1, \dots, V_d)$.
- $\beta, \Lambda \in \mathbb{R}^{d \times d}$, Λ is positive semi-definite.
- V follows a **Gaussian linear system** if $V = \beta^T + E$ where $E \sim \mathrm{N}(0, \Lambda)$.
- If β is non-singular, $V = (I - \beta)^{-T} E \sim \mathrm{N}(0, \Sigma)$, where $\Sigma = (I - \beta)^{-T} \Lambda (I - \beta)^{-1}$.
- **Marginalization** of Gaussian variables: take subvectors of mean and submatrices of covariance matrix.
- **Unconditional independence**: $V_{\mathcal{J}} \perp\!\!\!\perp V_{\mathcal{K}} \iff \Sigma_{\mathcal{J},\mathcal{K}} = 0$.
- **Conditional independence**: $V_j \perp\!\!\!\perp V_k \mid V_{[d] \setminus \{j,k\}} \iff (\Sigma^{-1})_{jk} = 0$.

Basic edge matrices

- $W[V \longrightarrow V]$ with (j,k)-entry given by $W[V_j \longrightarrow V_k] = \{V_j \longrightarrow V_k\}$ or $\emptyset$.
- Similarly for $W[V \longleftrightarrow V]$.

Example

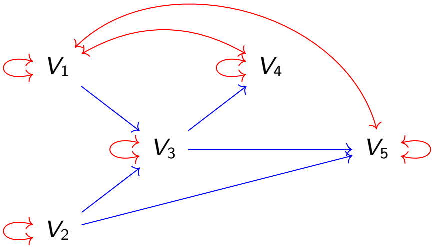


$$W[V \longrightarrow V] = \begin{pmatrix} \emptyset & \emptyset & \{V_1 \longrightarrow V_3\} & \emptyset & \emptyset \\ \emptyset & \emptyset & \{V_2 \longrightarrow V_3\} & \emptyset & \{V_2 \longrightarrow V_5\} \\ \emptyset & \emptyset & \emptyset & \{V_3 \longrightarrow V_4\} & \{V_2 \longrightarrow V_5\} \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \end{pmatrix}$$

Basic edge matrices

- $W[V \longrightarrow V]$ with (j,k)-entry given by $W[V_j \longrightarrow V_k] = \{V_j \longrightarrow V_k\}$ or $\emptyset$.
- Similarly for $W[V \longleftrightarrow V]$.

Example



$$W[V \longleftrightarrow V] = \begin{pmatrix} \{V_1 \longleftrightarrow V_1\} & \emptyset & \emptyset & \{V_1 \longleftrightarrow V_4\} & \{V_1 \longleftrightarrow V_5\} \\ \emptyset & \{V_2 \longleftrightarrow V_2\} & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \{V_3 \longrightarrow V_3\} & \emptyset & \emptyset \\ \{V_4 \longleftrightarrow V_1\} & \emptyset & \emptyset & \{V_4 \longleftrightarrow V_4\} & \emptyset \\ \{V_5 \longleftrightarrow V_1\} & \emptyset & \emptyset & \{V_5 \longleftrightarrow V_4\} & \{V_5 \longleftrightarrow V_5\} \end{pmatrix}.$$

Matrix algebra for walks on directed mixed graphs

- Binary operations on (sets of) walks: set union ($+$), concatenation ($\cdot$), transpose (^{T}).

Examples

$$\{V_2 \longrightarrow V_5\} + \{V_2 \longrightarrow V_3 \longrightarrow V_5\} = \{V_2 \longrightarrow V_5,\ V_2 \longrightarrow V_3 \longrightarrow V_5\}.$$

$$\{V_2 \longleftrightarrow V_2\} \cdot \{V_2 \longrightarrow V_5,\ V_2 \longrightarrow V_3 \longrightarrow V_5\} = \{V_2 \longleftrightarrow V_2 \longrightarrow V_5,\ V_2 \longleftrightarrow V_2 \longrightarrow V_3 \longrightarrow V_5\}.$$

$$\{V_2 \longrightarrow V_5,\ V_2 \longrightarrow V_3 \longrightarrow V_5\}^T = \{V_5 \longleftarrow V_2,\ V_5 \longleftarrow V_3 \longleftarrow V_2\}.$$

- Matrix addition, multiplication, and transpose are defined accordingly:

$$(W + W')[V_j, V_k] = W[V_j, V_k] + W'[V_j, V_k],$$

$$(W \cdot W')[V_j, V_k] = \sum_{V_l \in V} W[V_j, V_l] \cdot W'[V_l, V_k],$$

$$(W^T)[V_j, V_k] = (W[V_k, V_j])^T = \{w^T : w \in W[V_k, V_j]\},$$

- Let I denote the identity matrix for multiplication (diagonal of "empty walks").

Remarks

This matrix algebra is a “dioid” in the terminology of Gondran and Minoux (2008):

1. $+$ is a commutative monoid (associative with identity element $\emptyset$);
2. $\cdot$ is a monoid (associative with identity element Id);
3. $\cdot$ is distributive with respect to $+$;
4. The pre-order defined by $+$ ($W \preceq W'$ if and only if $W' = W + W''$ for some W'') is anti-symmetric: $W \preceq W'$ and $W' \preceq W$ imply that $W = W'$.

Dioids vs. rings

- In rings, $+$ is a **commutative group** (e.g. square matrices).
- In dioids, $+$ is an **ordered monoid** (e.g. nonnegative square matrices).

- Dioids are extensively studied in formal languages/algebraic combinatorics (I am a novice).
- What’s likely “new” are probabilistic concepts such as marginalization and conditioning.

"Words" in ADMGs—Special matrices

Notation

- Squiggly line $\rightsquigarrow$ means "arcs" (no collider).
- Endpoint arrowheads are important

Arcs

- **(Right-)Directed walks**: $W[V \rightsquigarrow V] = \sum_{q=1}^{\infty} (W[V \longrightarrow V])^q$.
- **Left-directed walks**: $W[V \leftsquigarrow V] = (W[V \rightsquigarrow V])^T$.
- **Treks (t-connected, exactly one bidirected edge)**:

$$W[V \overset{\text{t}}{\leftrightsquigarrow} V] = (I + W[V \leftsquigarrow V]) \cdot W[V \longleftrightarrow V] \cdot (I + W[V \rightsquigarrow V]).$$

- **d-connected** (no bidirected edge):

$$W[V \overset{\text{d}}{\leftrightsquigarrow} V] = W[V \leftsquigarrow V] + W[V \rightsquigarrow V] + W[V \leftsquigarrow V \rightsquigarrow V].$$

- **Arc (m-connected)**: $W[V \leftrightsquigarrow V] = W[V \overset{\text{t}}{\leftrightsquigarrow} V] + W[V \overset{\text{d}}{\leftrightsquigarrow} V]$.

Trek and covariance

$$W[V \overset{t}{\leftrightsquigarrow} V] = (I + W[V \leftsquigarrow V]) \cdot W[V \longleftrightarrow V] \cdot (I + W[V \rightsquigarrow V]).$$

- The name "trek" is due to Spirtes et al. (1993).
- This comes up due to the **trek rule**: recall $\Sigma = (I - \beta)^{-T} \Lambda (I - \beta)^{-1}$.
- Formally, let σ be the **weight function** generated by

$$\beta = \sigma(W[V \longrightarrow V]) \quad \text{and} \quad \Lambda = \sigma(W[V \longleftrightarrow V]).$$

- Example: $\sigma(\{V_1 \longleftrightarrow V_4, V_1 \longleftrightarrow V_1 \longrightarrow V_3 \longrightarrow V_4\}) = \lambda_{14} + \lambda_{11}\beta_{13}\beta_{34}$.

Theorem (Trek rule)

For any $V_j, V_k \in V$ in a Gaussian linear system, if β is non-singular, then

$$\textbf{not } V_j \overset{t}{\leftrightsquigarrow} V_k \textbf{ in } \mathrm{G} \Longrightarrow V_j \perp\!\!\!\perp V_k \textbf{ under } \mathrm{P}\,.$$

If β is stable, then

$$\mathrm{Cov}_{\mathrm{P}}(V) = \sigma(W[V \overset{t}{\leftrightsquigarrow} V \textbf{ in } \mathrm{G}]).$$

Marginalization

- This is essential but not emphasized enough in the literature.
- Partition the linear system by $\tilde{V} \subset V$ and $U = V \setminus \tilde{V}$:

$$\tilde{V} = \beta_{\tilde{V}\tilde{V}}^T \tilde{V} + \beta_{U\tilde{V}}^T U + E_{\tilde{V}}, \quad U = \beta_{\tilde{V}U}^T \tilde{V} + \beta_{UU}^T U + E_U.$$

- By eliminating U, we obtain

$$\begin{aligned}\tilde{V} &= \left\{\beta_{\tilde{V}\tilde{V}}^T + \beta_{U\tilde{V}}^T (\mathrm{Id} - \beta_{UU})^{-T} \beta_{\tilde{V}U}^T\right\} \tilde{V} + \left\{\beta_{U\tilde{V}}^T (\mathrm{Id} - \beta_{UU})^{-T} E_U + E_{\tilde{V}}\right\} \\ &= \sigma[\tilde{V} \leftsquigarrow \tilde{V} \mid \tilde{V}]\, \tilde{V} + \left\{\sigma[\tilde{V} \leftsquigarrow U \mid \tilde{V}]\, E_U + E_{\tilde{V}}\right\},\end{aligned}$$

where $[\cdots \mid \tilde{V}]$ collects all such walks without non-endpoints in $\tilde{V}$.

Theorem (Marginalization of linear systems)

If all principal submatrices of β are stable, $\tilde{V}$ is a Gausslian linear system wrt $\tilde{\mathsf{G}} = \mathrm{margin}_{\tilde{V}}(\mathsf{G})$:

$$\tilde{\beta} = \tilde{\sigma}(W[\tilde{V} \longrightarrow \tilde{V} \textbf{ in } \tilde{\mathsf{G}}]) = \sigma(W[\tilde{V} \rightsquigarrow \tilde{V} \mid \tilde{V} \textbf{ in } \mathsf{G}]),$$

$$\tilde{\Lambda} = \tilde{\sigma}(W[\tilde{V} \longleftrightarrow \tilde{V} \textbf{ in } \tilde{\mathsf{G}}]) = \sigma(W[\tilde{V} \overset{\mathrm{t}}{\leftrightsquigarrow} \tilde{V} \mid \tilde{V} \textbf{ in } \mathsf{G}]).$$

Conditional independences

- We have $\Sigma^{-1} = (\mathrm{Id} - \beta)\Lambda^{-1}(\mathrm{Id} - \beta)^T$, so

$$(\Sigma^{-1})_{jk} = \sum_{V_l, V_m \in V} (\delta_{jl} - \beta_{jl})(\Lambda^{-1})_{lm}(\delta_{km} - \beta_{km})$$

- A sufficient condition for $(\Sigma^{-1})_{jk} = 0$ is when all RHS summands vanish.
- Key observation: $\textbf{not } V_l \longleftrightarrow * \longleftrightarrow V_m \textbf{ in } \mathrm{G} \implies (\Lambda^{-1})_{lm} = 0$.

Proposition

If β is non-singular and Λ is positive definite, then for any $V_j \neq V_k$,

$$\textbf{not } V_j \longrightarrow * \longleftarrow V_k \textbf{ in } \mathrm{G} \implies V_j \perp\!\!\!\perp V_k \mid V \setminus \{V_j, V_k\} \textbf{ under } \mathrm{P}\,.$$

Challenge

How can this be extended when just some variables are conditioned on?

Blocking

- We say a walk is (**ancestrally**) **blocked** by $L \subseteq V$ if it contains a collider $V_m \notin L$ ($V_m \notin L$ **and** $V_m \not\rightsquigarrow L$) or a non-colliding non-enpoint $V_m \in L$.
  - The Bayes ball algorithm (Shachter 1998).
- Let $W[V_j \rightsquigarrow * \leftsquigarrow V_k \mid L$ **in** $\mathrm{G}]$ denote all walks from V_j to V_k not blocked by L:

$$W[V \rightsquigarrow * \leftsquigarrow V \mid L] = W[V \leftrightsquigarrow V \mid L] + W[V \rightsquigarrow L \mid L] \cdot \left\{ \mathrm{Id} + \sum_{q=1}^{\infty} (W[L \leftrightsquigarrow L \mid L])^q \right\} \cdot W[L \leftsquigarrow V \mid L].$$

- Wildcard character $*$ means arbitrary number of colliders.

Example

$$\begin{array}{ccccc} V_1 & \longrightarrow & V_3 & \longleftarrow & V_2 \\ & & \big\downarrow & & \\ & & V_4 & & \end{array}$$

- $V_1 \longrightarrow V_3 \longleftarrow V_2$ is blocked (but not ancestrally blocked) by V_4.
- Nevertheless, $V_1 \rightsquigarrow * \leftsquigarrow V_2 \mid V_4$ because $V_1 \longrightarrow V_3 \longrightarrow V_4 \longleftarrow V_3 \longleftarrow V_2$.

Presevation of "words" by marginalization

Fundamental Lemma 1

For any directed mixed graph G (possibly cyclic and any bidirected loops) and $L \subseteq \tilde{V} \subseteq V$,

$$\mathrm{margin}_{\tilde{V}}\left(W\left[\tilde{V} \left\{ \begin{matrix} \rightsquigarrow \\ \leftsquigarrow \\ \overset{\mathrm{t}}{\leftrightsquigarrow} \\ \overset{\mathrm{t}}{\leftrightsquigarrow} * \overset{\mathrm{t}}{\leftrightsquigarrow} \end{matrix} \right\} \tilde{V} \mid L \textbf{ in } \mathrm{G} \right] \right) = W\left[\tilde{V} \left\{ \begin{matrix} \rightsquigarrow \\ \leftsquigarrow \\ \overset{\mathrm{t}}{\leftrightsquigarrow} \\ \overset{\mathrm{t}}{\leftrightsquigarrow} * \overset{\mathrm{t}}{\leftrightsquigarrow} \end{matrix} \right\} \tilde{V} \mid L \textbf{ in } \mathrm{margin}_{\tilde{V}}(\mathrm{G}) \right],$$

Fundamental Lemma 2

Consider any disjoint $\{V_j\}, \{V_k\}, L \subset V$. If G is **canonical** (contains all bidirected loops), then

$$\text{(i) } V_j \overset{\mathrm{t}}{\leftrightsquigarrow} * \overset{\mathrm{t}}{\leftrightsquigarrow} V_k \mid L \iff \text{(ii) } V_j \rightsquigarrow * \leftsquigarrow V_k \mid L \iff \text{(iii) } P[j \rightsquigarrow * \leftsquigarrow k \mid_a L] \neq \emptyset.$$

Furthermore, if G is **canonically directed** (no other bidirected edges), then

$$\text{(i), (ii), (iii)} \iff \text{(iv) } V_j \overset{\mathrm{d}}{\rightsquigarrow} * \overset{\mathrm{d}}{\leftsquigarrow} V_k \mid L \iff \text{(v) } P[V_j \overset{\mathrm{d}}{\rightsquigarrow} * \overset{\mathrm{d}}{\leftsquigarrow} V_k \mid_a L] \neq \emptyset.$$

Here $P[\cdots \mid_a L]$ collects all such paths not ancestrally blocked by L.

Graph separation and conditional independence

Proposition (from 3 slides ago)

If β is non-singular and Λ is positive definite, then for any $V_j \neq V_k$,

$$\textbf{not } V_j \longleftrightarrow * \longleftrightarrow V_k \textbf{ in } \mathrm{G} \Longrightarrow V_j \perp\!\!\!\perp V_k \mid V \setminus \{V_j, V_k\} \textbf{ under } \mathrm{P}\,.$$

Theorem (m-separation implies CI in Gaussian linear systems)

If β is non-singular and Λ is positive definite, then for all disjoint $J, K, L \subseteq V$, we have

$$\textbf{not } J \leftrightsquigarrow * \leftrightsquigarrow K \mid L \textbf{ in } \mathrm{G} \Longrightarrow J \perp\!\!\!\perp K \mid L \textbf{ under } \mathrm{P}\,.$$

Proof. Because conditional independence is compositional for Gaussian variables, it suffices to consider $J = V_j$ and $K = V_k$. Let $\tilde{\mathrm{G}} = \mathrm{margin}_{\{V_j, V_k\} \cup L}(\mathrm{G})$, then

$$\begin{aligned} \textbf{not } V_j \leftrightsquigarrow * \leftrightsquigarrow V_k \mid L \textbf{ in } \mathrm{G} &\Longleftrightarrow \textbf{not } V_j \longleftrightarrow * \longleftrightarrow V_k \textbf{ in } \tilde{\mathrm{G}} \\ &\Longrightarrow V_j \perp\!\!\!\perp V_k \mid L \textbf{ under } \mathrm{P}\,. \end{aligned}$$

Remarks

- This result was obtained (independently) by Koster (1996) and Spirtes (1995) in (cyclic) DAGs. A full proof using directed mixed graphs is given in Koster (1999).
- What's different: we **do not just prove** d/m-separation implies conditional independence, **but derive** these graphical criteria from scratch.
- The paper further shows how to **derive** the **generalized backdoor criterion** in Shpitser et al. (2010) and, gives a **visual proof** of the **moralization/augmentation criterion**.

Other "words" in (canonical) ADMGs

1. $\longrightarrow * \longleftrightarrow$: Markov blanket/background (Richardson et al. 2023).
2. $\overset{d}{\rightsquigarrow} * \overset{d}{\leftrightsquigarrow}$: δ-connection (Didelez 2008).
3. $\rightsquigarrow * \leftrightsquigarrow$: μ-connection (Mogensen and Hansen 2020).
4. $\leftrightsquigarrow$: confounding arc (Guo and Zhao 2023).
5. $\leftrightsquigarrow * \leftrightsquigarrow$: confounding path/symmetric backdoor (Guo and Zhao 2023).

Outline

Introduction

Some interpretations of canonical ADMGs

1. **Latent DAG**: ADMG represents a unspecified DAG with latent variables.
   - Implicit in Pearl's writing.
   - See also Richardson et al. (2023, Sec. 4.1).
2. **Nonparametric SEM**: ADMG represents a nonparametric system with correlated noise.
   - Used by Pearl and his followers, with some ambiguity about the meaning of bidirected edges.
   - Sometimes called the **semi-Markovian** model.
3. **Specific expansion of bidirected edges**.
4. **Global Markov**.
5. **Nested Markov**.

Challenges

- These interpretations are generally not the same.
- Complicated inequality constraints in the latent variable interpretations.
- Statistical model vs causal model.

Main thesis

By default, we should use the nonparametric system of equations (**E model** below).

Definition

The E model collects all distributions P of V such that

$$V_j = f_j(V_{\mathrm{pa}(j)}, E_j), \quad \text{for all } V_j \in V.$$

- $f_1, \dots, f_d$ are some functions. $\mathrm{pa}(j) = \{k : V_k \longrightarrow V_j \textbf{ in } \mathsf{G}\}$.
- **The distribution of $E = (E_1, \dots, E_d) \in [0,1]^d$ is Markov wrt bidirected part of** G**:**

$$V_{\mathcal{J}} \not\leftrightarrow V_{\mathcal{K}} \textbf{ in } \mathsf{G} \Longrightarrow E_{\mathcal{J}} \perp\!\!\!\perp E_{\mathcal{K}}.$$

- This corresponds to a **noise expansion** of the ADMG.

Why should this be the default?

Two common ways to define "natural" mathematical concepts: **Equivalence** and **Completion**.

Equivalence fails for general ADMGs

The following implications/equivalences hold for **any (canonical) ADMG**.

$$\begin{array}{ccccc}
\text{Pairwise expansion (PE)} & & & & \\
\Downarrow & & & & \\
\text{Clique expansion (CE)} & & & & \\
\Downarrow & & & & \\
\text{Noise expansion (NE)} & \Longleftrightarrow & \text{Nonparametric equation (E)} & & \\
\Downarrow & & & & \\
\text{Nested Markov (NM)} & & & & \\
\Downarrow & & & & \\
\text{Ordered local Markov (LM)} & \Longleftrightarrow & \text{Global Markov (GM)} & \Longleftrightarrow & \text{Augmentation (A)} \\
\Downarrow & & & & \\
\text{Unconditional Markov (UM)} & & & &
\end{array}$$

The reverse implications don't hold in general.

Equivalence works for simpler subclasses

DAGs

$$\begin{array}{l} \text{PE} \Leftrightarrow \text{CE} \Leftrightarrow \text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{Factorization (F)} \Leftrightarrow \text{NM} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \\ \Downarrow \\ \text{UM} \end{array}$$

Bidirected graphs

$$\begin{array}{l} \text{PE} \\ \Downarrow \\ \text{CE} \\ \Downarrow \\ \text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{NM} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \Leftrightarrow \text{UM} \end{array}$$

A definition of completeness

- An "interpretation" of a ADMG is a collection $\mathbb{P}(\mathsf{G})$ of probability distributions.
- Denote $\text{expand}_{V'}(\mathsf{G}) = \{\mathsf{G}' \text{ is ADMG with vertex set } V' : \text{margin}_V(\mathsf{G}') = \mathsf{G}\}$.
- For each vertex set V, let $\mathbb{G}_0(V)$ be a subclass of ADMGs.

Definition

We say a collection of models $\mathbb{P}(\mathsf{G})$ for different ADMGs G is **complete** (wrt $\mathbb{G}_0$) if

$$\mathbb{P}(\mathsf{G}) = \bigcup_{V' \supset V} \bigcup_{\mathsf{G}'} \text{margin}_V(\mathbb{P}(\mathsf{G}')),$$

where the second union is over $\mathsf{G}' \in \text{expand}_{V'}(\mathsf{G}) \cap \mathbb{G}_0(V')$.

Why using the E model?

- Denote the set of **exogenous** vertices in G as $E = \{V_j \in V : V_k \not\rightarrow V_j \text{ for all } V_k \in V\}$.
- An ADMG is called **unconfounded** if

$$V_j \longleftrightarrow V_k \textbf{ in } \mathrm{G} \Longrightarrow V_j, V_k \in E, \quad \text{for all } V_j, V_k \in V, V_j \neq V_k.$$

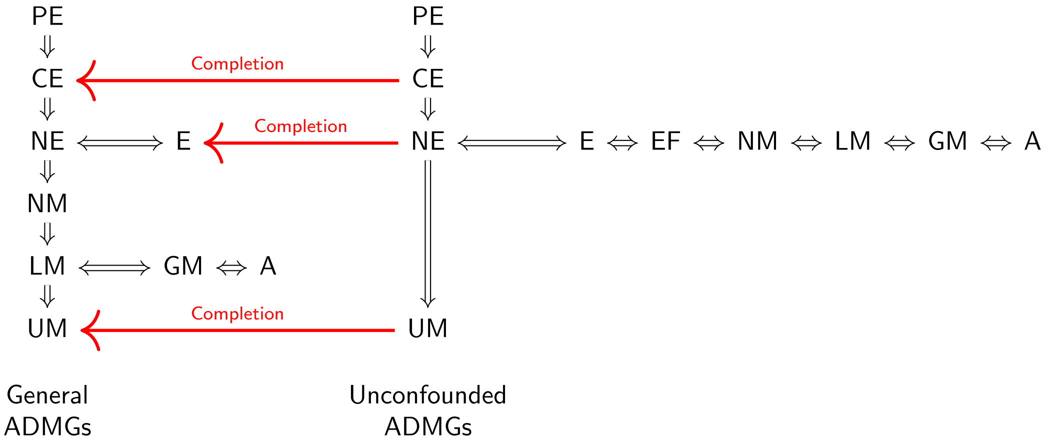
Unconfounded ADMGs generalize DAGs and bidirected graphs

$$\begin{array}{l} \text{PE} \\ \Downarrow \\ \text{CE} \\ \Downarrow \\ \text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{Exogenous Factorization (EF)} \Leftrightarrow \text{NM} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \\ \Downarrow \\ \text{UM} \end{array}$$

Theorem

When $\mathbb{G}_0(V)$ is all unconfounded ADMGs, only CE, E, NE, and UM models are complete.

A visualization of Equivalence + Completion



- So the latent DAG/CE interpretation does not consider bidirected/unconfounded graphs as interesting objects.

Causal model associated with ADMGs

- The **potential outcome schedule** is the collection of all potential outcomes:

$$V(\cdot) = (V_j(v_{\mathcal{I}}) : j \in [d], \mathcal{I} \subseteq [d], v_{\mathcal{I}} \in \mathbb{V}_{\mathcal{I}}).$$

Definition (Causal Markov model)

We say a distribution P of the potential outcome schedule of V is **causal Markov** wrt G if

1. The potential outcomes are **consistent**:

$$V_j(v_{\mathcal{I}}) = V_j(v_{\mathrm{pa}(j)\cap\mathcal{I}}, V_{\mathrm{pa}(j)\setminus\mathcal{I}}(v_{\mathcal{I}})), \text{ for all } j \in [d], \mathcal{I} \subseteq [d], v \in \mathbb{V}.$$

2. The distribution of **basic potential outcomes are Markov** wrt bidirected part of G:

$$V_{\mathcal{J}} \not\leftrightarrow V_{\mathcal{K}} \textbf{ in } \mathrm{G} \Longrightarrow V_{\mathcal{J}}(v) \perp\!\!\!\perp V_{\mathcal{K}}(v) \textbf{ under } \mathrm{P} \text{ for all } v \in \mathbb{V}.$$

- This definition does not rely on structural equations.
- $\longrightarrow$ means direct causal infeluence and $\longleftrightarrow$ means exogenous correlation.

Causal identification and Nested Markov (NM) property

- V_j is called **fixable** if there exists no V_k such that $V_j \rightsquigarrow V_k$ and $V_j \longleftrightarrow * \longleftrightarrow V_k$.
- NM requires that if V_j is fixable, the next distribution is global Markov wrt $\mathsf{G}_{V_{-j}}$:

$$(\mathrm{fix}_{V_j = v_j}(\mathrm{p}))(v_{-j}) = \frac{\mathrm{p}(v)}{\mathrm{p}(v_j \mid v_{\mathrm{mbg}_{\mathsf{G}}(j)})},$$

  and this needs to hold recursively. (This is the basic step in the do-calculus/ID algorithm.)
- Richardson et al. (2023, Sec. 4.1) give a proof of CE $\Rightarrow$ NM in general ADMGs.

A more intuitive proof sketch for E/NE $\Rightarrow$ NM

- Consider $\mathrm{P}_V \in$ E model. There exists causal Markov P on $V(\cdot)$ s.t. $\mathrm{margin}_V(\mathrm{P}) = \mathrm{P}_V$.
- It can be shown that if V_j is fixable, then

$$\mathrm{fix}_{V_j = v_j}(\mathrm{P}_V) = \mathrm{margin}_{V_{-j}(v_j)}(\mathrm{P}).$$

- A natural generalization of SWIGs (Richardson and Robins 2013) from DAGs to ADMGs shows that the RHS is global Markov wrt $\mathsf{G}_{V_{-j}}$.
- Repeatedly applying this shows that E/NE $\Rightarrow$ NM.

Use **ADMGs**, not **DAGs**

- DAG-based theory: ADMG represents a DAG with latent variables.
- ADMG-based theory: ADMG is a generalization of DAG.

ADMG-based theory is better because...

- No mysterious latent variables.
- More convincing in unconfounded graphs.
- Fewer (but still some) inequality constraints.
- By drawing ADMGs instead of DAGs, practitioners are encouraged to think about the **missing bidirected** and directed edges which really drive causal identification.

Outline

Two common heuristics

The conjunction heuristic/common cause principle)

Contrlling for all covariates "related" to both the treatment and the outcome.

- Very common in practice (Glymour et al. 2008) and methodological development (Koch et al. 2020; Shortreed and Ertefaie 2017).
- Well known that this may select too few.

The pre-treatment heuristic

Controlling for all covariates that precede the treatment temporally.

- Defended in Rubin (2009): "I cannot think of a credible real-life situation where I would intentionally allow substantially different observed distributions of a true covariate in the treatment and control groups."
- Counter-examples from graphical models: e.g. M-bias (Greenland et al. 1999).

Graphical approaches

Theorem (Back-door criterion (Pearl 1993, 2009))

Given a treatment X and an outcome Y, a set of covariates S controls for confounding if

1. *S contains no descendant of X (**not** $X \rightsquigarrow S$);*
2. *S blocks all back-door paths from X to Y (**not** $X \leftsquigarrow\rightsquigarrow * \leftsquigarrow\rightsquigarrow Y \mid S$).*

- Basically complete (Shpitser et al. 2010), but requires full structural knowledge.

Theorem (Disjunctive criterion (VanderWeele and Shpitser 2011))

Suppose the causal graph is faithful. If at least one subset of S controls for confounding, then $\{V_j \in S : V_j \rightsquigarrow X \text{ or } V_j \rightsquigarrow Y\}$ controls for confounding.

- Easy to use, but assumption is almost wishful.

Simplifying the back-door criterion

- Let $\tilde{G} = \text{margin}_{\tilde{V}}(G)$ and G^* be G without bidirected loops.
- Simplified marginalization of ADMGs:

$$V_j \longrightarrow V_k \textbf{ in } \tilde{G}^* \iff P[V_j \rightsquigarrow V_k \mid \tilde{V} \textbf{ in } G^*] \neq \emptyset,$$

$$V_j \longleftrightarrow V_k \textbf{ in } \tilde{G}^* \iff P[V_j \leftrightsquigarrow V_k \mid \tilde{V} \textbf{ in } G^*] \neq \emptyset. \text{ (confounding arcs)}$$

Definition

- $C \subseteq V \setminus \{A, B\}$ is an **adjustment set** for $A, B \in V$ if **not** $\{A, B\} \rightsquigarrow C$.
- C is **sufficient** if **not** $A \leftrightsquigarrow * \leftrightsquigarrow B \mid C$.
- C is **minimal sufficient** if none of its proper subsets is still sufficient.

Proposition (Symmetric back-door criterion)

- If S is a sufficient adjustment set for (X, Y), then S satisfies the back-door criterion.
- If an adjustment set S satisfies the back-door criterion and $X \rightsquigarrow Y$, then S is sufficient.

Main idea

- Confounder selection/blocking confounding paths is complicated because

$$\textbf{not } A \leftrightsquigarrow * \leftrightsquigarrow B \mid C \not\Rightarrow \textbf{not } A \leftrightsquigarrow * \leftrightsquigarrow B \mid \tilde{C} \text{ for } C \subset \tilde{C}.$$

- But observe that

$$\textbf{not } A \leftrightsquigarrow B \mid C \Rightarrow \textbf{not } A \leftrightsquigarrow B \mid \tilde{C} \text{ for } C \subset \tilde{C}.$$

- This motivates us to **block confounding arcs recursively**.

Definition

- An adjustment set C for A, B is called **primary** given another adjustment set S if

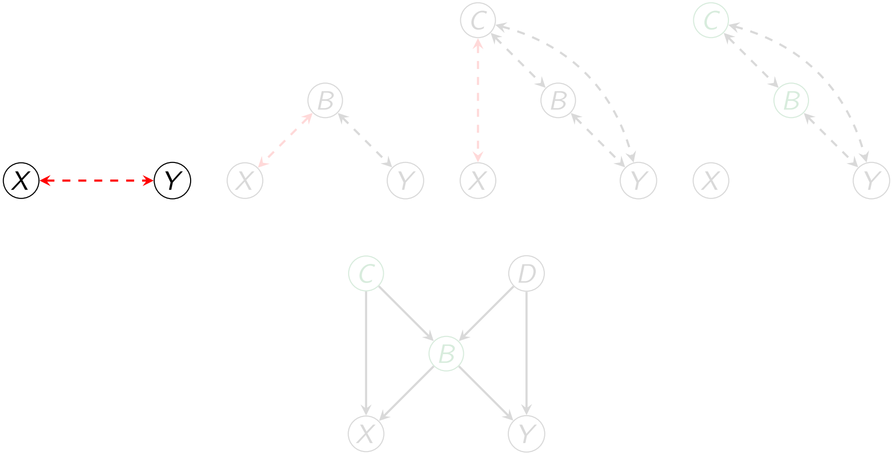
$$\textbf{not } A \leftrightsquigarrow B \mid S \cup C$$

- We say C is **minimal primary** if none of its proper subsets is primary.

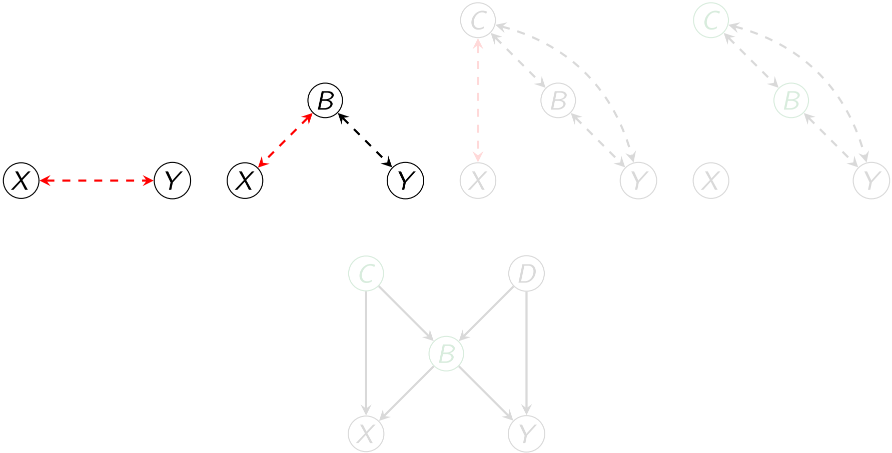
Iterative graph expansion (recursive version)

procedure $\textsc{ConfounderSelect}(X, Y)$
    $\mathcal{R} = \emptyset$
    procedure GraphExpand$(S, \mathcal{B}_y, \mathcal{B}_n)$
        $\bar{S} = S \cup X \cup Y$
        if $X \longleftrightarrow * \longleftrightarrow Y$ by edges in $\mathcal{B}_y$ **then**
            return
        else if not $X \longleftrightarrow * \longleftrightarrow Y$ by edges in $(\bar{S} \times \bar{S}) \setminus \mathcal{B}_n$ **then**
            $\mathcal{R} = \mathcal{R} \cup \{S\}$
            return
        end if
        $(A \leftrightarrow B) = \textsc{SelectEdge}(X, Y, S, \mathcal{B}_y, \mathcal{B}_n)$
        for C in $\textsc{FindPrimary}(A \leftrightarrow B, S)$ **do**
            GraphExpand$(S \cup C, \mathcal{B}_y, \mathcal{B}_n \cup \{A \leftrightarrow B\})$
        end for
        GraphExpand$(S, \mathcal{B}_y \cup \{A \leftrightarrow B\}, \mathcal{B}_n)$
    end procedure
    GraphExpand$(\emptyset, \emptyset, \emptyset)$
    return $\mathcal{R}$
end procedure

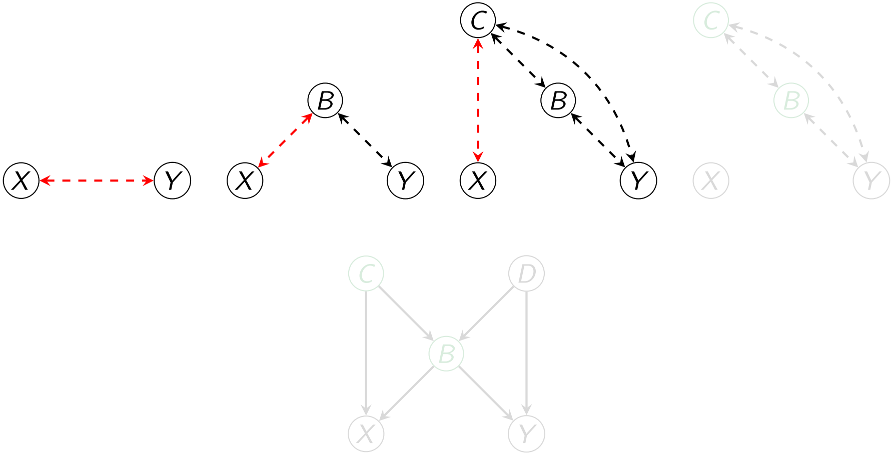
Illustration



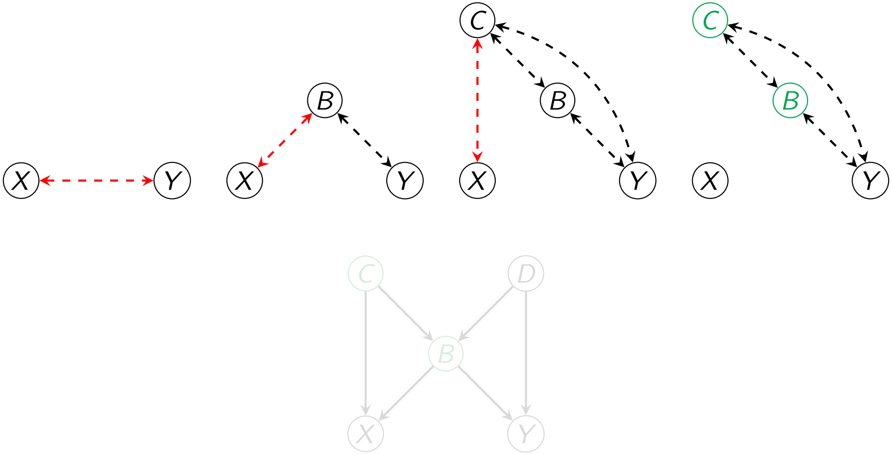
Illustration



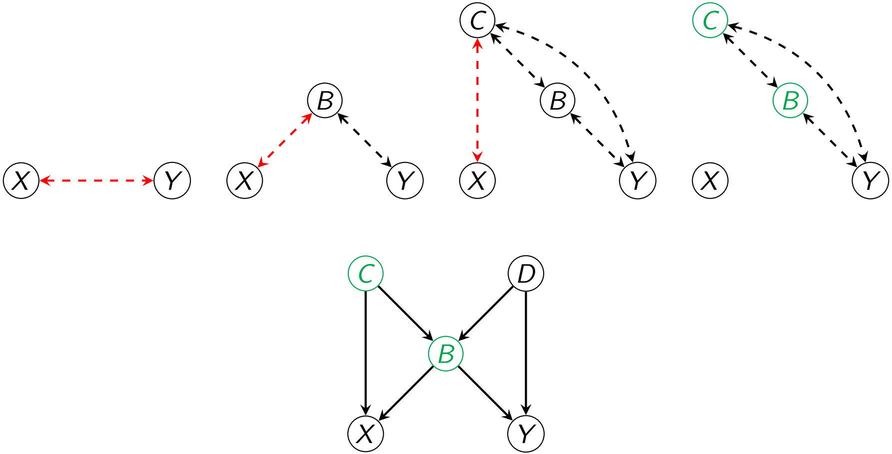
Illustration



Illustration



Illustration



Remarks

- This procedure is **sound** (primary $\Rightarrow$ sufficient) and **complete** (all minimal primary $\Rightarrow$ all minimal sufficient).
- Confounder selection **only requries thinking about bidirected edges ($\longleftrightarrow$)**.
- Primary adjustment sets can be found by a more basic routine.
- Try it: `https://ricguo.shinyapps.io/InteractiveConfSel/`

Conclusions

- Recap: **matrix algebra**, **statistical models**, **confounder selection** for ADMGs.
- Many interesting (possibly open) problems (email me at qyzhao@statslab.cam.ac.uk).
- Main take-away:

Use ADMGs, not DAGs.

References I

- Didelez, Vanessa (2008). "Graphical Models for Marked Point Processes Based on Local Independence". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.1, pp. 245–264.
- Drton, Mathias (Jan. 2018). "Algebraic Problems in Structural Equation Modeling". In: *The 50th Anniversary of Gröbner Bases*. Vol. 77. Mathematical Society of Japan, pp. 35–87.
- Glymour, M. Maria et al. (2008). "Methodological Challenges in Causal Research on Racial and Ethnic Patterns of Cognitive Trajectories: Measurement, Selection, and Bias". In: *Neuropsychology Review* 18.3, pp. 194–213.
- Gondran, Michel and Michel Minoux (2008). *Graphs, Dioids and Semirings*. Vol. 41. Operations Research/Computer Science Interfaces. Boston, MA: Springer US.
- Greenland, Sander et al. (1999). "Confounding and Collapsibility in Causal Inference". In: *Statistical Science* 14.1, pp. 29–46.
- Guo, F. Richard and Qingyuan Zhao (2023). *Confounder Selection via Iterative Graph Expansion*. arXiv: `2309.06053 [math, stat]`.
- Koch, Brandon et al. (2020). "Variable Selection and Estimation in Causal Inference Using Bayesian Spike and Slab Priors". In: *Statistical Methods in Medical Research* 29.9, pp. 2445–2469.

References II

- Koster, J. T. A. (Oct. 1996). "Markov Properties of Nonrecursive Causal Models". In: *The Annals of Statistics* 24.5, pp. 2148–2177.
- Koster, Jan T. A. (1999). "On the Validity of the Markov Interpretation of Path Diagrams of Gaussian Structural Equations Systems with Correlated Errors". In: *Scandinavian Journal of Statistics* 26.3, pp. 413–431.
- Mogensen, Søren Wengel and Niels Richard Hansen (Feb. 2020). "Markov Equivalence of Marginalized Local Independence Graphs". In: *The Annals of Statistics* 48.1, pp. 539–559.
- Pearl, Judea (1993). "Comment: Graphical Models, Causality and Intervention". In: *Statistical Science* 8.3, pp. 266–269.
- — (2009). *Causality*. 2nd ed. Cambridge: Cambridge University Press.
- — (May 2013). "Linear Models: A Useful "Microscope" for Causal Analysis". In: *Journal of Causal Inference* 1.1, pp. 155–170.
- Richardson, Thomas S and James M Robins (2013). *Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality*. Tech. rep. 128. Center for the Statistics and the Social Sciences, University of Washington Series.

References III

- Richardson, Thomas S. et al. (Feb. 2023). "Nested Markov Properties for Acyclic Directed Mixed Graphs". In: *The Annals of Statistics* 51.1, pp. 334–361.
- Rubin, Donald B. (2009). "Should Observational Studies Be Designed to Allow Lack of Balance in Covariate Distributions across Treatment Groups?" In: *Statistics in Medicine* 28.9, pp. 1420–1423.
- Shachter, Ross D. (July 1998). "Bayes-Ball: Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)". In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. UAI'98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 480–487.
- Shortreed, Susan M. and Ashkan Ertefaie (2017). "Outcome-Adaptive Lasso: Variable Selection for Causal Inference". In: *Biometrics. Journal of the International Biometric Society* 73.4, pp. 1111–1122.
- Shpitser, Ilya et al. (July 2010). "On the Validity of Covariate Adjustment for Estimating Causal Effects". In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. UAI'10. Arlington, Virginia, USA: AUAI Press, pp. 527–536.

References IV

- Spirtes, Peter (Aug. 1995). "Directed Cyclic Graphical Representations of Feedback Models". In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. UAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 491–498.
- Spirtes, Peter et al. (1993). *Causation, Prediction, and Search*. Ed. by J. Berger et al. Vol. 81. Lecture Notes in Statistics. New York, NY: Springer.
- VanderWeele, Tyler and Ilya Shpitser (2011). "A New Criterion for Confounder Selection". In: *Biometrics. Journal of the International Biometric Society* 67.4, pp. 1406–1413.
- Zhao, Qingyuan (2024a). *A Matrix Algebra for Graphical Statistical Models*. arXiv: 2407.15744 [math, stat].
- — (2024b). *On Statistical Models Associated with Acyclic Directed Mixed Graphs*. (working draft available upon request).