Towards Markov Properties for Continuous-Time Dynamical Systems

Joris Mooij

j.m.mooij@uva.nl

University of Amsterdam

December 3rd, 2022

Part I

Introduction

Motivation

Why **Markov properties?**

- Key concept in graphical approaches to causality.
- Allow to read off (conditional) **independences/invariances** from the (causal) graph.
- For example: *d*-**separation criterion** [Pearl, 1986] for (acyclic, causally sufficient, unconditioned, static) causal Bayesian networks and structural causal models.
- Powerful consequences:
  - **causal interpretation**: graphical definitions of indirect/direct causal relations and confounders,
  - **causal reasoning**: Pearl's do-calculus for causal domain adaptation,
  - **causal identification**: Tian's ID algorithm for identification of causal effects,
  - **causal discovery**: constraint-based approaches like PC and FCI algorithms,

  are all "corollaries" of the Markov property (and its completeness).

Extensions

This motivates the search for more general, powerful Markov properties.

- Various **notions of independence**:
  - purely probabilistic [Dawid, 1979]
  - purely deterministic (variation independence [Dawid, 2001])
  - mixed (e.g., transition independence [Forré, 2021]).

  The latter in particular allows to rigorously setup a decision-theoretic approach to causality [Dawid, 2002] where we distinguish **action** (context/regime/intervention) variables from **observation** variables and represent both graphically.
- Various **graphical representations**: DAGs, ADMGs, DGs, DMGs, AGs, CGs, BGs, ...
- Additional structure can be exploited (deterministic relations, context-specific independences, . . .).
- For **cyclic** causal systems, the d-separation criterion is not valid in general [Spirtes, 1995]. The (weaker) σ-**separation criterion** is more generally valid [Forré and Mooij, 2017, Bongers et al., 2021].

Causal modeling

- Causal Bayesian networks and structural causal models have fundamental limitations.
- More general alternative: Simon's **causal ordering** approach to causality [Simon, 1953].
- 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 usually deal with causality.
- Combining causal ordering with the σ-separation criterion provides a general Markov property for static causal systems represented as systems of equations [Blom et al., 2021].

But what about dynamics?

Continuous-time dynamical systems

Goal

Derive Markov properties for **continuous-time dynamical systems** represented as systems of **differential-algebraic equations** with (possibly random) initial conditions and (possibly random) exogenous processes.

Definition

Differential-algebraic equations (DAEs) are systems of equations involving processes and their time derivatives.

Example

Algebraic Equations:	Ordinary Differential Equations:	Differential-Algebraic Equations:
$X = f(Y)$	$\dot{X} = f(Y)$	$X = f(Y)$
$Y = g(X)$	$\dot{Y} = g(X)$	$\dot{Y} = g(X)$

- DAEs generalize ODEs and AEs;
- often encountered in engineering for modeling electrical circuits, constrained mechanical systems, chemical reactions, ...;
- inherently more complicated than ODEs.

Related work

Some sources of inspiration:

- Extensions of the causal ordering algorithm [Iwasaki and Simon, 1994] for DAEs.
- Application of causal ordering approach to perfectly adaptive systems [Blom and Mooij, 2022].
- Markov property for Structural Dynamical Causal Models [Bongers et al., 2022] (an extension of structural causal models to continuous-time dynamics).
- Other rich sources of ideas:
  - Mathematical literature on existence and uniqueness of solutions of DAEs;
  - Applied mathematics literature on automated solution of DAEs;
  - Engineering literature on DAEs.

Task: combine all these ideas to derive Markov properties for DAEs.

Part II

Causal Ordering for Static Systems

Markov property for recursive equations

For a system of algebraic equations of the form

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

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

Idea

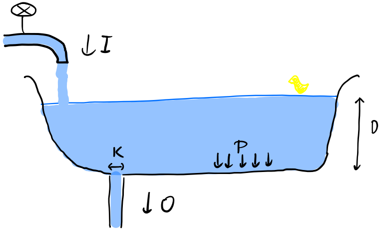
For any system of equations that **can be rewritten** in this canonical form, we obtain a Markov property.

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

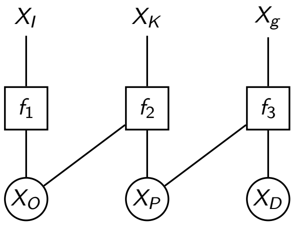
Assumption: endogenous variables do not cause exogenous variables.

Bipartite Graphical Representation

The structure of the equations:

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

can be represented with a bipartite graph:



Exogenous variables

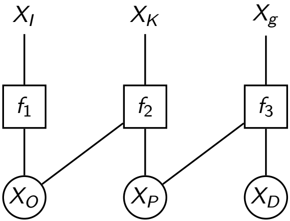
Equations

Endogenous variables

Solving systems of equations

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

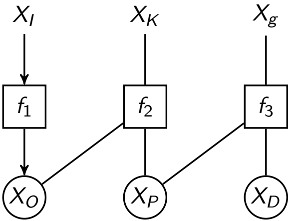
$$
\begin{aligned}
f_1 : &\quad 0 = X_I - X_O \\
f_2 : &\quad 0 = X_K X_P - X_O \\
f_3 : &\quad 0 = X_g X_D - X_P
\end{aligned}
$$



Solve in the following **ordering**:

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

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

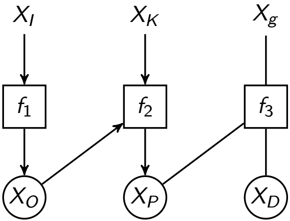


Solve in the following **ordering**:

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

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

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



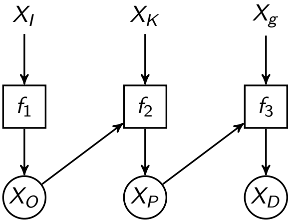
Solve in the following **ordering**:

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

Solving systems of equations

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

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



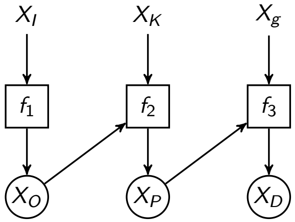
Solve in the following **ordering**:

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

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



Solve in the following **ordering**:

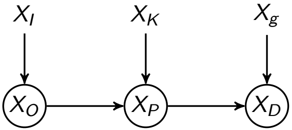
1. Solve f_1 for X_O in terms of X_I: $X_O = X_I$
2. Solve f_2 for X_P in terms of X_O and X_K: $X_P = \frac{X_O}{X_K}$
3. Solve f_3 for X_D in terms of X_P and X_g: $X_D = \frac{X_P}{X_g}$

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

Markov property from causal ordering

It also establishes a Markov property, as we have rewritten the equations in canonical form.

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



to read off (for example):

- $X_D \perp\!\!\!\perp X_O \mid X_P$;
- X_K does not cause X_O;
- X_g does not cause X_O, X_P.

(function nodes f_1, f_2, f_3 marginalized out for clarity)

Modeling interventions beyond SCMs/CBNs

Causality is about **change**.

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

How does a

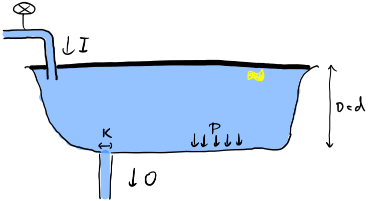
1. change of (distributions of) exogenous variables, or
2. change of equations

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!

Consider a "hard" intervention that enforces $X_D = d$ by replacing f_3.

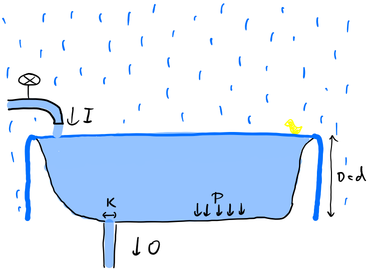


The mechanisms become:

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~~
$\tilde{f}_3$:	$0 = X_D - d$	water level equals bathtub height

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

Consider a "hard" intervention that enforces $X_D = d$ by replacing f_1.

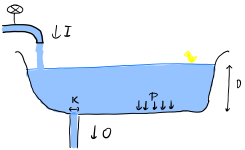


The mechanisms become:

f_1 :	~~$0 = X_I - X_O$~~	~~at equilibrium, outflow equals inflow~~
$\tilde{f}_1$:	$0 = X_D - d$	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?

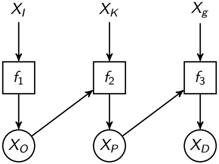
No intervention:



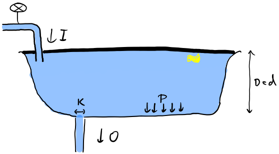
$f_1 : \quad 0 = X_I - X_O$

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

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



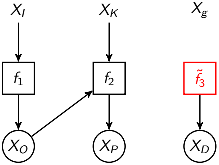
$\mathrm{do}(f_3 : X_D = d)$:



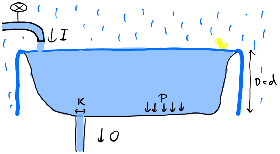
$f_1 : \quad 0 = X_I - X_O$

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

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



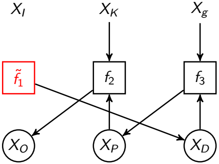
$\mathrm{do}(f_1 : X_D = d)$:



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

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

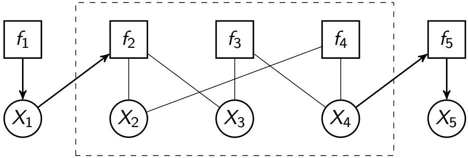
$f_3 : \quad 0 = X_g X_D - X_P$



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

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.



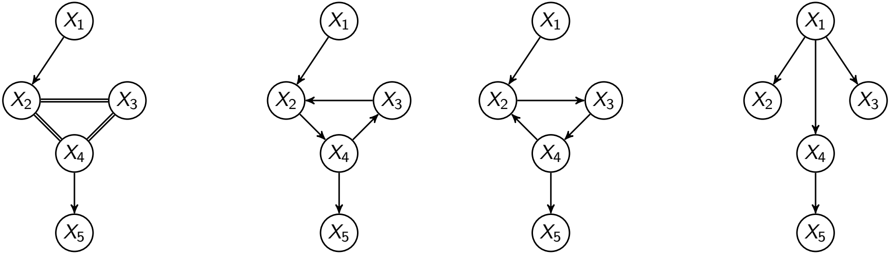
We can solve as follows:

- 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.

Local existence and uniqueness of the solutions for each cluster ($\{f_1, X_1\}$, $\{f_2, f_3, f_4, X_2, X_3, X_4\}$, and $\{f_5, X_5\}$) again implies a Markov property.

There are several equivalent formulations of the σ-separation criterion [Spirtes, 1995, Forré and Mooij, 2017, Bongers et al., 2021]:

s-separation: σ-separation: *d*-separation:



Necessity of solvability assumptions

Local existence and uniqueness for each cluster are necessary:

- 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:

- In case of **overcomplete subsystems** (more equations than variables) or **undercomplete subsystems** (more variables than equations), one can use the Dulmage-Mendelsohn decomposition [Dulmage and Mendelsohn, 1958] to get a Markov property [Blom et al., 2021].

Part III

Extension to Dynamical Systems

Main idea

Replace (static) variables with (dynamic) processes.

Fix a finite time interval $\mathbb{T} = [t_0, t_1] \subseteq \mathbb{R}$ and a probability space $(\Omega, \Sigma, \mathbb{P})$.

Static		Dynamic	
Variable	$X_i \in \mathcal{X}_i$	Trajectory	$X_i : \mathbb{T} \to \mathcal{X}_i$
Value space	$\mathcal{X}_i$	Trajectory space	$\mathcal{X}_i^{\mathbb{T}}$
Random variable	$X_i : \Omega \to \mathcal{X}_i$	Stochastic process	$X_i : \mathbb{T} \times \Omega \to \mathcal{X}_i$
		or random trajectory	$X_i : \Omega \to \mathcal{X}_i^{\mathbb{T}}$

Intuition

By replacing the spaces $\mathcal{X}_i$ by $\mathcal{X}_i^{\mathbb{T}}$ we **reduce** the dynamic case to the static case.

Mathematical details

We will typically assume that processes satisfy certain continuity or differentiability assumptions.

Denote by $C^m(\mathbb{T}, \mathbb{R}^n)$ the m-times continuously differentiable functions $\mathbb{T} \to \mathbb{R}^n$. Equipping this with the C^m-norm

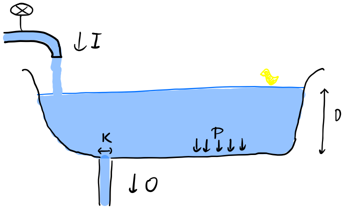
$$\|X\|^{(m)} := \sum_{i=1}^{m} \sup_{t \in \mathbb{T}} \|X^{(i)}(t)\|$$

(with $X^{(i)}$ the i'th derivative of X, and $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^n$) gives a Polish space, and with its Borel σ-algebra forms a standard measurable space.

Common operations (integration, differentiation, and evaluation) are continuous (and hence measurable).

Upshot

By restricting to sufficiently smooth trajectories we don't need to worry about measure theory.



Endogenous processes:

X_O	water outflow through drain
$X_{O'}$	its time-derivative
X_D	water depth
$X_{D'}$	its time-derivative
X_P	pressure at drain
$X_{P'}$	its time-derivative

Exogenous processes:

X_I	water inflow from faucet
X_K	drain size
X_g	gravitational acceleration

Exogenous variables:

$X_O(t_0)$	initial value for X_O
$X_D(t_0)$	initial value for X_D
$X_P(t_0)$	initial value for X_P

Mechanisms:

$$f_1' : X_{D'}(t) = \alpha_1 \big(X_I(t) - X_O(t)\big)$$

$$f_2' : X_{O'}(t) = \alpha_2 \big(\alpha_4 X_K(t) X_P(t) - X_O(t)\big)$$

$$f_3' : X_{P'}(t) = \alpha_3 \big(X_g(t) X_D(t) - X_P(t)\big)$$

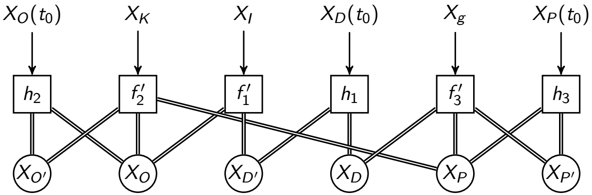
$$h_1 : \quad X_D(t) = X_D(t_0) + \int_{t_0}^{t} X_{D'}(\tau)\, d\tau$$

$$h_2 : \quad X_O(t) = X_O(t_0) + \int_{t_0}^{t} X_{O'}(\tau)\, d\tau$$

$$h_3 : \quad X_P(t) = X_P(t_0) + \int_{t_0}^{t} X_{P'}(\tau)\, d\tau$$

Result of causal ordering for dynamical bathtub

Causal ordering gives:

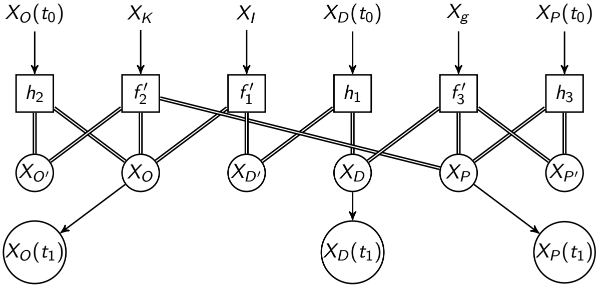


The Picard-Lindelöf theorem tells us that the initial value problem

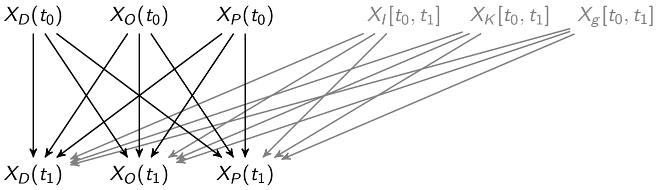
$$
\begin{aligned}
f_1' : \quad & \frac{d}{dt} X_D(t) = \alpha_1 \big(X_I(t) - X_O(t) \big) \\
f_2' : \quad & \frac{d}{dt} X_O(t) = \alpha_2 \big(\alpha_4 X_K(t) X_P(t) - X_O(t) \big) \\
f_3' : \quad & \frac{d}{dt} X_P(t) = \alpha_3 \big(X_g(t) X_D(t) - X_P(t) \big)
\end{aligned}
$$

have a unique global solution for any value of the initial condition $(X_D(t_0), X_O(t_0), X_P(t_0))$.
So we get a Markov property! But no new (conditional) independences...

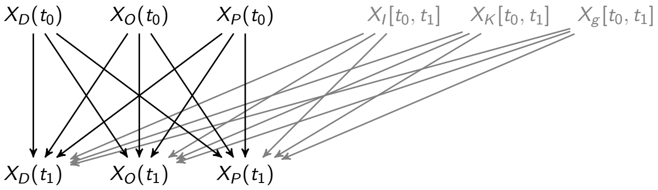
But we can do more.
Let us add "evaluation" variables $X_O(t_1), X_D(t_1), X_P(t_1)$ that evaluate the processes X_O, X_D, X_P at time t_1. We get:



After marginalizing out the middle layers:



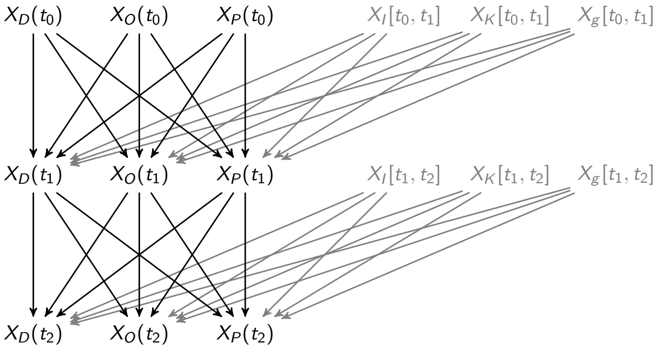
After marginalizing out the middle layers:



- We can add variables and processes for the time interval $[t_1, t_2]$.

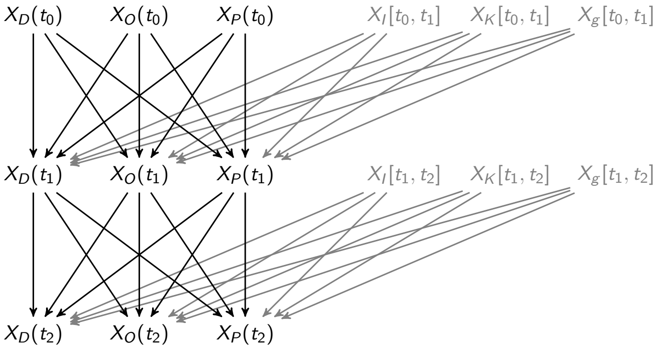
Markov property for dynamical bathtub II

After marginalizing out the middle layers:

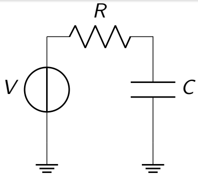


- We can add variables and processes for the time interval $[t_1, t_2]$.
- With d-separation, we get $X_{D,O,P}(t_2) \perp\!\!\!\perp X_{D,O,P}(t_0) \mid X_{D,O,P}(t_1)$.

After marginalizing out the middle layers:



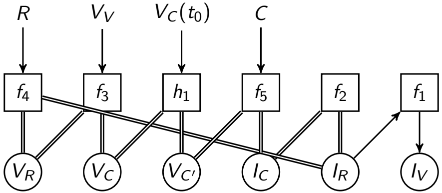
- We can add variables and processes for the time interval $[t_1, t_2]$.
- With d-separation, we get $X_{D,O,P}(t_2) \perp\!\!\!\perp X_{D,O,P}(t_0) \mid X_{D,O,P}(t_1)$.
- We may repeat this for additional time intervals $[t_2, t_3]$, $[t_3, t_4]$, etc.



The voltage source V_V is considered exogenous; the resistance R and capacitance C are considered exogenous and static; the initial condition $V_C(t_0)$ is considered exogenous.

f_1 :	$I_V(t) = I_R(t)$	Kirchoff's current law
f_2 :	$I_R(t) = I_C(t)$	Kirchoff's current law
f_3 :	$V_V(t) = V_R(t) + V_C(t)$	Kirchoff's voltage law
f_4 :	$RI_R(t) = V_R(t)$	Ohm's law
f_5 :	$CV_{C'}(t) = I_C(t)$	Ideal capacitor
h_1 :	$V_C(t) = V_C(t_0) + \int_{t_0}^{t} V_{C'}(\tau)\, d\tau$	

Causal ordering yields:

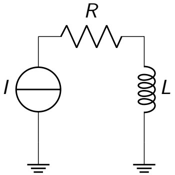


Solve clusters:

$$V_C(t) = V_C(t_0) + \int_{t_0}^{t} \frac{1}{RC}\big(V_V(\tau) - V_C(\tau)\big)\, d\tau$$
$$I_V(t) = R^{-1}\big(V_V(t) - V_C(t)\big)$$

Unique solutions exist (again by Picard-Lindelöf).

Conclusion: $I_V \perp\!\!\!\perp \{R, V_R, V_V, V_C, V_C(t_0), C, I_C\} \mid I_R$.



The current source I_I is considered exogenous; the resistance R and inductance L are considered exogenous and static.

$$f_1: \quad I_I(t) = I_R(t)$$ Kirchoff's current law

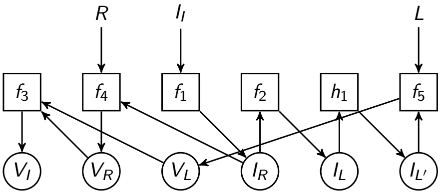
$$f_2: \quad I_R(t) = I_L(t)$$ Kirchoff's current law

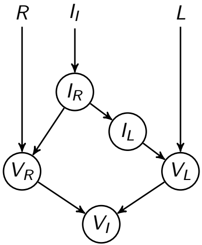
$$f_3: \quad V_I(t) = V_R(t) + V_L(t)$$ Kirchoff's voltage law

$$f_4: \quad RI_R(t) = V_R(t)$$ Ohm's law

$$f_5: \quad LI_{L'}(t) = V_L(t)$$ Ideal inductor

$$h_1: \quad I_{L'}(t) = \frac{d}{dt} I_L(t)$$





Surprise: We get an acyclic causal ordering yielding a non-trivial Markov property! For example, we read off that:

- L does not cause V_R,
- R does not cause V_L,
- $V_I \perp\!\!\!\perp L \mid V_L$,
- ...

Note: $I_L(t_0)$ is *not* an exogenous degree of freedom! Hence, [Iwasaki and Simon, 1994]'s dynamical causal ordering algorithm does not handle this case correctly.

Conclusion and Discussion

- By **replacing variables by processes** we generalized existing **Markov properties** to apply to **continuous-time dynamical systems** modeled by differential-algebraic equations.
- This yields more explicit (sometimes surprising) **causal interpretations** of such systems.
- Our framework also allows one to reason graphically about **(partial) equilibration**, **interventions** and **domain adaptation**.
- Two disadvantages of dynamical systems compared to static (equilibrium) systems:
  - Typical systems entail more (conditional) independences at equilibrium (because all time derivatives vanish at equilibrium);
  - Any conditional independence for processes requires one to condition on the **entire trajectory** (for example, $V_I \perp\!\!\!\perp L \mid V_L$ means we condition on $V_L(t)$ for $t \in [t_0, t_1]$). Challenging to test with finite samples!
- We could proceed and formulate a do-calculus, an ID algorithm, and causal discovery algorithms, but is it worth doing this. . . ?

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., Blom, T., and Mooij, J. M. (2022).
Causal modeling of dynamical systems.
arXiv.org preprint, arXiv:1803.08784v4 [cs.AI].

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.

Dawid, A. P. (1979).
Conditional independence in statistical theory.
Journal of the Royal Statistical Society: Series B (Methodological), 41(1):1–15.

References II

Dawid, A. P. (2001).
Some variations on variation independence.
In *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics*, volume R3 of *Proceedings of Machine Learning Research*, pages 83–86.

Dawid, A. P. (2002).
Influence diagrams for causal modelling and inference.
International Statistical Review, 70(2):161–189.

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].

References III

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

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

Pearl, J. (1986).
A contstraint propagation approach to probabilistic reasoning.
Uncertainty in Artificial Intelligence, pages 357–370.

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

Spirtes, P. (1995).
Directed cyclic graphical representations of feedback models.
In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pages 499–506.