

Figure 1: The green arrow is the directed global Markov property (validity of d -separation) when M has a unique solution (TODO: check this). The converse of the green arrow is d -faithfulness.

Dressing Room: Cyclic Causality: D-separation and Nonlinearity

David Reber

Columbia University

david.reber@columbia.edu

Abstract: Overall Outline © 2022 The Author(s)

1. Motivation and Problem Formulation

To the extent that feedback is present in a system (i.e. game theory, economics, physical sciences) a natural modeling choice would be to represent this feedback as a cycle in the structural causal model. However, the introduction of cycles breaks many of the convenient guarantees of acyclic models: the lack of a unique equilibrium means the potential response function is not well-defined; the observational, interventional, and counterfactual distributions may not exist, or if they do, they may not be unique; marginalizing over variables may not be possible, and even if it is, the causal semantics may not be preserved; d -separation may not hold (aka. the “directed global Markov property”); even the weaker variant of σ -separation may not hold (the “general directed global Markov property”).

In this report, we outline a promising direction for ensuring that a general class of nonlinear SCMs (intrinsically-stable SCMs, or ‘intrinsic SCMs’ for short) avoids the loss of the directed global Markov property. The directed global Markov property ensures the validity of d -separation, and can be visualized as in Figure 1 (for a precise formulation of d -separation in the cyclic setting, see Section 5 in the Appendix). d -separation is a vital component of the do-calculus, and consequently for inferring interventional effects and counterfactual queries, ensuring transportability, and adjusting for sampling bias, to name a few.

1.1. counterexample

For an illustration of how d -separation may not hold for cyclic SCMs, consider the following example from [4] (and further discussed in [1]).

Example 1 (Counterexample: Loss of d -separation validity) Consider the SCM $M = (\mathbf{U}, \mathbf{V}, F, P(\mathbf{U}))$ with $\mathbf{u} = [u_1, u_2, u_3, u_4] \in \mathbb{R}^4$, $\mathbf{v} = [x_1, x_2, x_3, x_4] \in \mathbb{R}^4$, where F is defined component-wise by

$$f_1(\mathbf{v}, \mathbf{u}) = u_1, \quad f_2(\mathbf{v}, \mathbf{u}) = u_2, \quad f_3(\mathbf{v}, \mathbf{u}) = x_1 x_4 + u_3, \quad f_4(\mathbf{v}, \mathbf{u}) = x_2 x_3 + u_4$$

and $P(\mathbf{U})$ is the standard-normal distribution on \mathbb{R}^4 . The causal graph G_M is shown in Figure 2 (left). For every equilibrium of M , the induced observational distribution $P_M(\mathbf{V})$ satisfies $(X_1 \not\perp\!\!\!\perp X_2 \mid X_3, X_4)_{P_M}$. However, X_1 and X_2 are d -separated given $\{X_3, X_4\}$ in the causal graph; that is, $(X_1 \perp\!\!\!\perp X_2 \mid X_3, X_4)_{G_M}$. Hence, the global directed Markov property does not hold for M .

When d -separation is not valid (the directed global Markov property does not hold), it may be that a weaker condition is satisfied. σ -separation is an extension of d -separation, which works by applying d -separation to the acyclicification of the original graph (see Figure 2 (right) for the acyclicification of Example 1). σ -separation implies d -separation; in other words, the general directed global Markov property is weaker than the directed global Markov property. The relationship between these two Markov properties can be seen in Figure 3.

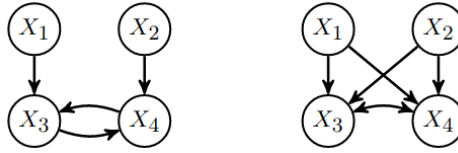


Figure 2: The causal graph of Example 1 (left) and its acyclification (right). Source: [1]

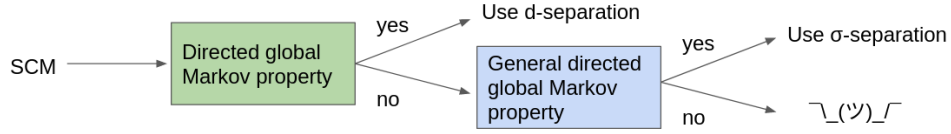


Figure 3: Relationship between d-separation and σ -separation.

Acyclic SCMs, discrete SCMs (with ancestrally unique solvability), and linear SCMs (with nontrivial dependencies and positive measure) are known to satisfy the directed global Markov property [3]. On the other hand, simple SCMs are known to satisfy the general directed global Markov property [1].

It seems reasonable that nonlinear SCMs which are asymptotically equivalent to linear SCMs should also satisfy the directed global Markov property (subject to the same restrictions of non-trivial dependencies and positive measure), and thus be able to read off conditional independencies from the graph via d-separation. I conjecture that intrinsically-stable SCMs (a class of cyclic, nonlinear, continuous-domained SCMs) are precisely this class of SCMs.

Problem 1 (Intrinsic SCMs and d-separation: Observational) *Prove whether the observational distribution of intrinsically-stable SCMs satisfy the directed global Markov property: that is, whether every conditional independence read-off by d-separation in the causal graph G_M holds in the observational distribution $P_M(V)$.*

Inputs/Outputs: Expressed by the green implication in Figure 1.

(Admittedly it seems awfully convenient that a topic I studied for over 4 years (undergraduate, Master's, and a bit thereafter) happens to be the right tool for the job. However, I think intrinsic SCMs can stand on their own merits):

Considered from a dynamical-systems perspective, intrinsic-SCMs have a number of promising properties relevant to causality. They have a unique equilibrium, so the potential response function of the SCM will be well-defined. They 'behave' like linear systems: they are asymptotically bounded by the dynamics of a linear system, and share the same equilibria if subject to the same forcing factor (exogenous distribution). Intuitively speaking, this means that we can go to linear-world, prove things about the system there, and the results will still hold in nonlinear-world. Furthermore, intrinsically-stable systems are quite general: they only require Lipschitz-continuity, and the domain to be a product of metric spaces (which could be something nice like \mathbb{R}^4 , or something abstract like language and shapes). If a general domain like metric spaces is used, the linearization uses the metrics to map to the real numbers: in this way the domain is simplified as well.

Of particular interest for causality, intrinsically-stable systems derive their name because they are closed under a surprising number of structural transformations; that is, the resulting system will still be intrinsically-stable, and often (if applicable) the equilibria will be preserved. Some of the transformations that come to mind are: 1. lengthening of paths (i.e. through time-delays), 2. collapsing portions of the graph, 3. duplicating portions of the graph (called specialization), 4. time-varying structural switching, and 5. any and all isospectral transformations (transformations which preserve the eigenvalues of the system).

Given these properties of intrinsic-SCMs, a negative result for Problem 1 becomes even more interesting. Since intrinsically-stable SCMs are asymptotically equivalent to linear SCMs, if d-separation holds for the linear but not the nonlinear SCM, this would indicate that something more than the equilibria of a cyclic SCM is needed to determine the validity of d-separation.

the previous two paragraphs are a great place to cite intrinsic papers

2. Literature Review

intrinsic stability!

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

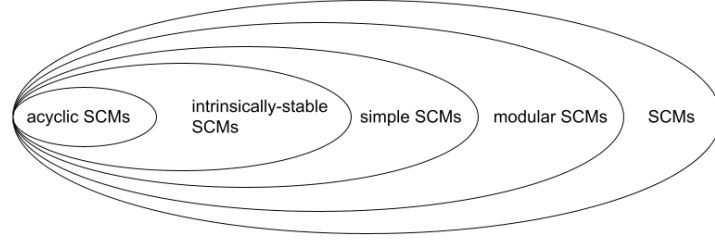


Figure 4: Conjectured set-inclusion relation of the class of intrinsically-stable SCMs relative to other classes of cyclic SCMs. **Move to be first figure?**

2.1. Frameworks for cyclic causality

- ‘Foundations’ (1 slide)
 - space inclusion diagram (int. stable between acyclic and simple)
- Settable Systems (1 slide)
 -

2.2. This problem specifically

- Foundations
 -
- Markov Paper?
 -
- Linear paper?
 -

3. Preliminary Results

3.1. Numerics

Consider again Example 1, the counterexample popular in the literature. If we restrict the domain of the exogenous variables to $\mathbf{u} \in [-0.5, 0.5]^4$ (which we may do by discarding any samples outside of this range), the resulting SCM is intrinsically-stable. Hence, if Problem 1 resolves in the affirmative, we should observe that the observational distribution of the restricted-domain SCM should have the conditional independence of $(X_1 \perp\!\!\!\perp X_2 \mid X_3, X_4)_{P_M}$, matching the independence $(X_1 \perp\!\!\!\perp X_2 \mid X_3, X_4)_{G_M}$ computed by d-separation from the graph.

The results of this test are shown in Figure 5. In order to test the conditional independence numerically over a continuous domain, the python package `fcit` was used [2]. The resulting p-values are distributed roughly uniformly on $[0, 1]$, consistent with an `fcit` result of conditional independence.

3.2. Proof sketch for Problem 1

My research strategy for proving the affirmative of Problem 1, comprises of two parts (see Figure 6): First, relate M to a linear SCM $\mathcal{L}(M)$ which satisfies the directed global Markov property. Secondly, show that M and $\mathcal{L}(M)$ respect the same graph and are observationally equivalent.

This can be broken down into greater detail via Figure 7. Each of these parts (with a rough proof sketch) are described below:

1. $M \rightarrow \mathcal{L}(M)$: Show that for every intrinsically-stable M , there exists a linear SCM $\mathcal{L}(M)$ which satisfies the directed global Markov property (the blue box).
2. $M \rightarrow G_M$: Always possible by definition, I believe.
3. $G_M = G_{\mathcal{L}(M)}$: Show that for every node, the parents are preserved (according to [1]’s definition of parents: see Appendix 5).

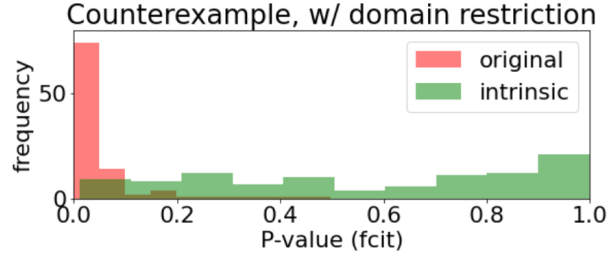


Figure 5: The counterexample discussed in [1, 4]. Here, conditional independence is tested while restricting the exogenous variables to be drawn from a truncated multivariate normal distribution (such that the SCM as a whole is intrinsically stable). The conditional independence identified by d-separation is now present in the observational distribution.

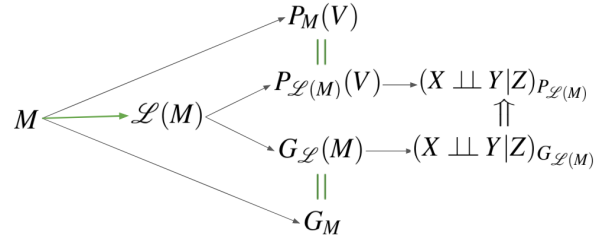


Figure 6: Conjectured proof sketch, for positive resolution of Problem 1.

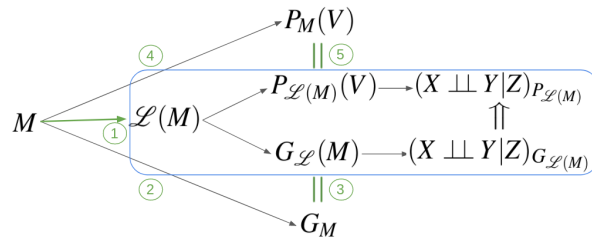


Figure 7: Identical as Figure 6, but with helpful labels of the conjectured relations. Note that the portion contained within the blue box simply expresses that the associated linear SCM $\mathcal{L}(M)$ satisfies the directed global Markov property (Figure 1).

4. $M \rightarrow P_M(V)$: Show that if M is intrinsically stable, it is uniquely solvable.
5. $P_M(V) = P_{\mathcal{L}(M)}(V)$: Translate intrinsic-stability property of M and $\mathcal{L}(M)$ have the same equilibrium solutions, to SCMs. Show that this means they induce the same obs. distributions.

3.3. Anticipated Implications

I consider it is worth checking early in one’s research whether hypothetical success would itself be ludicrous, even if this means initially high exposure to unverified speculation. To keep this report appropriately rigorous, however, my speculation about these potential implications is located in Appendix 4.1. In a sentence: While I do expect positive resolutions to Problems 1 and 2 to mean that nonlinear, intrinsic SCMs inherit many of the nice properties of linear cyclic SCMs, I also expect this will not violate the ‘No Free Lunch’ Principle: the resulting linear SCMs will be no easier to use for modeling purposes.

3.4. Potential (future) Collaborators

I recently met with both Lewis Hammond (Oxford) and Christian Kroer (Columbia, IOER department). Both are very interested in ‘game theory + causality’, which I find the most compelling motivator for cyclic causality research. To the extent that nonlinear, cyclic SCMs are ‘locally intrinsic’ (as Example 1 was shown to be here) there may be some valuable collaborations to be had. For now, I’ll just keep occasional dialogue with them to see what arises.

References

1. 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.
2. Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. Fast conditional independence test for vector variables with large sample sizes, 2018.
3. Patrick Forré and Joris M. Mooij. Markov properties for graphical models with cycles and latent variables. *arXiv: Statistics Theory*, 2017.
4. Peter Spirtes. Directed cyclic graphical representations of feedback models. *CoRR*, abs/1302.4982, 2013.

4. Appendix: Musings Beyond Observational d-seperation

4.1. d-seperation and Interventional Distributions

I’m very 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 the operations of atomic intervention $M \rightarrow M_x$ and linearization $M \rightarrow \mathcal{L}(M)$ commute for intrinsic SCMs. Thus, if Problem 1 resolves in the affirmative, I conjecture it should be straightforward to demonstrate (atomic) interventional equivalence between a nonlinear, intrinsic SCM and its associated linear SCM $\mathcal{L}(M)$.

Problem 2 (Intrinsic SCMs and d-seperation: Interventional) *Prove whether the interventional distribution of intrinsically-stable SCMs satisfy the directed global Markov property: that is, whether every conditional independence read-off by d-seperation in the causal graph $(G_M)_{\bar{X}}$ holds in the interventional distribution $P_{M_x}(V)$.*

Note the order of operations for $(G_M)_{\bar{X}}$: first we generate the causal graph of M , then intervene with an atomic intervention on that graph. This is in contrast to $P_{M_x}(V)$, where we first perform an atomic structural intervention on M , then generate the distribution of this new SCM.

If Problem 2 answers in the affirmative, I believe it follows that the do-calculus is valid for intrinsic SCMs.

4.2. d-seperation and Counterfactual distributions

I also confidently conjecture that (if Problem 1 resolves positively) intrinsic SCMs are also closed under the “mirror” transformation. (The mirror operation is useful for defining cyclic counterfactuals: you construct a ‘twin SCM’ using the twin network method). This operation, if I understand correctly, should merely reduce to the duplication of a strongly connected component, but otherwise leave strongly connected components unchanged. Since intrinsic-stability is based on a spectral criteria, one can without loss of generality simply consider each strongly connected component of the system independently. (Indeed, almost all my remaining uncertainty derives from wondering whether I’ve missed some subtle-yet-crucial detail in the definitions of cyclic SCMs).

Problem 3 (Intrinsic SCMs and d-separation: Counterfactual) *Prove whether the counterfactual distribution of intrinsically-stable SCMs satisfy the directed global Markov property: that is, whether every conditional independence read-off by d-separation in the twin network holds in the corresponding counterfactual distribution of the twin SCM.*

If Problem 3 resolves positively, then this would mean that the M and $\mathcal{L}(M)$ are not merely observationally or interventionally equivalent, but *counterfactually equivalent*. (However, they are precluded from being *structurally equivalent* according to the definition in [1], since this requires not only almost-everywhere equivalence of equilibrium solutions, but of the transition functions themselves.)

I do not expect this counterfactual equivalence (should it actually hold) to violate the ‘No Free Lunch’ principle. The counterfactual equivalence of an intrinsic SCM M with a linear SCM $\mathcal{L}(M)$ may do little more than imply that nice properties (like d-separation) are valid for M , as far as modeling is concerned, because I expect that the complexity of F_M ’s nonlinearity simply gets pushed into the linear SCM’s exogenous distribution $P_{\mathcal{L}(M)}(U)$.

5. Appendix: Relevant Definitions for Cyclic SCMs

The following definitions are copied verbatim from [1], and provided here only for ease of reference. Consequently they differ somewhat from the notation of this report.

DEFINITION 2.6 (Parent). Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ be an SCM. We call $k \in \mathcal{I} \cup \mathcal{J}$ a *parent* of $i \in \mathcal{I}$ if and only if there does not exist a measurable function⁷ $\tilde{f}_i : \mathcal{X}_{\setminus k} \times \mathcal{E}_{\setminus k} \rightarrow \mathcal{X}_i$ such that for $\mathbb{P}_{\mathcal{E}}$ -almost every $e \in \mathcal{E}$ and for all $x \in \mathcal{X}$,

$$x_i = f_i(x, e) \iff x_i = \tilde{f}_i(x_{\setminus k}, e_{\setminus k}).$$

Exogenous variables have no parents by definition. These parental relations are preserved under the equivalence relation \equiv on SCMs. They can be represented by a directed graph or a directed mixed graph.⁸

DEFINITION A.3 (Collider). Let $\pi = (i_0, \epsilon_1, i_1, \epsilon_2, i_2, \dots, \epsilon_n, i_n)$ 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 $i_{k-1} \rightarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \leftrightarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \rightarrow i_k \leftrightarrow i_{k+1}$ or $i_{k-1} \leftarrow i_k \leftrightarrow i_{k+1}$). 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 $i_{k-1} \rightarrow i_k \rightarrow i_{k+1}$, $i_{k-1} \leftarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \leftarrow i_k \rightarrow i_{k+1}$, $i_{k-1} \rightarrow i_k \rightarrow i_{k+1}$ or $i_{k-1} \leftarrow i_k \leftrightarrow i_{k+1}$.

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) $\pi = (i_0, \epsilon_1, i_1, \dots, \epsilon_n, i_n)$ in \mathcal{G} is said to be *C-d-blocked* or *d-blocked* by C if

1. it contains a collider $i_k \notin \text{an}_{\mathcal{G}}(C)$, 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

$$A \overset{d}{\perp\!\!\!\perp}_{\mathcal{G}} B \mid C.$$

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 $\mathcal{X}_{\mathcal{V}} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$, 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

$$A \overset{d}{\perp\!\!\!\perp}_{\mathcal{G}} B \mid C \implies \mathbf{X}_A \overset{\mathbb{P}_{\mathcal{V}}}{\perp\!\!\!\perp} \mathbf{X}_B \mid \mathbf{X}_C,$$

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 $X_i : \mathcal{X}_{\mathcal{V}} \rightarrow \mathcal{X}_i$ as random variables.