The algebraic geometry of Ornstein–Uhlenbeck processes in equilibrium

Tobias Boege

Department of Mathematics and Statistics
UiT The Arctic University of Norway

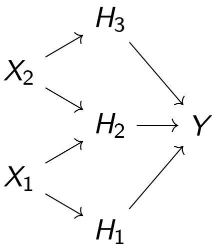
Oberseminar Kombinatorische Algebraische Geometrie,
Paderborn, 26 May 2026

Based on joint works **arXiv:2408.00583** and **arXiv:2510.04985** with Carlos Améndola, Mathias Drton, Benjamin Hollering, Sarah Lumpp, Pratik Misra, and Daniela Schkoda.

Funded by the European Union

Causal modeling:

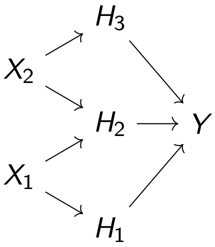
- We want a statistical model which captures the causal structure encoded in a directed graph $\mathcal{G} = (V, E)$.



Causal modeling with directed graphs

Causal modeling:

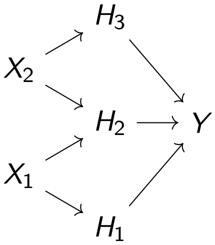
- We want a statistical model which captures the causal structure encoded in a directed graph $\mathcal{G} = (V, E)$.
- Parents of a node are regarded as its direct causes, further-up ancestors are only indirect causes.



Causal modeling with directed graphs

Causal modeling:

- We want a statistical model which captures the causal structure encoded in a directed graph $\mathcal{G} = (V, E)$.
- Parents of a node are regarded as its direct causes, further-up ancestors are only indirect causes.
- The causal diagram may come from expert knowledge or it might be learned from observational data.



What is a statistical model?

In this talk:

- A statistical model is a set of multivariate normal distributions $X = (X_1, \ldots, X_n)$.

What is a statistical model?

In this talk:

- A statistical model is a set of multivariate normal distributions $X = (X_1, \ldots, X_n)$.
- The distribution of X is entirely described by its mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathrm{PD}_n$:

$$\mathrm{pdf}_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right).$$

What is a statistical model?

In this talk:

- A statistical model is a set of multivariate normal distributions $X = (X_1, \ldots, X_n)$.
- The distribution of X is entirely described by its mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathsf{PD}_n$:

$$\mathrm{pdf}_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right).$$

- We assume $\mu = 0$ and identify X with its covariance matrix.

What is a statistical model?

In this talk:

- A statistical model is a set of multivariate normal distributions $X = (X_1, \ldots, X_n)$.
- The distribution of X is entirely described by its mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathrm{PD}_n$:

$$\mathrm{pdf}_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right).$$

- We assume $\mu = 0$ and identify X with its covariance matrix.
- So, a statistical model is a subset of PD_n.

What is a statistical model?

In this talk:

- A statistical model is a set of multivariate normal distributions $X = (X_1, \ldots, X_n)$.
- The distribution of X is entirely described by its mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathrm{PD}_n$:

$$\mathrm{pdf}_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right).$$

- We assume $\mu = 0$ and identify X with its covariance matrix.
- So, a statistical model is a subset of PD_n.

Fundamental theorem of algebraic statistics

Statistical models are semialgebraic sets.*

What is a statistical model?

In this talk:

- A statistical model is a set of multivariate normal distributions $X = (X_1, \ldots, X_n)$.
- The distribution of X is entirely described by its mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathrm{PD}_n$:

$$\mathrm{pdf}_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right).$$

- We assume $\mu = 0$ and identify X with its covariance matrix.
- So, a statistical model is a subset of PD_n.

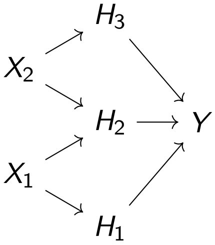
Fundamental theorem of algebraic statistics

Statistical models are semialgebraic sets.*

*sometimes

Linear structural equation models

- Consider a directed acyclic graph $\mathcal{G} = (V, E)$.



Linear structural equation models

- Consider a directed acyclic graph $\mathcal{G} = (V, E)$.
- The linear structural equation model for $\mathcal{G}$ encodes distributions of $X = (X_i : i \in V)$ where each X_j is a linear function of its parents, up to some noise:

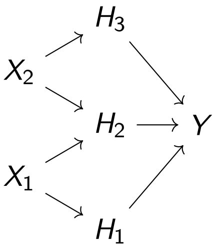
$$X_j = \sum_{i \in \mathrm{pa}(j)} \lambda_{ij} X_i + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \omega_j).$$

$$\begin{tikzcd}
 & H_3 \arrow[ddr] & \\
X_2 \arrow[ur] \arrow[dr] & & \\
 & H_2 \arrow[r] & Y \\
X_1 \arrow[ur] \arrow[dr] & & \\
 & H_1 \arrow[uur] &
\end{tikzcd}$$

Linear structural equation models

- Consider a directed acyclic graph $\mathcal{G} = (V, E)$.
- The linear structural equation model for $\mathcal{G}$ encodes distributions of $X = (X_i : i \in V)$ where each X_j is a linear function of its parents, up to some noise:

$$X_j = \sum_{i \in \mathrm{pa}(j)} \lambda_{ij} X_i + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \omega_j).$$

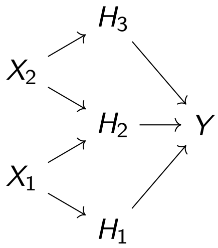


- In short: $X = \Lambda X + \varepsilon$ where Λ is $\mathcal{G}$-sparse and $\varepsilon \sim \mathcal{N}(0, \Omega)$ with Ω diagonal.

Linear structural equation models

- Consider a directed acyclic graph $\mathcal{G} = (V, E)$.
- The linear structural equation model for $\mathcal{G}$ encodes distributions of $X = (X_i : i \in V)$ where each X_j is a linear function of its parents, up to some noise:

$$X_j = \sum_{i \in \mathrm{pa}(j)} \lambda_{ij} X_i + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \omega_j).$$



- In short: $X = \Lambda X + \varepsilon$ where Λ is $\mathcal{G}$-sparse and $\varepsilon \sim \mathcal{N}(0, \Omega)$ with Ω diagonal.
- Solutions to this system are multivariate normal distributions with zero mean and covariance matrix Σ satisfying the congruence

$$\boxed{(\mathrm{Id} - \Lambda)^{\mathsf{T}} \Sigma (\mathrm{Id} - \Lambda) = \Omega.}$$

Example

$$\mathcal{G} = 1 \longrightarrow 2 \longrightarrow 3$$

Example

$$\mathcal{G} = 1 \longrightarrow 2 \longrightarrow 3 \qquad \Lambda = \begin{pmatrix} 0 & \lambda_{12} & 0 \\ 0 & 0 & \lambda_{13} \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad \Omega = \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & 0 \\ 0 & 0 & \omega_{33} \end{pmatrix} \in \mathrm{PD}_3$$

Example

$$\mathcal{G} = 1 \longrightarrow 2 \longrightarrow 3 \qquad \Lambda = \begin{pmatrix} 0 & \lambda_{12} & 0 \\ 0 & 0 & \lambda_{13} \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad \Omega = \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & 0 \\ 0 & 0 & \omega_{33} \end{pmatrix} \in \mathrm{PD}_3$$

$$\begin{aligned} \Sigma &= \phi_{\mathcal{G}}(\Lambda, \Omega) = (\mathrm{Id} - \Lambda)^{-T}\Omega(\mathrm{Id} - \Lambda)^{-1} \\ &= \begin{pmatrix} \omega_{11} & \lambda_{12}\omega_{11} & \lambda_{12}\lambda_{23}\omega_{11} \\ \lambda_{12}\omega_{11} & \lambda_{12}^2\omega_{11} + \omega_{22} & \lambda_{12}^2\lambda_{23}\omega_{11} + \lambda_{23}\omega_{22} \\ \lambda_{12}\lambda_{23}\omega_{11} & \lambda_{12}^2\lambda_{23}\omega_{11} + \lambda_{23}\omega_{22} & \lambda_{12}^2\lambda_{23}^2\omega_{11} + \lambda_{23}^2\omega_{22} + \omega_{33} \end{pmatrix} \end{aligned}$$

Example

$$\mathcal{G} = 1 \longrightarrow 2 \longrightarrow 3 \qquad \Lambda = \begin{pmatrix} 0 & \lambda_{12} & 0 \\ 0 & 0 & \lambda_{13} \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad \Omega = \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & 0 \\ 0 & 0 & \omega_{33} \end{pmatrix} \in \mathrm{PD}_3$$

$$\begin{aligned} \Sigma &= \phi_{\mathcal{G}}(\Lambda, \Omega) = (\mathrm{Id} - \Lambda)^{-T} \Omega (\mathrm{Id} - \Lambda)^{-1} \\ &= \begin{pmatrix} \omega_{11} & \lambda_{12}\omega_{11} & \lambda_{12}\lambda_{23}\omega_{11} \\ \lambda_{12}\omega_{11} & \lambda_{12}^2\omega_{11} + \omega_{22} & \lambda_{12}^2\lambda_{23}\omega_{11} + \lambda_{23}\omega_{22} \\ \lambda_{12}\lambda_{23}\omega_{11} & \lambda_{12}^2\lambda_{23}\omega_{11} + \lambda_{23}\omega_{22} & \lambda_{12}^2\lambda_{23}^2\omega_{11} + \lambda_{23}^2\omega_{22} + \omega_{33} \end{pmatrix} \end{aligned}$$

- Since $\mathcal{G}$ is acyclic, $\det(\mathrm{Id} - \Lambda) = 1$ and $\phi_{\mathcal{G}}$ is a polynomial map!

Example

$$\mathcal{G} = 1 \longrightarrow 2 \longrightarrow 3 \qquad \Lambda = \begin{pmatrix} 0 & \lambda_{12} & 0 \\ 0 & 0 & \lambda_{13} \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad \Omega = \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & 0 \\ 0 & 0 & \omega_{33} \end{pmatrix} \in \mathrm{PD}_3$$

$$\begin{aligned} \Sigma = \phi_{\mathcal{G}}(\Lambda, \Omega) &= (\mathrm{Id} - \Lambda)^{-\mathsf{T}} \Omega (\mathrm{Id} - \Lambda)^{-1} \\ &= \begin{pmatrix} \omega_{11} & \lambda_{12}\omega_{11} & \lambda_{12}\lambda_{23}\omega_{11} \\ \lambda_{12}\omega_{11} & \lambda_{12}^2\omega_{11} + \omega_{22} & \lambda_{12}^2\lambda_{23}\omega_{11} + \lambda_{23}\omega_{22} \\ \lambda_{12}\lambda_{23}\omega_{11} & \lambda_{12}^2\lambda_{23}\omega_{11} + \lambda_{23}\omega_{22} & \lambda_{12}^2\lambda_{23}^2\omega_{11} + \lambda_{23}^2\omega_{22} + \omega_{33} \end{pmatrix} \end{aligned}$$

- Since $\mathcal{G}$ is acyclic, $\det(\mathrm{Id} - \Lambda) = 1$ and $\phi_{\mathcal{G}}$ is a polynomial map!
- Using elimination theory we find that the image is cut out of PD_3 by a single quadratic: $\sigma_{13}\sigma_{22} - \sigma_{12}\sigma_{23}$.

LSEMs as statistical models

- The model $\mathcal{M}(\mathcal{G})$ is the image of the polynomial map $\phi_{\mathcal{G}} \colon \mathbb{R}^E \times \mathbb{R}^V_{>0} \to \mathrm{PD}_n$

$$\phi_{\mathcal{G}}(\Lambda, \Omega) := \Sigma = (I - \Lambda)^{-\mathsf{T}} \Omega (I - \Lambda)^{-1}$$

on a semialgebraic domain. Hence, $\mathcal{M}(\mathcal{G})$ is a semialgebraic set.

LSEMs as statistical models

- The model $\mathcal{M}(\mathcal{G})$ is the image of the polynomial map $\phi_{\mathcal{G}} \colon \mathbb{R}^E \times \mathbb{R}^V_{>0} \to \mathrm{PD}_n$

$$\phi_{\mathcal{G}}(\Lambda, \Omega) := \Sigma = (I - \Lambda)^{-\mathsf{T}} \Omega (I - \Lambda)^{-1}$$

  on a semialgebraic domain. Hence, $\mathcal{M}(\mathcal{G})$ is a semialgebraic set.

- Let $\mathcal{V}(\mathcal{G}) = \overline{\operatorname{im} \widehat{\phi}_{\mathcal{G}}}$ where $\widehat{\phi}_{\mathcal{G}} \colon \mathbb{C}^E \times \mathbb{C}^V \to \mathbb{C}^{\binom{n+1}{2}}$ extends the domain of $\phi_{\mathcal{G}}$.

LSEMs as statistical models

- The model $\mathcal{M}(\mathcal{G})$ is the image of the polynomial map $\phi_{\mathcal{G}} \colon \mathbb{R}^E \times \mathbb{R}^V_{>0} \to \mathrm{PD}_n$

$$\phi_{\mathcal{G}}(\Lambda, \Omega) \coloneqq \Sigma = (I - \Lambda)^{-\mathsf{T}} \Omega (I - \Lambda)^{-1}$$

  on a semialgebraic domain. Hence, $\mathcal{M}(\mathcal{G})$ is a semialgebraic set.

- Let $\mathcal{V}(\mathcal{G}) = \overline{\operatorname{im} \widehat{\phi}_{\mathcal{G}}}$ where $\widehat{\phi}_{\mathcal{G}} \colon \mathbb{C}^E \times \mathbb{C}^V \to \mathbb{C}^{\binom{n+1}{2}}$ extends the domain of $\phi_{\mathcal{G}}$.

Observations

- As the image of an irreducible variety under a polynomial map, $\mathcal{V}(\mathcal{G})$ is irreducible.

LSEMs as statistical models

- The model $\mathcal{M}(\mathcal{G})$ is the image of the polynomial map $\phi_{\mathcal{G}} \colon \mathbb{R}^E \times \mathbb{R}^V_{>0} \to \mathrm{PD}_n$

$$\phi_{\mathcal{G}}(\Lambda, \Omega) := \Sigma = (I - \Lambda)^{-\mathsf{T}} \Omega (I - \Lambda)^{-1}$$

  on a semialgebraic domain. Hence, $\mathcal{M}(\mathcal{G})$ is a semialgebraic set.

- Let $\mathcal{V}(\mathcal{G}) = \overline{\mathrm{im}\, \widehat{\phi}_{\mathcal{G}}}$ where $\widehat{\phi}_{\mathcal{G}} \colon \mathbb{C}^E \times \mathbb{C}^V \to \mathbb{C}^{\binom{n+1}{2}}$ extends the domain of $\phi_{\mathcal{G}}$.

Observations

- As the image of an irreducible variety under a polynomial map, $\mathcal{V}(\mathcal{G})$ is irreducible.
- We have $\mathcal{M}(\mathcal{G}) = \mathrm{PD}_n \cap \mathcal{V}(\mathcal{G})$. Conversely, $\mathcal{M}(\mathcal{G})$ is dense in $\mathcal{V}(\mathcal{G})$.

LSEMs as statistical models

- The model $\mathcal{M}(\mathcal{G})$ is the image of the polynomial map $\phi_{\mathcal{G}} \colon \mathbb{R}^E \times \mathbb{R}^V_{>0} \to \mathrm{PD}_n$

$$\phi_{\mathcal{G}}(\Lambda, \Omega) := \Sigma = (I - \Lambda)^{-\mathsf{T}} \Omega (I - \Lambda)^{-1}$$

  on a semialgebraic domain. Hence, $\mathcal{M}(\mathcal{G})$ is a semialgebraic set.

- Let $\mathcal{V}(\mathcal{G}) = \overline{\mathrm{im}\, \widehat{\phi}_{\mathcal{G}}}$ where $\widehat{\phi}_{\mathcal{G}} \colon \mathbb{C}^E \times \mathbb{C}^V \to \mathbb{C}^{\binom{n+1}{2}}$ extends the domain of $\phi_{\mathcal{G}}$.

Observations

- As the image of an irreducible variety under a polynomial map, $\mathcal{V}(\mathcal{G})$ is irreducible.
- We have $\mathcal{M}(\mathcal{G}) = \mathrm{PD}_n \cap \mathcal{V}(\mathcal{G})$. Conversely, $\mathcal{M}(\mathcal{G})$ is dense in $\mathcal{V}(\mathcal{G})$.

For many purposes we may study $\mathcal{V}(\mathcal{G})$ instead of $\mathcal{M}(\mathcal{G})$.

Statistically meaningful things to investigate (algebraically)

Hypothesis testing [SDL24]

- Need to describe $\mathcal{M}(\mathcal{G})$ implicitly using polynomial equations (and inequalities).

Statistically meaningful things to investigate (algebraically)

Hypothesis testing [SDL24]

- Need to describe $\mathcal{M}(\mathcal{G})$ implicitly using polynomial equations (and inequalities).
- Some tests deteriorate near singularities. Is the model smooth? [Drt09]

Statistically meaningful things to investigate (algebraically)

Hypothesis testing [SDL24]

- Need to describe $\mathcal{M}(\mathcal{G})$ implicitly using polynomial equations (and inequalities).
- Some tests deteriorate near singularities. Is the model smooth? [Drt09]

Parameter identifiability: learning (Λ, Ω) from observations

- Is $\phi_{\mathcal{G}}$ injective or generically injective (or at least generically finite-to-one)?

Statistically meaningful things to investigate (algebraically)

Hypothesis testing [SDL24]

- Need to describe $\mathcal{M}(\mathcal{G})$ implicitly using polynomial equations (and inequalities).
- Some tests deteriorate near singularities. Is the model smooth? [Drt09]

Parameter identifiability: learning (Λ, Ω) from observations

- Is $\phi_{\mathcal{G}}$ injective or generically injective (or at least generically finite-to-one)?

Structure identifiability: learning $\mathcal{G}$ from observations

- At the very least need to know model equivalence: when is $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$?

Statistically meaningful things to investigate (algebraically)

Hypothesis testing [SDL24]

- Need to describe $\mathcal{M}(\mathcal{G})$ implicitly using polynomial equations (and inequalities).
- Some tests deteriorate near singularities. Is the model smooth? [Drt09]

Parameter identifiability: learning (Λ, Ω) from observations

- Is $\phi_{\mathcal{G}}$ injective or generically injective (or at least generically finite-to-one)?

Structure identifiability: learning $\mathcal{G}$ from observations

- At the very least need to know model equivalence: when is $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$?
- Notice that for LSEMs this is the case if and only if $\mathcal{V}(\mathcal{G}) = \mathcal{V}(\mathcal{H})$!

Two shortcomings of LSEMs

- If $\mathcal{G}$ is acyclic, then $\phi_{\mathcal{G}}$ has a rational inverse (global parameter identifiability).

Two shortcomings of LSEMs

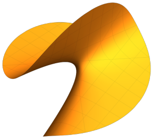
- If $\mathcal{G}$ is acyclic, then $\phi_{\mathcal{G}}$ has a rational inverse (global parameter identifiability).
- Identifiability is lost when $\mathcal{G}$ contains feedback loops (cycles).

Two shortcomings of LSEMs

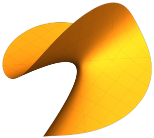
- If $\mathcal{G}$ is acyclic, then $\phi_{\mathcal{G}}$ has a rational inverse (global parameter identifiability).
- Identifiability is lost when $\mathcal{G}$ contains feedback loops (cycles).
- Different causal structures may define the same model due to Markov equivalence.

Two shortcomings of LSEMs

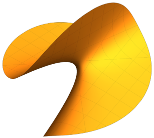
- If $\mathcal{G}$ is acyclic, then $\phi_{\mathcal{G}}$ has a rational inverse (global parameter identifiability).
- Identifiability is lost when $\mathcal{G}$ contains feedback loops (cycles).
- Different causal structures may define the same model due to Markov equivalence.



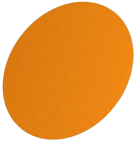
$\circ \to \circ \to \circ$



$\circ \leftarrow \circ \leftarrow \circ$



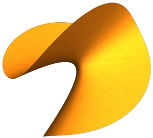
$\circ \leftarrow \circ \to \circ$



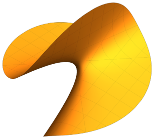
$\circ \to \circ \leftarrow \circ$

Two shortcomings of LSEMs

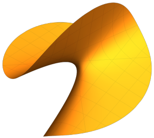
- If $\mathcal{G}$ is acyclic, then $\phi_{\mathcal{G}}$ has a rational inverse (global parameter identifiability).
- Identifiability is lost when $\mathcal{G}$ contains feedback loops (cycles).
- Different causal structures may define the same model due to Markov equivalence.



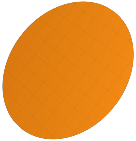
$\circ \rightarrow \circ \rightarrow \circ$



$\circ \leftarrow \circ \leftarrow \circ$



$\circ \leftarrow \circ \rightarrow \circ$



$\circ \rightarrow \circ \leftarrow \circ$

- Same model despite **opposite** claims of causation.

Markov / model equivalence

Verma & Pearl [VP90; AMP97]

It holds that $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if $\mathcal{G}$ and $\mathcal{H}$ have the same skeleton and $\mathcal{M}(\mathcal{G}[K]) = \mathcal{M}(\mathcal{H}[K])$ for all $K \in \binom{V}{3}$.

Markov / model equivalence

Verma & Pearl [VP90; AMP97]

It holds that $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if $\mathcal{G}$ and $\mathcal{H}$ have the same skeleton and $\mathcal{M}(\mathcal{G}[K]) = \mathcal{M}(\mathcal{H}[K])$ for all $K \in \binom{V}{3}$.

- The criterion boils down to:

$$\mathcal{G}[K] \text{ is a v-structure } (\circ \to \circ \leftarrow \circ) \text{ if and only if } \mathcal{H}[K] \text{ is.}$$

Markov / model equivalence

Verma & Pearl [VP90; AMP97]

It holds that $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if $\mathcal{G}$ and $\mathcal{H}$ have the same skeleton and $\mathcal{M}(\mathcal{G}[K]) = \mathcal{M}(\mathcal{H}[K])$ for all $K \in \binom{V}{3}$.

- The criterion boils down to:

$$\mathcal{G}[K] \text{ is a v-structure } (\circ \rightarrow \circ \leftarrow \circ) \text{ if and only if } \mathcal{H}[K] \text{ is.}$$

Chickering [Chi95]

It holds that $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if $\mathcal{G}$ and $\mathcal{H}$ can be transformed into one another by a sequence of covered edge flips.

Markov / model equivalence

Verma & Pearl [VP90; AMP97]

It holds that $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if $\mathcal{G}$ and $\mathcal{H}$ have the same skeleton and $\mathcal{M}(\mathcal{G}[K]) = \mathcal{M}(\mathcal{H}[K])$ for all $K \in \binom{V}{3}$.

- The criterion boils down to:

$$\mathcal{G}[K] \text{ is a v-structure } (\circ \to \circ \leftarrow \circ) \text{ if and only if } \mathcal{H}[K] \text{ is.}$$

Chickering [Chi95]

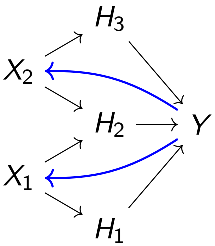
It holds that $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if $\mathcal{G}$ and $\mathcal{H}$ can be transformed into one another by a sequence of covered edge flips.

- An edge $i \to j$ is covered if $\mathrm{pa}(j) = \mathrm{pa}(i) \cup \{i\}$.

$$\{\circ \overset{*}{\to} \circ \to \circ,\ \ \circ \overset{*}{\leftarrow} \circ \overset{*}{\to} \circ,\ \ \circ \leftarrow \circ \overset{*}{\leftarrow} \circ\} \qquad \{\circ \to \circ \leftarrow \circ\}$$

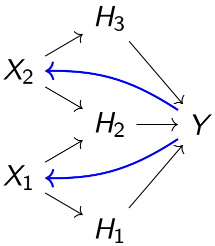
Lyapunov models

- We want a statistical model which captures the causal structure of a directed graph $\mathcal{G} = (V, E)$ with cycles.



Lyapunov models

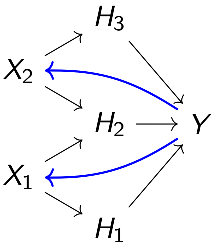
- We want a statistical model which captures the causal structure of a directed graph $\mathcal{G} = (V, E)$ with cycles.
- Idea of [Fit19] and independently [VH20]: unroll the feedback loop over time:



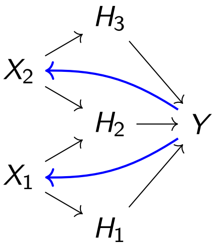
Lyapunov models

- We want a statistical model which captures the causal structure of a directed graph $\mathcal{G} = (V, E)$ with cycles.
- Idea of [Fit19] and independently [VH20]: unroll the feedback loop over time:

$$
\begin{tikzcd}
\cdots \arrow[r] \arrow[dr] & X(t) \arrow[r] \arrow[dr] & X(t+1) \arrow[r] \arrow[dr] & \cdots \\
\cdots \arrow[r] \arrow[ur] & Y(t) \arrow[r] \arrow[ur] & Y(t+1) \arrow[r] \arrow[ur] & \cdots
\end{tikzcd}
$$



Lyapunov models



- We want a statistical model which captures the causal structure of a directed graph $\mathcal{G} = (V, E)$ with cycles.
- Idea of [Fit19] and independently [VH20]: unroll the feedback loop over time:

$$\begin{tikzcd}
\cdots \arrow[r] \arrow[dr] & X(t) \arrow[r] \arrow[dr] & X(t+1) \arrow[r] \arrow[dr] & \cdots \\
\cdots \arrow[r] \arrow[ur] & Y(t) \arrow[r] \arrow[ur] & Y(t+1) \arrow[r] \arrow[ur] & \cdots
\end{tikzcd}$$

- Formally, take the stationary distribution of the Ornstein–Uhlenbeck process

$$\mathrm{d}\mathbb{X}(t) = M\mathbb{X}(t)\mathrm{d}t + D\mathrm{d}\mathbb{W}(t), \qquad \text{(cf. } X = \Lambda X + \varepsilon\text{)}$$

where M is $\mathcal{G}$-sparse, $\mathbb{W}$ a standard Brownian motion, and D diagonal.

Parametrization of Lyapunov models

- The stationary distribution is Gaußian with covariance matrix satisfying the Lyapunov equation $\boxed{M\Sigma + \Sigma M^{\mathsf{T}} + C = 0}$ where $C = DD^{\mathsf{T}}$.

Parametrization of Lyapunov models

- The stationary distribution is Gaußian with covariance matrix satisfying the Lyapunov equation $\boxed{M\Sigma + \Sigma M^{\mathsf{T}} + C = 0}$ where $C = DD^{\mathsf{T}}$.
- If M is stable (all eigenvalues have negative real part) and C is positive definite, then there exists a unique positive definite solution Σ.

Parametrization of Lyapunov models

- The stationary distribution is Gaußian with covariance matrix satisfying the Lyapunov equation $\boxed{M\Sigma + \Sigma M^{\mathsf{T}} + C = 0}$ where $C = DD^{\mathsf{T}}$.
- If M is stable (all eigenvalues have negative real part) and C is positive definite, then there exists a unique positive definite solution Σ.
- The Lyapunov equation is a linear matrix equation in Σ, so it can be rewritten via vectorization and Kronecker products:

$$(\mathrm{Id} \otimes M + M \otimes \mathrm{Id}) \operatorname{vec} \Sigma = -\operatorname{vec} C.$$

Parametrization of Lyapunov models

- The stationary distribution is Gaußian with covariance matrix satisfying the Lyapunov equation $M\Sigma + \Sigma M^\mathsf{T} + C = 0$ where $C = DD^\mathsf{T}$.
- If M is stable (all eigenvalues have negative real part) and C is positive definite, then there exists a unique positive definite solution Σ.
- The Lyapunov equation is a linear matrix equation in Σ, so it can be rewritten via vectorization and Kronecker products:

$$(\mathrm{Id} \otimes M + M \otimes \mathrm{Id}) \operatorname{vec} \Sigma = - \operatorname{vec} C.$$

- The unique solution is obtained via Cramer's rule.

Parametrization of Lyapunov models

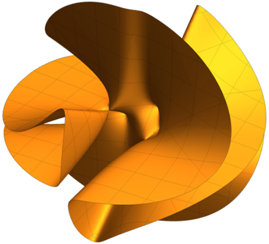
- The stationary distribution is Gaußian with covariance matrix satisfying the Lyapunov equation $\boxed{M\Sigma + \Sigma M^{\mathsf{T}} + C = 0}$ where $C = DD^{\mathsf{T}}$.
- If M is stable (all eigenvalues have negative real part) and C is positive definite, then there exists a unique positive definite solution Σ.
- The Lyapunov equation is a linear matrix equation in Σ, so it can be rewritten via vectorization and Kronecker products:

$$(\mathrm{Id} \otimes M + M \otimes \mathrm{Id}) \operatorname{vec} \Sigma = -\operatorname{vec} C.$$

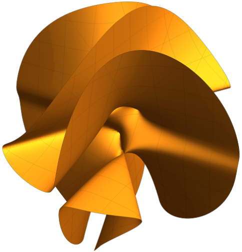
- The unique solution is obtained via Cramer's rule.

The Lyapunov model is an irreducible algebraic variety (inside PD_n)!

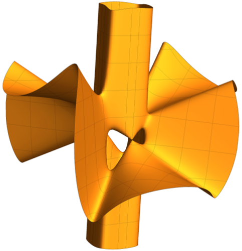
Beauty is in the eye of the beholder...



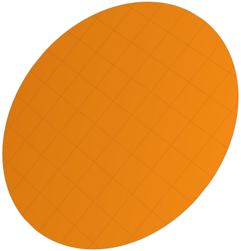
$\circ \rightarrow \circ \rightarrow \circ$



$\circ \leftarrow \circ \leftarrow \circ$



$\circ \leftarrow \circ \rightarrow \circ$



$\circ \rightarrow \circ \leftarrow \circ$

Parameter identifiability

$$M\Sigma + \Sigma M^{\mathsf{T}} + C = 0.$$

- The Lyapunov equation is **also** a linear matrix equation in M, equivalent to:

$$(\Sigma \otimes \mathrm{Id} + (\mathrm{Id} \otimes \Sigma) K_n) \operatorname{vec} M = -\operatorname{vec} C,$$

where K_n is the commutation matrix satisfying $K_n \operatorname{vec}(M) = \operatorname{vec}(M^{\mathsf{T}})$.

Parameter identifiability

$$M\Sigma + \Sigma M^{\mathsf{T}} + C = 0.$$

- The Lyapunov equation is **also** a linear matrix equation in M, equivalent to:

$$(\Sigma \otimes \mathrm{Id} + (\mathrm{Id} \otimes \Sigma) K_n) \operatorname{vec} M = -\operatorname{vec} C,$$

  where K_n is the commutation matrix satisfying $K_n \operatorname{vec}(M) = \operatorname{vec}(M^{\mathsf{T}})$.

- It has redundant rows since Σ and C are symmetric, but:

Parameter identifiability

$$M\Sigma + \Sigma M^{\mathsf{T}} + C = 0.$$

- The Lyapunov equation is **also** a linear matrix equation in M, equivalent to:

$$(\Sigma \otimes \mathrm{Id} + (\mathrm{Id} \otimes \Sigma) K_n) \operatorname{vec} M = -\operatorname{vec} C,$$

where K_n is the commutation matrix satisfying $K_n \operatorname{vec}(M) = \operatorname{vec}(M^{\mathsf{T}})$.

- It has redundant rows since Σ and C are symmetric, but:

Theorem ([DHADH23])

Let $\mathcal{G}$ be a simple directed graph (i.e., having no directed 2-cycles). Given Σ in the Lyapunov model of $\mathcal{G}$ with fixed diffusion matrix C, the parameter matrix M is uniquely recoverable as a rational function of Σ.

Parameter identifiability

$$M\Sigma + \Sigma M^{\mathsf{T}} + C = 0.$$

- The Lyapunov equation is **also** a linear matrix equation in M, equivalent to:

$$(\Sigma \otimes \mathrm{Id} + (\mathrm{Id} \otimes \Sigma)K_n)\,\mathrm{vec}\,M = -\,\mathrm{vec}\,C,$$

where K_n is the commutation matrix satisfying $K_n\,\mathrm{vec}(M) = \mathrm{vec}(M^{\mathsf{T}})$.

- It has redundant rows since Σ and C are symmetric, but:

Theorem ([DHADH23])

Let $\mathcal{G}$ be a simple directed graph (i.e., having no directed 2-cycles). Given Σ in the Lyapunov model of $\mathcal{G}$ with fixed diffusion matrix C, the parameter matrix M is uniquely recoverable as a rational function of Σ.

Lyapunov models are more often globally identifiable than LSEMs!

Model equivalence for Lyapunov models

- For simplicity we fix $C = 2\,\mathrm{Id}$. Arguments work for any diagonal C.

Model equivalence for Lyapunov models

- For simplicity we fix $C = 2\,\mathrm{Id}$. Arguments work for any diagonal C.
- We have a family of semialgebraic sets $\mathcal{M}(\mathcal{G})$ indexed by directed graphs.

Model equivalence for Lyapunov models

- For simplicity we fix $C = 2\,\mathrm{Id}$. Arguments work for any diagonal C.
- We have a family of semialgebraic sets $\mathcal{M}(\mathcal{G})$ indexed by directed graphs.
- Analogously to LSEMs we get irreducible varieties $\mathcal{V}(\mathcal{G})$ such that:

$$\mathcal{M}(\mathcal{G}) = \mathrm{PD}_n \cap \mathcal{V}(\mathcal{G}) \quad \text{and} \quad \mathcal{M}(\mathcal{G}) \text{ is dense in } \mathcal{V}(\mathcal{G}).$$

Model equivalence for Lyapunov models

- For simplicity we fix $C = 2\,\mathrm{Id}$. Arguments work for any diagonal C.
- We have a family of semialgebraic sets $\mathcal{M}(\mathcal{G})$ indexed by directed graphs.
- Analogously to LSEMs we get irreducible varieties $\mathcal{V}(\mathcal{G})$ such that:

$$\mathcal{M}(\mathcal{G}) = \mathrm{PD}_n \cap \mathcal{V}(\mathcal{G}) \quad \text{and} \quad \mathcal{M}(\mathcal{G}) \text{ is dense in } \mathcal{V}(\mathcal{G}).$$

- Hence, $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if $\mathcal{V}(\mathcal{G}) = \mathcal{V}(\mathcal{H})$.

Model equivalence for Lyapunov models

- For simplicity we fix $C = 2\,\mathrm{Id}$. Arguments work for any diagonal C.
- We have a family of semialgebraic sets $\mathcal{M}(\mathcal{G})$ indexed by directed graphs.
- Analogously to LSEMs we get irreducible varieties $\mathcal{V}(\mathcal{G})$ such that:

$$\mathcal{M}(\mathcal{G}) = \mathrm{PD}_n \cap \mathcal{V}(\mathcal{G}) \quad \text{and} \quad \mathcal{M}(\mathcal{G}) \text{ is dense in } \mathcal{V}(\mathcal{G}).$$

- Hence, $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if $\mathcal{V}(\mathcal{G}) = \mathcal{V}(\mathcal{H})$.

☞ The goal is to learn exactly what the function field $\mathbb{C}(\mathcal{V}(\mathcal{G}))$ reveals about the graph $\mathcal{G}$.

The essential ingredient: missing-edge relations

- Let $\mathcal{G}$ be a directed acyclic graph and fix a complete DAG $\mathcal{G}'$ containing $\mathcal{G}$.

The essential ingredient: missing-edge relations

- Let $\mathcal{G}$ be a directed acyclic graph and fix a complete DAG $\mathcal{G}'$ containing $\mathcal{G}$.
- $\mathrm{Stab}_n(\mathcal{G})$ is a coordinate subspace of $\mathrm{Stab}_n(\mathcal{G}')$.

The essential ingredient: missing-edge relations

- Let $\mathcal{G}$ be a directed acyclic graph and fix a complete DAG $\mathcal{G}'$ containing $\mathcal{G}$.
- $\mathrm{Stab}_n(\mathcal{G})$ is a coordinate subspace of $\mathrm{Stab}_n(\mathcal{G}')$.
- We want to describe $\mathcal{M}(\mathcal{G})$ as a subvariety of $\mathrm{PD}_n = \mathcal{M}(\mathcal{G}')$.

The essential ingredient: missing-edge relations

- Let $\mathcal{G}$ be a directed acyclic graph and fix a complete DAG $\mathcal{G}'$ containing $\mathcal{G}$.
- $\mathrm{Stab}_n(\mathcal{G})$ is a coordinate subspace of $\mathrm{Stab}_n(\mathcal{G}')$.
- We want to describe $\mathcal{M}(\mathcal{G})$ as a subvariety of $\mathrm{PD}_n = \mathcal{M}(\mathcal{G}')$.
- Take the parameter identification map $\psi_{\mathcal{G}'}$ and observe that if $i \to j \in E(\mathcal{G}') \setminus E(\mathcal{G})$ then

$$\psi_{\mathcal{G}'}(\Sigma)_{ij} = 0 \text{ for all } \Sigma \in \mathcal{M}(\mathcal{G}) \text{ because } m_{ij} = 0 \text{ on } \mathrm{Stab}_n(\mathcal{G}).$$

The essential ingredient: missing-edge relations

- Let $\mathcal{G}$ be a directed acyclic graph and fix a complete DAG $\mathcal{G}'$ containing $\mathcal{G}$.
- $\mathrm{Stab}_n(\mathcal{G})$ is a coordinate subspace of $\mathrm{Stab}_n(\mathcal{G}')$.
- We want to describe $\mathcal{M}(\mathcal{G})$ as a subvariety of $\mathrm{PD}_n = \mathcal{M}(\mathcal{G}')$.
- Take the parameter identification map $\psi_{\mathcal{G}'}$ and observe that if $i \to j \in E(\mathcal{G}') \setminus E(\mathcal{G})$ then

$$\psi_{\mathcal{G}'}(\Sigma)_{ij} = 0 \text{ for all } \Sigma \in \mathcal{M}(\mathcal{G}) \text{ because } m_{ij} = 0 \text{ on } \mathrm{Stab}_n(\mathcal{G}).$$

Proposition

The numerators of the $\psi_{\mathcal{G}'}(\Sigma)_{ij}$ for $i \to j \in E(\mathcal{G}') \setminus E(\mathcal{G})$ are the missing-edge relations of $\mathcal{G}$ wrt $\mathcal{G}'$. They cut out a variety $\mathcal{V}_{\mathcal{G}'}(\mathcal{G})$ such that $\mathcal{M}(\mathcal{G}) = \mathrm{PD}_n \cap \mathcal{V}_{\mathcal{G}'}(\mathcal{G})$.

The essential ingredient: missing-edge relations

- Let $\mathcal{G}$ be a directed acyclic graph and fix a complete DAG $\mathcal{G}'$ containing $\mathcal{G}$.
- $\mathrm{Stab}_n(\mathcal{G})$ is a coordinate subspace of $\mathrm{Stab}_n(\mathcal{G}')$.
- We want to describe $\mathcal{M}(\mathcal{G})$ as a subvariety of $\mathrm{PD}_n = \mathcal{M}(\mathcal{G}')$.
- Take the parameter identification map $\psi_{\mathcal{G}'}$ and observe that if $i \to j \in E(\mathcal{G}') \setminus E(\mathcal{G})$ then

$$\psi_{\mathcal{G}'}(\Sigma)_{ij} = 0 \text{ for all } \Sigma \in \mathcal{M}(\mathcal{G}) \text{ because } m_{ij} = 0 \text{ on } \mathrm{Stab}_n(\mathcal{G}).$$

Proposition

The numerators of the $\psi_{\mathcal{G}'}(\Sigma)_{ij}$ for $i \to j \in E(\mathcal{G}') \setminus E(\mathcal{G})$ are the missing-edge relations of $\mathcal{G}$ wrt $\mathcal{G}'$. They cut out a variety $\mathcal{V}_{\mathcal{G}'}(\mathcal{G})$ such that $\mathcal{M}(\mathcal{G}) = \mathrm{PD}_n \cap \mathcal{V}_{\mathcal{G}'}(\mathcal{G})$.

- We have $\mathcal{V}_{\mathcal{G}}^{\mathcal{G}'} \supseteq \mathcal{V}(\mathcal{G})$ and $\dim \mathcal{V}_{\mathcal{G}'}(\mathcal{G}) = \dim \mathcal{V}(\mathcal{G})$ but $\mathcal{V}_{\mathcal{G}'}(\mathcal{G})$ may be reducible.

Model equivalence via induced subgraphs

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if they have the same skeleton and $\mathcal{M}(\mathcal{G}[K]) = \mathcal{M}(\mathcal{H}[K])$ for all $K \in \binom{V}{4}$.

☞ Bayesian networks require all 3-induced subgraphs to be equivalent.

Model equivalence via induced subgraphs

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if they have the same skeleton and $\mathcal{M}(\mathcal{G}[K]) = \mathcal{M}(\mathcal{H}[K])$ for all $K \in \binom{V}{\mathbf{4}}$.

☞ Bayesian networks require all 3-induced subgraphs to be equivalent.

Corollary

Lyapunov models refine the Markov equivalence classes of LSEMs.

Model equivalence via induced subgraphs

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if they have the same skeleton and $\mathcal{M}(\mathcal{G}[K]) = \mathcal{M}(\mathcal{H}[K])$ for all $K \in \binom{V}{\mathbf{4}}$.

☞ Bayesian networks require all 3-induced subgraphs to be equivalent.

Corollary

Lyapunov models refine the Markov equivalence classes of LSEMs.

Corollary

Model equivalence can be checked in polynomial time.

Model equivalence via edge flips

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if there is a sequence of super-covered edge flips that transforms $\mathcal{G}$ into $\mathcal{H}$.

Model equivalence via edge flips

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if there is a sequence of super-covered edge flips that transforms $\mathcal{G}$ into $\mathcal{H}$. The edge $i \to j$ is super-covered if:

- $\mathrm{pa}(i) \cup \{i\} = \mathrm{pa}(j)$ *and* $\mathrm{ch}(i) = \mathrm{ch}(j) \cup \{j\}$,

Model equivalence via edge flips

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if there is a sequence of super-covered edge flips that transforms $\mathcal{G}$ into $\mathcal{H}$. The edge $i \to j$ is super-covered if:

- $\mathrm{pa}(i) \cup \{i\} = \mathrm{pa}(j)$ *and* $\mathrm{ch}(i) = \mathrm{ch}(j) \cup \{j\}$,
- *for all* $k \in \mathrm{pa}(j)$ *and* $l \in \mathrm{ch}(i)$ *the edge* $k \to l$ *exists in* $\mathcal{G}$,

Model equivalence via edge flips

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if there is a sequence of super-covered edge flips that transforms $\mathcal{G}$ into $\mathcal{H}$. The edge $i \to j$ is super-covered if:

- $\mathrm{pa}(i) \cup \{i\} = \mathrm{pa}(j)$ *and* $\mathrm{ch}(i) = \mathrm{ch}(j) \cup \{j\}$,
- *for all* $k \in \mathrm{pa}(j)$ *and* $l \in \mathrm{ch}(i)$ *the edge* $k \to l$ *exists in* $\mathcal{G}$,
- *if* $i \wedge k$ *or* $k \wedge j$ *(= have a common ancestor in* $\mathcal{G}$*) then* $i \to k$ *or* $k \to j$ *in* $\mathcal{G}$.

Model equivalence via edge flips

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if there is a sequence of super-covered edge flips that transforms $\mathcal{G}$ into $\mathcal{H}$. The edge $i \to j$ is super-covered if:

- $\mathrm{pa}(i) \cup \{i\} = \mathrm{pa}(j)$ *and* $\mathrm{ch}(i) = \mathrm{ch}(j) \cup \{j\}$,
- *for all* $k \in \mathrm{pa}(j)$ *and* $l \in \mathrm{ch}(i)$ *the edge* $k \to l$ *exists in* $\mathcal{G}$,
- *if* $i \wedge k$ *or* $k \wedge j$ *(= have a common ancestor in* $\mathcal{G}$*) then* $i \to k$ *or* $k \to j$ *in* $\mathcal{G}$.

☞ $A \to B \to C$ and $C \to B \to A$ define different Lyapunov models!

Model equivalence via edge flips

Theorem ([ABHM25])

If $\mathcal{G}$ and $\mathcal{H}$ are acyclic then $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{H})$ if and only if there is a sequence of super-covered edge flips that transforms $\mathcal{G}$ into $\mathcal{H}$. The edge $i \to j$ is super-covered if:

- $\mathrm{pa}(i) \cup \{i\} = \mathrm{pa}(j)$ *and* $\mathrm{ch}(i) = \mathrm{ch}(j) \cup \{j\}$,
- *for all* $k \in \mathrm{pa}(j)$ *and* $l \in \mathrm{ch}(i)$ *the edge* $k \to l$ *exists in* $\mathcal{G}$,
- *if* $i \wedge k$ *or* $k \wedge j$ *(= have a common ancestor in $\mathcal{G}$) then* $i \to k$ *or* $k \to j$ *in* $\mathcal{G}$.

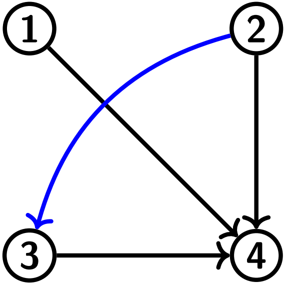
☞ $A \to B \to C$ and $C \to B \to A$ define different Lyapunov models!

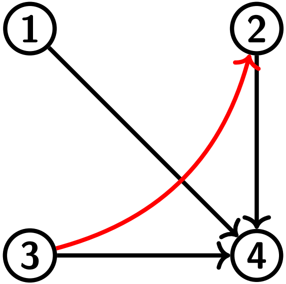
Corollary

A DAG is uniquely identifiable from data (within the class of DAGs) if and only if it does not contain a super-covered edge.

A sparsest (connected) example of non-identifiability

- $\mathrm{pa}(i) \cup \{i\} = \mathrm{pa}(j)$ and $\mathrm{ch}(i) = \mathrm{ch}(j) \cup \{j\}$,
- for all $k \in \mathrm{pa}(j)$ and $l \in \mathrm{ch}(i)$ the edge $k \to l$ exists in $\mathcal{G}$,
- if $i \wedge k$ or $k \wedge j$ (= have a common ancestor in $\mathcal{G}$) then $i \to k$ or $k \to j$ in $\mathcal{G}$.





Reservations and open problems

- We assume that the diffusion matrix D (and hence $C = DD^{\mathsf{T}}$) is diagonal corresponding to independent noise from the Brownian motion.

Reservations and open problems

- We assume that the diffusion matrix D (and hence $C = DD^\top$) is diagonal corresponding to independent noise from the Brownian motion.
- Our characterization holds only within the class of directed acyclic graphs.

Reservations and open problems

- We assume that the diffusion matrix D (and hence $C = DD^\mathsf{T}$) is diagonal corresponding to independent noise from the Brownian motion.
- Our characterization holds only within the class of directed acyclic graphs.
- How to learn the graph in practice from observational data?

Reservations and open problems

- We assume that the diffusion matrix D (and hence $C = DD^{\mathsf{T}}$) is diagonal corresponding to independent noise from the Brownian motion.
- Our characterization holds only within the class of directed acyclic graphs.
- How to learn the graph in practice from observational data?
- What is the asymptotic number of structurally identifiable DAGs? How does it compare to LSEMs?

Reservations and open problems

- We assume that the diffusion matrix D (and hence $C = DD^\top$) is diagonal corresponding to independent noise from the Brownian motion.
- Our characterization holds only within the class of directed acyclic graphs.
- How to learn the graph in practice from observational data?
- What is the asymptotic number of structurally identifiable DAGs? How does it compare to LSEMs?

		Lyapunov DAG models		Bayesian network models	
n	DAGs	Distinct models	Identifiable	Distinct models	Identifiable
3	25	17	13	11	4
4	543	461	423	185	59
5	29 281	27 697	26 761	8 782	2 616
6	3 781 503	3 715 745	3 665 673	1 067 825	306 117

References I

[ABHM25] Carlos Améndola, Tobias Boege, Benjamin Hollering, and Pratik Misra. *Structural Identifiability of Graphical Continuous Lyapunov Models*. 2025. arXiv: 2510.04985 [math.ST].

[AMP97] Steen A. Andersson, David Madigan, and Michael D. Perlman. "A characterization of Markov equivalence classes for acyclic digraphs". In: *The Annals of Statistics* 25.2 (1997), pp. 505–541. ISSN: 0090-5364. DOI: 10.1214/aos/1031833662.

[Chi95] David Maxwell Chickering. "A transformational characterization of equivalent Bayesian network structures". In: *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. 1995, pp. 87–98.

[DHADH23] Philipp Dettling, Roser Homs, Carlos Améndola, Mathias Drton, and Niels Richard Hansen. "Identifiability in Continuous Lyapunov Models". In: *SIAM Journal on Matrix Analysis and Applications* 44.4 (2023), pp. 1799–1821.

[Drt09] Mathias Drton. "Likelihood ratio tests and singularities". In: *Ann. Stat.* 37.2 (2009), pp. 979–1012. ISSN: 0090-5364. DOI: 10.1214/07-AOS571.

References II

[Fit19] Katherine E. Fitch. *Learning Directed Graphical Models from Gaussian Data*. Preprint, arXiv:1906.08050 [cs.LG]. 2019. arXiv: 1906.08050. URL: https://arxiv.org/abs/1906.08050.

[SDL24] Nils Sturma, Mathias Drton, and Dennis Leung. "Testing many constraints in possibly irregular models using incomplete U-statistics". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 86.4 (2024), pp. 987–1012. ISSN: 1369-7412. DOI: 10.1093/jrsssb/qkae022.

[VH20] Gherardo Varando and Niels Richard Hansen. "Graphical continuous Lyapunov models". In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*. Ed. by Jonas Peters and David Sontag. Vol. 124. Proceedings of Machine Learning Research. PMLR, 2020, pp. 989–998.

[VP90] Thomas Verma and Judea Pearl. "Equivalence and synthesis of causal models". In: *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*. UAI '90. Elsevier, 1990, pp. 255–270.