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) \textit{independences/invariances} from the (causal) graph.
- For example: \textit{d-separation criterion} [Pearl, 1986] for (acyclic, causally sufficient, unconditioned, static) causal Bayesian networks and structural causal models.
- Powerful consequences:
  - \textbf{causal interpretation}: graphical definitions of indirect/direct causal relations and confounders,
  - \textbf{causal reasoning}: Pearl’s do-calculus for causal domain adaptation,
  - \textbf{causal identification}: Tian’s ID algorithm for identification of causal effects,
  - \textbf{causal discovery}: constraint-based approaches like PC and FCI algorithms,

are all “corollaries” of the Markov property (and its completeness).
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) $\sigma$-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 $\sigma$-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

<table>
<thead>
<tr>
<th>Algebraic Equations:</th>
<th>Ordinary Differential Equations:</th>
<th>Differential-Algebraic Equations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X = f(Y)$</td>
<td>$\dot{X} = f(Y)$</td>
<td>$X = f(Y)$</td>
</tr>
<tr>
<td>$Y = g(X)$</td>
<td>$\dot{Y} = g(X)$</td>
<td>$\dot{Y} = g(X)$</td>
</tr>
</tbody>
</table>

- 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
For a system of algebraic equations of the form

\[ X_1 = f_1(E_1) \]
\[ X_2 = f_2(X_1, E_2) \]
\[ X_3 = f_3(X_1, X_2, E_3) \]
\[ X_4 = f_4(X_1, X_2, X_3, E_4) \]
\[ \ldots \]
\[ X_p = f_p(X_1, X_1, \ldots, X_{p-1}, E_p) \]

with \( E_1, \ldots, E_p \) 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.
Example: Bathtub (Static)

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:

\[ f_1 : \quad 0 = X_I - X_O \quad \text{at equilibrium, outflow equals inflow} \]
\[ f_2 : \quad 0 = X_K X_P - X_O \quad \text{outflow is proportional to pressure and drain diameter} \]
\[ f_3 : \quad 0 = X_g X_D - X_P \quad \text{pressure at drain proportional to depth and gravitational acceleration} \]

Assumption: endogenous variables do not cause exogenous variables.
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:

- **Exogenous variables**
  - \( X_I \), \( X_K \), \( X_g \)
  - \( f_1 \), \( f_2 \), \( f_3 \)
- **Endogenous variables**
  - \( X_O \), \( X_P \), \( X_D \)
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**:
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 \( X_I \): \( X_O = X_I \)
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 \( X_K \): \( X_P = \frac{X_O}{X_K} \)
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 \( X_g \): \( X_D = \frac{X_P}{X_g} \)
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 \( (\forall x_I,x_K,x_g > 0) \).
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:

![Graph with nodes labeled \(X_I, X_K, X_g, X_O, X_P, X_D\)]

...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)
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 ("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 $f_3$. 

\[ f_1 : 0 = X_I - X_O \]  
\[ f_2 : 0 = X_K X_P - X_O \]  
\[ f_3 : 0 = X_g X_D - X_P \]  
\[ \tilde{f}_3 : 0 = X_D - d \]  

The mechanisms become:

- $f_1$ at equilibrium, outflow equals inflow
- $f_2$ outflow is proportional to pressure and drain diameter
- $f_3$ pressure at drain proportional to depth and gravitational acceleration
- $\tilde{f}_3$ water level equals bathtub height
Modeling Interventions: $\text{do}(f_1 : X_D = d)$

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

The mechanisms become:

\begin{align*}
    f_1 &: \quad 0 = X_I - X_O \\
    \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
\end{align*}

- at equilibrium, outflow equals inflow
- water level equals bathtub height
- outflow is proportional to pressure and drain diameter
- pressure at drain proportional to depth and gravitational acceleration
What changes due to the intervention?

No intervention:

\[ f_1 : 0 = X_I - X_O \]
\[ f_2 : 0 = X_K X_P - X_O \]
\[ f_3 : 0 = X_g X_D - X_P \]

\[ \hat{f}_3 : 0 = X_D - d \]

\[ \tilde{f}_1 : 0 = X_D - d \]
\[ \tilde{f}_3 : 0 = X_D - d \]

For intervention \( \text{do}(f_1 : X_D = x_d) \), 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 $X_4$. 
Several formulations of the Markov property

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 \(\sigma\)-separation criterion [Spirtes, 1995, Forré and Mooij, 2017, Bongers et al., 2021]:

- **s-separation:**
  
- **\(\sigma\)-separation:**
  
- **d-separation:**
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
Replacing variables by processes

Main idea

Replace (static) variables with (dynamic) processes.

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

<table>
<thead>
<tr>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Trajectory</td>
</tr>
<tr>
<td>$X_i \in \mathcal{X}_i$</td>
<td>$X_i : \mathbb{T} \rightarrow \mathcal{X}_i$</td>
</tr>
<tr>
<td>Value space</td>
<td>Trajectory space</td>
</tr>
<tr>
<td>$\mathcal{X}_i$</td>
<td>$\mathcal{X}_i^\mathbb{T}$</td>
</tr>
<tr>
<td>Random variable</td>
<td>Stochastic process or random</td>
</tr>
<tr>
<td>$X_i : \Omega \rightarrow \mathcal{X}_i$</td>
<td>trajectory $X_i : \Omega \rightarrow \mathcal{X}_i^\mathbb{T}$</td>
</tr>
</tbody>
</table>

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 $\sigma$-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.
Example: Bathtub (Dynamic)

Endogenous processes:

\[ X_O \text{ water outflow through drain} \]
\[ X_O' \text{ its time-derivative} \]
\[ X_D \text{ water depth} \]
\[ X_D' \text{ its time-derivative} \]
\[ X_P \text{ pressure at drain} \]
\[ X_P' \text{ its time-derivative} \]

Exogenous processes:

\[ X_I \text{ water inflow from faucet} \]
\[ X_K \text{ drain size} \]
\[ X_g \text{ gravitational acceleration} \]

Exogenous variables:

\[ X_O(t_0) \text{ initial value for } X_O \]
\[ X_D(t_0) \text{ initial value for } X_D \]
\[ X_P(t_0) \text{ initial value for } X_P \]

Mechanisms:

\[ f_1' : X_D'(t) = \alpha_1 (X_I(t) - X_O(t)) \]
\[ f_2' : X_O'(t) = \alpha_2 (\alpha_4 X_K(t) X_P(t) - X_O(t)) \]
\[ f_3' : X_P'(t) = \alpha_3 (X_g(t) X_D(t) - X_P(t)) \]

\[ h_1 : X_D(t) = X_D(t_0) + \int_{t_0}^{t} X_D'(\tau) \, d\tau \]
\[ h_2 : X_O(t) = X_O(t_0) + \int_{t_0}^{t} X_O'(\tau) \, d\tau \]
\[ h_3 : X_P(t) = X_P(t_0) + \int_{t_0}^{t} X_P'(\tau) \, d\tau \]
The Picard-Lindelöf theorem tells us that the initial value problem

\[ f_1' : \quad \frac{d}{dt} X_D(t) = \alpha_1 (X_I(t) - X_O(t)) \]
\[ f_2' : \quad \frac{d}{dt} X_O(t) = \alpha_2 (\alpha_4 X_K(t) X_P(t) - X_O(t)) \]
\[ f_3' : \quad \frac{d}{dt} X_P(t) = \alpha_3 (X_g(t) X_D(t) - X_P(t)) \]

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:

$$X_D(t_0), X_O(t_0), X_P(t_0)$$

$$X_D(t_1), X_O(t_1), X_P(t_1)$$

$$X_I[t_0, t_1], X_K[t_0, t_1], X_g[t_0, t_1]$$
After marginalizing out the middle layers:

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

- We have $X_D(t_0), X_O(t_0), X_P(t_0)$.
- We have $X_I[t_0, t_1], X_K[t_0, t_1], X_g[t_0, t_1]$.
- We have $X_D(t_1), X_O(t_1), X_P(t_1)$. 

We may repeat this for additional time intervals $[t_2, t_3], [t_3, t_4], \text{etc.}$.
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.
Example: RC Electric Circuit

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.

\begin{align*}
    f_1 & : \quad I_V(t) = I_R(t) \quad \text{Kirchoff’s current law} \\
    f_2 & : \quad I_R(t) = I_C(t) \quad \text{Kirchoff’s current law} \\
    f_3 & : \quad V_V(t) = V_R(t) + V_C(t) \quad \text{Kirchoff’s voltage law} \\
    f_4 & : \quad R I_R(t) = V_R(t) \quad \text{Ohm’s law} \\
    f_5 & : \quad C V_C'(t) = I_C(t) \quad \text{Ideal capacitor} \\
    h_1 & : \quad V_C(t) = V_C(t_0) + \int_{t_0}^{t} V_C'(\tau) \, d\tau
\end{align*}
Example: RC Electric Circuit

Causal ordering yields:

\[ V_{C}(t) = V_{C}(t_0) + \int_{t_0}^{t} \frac{1}{RC} (V_{V}(\tau) - V_{C}(\tau)) \, d\tau \]

\[ I_{V}(t) = R^{-1}(V_{V}(t) - V_{C}(t)) \]

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 \( V_L \),
- \( V_I \perp\!\!\!\!\perp L \mid V_L \),
- \( \ldots \)

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.
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...?


Some variations on variation independence.

Influence diagrams for causal modelling and inference.

Coverings of bipartite graphs.

Transitional conditional independence.

Markov properties for graphical models with cycles and latent variables.


