[1] University of Amsterdam

# A Markov Property for Sample Paths of Stochastic Processes

Philip Boeken[12] Joris M. Mooij[1]

[2] Booking.com

tl;dr: We map a system of SDEs to a *Dynamic SCM*, which naturally facilitates causal reasoning, causal effect identification and causal discovery in a wide class of dynamical systems.

## Equipping SDEs with causal semantics

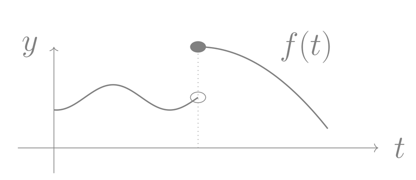
**Example (System of SDEs)** Consider the system of Stochastic Differential Equations:

$$\mathcal{D}: \begin{cases} X_1(t) = \xi_1 + \int_0^t g_1(s-, X_1, X_3)\mathrm{d}W_1(s) \\ X_2(t) = \xi_2 + \int_0^t g_2(s-, X_1, X_2)\mathrm{d}W_1(s) \\ X_3(t) = \xi_3 + \int_0^t g_3(s-, X_2, X_3)\mathrm{d}N_3(s) \\ X_4(t) = \xi_4 + \int_0^t g_4(s-, X_4)\mathrm{d}X_3(s). \end{cases}$$

**Theorem 1 (DSCM induced by SDE)** Under regularity assumptions, a system of SDEs $\mathcal{D}$ can be interpreted as a *Dynamic SCM* $\mathcal{M}_\mathcal{D}$ with structural equations

$$\mathcal{M}_\mathcal{D}: \begin{cases} X_1 = f_1(\xi_1, X_3, W_1) \\ X_2 = f_2(\xi_2, X_1, W_1) \\ X_3 = f_3(\xi_3, X_2, N_3) \\ X_4 = f_4(\xi_4, X_3) \end{cases}$$

with $\xi_i \in \mathbb{R}, X_i, W_j, N_2 \in D(\mathcal{T}, \mathbb{R})$ and exogenous distributions $\mathbb{P}(\xi_i), \mathbb{P}(W_j), \mathbb{P}(N_2)$.
This result builds upon a representation theorem by [8]. The space of sample paths $D(\mathcal{T}, \mathbb{R})$ is the (separable complete metric) space of càdlàg functions:




## Definition (Dynamic SCM)

Given a continuous time index $\mathcal{T} = [0, T]$ or discrete time index $\mathcal{T} = \{1, ..., T\}$, a *Dynamic Structural Causal Model* is an SCM (as in [1]):

$$\mathcal{M} = \langle V_0 \cup V_p, W_0 \cup W_p, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_\mathcal{E} \rangle$$

1. Endogenous parameters $V_0$, endogenous processes $V_p$
2. Exogenous parameters $W_0$, exogenous processes $W_p$
3. Standard Borel spaces $\mathcal{X} = \mathbb{R}^{|V_0|} \times D(\mathcal{T}, \mathbb{R})^{|V_p|}$ and $\mathcal{E} = \mathbb{R}^{|W_0|} \times D(\mathcal{T}, \mathbb{R})^{|W_p|}$
4. Structural equations $f_v : \mathcal{X} \times \mathcal{E} \to \mathcal{X}_v$ for all $v \in V$ (that are adapted for $v \in V_p$)
5. Exogenous distribution $\mathbb{P}_\mathcal{E} = \left(\bigotimes_{w \in W_0} \mathbb{P}(X_w)\right) \otimes \left(\bigotimes_{w \in W_p} \mathbb{P}(X_w)\right)$

**Definition (Intervention)**
For $T \subseteq V$ and $x_T \in \mathcal{X}_T$, the intervened DSCM is

$$\mathcal{M}_{\mathrm{do}(X_T = x_T)} : \begin{cases} X_v = f_v(X_V, X_W) & \text{if } v \in V \setminus T \\ X_v = x_v & \text{if } v \in T. \end{cases}$$

**Definition (Simple DSCM)** A DSCM is *simple* if its structural equations have a unique solution under all interventions, giving well-defined distributions $\mathbb{P}(X_V | \mathrm{do}(X_T))$.
Simple DSCMs can be cyclic. The class of simple DSCMs is closed under intervention and marginalisation [1]. The DSCM $\mathcal{M}_\mathcal{D}$ from Theorem 1 is simple.

## Implications (overview)

The following existing notions and results for SCMs naturally apply to DSCMs:

1. The graph of the DSCM, $\sigma$-separation, and a Global Markov property
2. Causal effect identification
3. Constraint-based causal discovery

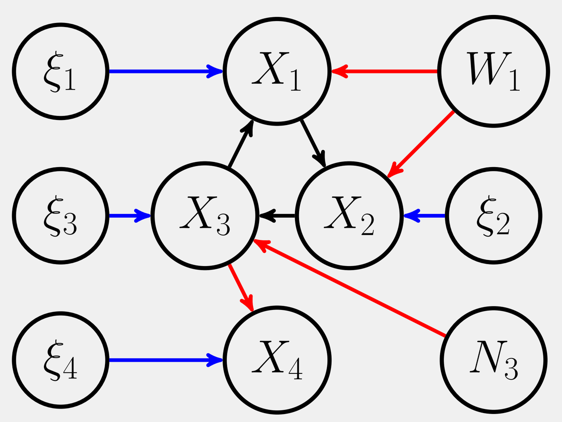
Additionally, we investigate:

4. Time-evaluations to map continuous time DSCMs to discrete time DSCMs

## (1) Markov property

**Definition (The graph $G(\mathcal{M})$)** See [1].

Example:




**Definition ($\sigma$-separation, [1, 2])**
The $\sigma$-separation criterion is a generalisation of $d$-separation to cyclic graphs. Works just like $d$-separation, with one extra condition: a path $v \to i \to j \to w$ is $\sigma$-blocked by *$i$ iff $i$ and $j$ are not in the same strongly connected component*.

Example: $\xi_1 \overset{d}{\underset{G(\mathcal{M})}{\perp}} X_4 | X_2$ $\quad \xi_1 \overset{\sigma}{\underset{G(\mathcal{M})}{\not\perp}} X_4 | X_2$ $\quad \xi_1 \overset{\sigma}{\underset{G(\mathcal{M})}{\perp}} X_4 | X_3$

**Theorem (Global Markov Property, [1, 2])**
For simple DSCM $\mathcal{M}$ with distribution $\mathbb{P}(X_V, X_W)$, graph $G(\mathcal{M})$ and (not necessarily disjoint) sets $A, B, C \subseteq V \cup W$, we have

$$A \perp^\sigma_{G(\mathcal{M})} B \,|\, C \implies X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \,|\, X_C.$$

## (2) Causal Effect Identification

The rules of do-calculus are valid for simple DSCMs [3], so we can reason about unconfoundedness:

$$\mathbb{P}(X_4 | \mathrm{do}(\xi_1)) = \mathbb{P}(X_4 | \xi_1)$$

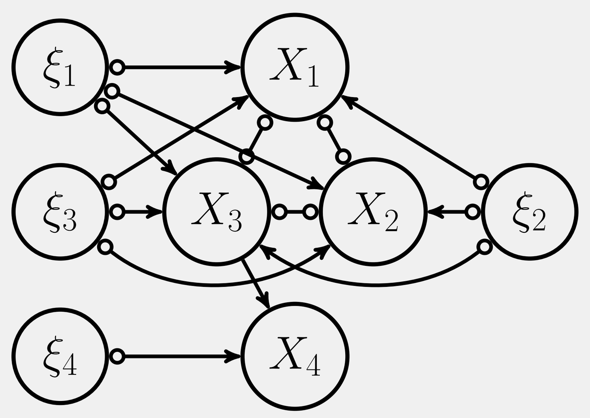
but we also have adjustment formulae, like backdoor adjustment:

$$\mathbb{P}(X_4 | \mathrm{do}(X_1 = x_1)) = \int_{D(\mathcal{T}, \mathbb{R})} \mathbb{P}(X_4 | X_1 = x_1, W_1 = w_1) \mathrm{d}\mathbb{P}(W_1 = w_1).$$

The ID algorithm [3] is valid as well.

## (3) Constraint-based Causal Discovery

**Example** FCI is sound and complete for simple DSCMs [7], and outputs the PAG:

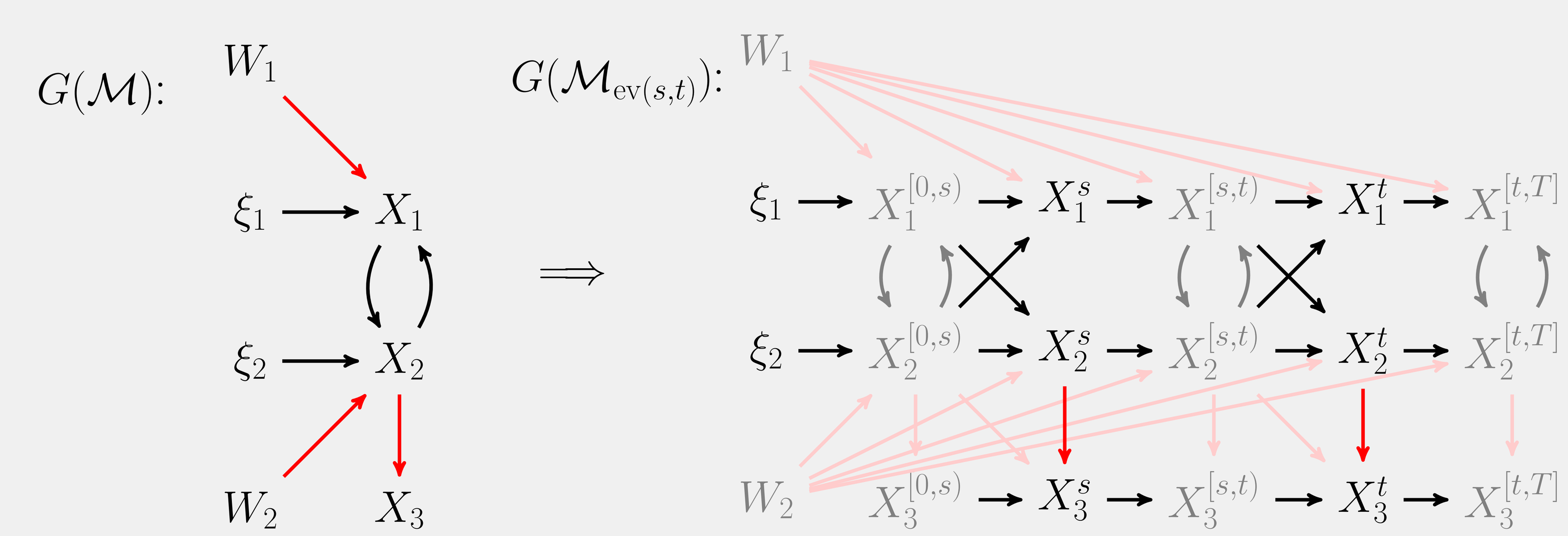



(This example is based on an independence oracle.)

Testing independence of *sample paths* (e.g. $X_1 \perp\!\!\!\perp_\mathbb{P} X_3 | X_2$) is an active area of research, with first results [4, 5].

## (4) Time-evaluations of DSCMs

**Definition (Time-evaluation)** Given time indices $s, t \in \mathcal{T}$ with $s < t$, the time-evaluated DSCM $\mathcal{M}_{\mathrm{ev}(s,t)}$ is a DSCM with endogenous variables for evaluations at $s$ and $t$.




(For graphical appeal we assume $X_1$ and $X_2$ to be temporally Markov.)

**Conjecture:** *Local independence* [6] (i.e. continuous-time Granger-noncausality) can be characterised in terms of time-evaluations of DSCMs:

$$X_B^t \perp\!\!\!\perp X_A^{[0,t)} | X_C^{[0,t)} \text{ for all } t \in \mathcal{T} \iff X_A \not\to X_B | X_C$$


[1] S. Bongers, P. Forré, J. Peters, and J.M. Mooij. Foundations of structural causal models with cycles and latent variables. *Ann. Stat.*, 49(5), 2021.
[2] P. Forré and J.M. Mooij. Markov Properties for Graphical Models with Cycles and Latent Variables, 2017.
[3] P. Forré and J.M. Mooij. Causal Calculus in the Presence of Cycles, Latent Confounders and Selection Bias. In *PMLR*, pages 71–80. PMLR, 2020.
[4] A.R. Lundborg, R.D. Shah, and J. Peters. Conditional Independence Testing in Hilbert Spaces with Applications to Functional Data Analysis. *J. R. Stat. Soc. Ser. B Methodol.*, 84(5):1821–1850, 2022.
[5] G. Manten, C. Casolo, E. Ferrucci, S.W. Mogensen, C. Salvi, and N. Kilbertus. Signature Kernel Conditional Independence Tests in Causal Discovery for Stochastic Processes, 2024.
[6] S.W. Mogensen and N. Richard Hansen. Markov equivalence of marginalized local independence graphs. *Ann. Stat.*, 48(1), 2020.
[7] J.M. Mooij and Tom Claassen. Constraint-Based Causal Discovery using Partial Ancestral Graphs in the presence of Cycles. In *UAI2020*, pages 1159–1168. PMLR, 2020.
[8] P. Przybyłowicz, V. Schwarz, A. Steinicke, and M. Szölgyenyi. A Skorohod measurable universal functional representation of solutions to semimartingale SDEs, 2023.


p.a.boeken@uva.nl