Dynamic Structural Causal Models

Philip Boeken[1,2] **Joris M. Mooij**[1]

[1]Korteweg-de Vries Institute for Mathematics, University of Amsterdam
[2]Booking.com, Amsterdam, The Netherlands

Abstract

We study a specific type of SCM, called a *Dynamic Structural Causal Model* (DSCM), whose endogenous variables represent functions of time, which is possibly cyclic and allows for latent confounding. As a motivating use-case, we show that certain systems of Stochastic Differential Equations (SDEs) can be appropriately represented with DSCMs. An immediate consequence of this construction is a graphical Markov property for systems of SDEs. We define a *time-splitting* operation, allowing us to analyse the concept of *local independence* (a notion of continuous-time Granger (non-)causality). We also define a *subsampling* operation, which returns a discrete-time DSCM, and which can be used for mathematical analysis of subsampled time-series. We give suggestions how DSCMs can be used for identification of the causal effect of time-dependent interventions, and how existing constraint-based causal discovery algorithms can be applied to time-series data.

1 INTRODUCTION

Many real-world systems exhibit a non-trivial dependence on time. In science and engineering, such systems are often modelled with Ordinary Differential Equations when the dynamics are deterministic, with Random Differential Equations (RDEs) when the dynamics of a population is described but every individual can have different (deterministic) dynamics, and with Stochastic Differential Equations (SDEs) when the dynamics of each individual are inherently stochastic (Banks et al., 2014). In this work, we will focus on the latter formalism, which generalises the former two. SDEs do not come formally equipped with causal semantics. As their mathematical formulation is rather technical, it is a nontrivial task to apply existing methods of causal reasoning to this model class. In this work, we investigate how SDEs can be mapped into a *Dynamic Structural Causal Models* (DSCMs), originally introduced by Rubenstein et al. (2018) and refined in this work, to naturally bridge this gap. More precisely, we

1. refine the definition of DSCMs to make them specific versions of SCMs as defined by Bongers et al. (2021), enhancing them with a Markov property based on σ-separation;
2. show how certain *uniquely solvable* systems of SDEs can be transformed into DSCMs;
3. define *time-splitting* and *subsampling* as formal tools to analyse the time-dependent structure of DSCMs;
4. show how time-split DSCMs allow for a graphical interpretation of local independence and that the graph of 'collapsed' DSCMs are local independence graphs, yielding a local independence Markov property;
5. give further suggestions of how existing results for 'static' causal effect identification and constraint-based causal discovery are naturally applicable to DSCMs.[1]

2 SYSTEMS OF STOCHASTIC DIFFERENTIAL EQUATIONS

Consider the time-index $\mathcal{T} := [0, T]$ for some $T \in \mathbb{R}$. Let $D(\mathcal{T}, \mathbb{R}^n)$ be the space of càdlàg[2] functions from $\mathcal{T}$ to $\mathbb{R}^n$. Throughout, we equip $D(\mathcal{T}, \mathbb{R}^n)$ with the J_1 topology (Skorokhod, 1956) and the Borel sigma-algebra, making it a standard Borel space. Let $(\Omega, \mathbb{F}, \mathbb{P})$ denote a filtered probability space with filtration $\mathbb{F} := (\mathcal{F}_t)_{t\in\mathcal{T}}$ that is right-continuous and contains all $\mathbb{P}$-null-sets. A process $X : \Omega \times \mathcal{T} \to \mathbb{R}^n$ is $\mathbb{F}$-adapted if $X(t) \in \mathcal{F}_t$ for all $t \in \mathcal{T}$; for predictable processes we have $X(t) \in \mathcal{F}_{t-}$, where $\mathcal{F}_{t-} := \bigcap_{s<t} \mathcal{F}_s$. Let $\mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R}^n)$ denote the class of $\mathbb{R}^n$-valued semimartingales, i.e. processes $X : \Omega \times \mathcal{T} \to \mathbb{R}^n$ that are $\mathbb{F}$-adapted with

[1]All proofs are provided in Appendix C.
[2]continue à droite, limité à gauche

càdlàg sample paths a.s., and with (unique) decomposition $X = A + M$ with A a predictable process of finite variation and M a local martingale. Let $\mathcal{P}_b$ denote the set of locally bounded predictably measurable processes. The *stochastic integral of G w.r.t. H*

$$J : \mathcal{P}_b \times \mathbb{S} \to \mathbb{S}, \quad (G, H) \mapsto \int_0^t G(s)\mathrm{d}H(s) \qquad (1)$$

is a well-defined notion (Protter (2005), Thm. IV.15).

For a formal definition of a system of stochastic differential equations, we require the following notion:

Assumption 1. Let $f : \mathcal{T} \times D(\mathcal{T}, \mathbb{R}^m) \to \mathbb{R}^n$ be a Borel-measurable function such that

i) $t \mapsto f(t, x) \in D(\mathcal{T}, \mathbb{R}^n)$ for all $x \in D(\mathcal{T}, \mathbb{R}^m)$,

ii) $f(t, x) = f(t, x^t)$ for all $x \in D(\mathcal{T}, \mathbb{R}^m)$ and $t \in \mathcal{T}$, where $x^t(s) := x(s \wedge t)$.

Remark 1. If f satisfies Assumption 1 and $X \in \mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R}^m)$, then the process $f(t, X)$ is in $\mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R}^n)$, and $f(t-, X) := \lim_{s\uparrow t} f(t, X)$ is an element of $\mathcal{P}_b$ (Przybyłowicz et al., 2023).

Definition 1 (System of SDEs, solutions)**.** *Given a filtered probability space* $(\Omega, \mathbb{F}, \mathbb{P})$*, a* system of stochastic differential equations (SDEs) *is a tuple* $\mathcal{D} = \langle V, W, (X_V^0, X_W), f\rangle$ *with*

i) V *and* W *finite disjoint index sets,*

ii) for each $v \in V$ *a random variable* $X_v^0 : \Omega \to \mathbb{R}$*, and for each* $w \in W$ *a stochastic process* $X_w \in \mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R})$*,*

iii) for each $v \in V$ *a stochastic differential equation*

$$f_v : X_v(t) = X_v^0 + \int_0^t g_v(s-, X_{\alpha(v)})\mathrm{d}h_v(s, X_{\beta(v)})$$

for all $t \in \mathcal{T}$*, for sets* $\alpha(v), \beta(v) \subseteq V \cup W$*, and functions* $g_v : \mathcal{T} \times D(\mathcal{T}, \mathbb{R}^{|\alpha(v)|}) \to \mathbb{R}^{m_v}$ *and* $h_v : \mathcal{T} \times D(\mathcal{T}, \mathbb{R}^{|\beta(v)|}) \to \mathbb{R}^{m_v}$ *for some* $m_v \in \mathbb{N}$*, both satisfying Assumption 1.*[3]

The processes $g_v(t, X_{\alpha(v)})$ *and* $h_v(t, X_{\beta(v)})$ *are called* integrands *and* integrators *respectively. A* solution[4] *of a system of SDEs* $\mathcal{D}$ *is a stochastic process* $X_V \in \mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R}^{|V|})$ *such that* $X_V = f_V(X_V^0, X_V, X_W)$ *holds* $\mathbb{P}$*-a.s.*

By Remark 1, we have for semimartingales $X_{\alpha(v)}$ and $X_{\beta(v)}$ that $g_v(s-, X_{\alpha(v)})$ is in $\mathcal{P}_b$ and $h_v(s, X_{\beta(v)})$ is in $\mathbb{S}$, making the stochastic integrals well-defined (see equation 1). Note that we typically have $v \in \alpha(v)$. Definition 1 generalises common definitions of SDEs, which typically only model the relations between the variables that appear in the integrands, and assume the integrators to be given. We allow for explicit modelling of the variables that appear in the integrators because it is mathematically feasible, has applications in controlled SDEs (Lyons et al., 2007), and allows for the modelling of instantaneous effects in time-split DSCMs (see Section 5).

Remark 2. For specific choices of integrators and integrands, SDEs can model some special cases:

1. If $g(t, X_\alpha) = g(t, X_\alpha(t))$ and h is a Lévy process independent of X^0, then any solution X is temporally Markov[5] (Protter, 2005, Theorem V.6.32). If additionally $g(t, X_\alpha) = g(X_\alpha(t))$, then the solution is strong Markov; such dynamics are called *time-invariant*.
2. If $h(s) = s$, the 'stochastic integral' $\int g(s, X_\alpha)\mathrm{d}h(s)$ reduces to the Riemann integral $\int g(s, X_\alpha)\mathrm{d}s$, so a system of Random Differential Equations is a special case of a system of SDEs.
3. If $h(t, W) = W(t)$ is a Brownian Motion, any solution X has continuous and possibly non-differentiable sample paths. If $h(t, W) = (t, W(t))$ and the dynamics are time-invariant, such an SDE is called an *Itô diffusion*.
4. If $h(t, N) = N(t)$ is a jump process (e.g. a Poisson process), the solution X can have jumps as well. If $h(t, W, N) = (t, W(t), N(t))$ and the dynamics are time-invariant, such an SDE is called a *jump-diffusion*.
5. If H is a process with differentiable sample paths, an RDE $X = \int g(H'(s), X(s))\mathrm{d}s$ with g linear in its first argument can be suitably modelled with an SDE $X = \int f(X)\mathrm{d}H(s)$ (Lyons et al., 2007). The SDE formulation generalises this concept to a large class of non-differentiable processes H.
6. The functional relationship $X(t) = h(t, X_\beta)$ can be equivalently expressed as $X(t) = h(0, X_\beta) + \int_0^t \mathrm{d}h(s, X_\beta)$.

Example 1. Consider the following system of Stochastic Differential Equations

$$\mathcal{D} : \begin{cases} X_1(t) = X_1^0 + \int_0^t g_1(s-, X_1, X_3)\mathrm{d}W(s) \\ X_2(t) = X_2^0 + \int_0^t g_2(s-, X_1, X_2)\mathrm{d}N(s) \\ X_3(t) = X_3^0 + \int_0^t g_3(s-, X_2, X_3)\mathrm{d}W(s) \\ X_4(t) = X_4^0 + \int_0^t g_4(s-, X_4)\mathrm{d}X_2(s) \end{cases}$$

with W a Brownian motion, N a Poisson process (independent from W), and $X_1^0, X_2^0, X_3^0, X_4^0$ independent random variables. Note that X_2 is the integrator of X_4.

To ensure existence and uniqueness of solutions of a system of SDEs, we require additional regularity of the function g:

[3]We write $\int_0^t g\mathrm{d}h$ for $\sum_{i=1}^m \int_0^t g_i \mathrm{d}h_i$.

[4]Commonly referred to as a *strong solution*.

[5]Process $X \in \mathbb{S}$ is called temporally *Markov* if $\mathbb{P}(X_{t+s}|\mathcal{F}_t) = \mathbb{P}(X_{t+s}|X_t)$ for all $s, t \in \mathcal{T}$ such that $s+t \in \mathcal{T}$; it is (temporally) *strong Markov* if this holds for any stopping time t.

Assumption 2. Let $g : \mathcal{T} \times D(\mathcal{T}, \mathbb{R}^m) \to \mathbb{R}^n$ satisfy Assumption 1, and let there exist a $K \in D(\mathcal{T}, (0, \infty))$ such that for all $x, x_1, x_2 \in D(\mathcal{T}, \mathbb{R}^m)$ we have

$$|g(t, x)| \leq K(t) \left(1 + \sup_{0 \leq s \leq t} |x(s)|\right)$$
$$|g(t, x_2) - g(t, x_1)| \leq K(t) \sup_{0 \leq s \leq t} |x_2(s) - x_1(s)|.$$

Theorem 1 (Przybyłowicz et al. (2023))**.** *For given functions $g : \mathcal{T} \times D(\mathcal{T}, \mathbb{R}^{k+1}) \to \mathbb{R}^m$ and $h : \mathcal{T} \times D(\mathcal{T}, \mathbb{R}^\ell) \to \mathbb{R}^m$ that satisfy Assumptions 2 and 1 respectively, there exists a Skorohod measurable mapping*

$$I : \mathbb{R} \times D(\mathcal{T}, \mathbb{R}^k) \times D(\mathcal{T}, \mathbb{R}^\ell) \to D(\mathcal{T}, \mathbb{R})$$

such that for any filtered probability space $(\Omega, \mathbb{F}, \mathbb{P})$ with square-integrable random variable $X^0 : \Omega \to \mathbb{R}$ and semimartingales $A \in \mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R}^k)$, $B \in \mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R}^\ell)$, the process $X := I(X^0, A, B) \in \mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R})$ is a solution of the SDE

$$X(t) = X^0 + \int_0^t g(s-, X, A) \mathrm{d}h(s, B).$$

We refer to such a solution function I as an *Itô map*. The above result straightforwardly extends to processes X taking values in $\mathbb{R}^n$.

2.1 TOWARDS CAUSAL REASONING WITH SDES

To analyse whether a system of SDEs can be solved, we use the dependency structure between the processes, which can be suitably depicted with a graph:

Definition 2 (Graph of system of SDEs)**.** *For a system of SDEs $\mathcal{D} = \langle V, W, (X_V^0, X_W), f \rangle$, let $G(\mathcal{D}) = (V \cup W, E)$ be the directed graph where $E := \{v \to w : w \in V, v \in \alpha(w) \cup \beta(w) \setminus \{w\}\}$.*

Example 2. Let $\mathcal{D}$ be the systems of SDEs from Example 1. The graph $G(\mathcal{D})$ is given in Figure 1. Outgoing edges of integrators are displayed in red.

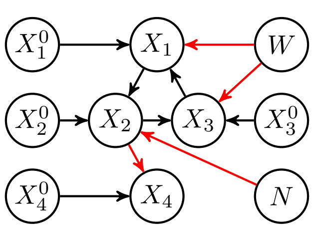

Figure 1: The directed graph $G(\mathcal{D})$.

By partitioning the set of endogenous processes V into *strongly connected components* (SCCs) (subsets whose every two vertices are connected by a directed path) and assuming that no element of such a SCC is the integrator of another element in the SCC (so no directed cycle contains a red edge), we can iteratively apply Theorem 1 along the topological order of the SCCs to obtain a unique solution.

Proposition 1 (Unique solvability of a system of SDEs)**.** *Let $\mathcal{D} = \langle V, W, (X_V^0, X_W), f \rangle$ be a system of SDEs such that for every SDE f_v, the parameters g_v and h_v satisfy Assumptions 2 and 1 respectively. If for every $v \in V$ we have $\beta(v) \cap \mathrm{Sc}(v) = \emptyset$,[6] there exists an Itô map I_V such that $X_V := I_V(X_V^0, X_W) \in \mathbb{S}_{\mathbb{F}}(\mathcal{T}, \mathbb{R}^{|V|})$ is the unique solution of $\mathcal{D}$.*

We refer to a system of SDEs $\mathcal{D}$ that satisfies Proposition 1 as *uniquely solvable*.

Remark 3. A system of SDEs is assumed to have a certain causal interpretation: for the SDEs in Definition 1, it is assumed that the LHS is *determined* by variables on the RHS. In Section 4 we will indeed interpret the Itô map I_v for and SDE f_v as a *structural equation* of a DSCM. This construction then gives a clear causal interpretation to the graph $G(\mathcal{D})$, where the processes V and W are interpreted as endogenous and exogenous respectively. Compared to existing work where for a *systems of constraints* one deduces the causal structure (e.g. Iwasaki and Simon, 1994; Blom et al., 2021), our setting makes relatively strong assumptions.

Another important causal aspect that is engrained in the system of SDEs is whether processes exhibit instantaneous dependencies. Although not explicitly stated in Przybyłowicz et al. (2023), one can verify via their proof of Theorem 1 that the Itô map I satisfies $I(x^0, a, b)^t = I(x^0, a^{t-}, b^t)^t$. As it will play an important role in the analysis of temporal dynamics of DSCMs, we formally define this concept:

Definition 3 (Adapted, predictable dependence)**.** *Let $g : D(\mathcal{T}, \mathbb{R}^m) \to D(\mathcal{T}, \mathbb{R}^n)$ be a Borel-measurable function. We say that the dependence of g on x is* adapted *if $g(x)^t = g(x^t)^t$ and* predictable *if $g(x)^t = g(x^{t-})^t$ for all $t \in \mathcal{T}$.*

Remark 4. For given process X and $Y := g(X)$ for measurable g, we have $Y \in \mathcal{F}_t^X$ if g is adapted, and $Y \in \mathcal{F}_{t-}^X$ if g is predictable,[7] motivating the nomenclature *adapted* and *predictable* (see Appendix A, Lemma 1). For any uniquely solvable system of SDEs $\mathcal{D}$ and any $v \in V$, the process X_v has predictable dependence on $X_{\alpha(v)}$ and adapted (and thus possibly instantaneous) dependence on $X_{\beta(v)}$. Graphically, we depict adapted dependencies with red edges.

3 DYNAMIC STRUCTURAL CAUSAL MODELS

Following Bongers et al. (2021) and Forré and Mooij (2023), a *Structural Causal Model* (SCM) is a tuple $\mathcal{M} = \langle V, W, \mathcal{X}_V, \mathcal{X}_W, f, \mathbb{P}(X_W) \rangle$ where V, W are disjoint finite index sets of *endogenous variables* and *exogenous variables* respectively, the *domains* $\mathcal{X}_V = \prod_{i \in V} \mathcal{X}_i$ and $\mathcal{X}_W = \prod_{i \in W} \mathcal{X}_i$ are products of *standard Borel spaces* $\mathcal{X}_i$, the

[6] $\mathrm{Sc}(v)$ denotes the strongly connected component of v.
[7] Here, $\mathcal{F}_t^X$ denotes the filtration generated by X.

exogenous distribution $\mathbb{P}(X_W) = \bigotimes_{w \in W} \mathbb{P}(X_w)$ is a product of probability distributions, and the *causal mechanism* $f : \mathcal{X} \to \mathcal{X}_V$ is a measurable function. Typically, SCMs are used to model 'static' systems, where variables take values in $\mathbb{R}$, for example. Discrete-time dynamical systems are often modelled with a copy of all variables X_V for every time-step; an approach that does not directly apply to continuous-time dynamical systems. Inspired by (Rubenstein et al., 2018) and Bongers et al. (2022), we define a specific type of SCM where variables are functions on some time-interval, and the structural equations respect temporal causality:

Definition 4 (DSCM). *Given a time-interval $\mathcal{T} = [0, T]$ for some $T \in \mathbb{R}_+$ or $\mathcal{T} \subseteq \{1, ..., T\}$ for some $T \in \mathbb{N}$, a* Dynamic Structural Causal Model *is an SCM $\mathcal{M} = \langle V, W, \mathcal{X}_V, \mathcal{X}_W, f, \mathbb{P} \rangle$ with*

i) $V := V_p \times \mathcal{E}$ with endogenous processes V_p and $\mathcal{E}$ a set of disjoint evaluation intervals *or* points *in $\mathcal{T}$;*[8]

ii) exogenous processes W;

iii) standard Borel spaces $\mathcal{X}_V = \bigtimes_{(v,\mathcal{I}) \in V \times \mathcal{E}} D(\mathcal{I}, \mathbb{R})$ and $\mathcal{X}_W = \bigtimes_{w \in W} D(\mathcal{T}, \mathbb{R})$;

iv) measurable structural equations $f : \mathcal{X}_V \times \mathcal{X}_W \to \mathcal{X}_V$ that are adapted *(see Definition 3);*

v) exogenous distribution $\mathbb{P}(X_W) = \bigotimes_{w \in W} \mathbb{P}(X_w)$.

The set $\mathcal{E}$ denotes the time intervals (or points) on which we evaluate the processes X_{V_p}. For example, when $\mathcal{E} = \{\{0\}, (0, T]\}$, for every $v \in V_p$ we will consider the variables X_v^0 and $X_v^{(0,T]}$. We refer to a DSCM as *collapsed* if $\mathcal{E} = \mathcal{T}$. For $v \in V_p$, we denote with $\alpha(v)$ and $\beta(v)$ the sets of parents of v on which it has predictable and adapted dependence respectively.

For notational simplicity, endogenous parameters like initial values or parameters that 'materialise' at some $t_0 \in \mathcal{T}$ (i.e. other variables are only allowed to depend on it from t_0 onwards) are modelled as elements of $D(\{t_0\}, \mathbb{R})$. For a function $x : (s, t) \to \mathbb{R}$ we use the convention that $x(r) = x(s+)$ for all $r < s$ and $x(u) = x(t-)$ for all $u > t$.[9]

We have refined the definition of DSCMs as given by Rubenstein et al. (2018) by 1) ensuring the structural equations respect temporal causality, 2) allowing for stochasticity through the exogenous variables, and 3) picking a standard Borel space ($D(\mathcal{T}, \mathbb{R})$) as space of trajectories, making for all $A, B \subseteq V \cup W$ the conditional distribution $\mathbb{P}(X_A \mid X_B)$ well-defined, and with that giving a well-defined notion of conditional independence. DSCMs differ from *Structural Dynamical Causal Models* (SDCMs) (Bongers et al., 2022) as we model entire sample paths as a whole, instead of separately at every time $t \in \mathcal{T}$. When $\mathcal{T}$ or $\bigcup \mathcal{E}$ is a discrete set, the DSCM reduces to a structural vector autoregression model (Peters et al., 2013; Malinsky and Spirtes, 2018) with possible serial dependence of the noise variables, or a Dynamic Bayesian Network (Dean and Kanazawa, 1989).

Perfect interventions on DSCMs are defined exactly the same as for SCMs:

Definition 5 (Perfect interventions on DSCMs). *Given a DSCM $\mathcal{M}$, intervention target $T \subseteq V$ and intervention value $x_T \in \mathcal{X}_T$, the perfectly intervened DSCM is defined as $\mathcal{M}_{\mathrm{do}(X_T = x_T)} := \langle V, W, \mathcal{X}_V, \mathcal{X}_W, f^\circ, \mathbb{P}(X_W) \rangle$ where the components of the intervened causal mechanism $f^\circ : \mathcal{X}_V \times \mathcal{X}_W \to \mathcal{X}_V$ are given by*

$$f_v^\circ(x_V, x_W) := \begin{cases} f_v(x_V, x_W) & \text{if } v \in V \setminus T \\ x_v & \text{if } v \in T. \end{cases}$$

For a given SCM $\mathcal{M}$, a random variable $X_V \in \mathcal{X}_V$ is a *solution* of the SCM if $X_V = f(X_V, X_W)$ holds $\mathbb{P}(X_W)$-a.s. Existence and uniqueness of solutions are generally not guaranteed for (intervened) cyclic SCMs. To this end, Bongers et al. (2022) introduced the following class of SCMs:

Definition 6 (Simple (D)SCM). *A (D)SCM is simple if for all $O \subseteq V$ (write $T := V \setminus O$) there exists a unique (up to modifications*[10]*) measurable function $g_O : \mathcal{X}_T \times \mathcal{X}_W \to \mathcal{X}_O$ such that for all $x_T \in \mathcal{X}_T$ and $\mathbb{P}(X_W)$-almost all $x_W \in \mathcal{X}_W$ we have $g_O(x_T, x_W) = f_O(x_T, g_O(x_T, x_W), x_W)$.*

For a simple DSCM $\mathcal{M}$, the random variable $g_O(x_T, X_W)$ is a solution for $\mathcal{M}_{\mathrm{do}(X_T = x_T)}$, implying the existence of a well-defined *interventional distribution* $\mathbb{P}(X_V \mid \mathrm{do}(X_T = x_T))$. The function g_O is referred to as a *solution function* for $\mathcal{M}_{\mathrm{do}(X_T = x_T)}$. The following proposition shows that solutions of DSCMs obey temporal causality:

Proposition 2. *Given a simple DSCM $\mathcal{M}$, the solution function $g_O(x_T, x_W)$ of the intervened DSCM $\mathcal{M}_{\mathrm{do}(X_T = x_T)}$ is adapted.*

Remark 5. Following Remark 4, adaptedness of the solution function $g_O(x_T, x_W)$ is equivalent to adaptedness of solution process $X_V := g_O(x_T, X_W)$ to a canonical filtration on the underlying probability space $\mathcal{X}_T \times \mathcal{X}_W$. For more information, see Appendix A.

In line with marginalisation of probability distributions, the marginalisation of a simple DSCM is defined as follows:

Definition 7 (Marginalisation). *For simple DSCM $\mathcal{M}$ and subset $L \subseteq V$ (write $O := V \setminus L$), the marginalised DSCM $\mathcal{M}_{\mathrm{marg}(L)} := \left\langle O, W, \mathcal{X}_O, \mathcal{X}_W, \tilde{f}, \mathbb{P}(X_W) \right\rangle$ has marginal causal mechanism*

$$\tilde{f}(x_O, x_W) := f_O(x_O, g_L(x_O, x_W), x_W),$$

[8] For $(v, \mathcal{I}) \in V \times \mathcal{E}$ we write $v^{\mathcal{I}} := (v, \mathcal{I})$ and $X_v^{\mathcal{I}} := X_{(v,\mathcal{I})}$ and $X_v := X_v^{\mathcal{T}}$.

[9] Here, $x(t+)$ denotes $\lim_{s \downarrow t} x(s)$.

[10] For every x_T, $g_O(x_T, x_W)$ is $\mathbb{P}(X_W)$ a.e. unique.

where g_L is the solution function in $\mathcal{M}_{\mathrm{do}(X_O=x_O)}$.

Proposition 3. *The class of simple DSCMs is closed under marginalisation.*

Bongers et al. (2021) show that marginalised simple SCMs are observationally, interventionally and counterfactually equivalent to the original SCM with respect to the observed variables. This makes marginalisation a powerful notion of abstraction, as any causal inference on the observed part of a system remains valid in the underlying system.

3.1 A MARKOV PROPERTY FOR SIMPLE DSCMS

For SCM $\mathcal{M}$ and nodes $i \in V \cup W$ and $j \in V$ we call i a *parent* of j if there does not exist a measurable function $\tilde{f}_j : \mathcal{X}_{V\cup W\setminus\{i\}} \to \mathcal{X}_j$ such that $f_j(x_V, x_W) = \tilde{f}_j(x_{V\cup W\setminus\{i\}})$ for $\mathbb{P}(X_W)$-almost all $x_W \in \mathcal{X}_W$ and for all $x_V \in \mathcal{X}_V$. We call a *directed graph* $G^+(\mathcal{M}) := (V\cup W, E)$ with directed edges $E = \{i \to j : i \in V \cup W, j \in V, i \text{ is a parent of } j \text{ in } \mathcal{M}\}$ the *augmented graph* of $\mathcal{M}$. Often we don't display the exogenous nodes in the graph, in which case we consider the *directed mixed graph* $G(\mathcal{M}) = (V, E')$ where $E' = E \cup \{i \leftrightarrow j : i, j \in V, (i \leftarrow k \to j) \in G(\mathcal{M}) \text{ for some } k \in W\}$, where E stems from $G^+(\mathcal{M})$. We call an SCM *cyclic* if its graph $G(\mathcal{M})$ has a directed cycle. Simple SCMs can be cyclic, but cannot have self-cycles. A suitable extension of d-separation (Pearl, 2009) from DAGs to DMGs is the following notion of σ-separation:

Definition 8 (σ-separation, Forré and Mooij (2017); Bongers et al. (2021))**.** *Given a DMG $G = (V, E)$, a set of nodes $C \subset V$ and a walk π in DMG G:*

1. *a non-collider is called* blockable *if it has a child on the walk that is not in the same strongly connected component,*
2. *the walk π is called* σ-blocked by C *if there is a collider on π that is not in $\mathrm{Anc}(C)$, or if there is a blockable non-collider on π in C.*

For sets of nodes $A, B, C \subseteq V$, we call A and B σ-separated *given C, written $A \perp^\sigma_G B \mid C$, if all paths between A and B are σ-blocked by C.*

In general σ-separation implies d-separation, and for acyclic DMGs the two notions coincide.

Example 3. In the DMG G from Figure 1 we have $X^0_1 \perp^d_G X^0_2 \mid X_1, X_2$, and $X^0_1 \not\perp^\sigma_G X^0_2 \mid X_1, X_2$.

Theorem 2 (Markov Property, Forré and Mooij (2017); Bongers et al. (2021))**.** *For simple (D)SCM $\mathcal{M}$ 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*

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

A simple (D)SCM $\mathcal{M}$ is called *σ-faithful* if the reverse implication holds; this need not hold in general.

4 CONSTRUCTING A DSCM FROM A SYSTEM OF SDES

As mentioned before, one of our goals is to map a system of SDEs $\mathcal{D}$ to a (D)SCM, so we can apply all tools that are available for SCMs to SDEs. However, an SDE cannot directly be interpreted as structural equation of a DSCM, as SDEs are not equations between variables taking values in $D(\mathcal{T}, \mathbb{R})$, but in $\mathbb{S}(\mathcal{T}, \mathbb{R})$. However, when $\mathcal{D}$ is uniquely solvable, we *can* unambiguously interpret the Itô maps of $\mathcal{D}$ as the structural equations of a DSCM $\mathcal{M}_\mathcal{D}$.

Definition 9 (DSCM induced by a system of SDEs)**.** *Given a uniquely solvable system of SDEs $\mathcal{D} = \left\langle V_p, W_p, (X^0_{V_p}, X_{W_p}), f \right\rangle$ on a filtered probability space $(\Omega, \mathbb{F}, \mathbb{P})$ and time-interval $[0, T]$, we define* the simple DSCM induced by $\mathcal{D}$ *as $\mathcal{M}_\mathcal{D} := \langle V, W, \mathcal{X}_V, \mathcal{X}_W, f^*, \mathbb{P}(X_W)\rangle$ with*

1. *$V := V_p \times \mathcal{E}$ with $\mathcal{E} := \mathcal{T}$;*
2. *$W := W_0 \cup W_p$ with $W_0 := \{v^0 : v \in V_p\}$ a copy of V_p as the indices of the initial values;*
3. *for every $v \in V_p$ the structural equation*

   $$f^*_v(x_V, x_W) := I_v(x^0_v, x_{\alpha(v)}, x_{\beta(v)})$$

   where I_v is the Itô map for the SDE f_v;
4. *$\mathcal{X}_V := D(\mathcal{T}, \mathbb{R}^{|V|})$ and $\mathcal{X}_W := D(\mathcal{T}, \mathbb{R}^{|W|})$;*
5. *$\mathbb{P}(X_W)$ is the law of $(X^0_{V_p}, X_{W_p})$.*

Note that $G^+(\mathcal{M}_\mathcal{D}) = G(\mathcal{D})$, and does not have any self-loops $v \to v$. Interventions on a SDE $\mathcal{D}$ can be defined similarly as for DSCMs, intervening and mapping to a DSCM commute, and $\mathcal{M}_\mathcal{D}$ is indeed simple (see Appendix B).

Example 4. Let $\mathcal{D}$ be the systems of SDEs from Example 1. The structural equations of the induced DSCM $\mathcal{M}_\mathcal{D}$ are given by:

$$\mathcal{M}_\mathcal{D} : \begin{cases} X_1 = f_1(X^0_1, X_3, W) \\ X_2 = f_2(X^0_2, X_1, N) \\ X_3 = f_3(X^0_3, X_2, W) \\ X_4 = f_4(X^0_4, X_2). \end{cases}$$

The augmented graph $G^+(\mathcal{M}_\mathcal{D}) = G(\mathcal{D})$ is given in Figure 1, and $G(\mathcal{M}_\mathcal{D})$ is given in Figure 2, where the bidirected edge represents the latent confounder W.

Remark 6. A technical but important detail is that the Itô maps f^*_v (and thus the structural equations of $\mathcal{M}_\mathcal{D}$) are well-defined *everywhere* on their domain, due to the result of Przybyłowicz et al. (2023). If one obtains such an Itô

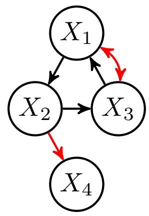

Figure 2: The directed mixed graph $G(\mathcal{M}_{\mathcal{D}})$.

map through the factorisation lemma (Kallenberg (2021), Lemma 1.14) it is not everywhere uniquely defined, rendering the effects of interventions outside this essential support ambiguous. It is also important to note that the Itô maps have the right measurability properties (with respect to the Borel sets generated by the J_1 topology).

Remark 7. The Markov property for DSCMs (Theorem 2) is closely related to the Markov property for the SDCM framework for RDEs (Bongers et al., 2022). We conjecture that for a given system $\mathcal{D}$ of uniquely solvable RDEs, the graph $G(\mathcal{M}_{\mathcal{D}})$ of DSCM $\mathcal{M}_{\mathcal{D}}$ is a subgraph of the graph $G(\mathcal{R}_{\mathcal{D}})$ of the SDCM $\mathcal{R}_{\mathcal{D}}$, which would make the DSCM Markov property at least as powerful as the SDCM Markov property.

5 TIME-EVALUATIONS OF DSCMS

Given a system of SDEs $\mathcal{D}$, we can construct a system of SDEs $\mathcal{D}_{\mathrm{ev}(r)}$ by splitting up the time interval $[0,T] = [0,r) \cup \{r\} \cup (r,T]$, and consider the SDEs

$$
\begin{aligned}
X_v(t) &= X_v^0 + \int_0^t g_v(s-, X_{\alpha(v)}) \mathrm{d}h_v(s, X_{\beta(v)}) \\
X_v(r) &= X_v^0 + \int_0^r g_v(s-, X_{\alpha(v)}) \mathrm{d}h_v(s, X_{\beta(v)}) \\
X_v(u) &= X_v(r) + \int_r^u g_v(s-, X_{\alpha(v)}) \mathrm{d}h_v(s, X_{\beta(v)})
\end{aligned}
$$

for all $v \in V$ and all $t \in [0,r)$ and $u \in (r,T]$. We can do this type of time-splitting also on the level of a DSCM, to be formally defined below. The graphs of an example DSCM $\mathcal{M}$ and its time-split version $\mathcal{M}_{\mathrm{ev}^+(0,s,t)}$ and subsampled version $\mathcal{M}_{\mathrm{ev}(0,s,t)}$ are given in Figure 3.

Definition 10 (time-split DSCM)**.** *Let a set of distinct time-indices $\tau = \{t_1, ..., t_m\} \subseteq \mathcal{T}$ and DSCM $\mathcal{M} = \langle V_p \times \mathcal{E}, W, \mathcal{X}_V, \mathcal{X}_W, f, \mathbb{P}(X_W)\rangle$ be given. The* time-split DSCM *$\mathcal{M}_{\mathrm{ev}^+(\tau)}$ is defined as $\mathcal{M}_{\mathrm{ev}^+(\tau)} = \left\langle \tilde{V}, W, \mathcal{X}_{\tilde{V}}, \mathcal{X}_W, \tilde{f}, \mathbb{P}(X_W)\right\rangle$ with as differences from $\mathcal{M}$:*

i) $\tilde{V} = V_p \times \tilde{\mathcal{E}}$, where $\tilde{\mathcal{E}}$ is a further subdivision of $\mathcal{E}$, where for every $t \in \tau$, an interval $[s,u] \in \mathcal{E}$ containing t is partitioned into $\{[s,t), \{t\}, (t,u]\} \subseteq \tilde{\mathcal{E}}$;[11]

ii) standard Borel space $\mathcal{X}_{\tilde{V}} = \bigtimes_{(v,\mathcal{I}) \in \tilde{V}} D(\mathcal{I}, \mathbb{R})$;

iii) for every $(v,\mathcal{I}) \in \tilde{V}$ a structural equation $\tilde{f}_{v^{\mathcal{I}}} : \mathcal{X}_{\tilde{V}} \times \mathcal{X}_W \to \mathcal{X}_{v^{\mathcal{I}}}$ with for all $t \in \mathcal{I}$

$$
\tilde{f}_{v^{\mathcal{I}}}(x_{\tilde{V}}, x_W)(t) := f_v\left(x_{\alpha(v)}^{\tilde{\mathcal{E}}_{<\mathcal{I}}}, x_{\beta(v)}^{\tilde{\mathcal{E}}_{\leq \mathcal{I}}}, x_W\right)(t),
$$

where $\tilde{\mathcal{E}}_{\leq \mathcal{I}} := \{\mathcal{I}' \in \tilde{\mathcal{E}} \mid \exists s \in \mathcal{I}', t \in \mathcal{I} : s \leq t\}$.

[11]A similar construction applies for (half-)open intervals.

The above definition can straightforwardly be extended to differing time-evaluations τ_v for each $v \in V$. Note the explicit modelling of the predictable dependence on α, and adapted dependence on β.

Definition 11 (Subsampled DSCM)**.** *Given a DSCM $\mathcal{M}$, consider the time-split DSCM $\mathcal{M}_{\mathrm{ev}^+(\tau)}$ with endogenous variables $V \times \mathcal{E}$. Let $\mathcal{E}' \subseteq \mathcal{E}$ denote the set of non-singleton intervals. The* subsampled DSCM *is defined as $\mathcal{M}_{\mathrm{ev}(\tau)} := (\mathcal{M}_{\mathrm{ev}^+(\tau)})_{\mathrm{marg}(V \times \mathcal{E}')}$.*

Example 5. Let $\mathcal{M}_{\mathcal{D}}$ be the DSCM from Example 4, and let $\mathcal{M} := (\mathcal{M}_{\mathcal{D}})_{\mathrm{marg}(L)}$ be the marginalisation of $\mathcal{M}_{\mathcal{D}}$ with respect to $L := \{X_3^0, X_3, X_4^0\}$. The graphs $G(\mathcal{M})$, time-split $G(\mathcal{M}_{\mathrm{ev}^+(s,t)})$ and subsampled $G(\mathcal{M}_{\mathrm{ev}(s,t)})$ are depicted in Figure 3, where we assume all processes to be temporally Markov. We assume the processes X_1, X_2 and X_4 to be temporally Markov for graphical appeal.

Definition 12 (Path concatenation)**.** *Let $\mathcal{I}_X, \mathcal{I}_Y$ be two adjacent intervals in $[0,T]$ with $s := \sup \mathcal{I}_X = \inf \mathcal{I}_Y$, and let $X \in D(\mathcal{I}_X, \mathbb{R})$ and $Y \in D(\mathcal{I}_Y, \mathbb{R})$. The concatenation of X and Y is the path $X * Y \in D(\mathcal{I}_X \cup \mathcal{I}_Y, \mathbb{R})$ defined by*

$$
(X * Y)(t) := \begin{cases} X(t) & \text{if } t \in [0,s) \cap \mathcal{I}_X \\ Y(t+) & \text{if } t \in [s,T] \cap \mathcal{I}_Y. \end{cases}
$$

Definition 13 (Collapsed DSCM)**.** *For a DSCM $\mathcal{M} = \langle V \times \mathcal{E}, W, \mathcal{X}_V, \mathcal{X}_W, f, \mathbb{P}(X_W)\rangle$, the* collapsed DSCM *is defined as $\mathcal{M}_* := \left\langle \tilde{V}, W, \mathcal{X}_{\tilde{V}}, \mathcal{X}_W, \tilde{f}, \mathbb{P}(X_W)\right\rangle$, where $\tilde{V} := V \times \tilde{\mathcal{E}}$ with $\tilde{\mathcal{E}} := \bigcup \mathcal{E}$ and structural equations*

$$
\tilde{f}_v(x_V, x_W) := \mathop{\ast}_{\mathcal{I} \in \mathcal{E}} f_{v^{\mathcal{I}}}(x_V, x_W).
$$

Note that this is consistent with the nomenclature that is earlier introduced, where we referred to a DSCM as collapsed if $\mathcal{E} = \mathcal{T}$.

The following proposition shows that no information is lost when time-splitting a DSCM, but not when subsampling a DSCM.

Proposition 4. *For a DSCM $\mathcal{M}$ with evaluation index $\mathcal{E}$ that partitions $\mathcal{T}$ and a sequence of increasing time-indices $\tau = (t_1, ..., t_m) \subseteq \mathcal{T}$, we have $\mathcal{M} = \mathcal{M}_{\mathrm{ev}^+(\tau)*}$. If $\tau \neq \mathcal{T}$, we have $\mathcal{M} \neq \mathcal{M}_{\mathrm{ev}(\tau)*}$.*

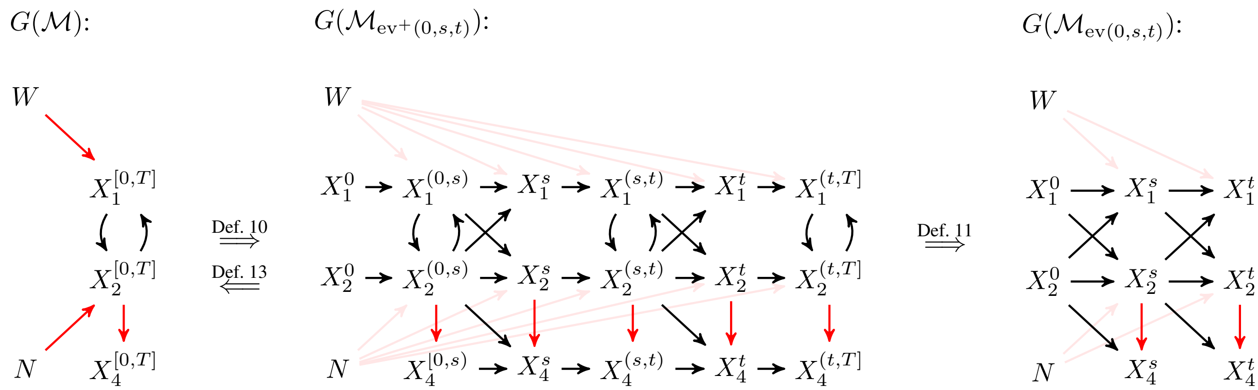

Figure 3: The graph $G(\mathcal{M})$, time-split $G(\mathcal{M}_{\mathrm{ev}^+(0,s,t)})$ and subsampled $G(\mathcal{M}_{\mathrm{ev}(0,s,t)})$.

5.1 LOCAL CONDITIONAL INDEPENDENCE

Throughout this section, let $\mathcal{M}$ be a simple DSCM with time-interval $\mathcal{T} = [0,T]$ for some $T \in \mathbb{R}$, with endogenous variables $A, B, C \in V$ such that $X_A, X_B, X_C \in \mathbb{S}_{\mathbb{F}_{\mathcal{X}_W}}$ (i.e. they are semimartingales). With respect to $\mathcal{F}^{A,B,C}$, such processes have a unique Doob-Meyer decomposition $X_B = \Lambda_B + M_B$, where 'the drift term' Λ_B is a predictable $\mathcal{F}_t^{A,B,C}$-adapted process, and 'the noise term' M_B is a $\mathcal{F}_t^{A,B,C}$-adapted martingale. We assume that the process Λ_B is absolutely continuous, i.e. $\Lambda_B = \int_0^t \lambda_B(s)\mathrm{d}s$ for some $\mathcal{F}_t^{A,B,C}$-predictable *intensity* process λ_B.

Definition 14 (Local conditional independence)**.** *We say that X_B is* locally conditionally independent *from X_A given X_C, written $X_A \not\to X_B \mid X_C$, if the intensity process λ_B has a $\mathcal{F}_t^{B,C}$-predictable version.*[12]

In the following theorem, we relate local independence to statistical relation between the variables in time-split DSCMs.

Theorem 3. *Considering DSCM $\mathcal{M}$ and for every $t \in \mathcal{T}$ the time-split DSCM $\mathcal{M}_{\mathrm{ev}^+(t)}$, we have that*

$$\begin{aligned} & X_B^t \perp\!\!\!\perp X_A^{[0,t)} | X_B^{[0,t)}, X_C^{[0,t)} \text{ for all } t \in \mathcal{T} \\ \Longrightarrow\ & X_A \not\to X_B \mid X_C \\ \Longrightarrow\ & \mathbb{E}[X_B^t \mid X_A^{[0,t)}, X_B^{[0,t)}, X_C^{[0,t)}] = \mathbb{E}[X_B^t \mid X_B^{[0,t)}, X_C^{[0,t)}] \\ & \text{for all } t \in \mathcal{T}. \end{aligned}$$

Definition 15 (Local independence graph, Mogensen and Hansen (2022))**.** *Given a set V of semimartingales X_v with joint distribution $\mathbb{P}(X_V)$, a directed mixed graph $G = (V, E)$ is a local independence graph for $\mathbb{P}(X_V)$ if $A \to B$ not in $G(\mathcal{M})$ implies $X_A \not\to_{\mathbb{P}} X_B | X_{V \setminus A}$ for all $A, B \in V$.*

[12] If one does not want to assume the existence of the intensity λ_B, local conditional independence can be defined more generally as Λ_B having a $\mathcal{F}_t^{B,C}$-adapted version, or equivalently, that M_B has a version that is a $\mathcal{F}_t^{B,C}$-adapted martingale (Mogensen and Hansen, 2020).

Existing work on local independence often assume some kind of *autonomy* of the processes X_A, X_B, X_C. Aalen et al. (2012) assumes that the martingales M_A, M_B, M_C are strongly orthogonal (i.e. so have zero quadratic covariation). Didelez (2008); Røysland et al. (2024) consider pure jump-type processes and assume orthogonality, ensuring they never jump simultaneously. Christgau et al. (2023) require X_A, X_C to be predictable, which ensures $M_A = M_C = 0$. We consider a slightly stronger notion of autonomy:

Assumption 3 (Independent integrators)**.** Let $\mathcal{M}$ be a collapsed DSCM with $\beta(v) \subseteq W$ and $\beta(v) \cap \beta(v') = \emptyset$ for all $v, v' \in V$.

In words, this means that there can be no instantaneous effects between the endogenous variables, nor any instantaneous dependencies between them through latent confounders.

Theorem 4. *For a collapsed DSCM $\mathcal{M}$ that satisfies Assumption 3 we have:*

1. *$G(\mathcal{M})$ is a local independence graph,*
2. *a σ-separation local independence global Markov property:*

$$A \underset{G(\mathcal{M})}{\overset{\sigma}{\perp}} B|C \implies X_A \not\to X_B \mid X_C.$$

Example 6. The graph $G(\mathcal{M}_{\mathcal{D}})$ in Figure 2 is not guaranteed to be a local independence graph, but $G((\mathcal{M}_{\mathcal{D}})_{\mathrm{marg}(L)})$ with $L := \{X_3, X_4\}$ is.

We note that the above Markov property is weaker than existing Markov properties for δ or μ separation (e.g. Didelez (2008) for orthogonal counting processes, and Mogensen et al. (2018) for Itô processes with independent noise), as δ- and μ-separation imply σ-separation. However, it applies to any simple DSCM induced by uniquely solvable systems of SDEs with independent integrators, so spans a much larger class of processes.[13]

[13] A notable exception is given in Mogensen and Hansen (2022) for Ornstein-Uhlenbeck processes with correlated noise.

6 FURTHER APPLICATIONS

In this section, we briefly elaborate how existing results for causal effect identification and causal discovery can directly be applied to DSCMs.

6.1 CAUSAL EFFECT IDENTIFICATION

We note that perfect interventions can be modelled for an entire trajectory $X_v^{[0,T]}$ in a collapsed DSCM, on a specific subinterval $X_v^{[s,t]}$ with $[s,t] \subseteq \mathcal{T}$ in a time-split DSCM, or on a time-point X_v^t in either a time-split or subsampled DSCM, e.g. in the DSCMs depicted in Figure 3.

The rules of do-calculus are valid for simple (D)SCMs (Forré and Mooij, 2020). These criteria are stated in terms of so-called *intervention nodes*, which for variable $X \in V$ are defined as vertices I_X in $G(\mathcal{M})$ with $I_X \to X$ as their only connected edge.

Theorem 5 (Do-calculus, Forré and Mooij (2020))**.** *Given a (D)SCM $\mathcal{M}$ with $X, Y, Z, W \subseteq V$, we have*

- *Insertion/deletion of observation:*
  $Y \perp^\sigma X \mid Z, \mathrm{do}(W) \implies$
  $\mathbb{P}(Y \mid X, Z, \mathrm{do}(W)) = \mathbb{P}(Y \mid Z, \mathrm{do}(W))$
- *Action/observation exchange:*
  $Y \perp^\sigma I_X \mid X, Z, \mathrm{do}(W) \implies$
  $\mathbb{P}(Y \mid \mathrm{do}(X), Z, \mathrm{do}(W)) = \mathbb{P}(Y \mid X, Z, \mathrm{do}(W))$
- *Insertion/deletion of actions:*
  $Y \perp^\sigma I_X \mid Z, \mathrm{do}(W) \implies$
  $\mathbb{P}(Y \mid \mathrm{do}(X), Z, \mathrm{do}(W)) = \mathbb{P}(Y \mid Z, \mathrm{do}(W))$.

As a consequence, generalised adjustment formulae (like backdoor adjustment) based on σ-separation criteria are also valid for simple DSCMs, as is the ID algorithm. For more information, we refer to Forré and Mooij (2020).

The above mentioned results consider the identification of certain estimands in terms of observational distributions. We note that the construction of *estimators* for such expressions can be highly nontrivial when the variables take values in function spaces.

6.2 CAUSAL DISCOVERY

For a given simple DSCM with graph G, let $\mathrm{IM}_\sigma(G)$ denote the σ-independence model of G, i.e. the set of all σ-separations that are present in G. Constraint-based causal discovery algorithms are methods that reconstruct (an equivalence class of) G from the independence model $\mathrm{IM}_\sigma(G)$. An example is *Fast Causal Discovery* (FCI) (Spirtes et al., 1995; Zhang, 2008), which outputs a *Partial Ancestral Graph* (PAG), and which is shown to be sound and complete for simple (and thus possibly cyclic) SCMs. In the following let F denote FCI as a mapping from an independence model to a PAG.

Theorem 6 (FCI, Mooij and Claassen (2020))**.** *Let $\mathcal{M}$ be a simple (D)SCM with graph G, then*

1. *FCI is sound: $G \in F(\mathit{IM}_\sigma(G))$*
2. *FCI is complete: let G' be another DMG, then $F(\mathit{IM}_\sigma(G)) = F(\mathit{IM}_\sigma(G'))$ iff $\mathit{IM}_\sigma(G) = \mathit{IM}_\sigma(G')$.*

Other constraint-based causal discovery algorithms that are known to be sound for (possibly cyclic) simple DSCMs are Local Causal Discovery (Cooper, 1997; Mooij et al., 2020) and Y-structures (Mani, 2006; Mooij et al., 2020).

In practice, one requires a conditional independence (CI) test to map a dataset with samples of an SCM $\mathcal{M}$ to an independence model (assuming faithfulness). If variables X, Y, Z take values in $D(\mathcal{T}, \mathbb{R})$, testing $X \perp\!\!\!\perp Y \mid Z$ is not straightforward. Recently, first results for such functional CI testing are proposed in Lundborg et al. (2022); Laumann et al. (2023) and Manten et al. (2024).

Example 7. Let $\mathcal{M}_\mathcal{D}$ be the DSCM from Example 4. The output of FCI on the σ-separation independence model of $G((\mathcal{M}_\mathcal{D})_{\mathrm{ev}^+(0)})$ is given in Figure 4.

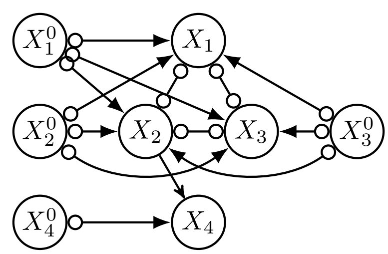

Figure 4: The output of FCI on $\mathrm{IM}_\sigma(G((\mathcal{M}_\mathcal{D})_{\mathrm{ev}^+(0)}))$.

We conjecture that 'Tiered FCI' (Andrews et al., 2020) is sound for cyclic DSCMs, and applying it on evaluated DSCM $\mathcal{M}_{\mathrm{ev}(t)}$ and subsequently mapping its output back to the PAG for $\mathcal{M}$ can be used to orient more edges in the output of regular FCI.

7 CONCLUSION

In this work, we have refined the concept of a *Dynamic Structural Causal Model*. We have shown that a large class of systems of differential equations can be mapped to a DSCM, giving it a Markov property based on σ-separation. We have formalised the concepts of *time-splitting* and *subsampling*, which both subsequently result in a DSCM, and thus remain to have clear causal semantics. We have shown how a time-split DSCM can shed some light on the concept of local conditional independence, and that the graph of a collapsed DSCM can be interpreted as a local independence graph if all integrators are exogenous and independent. Additionally, we have shown how existing results for causal effect identification and constraint-based causal discovery for simple SCMs can directly be applied to DSCMs.

REFERENCES

Aalen, O. (1978). Nonparametric Inference for a Family of Counting Processes. *The Annals of Statistics*, 6(4):701–726.

Aalen, O. O., Røysland, K., Gran, J. M., and Ledergerber, B. (2012). Causality, mediation and time: A dynamic viewpoint. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(4):831–861.

Andrews, B., Spirtes, P., and Cooper, G. F. (2020). On the Completeness of Causal Discovery in the Presence of Latent Confounding with Tiered Background Knowledge. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 4002–4011. PMLR.

Banks, H. T., Hu, S., and Thompson, W. C. (2014). *Modeling and Inverse Problems in the Presence of Uncertainty*. Chapman and Hall/CRC, 0 edition.

Blom, T., Van Diepen, M. M., and Mooij, J. (2021). Conditional independences and causal relations implied by sets of equations. *The Journal of Machine Learning Research*, 22(1):178:8044–178:8105.

Bongers, S., Blom, T., and Mooij, J. (2022). Causal Modeling of Dynamical Systems.

Bongers, S., Forré, P., Peters, J., and Mooij, J. (2021). Foundations of structural causal models with cycles and latent variables. *Ann. Stat.*, 49(5).

Christgau, A. M., Petersen, L., and Hansen, N. R. (2023). Nonparametric conditional local independence testing. *The Annals of Statistics*, 51(5).

Cooper, G. F. (1997). A Simple Constraint-Based Algorithm for Efficiently Mining Observational Databases for Causal Relationships. *Data Mining and Knowledge Discovery*.

Dean, T. and Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence*, 5(2):142–150.

Didelez, V. (2008). Graphical Models for Marked Point Processes Based on Local Independence. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(1):245–264.

Forré, P. and Mooij, J. (2017). Markov Properties for Graphical Models with Cycles and Latent Variables.

Forré, P. and Mooij, J. (2020). Causal Calculus in the Presence of Cycles, Latent Confounders and Selection Bias. In *PMLR*, pages 71–80. PMLR.

Forré, P. and Mooij, J. (2023). A Mathematical Introduction to Causality.

Hansen, N. and Sokol, A. (2014). Causal interpretation of stochastic differential equations. *Electronic Journal of Probability*, 19(none):1–24.

Iwasaki, Y. and Simon, H. A. (1994). Causality and model abstraction. *Artificial Intelligence*, 67(1):143–194.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren Der Mathematischen Wissenschaften*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kallenberg, O. (2021). *Foundations of Modern Probability*, volume 99 of *Probability Theory and Stochastic Modelling*. Springer International Publishing, Cham.

Laumann, F., von Kügelgen, J., Park, J., Schölkopf, B., and Barahona, M. (2023). Kernel-Based Independence Tests for Causal Structure Learning on Functional Data. *Entropy*, 25(12):1597.

Lundborg, A., Shah, R., and Peters, J. (2022). Conditional Independence Testing in Hilbert Spaces with Applications to Functional Data Analysis. *J. R. Stat. Soc. Ser. B Methodol.*, 84(5):1821–1850.

Lyons, T. J., Caruana, M., and Lévy, T. (2007). *Differential Equations Driven by Rough Paths: École d'Été de Probabilités de Saint-Flour XXXIV - 2004*, volume 1908 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Malinsky, D. and Spirtes, P. (2018). Causal Structure Learning from Multivariate Time Series in Settings with Unmeasured Confounding. In *Proceedings of 2018 ACM SIGKDD Workshop on Causal Discovery*, pages 23–47. PMLR.

Mani, S. (2006). *A Bayesian Local Causal Discovery Framework*. University of Pittsburgh ETD, University of Pittsburgh.

Manten, G., Casolo, C., Ferrucci, E., Mogensen, S., Salvi, C., and Kilbertus, N. (2024). Signature Kernel Conditional Independence Tests in Causal Discovery for Stochastic Processes.

Mogensen, S. and Hansen, N. R. (2020). Markov equivalence of marginalized local independence graphs. *Ann. Stat.*, 48(1).

Mogensen, S. W. and Hansen, N. R. (2022). Graphical modeling of stochastic processes driven by correlated noise. *Bernoulli*, 28(4).

Mogensen, S. W., Malinsky, D., and Hansen, N. R. (2018). Causal Learning for Partially Observed Stochastic Dynamical Systems.

Mooij, J. and Claassen, T. (2020). Constraint-Based Causal Discovery using Partial Ancestral Graphs in the presence of Cycles. In *UAI2020*, pages 1159–1168. PMLR.

Mooij, J. M., Magliacane, S., and Claassen, T. (2020). Joint causal inference from multiple contexts. *The Journal of Machine Learning Research*, 21(1):99:3919–99:4026.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Peters, J., Janzing, D., and Schölkopf, B. (2013). Causal Inference on Time Series using Restricted Structural Equation Models. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Przybyłowicz, P., Schwarz, V., Steinicke, A., and Szölgyenyi, M. (2023). A Skorohod measurable universal functional representation of solutions to semimartingale SDEs.

Røysland, K., Ryalen, P., Nygård, M., and Didelez, V. (2024). Graphical criteria for the identification of marginal causal effects in continuous-time survival and event-history analyses.

Rubenstein, P. K., Bongers, S., Schölkopf, B., and Mooij, J. M. (2018). From Deterministic ODEs to Dynamic Structural Causal Models.

Skorokhod, A. V. (1956). Limit Theorems for Stochastic Processes. *Theory of Probability & Its Applications*, 1(3):261–290.

Spirtes, P. L., Meek, C., and Richardson, T. S. (1995). Causal Inference in the Presence of Latent Variables and Selection Bias.

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17):1873–1896.

A ADAPTEDNESS OF SOLUTIONS OF SIMPLE DSCMS

Naturally, for a solution $(X_V, X_W) : \Omega \times \mathcal{T} \to \mathbb{R}^{|V \cup W|}$ of a DSCM $\mathcal{M}$ and $\mathbb{F}_{X_W}$ the filtration on Ω consisting of sigma algebras $\mathcal{F}_t := \sigma(\{X_W(s) : s \le t\})$, one would require X_{V_p} to be $\mathbb{F}_{X_W}$-adapted. However, for intervened DSCMs $\mathcal{M}_{\mathrm{do}(X_T = x_T)}$ we also want the solution to not depend on future values of x_T; a property that cannot be expressed in terms of a filtration on Ω, since the intervened variable X_T does not depend on Ω. To mitigate this problem, we require adaptedness of the solution functions with respect to the *canonical filtration on* $\mathcal{X}_W$, which we will define shortly.

The classical approach for modelling stochastic processes is to consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and models for example a Brownian motion $X : \Omega \times \mathcal{T} \to \mathbb{R}$ and a Poisson process $Y : \Omega \times \mathcal{T} \to \mathbb{R}$, for which the induced filtrations $\mathbb{F}_X$ and $\mathbb{F}_Y$ on Ω are different. We can equivalently model the exogenous processes directly on $D(\mathcal{T}, \mathbb{R})$ via the laws $\mathbb{P}(X)$ and $\mathbb{P}(Y)$ (so without some underlying measurable space Ω): the corresponding stochastic processes $\tilde{X}$ and $\tilde{Y}$ on $D(\mathcal{T}, \mathbb{R})$ are modelled as the *coordinate mapping process* $Z : D(\mathcal{T}, \mathbb{R}) \times \mathcal{T} \to \mathbb{R}, (\omega, t) \mapsto \omega(t)$, and equipping the measurable space $D(\mathcal{T}, \mathbb{R})$ either with $\mathbb{P}(X)$ or $\mathbb{P}(Y)$ determines whether Z is a Brownian motion or Poisson process. The *canonical filtration* $\mathbb{F}$ on $D(\mathcal{T}, \mathbb{R})$ is the filtration generated by Z.

Definition 16 (Canonical filtration)**.** *The* canonical filtration *on $D(\mathcal{T}, \mathbb{R})$ is defined as $\mathbb{F} = \{\mathcal{F}_t : t \in \mathcal{T}\}$ with*

$$\mathcal{F}_t = \sigma\left(\{z \in D(\mathcal{T}, \mathbb{R}) : z(s) \in B, s \le t, B \in \mathcal{B}(\mathbb{R})\}\right).$$

An adapted, measurable function $Y = g(Z)$ of such a coordinate process Z then generates a subfiltration $\mathcal{F}_t^Y \subseteq \mathcal{F}_t$. In Proposition 5, we show that solutions of simple DSCMs are adapted to the canonical filtration on the outcome spaces of the exogenous (and intervened, if applicable) variables.

Lemma 1. *For given process X and $Y := g(X)$ with g measurable, we have $Y \in \mathcal{F}_t^X$ if g is adapted, and $Y \in \mathcal{F}_{t-}^X$ if g is predictable, where $\mathcal{F}_t^X$ denotes the filtration generated by X.*

Proof. We have

$$\begin{aligned}\mathcal{F}_t^X &= \sigma(\{X(s) : s \le t\}) \\ &= \{\pi_s(X) \in B \mid s \le t, B \in \mathcal{B}(\mathbb{R})\}\end{aligned}$$

and

$$\mathcal{F}_{t-}^X = \{\pi_s(X) \in B \mid s < t, B \in \mathcal{B}(\mathbb{R})\}$$

similarly. Let $Z = g(X)$ with g adapted. For $t \in \mathcal{T}$ we have

$$\begin{aligned}&\{Z(t) \in B \mid B \in \mathcal{B}(\mathbb{R})\} \\ &\quad = \{\pi_t \circ g \circ X \in B \mid B \in \mathcal{B}(\mathbb{R})\} \\ &\quad = \{X \in g^{-1}(\pi_t^{-1}(B)) \mid B \in \mathcal{B}(\mathbb{R})\} \\ &\quad = \{\pi_s(X) \in \pi_s(g^{-1}(\pi_t^{-1}(B))) \mid s \le t, B \in \mathcal{B}(\mathbb{R})\}.\end{aligned}$$

where the last equality holds since g is adapted. This gives

$$\begin{aligned}&\{Z(t) \in B \mid B \in \mathcal{B}(\mathbb{R})\} \\ &\quad \subseteq \{\pi_s(X) \in B \mid s \le t, B \in \mathcal{B}(\mathbb{R})\} = \mathcal{F}_t^X.\end{aligned}$$

Similarly, if $Z = g(X)$ with g predictable, we have

$$\begin{aligned}&\{Z(t) \in B \mid B \in \mathcal{B}(\mathbb{R})\} \\ &\quad = \{\pi_s(X) \in \pi_s(g^{-1}(\pi_t^{-1}(B))) \mid s < t, B \in \mathcal{B}(\mathbb{R})\} \\ &\quad \subseteq \{\pi_s(X) \in B \mid s < t, B \in \mathcal{B}(\mathbb{R})\} = \mathcal{F}_{t-}^X.\end{aligned}$$

■

Proposition 5 (Adaptedness of solutions)**.** *Given an intervened simple DSCM $\mathcal{M}_{\mathrm{do}(X_T = x_t)}$ (write $O := V \setminus T$), the solution X_O of the is adapted to the canonical filtration on $\mathcal{X}_{T \cup W}$.*

Proof. Since $\mathcal{M}_{\mathrm{do}(X_T = x_t)}$ is simple there is a solution function g_O such that $X_O = g_O(x_T, X_W)$. By Proposition 2 we have that g_O is adapted, so by Lemma 1 we have that X_O is adapted to $\mathcal{F}_t^{X_T, X_W}$. Since we consider (X_T, X_W) to be the coordinate process on $\mathcal{X}_T \times \mathcal{X}_W$, the filtration $\mathcal{F}_t^{X_T, X_W}$ is the canonical filtration on $\mathcal{X}_T \times \mathcal{X}_W$. ■

B DSCM INDUCED BY A SYSTEM OF SDES

To prove that $\mathcal{M}_{\mathcal{D}}$ is simple, we use the following notion:

Definition 17 (Perfect interventions on SDEs)**.** *Given a system of SDEs $\mathcal{D} = \langle V, W, (X_V^0, X_W), f \rangle$, intervention target $T \subseteq V$ and intervention value $x_T \in D(\mathcal{T}, \mathbb{R}^{|T|})$, the* intervened system of SDEs *is defined as $\mathcal{D}_{\mathrm{do}(X_T = x_T)} = \langle V, W, (X_V^0, X_W), f^\circ \rangle$ with*

$$f_v^\circ : \begin{cases} X_v(t) = X_v^0 \\ \quad + \int_0^t g_v(s-, X_{\alpha(v)}) \mathrm{d}h_v(s, X_{\beta(v)}) & \text{if } v \in V \setminus T \\ X_v(t) = x_v(t) & \text{if } v \in T, \end{cases}$$

where x_T is to be interpreted as an element of $\mathbb{S}_F$, so as function $x_T : \Omega \times \mathcal{T} \to \mathbb{R}^{|T|}$ that is constant in Ω.

A definition of a soft-intervention on a system of SDEs can be found in Hansen and Sokol (2014).

Proposition 6. *Given a system of SDEs $\mathcal{D} = \langle V, W, (X_V^0, X_W), f \rangle$, intervention target $T \subseteq V$ and intervention value $x_T \in \mathcal{X}_T$, we have*

$$\mathcal{M}_{(\mathcal{D}_{\mathrm{do}(X_T=x_T)})} = (\mathcal{M}_\mathcal{D})_{\mathrm{do}(X_T=x_T)}.$$

Proof. We have $\mathcal{M}_\mathcal{D} = \langle V, W, \mathcal{X}_V, \mathcal{X}_W, f^*, \mathbb{P}(X_W)\rangle$ and $(\mathcal{M}_\mathcal{D})_{\mathrm{do}(X_T=x_T)} = \langle V, W, \mathcal{X}_V, \mathcal{X}_W, (f^*)^\circ, \mathbb{P}(X_W)\rangle$, and on the other hand $\mathcal{D}_{\mathrm{do}(X_T=x_T)} = \langle V, W, (X_V^0, X_W), f^\circ \rangle$ and $\mathcal{M}_{(\mathcal{D}_{\mathrm{do}(X_T=x_T)})} = \langle V, W, \mathcal{X}_V, \mathcal{X}_W, (f^\circ)^*, \mathbb{P}(X_W)\rangle$. Since the trivial SDE $X_T(t) = x_T(t)$ has Itô map $I_T(x_T) = x_T$, it is clear that taking the Itô map and intervention commute, i.e. $(f^*)^\circ = (f^\circ)^*$, so we have the required result. ∎

Proposition 7. *Given a uniquely solvable system of SDEs $\mathcal{D}$, the induced DSCM $\mathcal{M}_\mathcal{D}$ is simple.*

Proof. For every $T \subseteq V$ and $x_T \in \mathcal{X}_T$, let $\mathcal{D}' := \left\langle V \setminus T, W \cup T, (X_{V\setminus T}^0, X_W, x_T), f_{V\setminus T} \right\rangle$, where x_T is to be interpreted as an element of $\mathbb{S}_F$ that is constant in Ω. Then $\mathcal{D}'$ is uniquely solvable, hence it has an Itô map $I_{V\setminus T}(x_V^0, x_T, x_W)$. Considering the intervened SDE $\mathcal{D}_{\mathrm{do}(X_T=x_T)}$ defined according to Definition 17, we have that $I_{V\setminus T}(x_V^0, x_T, x_W)$ is the solution function of $\mathcal{M}_{(\mathcal{D}_{\mathrm{do}(X_T=x_T)})}$. By Proposition 6 we have that $I_{V\setminus T}$ is the solution function of $(\mathcal{M}_\mathcal{D})_{\mathrm{do}(X_T=x_T)}$ as well, so we have unique solvability of $\mathcal{M}_\mathcal{D}$. ∎

C PROOFS

Proposition 1 (Unique solvability of a system of SDEs)**.** *Let $\mathcal{D} = \langle V, W, (X_V^0, X_W), f \rangle$ be a system of SDEs such that for every SDE f_v, the parameters g_v and h_v satisfy Assumptions 2 and 1 respectively. If for every $v \in V$ we have $\beta(v) \cap \mathrm{Sc}(v) = \emptyset$, there exists an Itô map I_V such that $X_V := I_V(X_V^0, X_W) \in \mathbb{S}_\mathbb{F}(\mathcal{T}, \mathbb{R}^{|V|})$ is the unique solution of $\mathcal{D}$.*

Proof. Let $(S_1, ..., S_k)$ with $k \leq |V|$ be a topologically ordered partition of V into strongly connected components of $G(\mathcal{D})$. For the first component S_1 write $\tilde{\mathrm{Pa}}(S_1) = \alpha(S_1) \cup \beta(S_1) \setminus S_1$, which has $\tilde{\mathrm{Pa}}(S_1) \subseteq W$, hence we can apply Theorem 1 to obtain $X_{S_1} := I_{S_1}(X_{S_1}^0, X_{\tilde{\mathrm{Pa}}(S_1)})$ as the solution of the SDE f_{S_1}. For given $i \in \{1, ..., k\}$ denote $S_{\leq i} := \bigcup_{j \leq i} S_j$, and assume that $X_{S_{\leq i}} := I_{S_{\leq i}}(X_{S_{\leq i}}^0, X_{\tilde{\mathrm{Pa}}(S_{\leq i}) \cap W})$ is the solution of the SDE $f_{S_{\leq i}}$. We have that $\tilde{\mathrm{Pa}}(S_{i+1}) \subseteq S_{\leq i} \cup W$. We apply Theorem 1 again to obtain the Itô map $I_{S_{i+1}}(X_{S_{i+1}}^0, X_{\tilde{\mathrm{Pa}}(S_{i+1}) \cap S_{\leq i}}, X_{\tilde{\mathrm{Pa}}(S_{i+1}) \cap W})$. The Itô map for the SDE $f_{S_{\leq i+1}}$ is then $I_{S_{\leq i+1}}(X_V^0, X_W)$ with as its S_{i+1}-component $I_{S_{i+1}}(X_{S_{i+1}}^0, I_{\tilde{\mathrm{Pa}}(S_{i+1}) \cap V}(X_V^0, X_W), X_{\tilde{\mathrm{Pa}}(S_{i+1}) \cap W})$, which gives a unique solution $X_{S_{\leq i+1}} := I_{S_{\leq i+1}}(X_V^0, X_W)$. Finally, set $I_V := I_{S_{\leq k}}$. ∎

Proposition 2. *Given a simple DSCM $\mathcal{M}$, the solution function $g_O(x_T, x_W)$ of the intervened DSCM $\mathcal{M}_{\mathrm{do}(X_T=x_T)}$ is adapted.*

Proof. By definition of the solution function g_O and adaptedness of f_O we have for all $x_T \in \mathcal{X}_T$ and $\mathbb{P}(X_W)$-almost all $x_W \in \mathcal{X}_W$ that

$$\begin{aligned} g_O(x_T, x_W)^t &= f_O(g_O(x_T, x_W), x_T, x_W)^t \\ &= f_O(g_O(x_T, x_W)^t, x_T^t, x_W^t)^t \\ &= f_O(g_O(x_T^t, x_W^t)^t, x_T^t, x_W^t)^t \\ &= g_O(x_T^t, x_W^t)^t, \end{aligned}$$

so g_O is adapted. ∎

Proposition 3. *The class of simple DSCMs is closed under marginalisation.*

Proof. By Bongers et al. (2021), Proposition 8.2 we have that $\mathcal{M}_{\mathrm{marg}(L)}$ is a simple SCM. Since $\tilde{f}(x_O, x_W)^t = f_O(x_O, g_L(x_O, x_W), x_W)^t = f_O(x_O^t, g_L(x_O^t, x_W^t)^t, x_W^t)^t = \tilde{f}(x_O^t, x_W^t)^t$ holds by adaptedness of f and g_L, we also have that its structural equation $\tilde{f}$ is adapted, and so $\mathcal{M}_{\mathrm{marg}(L)}$ is indeed a simple DSCM. ∎

Proposition 4. *For a DSCM $\mathcal{M}$ with evaluation index $\mathcal{E}$ that partitions $\mathcal{T}$ and a sequence of increasing time-indices $\tau = (t_1, ..., t_m) \subseteq \mathcal{T}$, we have $\mathcal{M} = \mathcal{M}_{\mathrm{ev}^+(\tau)*}$. If $\tau \neq \mathcal{T}$, we have $\mathcal{M} \neq \mathcal{M}_{\mathrm{ev}(\tau)*}$.*

Proof. The first claim follows directly from Definition 13. For the second claim, let $V \times \mathcal{E}$ and $V \times \mathcal{E}'$ denote the endogenous indices of $\mathcal{M}$ and $\mathcal{M}_{\mathrm{ev}(\tau)*}$ respectively. If $\tau \neq \mathcal{T}$ we have $\mathcal{E}' \neq \mathcal{E}$, which proves the result. ∎

C.1 LOCAL INDEPENDENCE

As a link between the notion of local independence and conditional independence of the sample paths of stochastic processes, we relate a conditional distribution $\mathbb{E}[X_B^t \mid \mathcal{F}_t^A] : \Omega \to \mathbb{R}$ to the conditional distribution $\mathbb{E}\left[X_B^t \mid X_A^{[0,t]}\right] : \mathbb{R} \to \mathbb{R}$.

Lemma 2. *Given a DSCM $\mathcal{M}$ with $A, B \subseteq V$, in $\mathcal{M}_{\mathrm{ev}(t)}$ we have $\mathbb{E}\left[X_B^t \mid \mathcal{F}_t^A\right] = \mathbb{E}\left[X_B^t \mid X_A^{[0,t]} = X_A^{[0,t]}\right]$ a.s. and $\mathbb{E}\left[X_B^t \mid \mathcal{F}_{t-}^A\right] = \mathbb{E}\left[X_B^t \mid X_A^{[0,t)} = X_A^{[0,t)}\right]$ a.s.*

Proof. First, considering the random variable $X_A^{[0,t]}: D(\mathcal{T},\mathbb{R}^{|W|}) \to D([0,t],\mathbb{R}^{|A|})$, we have

$$\begin{aligned}
&\sigma(X_A^{[0,t]}(s) : s \le t) \\
&= \sigma(\{\{x_W : g_A^{[0,t]}(x_W)(s) \in F\} \mid s \le t, F \in \mathcal{B}(\mathbb{R})\}) \\
&= \sigma(\{\{x_W : g_A^{[0,t]}(x_W) \in F'\} \mid F' \in \mathcal{B}(D([0,t],\mathbb{R}^{|A|}))\}) \\
&= \sigma(X_A^{[0,t]}),
\end{aligned}$$

since $\mathcal{B}(D([0,t],\mathbb{R})) = \sigma(\omega_s : s \in [0,t]) = \sigma(\{\{x_A : x_A(s) \in F\} : s \in [0,t], F \in \mathcal{B}(\mathbb{R})\})$. We then have

$$\begin{aligned}
\mathbb{E}\left[X_B^t \mid \mathcal{F}_t^A\right] &= \mathbb{E}\left[X_B^t \mid \sigma(X_A^{[0,t]}(s) : s \le t)\right] \\
&= \mathbb{E}\left[X_B^t \mid \sigma(X_A^{[0,t]})\right] \\
&= \mathbb{E}\left[X_B^t \mid X_A^{[0,t]} = X_A^{[0,t]}\right],
\end{aligned}$$

where the last equality holds by the factorisation lemma (Kallenberg (2021), Lemma 1.14). We also have

$$\mathcal{F}_{t-}^A = \sigma\left(\mathcal{F}_s^A : s < t\right) = \sigma(X_A^{[0,t]}(s) : s < t) = \sigma(X_A^{[0,t)}),$$

and so by the factorisation lemma $\mathbb{E}\left[X_B^t \mid \mathcal{F}_{t-}^A\right] = \mathbb{E}\left[X_B^t \mid X_A^{[0,t)} = X_A^{[0,t)}\right]$. ■

Theorem 3. *Considering DSCM $\mathcal{M}$ and for every $t \in \mathcal{T}$ the time-split DSCM $\mathcal{M}_{\mathrm{ev}^+(t)}$, we have for all $A, B, C \subseteq V$ that*

$$\begin{aligned}
&X_B^t \perp\!\!\!\perp X_A^{[0,t)} | X_B^{[0,t)}, X_C^{[0,t)} \textit{ for all } t \in \mathcal{T} \\
&\implies X_A \not\to X_B \mid X_C \\
&\implies \mathbb{E}[X_B^t \mid X_A^{[0,t)}, X_B^{[0,t)}, X_C^{[0,t)}] = \mathbb{E}[X_B^t \mid X_B^{[0,t)}, X_C^{[0,t)}] \\
&\qquad\qquad \textit{for all } t \in \mathcal{T}.
\end{aligned}$$

Proof. First, by Lemma 2 we have that $\mathbb{E}[X_B(t) \mid X_A^{[0,t)}, X_B^{[0,t)}, X_C^{[0,t)}] = \mathbb{E}[X_B(t) \mid X_B^{[0,t)}, X_C^{[0,t)}]$ holds a.s. iff $\mathbb{E}\left[X_B(t) \mid \mathcal{F}_{t-}^{A,B,C}\right] = \mathbb{E}\left[X_B(t) \mid \mathcal{F}_{t-}^{B,C}\right]$ a.s. By the innovation theorem (see e.g. Aalen (1978)), we can pick a version $\lambda^{B,C}(t)$ of $\mathbb{E}[\lambda^{A,B,C}(t) \mid \mathcal{F}_{t-}^{B,C}]$ that is a $\mathcal{F}_t^{B,C}$-intensity of X_B. By assumption, we can write

$$\begin{aligned}
X_B(t) &= \int_0^t \lambda^{A,B,C}(s)\mathrm{d}s + M^{A,B,C}(t) \\
&= \int_0^t \lambda^{B,C}(s)\mathrm{d}s + M^{B,C}(t)
\end{aligned}$$

for $M^{A,B,C}$ ($M^{B,C}$) a $\mathcal{F}_t^{A,B,C}$ (resp. $\mathcal{F}_t^{B,C}$) martingale. By Aalen (1978), we have $\lambda_t^{A,B,C} = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}[X_B(t+h) - X_B(t) \mid \mathcal{F}_{t-}^{A,B,C}]$. Since $\lambda^{A,B,C}$ and $\lambda^{B,C}$ are bounded, by dominated convergence we have $\lambda_t^{A,B,C} = \mathbb{E}[\lim_{h\downarrow 0} \frac{1}{h}(X_B(t+h) - X_B(t)) \mid \mathcal{F}_{t-}^{A,B,C}]$. Since $\lim_{h\downarrow 0} \frac{1}{h}(X_B(t+h) - X_B(t)) \in \mathcal{F}_t^B$ and the independence assertion can be equivalently stated as $\mathcal{F}_t^B \perp\!\!\!\perp \mathcal{F}_{t-}^A | \mathcal{F}_{t-}^{B,C}$, we get

$$\begin{aligned}
\lambda_t^{A,B,C} &= \mathbb{E}\left[\lim_{h\downarrow 0} \frac{1}{h}(X_B(t+h) - X_B(t)) \mid \mathcal{F}_{t-}^{A,B,C}\right] \\
&= \mathbb{E}\left[\lim_{h\downarrow 0} \frac{1}{h}(X_B(t+h) - X_B(t)) \mid \mathcal{F}_{t-}^{B,C}\right] \\
&= \lambda_t^{B,C},
\end{aligned}$$

proving local independence.

Under local conditional independence we have $\lambda^{A,B,C} = \lambda^{B,C}$ a.s., and by the above equation also $M^{A,B,C} = M^{B,C}$ a.s., so by Jacod and Shiryaev (2003), Lemma I.2.27 we have that

$$\begin{aligned}
\mathbb{E}\left[M^{A,B,C}(t)|\mathcal{F}_{t-}^{A,B,C}\right] &= M^{A,B,C}(t-) \\
&= M^{B,C}(t-) = \mathbb{E}\left[M^{B,C}(t)|\mathcal{F}_{t-}^{B,C}\right]
\end{aligned}$$

and so

$$\begin{aligned}
&\mathbb{E}\left[X_B(t) \mid \mathcal{F}_{t-}^{A,B,C}\right] \\
&= \mathbb{E}\left[\int_0^t \lambda^{A,B,C}(s)\mathrm{d}s + M^{A,B,C}(t) \mid \mathcal{F}_{t-}^{A,B,C}\right] \\
&= \mathbb{E}\left[\int_0^t \lambda^{B,C}(s)\mathrm{d}s + M^{B,C}(t) \mid \mathcal{F}_{t-}^{B,C}\right] \\
&= \mathbb{E}\left[X_B(t) \mid \mathcal{F}_{t-}^{B,C}\right],
\end{aligned}$$

so $\mathbb{E}\left[X_B(t) \mid X_A^{[0,t)}, X_B^{[0,t)}, X_C^{[0,t)}\right] = \mathbb{E}\left[X_B(t) \mid X_B^{[0,t)}, X_C^{[0,t)}\right]$. ■

Theorem 4. *For a collapsed DSCM $\mathcal{M}$ that satisfies Assumption 3 we have:*

1. *$G(\mathcal{M})$ is a local independence graph,*
2. *a σ-separation local independence global Markov property:*

$$A \underset{G(\mathcal{M})}{\overset{\sigma}{\perp}} B|C \implies X_A \not\to X_B \mid X_C$$

for all $A, B, C \subseteq V$.

Proof. Throughout this proof, write $G_t := G(\mathcal{M}_{\mathrm{ev}^+(t)})$.
1) Let $A, B \in V$. By Assumption 3 we have that $\mathrm{Pa}_{G(\mathcal{M})}(B) = \alpha(B)$. If $A \to B$ not in $G(\mathcal{M})$, we have $A \notin \alpha(B)$, so $X_A^{[0,t)} \notin \mathrm{Pa}_{G_t}(X_B^t) = X_{\alpha(B)}^{[0,t)}$ for all t. If we let $N := (V \setminus A) \setminus \alpha(B)$, then since $X_A^{[0,t)} \in \mathrm{Nd}(X_B^t)$ and $X_N^{[0,t)} \subset \mathrm{Nd}(X_B^t)$ we have by the local Markov property[14] that $X_B^t \perp\!\!\!\perp X_A^{[0,t)} | X_{\alpha(B)}^{[0,t)}$ and $X_B^t \perp\!\!\!\perp X_N^{[0,t)} | X_{\alpha(B)}^{[0,t)}$.

[14]For SCM $\mathcal{M}$ with $B \in V$, we have $X_B \perp\!\!\!\perp X_{\mathrm{Nd}(B)} | X_{\mathrm{Pa}(B)}$, where $\mathrm{Nd}(B) = V \setminus \mathrm{Desc}(B)$ is the set of non-descendants of B.

By the weak contraction property of conditional independence (and the fact that $V \setminus A = \alpha(B) \cup N$) this gives $X_B^t \perp\!\!\!\perp X_A^{[0,t)} | X_{V\setminus A}^{[0,t)}$ so by Theorem 3 we have $X_A \not\to_{\mathbb{P}} X_B | X_{V\setminus A}$.

2) Let G' be the $[0,t)$-slice of G_t, so $G' := G((\mathcal{M}_{\mathrm{ev}^+(t)})_{\mathrm{marg}(V\times(\mathcal{E}\setminus\{[0,t)\}))})$, for which we have $G' \subseteq G(\mathcal{M})$, hence $X_B^{[0,t)} \perp^\sigma_{G'} X_A^{[0,t)} | X_C^{[0,t)}$. Since all vertices in G' have no additional ancestors in G_t, we have $X_B^{[0,t)} \perp^\sigma_{G_t} X_A^{[0,t)} | X_C^{[0,t)}$. We have that every walk π from $X_A^{[0,t)}$ to X_B^t that is σ-open given $X_C^{[0,t)}$ must end in $v \to X_B^t$ for some $v \in \{X_B^{[0,t)}\} \cup \mathrm{Pa}(X_B^{[0,t)})$, hence a copy π' of that walk that ends in $v \to X_B^{[0,t)}$ must be σ-open given $X_C^{[0,t)}$ as well, contradicting $X_B^{[0,t)} \perp^\sigma_{G_t} X_A^{[0,t)} | X_C^{[0,t)}$, so we must have $X_B^t \perp^\sigma_{G_t} X_A^{[0,t)} | X_C^{[0,t)}$ for all t. Since this implies $X_B^t \perp^\sigma_{G_t} X_A^{[0,t)} | X_B^{[0,t)}, X_C^{[0,t)}$, Theorem 3 gives the result. ■