stable-Lipschitz SCMs

d-separation and Cyclic Causality

David Reber

Outline

- Introduction
  - Motivation and Challenges
  - Problem Statement
  - Framework Selection
  - Overview of Main Results
- Intuition behind Dynamics and Potential Responses
  - 3 examples of increasing complexity
- Case Study for d-separation
  - Why d-separation fails, from several perspectives
  - Example of the main result
- Results and Future Work
  - Properties of stable-Lipschitz SCMs
  - Observational/Interventional validity of d-separation
  - Supporting Numerics
  - Future directions

Why model cyclic causality?

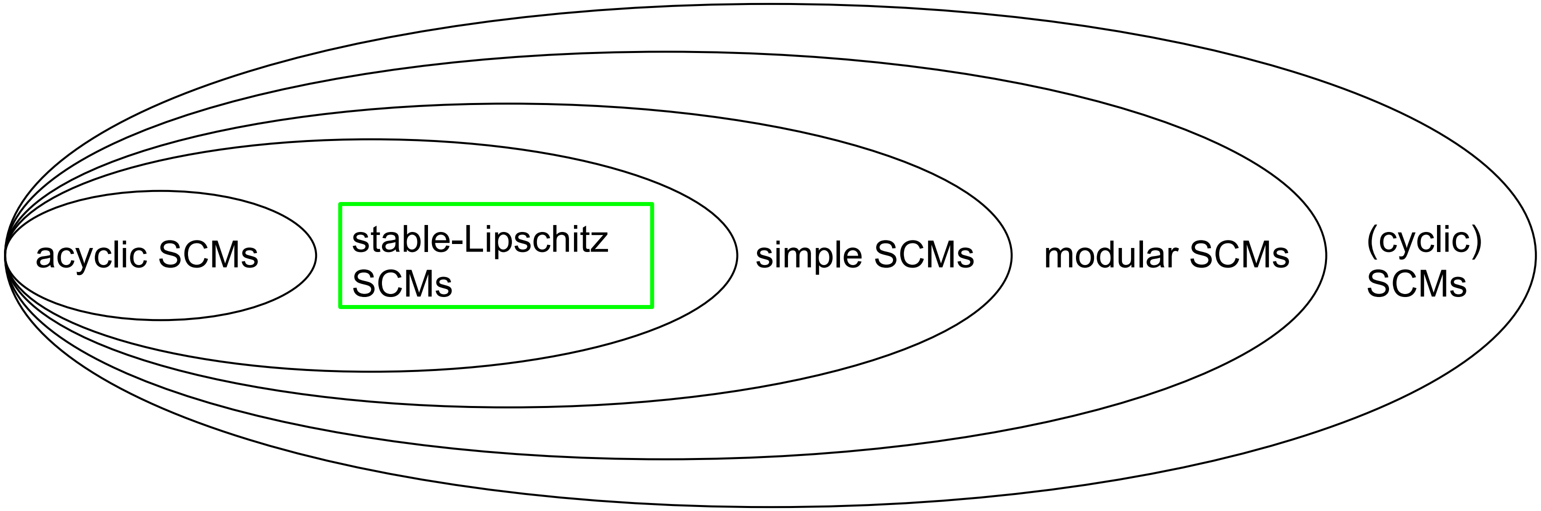
- Feedback is everywhere!
  - Game Theory
  - Economics
  - Neuroscience
  - Physical Sciences
- Creating acyclic models for cyclic phenomena is…challenging
  - Amounts to trying to guess the analytic solution to a differential equation straight from data…without even asking what kind of differential equation it is.
    - Even if it’s possible to fit a square peg in a round hole, it’s not easy; which is yet another hurdle a practitioner must overcome
  - *Maybe if you zoom in enough, everything looks acyclic?*
    - High resolution (in time, or in space) is computationally expensive
      - …if that data is even possible to gather

Why (not) Cyclic Causality?

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 (!)

Big Picture: Which generalizations of **acyclic SCMs** preserve their nice properties?



Problem Statement

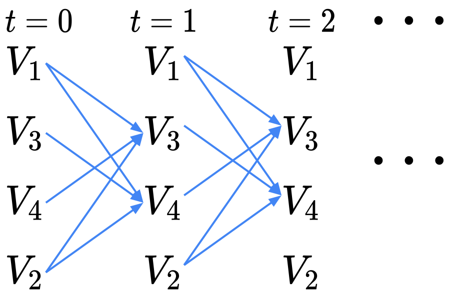
Which cyclic SCMs have the Directed Global Markov Property? ("Validity of d-separation")

SCM	d-separation valid?
Acylic	Yes
Discrete	Yes
Linear	Yes
Nonlinear, Continuous	Not in general (non-Lipschitz counterexample)
Lipschitz, Continuous	????

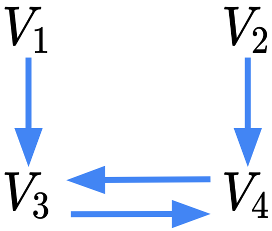
Importantly, we'd like the class to be closed under marginalizations and interventions

- so we can perform do-calculus

Modeling Options for Cyclic Causality



- Unravel with time
  - Low-level computational perspective
  - Analysis breaks down if sampling frequency << feedback frequency
    - [Hyttinen et. al]
  - Less compact
- Settable Systems [White&Chalak 2009]
  - Leans towards low-level optimization perspective
    - Most useful when you know the optimization process
  - Introduce new variables to index multiple potential responses
  - Applicable for game theory [White, Xu, Chalak]
- Cyclic SCMs [Bongers et al. 2021]
  - More abstract representations (for better and for worse)
    - Easier to show inheritance of properties
    - Cumbersome almost-everywhere equivalence
      - Only use ‘structurally minimal’ SCMs, no trivial parents, etc.
  - Very similar to acyclic SCM framework of [Pearl 2009]
  - Applications to econometrics, game theory



Status of Main Results

- Theorems
  - acyclic SCMs $\subset$ stable-Lipschitz SCMs $\subset$ simple SCMs
  - stable-Lipschitz SCMs are closed under interventions
  - **Obs. D-separation:** The observational distribution of stable-Lipschitz SCMs satisfies dGMP
  - Validity of Backdoor Criterion on stable-Lipschitz SCMs
  - **Int. D-separation:** The interventional distribution of stable-Lipschitz SCMs satisfies dGMP
  - stable-Lipschitz SCMs are closed under twin-operation
- Numerics
  - Obs. D-separation (above)
  - The observational distribution of **(non-stable)** Lipschitz SCMs also satisfy dGMP

dGMP := Directed Global Markov Property (a.k.a. "Validity of d-separation")

Outline: Intuition behind Dynamics and Potential Responses

As we gradually increase the complexity of the SCM:

- Definitions
  - Unique Solvability
  - Potential Response
- Acyclic + Linear
- Cyclic + Linear
  - Pearl Causal Hierarchy
  - Lipschitz matrix
  - stable-Lipschitz SCMs
- Cyclic + Nonlinear
  - d-separation

[Bongers et al. 2021]

Unique Solvability

Definition 1 (Unique Solvability). *Let $M = \langle V, U, F, P(U) \rangle$ be an SCM and $\mathbf{Z} \subseteq \mathbf{V}$ a subset of endogenous variables. We say that M is uniquely solvable if for almost every $\mathbf{u} \in dom(\mathbf{U})$ the equations*

$$\mathbf{V} = F(\mathbf{V}, \mathbf{u})$$

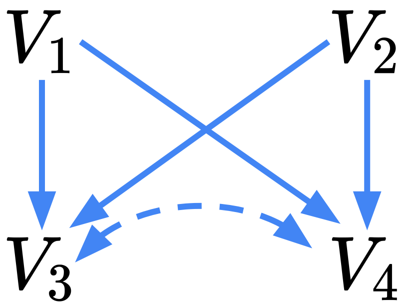
have a unique solution.

Potential Response

Definition 2 (Potential Response)**.** *Let* $M = \langle V, U, F, P(U) \rangle$ *be a uniquely solvable SCM. The potential response function is the mapping* $\overline{\mathbf{V}}(\mathbf{U}) : dom(\mathbf{U}) \rightarrow dom(\mathbf{V})$ *which associates each* $\mathbf{u} \in dom(\mathbf{U})$ *with the unique solution* $\mathbf{v}^*$ *of* $\mathbf{V} = F(\mathbf{V}, \mathbf{u})$.

Acyclic + Linear SCM (1 of 3)

$$M_1 \begin{cases} \mathbf{u}, \mathbf{v} \in \mathbb{R}^4 \\ U_i \sim \mathcal{N}(0,1) \quad iid. \\ f_1 : V_1 \leftarrow U_1 \\ f_2 : V_2 \leftarrow U_2 \\ f_3 : V_3 \leftarrow \frac{2}{3}V_1 + \frac{1}{3}V_2 + \frac{4}{3}U_3 + \frac{2}{3}U_4 \\ f_4 : V_4 \leftarrow \frac{1}{3}V_1 + \frac{2}{3}V_2 + \frac{2}{3}U_3 + \frac{4}{3}U_4 \end{cases}$$



Sample $U_i \sim \mathcal{N}(0,1)$

For every initialization $\mathbf{V}_0$

Evolve until convergence

$$\overline{\mathbf{V}}(\mathbf{u}) = \lim_{k\to\infty} F^{(k)}_{\mathbf{U}=\mathbf{u}}(\mathbf{V}_0)$$

Potential Response

Dynamics

(For acyclic SCMs, convergence occurs within a single timestep, and the initialization is immediately “washed out”)

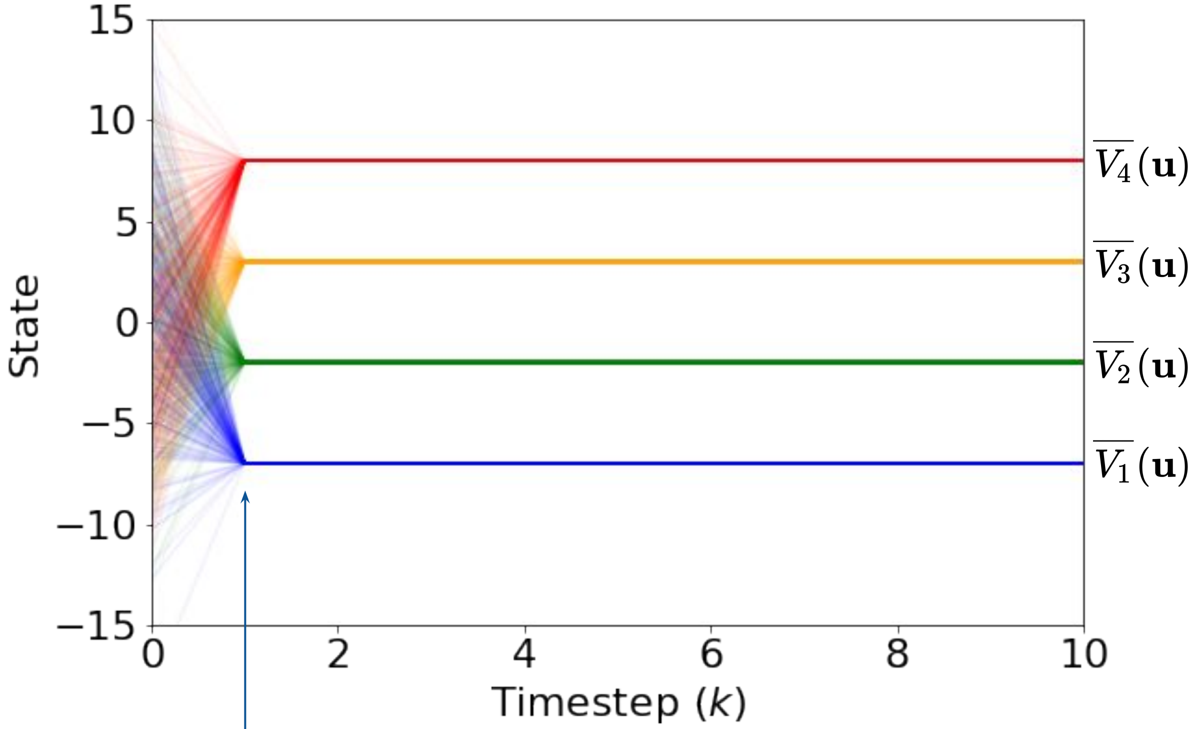
Acyclic + Linear SCM (2 of 3)

Fixed $\mathbf{u} = (-7, -2, 2.5, 7.5)$

$$\overline{\mathbf{V}}(\mathbf{u}) = \lim_{k\to\infty} F^{(k)}_{\mathbf{U}=\mathbf{u}}(\mathbf{V}_0)$$

Potential Response

Dynamics



Converges within 1 timestep, no matter the initialization

Acyclic + Linear SCM (3 of 3)

$$M_1 \quad \xrightarrow{k \to \infty} \quad \text{Potential Response}$$

M_1	Potential Response
$f_1 : V_1 \leftarrow U_1$	$\overline{V_1}(\mathbf{U}) = U_1$
$f_2 : V_2 \leftarrow U_2$	$\overline{V_2}(\mathbf{U}) = U_2$
$f_3 : V_3 \leftarrow \frac{2}{3}V_1 + \frac{1}{3}V_2 + \frac{4}{3}U_3 + \frac{2}{3}U_4$	$\overline{V_3}(\mathbf{U}) = \frac{2}{3}U_1 + \frac{1}{3}U_2 + \frac{4}{3}U_3 + \frac{2}{3}U_4$
$f_4 : V_4 \leftarrow \frac{1}{3}V_1 + \frac{2}{3}V_2 + \frac{2}{3}U_3 + \frac{4}{3}U_4$	$\overline{V_4}(\mathbf{U}) = \frac{1}{3}U_1 + \frac{2}{3}U_2 + \frac{2}{3}U_3 + \frac{4}{3}U_4$

Cyclic + Linear SCM (1 of 3)

$$M_2 \begin{cases} \mathbf{u}, \mathbf{v} \in \mathbb{R}^4 \\ U_i \sim \mathcal{N}(0,1) \quad iid. \\ f_1 : V_1 \leftarrow U_1 \\ f_2 : V_2 \leftarrow U_2 \\ f_3 : V_3 \leftarrow \frac{1}{2}V_1 + \frac{1}{2}V_4 + U_3 \\ f_4 : V_4 \leftarrow \frac{1}{2}V_2 + \frac{1}{2}V_3 + U_4 \end{cases}$$

$$\begin{tikzcd} V_1 \arrow[d] & V_2 \arrow[d] \\ V_3 \arrow[r, shift right] & V_4 \arrow[l, shift right] \end{tikzcd}$$

Spoiler alert: $P_{M_1}(\mathbf{V}) = P_{M_2}(\mathbf{V})$!
We'll see why later.

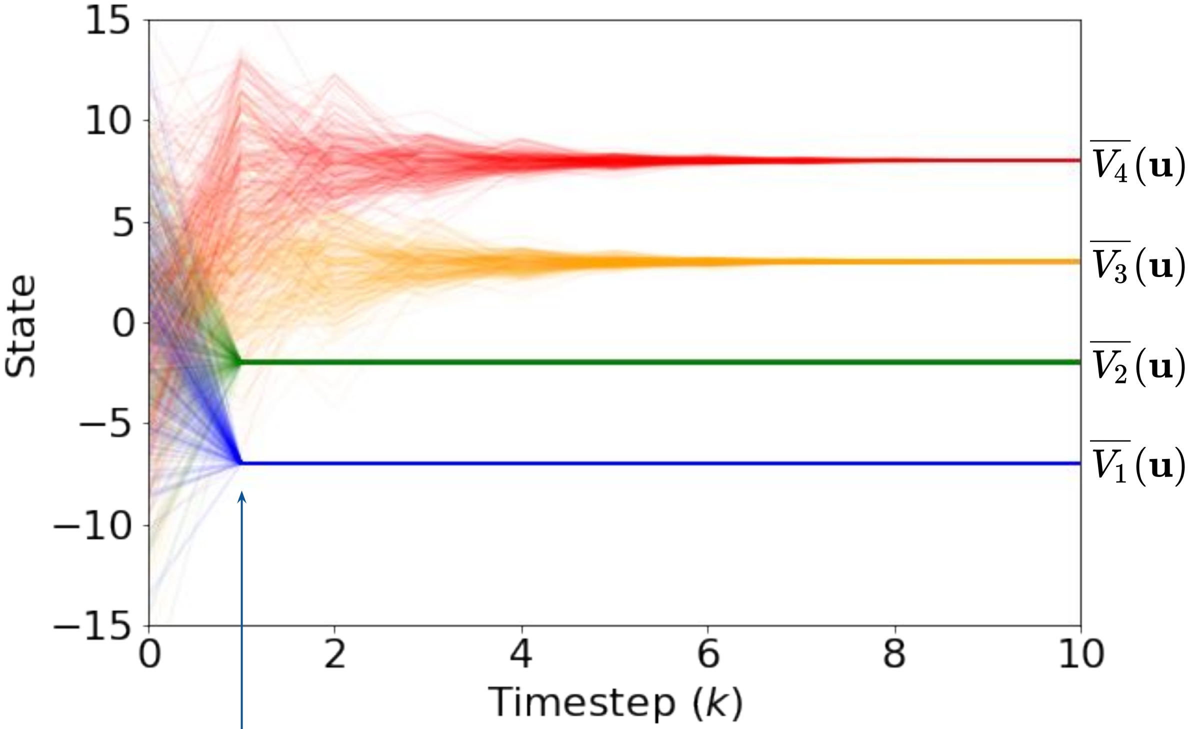
Cyclic + Linear SCM (2 of 3)

Fixed $\mathbf{u} = (-7, -2, 2.5, 7.5)$

$$\overline{\mathbf{V}}(\mathbf{u}) = \lim_{k\to\infty} F^{(k)}_{\mathbf{U}=\mathbf{u}}(\mathbf{V}_0)$$

Potential Response

Dynamics



Nodes 1, 2 converge immediately (as if acyclic)

Cyclic + Linear SCM (3 of 3)

$$M_2 \qquad \xrightarrow{k \to \infty} \qquad \text{Potential Response}$$

$$f_1 : V_1 \leftarrow U_1$$
$$f_2 : V_2 \leftarrow U_2$$
$$f_3 : V_3 \leftarrow \frac{1}{2}V_1 + \frac{1}{2}V_4 + U_3$$
$$f_4 : V_4 \leftarrow \frac{1}{2}V_2 + \frac{1}{2}V_3 + U_4$$

$$\overline{V_1}(\mathbf{U}) = U_1$$
$$\overline{V_2}(\mathbf{U}) = U_2$$
$$\overline{V_3}(\mathbf{U}) = \frac{2}{3}U_1 + \frac{1}{3}U_2 + \frac{4}{3}U_3 + \frac{2}{3}U_4$$
$$\overline{V_4}(\mathbf{U}) = \frac{1}{3}U_1 + \frac{2}{3}U_2 + \frac{2}{3}U_3 + \frac{4}{3}U_4$$

By construction this matches the potential response of M_1

If an SCM is linear, so is its potential response: $F(\mathbf{V}) = A\mathbf{V} + \mathbf{U}$

$$\Rightarrow \overline{\mathbf{V}} = (I - A)^{-1}\mathbf{U}$$

Pearl Causal Hierarchy (informal)

$$M_1$$

$$f_1 : V_1 \leftarrow U_1$$
$$f_2 : V_2 \leftarrow U_2$$
$$f_3 : V_3 \leftarrow \tfrac{2}{3}V_1 + \tfrac{1}{3}V_2 + \tfrac{4}{3}U_3 + \tfrac{2}{3}U_4$$
$$f_4 : V_4 \leftarrow \tfrac{1}{3}V_1 + \tfrac{2}{3}V_2 + \tfrac{2}{3}U_3 + \tfrac{4}{3}U_4$$

$$M_2$$

$$f_1 : V_1 \leftarrow U_1$$
$$f_2 : V_2 \leftarrow U_2$$
$$f_3 : V_3 \leftarrow \tfrac{1}{2}V_1 + \tfrac{1}{2}V_4 + U_3$$
$$f_4 : V_4 \leftarrow \tfrac{1}{2}V_2 + \tfrac{1}{2}V_3 + U_4$$

Potential Response (both)

$$\overline{V_1}(\mathbf{U}) = U_1$$
$$\overline{V_2}(\mathbf{U}) = U_2$$
$$\overline{V_3}(\mathbf{U}) = \tfrac{2}{3}U_1 + \tfrac{1}{3}U_2 + \tfrac{4}{3}U_3 + \tfrac{2}{3}U_4$$
$$\overline{V_4}(\mathbf{U}) = \tfrac{1}{3}U_1 + \tfrac{2}{3}U_2 + \tfrac{2}{3}U_3 + \tfrac{4}{3}U_4$$

By construction, same for both

Pearl Causal Hierarchy (informal)

$$M_1$$

$$\begin{aligned} f_1 &: V_1 \leftarrow U_1 \\ f_2 &: V_2 \leftarrow U_2 \\ f_3 &: V_3 \leftarrow x \\ f_4 &: V_4 \leftarrow \tfrac{1}{3}V_1 + \tfrac{2}{3}V_2 + \tfrac{2}{3}U_3 + \tfrac{4}{3}U_4 \end{aligned}$$

$$M_2$$

$$\begin{aligned} f_1 &: V_1 \leftarrow U_1 \\ f_2 &: V_2 \leftarrow U_2 \\ f_3 &: V_3 \leftarrow x \\ f_4 &: V_4 \leftarrow \tfrac{1}{2}V_2 + \tfrac{1}{2}V_3 + U_4 \end{aligned}$$

Potential Responses under intervention

$$\begin{aligned} \overline{V_1}(\mathbf{U}) &= U_1 \\ \overline{V_2}(\mathbf{U}) &= U_2 \\ \overline{V_3}(\mathbf{U}) &= x \\ \overline{V_4}(\mathbf{U}) &= \tfrac{1}{3}U_1 + \tfrac{2}{3}U_2 + \tfrac{2}{3}U_3 + \tfrac{4}{3}U_4 \end{aligned}$$

Not equal under interventions!

$$\begin{aligned} \overline{V_1}(\mathbf{U}) &= U_1 \\ \overline{V_2}(\mathbf{U}) &= U_2 \\ \overline{V_3}(\mathbf{U}) &= x \\ \overline{V_4}(\mathbf{U}) &= \tfrac{1}{2}U_1 + \tfrac{1}{2}x + U_4 \end{aligned}$$

Lipschitz Matrix

Definition 4 (Lipschitz Matrix). *Let* $M = \langle V, U, F, P(U) \rangle$ *be an SCM, with* $dom(\mathbf{U}) = \mathbb{R}^m$, $dom(\mathbf{V}) = \mathbb{R}^n$, *and* $F : dom(\mathbf{U}) \times dom(\mathbf{V}) \rightarrow dom(\mathbf{V})$ *differentiable and Lipschitz.*

Let $\mathbf{Z} \subseteq \mathbf{V}$ *be a subset of endogenous variables. For each pair of vertex indices* i, $j \in \mathbf{Z}$, *define*

$$a_{ij} = \sup_{\mathbf{u},\mathbf{v}} \left| \frac{\partial f_i}{\partial v_j}(\mathbf{u}, \mathbf{v}) \right|$$

We call the matrix $A_{\mathbf{z}} = [a_{ij}]_{i,j \in \mathbf{Z}}$ *the Lipschitz matrix of* $F_{\mathbf{Z}}$.

When $\mathbf{Z} = \mathbf{V}$, *we simply call* $A = [a_{ij}]$ *the Lipschitz matrix of* F.

Lipschitz Matrix

$$M_2 \begin{cases} \mathbf{u}, \mathbf{v} \in \mathbb{R}^4 \\ U_i \sim \mathcal{N}(0,1) \quad iid. \\ f_1 : V_1 \leftarrow U_1 \\ f_2 : V_2 \leftarrow U_2 \\ f_3 : V_3 \leftarrow \frac{1}{2}V_1 + \frac{1}{2}V_4 + U_3 \\ f_4 : V_4 \leftarrow \frac{1}{2}V_2 + \frac{1}{2}V_3 + U_4 \end{cases}$$

$$\begin{tikzcd} V_1 \arrow[d] & V_2 \arrow[d] \\ V_3 \arrow[r, shift right] & V_4 \arrow[l, shift right] \end{tikzcd}$$

$$A_{2,3} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

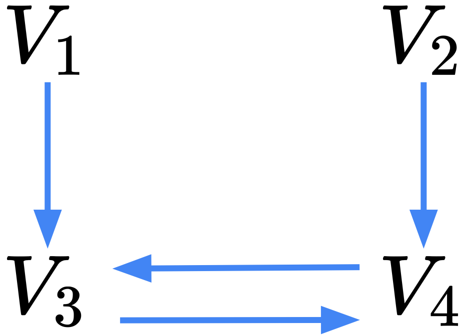
stable-Lipschitz SCM

Definition 5 (stable-Lipschitz SCM). *Let $M = \langle V, U, F, P(U) \rangle$ be an SCM with causal diagram G. We say M is stable-Lipschitz if, for every strongly connected component $\mathbf{Z}$ of G, the following conditions hold:*

- *$F_{\mathbf{Z}}$ is differentiable and Lipschitz.*
- *$\rho(A_{\mathbf{Z}}) < 1$.*

stable-Lipschitz SCM

$$M_2 \begin{cases} \mathbf{u}, \mathbf{v} \in \mathbb{R}^4 \\ U_i \sim \mathcal{N}(0,1) \quad iid. \\ f_1 : V_1 \leftarrow U_1 \\ f_2 : V_2 \leftarrow U_2 \\ f_3 : V_3 \leftarrow \frac{1}{2}V_1 + \frac{1}{2}V_4 + U_3 \\ f_4 : V_4 \leftarrow \frac{1}{2}V_2 + \frac{1}{2}V_3 + U_4 \end{cases}$$



$$A_{2,3} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

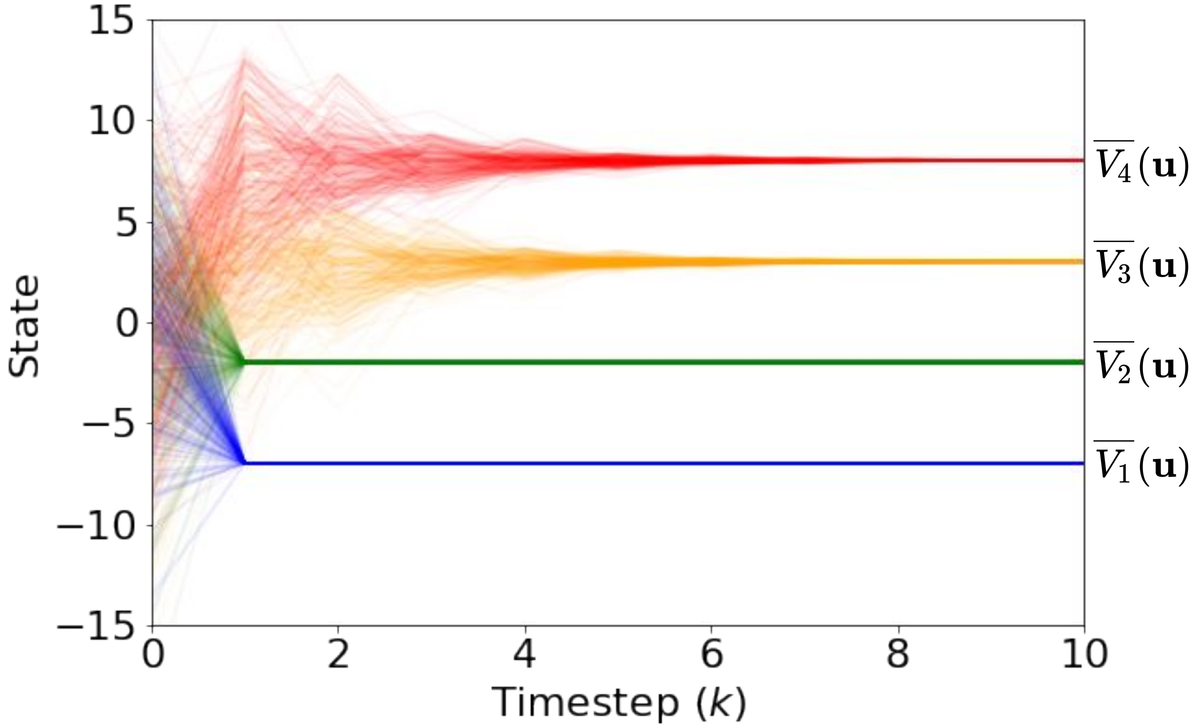
$$\rho(A_{2,3}) = 1$$

Not stable-Lipschitz

Limitations: stable-Lipschitz SCM

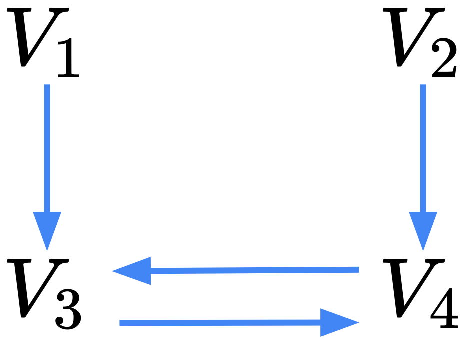
Fixed $\mathbf{u} = (-7, -2, 2.5, 7.5)$

This SCM still converges, even though it's **not** stable-Lipschitz. This is a limitation of the current stable-Lipschitz definition (but for now sticking with it because it makes it easier to prove results).



Cyclic + Nonlinear SCM (1 of 2)

$$M_3 \begin{cases} \mathbf{u}, \mathbf{v} \in \mathbb{R}^4 \\ U_i \sim \mathcal{N}(0,1) \quad iid. \\ f_1 : V_1 \leftarrow U_1 \\ f_2 : V_2 \leftarrow U_2 \\ f_3 : V_3 \leftarrow V_1 V_4 + U_3 \\ f_4 : V_4 \leftarrow V_2 V_3 + U_4 \end{cases}$$



No longer linear (or even Lipschitz-continuous), what does the potential response look like?

Cyclic + Nonlinear SCM (2 of 2)

M_3	$\xrightarrow{k \to \infty}$	Potential Response
$f_1 : V_1 \leftarrow U_1$		$\overline{V_1}(\mathbf{U}) = U_1$
$f_2 : V_2 \leftarrow U_2$		$\overline{V_2}(\mathbf{U}) = U_2$
$f_3 : V_3 \leftarrow V_1 V_4 + U_3$		$\overline{V_3}(\mathbf{U}) = \frac{U_1 U_4 + U_3}{1 - U_1 U_2}$
$f_4 : V_4 \leftarrow V_2 V_3 + U_4$		$\overline{V_4}(\mathbf{U}) = \frac{U_1 U_4 + U_3}{1 - U_1 U_2} U_2 + U_4$

When an SCM is nonlinear, its potential response may have singularities. Here, when $U_1 U_2 = 1$

If I see that nodes 3 and 4 are large I must be close to the singularity…In which case knowing something about node 1 (it's large) tells me something about node 2 (it's small)! So not conditionally independent

Example: d-separation

[Spirtes 2013
Bongers et al. 2021
Forr’e et al. 2017]

$$\boxed{\begin{aligned} V_1 &\leftarrow U_1 \\ V_2 &\leftarrow U_2 \\ V_3 &\leftarrow V_1 V_4 + U_3 \\ V_4 &\leftarrow V_2 V_3 + U_4 \end{aligned}}$$

$$\begin{tikzcd} V_1 \arrow[d] & V_2 \arrow[d] \\ V_3 \arrow[r, shift right] & V_4 \arrow[l, shift right] \end{tikzcd}$$

2 edges means 2 paths

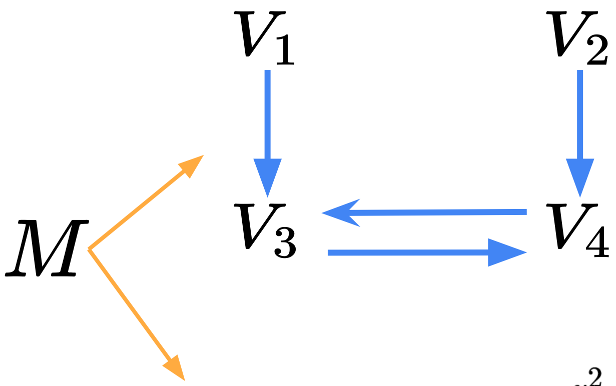
$$(V_1 \perp\!\!\!\perp V_2 | V_3, V_4)_G \checkmark$$

Path 1: $V_1 \rightarrow \boxed{V_2} \rightarrow \boxed{V_3} \leftarrow V_4$ *Blocked*

Path 2: $V_1 \rightarrow \boxed{V_2} \leftarrow \boxed{V_3} \leftarrow V_4$ *Blocked*

[Spirtes 2013]

Example: Directed Global Markov Property



$\xrightarrow{\text{d-separation}} (V_1 \perp\!\!\!\perp V_2 | V_3, V_4)_G$

$$P(\mathbf{V}) = \left(\frac{1}{4\pi^2}\right) e^{-\frac{v_1^2}{2}} e^{-\frac{v_2^2}{2}} \underline{e^{-\frac{(v_3 - v_1 v_4)^2}{2}} e^{-\frac{(v_4 - v_2 v_3)^2}{2}} |(1 - v_1 v_2)^{-1}|}$$

Not Possible to Factor!

So $(V_1 \not\!\perp\!\!\!\perp V_2 | V_3, V_4)_P$

Let's double check with numerics…

Numerics: Directed Global Markov Property

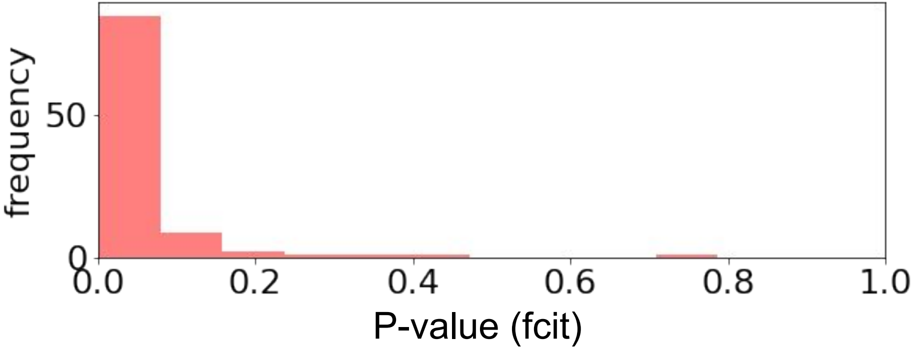
$$V_1(\mathbf{U}) = U_1$$
$$V_2(\mathbf{U}) = U_2$$
$$V_3(\mathbf{U}) = \frac{U_1U_4+U_3}{1-U_1U_2}$$
$$V_4(\mathbf{U}) = \frac{U_1U_4+U_3}{1-U_1U_2}U_2 + U_4$$

$$P(\mathbf{U}) \xrightarrow{V_i(\mathbf{U})} P(\mathbf{V})$$

- Monte Carlo: sample from $P(\mathbf{U})$ and apply $V_i(\mathbf{U})$

- Test $(V_1 \overset{?}{\perp\!\!\!\perp} V_2 | V_3, V_4)_P$ numerically using FCIT

$(V_1 \not\!\perp\!\!\!\perp V_2|V_3, V_4)$



Low p-values $\longrightarrow$ conditional dependence

$$(V_1 \not\!\perp\!\!\!\perp V_2 | V_3, V_4)_P$$

$$(V_1 \perp\!\!\!\perp V_2|V_3, V_4)_G \not\Rightarrow (V_1 \perp\!\!\!\perp V_2|V_3, V_4)_P$$

Recap: Asymptotics

$$V_1(\mathbf{U}) = U_1$$
$$V_2(\mathbf{U}) = U_2$$
$$V_3(\mathbf{U}) = \frac{U_1 U_4 + U_3}{1 - U_1 U_2}$$
$$V_4(\mathbf{U}) = \frac{U_1 U_4 + U_3}{1 - U_1 U_2} U_2 + U_4$$

$$P(\mathbf{U}) \xrightarrow{V_i(\mathbf{U})} P(\mathbf{V})$$

- 🤔 $V_i(\mathbf{U})$ is found by fixing $\mathbf{u}$ and solving for the equilibrium of the system
- 🤔 For linear SCMs*, these $V_i(\mathbf{U})$ at equilibrium preserve the conditional independencies of G (hence, d-separation is valid)

Perhaps nonlinear SCMs whose *asymptomatic dynamics* are bounded by a linear SCM will also inherit these nice properties of $V_i(\mathbf{U})$

*If $P(\mathbf{V}) > 0$ everywhere, and there's no trivial dependencies

Example: Lipschitz Matrix

Bound each f_i by $g_i = (\sum_j a_{ij} V_j) + U_i$ $a_{ij} = \max_{\mathbf{v}} |\frac{\partial f_i}{\partial V_j}(\mathbf{v}, \mathbf{u})|$

$$
\begin{aligned}
f_1 &: V_1 \leftarrow U_1 \\
f_2 &: V_2 \leftarrow U_2 \\
f_3 &: V_3 \leftarrow V_1 V_4 + U_3 \\
f_4 &: V_4 \leftarrow V_2 V_3 + U_4
\end{aligned}
$$

$$
\begin{aligned}
g_1 &: V_1 \leftarrow U_1 \\
g_2 &: V_2 \leftarrow U_2 \\
g_3 &: V_3 \leftarrow \boxed{a_{31}} V_1 + \boxed{a_{34}} V_4 + U_3 \\
g_4 &: V_4 \leftarrow \boxed{a_{42}} V_2 + \boxed{a_{43}} V_3 + U_4
\end{aligned}
$$

Example: Lipschitz Matrix

Bound each f_i by $\quad g_i = \left(\sum_j a_{ij} V_j\right) + U_i \qquad a_{ij} = \max_{\mathbf{v}} \left|\frac{\partial f_i}{\partial V_j}(\mathbf{v}, \mathbf{u})\right|$

$$f_1 : V_1 \leftarrow U_1 \longrightarrow a_{1j} = 0 \longrightarrow g_1 : V_1 \leftarrow U_1$$

$$f_2 : V_2 \leftarrow U_2 \longrightarrow a_{2j} = 0 \longrightarrow g_2 : V_2 \leftarrow U_2$$

$$f_3 : V_3 \leftarrow V_1 V_4 + U_3 \longrightarrow g_3 : V_3 \leftarrow \boxed{a_{31}} V_1 + \boxed{a_{34}} V_4 + U_3$$

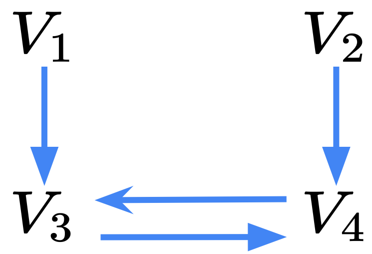
$$f_4 : V_4 \leftarrow V_2 V_3 + U_4 \longrightarrow g_4 : V_4 \leftarrow \boxed{a_{42}} V_2 + \boxed{a_{43}} V_3 + U_4$$

$$a_{31} = \max_{\mathbf{v}} |V_4| \quad a_{34} = \max_{\mathbf{v}} |V_1| \quad a_{42} = \max_{\mathbf{v}} |V_3| \quad a_{43} = \max_{\mathbf{v}} |V_2|$$

- Some terms go to infinity, so we can't bound this example!
  - (these correspond directly to unboundedness of the potential response function)
- But, there's a subset of U such that V is bounded…

Example: stable-Lipschitz SCM

$$M_4 \begin{cases} \mathbf{u}, \mathbf{v} \in [-0.5, 0.5] \\ U_i \sim \mathcal{N}(0,1) \cap [-0.5, 0.5] \quad iid. \\ f_1 : V_1 \leftarrow U_1 \\ f_2 : V_2 \leftarrow U_2 \\ f_3 : V_3 \leftarrow V_1 V_4 + U_3 \\ f_4 : V_4 \leftarrow V_2 V_3 + U_4 \end{cases}$$



$$(V_1 \overset{?}{\perp\!\!\!\perp} V_2 \mid V_3, V_4)_P$$?

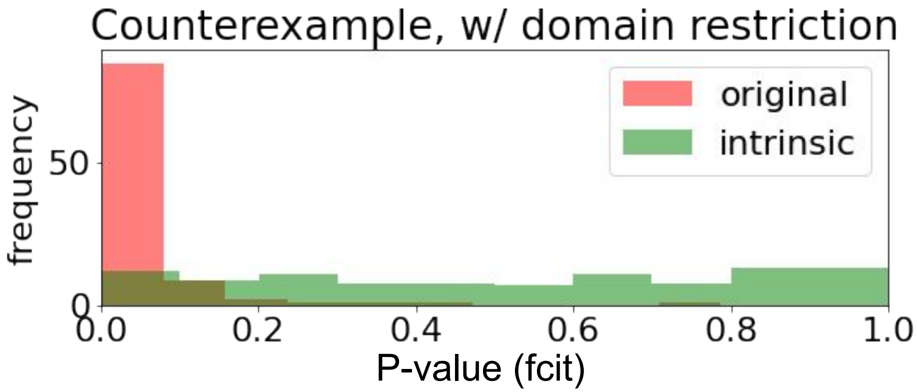
Numerics: Lipschitz SCM

$U_i \sim \mathcal{N}(0,1) \cap [-0.5, 0.5] \quad iid.$

$$M \quad \boxed{\begin{aligned} V_1(\mathbf{U}) &= U_1 \\ V_2(\mathbf{U}) &= U_2 \\ V_3(\mathbf{U}) &= \frac{U_1U_4+U_3}{1-U_1U_2} \\ V_4(\mathbf{U}) &= \frac{U_1U_4+U_3}{1-U_1U_2}U_2 + U_4 \end{aligned}}$$

Test $(V_1 \overset{?}{\perp\!\!\!\perp} V_2 | V_3, V_4)_P$ again, numerically

- Using $P(\mathbf{U}) \xrightarrow{V_i(\mathbf{U})} P(\mathbf{V})$ but with restricted domain



Uniform p-values $\longrightarrow$ conditional independence

$$(V_1 \perp\!\!\!\perp V_2 | V_3, V_4)_P$$

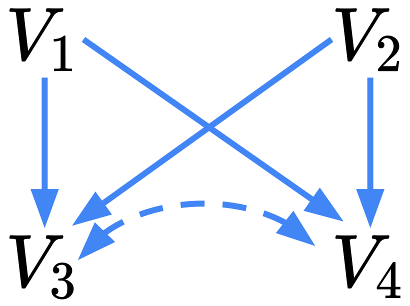
$$(V_1 \perp\!\!\!\perp V_2 | V_3, V_4)_G \implies (V_1 \perp\!\!\!\perp V_2 | V_3, V_4)_P$$

was inherited by the linear bound!

Main Results

Acyclic SCMs are stable-Lipschitz

Theorem 1 (acylic $\subset$ stable-Lipschitz). *Let M be an acyclic SCM. Then M is stable-Lipschitz.*



stable-Lipschitz SCMs are Simple

[Bongers et al. 2021]

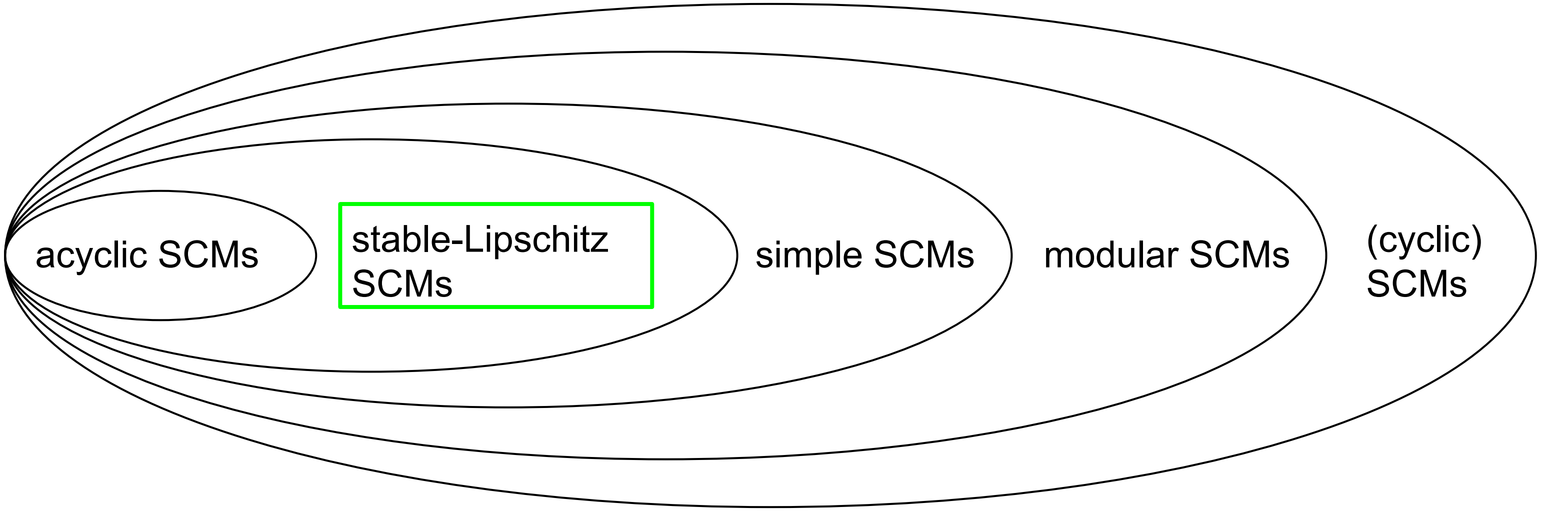
Definition 3 (Simple SCM). *Let $M = \langle V, U, F, P(U) \rangle$ be an SCM. We call M simple if M is uniquely solvable w.r.t. every subset $\mathbf{Z} \subseteq \mathbf{V}$.*

$$\mathbf{V} = F(\mathbf{V}, \mathbf{u}) \qquad \mathbf{V}_{\mathbf{Z}} = F_{\mathbf{Z}}(\mathbf{V}, \mathbf{u})$$

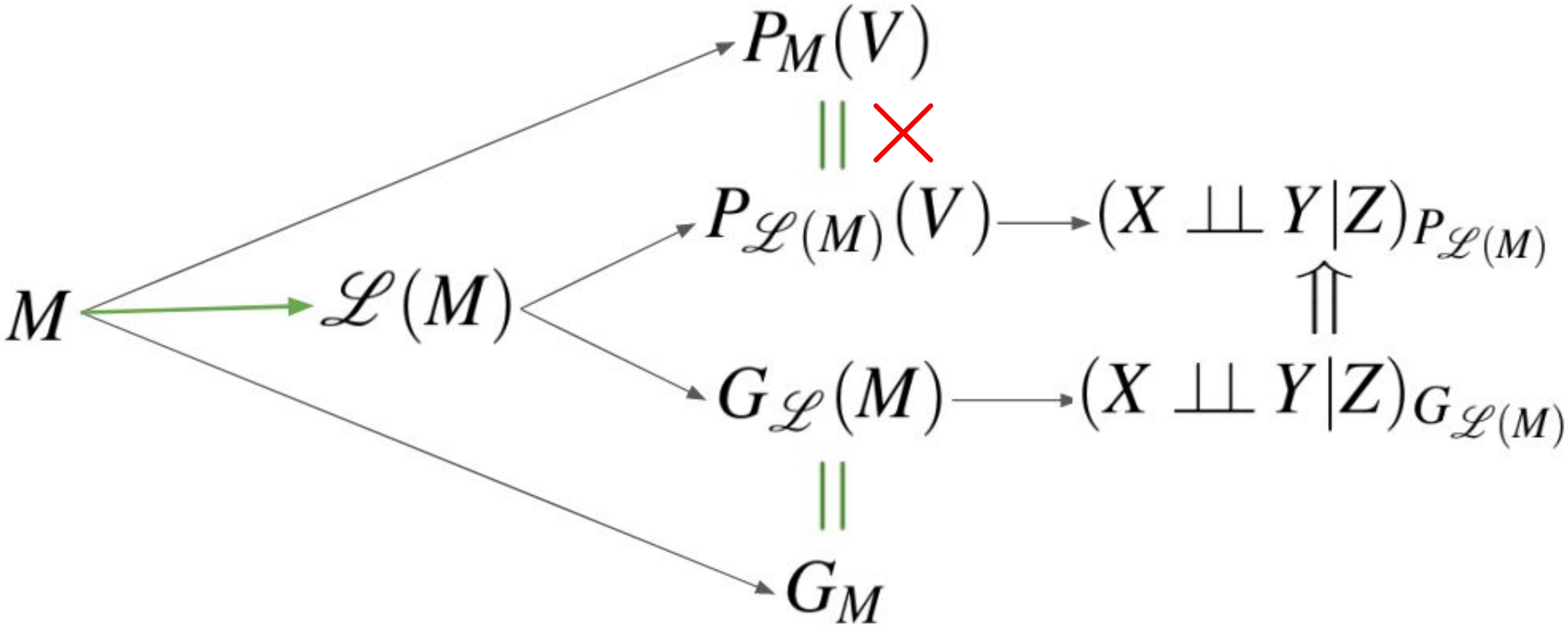
Theorem 4 (stable-Lipschitz $\subset$ simple). *Let M be a stable-Lipschitz SCM. Then M is simple.*

[Bongers et al. 2021]

Relation between spaces



Attempt 1: Observational Markov (failed)



Numerics: Observational d-separation

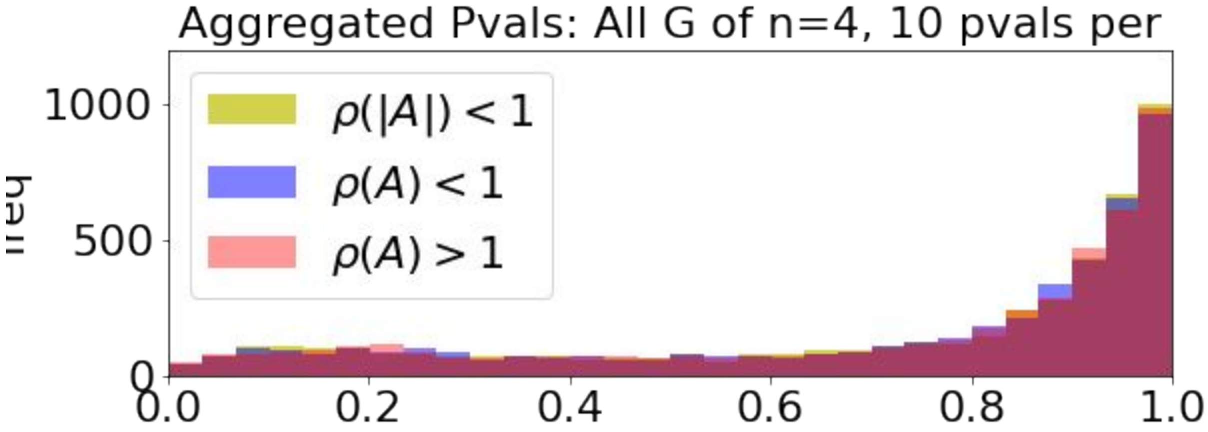
Conjectures:

1. Every stable-Lipschitz SCM satisfies the dGMP
   a. that is, the observational distribution respects every conditional independence in the causal graph.
2. Lipschitz SCMs won't generally satisfy the dGMP

Numerics: Observational d-separation

Numerics Results:

1. Every stable-Lipschitz SCM satisfies the dGMP ✓
   a. that is, the observational distribution respects every conditional independence in the causal graph.
2. Lipschitz SCMs won't generally satisfy the dGMP ✗



Method:

- All cyclic graphs with 4 nodes
  - 107 non-trivial graphs
- Given a graph, sample Relu neural network (with appropriate spectral radius)
- Enumerate all d-separations
- 10x:
  - Sample obs. dist.
    - Root-finding
  - Test independence (fcit)

Observational Markov

Theorem 7 (Observational dGMP). *Let M be stable Lipschitz with 1. structural equations of the form $F(\mathbf{V}, \mathbf{U}) = H(\mathbf{V}) + \mathbf{U}$ (additive noise), 2. each $U_i \cap U_j = \emptyset$ for $i \neq j$ (independent noise), and 3. $P_M(\mathbf{V})$ has density according the the Legesgue measure on $\mathbb{R}^{|\mathbf{V}|}$ (positivity).*

Then M satisfies the directed global Markov property.

I am confident that conditions 1 and 2 in the hypothesis can be substantially weakened with further research.

This is a very new result, so while I believe the proof to be accurate and comprehensive, I'm still vetting it for errors: I'd place 5:1 odds against finding an irrecoverable error in the proof.

stable-Lipschitz SCMs are Closed under Interventions

Theorem 2 (Closed under Interventions). *Let M be a stable-Lipshitz SCM, $\mathbf{X} \subseteq \mathbf{V}$, and $\mathbf{x} \in dom(\mathbf{X})$. Then $M_{do(\mathbf{X}=\mathbf{x})}$ is stable-Lipschitz.*

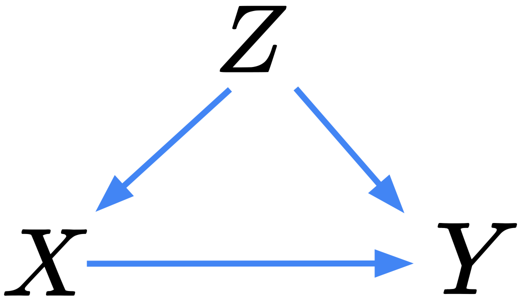
$$M_2 \left\{ \begin{array}{l} \mathbf{u}, \mathbf{v} \in \mathbb{R}^4 \\ U_i \sim \mathcal{N}(0,1) \quad iid. \\ f_1 : V_1 \leftarrow U_1 \\ f_2 : V_2 \leftarrow U_2 \\ f_3 : V_3 \leftarrow \frac{1}{2}V_1 + \frac{1}{2}V_4 + U_3 \\ f_4 : V_4 \leftarrow \frac{1}{2}V_2 + \frac{1}{2}V_3 + U_4 \end{array} \right. \longrightarrow \begin{array}{l} f_1 : V_1 \leftarrow U_1 \\ f_2 : V_2 \leftarrow U_2 \\ f_3 : V_3 \leftarrow x \\ f_4 : V_4 \leftarrow \frac{1}{2}V_2 + \frac{1}{2}V_3 + U_4 \end{array}$$

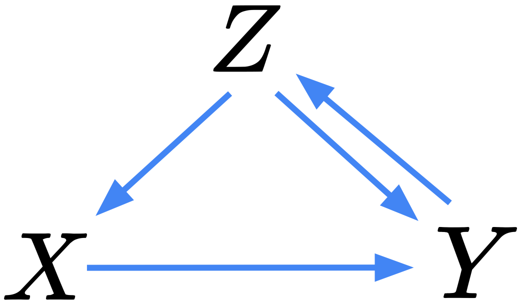
$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 0 & & 0 \\ 0 & 0 & & 0 \\ & & & \\ 0 & \frac{1}{2} & & 0 \end{bmatrix}$$

Nonnegative matrix theory

Backdoor for stable-Lipschitz SCMs

Corollary 1 (Adjustment Formula). *Let M be as (before) and $Q = P(y|do(x))$ a causal query. If the BDC is satisfied, then Q can be found via backdoor adjustment.*





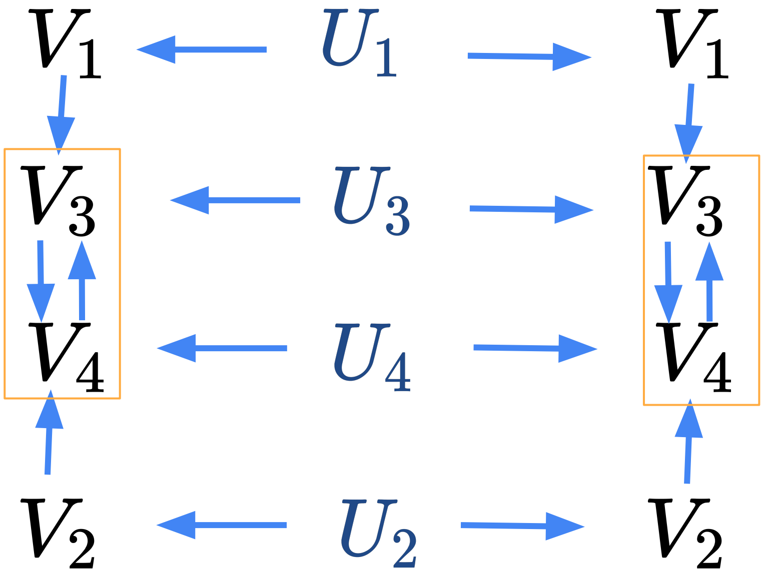
One of the motivations for weakening the condition of independent noise is to be able to similarly prove that the front-door criteria is also valid.

Interventional Markov

Theorem 8 (Interventional dGMP). *Let M be as in Theorem 7. For any $\mathbf{X} \subseteq \mathbf{Z}$, $M_{do(\mathbf{X}=\mathbf{x})}$ satisfies the directed global Markov property.*

stable-Lipschitz SCMs are Closed under Counterfactuals

Theorem 9 (Closure under Twin Operation). *Let M be stable-Lipschitz. Then M^{twin} is also stable-Lipschitz.*



$$A_{2,3} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Counterfactual Markov (Conjecture)

Conjecture 1 (Counterfactual dGMP). *Let M be stable-Lipschtiz. Then the counterfactual distributions of M satisfy the directed global Markov property relative to the corresponding twin network.*

I believe the counterfactual dGMP holds if the condition of independent noise can weakened, as an immediate consequence of the other results so far. However, I would place 2:1 odds that I'm missing some additional aspect of the proof.

Future Work

- Pearl Causal Hierarchy
  - [Bongers. et. al] sets it up, but fails to make any statement about “collapse on a set of measure 0”
  - **Belief that the PCH holds for cyclic SCMs is the primary motivation for all my results**
- Weakening hypothesis of Observational dGMP
  - Allow for latent confounding (Front-door adjustment)
  - general Lipschitz SCMs?
- Counterfactual dGMP
- Do-calculus
  - To extend the Backdoor Adjustment result
- **Multiple Equilibria**
  - Weakening “uniquely solvable” to “solvable”, by showing that d-separation holds in the neighborhood of equilibria of “locally stable-Lipschitz” SCMs
    - **A completely new alternative to settable systems!**
      - In some sense, knowing which equilibria we’re at is SCM-level knowledge – more than we should need for identification from the causal graph G and P(V)

References

- Antti Hyttinen, Frederick Eberhardt, and Patrik O. Hoyer. Learning linear cyclic causal models with latent variables. J. Mach. Learn. Res., 13:3387–3439, 2012.
- David Reber and Benjamin Z. Webb. Intrinsic stability: Global stability of dynamical networks and switched systems resilient to any type of time-delays. ArXiv, abs/1906.04000, 2019
- Halbert White and Karim Chalak. Settable systems: An extension of pearl's causal model with optimization, equilibrium, and learning. Journal of Machine Learning Research, 10(61):1759–1799, 2009.
- Halbert White, Haiqing Xu, and Karim Chalak. Causal discourse in a game of incomplete information. Journal of Econometrics, 182(1):45–58, 2014. Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions.
- Judea Pearl. Causality. Cambridge University Press, 2 edition, 2009.
- Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. Fast conditional independence test for vector variables with large sample sizes, 2018.
- Patrick Forr'e and Joris M. Mooij. Markov properties for graphical models with cycles and latent variables. arXiv:Statistics Theory, 2017.
- Peter Spirtes. Directed cyclic graphical representations of feedback models. CoRR, abs/1302.4982, 2013.
- Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M. Mooij. Foundations of structural causal models with cycles and latent variables. The Annals of Statistics, 49(5):2885 – 2915, 2021.