Bipartite Graphical Causal Models

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

j.m.mooij@uva.nl

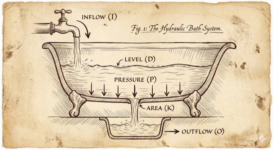
University of Amsterdam

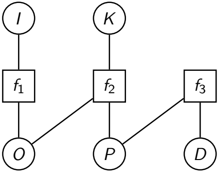
January 24th, 2026

This version has been updated slightly after the presentation (fixed Bernoulli's law by replacing "X_P" by "$\sqrt{X_P}$", renamed "non-collider*" into "blockable non-collider*" to avoid confusion, fixed typo in definition of "d^*-blocked")

Overview

- Causal Bayesian Networks (CBNs) and Structural Causal Models (SCMs) are popular causal modeling frameworks.
- But: systems may have "pathological" causal semantics.
- Example: bathtub or sink at equilibrium [Iwasaki and Simon, 1994].





- We propose to use **bipartite** causal graphs that contain variable nodes **and equation nodes**.
  1. Offers richer causal semantics;
  2. Reduces ambiguity surrounding the notion of perfect intervention;
  3. Comes with Markov property and do-calculus;
  4. Models various systems that exhibit "pathological" causal semantics.

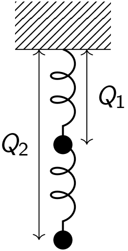
Part I

Context: Causal Modeling

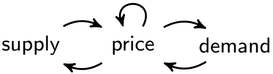
Cyclic causal relations

- Feedback in dynamical systems may (but need not!) induce cyclic causality at equilibrium.
- Fast dynamical interactions can lead to “instantaneous” causal cycles in dynamical systems.

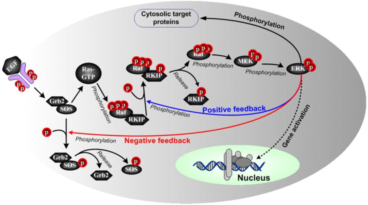
physics:



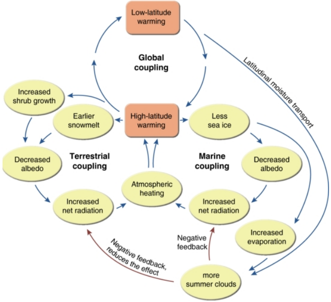
econometry:



biology, chemistry:

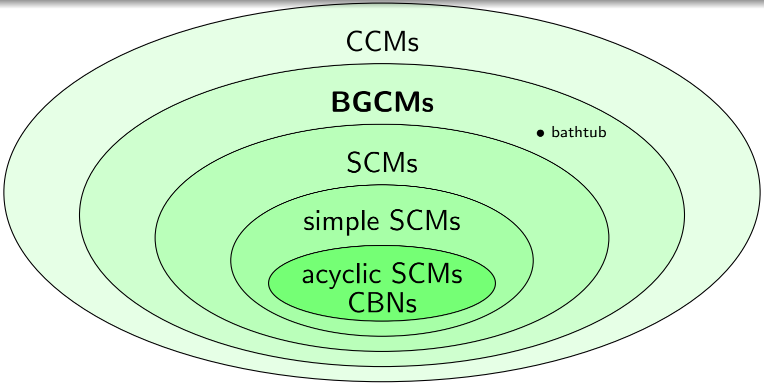


climatology:



In many applications, modeling causal cycles is essential.

Relations between (static) causal modeling frameworks



Acronym	Model class	Cycles?	Reference
CBN	causal Bayesian network	−	[Pearl, 2009]
acyclic SCM	acyclic structural causal model	−	[Pearl, 2009]
simple SCM	simple structural causal model	+	[Bongers et al., 2021]
SCM	structural causal model	+	[Bongers et al., 2021]
BGCM	**bipartite graphical causal model**	+	extends [Blom et al., 2021]
CCM	causal constraint models	+	[Blom et al., 2020]

BGCMs optimally balance model flexibility and causal reasoning power.

Context & Contribution

- Simon's **causal ordering** approach to causality [Simon, 1953] provides a fundamentally different perspective.
  - Given a **system of equations**, one deduces possible **causal interpretations** of the equations.
  - Each causal interpretation corresponds with a partitioning of the variables into **exogenous** and **endogenous** variables ("inputs" and "outputs") and a **partial causal ordering** of the variables.
- This matches notions of causality used by engineers and applied scientists (e.g. in Modelica).
- Combining causal ordering with the σ-separation criterion for SCMs [Forré and Mooij, 2017] provides a Markov property for causal systems represented as systems of equations [Blom et al., 2021].

Contributions of this presentation

- Formulate Markov property in terms of bipartite graph;
- Formulate causal reasoning (domain invariances/"do-calculus");
- Case study: Complete analysis of causal semantics of bathtub system.

Part II

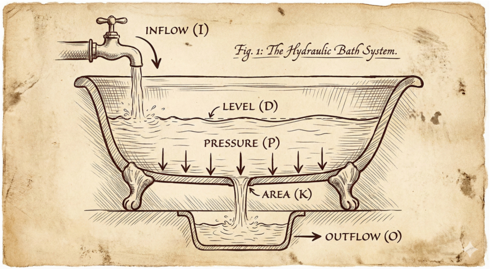
Causal Ordering Algorithm [Simon, 1953]

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



All bathtub drawings created by Google Gemini

Independent/modular/autonomous **mechanisms**:

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

$f_2: \quad 0 = X_K \sqrt{X_P} - X_O$ — Bernoulli's law: outflow is proportional to drain area and square root of pressure

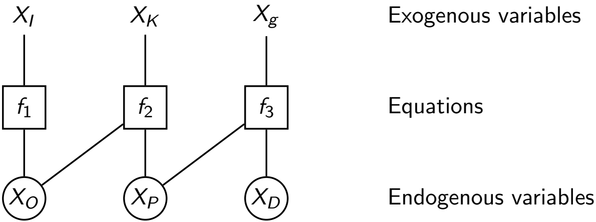
$f_3: \quad 0 = X_g X_D - X_P$ — Stevin's law: pressure is proportional to depth and gravitational acceleration

Bipartite Graphical Representation

The structure of the equations:

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

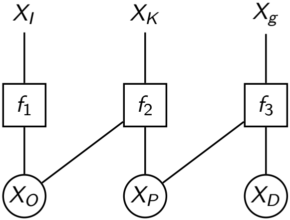
can be represented with an undirected bipartite graph:



Solving systems of equations

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

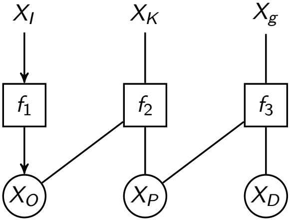
$$f_1: \quad 0 = X_I - X_O$$
$$f_2: \quad 0 = X_K\sqrt{X_P} - X_O$$
$$f_3: \quad 0 = X_g X_D - X_P$$



Solve in the following **ordering**:

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

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



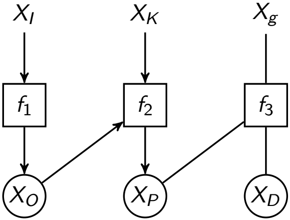
Solve in the following **ordering**:

1. Solve f_1 for X_O in terms of X_I: $\quad X_O = X_I$

Solving systems of equations

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

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



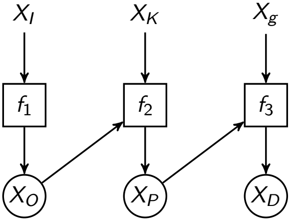
Solve in the following **ordering**:

1. Solve f_1 for X_O in terms of X_I: $X_O = X_I$
2. Solve f_2 for X_P in terms of X_O and X_K: $X_P = \frac{X_O^2}{X_K^2}$

Solving systems of equations

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

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



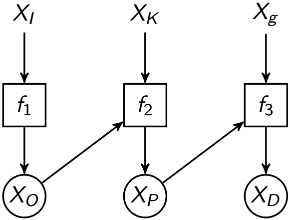
Solve in the following **ordering**:

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

Solving systems of equations

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

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



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: $X_P = \frac{X_O^2}{X_K^2} = \frac{X_I^2}{X_K^2}$
3. Solve f_3 for X_D in terms of X_P and X_g: $X_D = \frac{X_P}{X_g} = \frac{X_I^2}{X_K^2 X_g}$

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:

$$F : (x_I, x_K, x_g) \mapsto (x_I, x_K, x_g, x_O, x_P, x_D) = \left(x_I, x_K, x_g, x_I, \frac{x_I^2}{x_K^2}, \frac{x_I^2}{x_K^2 x_g} \right)$$

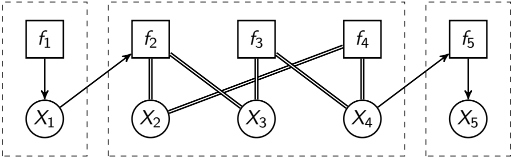
- If we assume that all exogenous variables are random variables that are independently distributed:

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

  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 the **exogenous distribution** $\mathbb{P}(X_I, X_K, X_g) = \mathbb{P}(X_I) \otimes \mathbb{P}(X_K) \otimes \mathbb{P}(X_g)$.
- 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.

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.

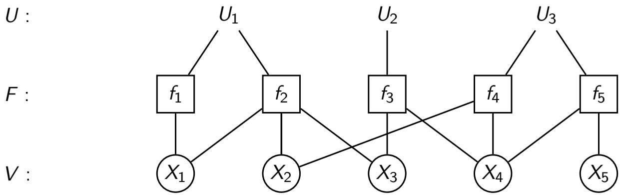


We can try to solve as follows:

- 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.

If each cluster can be solved uniquely, we get a unique global solution function (and hence joint distribution and Markov kernels).

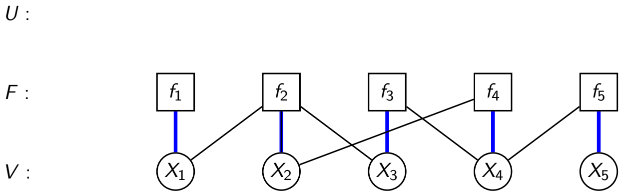
Input: bipartite graph $G = (V, F, E)$, exogenous variable nodes $U \subseteq V$.



Causal Ordering Algorithm [Simon, 1953, Nayak, 1995]

Input: bipartite graph $G = (V, F, E)$, exogenous variable nodes $U \subseteq V$.

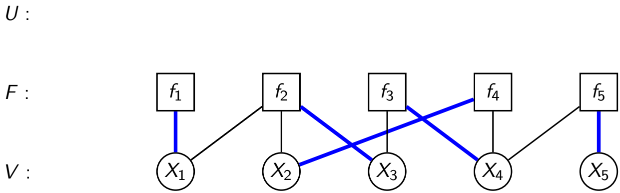
1. Pick perfect matching M of $G_{\setminus U}$ (Hopcroft-Carp algorithm)



Causal Ordering Algorithm [Simon, 1953, Nayak, 1995]

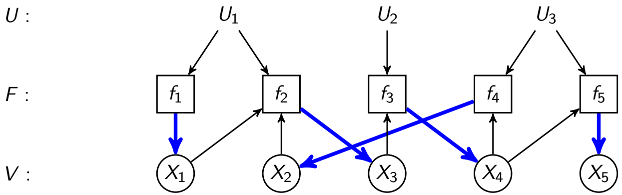
Input: bipartite graph $G = (V, F, E)$, exogenous variable nodes $U \subseteq V$.

1. Pick perfect matching M of $G_{\setminus U}$ (Hopcroft-Carp algorithm)



Input: bipartite graph $G = (V, F, E)$, exogenous variable nodes $U \subseteq V$.

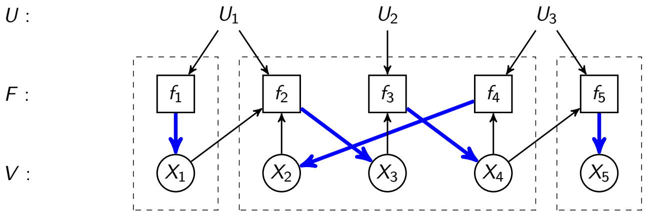
1. Pick perfect matching M of $G_{\setminus U}$ (Hopcroft-Carp algorithm)
2. Orient edges of G as $f \to v$ if in M, $f \leftarrow v$ if not in M



Causal Ordering Algorithm [Simon, 1953, Nayak, 1995]

Input: bipartite graph $G = (V, F, E)$, exogenous variable nodes $U \subseteq V$.

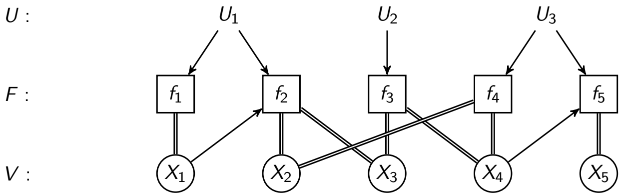
1. Pick perfect matching M of $G_{\setminus U}$ (Hopcroft-Carp algorithm)
2. Orient edges of G as $f \to v$ if in M, $f \leftarrow v$ if not in M
3. For $v \in V \setminus U$, define cluster $[v]$ as the strongly-connected component of v together with its F-parents



Causal Ordering Algorithm [Simon, 1953, Nayak, 1995]

Input: bipartite graph $G = (V, F, E)$, exogenous variable nodes $U \subseteq V$.

1. Pick perfect matching M of $G_{\setminus U}$ (Hopcroft-Carp algorithm)
2. Orient edges of G as $f \to v$ if in M, $f \leftarrow v$ if not in M
3. For $v \in V \setminus U$, define cluster $[v]$ as the strongly-connected component of v together with its F-parents
4. Replace directed edges within clusters by $=$ edges



Output: partially oriented bipartite graph $\vec{G}$.

Part III

Causal Semantics of the Bathtub

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?

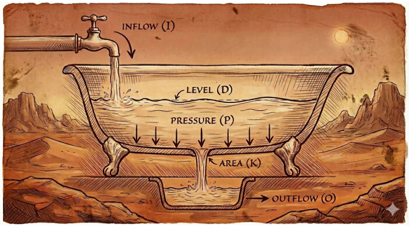
Caveat [Blom et al., 2021]

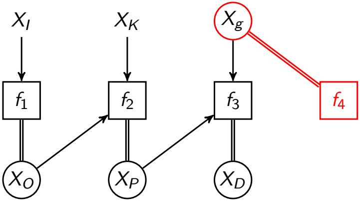
$\mathrm{do}(X = \xi)$ may be ambiguous in general!

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 : $0 = X_K \sqrt{X_P} - X_O$ — Bernoulli's law: outflow is proportional to drain area and square root of pressure

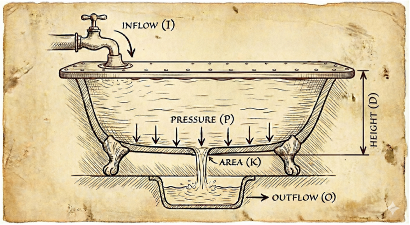
f_3 : $0 = X_g X_D - X_P$ — Stevin's law: pressure is proportional to depth and gravitational acceleration

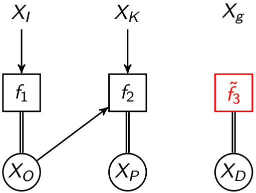
f_4 : $0 = X_g - g_{\mathrm{Mars}}$ — gravitational acceleration set to Mars

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

What-if...?

...we seal off the bathtub at height ξ_D and ensure it is completely filled?





We replace mechanism f_3:

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

$f_2 : \quad 0 = X_K \sqrt{X_P} - X_O$ — Bernoulli's law: outflow is proportional to drain area and square root of pressure

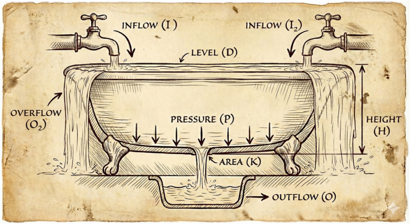
$f_3 : \quad$ ~~$0 = X_g X_D - X_P$~~ — ~~Stevin's law: pressure is proportional to depth and gravitational acceleration~~

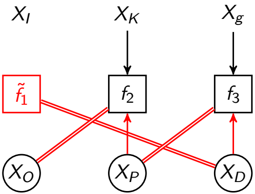
$\tilde{f}_3 : \quad 0 = X_D - \xi_D$ — water level equals bathtub height

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

What-if. . . ?

. . . we cut off a bathtub at height ξ_D and ensure it overflows?



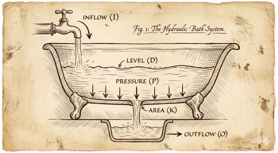


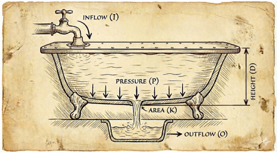
We replace mechanism f_1:

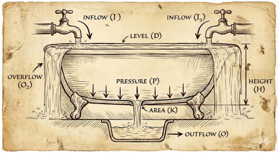
f_1 :	~~$0 = X_I - X_O$~~	~~at equilibrium, outflow equals inflow~~
$\tilde{f}_1$:	$0 = X_D - \xi_D$	water level equals bathtub height
f_2 :	$0 = X_K\sqrt{X_P} - X_O$	Bernoulli's law: outflow is proportional to drain area and square root of pressure
f_3 :	$0 = X_g X_D - X_P$	Stevin's law: pressure is proportional to depth and gravitational acceleration

What changes due to the intervention?

No intervention:	$\mathrm{do}(f_3 : X_D = \xi_D)$:	$\mathrm{do}(f_1 : X_D = \xi_D)$:







No intervention:

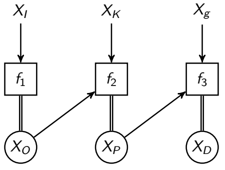
$$f_1 : \quad 0 = X_I - X_O$$
$$f_2 : \quad 0 = X_K \sqrt{X_P} - X_O$$
$$f_3 : \quad 0 = X_g X_D - X_P$$

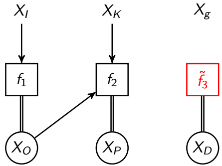
$\mathrm{do}(f_3 : X_D = \xi_D)$:

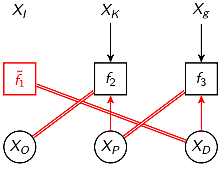
$$f_1 : \quad 0 = X_I - X_O$$
$$f_2 : \quad 0 = X_K \sqrt{X_P} - X_O$$
$$\tilde{f}_3 : \quad 0 = X_D - \xi_D$$

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

$$\tilde{f}_1 : \quad 0 = X_D - \xi_D$$
$$f_2 : \quad 0 = X_K \sqrt{X_P} - X_O$$
$$f_3 : \quad 0 = X_g X_D - X_P$$







For intervention $\mathrm{do}(f_1 : X_D = \xi_D)$, the causal ordering reverses!

Solutions and intervention effects

Solving the intervened systems of equations yields solution functions:

	X_O	X_P	X_D
observational	X_I	$\frac{X_I^2}{X_K^2}$	$\frac{X_I^2}{X_K^2 X_g}$
$\mathrm{do}(X_g = \xi_g)$	X_I	$\frac{X_I^2}{X_K^2}$	$\frac{X_I^2}{X_K^2 \xi_g}$
$\mathrm{do}(f_3 : X_D = \xi_D)$	X_I	$\frac{X_I^2}{X_K^2}$	ξ_D
$\mathrm{do}(f_1 : X_D = \xi_D)$	$X_K \sqrt{X_g \xi_D}$	$X_g \xi_D$	ξ_D

- This implies corresponding changes in the endogenous distribution $\mathbb{P}(X_P, X_O, X_D)$ or Markov kernel (e.g. $\mathbb{P}(X_P, X_O, X_D \,\|\, X_I)$).
- Note: the two interventions that set X_D to ξ_D have different effects: $\mathrm{do}(X_D = \xi_D)$ **is ambiguous**.
- Solution: specify hard interventions as $\mathrm{do}(f_i : X_j = \xi_j)$ [Blom et al., 2021].

All hard interventions for the bathtub

The following hard interventions can be implemented:

$\mathrm{do}(X_u = \xi_u)$	
$X_I = \xi_I$	✓
$X_K = \xi_K$	✓
$X_g = \xi_g$	✓

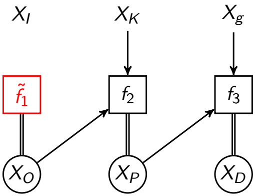
$\mathrm{do}(f_j : X_v = \xi_v)$	f_1	f_2	f_3
$X_O = \xi_O$	✓	$\lightning$	$\lightning$
$X_P = \xi_P$	✓	✓	$\lightning$
$X_D = \xi_D$	✓	✓	✓

Note:

- The interventions marked with $\lightning$ have no solutions.
- What would $\mathrm{do}(X_P = \xi_P)$, $\mathrm{do}(X_D = \xi_D)$ refer to?
- The bathtub **cannot** be modeled as a CBN or an SCM.

Implementing hard intervention $\mathrm{do}(f_1 : X_O = \xi_O)$

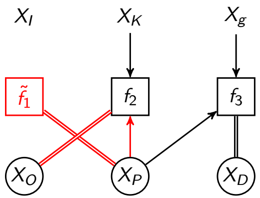
$$\tilde{f}_1 : \quad 0 = X_O - \xi_O$$
$$f_2 : \quad 0 = X_K \sqrt{X_P} - X_O$$
$$f_3 : \quad 0 = X_g X_D - X_P$$



Physical implementation:

- Divert away the inflow
- Add faucet with inflow $X_{I_2} = \xi_O$.

$$\tilde{f}_1 : \quad 0 = X_P - \xi_P$$
$$f_2 : \quad 0 = X_K \sqrt{X_P} - X_O$$
$$f_3 : \quad 0 = X_g X_D - X_P$$



Physical implementation:

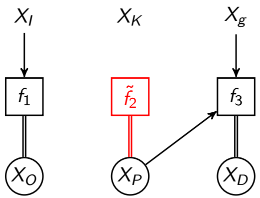
- Divert away the inflow
- Add faucet with inflow X_{I_2} (sufficiently large)
- Connect a pressure relieve valve to the bottom of the tub that activates if $X_P > \xi_P$.

Implementing hard intervention $\mathrm{do}(f_2 : X_P = \xi_P)$

$$f_1 : \quad 0 = X_I - X_O$$

$$\tilde{f}_2 : \quad 0 = X_P - \xi_P$$

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



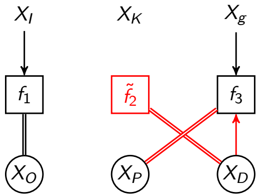
Physical implementation:

- Clogg the drain
- Reroute inflow directly to outflow through pipe, bypassing the tub and the drain
- Add another inflow (sufficiently large)
- Connect a pressure relieve valve to the bottom of the tub that activates if $X_P > \xi_P$.

$$f_1 : \quad 0 = X_I - X_O$$

$$\tilde{f}_2 : \quad 0 = X_D - \xi_D$$

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



Physical implementation:

- Clogg the drain
- Reroute inflow directly to outflow through pipe, bypassing the tub and the drain
- Add another inflow (sufficiently large)
- Cut off the bathtub at height ξ_D.

All Hard Intervention Effects

	X_O	X_P	X_D
observational	X_I	$\frac{X_I^2}{X_K^2}$	$\frac{X_I^2}{X_K^2 X_g}$
$\mathrm{do}(X_g = \xi_g)$	X_I	$\frac{X_I^2}{X_K^2}$	$\frac{X_I^2}{X_K^2 \xi_g}$
$\mathrm{do}(f_3 : X_D = \xi_D)$	X_I	$\frac{X_I^2}{X_K^2}$	ξ_D
$\mathrm{do}(X_K = \xi_K)$	X_I	$\frac{X_I^2}{\xi_K^2}$	$\frac{X_I^2}{\xi_K^2 X_g}$
$\mathrm{do}(f_2 : X_D = \xi_D)$	X_I	$X_g \xi_D$	ξ_D
$\mathrm{do}(f_2 : X_P = \xi_P)$	X_I	ξ_P	$\frac{\xi_P}{X_g}$
$\mathrm{do}(X_I = \xi_I)$	ξ_I	$\frac{\xi_I^2}{X_K^2}$	$\frac{\xi_I^2}{X_K^2 X_g}$
$\mathrm{do}(f_1 : X_O = \xi_O)$	ξ_O	$\frac{\xi_O^2}{X_K^2}$	$\frac{\xi_O^2}{X_K^2 X_g}$
$\mathrm{do}(f_1 : X_P = \xi_P)$	$\sqrt{\xi_P}\xi_K$	ξ_P	$\frac{\xi_P}{X_g}$
$\mathrm{do}(f_1 : X_D = \xi_D)$	$X_K \sqrt{X_g \xi_D}$	$X_g \xi_D$	ξ_D

Part IV

Markov Property for BGCMs

Extension of d-separation, I

Let $\overrightarrow{G} = (V, F, E)$ be a partially oriented bipartite graph, a, b nodes in $\overrightarrow{G}$, C a set of nodes in $\overrightarrow{G}$.

Definition (Ancestor*)

a is called **ancestor*** of b if there is a path in $\overrightarrow{G}$ starting at a ending at b consisting only of $\{\to, =\}$ edges. Example:

$$x_1 \to f_2 = x_3 = f_3 = x_4 \to f_5 \to x_5.$$

Definition ((Non)collider* patterns)

We extend the notions of (non)collider to the following subpaths:

collider*	$\to k \leftarrow$, $\to k_1 = \ldots = k_n \leftarrow$
blockable noncollider*	$\to k \to$, $\leftarrow k \leftarrow$, $\leftarrow k \to$, start/end node, $= k \to$, $\leftarrow k =$
unblockable noncollider*	all remaining patterns

Definition (d*-blocking)

A path (walk) in $\overrightarrow{G}$ is **d*-blocked by** C if it contains a blockable noncollider* in C, or a collider* that is not ancestor* of C.

Extension of d-separation, II

Let $\overrightarrow{G} = (V, F, E)$ be a partially oriented bipartite graph and A, B, C sets of nodes in $\overrightarrow{G}$.

Definition (d*-separation)

A **is d*****-separated from** B **given** C **in** $\overrightarrow{G}$, in symbols:

$$A \underset{\overrightarrow{G}}{\overset{d^*}{\perp}} B \mid C,$$

if every path from a node in A to a node in B is d*-blocked by C in $\overrightarrow{G}$.

We will usually only consider $A, B, C \subseteq V$.

Local existence and uniqueness conditions

Definition (Clusterwise unique solvability)

A cluster $[c]$ is **uniquely solvable** if the equations in $F \cap [c]$ can be solved for the endogenous variables $(V \setminus U) \cap [c]$ in terms of $\mathrm{pa}_{\vec{G}}([c])$, and this solution is unique.

Assumption (Local existence and uniqueness)

- *Exogenous variables are* **variation independent**: *their joint value space is a Cartesian product* $\prod_{u \in U} \mathcal{X}_u$.
- *The equations are* **clusterwise uniquely solvable**: *each cluster* $[c]$ *is uniquely solvable.*

Proposition

The assumption implies the existence and uniqueness of a global solution function, and hence of the joint distribution/Markov kernels.

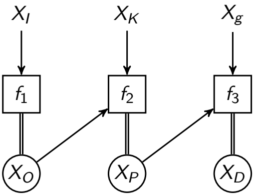
Global Markov Property

Theorem (Corollary of [Blom et al., 2021])

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:*

$$\forall A, B, C \subseteq V : \qquad A \underset{\vec{G}}{\overset{d^*}{\perp}} B \mid C \implies X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \mid X_C.$$

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



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

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

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

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

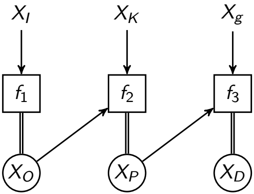
A more general Markov property allows treating some of the exogenous variables as non-random, using an **extended notion of conditional independence** [Constantinou and Dawid, 2017, Forré, 2021].

Theorem

If a system of equations is clusterwise uniquely solvable, and we treat exogenous variables $J \subseteq U$ as non-random and only put independent distributions on exogenous variables $U \setminus J$, we obtain a unique Markov kernel $\mathbb{P}(X_V \| X_J)$ that satisfies:

$$\forall A, B, C \subseteq V \text{ s.t. } A \cap J = \emptyset, J \subseteq (B \cup C): \\ A \overset{d^*}{\underset{\vec{G}}{\perp}} B \mid C \implies X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \mid X_C.$$

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



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

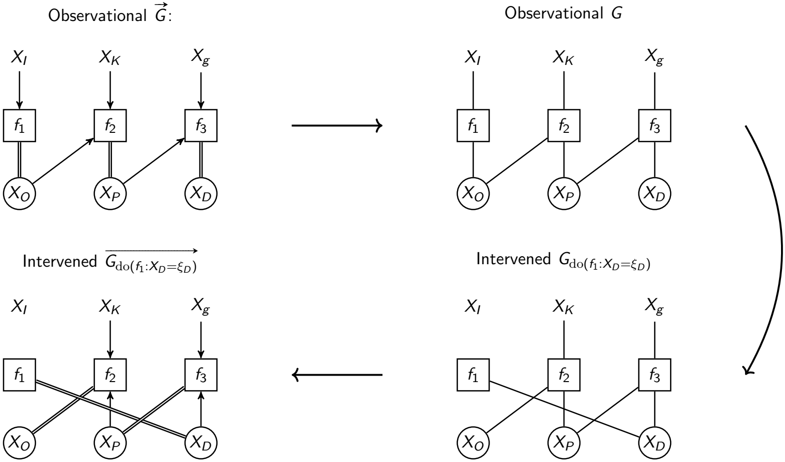
The extended Markov property applied to the bathtub states e.g.:

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

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

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

Hard interventions change the bipartite graph and the partial orientation, and hence the conditional independences.



Part V

Domain invariance ("do-calculus")

Domain invariances

- Goal: Relate the solutions in domain A with those in domain B. Which solution properties are **invariant** across domains?
- For causal Bayesian networks, Pearl's "do-calculus" formulates three rules for invariances of Markov kernels across domains:

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

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

Domain A	Domain B
observational	$\mathrm{do}(X_g = \xi_g)$
observational	$\mathrm{do}(f_1 : X_D = \xi_D)$
observational	$\mathrm{do}(f_3 : X_D = \xi_D)$
$\mathrm{do}(f_1 : X_D = \xi_D)$	$\mathrm{do}(f_1 : X_D = \xi'_D)$

Domain invariances 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 invariances: 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 unique solvability;
5. Apply the Markov property to $\overrightarrow{G^*}$.

Note: Apart from the check of the clusterwise unique solvability, this is a purely graphical procedure.

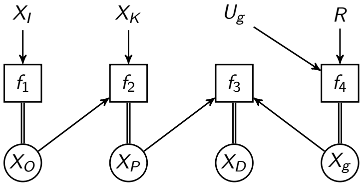
$X_K \sim \mathbb{P}(X_K),\ X_I \sim \mathbb{P}(X_I),\ U_g \sim \mathbb{P}(U_g)$

f_1: $0 = X_I - X_O$

f_2: $0 = X_K\sqrt{X_P} - X_O$

f_3: $0 = X_g X_D - X_P$

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



Applying the Markov property (using transition independence):

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

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

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

An answer to what-if question

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

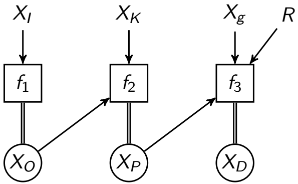
Bathtub Example II: observational vs. $\mathrm{do}(f_3 : X_D = \xi_D)$

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

f_1: $0 = X_I - X_O$

f_2: $0 = X_K\sqrt{X_P} - X_O$

f_3: $0 = \begin{cases} X_g X_D - X_P & R = * \\ X_D - \xi_D & R = \xi_D \end{cases}$



$$O \underset{\overrightarrow{G^*}}{\overset{d^*}{\perp}} R \,|\, D, P \implies X_O \perp\!\!\!\perp X_R \,|\, X_D, X_P \implies$$

$$\begin{aligned} \mathbb{P}_A(X_O \,|\, X_D = \xi_D, X_P) &= \mathbb{P}_{AB}(X_O \,|\, X_D = \xi_D, X_P \,\|\, R = *) \\ &= \mathbb{P}_{AB}(X_O \,|\, X_D = \xi_D, X_P \,\|\, R = \xi_D) \\ &= \mathbb{P}_B(X_O \,|\, \mathrm{do}(f_3 : X_D = \xi_D), X_P) \end{aligned}$$

An answer to what-if question

If we seal off the bathtub at height ξ_D and ensure it is completely filled, the conditional equilibrium distribution of outflow given pressure is the same as if we would just have observed a depth ξ_D.

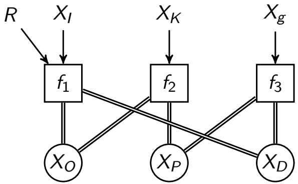
Bathtub Example IIIa: observational vs. $\mathrm{do}(f_1 : X_D = \xi_D)$

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

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

$$f_2: \quad 0 = X_K \sqrt{X_P} - X_O$$

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



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 let a bathtub that is cut off at height ξ_D overflow, the entire equilibrium distribution may change.

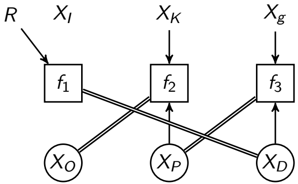
Bathtub Example IIIb: $\mathrm{do}(f_1 : X_D = \xi_D)$ vs. $\mathrm{do}(f_1 : X_D = \xi'_D)$

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

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

$$f_2: \quad 0 = X_K \sqrt{X_P} - X_O$$

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



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

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

An answer to what-if question

Overflowing bathtubs will yield the same conditional distribution of outflow given pressure, independent of their height.

Conclusion & Discussion

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

- Reduce **ambiguity** when specifying **interventions**;
- Simon's causal ordering algorithm defines **partial orientation**;
- Formulated **Markov property** that propagates conditional independences along the partial ordering;
- Markov property facilitates causal reasoning about **domain invariances** (extended "do-calculus");
- BGCMs **extend** CBNs, acylic SCMs, simple SCMs, SCMs;
- BGCMs naturally model **equilibrium systems** like the bathtub, and many more
- BGCMs may also naturally model dynamical systems (e.g., price-supply-demand, electronic circuits, enzyme reaction, chemical reactions) [Blom and Mooij, 2022];
- Future work: dynamical extensions to incorporate (stochastic) differential equations.

References I

Blom, T., Bongers, S., and Mooij, J. M. (2020).
Beyond structural causal models: Causal constraints models.
In Adams, R. P. and Gogate, V., editors, *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference (UAI-19)*, volume 115 of *Proceedings of Machine Learning Research*, pages 585–594. PMLR.

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.

Constantinou, P. and Dawid, A. P. (2017).
Extended conditional independence and applications in causal inference.
Annals of Statistics, 45(6):2618–2653.

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].

Forré, P. and Mooij, J. M. (2017).
Markov properties for graphical models with cycles and latent variables.
arXiv.org preprint, arXiv:1710.08775 [math.ST].

Iwasaki, Y. and Simon, H. A. (1994).
Causality and model abstraction.
Artificial intelligence, 67:143–194.

References III

Nayak, P. (1995).
Automated modeling of physical systems.
Springer-Verlag, Berlin.

Pearl, J. (2009).
Causality: Models, Reasoning and Inference.
Cambridge University Press.

Simon, H. A. (1953).
Causal ordering and identifiability.
In *Studies in Econometric Methods*, pages 49–74. John Wiley & Sons.

Dulmage-Mendelsohn decomposition

Clusterwise unique solvability is **necessary** for a global Markov property in the sense that:

- without local existence, no global existence;
- without local uniqueness, multiple solutions are possible, which allows for dependence with any variable in the model (the model is incomplete).

A useful generalization:

- If the bipartite graph has no perfect matching, one can choose a maximum matching and perform the Dulmage-Mendelsohn decomposition [Dulmage and Mendelsohn, 1958].
- This can be used to represent **overcomplete subsystems** (more equations than variables) and **undercomplete subsystems** (more variables than equations).
- A marginal Markov property for the (over)complete subsystems can be derived [Blom et al., 2021].