Title: Bipartite graphical causal models: beyond causal Bayesian networks and structural causal models

Speakers: Joris M. Mooij

Series: Quantum Foundations, Quantum Information

Date: September 16, 2024 - 11:30 AM

URL: https://pirsa.org/24090085

Abstract: Based on the immense popularity of causal Bayesian networks and structural causal models, one might expect that these representations are appropriate to describe the causal semantics of any real-world system, at least in principle. In this talk, I will argue that this is not the case, and motivate the study of more general causal modeling frameworks. In particular, I will discuss bipartite graphical causal models.

Real-world complex systems are often modelled by systems of equations with endogenous and independent exogenous random variables. Such models have a long tradition in physics and engineering. The structure of such systems of equations can be encoded by a bipartite graph, with variable and equation nodes that are adjacent if a variable appears in an equation. I will show how one can use Simonâ€™s causal ordering algorithm and the Dulmage-Mendelsohn decomposition to derive a Markov property that states the conditional independence for (distributions of) solutions of the equations in terms of the bipartite graph. I will then show how this Markov property gives rise to a do-calculus for bipartite graphical causal models, providing these with a refined causal interpretation.

Bipartite Graphical Causal Models

Joris Mooij
(collaborators: Tineke Blom, Mirthe van Diepen, Patrick Forré)

j.m.mooij@uva.nl

University of Amsterdam

September 16th, 2024

Part I

Introduction

Motivation

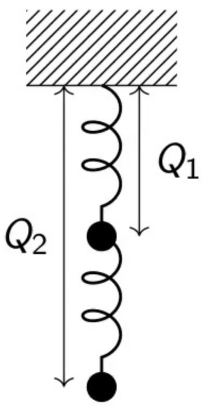
- Causal Bayesian Networks (CBNs) and Structural Causal Models (SCMs) are very popular.
- But these are not always appropriate.
- Example: bathtub or sink at equilibrium [Iwasaki and Simon, 1994].

- A more general causal modeling framework is needed.
- Here, we propose **bipartite** causal graphs that include both variable vertices **and equation vertices**.
- These reduce ambiguity of the notion of perfect intervention.
- We provide a Markov property and sketch a do-calculus.

Let us not ignore cycles!

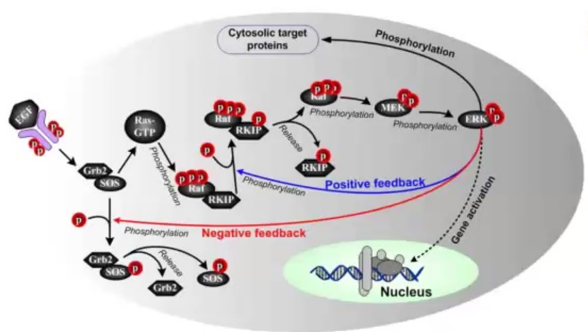
- Feedback in dynamical systems may induce cyclic causality at equilibrium.
- Fast dynamical interactions can lead to "instantaneous" causal cycles in time-series modeling.

physics:

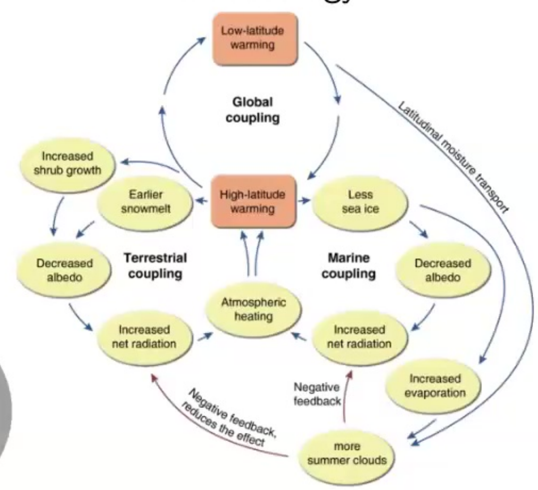


econometry:

biology, chemistry:

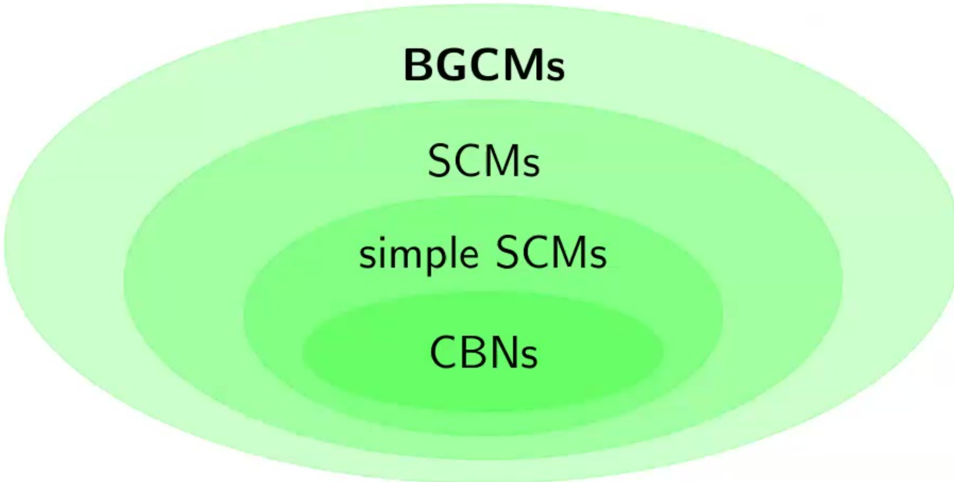


climatology:



In many applications, modeling causal cycles is essential.

Relations between causal models



Acronym	Model class	Cycles?	Reference
CBN	causal Bayesian network	−	[Pearl, 2009]
SCM	structural causal model	+	[Bongers et al., 2021]
simple SCM	simple structural causal model	+	[Bongers et al., 2021]
BGCM	**bipartite graphical causal model**	+	[Blom et al., 2021]

Bipartite graphical causal models are the most expressive models for cyclic causal systems.

Context

- Causal Bayesian networks and structural causal models have limitations when modeling cyclic causality.
- Simon's **causal ordering** approach to causality [Simon, 1953] provides a fundamentally different perspective.
- Given a **system of equations**, it provides possible **causal interpretations** of the equations.
- Each causal interpretation corresponds with a possible partitioning of the variables into **exogenous** and **endogenous** variables.
- This matches with how engineers and applied scientists often deal with causality.
- Combining causal ordering with the σ-separation criterion [Forré and Mooij, 2017] provides a general Markov property for causal systems represented as systems of equations [Blom et al., 2021].

This talk

Formulate Markov property and do-calculus in terms of the bipartite graph *only*.

Part II

Causal Ordering Algorithm

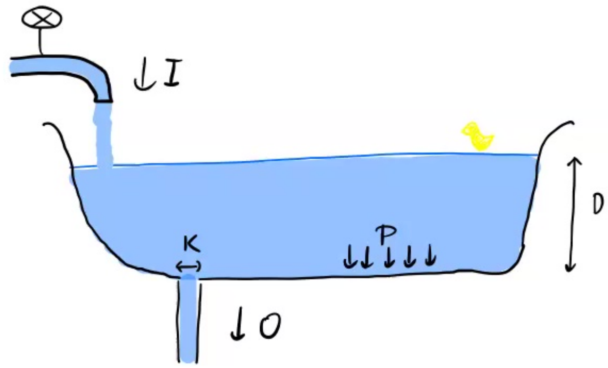
Example: Bathtub [Iwasaki and Simon, 1994]

Endogenous variables:

- X_O water outflow through drain
- X_D water depth
- X_P pressure at drain

Exogenous variables:

- X_I water inflow from faucet
- X_K drain size
- X_g gravitational acceleration



Independent/modular/autonomous **mechanisms**:

$f_1: \quad 0 = X_I - X_O$ at equilibrium, outflow equals inflow

$f_2: \quad 0 = X_K X_P - X_O$ outflow is proportional to pressure and drain diameter

$f_3: \quad 0 = X_g X_D - X_P$ pressure at drain proportional to depth and gravitational acceleration

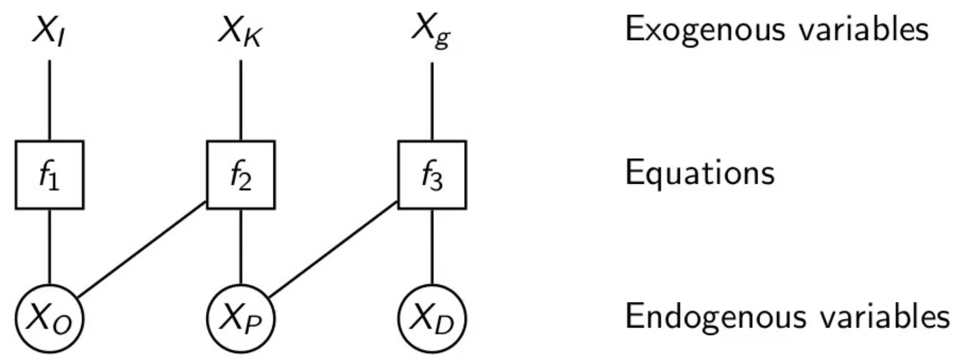
Assumption: endogenous variables do not cause exogenous variables.

Bipartite Graphical Representation

The structure of the equations:

$$f_1: \quad 0 = X_I - X_O$$
$$f_2: \quad 0 = X_K X_P - X_O$$
$$f_3: \quad 0 = X_g X_D - X_P$$

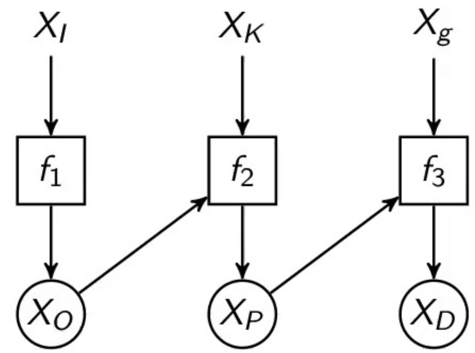
can be represented with a bipartite graph:



Solving systems of equations

The bipartite graph is helpful when solving a system of equations!

$$
\begin{aligned}
f_1 &: \quad 0 = X_I - X_O \\
f_2 &: \quad 0 = X_K X_P - X_O \\
f_3 &: \quad 0 = X_g X_D - X_P
\end{aligned}
$$



Solve in the following **ordering**:

1. Solve f_1 for X_O in terms of X_I: $\quad X_O = X_I$
2. Solve f_2 for X_P in terms of X_O and X_K: $X_P = \dfrac{X_O}{X_K}$
3. Solve f_3 for X_D in terms of X_P and X_g: $X_D = \dfrac{X_P}{X_g}$

This establishes **existence and uniqueness** of the solution $(\forall_{X_I, X_K, X_g > 0})$.

Solutions, distributions, Markov kernels

- By solving the equations we obtain **solution functions** that express all variables in terms of the exogenous variables:

$$F : (x_I, x_K, x_g) \mapsto (x_I, x_K, x_g, x_O, x_P, x_D) = \left(x_I, x_K, x_g, x_I, \frac{x_I}{x_K}, \frac{x_I}{x_K x_g} \right)$$

- We can assume all exogenous random variables to be independently distributed:

$$X_I \sim \mathbb{P}(X_I) \qquad X_K \sim \mathbb{P}(X_K) \qquad X_g \sim \mathbb{P}(X_g);$$

the **joint distribution** $\mathbb{P}(X_I, X_K, X_g, X_O, X_P, X_D)$ of all variables is obtained as the **push-forward** through the solution function F of $\mathbb{P}(X_I, X_K, X_g) = \mathbb{P}(X_I) \otimes \mathbb{P}(X_K) \otimes \mathbb{P}(X_g)$.

- We can also treat some exogenous variables as random, and others as non-random. This yields a **Markov kernel**, e.g., $\mathbb{P}(X_K, X_g, X_O, X_P, X_D \,\|\, X_I)$ if only X_I is treated as non-random.

Markov property for recursive equations

For a system of equations of the form

$$
\begin{aligned}
X_1 &= f_1(E_1) & & \\
X_2 &= f_2(X_{\mathrm{pa}(2)}, E_2) & \mathrm{pa}(2) &\subseteq \{1\} \\
X_3 &= f_3(X_{\mathrm{pa}(3)}, E_3) & \mathrm{pa}(3) &\subseteq \{1,2\} \\
X_4 &= f_4(X_{\mathrm{pa}(4)}, E_4) & \mathrm{pa}(4) &\subseteq \{1,2,3\} \\
\dots & & & \\
X_p &= f_p(X_{\mathrm{pa}(p)}, E_p) & \mathrm{pa}(p) &\subseteq \{1,2,3,\dots,p-1\}
\end{aligned}
$$

with $E_1, \dots, E_p$ independent, the *d*-separation criterion (global directed Markov property) holds for the corresponding DAG.

From causal ordering to Markov property

For any system of equations that **can be rewritten** in this canonical form, we obtain a Markov property.

Example: Markov property from causal ordering

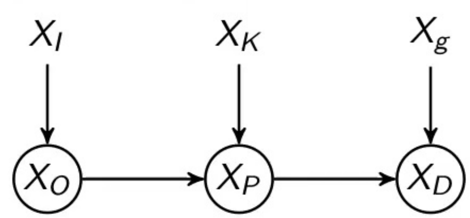
The bathtub equations

$$\begin{aligned} f_1 : \quad & 0 = X_I - X_O \\ f_2 : \quad & 0 = X_K X_P - X_O \\ f_3 : \quad & 0 = X_g X_D - X_P \end{aligned}$$

end up in canonical form by ordering and solving:

$$\begin{aligned} X_O &= X_I \\ X_P &= X_O / X_K \\ X_D &= X_P / X_g . \end{aligned}$$

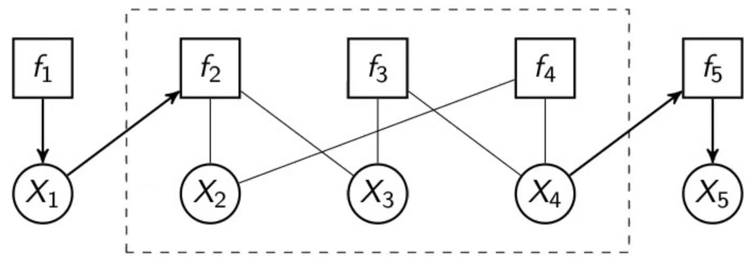
Assuming that exogenous variables (X_I, X_K, X_g) are independent, we can therefore apply the d-separation criterion to the DAG:



to read off (for example) that $X_D \perp\!\!\!\perp X_O \mid X_P$.

Loops in the bipartite graph

- Often we can only find an acyclic causal ordering after **clustering** some variables and equations.
- We then end up with subsets of equations that have to be solved simultaneously for subsets of variables.



We can solve as follows:

- Solve f_1 for X_1;
- Solve $\{f_2, f_3, f_4\}$ for $\{X_2, X_3, X_4\}$ in terms of X_1;
- Solve f_5 for X_5 in terms of X_4.

This requires a modification of the *d*-separation criterion [Spirtes, 1995, Forré and Mooij, 2017, Bongers et al., 2021].

Part III

Causal Semantics

Modeling interventions beyond SCMs/CBNs

Causality is about **change**.

How does the system react to interventions (externally imposed changes)?

How does a

1. change of (distributions of) exogenous variables, or
2. change of equations

affect the solution?

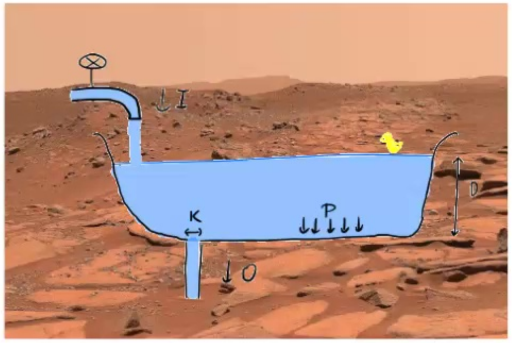
Caveat [Blom et al., 2021]

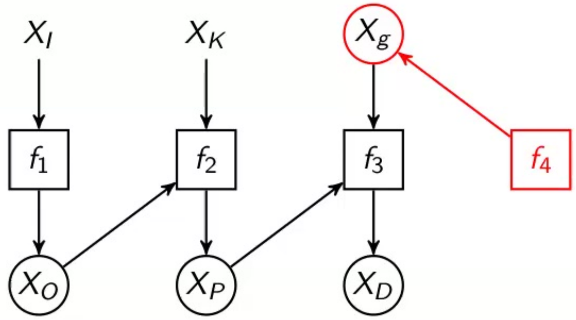
While it is common to consider perfect/surgical/hard interventions that set a certain endogenous variable to a certain value ("$\mathrm{do}(X = x)$"), we note that this notion is not well-defined in general, because there can be different ways of changing the equations to achieve this!

Modeling Interventions: $\mathrm{do}(X_g = g_{\mathrm{Mars}})$

What-if. . . ?

. . . we move the bathtubs to Mars?





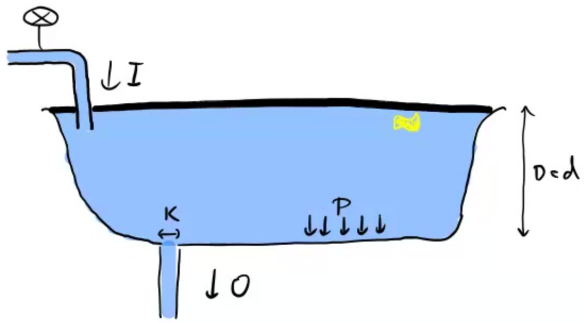
We can add one mechanism:

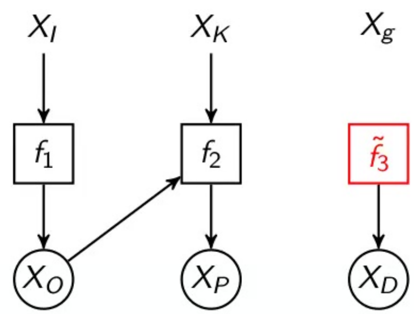
f_1 :	$0 = X_I - X_O$	at equilibrium, outflow equals inflow
f_2 :	$0 = X_K X_P - X_O$	outflow is proportional to pressure and drain diameter
f_3 :	$0 = X_g X_D - X_P$	pressure at drain proportional to depth and gravitational acceleration
f_4 :	$0 = X_g - g_{\mathrm{Mars}}$	gravitational acceleration set to Mars

Modeling Interventions: $\mathrm{do}(f_3 : X_D = x_D)$

What-if. . . ?

. . . we seal off the bathtub at height x_D and ensure the inflow is sufficiently large?





The mechanisms become:

$f_1 : \quad 0 = X_I - X_O$ — at equilibrium, outflow equals inflow

$f_2 : \quad 0 = X_K X_P - X_O$ — outflow is proportional to pressure and drain diameter

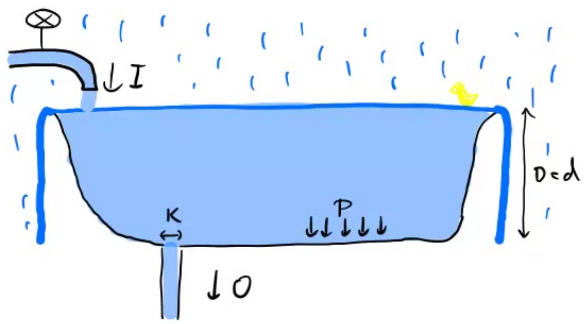
$f_3 : \quad$ ~~$0 = X_g X_D - X_P$~~ — ~~pressure at drain proportional to depth and gravitational acceleration~~

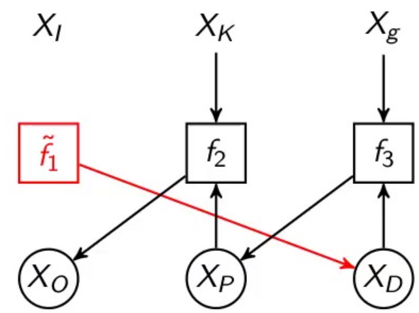
$\tilde{f}_3 : \quad 0 = X_D - x_D$ — water level equals bathtub height

Modeling Interventions: $\mathrm{do}(f_1 : X_D = x_D)$

What-if. . . ?

. . . we cut off a bathtub at height x_D and place it outside during heavy rainfall?



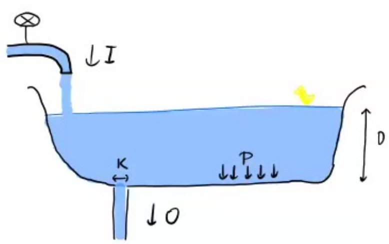


The mechanisms become:

f_1 :	~~$0 = X_I - X_O$~~	~~at equilibrium, outflow equals inflow~~
$\tilde{f}_1$:	$0 = X_D - x_D$	water level equals bathtub height
f_2 :	$0 = X_K X_P - X_O$	outflow is proportional to pressure and drain diameter
f_3 :	$0 = X_g X_D - X_P$	pressure at drain proportional to depth and gravitational acceleration

What changes due to the intervention?

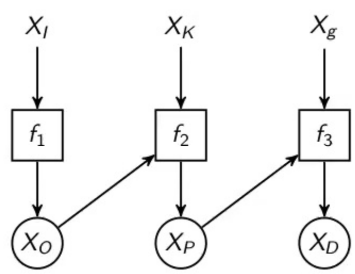
No intervention:



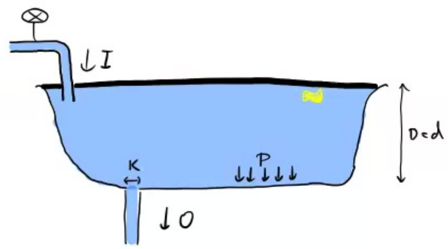
$f_1: \quad 0 = X_I - X_O$

$f_2: \quad 0 = X_K X_P - X_O$

$f_3: \quad 0 = X_g X_D - X_P$



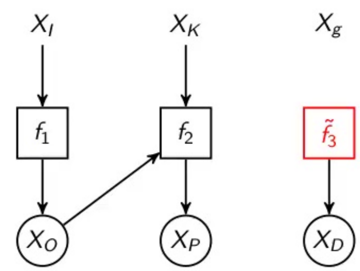
$\mathrm{do}(f_3 : X_D = x_D)$:



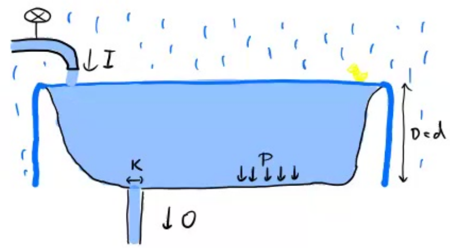
$f_1: \quad 0 = X_I - X_O$

$f_2: \quad 0 = X_K X_P - X_O$

$\tilde{f}_3: \quad 0 = X_D - x_D$



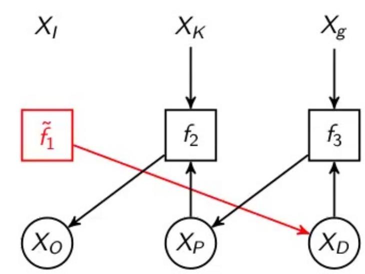
$\mathrm{do}(f_1 : X_D = x_D)$:



$\tilde{f}_1: \quad 0 = X_D - x_D$

$f_2: \quad 0 = X_K X_P - X_O$

$f_3: \quad 0 = X_g X_D - X_P$



For intervention $\mathrm{do}(f_1 : X_D = x_D)$, the causal ordering reverses and **the causal relations between the variables change drastically!**

Solutions and intervention effects

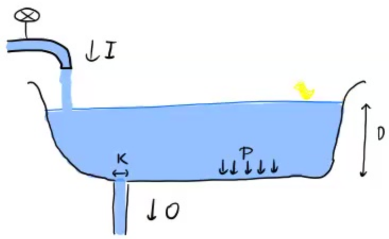
By solving the (intervened) systems of equations by hand, we can obtain the following solution functions.

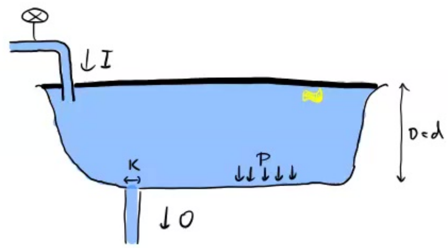
	X_P	X_O	X_D
observational	$\frac{X_I}{X_K}$	X_I	$\frac{X_I}{X_K X_g}$
$\mathrm{do}(X_g = x_g)$	$\frac{X_I}{X_K}$	X_I	$\boldsymbol{\frac{X_I}{X_K x_g}}$
$\mathrm{do}(f_3 : X_D = x_D)$	$\frac{X_I}{X_K}$	X_I	$\boldsymbol{x_D}$
$\mathrm{do}(f_1 : X_D = x_D)$	$\boldsymbol{X_g x_D}$	$\boldsymbol{X_K X_g x_D}$	$\boldsymbol{x_D}$

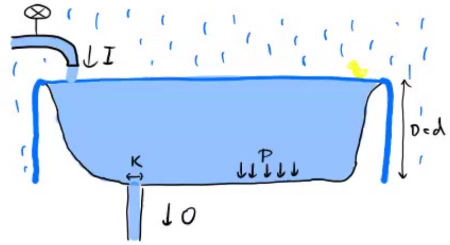
- Different interventions on exogenous distributions or mechanisms of the system lead to different changes in the values of some variables (the **effects** of the interventions).
- The endogenous distribution $\mathbb{P}(X_P, X_O, X_D)$ (or Markov kernel) changes as a result of the interventions.
- Note: the two interventions that set X_D to x_D are not equivalent!

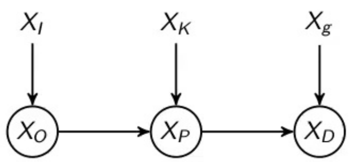
Can we model this with a CBN / acyclic SCM?

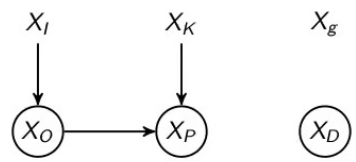
No intervention: $\mathrm{do}(X_D = x_D)$:

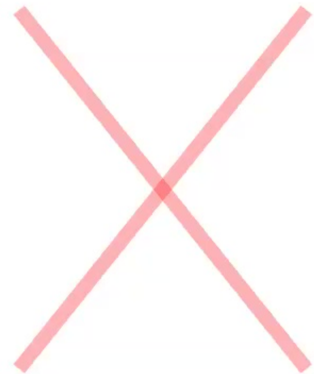












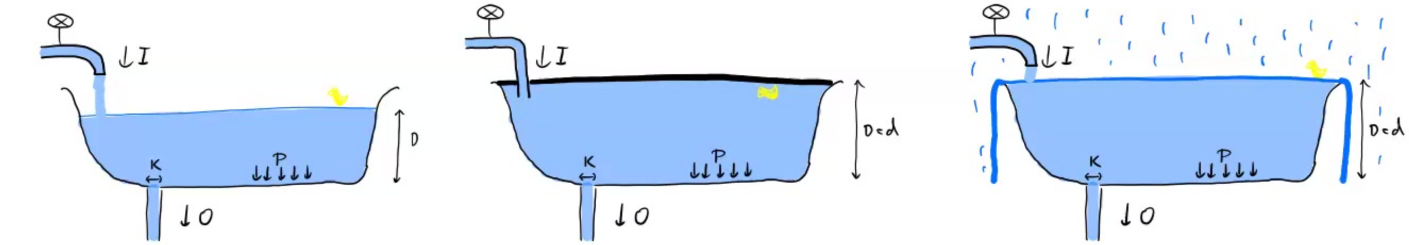
$$X_O = X_I$$
$$X_P = X_O / X_K$$
$$X_D = X_P / X_g$$

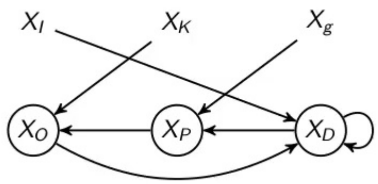
$$X_O = X_I$$
$$X_P = X_O / X_K$$
$$X_D = x_D$$

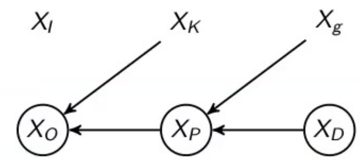
The reversal of the causal ordering under the intervention $\mathrm{do}(f_1 : X_D = x_D)$ cannot be represented appropriately!

Can we model this with an SCM?

No intervention: $\mathrm{do}(X_D = x_D)$:







$$X_O = X_K X_P$$
$$X_P = X_g X_D$$
$$X_D = X_D + (X_I - X_O)$$

$$X_O = X_K X_P$$
$$X_P = X_g X_D$$
$$X_D = x_D$$

Also a cyclic SCM cannot represent both interventions $\mathrm{do}(f_1 : X_D = x_D)$ and $\mathrm{do}(f_3 : X_D = x_D)$.

Conclusion

This shows that for certain cyclic causal systems,

- [Pearl, 2009]'s notion of "atomic/hard/perfect" intervention $\mathrm{do}(X_j = x_j)$ is ambiguous / inappropriate;
- CBNs and SCMs fail to represent how the system reacts to interventions.

To address this, we propose:

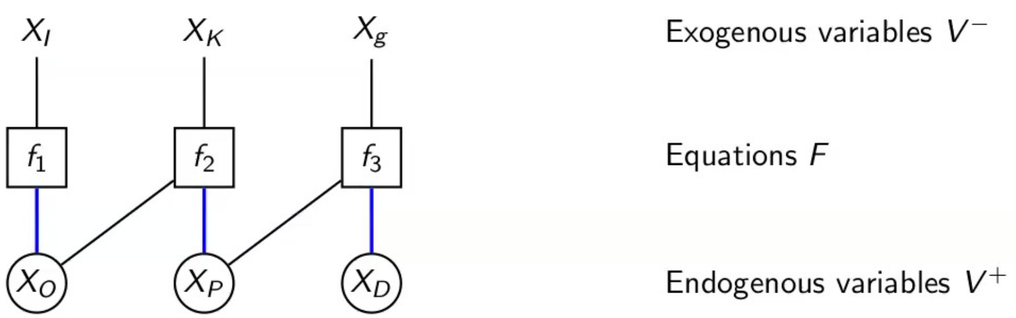
- to use a bipartite graphical model which also explicitly represents the causal mechanisms;
- to consider "atomic/hard/perfect" interventions $\mathrm{do}(f_i : X_j = x_j)$ which explicitly refer to the causal mechanism f_i that is replaced when setting X_j to the value x_j.

Part IV

General Theory

Bipartite Graph Terminology

Let $G = (V, F, E)$ be a bipartite graph with variable nodes V and equation nodes F and (undirected) edges $E \subseteq V \times F$. Partition $V = V^- \dot{\cup} V^+$ into **exogenous** variables V^- and **endogenous** variables V^+.



Walk	Sequence of adjacent edges on a graph.
Matching	Subset M of edges $v — f$ with $v \in V^+, f \in F$ such that no node occurs more than once.
Perfect matching	Matching such that each node in $V^+ \cup F$ is matched.
Alternating walk	Walk with edges that are alternatingly in M and not in M.
Closed alternating walk	Alternating walk that starts and ends in the same node.

Equivalence relation

We introduce an equivalence relation on the nodes of the bipartite graph.

Definition

Given a bipartite graph $G = (V^- \dot\cup V^+, F, E)$ with perfect matching M of $G_{V^+ \dot\cup F}$, define an equivalence relation $\sim$ on $V \dot\cup F$ as follows:
$a \sim b$ if $a — b \in M$, or if a and b lie on a closed alternating walk.

Lemma

The equivalence relation only depends on the bipartite graph G, but is independent of the choice of the perfect matching M.

Denote the equivalence class of a node $a \in V \dot\cup F$ as $[a]$.

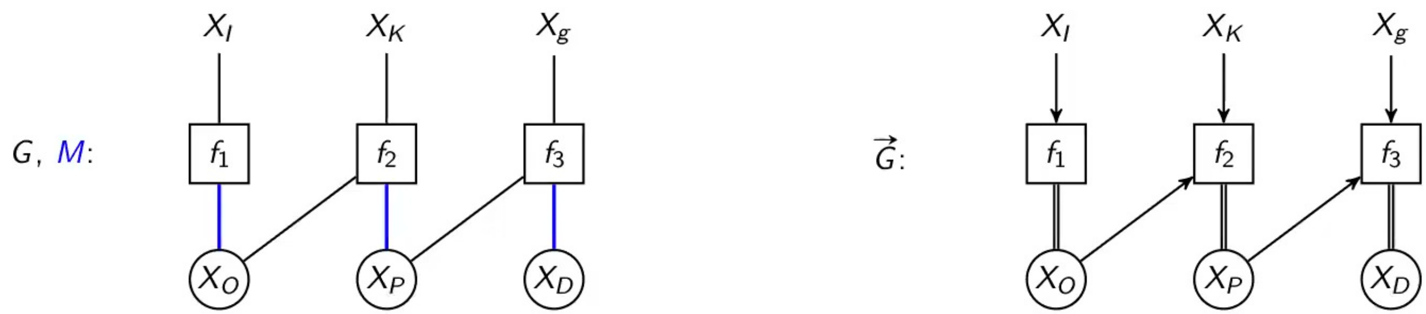
Partial Orientation

Use the equivalence relation to partially orient the bipartite graph G as $\vec{G}$:

Definition

For each edge $v - f \in E$ of G with $v \in V, f \in F$, "orient" it in $\vec{G}$ as:

$$\begin{cases} v \to f & \text{if } v \not\sim f, \\ v = f & \text{if } v \sim f. \end{cases}$$

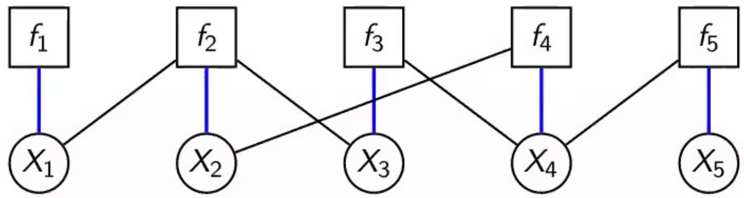


The mapping $G \mapsto \vec{G}$ is equivalent to Simon's causal ordering algorithm [Simon, 1953].

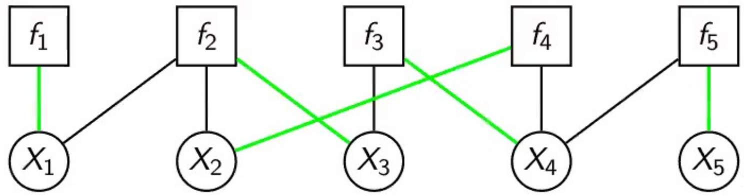
Example with Cycles

In case of cycles, multiple perfect matchings exist.

G, M_1:

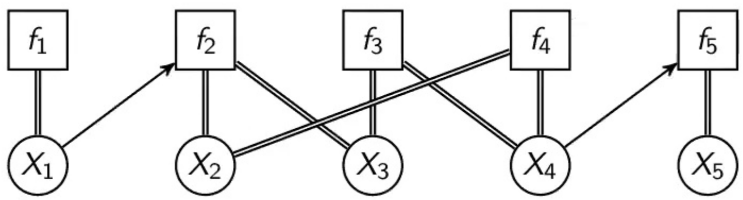


G, M_2:



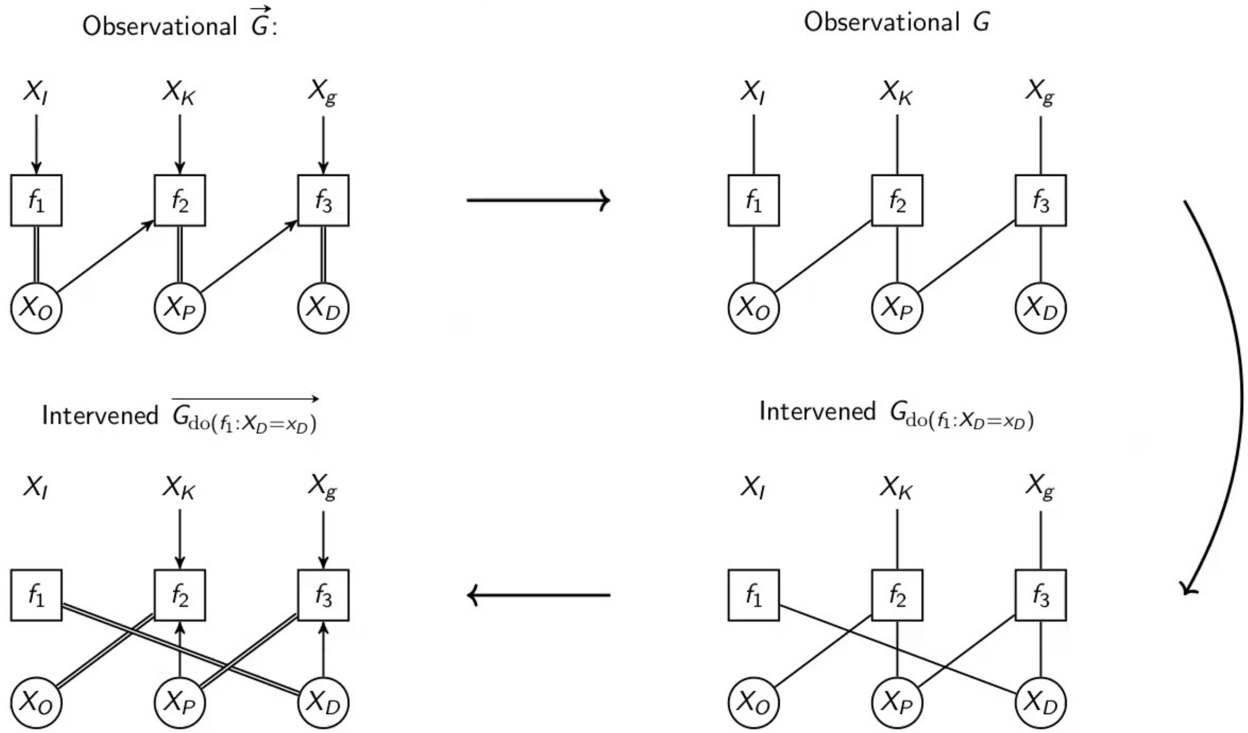
Both choices lead to the partial orientation:

$\vec{G}$:



Interventions may change the partial orientation

Interventions change the bipartite graph and the partial orientation.
Example: $\mathrm{do}(f_1 : X_D = x_D)$.



(Note: in CBNs and SCMs, the orientation is not changed by interventions!)

Local existence and uniqueness conditions

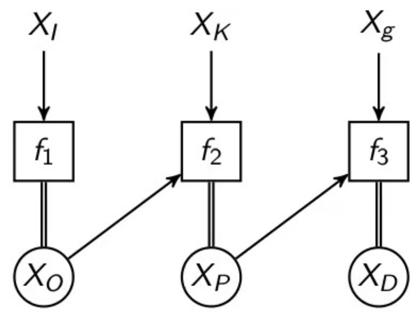
Definition

The *parents* of $[c]$ (for $c \in V \dot{\cup} F$) are the nodes
$\mathrm{pa}([c]) := \{b \in V \dot{\cup} F : \exists \tilde{c} \in [c] : b \to \tilde{c} \in \vec{G}\}$.

Definition (Clusterwise unique solvability)

A system of equations corresponding to $\vec{G}$ is **clusterwise uniquely solvable** if for each equivalence class $[c]$, the equations in $F \cap [c]$ can be solved for the endogenous variables $V^+ \cap [c]$ in terms of $\mathrm{pa}([c])$.

Example of clusterwise unique solvability



Equivalence classes:

$$[X_O] = \{X_O, f_1\} \quad \mathrm{pa}([X_O]) = \{X_I\}$$
$$[X_P] = \{X_P, f_2\} \quad \mathrm{pa}([X_P]) = \{X_K, X_O\}$$
$$[X_D] = \{X_D, f_3\} \quad \mathrm{pa}([X_D]) = \{X_g, X_P\}$$

f_1: $0 = X_I - X_O$ can be solved uniquely for X_O in terms of X_I

f_2: $0 = X_K X_P - X_O$ can be solved uniquely for X_P in terms of X_K and X_O

f_3: $0 = X_g X_D - X_P$ can be solved uniquely for X_D in terms of X_g and X_P

Assuming positivity, the bathtub is clusterwise uniquely solvable.

S-blocking

Definition (S-blocking [Forré and Mooij, 2017])

Consider a partially oriented bipartite graph $\vec{G}$. Consider a walk on $\vec{G}$. We can partition it into maximal segments $\sigma_1, \dots, \sigma_m$ such that each segment σ_i is a subwalk $\sigma_{i,l} \dots \sigma_{i,r}$ of maximal length that is entirely contained within one equivalence class of $\vec{G}$. We will call σ_1 and σ_m the end segments of the walk. For $Z \subseteq V$, the walk will be called *Z-s-blocked* or *s-blocked by Z* if:

1. at least one of the end nodes $\sigma_{1,l}$ or $\sigma_{m,r}$ is in Z, or
2. there is a non-collider segment σ_i with an outgoing directed edge (e.g., $\leftarrow \sigma_i$ or $\sigma_i \rightarrow$) and its corresponding endnode (i.e., $\sigma_{i,l}$ or $\sigma_{i,r}$, respectively) is in Z, or
3. there is a collider segment $\rightarrow \sigma_i \leftarrow$ and $[\sigma_i] \cap Z = \emptyset$.

Otherwise, the walk is called *Z-s-open* or *s-open given Z*.

S-separation

We can now define *s-separation* (in the usual way).

Definition (S-separation)

Let $\vec{G} = (V, F, E)$ be a partially oriented bipartite graph and $A, B, C \subseteq V$ (not necessarily disjoint) subset of nodes. We then say that: *A is s-separated from B given C in $\vec{G}$*, in symbols:

$$A \underset{\vec{G}}{\overset{s}{\perp}} B \mid C,$$

if every walk from a node in A to a node in B is s-blocked by C in $\vec{G}$.

This notion was already proposed as the "segment version of σ-separation" [Forré and Mooij, 2017] in another setting.

Global Markov Property

We can now prove:

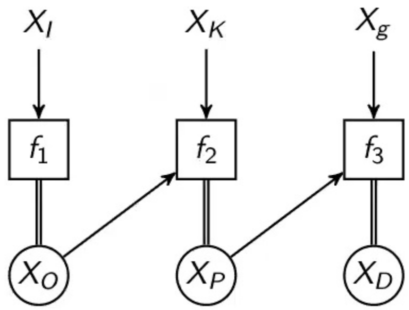
Theorem

If a system of equations is clusterwise uniquely solvable, and we put independent distributions on the exogenous variables, then we obtain a unique joint distribution $\mathbb{P}(X_V)$ that satisfies: for all $A, B, C \subseteq V$:

$$A \underset{\vec{G}}{\overset{s}{\perp}} B \mid C \implies X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \mid X_C.$$

The Markov property "propagates" independences through the equations following the partial ordering.

Example: Markov Property for the Bathtub



$$
\begin{aligned}
&X_K \sim \mathbb{P}(X_K) \\
&X_I \sim \mathbb{P}(X_I) \\
&X_g \sim \mathbb{P}(X_g) \\
f_1: \quad & 0 = X_I - X_O \\
f_2: \quad & 0 = X_K X_P - X_O \\
f_3: \quad & 0 = X_g X_D - X_P
\end{aligned}
$$

The Markov property applied to the bathtub states e.g.:

$$D \underset{\vec{G}}{\overset{s}{\perp}} O \mid P \implies X_D \underset{\mathbb{P}}{\perp\!\!\!\perp} X_O \mid X_P$$

which means

$$\mathbb{P}(X_D, X_O, X_P) = \mathbb{P}(X_D \mid X_P) \otimes \mathbb{P}(X_O, X_P)$$

Extended Global Markov Property

We can also derive a more general Markov property that treats some of the exogenous variables as non-random, using an extended notion of conditional independence [Forré, 2021].

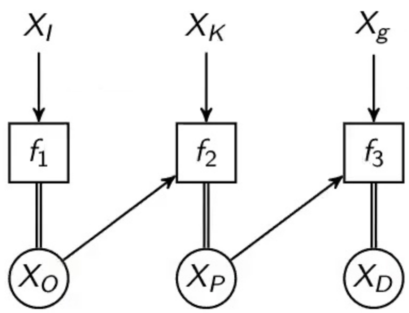
Theorem

If a system of equations is clusterwise uniquely solvable, and we treat exogenous variables $V^J \subseteq V^-$ as non-random and only put independent distributions on exogenous variables $V^- \setminus V^J$, we obtain a unique Markov kernel $\mathbb{P}(X_V \| X_{V^J})$ that satisfies: for all $A, B, C \subseteq V$ with $A \cap V^J = \emptyset$ and $V^J \subseteq (B \cup C)$:

$$A \overset{s}{\underset{\vec{G}}{\perp}} B \mid C \implies X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \mid X_C.$$

Here, independence of a non-random variable means that the Markov kernel is constant in that variable.

Example: Extended Markov Property for the Bathtub



$$
\begin{aligned}
& X_K \sim \mathbb{P}(X_K) \\
& X_I \text{ is exogenous non-random} \\
& X_g \sim \mathbb{P}(X_g) \\
f_1: \quad & 0 = X_I - X_O \\
f_2: \quad & 0 = X_K X_P - X_O \\
f_3: \quad & 0 = X_g X_D - X_P
\end{aligned}
$$

The extended Markov property applied to the bathtub states e.g.:

$$D \underset{\vec{G}}{\overset{s}{\perp}} I \mid P \implies X_D \underset{\mathbb{P}}{\perp\!\!\!\perp} X_I \mid X_P$$

which means there exists a Markov kernel $\mathbb{P}(X_D \,\|\, X_P)$ such that

$$\mathbb{P}(X_D, X_P \,\|\, X_I) = \mathbb{P}(X_D \,\|\, X_P) \otimes \mathbb{P}(X_P \,\|\, X_I)$$

Part V

Domain adaptation

Domain adaptation

- Simply put: the goal of domain adaptation is to relate the solution (or their distribution) in domain A with the solution (or their distribution) in domain B.
- Pearl's "do-calculus" formulates three rules for domain adaptation using causal Bayesian networks:

	Domain A	Domain B
Rule 1 (adding/removing observation)	observational	observational
Rule 2 (action/observation exchange)	observational	$\mathrm{do}(X_v = x_v)$
Rule 3 (adding/removing action)	observational	$\mathrm{do}(X_v = x_v)$

- We provide some examples of similar causal reasoning for bipartite causal graphs, for the equilibrated bathtub:

Domain A	Domain B
observational	$\mathrm{do}(X_g = x_g)$
observational	$\mathrm{do}(f_1 : X_D = x_D)$
observational	$\mathrm{do}(f_3 : X_D = x_D)$
$\mathrm{do}(f_1 : X_D = x_D)$	$\mathrm{do}(f_1 : X_D = x'_D)$

Domain adaptation in bipartite graphical causal models

By **jointly** modeling domains A and B, and **adding a domain indicator** R, we can relate the distributions via the Markov property. This provides a generalization of Pearl's do-calculus.

The general recipe is:

Domain adaptation: the recipe

1. Construct the joint model with an exogenous domain indicator R;
2. Construct a bipartite graph G^* representation of the joint model;
3. Run causal ordering to construct its partial orientation $\overrightarrow{G^*}$;
4. Check for clusterwise existence and uniqueness of solutions;
5. Apply the Markov property to $\overrightarrow{G^*}$.

Note: Apart from the check of the clusterwise existence and uniqueness, this is a purely graphical procedure.

Bathtub Example I: observational vs. $\mathrm{do}(X_g = x_g)$

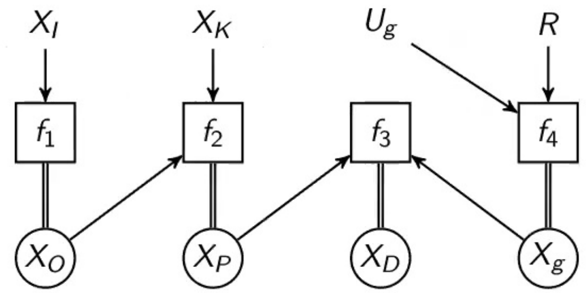
$X_K \sim \mathbb{P}(X_K)$, $X_I \sim \mathbb{P}(X_I)$, $U_g \sim \mathbb{P}(U_g)$

f_1: $0 = X_I - X_O$

f_2: $0 = X_K X_P - X_O$

f_3: $0 = X_g X_D - X_P$

f_4: $X_g = \begin{cases} U_g & R = A \\ x_g & R = B \end{cases}$



Applying the Markov property (using transition independence):

$$P, O \underset{\overrightarrow{G^*}}{\overset{s}{\perp}} R \implies X_P, X_O \perp\!\!\!\perp X_R \implies \mathbb{P}_A(X_P, X_O) = \mathbb{P}_B(X_P, X_O).$$

In Pearl's notation, the invariance under this intervention could be written:

$$\mathbb{P}(X_P, X_O) = \mathbb{P}(X_P, X_O \,|\, \mathrm{do}(X_g = x_g)).$$

An answer to what-if question

The equilibrium distribution of pressure and outflow does not change if we move the bathtubs to Mars.

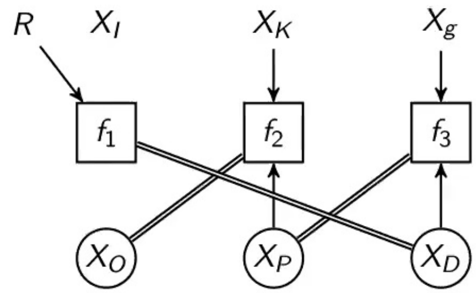
Bathtub Example IIIb: $\mathrm{do}(f_1 : X_D = x_D)$ vs. $\mathrm{do}(f_1 : X_D = x'_D)$

$X_K \sim \mathbb{P}(X_K)$, $X_I \sim \mathbb{P}(X_I)$, $X_g \sim \mathbb{P}(X_g)$

$$f_1: \quad 0 = \begin{cases} X_D - x_D & R = A \\ X_D - x'_D & R = B \end{cases}$$

$$f_2: \quad 0 = X_K X_P - X_O$$

$$f_3: \quad 0 = X_g X_D - X_P$$



$$O \underset{\overrightarrow{G^*}}{\overset{s}{\perp}} R \mid P \implies X_O \perp\!\!\!\perp X_R \mid X_P \implies$$

$$\begin{aligned} \mathbb{P}_A(X_O \mid \mathrm{do}(f_1 : X_D = x_D), X_P) &= \mathbb{P}_{AB}(X_O \mid X_P \,\|\, R = A) \\ &= \mathbb{P}_{AB}(X_O \mid X_P \,\|\, R = B) \\ &= \mathbb{P}_B(X_O \mid \mathrm{do}(f_1 : X_D = x'_D), X_P) \end{aligned}$$

An answer to what-if question

Bathtubs placed outside during heavy rainfall will yield the same conditional distribution of outflow given pressure, independent of their height.

Conclusion

We proposed a novel causal modeling framework using bipartite graphs that have **equation nodes** in addition to variable nodes.

- This allows us to **avoid ill-posedness of interventions**;
- We employ Simon's causal ordering algorithm to obtain a **partial orientation**;
- We stated a **Markov property** that propagates independences through the solutions of the equations, following the partial ordering;
- The Markov property enables causal reasoning about **domain adaptation** (extended do-calculus);
- The bipartite causal graphs allow us to naturally model equilibrium systems like the bathtub and other equilibrated systems;
- The framework reduces to causal Bayesian networks and (a)cyclic Structural Causal Models as **special cases**.
- There are many more systems like the bathtub (price-supply-demand, enzyme reaction, chemical reactions, ...) that can be modeled in this way; see also [Blom and Mooij, 2022].

References I

Blom, T. and Mooij, J. M. (2022).
Causality and independence in perfectly adapted dynamical systems.
arXiv.org preprint, arXiv:2101.11885v2 [stat.ML].

Blom, T., van Diepen, M. M., and Mooij, J. M. (2021).
Conditional independences and causal relations implied by sets of equations.
Journal of Machine Learning Research, 22(178):1–62.

Bongers, S., Forré, P., Peters, J., and Mooij, J. M. (2021).
Foundations of structural causal models with cycles and latent variables.
Annals of Statistics, 49(5):2885–2915.

Dulmage, A. L. and Mendelsohn, N. S. (1958).
Coverings of bipartite graphs.
Canadian Journal of Mathematics, 10:517–534.

Forré, P. (2021).
Transitional conditional independence.
arXiv.org preprint, arXiv:2104.11547 [math.ST].

Bathtub Example I: observational vs. $\mathrm{do}(X_g = x_g)$

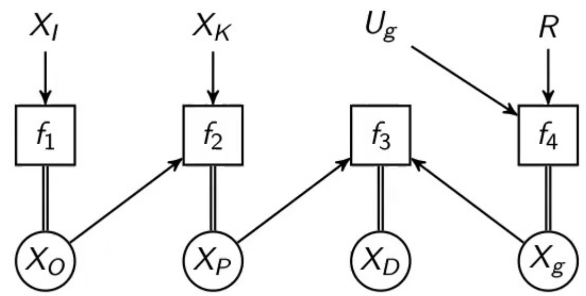
$X_K \sim \mathbb{P}(X_K)$, $X_I \sim \mathbb{P}(X_I)$, $U_g \sim \mathbb{P}(U_g)$

f_1: $0 = X_I - X_O$

f_2: $0 = X_K X_P - X_O$

f_3: $0 = X_g X_D - X_P$

f_4: $X_g = \begin{cases} U_g & R = A \\ x_g & R = B \end{cases}$



Applying the Markov property (using transition independence):

$$P, O \underset{\overrightarrow{G^*}}{\overset{s}{\perp}} R \implies X_P, X_O \perp\!\!\!\perp X_R \implies \mathbb{P}_A(X_P, X_O) = \mathbb{P}_B(X_P, X_O).$$

In Pearl's notation, the invariance under this intervention could be written:

$$\mathbb{P}(X_P, X_O) = \mathbb{P}(X_P, X_O \mid \mathrm{do}(X_g = x_g)).$$

An answer to what-if question

The equilibrium distribution of pressure and outflow does not change if we move the bathtubs to Mars.

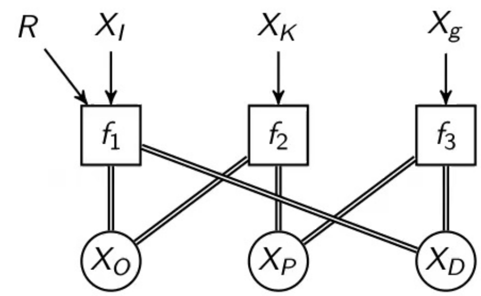
Bathtub Example IIIa: observational vs. $\mathrm{do}(f_1 : X_D = x_D)$

$X_K \sim \mathbb{P}(X_K)$, $X_I \sim \mathbb{P}(X_I)$, $X_g \sim \mathbb{P}(X_g)$

$$f_1: \quad 0 = \begin{cases} X_I - X_O & R = * \\ X_D - x_D & R = x_D \end{cases}$$

$$f_2: \quad 0 = X_K X_P - X_O$$

$$f_3: \quad 0 = X_g X_D - X_P$$



In this case, the Markov property does not yield non-trivial independences. Thus we cannot use it to relate the distributions in these two domains.

An answer to what-if question

If we place a bathtub cut off at height x_D outside during heavy rainfall, the entire equilibrium distribution may change.

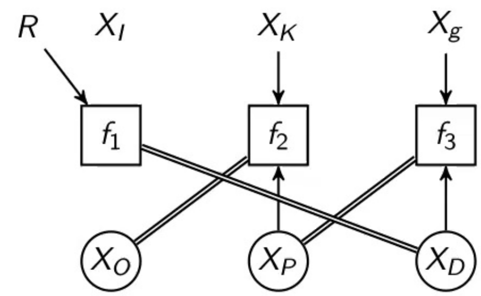
Bathtub Example IIIb: $\mathrm{do}(f_1 : X_D = x_D)$ vs. $\mathrm{do}(f_1 : X_D = x'_D)$

$X_K \sim \mathbb{P}(X_K)$, $X_I \sim \mathbb{P}(X_I)$, $X_g \sim \mathbb{P}(X_g)$

$$f_1: \quad 0 = \begin{cases} X_D - x_D & R = A \\ X_D - x'_D & R = B \end{cases}$$

$$f_2: \quad 0 = X_K X_P - X_O$$

$$f_3: \quad 0 = X_g X_D - X_P$$



$$O \underset{\overrightarrow{G^*}}{\overset{s}{\perp}} R \mid P \implies X_O \perp\!\!\!\perp X_R \mid X_P \implies$$

$$\begin{aligned} \mathbb{P}_A(X_O \mid \mathrm{do}(f_1 : X_D = x_D), X_P) &= \mathbb{P}_{AB}(X_O \mid X_P \parallel R = A) \\ &= \mathbb{P}_{AB}(X_O \mid X_P \parallel R = B) \\ &= \mathbb{P}_B(X_O \mid \mathrm{do}(f_1 : X_D = x'_D), X_P) \end{aligned}$$

An answer to what-if question

Bathtubs placed outside during heavy rainfall will yield the same conditional distribution of outflow given pressure, independent of their height.