Cyclic Causality: D-separation and Nonlinearity

David Reber

3/16 - 3/24 update

Outline

- To-Do Review
- Motivation and Problem Formulation
- Literature Review
- Proposed Research Path
  - Breaking down the problem
  - Other (research) musings
- Potential Collaborations
- Next tasks

Review: Previous To-do’s (discarded)

Main Objective: Get a good enough lit. review on cyclic causality that I stop being surprised by what I read

- Finish distilling "main highlights" from:
  - Foundations of structural causal models with cycles and latent variables (Bongers, Forré, Peters, Mooij: Oct. 2021)
  - Learning Linear Cyclic Causal Models with Latent Variables (Hyttinen, Eberhardt, Hoyer: 2012)
  - Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning (White and Chalak: 2009)
- Literature review specifically on any nonlinear, cyclic d-separation results
  - Read: Markov Properties for Graphical Models with Cycles and Latent Variables (Forré, Mooij: 2017)
- (Time permitting) Read: Causal Modeling of Dynamical Systems (Bongers, Blom, Mooij: original in 2018, updated Dec. 2021)
- (Time permitting) Switch to "breadth-first" search of literature, only reading abstracts/conclusions
  - with the goal of making sure I've identified all "landmark" papers on cyclic causality

Reprioritized To-do's (after 3/20 emails with Alexis)

Main Objective: **Clarify Problem Statement**

- ✓ Articulate clear problem statement (by Wednesday mtn with Alexis)
- ✓ 'Sanity-check' numerics
- Relate problem to the literature I've already read
- Get a 'workable' mid-semester report
- (if time) look for more readings *directly related to my problem*

Motivation and Problem Formulation

Why (not) Cyclic Causality?

Motivation

- Feedback is everywhere! (game theory, economics, physical sciences)

Why Avoided

- No unique equilibrium -> No potential response function (!)
- The observational / interventional / counterfactual distributions may not exist, or if they do, they may not be unique
- Marginalizing over variables may not be possible, or sensible
- d-separation may not hold (aka. the "directed global Markov property")
    - Or even the weaker variant of σ -separation (the “general directed global Markov property”)
- the induced causal graph may not be consistent with the SCM's causal semantics (!)

A closer look: d-separation

- d-separation underlies much of acyclic causal inference
  - interventional/counterfactual Inference, transportability, structure learning
- But in the cyclic setting, validity of d-separation does not always hold (!)
  - Even after we tweak the definition to make sense for cyclic SCMs (see Appendix)
- Sometimes a weaker variant (σ-separation) may hold for cyclic SCMs

$$\begin{tikzcd} & P_M(V) \arrow[r] & (X \perp\!\!\!\perp Y | Z)_{P_M} \\ M \arrow[ur] \arrow[dr] & & \\ & G_M \arrow[r] & (X \perp\!\!\!\perp Y | Z)_{G_M} \arrow[uu, Rightarrow] \end{tikzcd}$$

directed global Markov property
(“validity of d-separation”)
when M has a unique solution

A closer look: d-separation counterexample

EXAMPLE A.8 (Directed global Markov property does not hold for cyclic SCM). *Consider the SCM $\mathcal{M} = \langle \mathbf{4}, \mathbf{4}, \mathbb{R}^4, \mathbb{R}^4, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^4} \rangle$ with causal mechanism given by*

$$f_1(\boldsymbol{x}, \boldsymbol{e}) = e_1, \quad f_2(\boldsymbol{x}, \boldsymbol{e}) = e_2, \quad f_3(\boldsymbol{x}, \boldsymbol{e}) = x_1 x_4 + e_3, \quad f_4(\boldsymbol{x}, \boldsymbol{e}) = x_2 x_3 + e_4$$

and $\mathbb{P}_{\mathbb{R}^4}$ is the standard-normal distribution on $\mathbb{R}^4$. The graph of $\mathcal{M}$ is depicted in Figure 1 on the left. The model is uniquely solvable (it is even simple). One can check that for every solution $\boldsymbol{X}$ of $\mathcal{M}$, X_1 is not independent of X_2 given $\{X_3, X_4\}$. However, the variables X_1 and X_2 are d-separated given $\{X_3, X_4\}$ in $\mathcal{G}(\mathcal{M})$. Hence the global directed Markov property does not hold here.

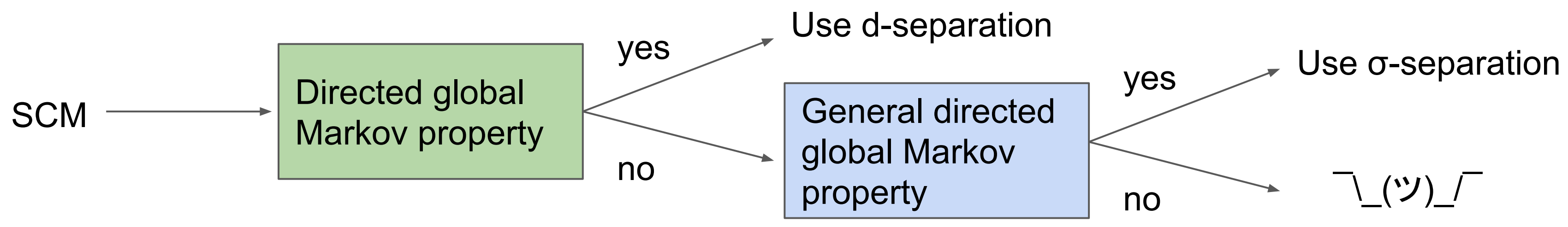
Source: Foundations of structural causal models with cycles and latent variables (Bongers, Forré, Peters, Mooij: Oct. 2021)

Originally from:

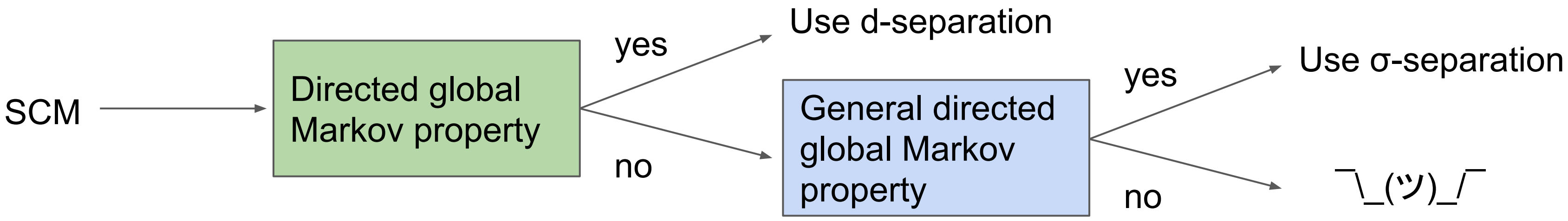
SPIRTES, P. (1994). Conditional Independence in Directed Cyclic Graphical Models for Feedback Technical Report No. CMU-PHIL-54, Carnegie Mellon University.

SPIRTES, P. (1995). Directed Cyclic Graphical Representations of Feedback Models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)* (P. BESNARD and S. HANKS, eds.) 499–506. Morgan Kaufmann, San Francisco, CA, USA.

Beyond d-separation: σ-separation

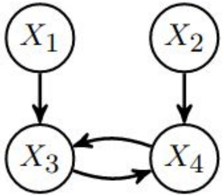


Beyond d-separation: σ-separation

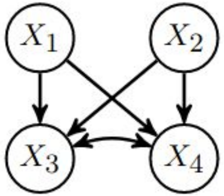


- The directed global Markov property implies the general directed global Markov property
- σ-separation: apply d-separation to the acyclification of the graph

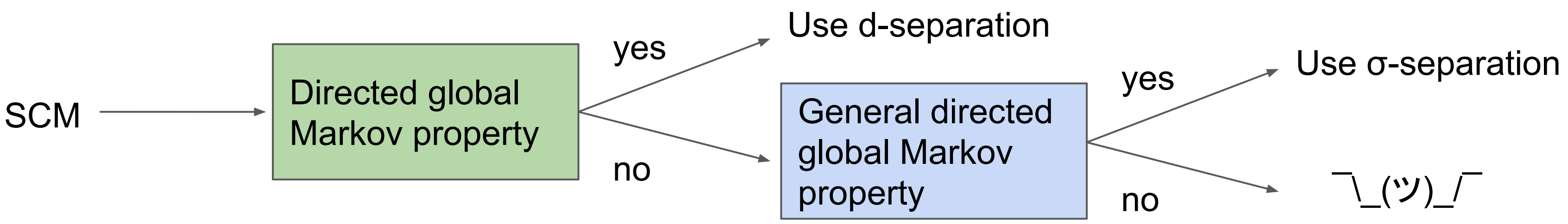
Example A.8



Acyclified graph



d-separation vs. σ-separation



SCMs known to satisfy d-separation:

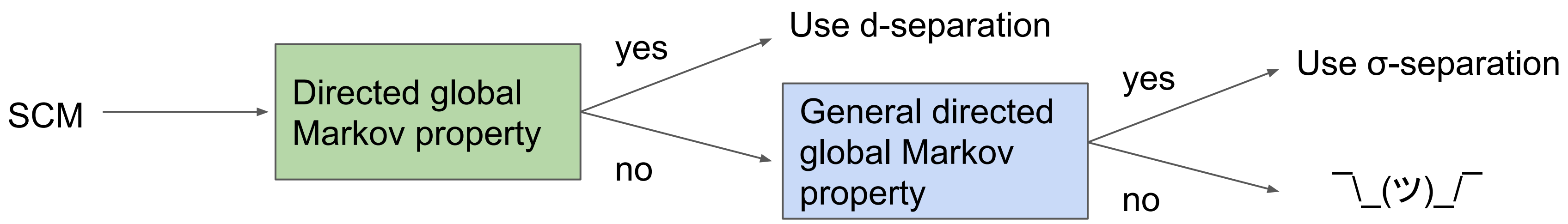
- Acyclic
- Discrete-domain (with ancestrally unique solvability)
- Linear (with non-trivial dependencies and positive measure)

SCMs known to satisfy σ-separation:

- Simple SCMs
- Any SCM “uniquely solvable w.rt. strongly connected components of G(M)”

Source: Foundations of structural causal models with cycles and latent variables (Bongers, Forré, Peters, Mooij: Oct. 2021)

d-separation vs. σ-separation



SCMs known to satisfy d-separation:

- Acyclic
- Discrete-domain (with ancestrally unique solvability)
- Linear (with non-trivial dependencies and positive measure)
- Nonlinear SCMs that behave like linear SCMs asymptotically? (sounds like intrinsic-stability…)

SCMs known to satisfy σ-separation:

- Simple SCMs
- Any SCM “uniquely solvable w.rt. strongly connected components of G(M)”

Source: Foundations of structural causal models with cycles and latent variables (Bongers, Forré, Peters, Mooij: Oct. 2021)

Problem Formulation

Problem: Prove whether intrinsically-stable SCMs satisfy the directed global Markov property.

- (for non-trivial SCMs: cyclic, nonlinear, continuous-domain)

$$
\begin{tikzcd}
 & P_M(V) \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{P_M} \\
M \arrow[ur] \arrow[dr] & & \\
 & G_M \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{G_M} \arrow[uu, Rightarrow]
\end{tikzcd}
$$

Problem Formulation

Problem: Prove whether intrinsically-stable SCMs satisfy the directed global Markov property.

- (for non-trivial SCMs: cyclic, nonlinear, continuous-domain)

Why intrinsically-stable systems?

- Unique equilibrium (the potential response function will be well-defined)
- They ‘behave’ like linear systems (asymptotically)
  - We can go to linear-world, analyze the system there, and the results hold in nonlinear-world
- Closed under a surprising number of structural transformations
  - Lengthening of paths, collapsing/duplication portions of the graph, isospectral transformations, time-varying structural switching,
- Quite general
  - Lipschitz-continuous. Domain: product of metric spaces (like language, shapes, rankings).

Possible Research Path

Numerics

Consider again the counterexample from earlier:

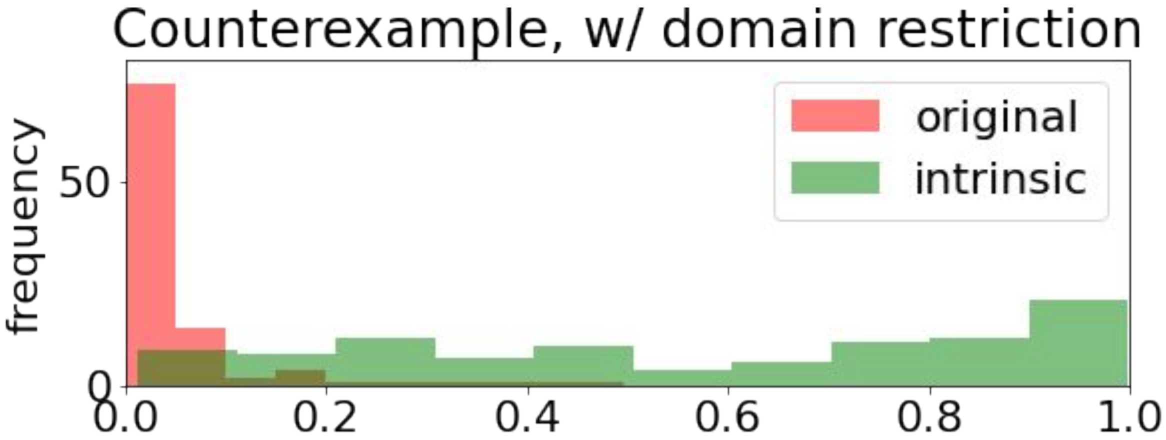
$$f_1(\boldsymbol{x},\boldsymbol{e}) = e_1\,, \quad f_2(\boldsymbol{x},\boldsymbol{e}) = e_2\,, \quad f_3(\boldsymbol{x},\boldsymbol{e}) = x_1 x_4 + e_3\,, \quad f_4(\boldsymbol{x},\boldsymbol{e}) = x_2 x_3 + e_4$$

If when we sample the exogenous variables e from a standard normal distribution over R^4, we restrict them to be in [-0.5,0.5]^4 (dropping samples outside this range), the resulting SCM is intrinsically-stable.

Hence, we should observe that the observational distribution of the restricted-domain SCM should have the conditional independence of (X1 indep X2 | X3, X4).

Numerics

$$f_1(\boldsymbol{x},\boldsymbol{e}) = e_1\,, \quad f_2(\boldsymbol{x},\boldsymbol{e}) = e_2\,, \quad f_3(\boldsymbol{x},\boldsymbol{e}) = x_1 x_4 + e_3\,, \quad f_4(\boldsymbol{x},\boldsymbol{e}) = x_2 x_3 + e_4$$



P-value (fcit package)

The distribution corresponding to the original, unrestricted SCM produces small p-values, consistent with an fcit result of conditional dependance.
Meanwhile, the restricted-domain (and intrinsically stable) SCM appears to have a uniform distribution of p-values over [0,1], consistent with an fcit result of conditional independence.

Possible Research Path: breaking up the problem

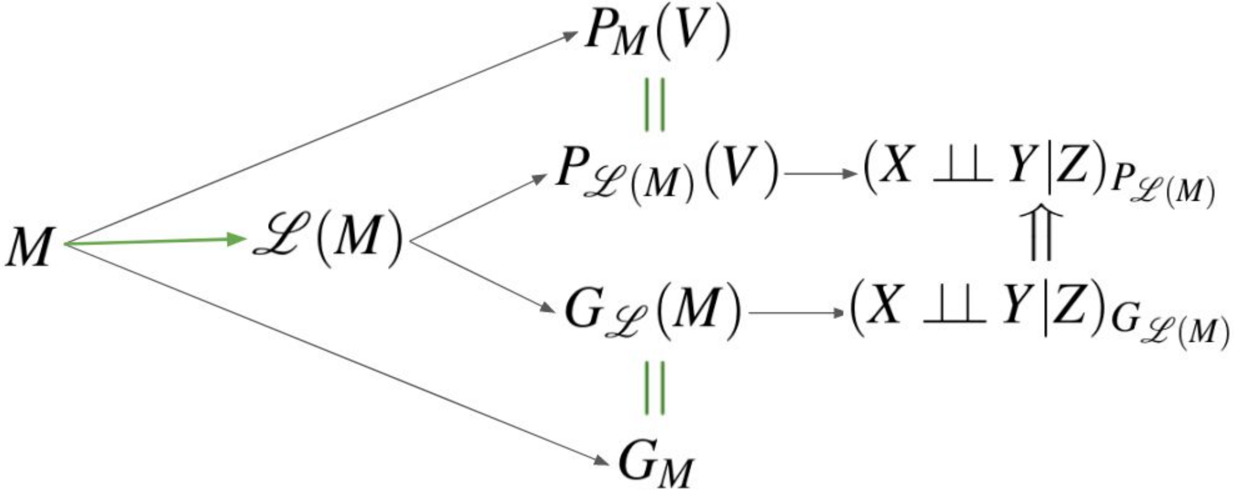
Strategy:

- Relate M to a linear SCM L(M) which satisfies the directed global Markov property.
- Show that M and L(M) respect the same graph, and are observationally equivalent.

Possible Research Path: breaking up the problem

Strategy:

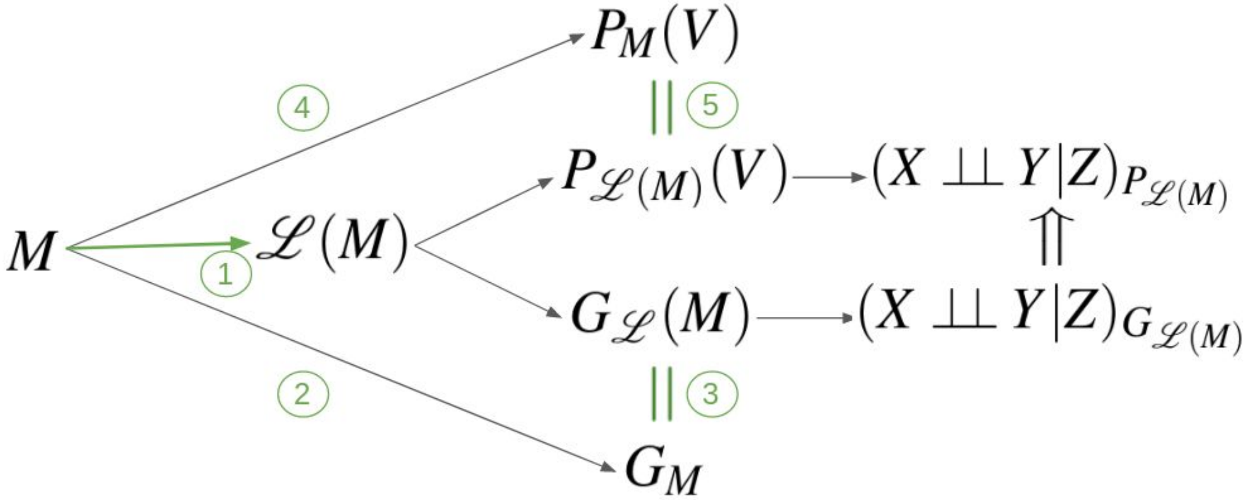
- Relate M to a linear SCM L(M) which satisfies the directed global Markov property.
- Show that M and L(M) respect the same graph, and are observationally equivalent.



Possible Research Path: breaking up the problem

Strategy:

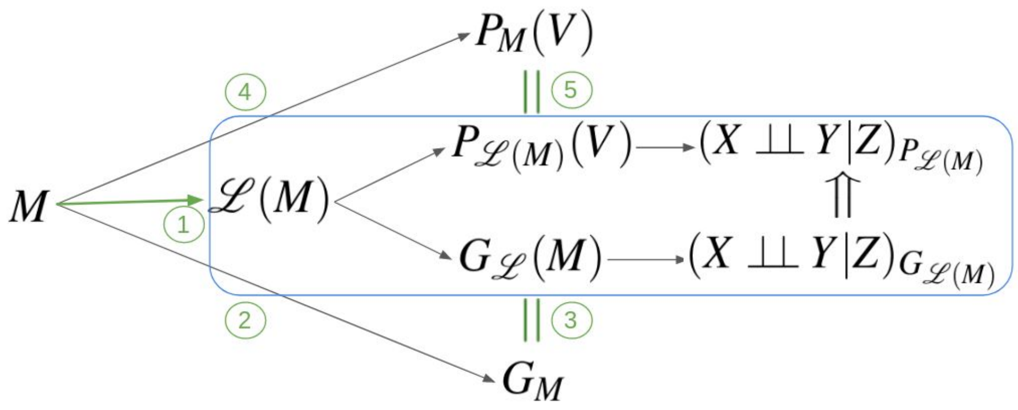
- Relate M to a linear SCM L(M) which satisfies the directed global Markov property.
- Show that M and L(M) respect the same graph, and are observationally equivalent.



Possible Research Path: breaking up the problem

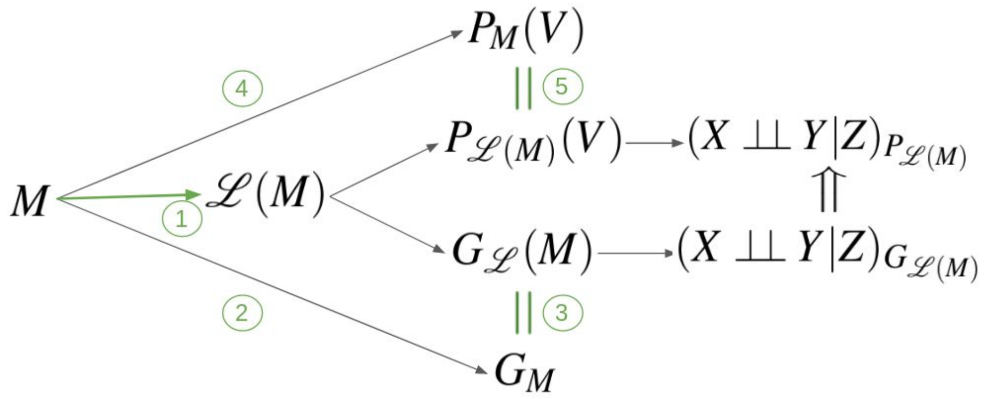
$$
\begin{tikzcd}
 & & P_M(V) \arrow[d, equal, "5"] & \\
 & & P_{\mathscr{L}(M)}(V) \arrow[r] & (X \perp\!\!\!\perp Y | Z)_{P_{\mathscr{L}(M)}} \\
M \arrow[uurr, "4"] \arrow[r, "1"'] \arrow[ddrr, "2"'] & \mathscr{L}(M) \arrow[ur] \arrow[dr] & & \\
 & & G_{\mathscr{L}}(M) \arrow[r] \arrow[d, equal, "3"] & (X \perp\!\!\!\perp Y | Z)_{G_{\mathscr{L}(M)}} \arrow[uu, Rightarrow] \\
 & & G_M &
\end{tikzcd}
$$

Possible Research Path: breaking up the problem



1. M -> L(M): Show that for every intrinsically-stable M, there exists a linear SCM L(M) which satisfies the directed global Markov property (which is the blue box)

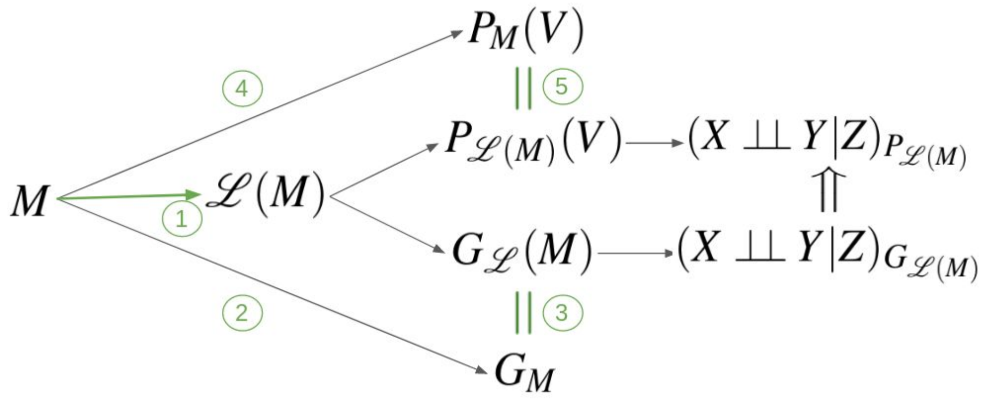
Possible Research Path: breaking up the problem



1. M -> L(M): Show that for every intrinsically-stable M, there exists a linear SCM L(M) which satisfies the directed global Markov property (which is the blue box)

2. M -> G(M): (by definition, I believe)

Possible Research Path: breaking up the problem



1. M -> L(M): Show that for every intrinsically-stable M, there exists a linear SCM L(M) which satisfies the directed global Markov property (which is the blue box)

2. M -> G(M): (by definition, I believe)

3. G_M = G_L(M): Show that for every node, the parents are preserved
   a. Using the definition of parents from “Foundations…”

Possible Research Path: breaking up the problem

$$\begin{tikzcd}
 & & P_M(V) \arrow[d, equal, "5"] & \\
 & & P_{\mathscr{L}(M)}(V) \arrow[r] & (X \perp\!\!\!\perp Y | Z)_{P_{\mathscr{L}(M)}} \\
M \arrow[uurr, "4"] \arrow[r, "1"'] \arrow[ddrr, "2"'] & \mathscr{L}(M) \arrow[ur] \arrow[dr] & & \\
 & & G_{\mathscr{L}}(M) \arrow[r] \arrow[d, equal, "3"] & (X \perp\!\!\!\perp Y | Z)_{G_{\mathscr{L}(M)}} \arrow[uu, Rightarrow] \\
 & & G_M &
\end{tikzcd}$$

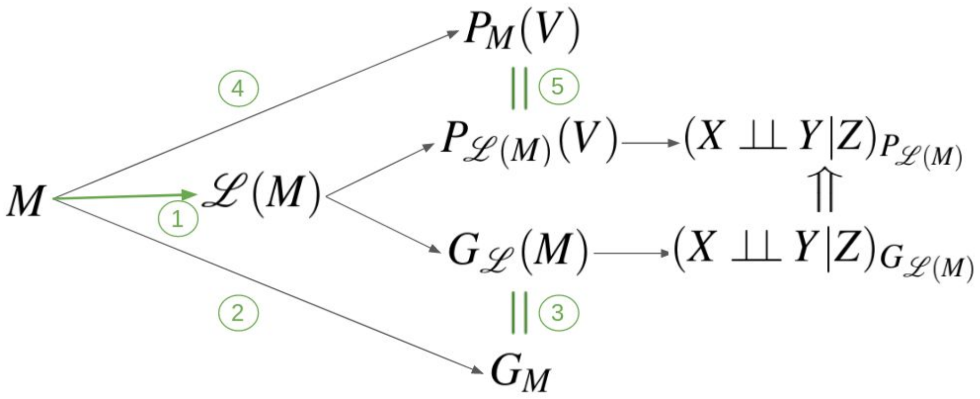
1. M -> L(M): Show that for every intrinsically-stable M, there exists a linear SCM L(M) which satisfies the directed global Markov property (which is the blue box)

2. M -> G(M): (by definition, I believe)

3. G_M = G_L(M): Show that for every node, the parents are preserved
   a. Using the definition of parents from “Foundations…”

4. M -> P_M(V): Show that if M is intrinsically stable, it is uniquely solvable

Possible Research Path: breaking up the problem



1. M -> L(M): Show that for every intrinsically-stable M, there exists a linear SCM L(M) which satisfies the directed global Markov property (which is the blue box)

2. M -> G(M): (by definition, I believe)

3. G_M = G_L(M): Show that for every node, the parents are preserved
   a. Using the definition of parents from “Foundations…”

4. M -> P_M(V): Show that if M is intrinsically stable, it is uniquely solvable

5. P_M(V) = P_L(M)(V)
   a. Translate previous result of M and L(M) having the same equilibrium solutions, to SCMs
   b. Show that this means they induce the same obs. distributions.

Other (research) musings

Validity of d-separation beyond Obs. graph/distribution?

$$
\begin{tikzcd}
 & P_M(V) \arrow[r] & (X \perp\!\!\!\perp Y | Z)_{P_M} \\
M \arrow[ur] \arrow[dr] & & \\
 & G_M \arrow[r] & (X \perp\!\!\!\perp Y | Z)_{G_M} \arrow[uu, Rightarrow]
\end{tikzcd}
$$

$$
\begin{tikzcd}
 & & P_M(V) \arrow[d, equal] & \\
 & & P_{\mathscr{L}(M)}(V) \arrow[r] & (X \perp\!\!\!\perp Y | Z)_{P_{\mathscr{L}(M)}} \\
M \arrow[r] \arrow[uurr] \arrow[ddrr] & \mathscr{L}(M) \arrow[ur] \arrow[dr] & & \\
 & & G_{\mathscr{L}}(M) \arrow[r] \arrow[d, equal] & (X \perp\!\!\!\perp Y | Z)_{G_{\mathscr{L}(M)}} \arrow[uu, Rightarrow] \\
 & & G_M &
\end{tikzcd}
$$

- Will these conjectures still hold if we replace M with a submodel corresponding to an intervention (or a counterfactual)?

Validity of d-separation beyond Obs. graph/distribution?

$$
\begin{tikzcd}
& P_M(V) \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{P_M} \\
M \arrow[ur] \arrow[dr] & & \\
& G_M \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{G_M} \arrow[uu, Rightarrow]
\end{tikzcd}
$$

$$
\begin{tikzcd}
& & P_M(V) \arrow[d, equal] & \\
& & P_{\mathscr{L}(M)}(V) \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{P_{\mathscr{L}(M)}} \\
M \arrow[uurr] \arrow[r] \arrow[ddrr] & \mathscr{L}(M) \arrow[ur] \arrow[dr] & & \\
& & G_{\mathscr{L}}(M) \arrow[r] \arrow[d, equal] & (X \perp\!\!\!\perp Y|Z)_{G_{\mathscr{L}(M)}} \arrow[uu, Rightarrow] \\
& & G_M &
\end{tikzcd}
$$

- Will these conjectures still hold if we replace M with a submodel corresponding to an intervention (or a counterfactual)?
- I'm quite confident that the space of intrinsic SCMs is closed under interventions (if you intervene on an intrinsic SCM, the resulting SCM is also intrinsic)

Validity of d-separation beyond Obs. graph/distribution?

$$\begin{tikzcd}
& P_M(V) \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{P_M} \\
M \arrow[ur] \arrow[dr] & & \\
& G_M \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{G_M} \arrow[uu, Rightarrow]
\end{tikzcd}$$

$$\begin{tikzcd}
& & P_M(V) \arrow[d, equal] & \\
& & P_{\mathscr{L}(M)}(V) \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{P_{\mathscr{L}(M)}} \\
M \arrow[r] \arrow[uurr] \arrow[ddrr] & \mathscr{L}(M) \arrow[ur] \arrow[dr] & & \\
& & G_{\mathscr{L}}(M) \arrow[r] \arrow[d, equal] & (X \perp\!\!\!\perp Y|Z)_{G_{\mathscr{L}(M)}} \arrow[uu, Rightarrow] \\
& & G_M &
\end{tikzcd}$$

- Will these conjectures still hold if we replace M with a submodel corresponding to an intervention (or a counterfactual)?

- I’m quite confident that the space of intrinsic SCMs is closed under interventions (if you intervene on an intrinsic SCM, the resulting SCM is also intrinsic)
- I’m also confident that intervening and linearizing commute for intrinsic SCMs.

Validity of d-separation beyond Obs. graph/distribution?

$$\begin{tikzcd}
 & P_M(V) \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{P_M} \\
M \arrow[ur] \arrow[dr] & & \\
 & G_M \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{G_M} \arrow[uu, Rightarrow]
\end{tikzcd}$$

$$\begin{tikzcd}
 & & P_M(V) \arrow[d, equal] & \\
 & & P_{\mathscr{L}(M)}(V) \arrow[r] & (X \perp\!\!\!\perp Y|Z)_{P_{\mathscr{L}(M)}} \\
M \arrow[r] \arrow[uurr] \arrow[ddrr] & \mathscr{L}(M) \arrow[ur] \arrow[dr] & & \\
 & & G_{\mathscr{L}}(M) \arrow[r] \arrow[d, equal] & (X \perp\!\!\!\perp Y|Z)_{G_{\mathscr{L}(M)}} \arrow[uu, Rightarrow] \\
 & & G_M &
\end{tikzcd}$$

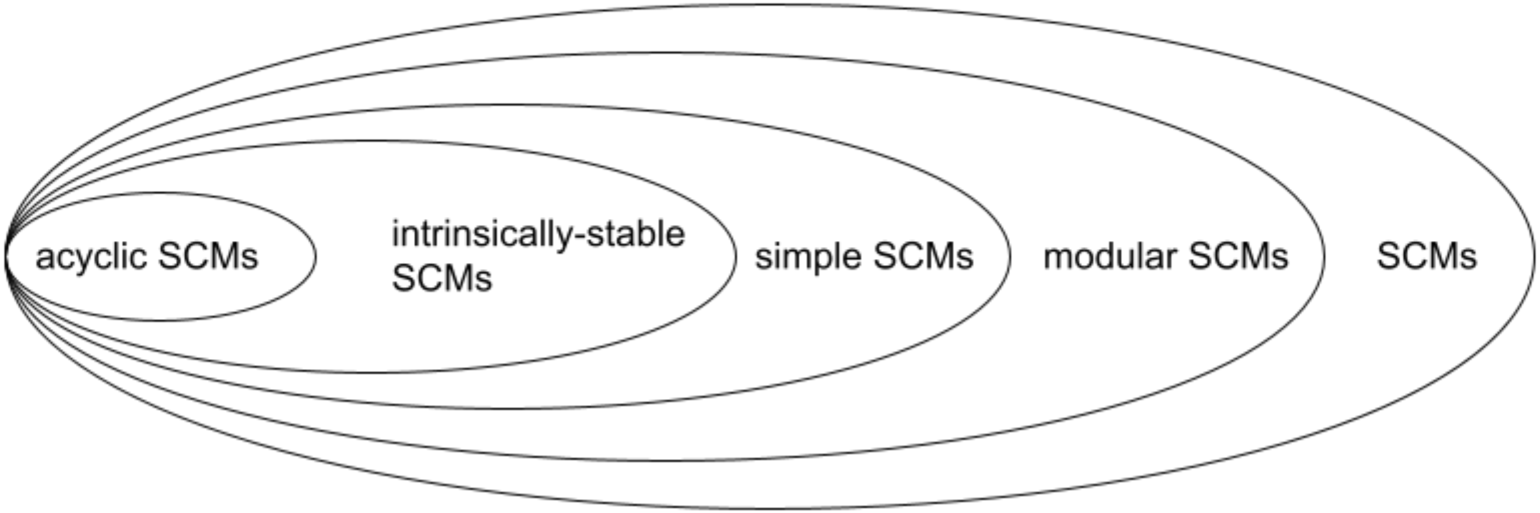
- Will these conjectures still hold if we replace M with a submodel corresponding to an intervention (or a counterfactual)?

- I'm quite confident that the space of intrinsic SCMs is closed under interventions (if you intervene on an intrinsic SCM, the resulting SCM is also intrinsic)
- I'm also confident that intervening and linearizing commute for intrinsic SCMs.

- I conjecture that intrinsic SCMs are also closed under the "mirror" transformation (necessary for defining cyclic counterfactuals).

Conjectured relation to other spaces of cyclic SCMs

- Intrinsic SCMs contain the space acyclic SCMs (so long as we tweak the definition slightly)
- I conjecture that intrinsic SCMs are a subset of simple SCMs



Potential (future) Collaborations: Game Theory + Causality

- Lewis Hammond (Oxford)
  - Research: Causality + Game Theory
  - Met (remotely) at Deepmind causal influence diagram, research agenda meeting (3/15/2022)
    - + Tom Everitt (Deepmind), Ryan Carey (Oxford) and a few others
  - Shared paper “Reasoning about Causality in Games” (submitted to AIJ)
    - Introduces “Structural causal games” framework
    - Extends Pearl’s causal hierarchy to the game-theoretic domain
- Christian Kroer (Columbia, IOER department)
  - Research: Game Theory (also, scalable optimization methods)
  - Interested in `Game theory + Causality’
  - Meet again in a couple of weeks to discuss further
- Relevance?
  - One of the primary motivations for cyclic causality
  - …Not directly relevant for current research direction (unless ‘intrinsic’ results hold locally?)

Next To-Do's

- 3/24 - 3/25
  - Relate problem to the literature I've already read
  - Writeup, submit mid-semester report
- 3/27 - 3/30
  - Reexamine definitions/theorems for cyclic SCMs and the directed global Markov property
    - I've already encountered several subtle but important differences in definitions compared to the acyclic setting, and want to ensure I caught them all
  - Translate "intrinsic systems" to "intrinsic SCM" notation…anything break?
  - (if time) either unit-test previous numerics, or come up with new numerics
    - Since the validity of the numerics is a crux for pursuing this direction

(It'll be a short week for me as I'll be out of town 3/31 - 4/3)

Appendix

d-separation for cyclic SCMs (1 of 2)

DEFINITION A.3 (Collider). *Let* $\pi = (i_0, \epsilon_1, i_1, \epsilon_2, i_2, \ldots, \epsilon_n, i_n)$ *be a walk (path) in a directed mixed graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$. *A node* i_k *on* π *is called a* collider on π *if it is a non-endpoint node* ($1 \leq k < n$) *and the two edges* $\epsilon_k, \epsilon_{k+1}$ *meet head-to-head on* i_k *(i.e., if the subwalk* $(i_{k-1}, \epsilon_k, i_k, \epsilon_{k+1}, i_{k+1})$ *is of the form* $i_{k-1} \to i_k \leftarrow i_{k+1}$, $i_{k-1} \leftrightarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \to i_k \leftrightarrow i_{k+1}$ *or* $i_{k-1} \leftrightarrow i_k \leftrightarrow i_{k+1}$*). The node* i_k *is called a* non-collider on π *otherwise, that is, if it is an endpoint node* ($k = 0$ *or* $k = n$) *or if the subwalk* $(i_{k-1}, \epsilon_k, i_k, \epsilon_{k+1}, i_{k+1})$ *is of the form* $i_{k-1} \to i_k \to i_{k+1}$, $i_{k-1} \leftarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \leftarrow i_k \to i_{k+1}$, $i_{k-1} \leftrightarrow i_k \to i_{k+1}$ *or* $i_{k-1} \leftarrow i_k \leftrightarrow i_{k+1}$.

Source: Foundations of structural causal models with cycles and latent variables (Bongers, Forré, Peters, Mooij: Oct. 2021)

d-separation for cyclic SCMs (2 of 2)

DEFINITION A.4 (d-separation). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph and let $C \subseteq \mathcal{V}$ be a subset of nodes. A walk (path) $\pi = (i_0, \epsilon_1, i_1, \dots, i_n)$ in $\mathcal{G}$ is said to be C-d-*blocked or d-blocked by C *if*

1. *it contains a collider $i_k \notin \mathrm{an}_{\mathcal{G}}(C)$, or*
2. *it contains a non-collider $i_k \in C$.*

*The walk (path) π is said to be C-d-*open *if it is not d-blocked by C. For two subsets of nodes $A, B \subseteq \mathcal{V}$, we say that A* is d-separated from B given C in $\mathcal{G}$ *if all paths between any node in A and any node in B are d-blocked by C, and write*

$$A \underset{\mathcal{G}}{\overset{d}{\perp}} B \mid C \,.$$

Source: Foundations of structural causal models with cycles and latent variables (Bongers, Forré, Peters, Mooij: Oct. 2021)

Directed global Markov property (validity of d-separation)

DEFINITION A.6 (Directed global Markov property). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph and $\mathbb{P}_\mathcal{V}$ a probability distribution on $\boldsymbol{\mathcal{X}}_\mathcal{V} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$, where each $\mathcal{X}_i$ is a standard probability space. The probability distribution $\mathbb{P}_\mathcal{V}$ satisfies the* directed global Markov property *relative to $\mathcal{G}$ if for all subsets $A, B, C \subseteq \mathcal{V}$ we have*

$$A \underset{\mathcal{G}}{\overset{d}{\perp}} B \,|\, C \quad \Longrightarrow \quad \boldsymbol{X}_A \underset{\mathbb{P}_\mathcal{V}}{\perp\!\!\!\perp} \boldsymbol{X}_B \,|\, \boldsymbol{X}_C \,,$$

that is, $(X_i)_{i \in A}$ and $(X_i)_{i \in B}$ are conditionally independent given $(X_i)_{i \in C}$ under $\mathbb{P}_\mathcal{V}$, where we take the canonical projections $X_i : \boldsymbol{\mathcal{X}}_\mathcal{V} \to \mathcal{X}_i$ as random variables.

Source: Foundations of structural causal models with cycles and latent variables (Bongers, Forré, Peters, Mooij: Oct. 2021)