Causal Graphs, Markov Properties and Do-calculus for Stochastic Differential Equations

Philip Boeken* Joris M. Mooij†

July 13, 2026

Abstract

Stochastic differential equations (SDEs) are widely used to model continuous-time dynamical systems, but graphical causal models for them are not yet well-understood. We consider *systems of causal SDEs* that are equipped with an explicit causal semantics. We pose solvability conditions for systems of causal SDEs such that they have well-defined observational and interventional distributions — even after marginalisation — and provide a general class of Lipschitz semimartingale SDEs that satisfies these conditions. As core results we establish the σ-separation Markov property and the do-calculus in terms of the system's *causal graph* for probabilistic independence and interventions on the level of sample paths. For a class of additive-noise SDEs we prove a stronger d-separation Markov property, even if the system is cyclic. As a corollary of the do-calculus, we obtain an explicit causal interpretation of the graph: that the absence of a directed path implies the absence of a causal effect. We further introduce *time-split* systems, which consider the causal relations between the processes when evaluated on disjoint intervals or time-points, and use them to reason about subsampled time-series, continuous-time Granger non-causality and local independence. Finally, we discuss how constraint-based causal discovery algorithms (PC, FCI, CCD, CCI) apply directly to SDEs within our framework when conditional independence between sample paths can be consistently tested.

1 Introduction

Stochastic differential equations (SDEs) are widely used to model continuous-time dynamical systems subject to randomness, with applications across the natural sciences and engineering. Causal graphs for these continuous-time models are not yet well understood. Celebrated tools from causal modelling are the causal graph, the *Markov Property* (graphical separations between vertices imply conditional independencies between the variables), the *do-calculus* (graphical criteria that relate observational distributions to interventional distributions), and *causal discovery* (estimating the causal graph from data). These tools have not yet been worked out for general (possibly cyclic) systems of SDEs.

Causal modelling of continuous-time systems has been worked out in specific settings. Interventional semantics for equilibrium states of ODEs are developed in Mooij et al. (2013), and for the dynamic regime of ODEs in Rubenstein et al. (2018) and Bongers et al. (2022). Hansen and Sokol (2014) defined causal semantics for SDEs. Graphical models with a Markov property for Granger non-causality have been introduced by Eichler and Didelez (2010); Eichler (2012). A continuous-time notion of Granger non-causality called *local independence* has been

*Department of Mathematics, VU Amsterdam, p.a.boeken@vu.nl

†Korteweg-de Vries Institute for Mathematics, University of Amsterdam, j.m.mooij@uva.nl

introduced by Aalen (1978); Aalen et al. (2008); Florens and Fougere (1996); Comte and Renault (1996) and developed graphically by Didelez (2008); Mogensen and Hansen (2020, 2022) with causal interpretations by Aalen et al. (2012); Røysland et al. (2024) and nonparametric testing in Christgau et al. (2023). Causal discovery for discrete-time systems has been investigated by Malinsky and Spirtes (2018); Runge et al. (2019); Reiter et al. (2024); see Assaad et al. (2022) for an overview. Causal discovery in continuous time has likewise been carried out for specific graphs: dynamic-independence graphs for stochastic kinetic networks (Bowsher, 2010), local-independence graphs via an FCI-type algorithm (Mogensen et al., 2018), and the directed graph of a stable process from its characteristic exponents (Brück et al., 2026). Other approaches to causal discovery for systems of SDEs include Manten et al. (2024), Guan et al. (2024), Engelke et al. (2024) and Nathaniel et al. (2025). None of these lines of work establishes a graphical Markov property in terms of σ- or d-separation for probabilistic independence between sample paths, or provides a do-calculus for general (possibly cyclic) SDEs.

The framework of Structural Causal Models (SCMs) provides many tools for causal reasoning in static and discrete-time settings: causal graphs encode conditional independence structure, the do-calculus can be employed to identify causal effects from observational distributions, and constraint-based algorithms recover causal structure from data. Building on the cyclic-SCM theory of Forré and Mooij (2017); Bongers et al. (2021); Forré and Mooij (2025) and the pathwise solution function for SDEs of Przybyłowicz et al. (2024), this paper develops such a framework for general systems of causal SDEs, allowing cycles, instantaneous relations, jumps, and non-Markovianity.

Causal models that abstract away time in an unspecified way are inherently ambiguous (Reisach et al., 2025). Most existing approaches that explicitly model time do so in discrete time, e.g., via structural vector autoregression models or dynamic Bayesian networks. We give a formal interpretation of these models as projections of underlying continuous-time systems observed at specific points in time. This projection can distort causal inference: conditional independencies may vanish, estimates of causal effects may be invalid, and inferred causal structure may not correspond to its true continuous-time counterpart. By grounding causal semantics directly in continuous-time dynamics, our framework provides a principled foundation for interpreting discretely observed data while remaining mindful of the underlying processes.

1.1 Contributions

We equip systems of SDEs with a structural causal semantics via the notion of perfect intervention, following Mooij et al. (2013); Hansen and Sokol (2014); Rubenstein et al. (2018); Peters et al. (2020), and we give a definition of the *causal graph* of such systems of causal SDEs. Although not strictly necessary for our analysis, we also give a pathwise, functional formulation of the stochastic integral (Definition 4): it lets one read the system of SDEs as a structural causal model (SCM), which we find conceptually clarifying, and it supplies the notation Φ_v for each causal mechanism (structural equation) that we rely on throughout the proofs. We then provide the following results:

(i) *Solvability.* We analyse the *essential unique solvability* (Definition 5) of a system of causal SDEs with respect to a subset of variables, which is a solvability condition for a subset of the system, under arbitrary adapted inputs for the remaining variables. We provide two model classes of particular interest that satisfy these conditions: a class of semimartingale SDEs that satisfies standard Lipschitz and linear-growth conditions (Assumption 1), and an additive noise SDE driven by Brownian motion (Assumption 2).

(ii) *Marginalisation.* We define the marginalisation of a system of causal SDEs on a subset of the variables (Definition 6) that is causally consistent with the original model, and show that essential unique solvability is preserved under this operation (Theorem 5).

(iii) *Markov properties.* Following the acyclification strategy of Forré and Mooij (2017); Bongers et al. (2021); Forré and Mooij (2025), we prove the σ-separation Markov property in the causal graph $G(\mathcal{D})$ for certain essentially uniquely solvable systems (Theorem 6). For the class of additive-noise SDEs, we strengthen this to the d-separation Markov property (Theorem 11).

(iv) *Do-calculus.* We establish the three rules of the do-calculus in terms of σ-separation in the causal graph for essentially uniquely solvable systems (Theorem 12). As a consequence, the absence of a directed path implies the absence of a causal effect, giving the graph a clear causal interpretation (Theorem 13).

(v) *Time-splitting and subsampling.* We introduce *time-split* systems of causal SDEs, which separately model the causal relations between processes on disjoint subintervals (Definition 8), and as a special case the *subsampled* system that marginalises the time-split system onto a set of time-point measurements (Definition 9). We transfer the σ- and d-separation Markov properties and the do-calculus to them, which formalises continuous-time Granger non-causality (Granger, 1969) and local independence (Schweder, 1970; Didelez, 2008; Mogensen and Hansen, 2020), and yields a Markov property for local independence.

(vi) *Causal discovery.* We give a proof-of-concept application of the cyclic constraint-based discovery algorithms FCI (Spirtes et al., 1999; Mooij and Claassen, 2020) and CCI (Strobl, 2019) to SDEs, relying on a conditional independence oracle.

All results are illustrated on a single running example, introduced in the following section. Proofs are given in the appendix.

1.2 Example: the repressilator

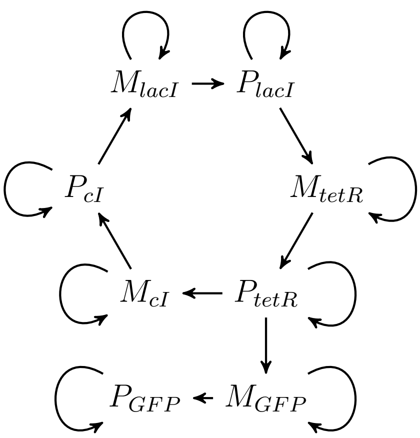
To make the exposition more concrete, we examine a biological system that has been extensively studied and lends itself naturally to a stochastic modelling perspective. Gene regulatory networks are a prototypical example: they consist of interacting components whose dynamics are driven both by feedback loops and by random fluctuations at the molecular level. Among such systems, the *repressilator* (Elowitz and Leibler, 2000) has become a canonical case study. It illustrates how cyclic interactions, noise, and causal mechanisms interact in a way that is both biologically relevant and mathematically tractable. We use this system as a running example of a system of SDEs that are *causal*, which means that the variable on the left-hand side of the equation is directly caused by those on the right-hand side of the equation.

Example 1 (Repressilator)**.** *The* repressilator *is an oscillator of gene expressions in the E. coli bacteria, specifically of the genes* lacI, tetR, cI *and* GFP*. Extending the original ODE model of Elowitz and Leibler (2000) by adding Brownian noise, the dynamics of the mRNA abundance M_i and protein values P_i for each gene i and its predecessor j for $(i, j) \in \{$(lacI, cI), (tetR, lacI), (cI, tetR), (GFP, tetR)$\}$ are given by*

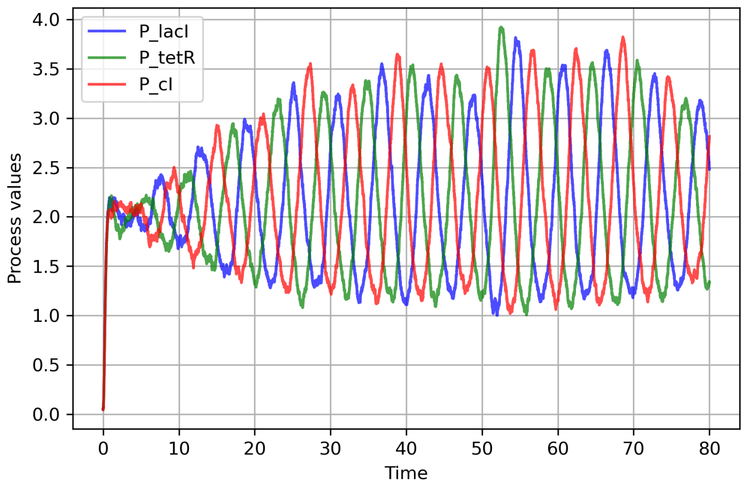
$$\begin{aligned} \mathrm{d}M_i(t) &= \left(\frac{\alpha}{1 + |P_j(t)|^n} + \alpha_0 - M_i(t)\right)\mathrm{d}t + \sigma \mathrm{d}Z_i^M(t) \\ \mathrm{d}P_i(t) &= \beta(M_i(t) - P_i(t))\mathrm{d}t + \sigma \mathrm{d}Z_i^P(t), \end{aligned}$$

with given exogenous random variables for the initial conditions $M_i(0)$ and $P_i(0)$, independent Brownian motions Z_i^M and Z_i^P, and values of $\alpha, \alpha_0, \beta, \sigma > 0$ and Hill coefficient $n \geq 1$. Setting $n = 1$ gives the well-known Michaelis-Menten kinetics. In this model, protein abundance P_j

inhibits the mRNA M_i of protein i, which is required for growth of the protein abundance P_i. The dependence structure is graphically depicted in Figure 1a, where $u \to v$ if variable u occurs in the SDE for variable v. In Definition 2 we will formalise this as a causal graph. A simulated sample path of the proteins in the cycle is provided in Figure 1b.



(a) Causal graph.



(b) Sample trajectories of the proteins P_{lacI}, P_{tetR} and P_{cI}.

Figure 1: The repressilator of the E. coli bacteria, modelled with additive noise.

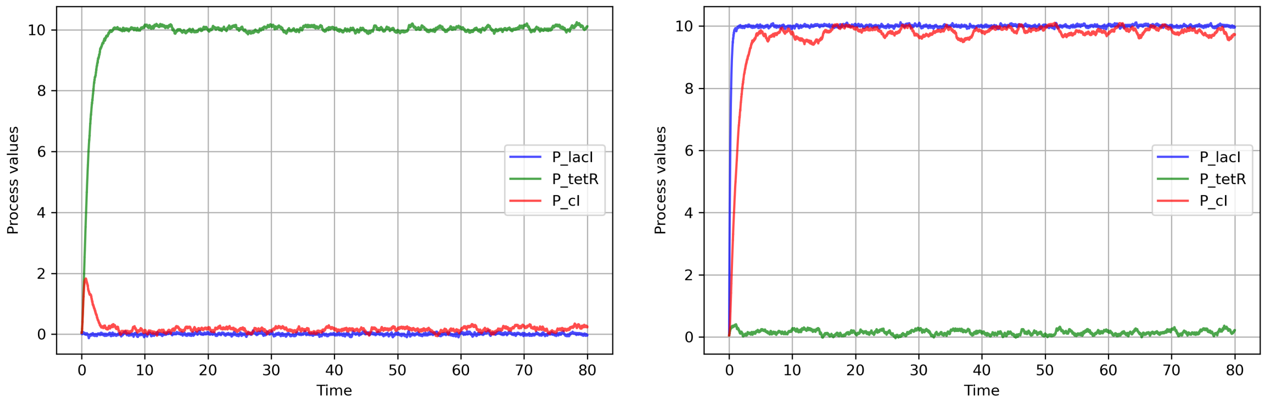
Experimentation is a cornerstone of the discovery of causal relations: by deliberately perturbing a system and observing the resulting changes, one can distinguish genuine cause-effect relationships from mere correlations. In the context of stochastic dynamics, this means modifying certain components of the governing equations and analysing how the system responds over time. Importantly, interventions are not only of interest for discovering causal structure, they also arise naturally in practice, where one aims to steer a system toward a desired state. Translating this intuition into the language of SDEs, a *perfect intervention* amounts to altering or overriding the dynamics of a particular variable by fixing that component to a pre-specified sample path, without any dependency on the remainder of the system, and without any side-effects. In the following example, we illustrate this idea by intervening on the repressilator.

Example 2 (Intervening on the repressilator)**.** *A typical way of experimenting with gene regulatory networks is to knock out a particular gene, disabling the production of the protein. Suppose we knock out the lacI gene, then we obtain the system where we set $M_{lacI}(t) = 0$ for all t. A sample of the resulting process for the proteins is depicted in Figure 2a. We see that tetR is not inhibited anymore and thus reaches high activity levels, highly suppressing cI abundance. Similarly, when exciting the lacI gene, this suppresses tetR, which in turn excites cI levels, as depicted in Figure 2b.*

In the following section, we make systems of causal SDEs and interventions on them (as demonstrated using the repressilator) mathematically precise, allowing for jumps in the sample paths, non-Markovianity, and instantaneous dependencies.

2 Systems of causal SDEs

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a filtered probability space with filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}$ for some $T < \infty$, satisfying *the usual conditions*: $\mathcal{F}$ is right-continuous and contains all $\mathbb{P}$-null-sets. Let $D([0,T], \mathbb{R}^n)$ be the space of functions from $[0,T]$ to $\mathbb{R}^n$ that are càdlàg (continue à droite,



(a) Gene 'lacI' knocked out.

(b) Gene 'lacI' excited.

Figure 2: The effects of interventions on the 'lacI' gene of the repressilator.

limité à gauche, i.e. right-continuous with left limits).[1] A stochastic process $X : \Omega \times [0,T] \to \mathbb{R}^n$ is adapted if $X(t) \in \mathcal{F}_t$ for all $t \in [0,T]$, and càdlàg if it has càdlàg sample paths almost surely, in which case we can represent it as a process $X : \Omega \to D([0,T],\mathbb{R}^n)$. Let $\mathbb{S}([0,T],\mathbb{R}^n)$ denote the class of $\mathbb{R}^n$-valued semimartingales, i.e. adapted processes $X : \Omega \to D([0,T],\mathbb{R}^n)$ that admit a decomposition $X = \Lambda + M$ with Λ a process of finite variation and M a local martingale.[2] Denoting with $\mathcal{P}([0,T],\mathbb{R}^m)$ the set of predictable processes on $\mathbb{R}^m$, the stochastic (Itô) integral of G w.r.t. H is a mapping

$$J : \mathcal{P}([0,T],\mathbb{R}^m) \times \mathbb{S}([0,T],\mathbb{R}^m) \to \mathbb{S}([0,T],\mathbb{R}), \quad (G,H) \mapsto \int_0^{(\cdot)} G(s)\mathrm{d}H(s), \tag{1}$$

where $\int_0^t G(s)\mathrm{d}H(s) = \sum_{i=1}^m \int_0^t G_i(s)\mathrm{d}H_i(s)$ (Protter, 2005, Chapter IV, Theorem 15).

We call a Borel-measurable function $f : [0,T] \times D([0,T],\mathbb{R}^m) \to \mathbb{R}^n$ *càdlàg* if $t \mapsto f(t,x) \in D([0,T],\mathbb{R}^n)$ for all $x \in D([0,T],\mathbb{R}^m)$, *adapted* if $f(t,x) = f(t,x^{\wedge t})$ for all $x \in D([0,T],\mathbb{R}^m)$ and $t \in [0,T]$, where $x^{\wedge t}(s) := x(s \wedge t)$. If the process X is càdlàg and adapted and the function f is measurable, càdlàg and adapted, then the processes $f(t,X)$ and $f(t-,X) := \lim_{s\uparrow t} f(s,X)$ are adapted, and the process $f(t-,X)$ is predictable (Przybyłowicz et al., 2024, Lemma 2.2).

Definition 1 (System of causal SDEs)**.** Given a filtered probability space $(\Omega,\mathcal{F},\mathbb{P})$ satisfying the usual conditions, a *system of causal stochastic differential equations (SDEs)* is a tuple $\mathcal{D} = (V,W,X_W,f,g,h)$ with V and W finite disjoint index sets, for each $w \in W$ a stochastic process $X_w \in \mathbb{S}([0,T],\mathbb{R})$ with the family $(X_w)_{w\in W}$ mutually independent, and for each $v \in V$ a *causal SDE*

$$X_v(t) = f_v(t,X_{\alpha(v)}) + \int_0^t g_v(s-,X_v,X_{\beta(v)})\mathrm{d}h_v(s,X_{\gamma(v)}) \quad \text{for all } t \in [0,T], \tag{2}$$

with given sets $\alpha(v) \subseteq V \cup W$, $\beta(v) \subseteq (V \cup W) \setminus \{v\}$ and $\gamma(v) \subseteq W$, and for some $m_v \in \mathbb{N} \cup \{0\}$,

[1]Throughout, we equip $D([0,T],\mathbb{R}^n)$ with the J_1 topology (Skorokhod, 1956) and the Borel σ-algebra, making it a standard Borel space. We make no distinction between $D([0,T],\mathbb{R})^n$ and $D([0,T],\mathbb{R}^n)$ since they are measurably isomorphic. The product topology on $D([0,T],\mathbb{R})^n$ is strictly weaker than the Skorokhod topology on $D([0,T],\mathbb{R}^n)$ (Jacod and Shiryaev, 2003, Section VI.1b), but this does not cause problems: the only continuity conditions are imposed in Assumptions 1 and 2 via Theorem 2, which are insensitive to this distinction.

[2]Note that $\mathbb{S}([0,T],\mathbb{R}^n) = \mathbb{S}([0,T],\mathbb{R})^n$.

measurable, càdlàg and adapted functions

$$\begin{aligned} f_v &: [0,T] \times D([0,T], \mathbb{R}^{|\alpha(v)|}) \to \mathbb{R} \\ g_v &: [0,T] \times D([0,T], \mathbb{R}^{|\beta(v)|+1}) \to \mathbb{R}^{m_v} \\ h_v &: [0,T] \times D([0,T], \mathbb{R}^{|\gamma(v)|}) \to \mathbb{R}^{m_v} \end{aligned}$$

such that $h_v(t, X_{\gamma(v)}) \in \mathbb{S}([0,T], \mathbb{R}^{m_v})$.

A *solution* (commonly referred to as a *strong solution*) of a system of SDEs $\mathcal{D}$ is an adapted process $X_V : \Omega \to D([0,T], \mathbb{R}^{|V|})$ such that for each $v \in V$ the equation (2) holds $\mathbb{P}$-a.s. The qualifier "causal" reflects that, like structural equations in SCMs (Pearl, 2009), each equation (2) encodes a causal mechanism rather than an algebraic constraint. Algebraically rewriting the system of SDEs might give the same solutions, but changes the causal graph and the effects of interventions (as defined below).

The function $f_v(t, X_{\alpha(v)})$ can be used to model the initial value of X_v and instantaneous functional dependencies. For the former, we say that f_v *models the initial condition* if $f_v(t, X_{\alpha(v)})$ is constant in t, so that it only describes the initial value $X_v^0 := f_v(0, X_{\alpha(v)}^{\wedge 0})$. In this case, equations of the form (2) are often suggestively written as

$$\mathrm{d}X_v(t) = g_v(t-, X_v, X_{\beta(v)})\mathrm{d}h_v(t, X_{\gamma(v)}) \qquad X_v(0) = X_v^0,$$

hence the name stochastic *differential* equation. Since h_v is typically not differentiable this has no other meaning than the integral equation. The processes $g_v(t-, X_v, X_{\beta(v)})$ and $h_v(t, X_{\gamma(v)})$ are called *integrands* and *integrators* respectively. We call g_v *Markov* if $g_v(t, X_v, X_{\beta(v)}) = g_v(t, X_v(t), X_{\beta(v)}(t))$, and *time-invariant* if $g_v(t, X_v, X_{\beta(v)}) = g_v(X_v(t), X_{\beta(v)}(t))$. If $X_{\gamma(v)}$ is a semimartingale and $h_v(t, X_{\gamma(v)}) = h_v(X_{\gamma(v)}(t))$ and the map $x \mapsto h_v(x)$ is twice continuously differentiable, then $h_v(t, X_{\gamma(v)})$ is also a semimartingale. If the process $h_v(t, X_{\gamma(v)})$ is deterministic, it is a semimartingale if and only if its path is of finite variation. If the integrator $h_v(t, X_{\gamma(v)})$ is $\mathbb{P}$-almost surely continuous, one obtains the same solutions without taking the left limit in the integrand.

Definition 1 generalises common definitions of SDEs, by allowing for functional relations via f_v. For specific choices of integrators and integrands, systems of causal SDEs model some special cases:

i) If $h_v(s, X_{\gamma(v)}) = s$, the 'stochastic integral' $\int g_v(s-, X_v, X_{\beta(v)})\mathrm{d}h_v(s, X_{\gamma(v)})$ reduces to the Riemann integral $\int g_v(s, X_v, X_{\beta(v)})\mathrm{d}s$, so Ordinary Differential Equations and Random Differential Equations are special cases of SDEs. If $h_v(s, X_{\gamma(v)})$ is of finite variation, the integral $\int g_v(s-, X_v, X_{\beta(v)})\mathrm{d}h_v(s, X_{\gamma(v)})$ reduces to the Stieltjes integral. If $g_v(s, X_v, X_{\beta(v)})$ and $h_v(s, X_{\gamma(v)})$ have no common discontinuities and finite p-variation and q-variation respectively with $p^{-1} + q^{-1} > 1$, the integral reduces to the Young integral (Lyons et al., 2007, Section 1.3).

ii) If f_v models the initial condition, g_v is Markov and $h_v(\cdot, X_{\gamma(v)})$ is a vector of independent Lévy processes for all $v \in V$, the solution X_V is temporally Markov[3] (see e.g. Protter, 2005, Chapter V, Theorem 32 for the case that g_v is Lipschitz). If additionally g_v is time-invariant, then the solution is temporally strong Markov.

iii) If f_v models the initial condition, g_v is Markov and $h_v(t, X_{\gamma(v)}) = (t, X_{\gamma(v)}(t))$ where $X_{\gamma(v)}$ is a Brownian motion, any solution X_v has continuous and possibly non-differentiable sample paths. Such an SDE is called an *Itô diffusion*.

[3]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 $s, t \in [0,T]$ such that $s + t \in [0,T]$; it is temporally *strong Markov* if this holds for any stopping time t.

iv) If f_v models the initial condition, g_v is Markov and $h_v(t, X_{\gamma(v)})$ is a jump process (e.g. a Poisson process), the solution X_v can have jumps as well. If $h_v(t, X_{\gamma_1(v)}, X_{\gamma_2(v)}) = (t, X_{\gamma_1(v)}(t), X_{\gamma_2(v)}(t))$ with $X_{\gamma_1(v)}$ and $X_{\gamma_2(v)}$ a Brownian motion and a jump process respectively, such an SDE is called a *jump-diffusion*.

Given a system of causal SDEs $\mathcal{D}$, we define its *causal graph* as follows:

Definition 2 (Causal graph)**.** Given a system of causal SDEs $\mathcal{D} = (V, W, X_W, f, g, h)$, the *augmented causal graph* is the directed graph $G^+(\mathcal{D}) = (V \cup W, E)$ where

$$\begin{aligned} E := &\{u \to v : v \in V, u \in \alpha(v) \text{ and } f_v \text{ is not constant in } X_u\} \\ &\cup \{u \to v : v \in V, u \in \{v\} \cup \beta(v) \text{ and } g_v \text{ is not constant in } X_u\} \\ &\cup \{u \to v : v \in V, u \in \gamma(v) \text{ and } h_v \text{ is not constant in } X_u\} \end{aligned}$$

The *causal graph* is the directed mixed graph $G(\mathcal{D}) = (V, E', L)$, where $L = \{i \leftrightarrow j : i \neq j \in V, (i \leftarrow k \to j) \in G^+(\mathcal{D}) \text{ for some } k \in W\}$ and E' is the restriction of E to V.

For a subset A of the vertices of a directed mixed graph G, we write $\mathrm{pa}(A) := \{u \notin A : u \to v \text{ in } G \text{ for some } v \in A\}$ for the *parents* of A (excluding A) and $\mathrm{Anc}(A) := A \cup \{u : u \to \cdots \to v \text{ in } G \text{ for some } v \in A\}$ for its *ancestors* (by definition including A), where u ranges over the vertices of G. Unless another graph is specified the ambient graph is $G(\mathcal{D})$, so in particular $\mathrm{pa}(A) \subseteq V$. We identify each vertex $v \in V$ with its associated process X_v. We sometimes abuse notation to write the labels of the random variables for the vertices in the graphs (as in Figure 1a) and for graphical separation statements (e.g. in Section 7). For the repressilator, the causal graph $G(\mathcal{D})$ is depicted in Figure 1a, where the noise processes and initial values are considered to be exogenous.

In line with Mooij et al. (2013), Hansen and Sokol (2014) and Peters et al. (2020), we define perfect interventions on systems of causal SDEs as follows.

Definition 3 (Perfect intervention)**.** Given a system of causal SDEs $\mathcal{D} = (V, W, X_W, f, g, h)$, intervention target $S \subseteq V$ and intervention value $x_S \in D([0,T], \mathbb{R})^{|S|}$, the *perfectly intervened system of causal SDEs* is defined as $\mathcal{D}_{\mathrm{do}(X_S = x_S)} = (V, W, X_W, f^\circ, g^\circ, h^\circ)$ with for each $v \in V$ the mechanisms $f_v^\circ(t, \cdot) := x_v(t)$, $g_v^\circ := 0$ and $h_v^\circ := 0$ if $v \in S$ and $f_v^\circ := f_v$, $g_v^\circ := g_v$ and $h_v^\circ := h_v$ otherwise, that is, we have for each $v \in V$ the *intervened causal SDE*

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

If one wants to model interventions on parameters that determine the dynamics of other variables, those parameters should be modelled as an endogenous process, which can be intervened upon.

Mathematically ‘intervening’ on a system of causal SDEs should be done with caution: it may not always reflect a realistic action. Suppose we can precisely determine the inflow and the outflow of atoms in the repressilator. Intervening on both inflow and outflow simultaneously could lead to physical contradictions (as the number of atoms is a conserved quantity in chemical reactions), so the mathematical intervention may have no realistic counterpart, and only interventions on a subset of the variables, for a subset of intervention values, may have a realistic interpretation.

2.1 Pathwise interpretation of systems of causal SDEs

The integrals in Definitions 1 and 3 are interpreted as Itô integrals as given by (1), defining each X_v as a random variable on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This ‘direct’ dependence on ω of all variables conceals an independence structure that is convenient for causal inference. The pathwise stochastic integral of Karandikar (1995) — improved upon by Przybyłowicz et al. (2024) — makes this independence structure explicit by replacing the ω-dependent stochastic integral with an explicit functional dependence between sample paths. In particular, Przybyłowicz et al. (2024) (Appendix A, Corollary A.1) provides an adapted measurable map

$$\Psi : D([0,T],\mathbb{R}^m) \times D([0,T],\mathbb{R}^m) \to D([0,T],\mathbb{R}),$$

with the property that for every càdlàg adapted process G and every semimartingale $H \in \mathbb{S}([0,T],\mathbb{R}^m)$, the process $\Psi(G,H)$ is a càdlàg version of the Itô integral $\int_0^{\cdot} G(s-)\mathrm{d}H(s)$. This allows for a pathwise expression of the SDE, where all endogenous variables only depend on ω though the exogenous variables, and not directly.

Definition 4 (Pathwise interpretation of systems of causal SDEs)**.** Given a system of causal SDEs $\mathcal{D}$, for each variable $v \in V$ the SDE for X_v can be represented by the equation

$$X_v = \Phi_v(X_v, X_{\alpha(v)}, X_{\beta(v)}, X_{\gamma(v)}) \quad \mathbb{P}\text{-a.s.}$$

where for all $t \in [0,T]$,

$$\Phi_v(x_v, x_{\alpha(v)}, x_{\beta(v)}, x_{\gamma(v)})(t) := f_v\big(t, x_{\alpha(v)}\big) + \Psi\big(g_v(\cdot, x_v, x_{\beta(v)}), h_v(\cdot, x_{\gamma(v)})\big)(t).$$

For $O \subseteq V, x_V \in D([0,T],\mathbb{R})^{|V|}$ and $x_W \in D([0,T],\mathbb{R})^{|W|}$, write

$$\Phi_O(x_V, x_W) := (\Phi_v(x_v, x_{\alpha(v)}, x_{\beta(v)}, x_{\gamma(v)}))_{v\in O}.$$

Although not strictly necessary for our purposes, this functional interpretation may guide conceptual understanding of how interventions propagate through the system, and it provides a notation Φ_v for each SDE that is useful in the proofs. The pathwise SDE formulation also clarifies why the processes $X_{\alpha(v)}$ and $X_{\beta(v)}$ are allowed to be endogenous (and hence intervenable), and the integrator processes $X_{\gamma(v)}$ are required to be exogenous (and hence non-intervenable). Namely, the map Ψ correctly represents the stochastic integral for all adapted measurable càdlàg integrand processes (so in particular also for all deterministic càdlàg values of the integrand), but it requires the integrator to be a semimartingale, and not every deterministic càdlàg sample path is a semimartingale – this only holds when the path is of finite variation.

Under this pathwise reading, a system of causal SDEs is itself a (possibly cyclic) *Structural Causal Model* (SCM) (Pearl, 2009; Bongers et al., 2021) on path space, i.e. a tuple $(V, W, \mathcal{X}_V, \mathcal{X}_W, \Phi, \mathbb{P}(X_W))$ where V and W are disjoint finite index sets of endogenous variables and exogenous variables respectively, the domains $\mathcal{X}_V = \prod_{i\in V} D([0,T],\mathbb{R})$ and $\mathcal{X}_W = \prod_{i\in W} D([0,T],\mathbb{R})$ are products of Skorokhod spaces, the exogenous distribution $\mathbb{P}(X_W) = \bigotimes_{w\in W} \mathbb{P}(X_w)$ is the product of the laws of the exogenous processes, and for each $v \in V$ the *causal mechanism* $\Phi_v : \mathcal{X}_V \times \mathcal{X}_W \to \mathcal{X}_v$ yields the *structural equation*

$$X_v = \Phi_v(X_V, X_W).$$

Such an SCM on path space is also known as a *Dynamic Structural Causal Model* (Rubenstein et al., 2018; Boeken and Mooij, 2024). The causal graph $G(\mathcal{D})$ of Definition 2 is a supergraph of the graph of the SCM as defined in Bongers et al. (2021), and the perfect intervention on the system of causal SDEs coincides with the corresponding perfect intervention on the SCM. The remainder of the paper leverages this pathwise SCM interpretation to prove various results that exist for SCMs (Markov properties, the do-calculus, constraint-based causal discovery) for systems of causal SDEs.

3 Solvability of systems of causal SDEs

The primary objects of causal inference are the observational and interventional distributions. If $\mathcal{D}$ has a solution X_V then the *observational distribution* is the law $\mathbb{P}(X_V)$, and if for $S \subseteq V$ and $x_S \in \mathcal{X}_S$ (write $O := V \setminus S$) the system $\mathcal{D}_{\mathrm{do}(X_S = x_S)}$ has solution X_O then the *interventional distribution* is the law of X_O, denoted by $\mathbb{P}(X_O \,|\, \mathrm{do}(X_S = x_S))$. However, these distributions are not always well-defined. To resolve this, Bongers et al. (2021) introduced for (possibly cyclic) SCMs solvability requirements such that the observational distribution and all interventional distributions are well-defined. However, these solvability requirements are too strong for our purposes, and not necessary to derive useful results. Instead, we consider the following solvability property of systems of causal SDEs.

Definition 5 (Essential unique solvability)**.** Let $O \subseteq V$ and write $S := V \setminus O$. The system of causal SDEs $\mathcal{D}$ is *essentially uniquely solvable w.r.t. O* if there exists a measurable adapted function $I^{[O]} : \mathcal{X}_S \times \mathcal{X}_W \to \mathcal{X}_O$ such that for every measurable adapted function $\varphi_S : \mathcal{X}_W \to \mathcal{X}_S$ the following two conditions hold:

- $I^{[O]}$ is a.s. a fixed point of the system of causal SDEs: we have $\mathbb{P}(X_W)$-almost surely

$$I^{[O]}(\varphi_S(X_W), X_W) = \Phi_O\big(I^{[O]}(\varphi_S(X_W), X_W), \varphi_S(X_W), X_W\big); \tag{3}$$

- $I^{[O]}$ is an a.s. unique fixed point, i.e., for every measurable adapted function $\varphi_O : \mathcal{X}_W \to \mathcal{X}_O$ satisfying $\varphi_O(X_W) = \Phi_O\big(\varphi_O(X_W), \varphi_S(X_W), X_W\big)$ $\mathbb{P}(X_W)$-a.s., we have $\mathbb{P}(X_W)$-a.s.

$$\varphi_O(X_W) = I^{[O]}(\varphi_S(X_W), X_W).$$

We refer to such a function $I^{[O]}$ as a *solution function (w.r.t. O)*.

A solution function may be constant on a large part of its domain: whenever $\mathcal{D}$ is essentially uniquely solvable w.r.t. O, one can choose a solution function $I^{[O]}(x_{\mathrm{pa}(O)}, x_W)$ that depends on X_S (with $S := V \setminus O$) only through the endogenous parents $\mathrm{pa}(O)$ of O (Lemma B.2).

This notion of essential unique solvability differs from the notion of Bongers et al. (2021). We require existence of the fixed point outside of a null set that may depend on the input process φ_S, and our uniqueness condition holds outside of a null set that may depend on the alternative solution φ_O. In Bongers et al. (2021), existence and uniqueness must hold simultaneously on a single, uniform null set, that only depends on Φ_O and not on φ_S and φ_O. In this regard our notion of solvability is weaker, and this weakening is suitable for obtaining solution functions for a certain class of systems of causal SDEs, as shown in Section 3.1 (Theorem 2).[4] Due to this weakening, we cannot rely on existing results for simple SCMs (uniqueness of interventional distributions, Markov property, do-calculus) but we prove these results from scratch.

First, we indeed have that essential unique solvability allows for clear expressions of the observational and interventional distributions.

Theorem 1. *Let $\mathcal{D}$ be a system of causal SDEs. If $\mathcal{D}$ is essentially uniquely solvable w.r.t. V with solution function $I^{[V]} : \mathcal{X}_W \to \mathcal{X}_V$, then the observational distribution of $\mathcal{D}$ is unique and satisfies*

$$\mathbb{P}(X_V) = \mathbb{P}(I^{[V]}(X_W)).$$

If $L, O, S \subseteq V$ partition V and $\mathcal{D}$ is essentially uniquely solvable w.r.t. $O \subseteq V$ with solution function $I^{[O]} : \mathcal{X}_S \times \mathcal{X}_L \times \mathcal{X}_W \to \mathcal{X}_O$, then for any $x_S \in \mathcal{X}_S$ the intervened system $\mathcal{D}_{\mathrm{do}(X_S = x_S)}$ is

[4] Note that Boeken and Mooij (2024) overlooked this nuance and assumed that the SDEs satisfy the solvability requirements of Bongers et al. (2021).

essentially uniquely solvable w.r.t. O with solution function $I^{[O]}(x_S, \cdot, \cdot) : \mathcal{X}_L \times \mathcal{X}_W \to \mathcal{X}_O$. If $L = \emptyset$, the interventional distribution satisfies

$$\mathbb{P}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = \mathbb{P}(I^{[O]}(x_S, X_W)),$$

is unique, and $x_S \mapsto \mathbb{P}(X_O \,|\, \mathrm{do}(X_S = x_S))$ is a Markov kernel.

Example 3. *Consider a linearised version of the repressilator with dynamical model*

$$\begin{aligned} \mathrm{d}M_i(t) &= (\gamma - M_i(t) - \lambda P_j(t))\,\mathrm{d}t + \sigma \mathrm{d}Z_i^M(t) \\ \mathrm{d}P_i(t) &= \beta(M_i(t) - P_i(t))\mathrm{d}t + \sigma \mathrm{d}Z_i^P(t) \end{aligned}$$

for each gene and its predecessor $(i, j) \in \{$(lacI, cI), (tetR, lacI), (cI, tetR), (GFP, tetR)$\}$. This system is for example essentially uniquely solvable with respect to M_i for each i, with solution function

$$I^{[M_i]}(M_i^0, P_j, Z_i^M)(t) := M_i^0 e^{-t} + \int_0^t e^{-(t-s)}\big(\gamma - \lambda P_j(s)\big)\mathrm{d}s + \sigma \int_0^t e^{-(t-s)} \mathrm{d}Z_i^M(s),$$

with the integrals interpreted as pathwise maps Ψ. This solution function determines the causal effect of the endogenous parent variable P_j on M_i: for any intervention path $x_{P_j} \in \mathcal{X}_{P_j}$ the interventional distribution is written as $\mathbb{P}\big(X_{M_i} \,|\, \mathrm{do}(X_{P_j} = x_{P_j})\big) = \mathbb{P}(I^{[M_i]}(M_i^0, x_{P_j}, Z_i^M))$.

3.1 Sufficient conditions for solvability

We provide two classes of systems of causal SDEs that are essentially uniquely solvable with respect to every $O \subseteq V$. Given a measurable map $g : [0,T] \times D([0,T], \mathbb{R}^k) \times D([0,T], \mathbb{R}^\ell) \to D([0,T], \mathbb{R}^m)$ consider the conditions that there exists a measurable càdlàg $K : [0,T] \times D([0,T], \mathbb{R}^\ell) \to (0, \infty)$ such that for all $x, x_1, x_2 \in D([0,T], \mathbb{R}^k)$, $y \in D([0,T], \mathbb{R}^\ell)$, $t \in [0,T]$,

$$\|g(t, x, y)\| \leq K(t, y)\big(1 + \sup_{0 \leq s \leq t} \|x(s)\|\big), \tag{4}$$

$$\|g(t, x_1, y) - g(t, x_2, y)\| \leq K(t, y) \sup_{0 \leq s \leq t} \|x_1(s) - x_2(s)\|. \tag{5}$$

We refer to (4) and (5) respectively as the *linear-growth* and *Lipschitz* conditions on g in x given y. These conditions can be verified component-wise, as shown in Lemma B.1 in the appendix. Under these conditions, Przybyłowicz et al. (2024) (see also Appendix A, Theorem A.1) prove the existence of a measurable, adapted solution function.[5] Translated to systems of causal SDEs, their result implies the following.

Theorem 2. *Let $\mathcal{D}$ be a system of causal SDEs and let $O \subseteq V$. If for every $v \in O$ we have $\alpha(v) \cap O = \emptyset$ and the integrand g_v satisfies the linear-growth and Lipschitz conditions in $(X_v, X_{\beta(v) \cap O})$ given $X_{\beta(v) \setminus O}$, then $\mathcal{D}$ is essentially uniquely solvable w.r.t. O.*

For simplicity, this sufficient condition assumes that there are no functional relations between the variables in O. To obtain essential unique solvability, we apply Theorem 2 separately to each *strongly connected component* (SCC) of $G(\mathcal{D})$ (subsets whose every two vertices are connected by a directed path) and compose the resulting solution functions along the topological order of SCCs; essential unique solvability w.r.t. each SCC of $G(\mathcal{D})$ then lifts to essential unique solvability w.r.t. V (Lemma B.3 in the appendix). However, this does not imply essential unique solvability with respect to every subset $O' \subseteq V$. The following assumption allows for this construction.

[5]For SDEs driven by Brownian motion, similar results have been shown by Yamada and Watanabe (1971) and Kallenberg (1996), and for general semimartingale SDEs this has been shown by Karandikar (1995), whose solution function was proven to be measurable with respect to Borel sets generated by a (non-Polish) topology that is strictly stronger than the Skorokhod topology.

Assumption 1. *Consider a system of causal SDEs $\mathcal{D} = (V, W, X_W, f, g, h)$ such that*

(i) for every SCC S of $G(\mathcal{D})$ and every $v \in S$, $\alpha(v) \cap S = \emptyset$, and

(ii) for every $v \in V$, writing S_v for the SCC of $G(\mathcal{D})$ containing v, g_v satisfies the linear-growth and Lipschitz conditions in $(X_v, X_{\beta(v)\cap S_v})$ given $X_{\beta(v)\setminus S_v}$.

This assumption is closed under perfect interventions, so we obtain the following result.

Theorem 3. *If a system of causal SDEs $\mathcal{D}$ satisfies Assumption 1, then it is essentially uniquely solvable w.r.t. every subset $O \subseteq V$.*

For the d-separation Markov property (in Section 5.2 below) we will need a more restricted model class:

Assumption 2. *Consider the additive-noise causal SDE*

$$X_V(t) = X_\alpha^0 + \int_0^t \mu(s, X_V(s))\mathrm{d}s + \int_0^t \Sigma \mathrm{d}X_\gamma(s), \tag{6}$$

for the endogenous index set V with $|V| =: d$ and exogenous index sets $W = \alpha \cup \gamma$ with α, γ disjoint such that $|\alpha| = |\gamma| = d$, with $v \mapsto \alpha(v)$ a bijection of V onto α, where X_α^0 is a vector of mutually independent exogenous random variables satisfying $\mathbb{E}[\|X_\alpha^0\|^2] < \infty$, and X_γ is a d-dimensional standard Brownian motion with independent components, independent of X_α^0. The drift $\mu : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ has $t \mapsto \mu(t, x_V(t))$ continuous in t and satisfies the linear-growth and Lipschitz conditions in x_V, and $\Sigma \in \mathbb{R}^{d\times d}$ is an invertible matrix.

Theorem 4. *If a system of causal SDEs $\mathcal{D}$ satisfies Assumption 2, then it satisfies Assumption 1.*

We now show that the repressilator satisfies Assumption 2.

Example 4 (Repressilator is essentially uniquely solvable)**.** *We may write the repressilator in vector form as*

$$\mathrm{d}X(t) = \mu(X(t))\mathrm{d}t + \sigma \mathrm{d}Z(t)$$

with drift

$$\mu(X) = \begin{pmatrix} \frac{\alpha}{1+|P_{cI}|^n} + \alpha_0 - M_{lacI} \\ \beta(M_{lacI} - P_{lacI}) \\ \frac{\alpha}{1+|P_{lacI}|^n} + \alpha_0 - M_{tetR} \\ \beta(M_{tetR} - P_{tetR}) \\ \frac{\alpha}{1+|P_{tetR}|^n} + \alpha_0 - M_{cI} \\ \beta(M_{cI} - P_{cI}) \\ \frac{\alpha}{1+|P_{tetR}|^n} + \alpha_0 - M_{GFP} \\ \beta(M_{GFP} - P_{GFP}) \end{pmatrix}, \quad X = \begin{pmatrix} M_{lacI} \\ P_{lacI} \\ M_{tetR} \\ P_{tetR} \\ M_{cI} \\ P_{cI} \\ M_{GFP} \\ P_{GFP} \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} Z_{lacI}^M \\ Z_{lacI}^P \\ Z_{tetR}^M \\ Z_{tetR}^P \\ Z_{cI}^M \\ Z_{cI}^P \\ Z_{GFP}^M \\ Z_{GFP}^P \end{pmatrix}.$$

We verify the linear-growth and Lipschitz conditions on μ componentwise. For each M_i component (i.e. rows 1, 3, 5, 7) we have $\left|\frac{\alpha}{1+|P_j(t)|^n} + \alpha_0 - M_i\right| \leq \max\{\alpha + \alpha_0, 1\}(1 + \|(P_j(t), M_i(t))\|)$, so it satisfies the linear-growth condition in (P_j, M_i), and the function $p \mapsto \frac{\alpha}{1+|p|^n}$ has Lipschitz constant upper bounded by αn, so one readily verifies that the component is $\sqrt{2}\max\{\alpha n, 1\}$-Lipschitz in (P_j, M_i). Each P_i component is linear in (M_i, P_i) and so satisfies both the linear-growth condition and the Lipschitz condition in (M_i, P_i). Hence μ satisfies the linear-growth and Lipschitz conditions jointly. Moreover, the diffusion matrix $\Sigma = \sigma I_8$ is invertible, with inverse $\sigma^{-1} I_8$. If the initial condition X^0 is a vector of mutually independent variables with $\mathbb{E}[\|X^0\|^2] < \infty$, independent of the driving Brownian motion Z, then the repressilator satisfies Assumption 2, and by Theorems 3 and 4 it is essentially uniquely solvable w.r.t. every $O \subseteq V$.

4 Marginalisation

When a subset $L \subseteq V$ of variables is not of direct interest or unobserved, one may want to reason about $\mathcal{D}$ on $O := V \setminus L$. The marginalisation of a system of causal SDEs substitutes a solution function for L into the SDE coefficients for $V \setminus L$.

Definition 6 (Marginalisation)**.** Let $L \subseteq V$ and let $\mathcal{D} = (V, W, X_W, f, g, h)$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. $L \subseteq V$ with solution function $I^{[L]} : \mathcal{X}_O \times \mathcal{X}_W \to \mathcal{X}_L$, writing $O := V \setminus L$. The *marginalised system of causal SDEs* $\mathcal{D}_{\mathrm{marg}(L)} := \left(O, W, X_W, \tilde{f}, \tilde{g}, \tilde{h}\right)$ has for each $v \in O$ an SDE of the form (2) given by the following parameters:

$$\begin{aligned}
\tilde{f}_v(t, x_O, x_W) &:= f_v\big(t, x_{\alpha(v)\cap O}, I^{[L]}_{\alpha(v)\cap L}(x_O, x_W), x_{\alpha(v)\cap W}\big),\\
\tilde{g}_v(s-, x_v, x_O, x_W) &:= g_v\big(s-, x_v, x_{\beta(v)\cap O}, I^{[L]}_{\beta(v)\cap L}(x_O, x_W), x_{\beta(v)\cap W}\big),\\
\tilde{h}_v(s, x_W) &:= h_v\big(s, x_{\gamma(v)}\big).
\end{aligned}$$

Since adapted solution functions $I^{[L]}$ are not unique, the marginalised system $\mathcal{D}_{\mathrm{marg}(L)}$ is itself not unique. The induced distributions are nonetheless unambiguous: whenever $\mathcal{D}$ is essentially uniquely solvable w.r.t. L and V, the observational distributions on $O := V \setminus L$ coincide, $\mathbb{P}_{\mathcal{D}}(X_O) = \mathbb{P}_{\mathcal{D}_{\mathrm{marg}(L)}}(X_O)$; and more generally, for $O' \subseteq O$ with $S := O \setminus O'$, whenever $\mathcal{D}$ is essentially uniquely solvable w.r.t. L and $L \cup O'$, the interventional distributions coincide, $\mathbb{P}_{\mathcal{D}}(X_{O'} \,|\, \mathrm{do}(X_S)) = \mathbb{P}_{\mathcal{D}_{\mathrm{marg}(L)}}(X_{O'} \,|\, \mathrm{do}(X_S))$ (Lemma C.1). Marginalisation is therefore a powerful notion of abstraction: (causal) inference on the observed part of a system remains valid in the underlying system. It is moreover compatible with itself and with intervention: sequentially marginalising over two disjoint subsets $L_1, L_2 \subseteq V$ yields the same observational and interventional distributions regardless of the order (Lemma C.2), and marginalisation commutes with intervention (Lemma C.3). These results rest on the following closure property: marginalising over L preserves essential unique solvability, in that solvability of $\mathcal{D}$ w.r.t. L and $L \cup O'$ transfers to solvability of $\mathcal{D}_{\mathrm{marg}(L)}$ w.r.t. O'.

Theorem 5. *Let $\mathcal{D}$ be a system of causal SDEs, and let $L, O \subseteq V$ be disjoint. If $\mathcal{D}$ is essentially uniquely solvable w.r.t. L and w.r.t. $L \cup O$, then the marginalised system $\mathcal{D}_{\mathrm{marg}(L)}$ is essentially uniquely solvable w.r.t. O. Moreover, for any solution function $I^{[L\cup O]}$ of $\mathcal{D}$ w.r.t. $L \cup O$, its restriction $I^{[L\cup O]}_O$ is a solution function of $\mathcal{D}_{\mathrm{marg}(L)}$ w.r.t. O.*

Any system of causal SDEs satisfying Assumption 1 or Assumption 2 is essentially uniquely solvable w.r.t. every $O \subseteq V$ (Theorems 3 and 4). By Theorem 5, any such system therefore remains essentially uniquely solvable w.r.t. every set of endogenous variables after any marginalisation.

5 Graphical Markov properties

A cornerstone of reasoning with causal graphs is the Markov property: that a d-separation or σ-separation in the graph implies a conditional independence in the distribution. The Markov property is the foundation of the do-calculus and constraint-based causal discovery algorithms, as will be treated in Sections 6 and 8.1.

Formally, a walk π in a DMG $G = (V, E, L)$, is *d-blocked* by $C \subseteq V$ if it has an end-point or a non-collider in C, or if there is a collider which is not in $\mathrm{Anc}(C)$. For sets of nodes $A, B, C \subseteq V$, we call A and B *d-separated* given C, written $A \perp^d_G B \,|\, C$, if all walks between A and B are d-blocked by C. The d-separation Markov property then means that for all $A, B, C \subseteq V$, we have the implication

$$A \underset{G}{\overset{d}{\perp}} B \,|\, C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C.$$

This is known to hold for acyclic SCMs, as well as for certain cyclic SCMs: if all variables are discrete and the SCM is ancestrally uniquely solvable (Pearl and Dechter, 1996; Neal, 2000; Forré and Mooij, 2017), or if the structural equations are linear and depend on an exogenous variable whose distribution has a density with respect to Lebesgue measure (Spirtes, 1994; Forré and Mooij, 2017), see also Bongers et al. (2021). In the next section, we will see that a class of cyclic additive-noise SDEs can be added to this list.

It is known that there are cyclic SCMs for which the d-separation Markov property does not hold, for example for certain nonlinear cyclic SCMs with Gaussian noise (Spirtes, 1994, 1995). A more suitable separation criterion for cyclic SCMs is the following notion of *σ-separation* (Forré and Mooij, 2017; Bongers et al., 2021; Forré and Mooij, 2025). Given a DMG $G = (V, E)$, a set of nodes $C \subseteq V$ and a walk π in DMG G:

- a non-collider v is called *blockable* if it points towards a neighbouring node on the walk that is not in the same strongly connected component as v,
- the walk π is called *σ-blocked by C* if it has an end-point or a blockable non-collider in C, or if there is a collider on π that is not in $\mathrm{Anc}(C)$.

For sets of nodes $A, B, C \subseteq V$, we call A and B *σ-separated* given C, written $A \perp_G^\sigma B \mid C$, if all walks between A and B are σ-blocked by C. For general DMGs, σ-separation implies d-separation, and for acyclic DMGs the two notions coincide (since every non-collider is automatically blockable). Self-loops $v \to v$ do not affect d- or σ-separations (Theorem D.1); we may therefore ignore them when reading off separations.

5.1 A σ-separation Markov property for systems of causal SDEs

The σ-separation Markov property was established for simple SCMs by Forré and Mooij (2017); Bongers et al. (2021); Forré and Mooij (2025). Here we extend this result to systems of causal SDEs that are essentially uniquely solvable w.r.t. every SCC of $G(\mathcal{D})$, following the same proof strategy: we construct an *acyclification* by replacing each SCC's SDE with its solution function, show that this acyclified system is observationally equivalent to the original and has a graph that is a subgraph of any graphical acyclification, and then invoke the d-separation Markov property for acyclic SCMs (Bongers et al., 2021, Theorem 6.3). The full proof is given in the appendix.

Theorem 6. *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every SCC of $G(\mathcal{D})$. For all $A, B, C \subseteq V \cup W$ we have*

$$A \underset{G^+(\mathcal{D})}{\overset{\sigma}{\perp}} B \mid C \implies X_A \perp\!\!\!\perp X_B \mid X_C. \tag{7}$$

For sets $A, B, C \subseteq V$ restricted to endogenous variables, this implies the σ-separation Markov property in the (non-augmented) causal graph $G(\mathcal{D})$, via the correspondence between σ-separation in $G(\mathcal{D})$ and $G^+(\mathcal{D})$. We call $\mathcal{D}$ *σ-faithful* if the reverse implication of (7) holds; this need not hold in general.

Example 5 (Graphical separation in the repressilator)**.** *Consider the cyclic graph in Figure 1a. In this graph we have the d-separation $P_{lacI} \perp^d P_{tetR} \mid M_{tetR}, M_{lacI}$, but a σ-connection $P_{lacI} \not\perp^\sigma P_{tetR} \mid M_{tetR}, M_{lacI}$: on the walk $P_{lacI} \to M_{tetR} \to P_{tetR}$ the non-collider M_{tetR} points to P_{tetR}, which lies in the same strongly connected component $\{M_{lacI}, P_{lacI}, M_{tetR}, P_{tetR}, M_{cI}, P_{cI}\}$, so M_{tetR} is not blockable and conditioning on it does not block the walk. Similarly, we have a σ-connection $P_{lacI} \not\perp^\sigma P_{GFP} \mid M_{tetR}, M_{lacI}$, and we have a σ-separation $P_{lacI} \perp^\sigma P_{GFP} \mid P_{tetR}$.*

For essentially uniquely solvable cyclic systems of causal SDEs the σ-separation Markov property holds (Theorem 6), but the question remains: can this be strengthened to a d-separation Markov property? Do we have a conditional independence $P_{lacI} \perp\!\!\!\perp P_{tetR} \mid M_{tetR}, M_{lacI}$?

5.2 A d-separation Markov property for additive-noise SDEs

In this section we will show that the d-separation Markov property holds for (possibly cyclic) additive noise SDEs satisfying Assumption 2. Our proof technique relies on a time-discretised process for which the d-separation Markov property holds by acyclicity. By showing that the law of the discretised process – suitably embedded in the Skorokhod space $D([0,T],\mathbb{R}^d)$ – converges in total variation to the law of the original process as a probability measure on $D([0,T],\mathbb{R}^d)$, conditional independence in the time-discretised process implies a conditional independence in the solution of the original process.

We consider the following discrete-time Euler-Maruyama-type scheme, with initial value $X_V^\Delta(0) = X_\alpha^0$ and

$$X_V^\Delta(t_{k+1}) = X_V^\Delta(t_k) + \mu(t_k, X_V^\Delta(t_k))(t_{k+1}-t_k) + \Sigma\big(X_\gamma(t_{k+1}) - X_\gamma(t_k)\big)$$

with $t_k := Tk/n$, $n \in \mathbb{N}$, $k = 0, ..., n$ and independent increment processes $X_\gamma^{[t_k,t_{k+1}]} := X_\gamma - X_\gamma(t_k)$ considered as a random variable in $D([t_k,t_{k+1}],\mathbb{R}^d)$. This induces a discrete-time *Euler SCM* $\mathcal{M}_\mathcal{D}^\Delta$ with endogenous variables $X_v^\Delta(t_0), ..., X_v^\Delta(t_n)$ for all $v \in V$ and exogenous independent variables X_w^0 for all $w \in \alpha$ and $X_w^{[0,t_1]}, ..., X_w^{[t_{n-1},t_n]}$ for all $w \in \gamma$ (Hansen and Sokol, 2014). We refer to the graph $G(\mathcal{M}_\mathcal{D}^\Delta)$ of $\mathcal{M}_\mathcal{D}^\Delta$ as the *Euler graph*, see Example 6 below. In contrast, we call the causal graph $G(\mathcal{D})$ the *summary graph*, since each of its vertices represents an entire process X_v on $[0,T]$ rather than the individual time-points $X_v^\Delta(t_k)$ as displayed by the Euler graph (or the subinterval processes of the time-split graph in Section 7). An important observation is that the Euler graph is acyclic, so the Euler SCM satisfies the d-separation Markov property.

The following lemma relates the causal graph to the Euler graph. It is an adaptation of Ferreira and Assaad (2024), who consider discrete-time stochastic processes with a corresponding summary graph.

Lemma 7. Let $\mathcal{D}$ be the SDE from Assumption 2 and let $\mathcal{M}_\mathcal{D}^\Delta$ be the Euler SCM (for any $n \in \mathbb{N}$), then for $A, B, C \subseteq V$ we have

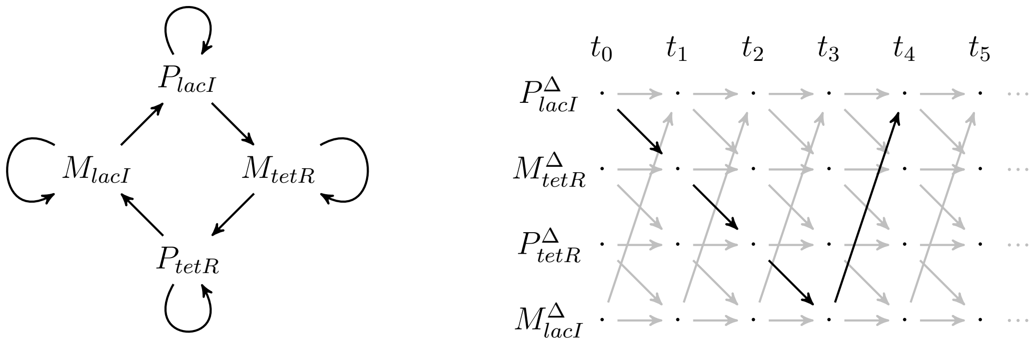
$$X_A \overset{d}{\underset{G(\mathcal{D})}{\perp}} X_B \,|\, X_C \implies X_A^\Delta \overset{d}{\underset{G(\mathcal{M}_\mathcal{D}^\Delta)}{\perp}} X_B^\Delta \,|\, X_C^\Delta \implies X_A^\Delta \perp\!\!\!\perp X_B^\Delta \,|\, X_C^\Delta$$

where we write $X_A^\Delta := (X_A^\Delta(t_0), ..., X_A^\Delta(t_n))$.

Example 6 (Euler graph of repressilator)**.** *Considering only the variables P_{lacI}, M_{tetR}, P_{tetR} and M_{lacI} of the repressilator, the causal graph of the marginalised SDE $\mathcal{D}' := \mathcal{D}_{\mathrm{marg}(M_{cI},P_{cI},M_{GFP},P_{GFP})}$ is given in Figure 3a, from which we read off the d-separation $P_{lacI} \perp^d P_{tetR} \,|\, M_{tetR}, M_{lacI}$. The Euler graph is depicted in Figure 3b, where we observe that $P_{lacI}^\Delta \perp^d P_{tetR}^\Delta \,|\, M_{tetR}^\Delta, M_{lacI}^\Delta$. For example, the displayed paths from $P_{lacI}^\Delta(t_0)$ to $P_{tetR}^\Delta(t_2)$ and from $P_{tetR}^\Delta(t_2)$ to $P_{lacI}^\Delta(t_4)$ are blocked by $M_{tetR}^\Delta(t_1)$ and $M_{lacI}^\Delta(t_3)$ respectively. Since this discrete-time SCM is acyclic, we have $P_{lacI}^\Delta \perp\!\!\!\perp P_{tetR}^\Delta \,|\, M_{tetR}^\Delta, M_{lacI}^\Delta$.*

It follows from Lemma 7 that the d-separation Markov property holds for any discrete-time SCM which does not have instantaneous cycles, similar to results shown by Ferreira and Assaad (2024), Niemiro (2024) and Reiter et al. (2024).

To obtain a d-separation Markov property for the SDE, we aim to deduce the conditional independence $X_A \perp\!\!\!\perp X_B \,|\, X_C$ from the conditional independence $X_A^\Delta \perp\!\!\!\perp X_B^\Delta \,|\, X_C^\Delta$. To do so, we first extend the discrete-time process X_V^Δ to a continuous-time process X_V^n, for which we show that $X_A^n \perp\!\!\!\perp X_B^n \,|\, X_C^n$ holds as well. Finally, we show that the distribution $\mathbb{P}(X_V^n)$ converges in



(a) The causal graph $G(\mathcal{D}')$.

(b) The Euler graph $G(\mathcal{M}^{\Delta}_{\mathcal{D}'})$.

Figure 3: The marginalisation of the repressilator onto $\{P_{lacI}, M_{tetR}, P_{tetR}, M_{lacI}\}$, showcasing that $P_{lacI} \perp^{d} P_{tetR} \,|\, M_{tetR}, M_{lacI}$ implies $P^{\Delta}_{lacI} \perp^{d} P^{\Delta}_{tetR} \,|\, M^{\Delta}_{tetR}, M^{\Delta}_{lacI}$.

total variation to $\mathbb{P}(X_V)$, which by Lauritzen (2024) implies the desired conditional independence $X_A \perp\!\!\!\perp X_B \,|\, X_C$.

Obtaining total variation convergence of the law of the Euler scheme to the law of the solution of the SDE is not straightforward. For example, if we extend X^{Δ}_V to continuous time by setting $X^n(t) := X^{\Delta}(t_k)$ for $t \in [t_k, t_{k+1})$, then $\mathbb{P}(X^n_V)$ will never converge in total variation to the law of the SDE since the set of continuous paths is a Borel measurable set in the Skorokhod space, giving total variation distance $d_{TV}(X^{\Delta}_V, X^n_V) = 1$. To mitigate this problem we consider for $t \in [t_k, t_{k+1})$ the continuously interpolated Euler scheme

$$X^n_V(t) = X^n_V(t_k) + \mu(t_k, X^n_V(t_k))(t - t_k) + \Sigma\big(X_\gamma(t) - X_\gamma(t_k)\big). \tag{8}$$

This Euler approximation has the same noise structure as the original SDE, which ensures that the processes both have a density with respect to a (transformed) Wiener measure on $D([0,T], \mathbb{R}^d)$ via Girsanov's theorem. Convergence of the density of $\mathbb{P}(X^n_V)$ to the density of $\mathbb{P}(X_V)$ finally implies the desired total variation convergence, via Scheffé's Theorem (Scheffé, 1947).

Having introduced a suitable continuous-time extension of the Euler scheme, we show that it inherits the conditional independencies implied by d-separation in the causal graph $G(\mathcal{D})$. The proof is in the appendix, but relies on the insight that one can write X^n_V as a function of X^{Δ}_V and a vector $\tilde{X}_V$ of Brownian bridges, with $X^{\Delta}_V \perp\!\!\!\perp \tilde{X}_V$ (Lemma D.2). By Lemma 7 and the d-separation Markov property we have $X^{\Delta}_A \perp\!\!\!\perp X^{\Delta}_B \,|\, X^{\Delta}_C$, and by Lemma D.3 we have $\tilde{X}_A \perp\!\!\!\perp \tilde{X}_B \,|\, \tilde{X}_C$. Combining these conditional independencies then gives the following result:

Theorem 8. *For all $A, B, C \subseteq V$, the continuous Euler scheme satisfies*

$$X_A \underset{G(\mathcal{D})}{\overset{d}{\perp}} X_B \,|\, X_C \implies X^n_A \perp\!\!\!\perp X^n_B \,|\, X^n_C.$$

The following result gives the desired total variation convergence of the Euler scheme.

Theorem 9. *Let $\mathcal{D}$ satisfy Assumption 2, then there is a subsequence n_m such that $\mathbb{P}(X^{n_m}_V)$ converges in total variation to $\mathbb{P}(X_V)$.*

Finally, the conditional independence $X^n_A \perp\!\!\!\perp X^n_B \,|\, X^n_C$ implies the desired conditional independence $X_A \perp\!\!\!\perp X_B \,|\, X_C$ by the following result.

Theorem 10 (Lauritzen (2024))**.** *Given a sequence of random variables $(X^n_A, X^n_B, X^n_C)_{n \in \mathbb{N}}$ such that $X^n_A \perp\!\!\!\perp X^n_B \,|\, X^n_C$ for all $n \in \mathbb{N}$ and given a random variable (X_A, X_B, X_C) such that $\mathbb{P}(X^n_A, X^n_B, X^n_C)$ converges in total variation to $\mathbb{P}(X_A, X_B, X_C)$, we have $X_A \perp\!\!\!\perp X_B \,|\, X_C$.*

We now state the main result of this section, that the additive-noise SDE satisfies the d-separation Markov property.

Theorem 11. *Let the system of causal SDEs $\mathcal{D}$ satisfy Assumption 2. Then for all $A, B, C \subseteq V$,*

$$X_A \underset{G(\mathcal{D})}{\overset{d}{\perp}} X_B \,|\, X_C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C.$$

Proof. Let $A \perp^d_{G(\mathcal{D})} B \,|\, C$, then by Theorem 8 the continuous Euler scheme satisfies $X^n_A \perp\!\!\!\perp X^n_B \,|\, X^n_C$. By Theorem 9 there is a subsequence n_m such that $\mathbb{P}(X^{n_m}_{A,B,C}) \xrightarrow{tv} \mathbb{P}(X_{A,B,C})$. By Theorem 10 this implies $X_A \perp\!\!\!\perp X_B \,|\, X_C$. □

Theorem 11 holds for $A, B, C \subseteq V$, so for conditional independencies between endogenous variables. While Theorem 8 (and the underlying lemmas) extend to $A, B, C \subseteq V \cup W$ on the augmented causal graph $G^+(\mathcal{D})$, the total variation convergence of Theorem 9 does not extend to the joint law of endogenous and exogenous variables, since the measures $\mathbb{P}(X^n_V, X_W)$ and $\mathbb{P}(X_V, X_W)$ are mutually singular and hence have total variation distance 1. Hence Theorem 11 is not straightforwardly extended to the augmented graph. For conditional independencies involving exogenous variables, the σ-separation Markov property (Theorem 6) still applies.

Example 7 (d-separation Markov property)**.** *By Theorem 11, the system of causal SDEs satisfies the d-separation Markov property, so the d-separation $P_{lacI} \perp^d P_{tetR} \,|\, M_{tetR}, M_{lacI}$ in the graph from Example 5 implies $P_{lacI} \perp\!\!\!\perp P_{tetR} \,|\, M_{tetR}, M_{lacI}$.*

6 Do-calculus

A question central to causal inference is whether a causal effect (an interventional distribution) can be identified from the observational distribution. When the underlying causal graph is known, such questions can be answered by the rules of do-calculus. We employ a formulation that uses the notion of a causal graph with *intervention variables* (Spirtes et al., 1993; Pearl, 1993; Forré and Mooij, 2020; Dawid, 2021), that is equivalent to the well-known formulation of the do-calculus in terms of *mutilated graphs* (Pearl, 1995, 2009).

Definition 7 (Graph with intervention variables)**.** Given a DMG $G = (V, E, L)$ and intervention target $S \subseteq V$, let $G_{\mathrm{do}(R_S)}$ be the graph G appended with the vertex R_v and edge $R_v \to v$ for each $v \in S$.

More details on *systems of causal SDEs with intervention variables* are given in the appendix (cf. Definition E.1). The following formulation is inspired by Forré and Mooij (2025), Theorem 5.1.2. In the main text we only focus on the graphical aspect of the intervention variables, as this is sufficient for formulating the do-calculus.

Theorem 12. *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every $O \subseteq V$, and let $A, B, C \subseteq V$ be pairwise disjoint. Then:*

1. Insertion/deletion of observation. *If $A \perp^\sigma_{G(\mathcal{D})} B \,|\, C$, then*

$$\mathbb{P}(X_A \,|\, X_B, X_C) = \mathbb{P}(X_A \,|\, X_C) \qquad \mathbb{P}(X_B, X_C)\text{-}a.s.$$

2. Action/observation exchange. *If $A \perp^\sigma_{G(\mathcal{D})_{\mathrm{do}(R_B)}} R_B \,|\, B \cup C$ and $\mathbb{P}(X_C \mid X_B = x_B) \ll \mathbb{P}(X_C \mid \mathrm{do}(X_B = x_B))$ for $\mathbb{P}(X_B)$-a.a. x_B, then*

$$\mathbb{P}(X_A \,|\, \mathrm{do}(X_B), X_C) = \mathbb{P}(X_A \,|\, X_B, X_C) \qquad \mathbb{P}(X_B, X_C)\text{-}a.s.$$

3. Insertion/deletion of action. *If* $A \perp^{\sigma}_{G(\mathcal{D})_{\mathrm{do}(R_B)}} R_B \,|\, C$, *then for every* $x_B \in \mathcal{X}_B$ *such that* $\mathbb{P}(X_C) \ll \mathbb{P}(X_C \mid \mathrm{do}(X_B = x_B))$,

$$\mathbb{P}(X_A \,|\, \mathrm{do}(X_B = x_B), X_C) = \mathbb{P}(X_A \,|\, X_C) \quad \mathbb{P}(X_C)\text{-}a.s.$$

In particular, any system of causal SDEs satisfying Assumption 1 is essentially uniquely solvable w.r.t. every $O \subseteq V$ (Theorem 3), and hence the σ-separation do-calculus applies in the causal graph $G(\mathcal{D})$.

A d-separation analogue of Theorem 12 does not follow automatically from the d-separation Markov property for systems of causal SDEs that satisfy Assumption 2: the proof of Theorem 12 for Rules 2 and 3 uses an auxiliary system $\mathcal{D}_{\mathrm{do}(R_B \sim \nu)}$ (cf. Definition E.1) for which the required d-separation Markov property does not automatically follow. We conjecture this d-separation do-calculus to hold. For acyclic graphs, σ-separation and d-separation coincide, so on acyclic systems the d-separation do-calculus follows trivially from Theorem 12.

We obtain a fundamental consequence of the do-calculus for inferring non-causation in systems of causal SDEs, and thereby establish a clear causal interpretation of the causal graph.

Theorem 13. *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every $O \subseteq V$ and let $u, v \in V$. If there is no directed path from u to v in $G(\mathcal{D})$, then* $\mathbb{P}(X_v \,|\, \mathrm{do}(X_u)) = \mathbb{P}(X_v)$.

Another consequence of Theorem 12 is that (generalised) adjustment formulae like backdoor adjustment, and the ID algorithm – which are all derived from the do-calculus – are also valid, provided the appropriate absolute continuity conditions hold (Pearl, 2009; Forré and Mooij, 2020, 2025).

Similar results have been derived in other frameworks. Røysland et al. (2024) provide identifiability criteria for causal effects based on local independence graphs. Gill and Robins (2001) provide a g-formula for continuous-time processes in the potential outcomes framework, that Ryalen et al. (2026) partially extend to a g-formula for time-varying treatment regimes for marked point processes. The do-calculus and subsequent adjustment formulae consider the identification of certain estimands in terms of observational distributions. We note that the construction of *estimators* for such expressions can be highly non-trivial when the variables take values in function spaces, see e.g. Gill and Robins (2001), Lok (2008) and Rytgaard et al. (2022).

7 Time-splitting and subsampling

When reasoning about causal dynamical systems, it can be useful to evaluate subprocesses on distinct time intervals, for example, when analysing how local modifications propagate through the system. Formally, this requires extending the framework of causal SDEs to time-split systems, in which the global time interval is partitioned into subintervals, each equipped with its own structural dynamics while remaining consistent with the overall system.

Example 8 (Time-split repressilator)**.** *Suppose that for some $t \in (0, T)$ we want to consider the repressilator on the time points $0, t, T$, and the intermediate intervals $(0, t)$ and (t, T). This can straightforwardly be modelled by considering the variables $X^0, X^{(0,t)}, X^t, X^{(t,T)}, X^T$ with* *dynamics described by the SDEs*

$$
\begin{aligned}
X^{(0,t)}(s) &= X^0 &&+ \int_0^s \mu(u, X^0, X^{(0,t)})du + \sigma \tilde{Z}^{(0,t)}(s) && s \in (0,t)\\
X^t &= X^{(0,t)}(t-) \\
X^{(t,T)}(s) &= X^t &&+ \int_t^s \mu(u, X^t, X^{(t,T)})du + \sigma \tilde{Z}^{(t,T)}(s) && s \in (t,T)\\
X^T &= X^{(t,T)}(T-)
\end{aligned}
$$

where $\tilde{Z}^{(0,t)}(s) := Z(s) - Z(0)$ *for* $s \in (0,t)$ *and* $\tilde{Z}^{(t,T)}(s) := Z(s) - Z(t)$ *for* $s \in (t,T)$ *are independent exogenous variables. The graph of the time-split system (marginalised onto* P_{lacI}, P_{tetR} *and* P_{GFP} *for visual clarity, denoted by* $\bar{\mathcal{D}}$*) is shown in Figure 4.*

To model the inhibition of P_{lacI} *during the first half of the process, we can consider the time-split SDE with* $t = T/2$ *and doing the perfect intervention* $\mathrm{do}(M^0_{lacI} = M^{(0,t)}_{lacI} = 0)$*. After releasing the intervention the system returns to its stable behaviour, as depicted in Figure 5.*

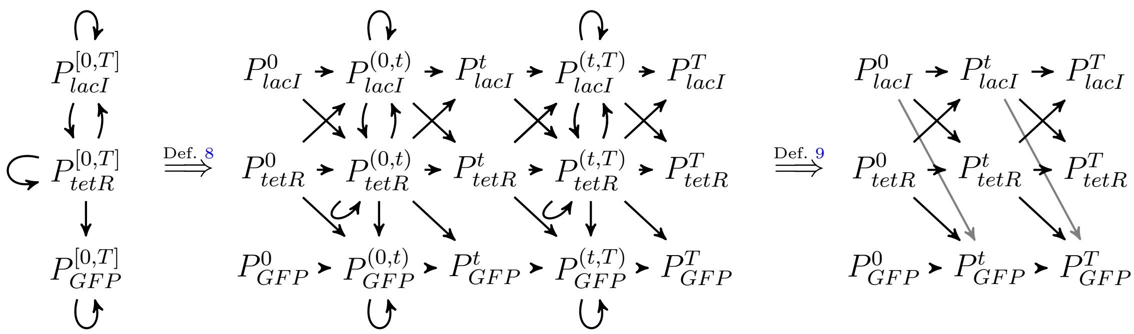


Figure 4: The causal graph $G(\bar{\mathcal{D}})$, the time-split causal graph $G(\bar{\mathcal{D}}^{\mathcal{E}})$ with $\mathcal{E} = \{\{0\}, (0,t), \{t\}, (t,T), \{T\}\}$, and the subsampled causal graph $G(\bar{\mathcal{D}}^{\{0,t,T\}})$ (see Definition 9 below).

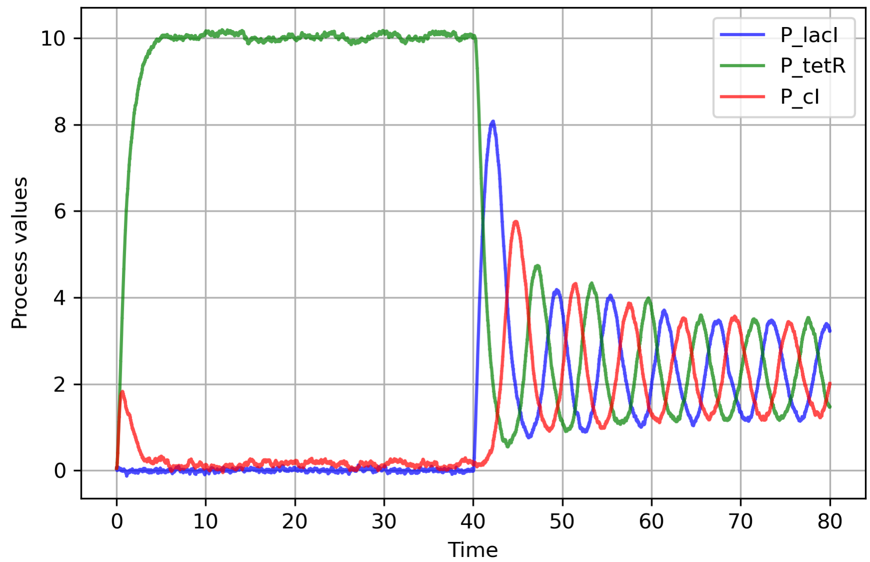


Figure 5: Repressilator with gene 'lacI' knocked out in the first half of the simulation.

We now formalise this construction as follows. Let $\mathcal{E} = \{\mathcal{I}_1, ..., \mathcal{I}_m\}$ be a finite partition of $[0,T]$ into intervals with *temporal ordering*, i.e., for $i < j$ we have $s < t$ for all $s \in \mathcal{I}_i$ and

$t \in \mathcal{I}_j$. For càdlàg path $x \in D([0,T],\mathbb{R}^n)$, if we let $x^{\mathcal{I}} := x|_{\mathcal{I}} \in D(\mathcal{I},\mathbb{R}^n)$ be the restriction of x to $\mathcal{I}$ and write $x^{\mathcal{E}} = (x^{\mathcal{I}_1},...,x^{\mathcal{I}_m})$, then let $\mathrm{cat}(x^{\mathcal{E}}) \in D([0,T],\mathbb{R}^n)$ be the concatenation $\mathrm{cat}(x^{\mathcal{E}})(t) := \sum_{\mathcal{I}_k} \mathbb{1}_{\mathcal{I}_k}(t) x^{\mathcal{I}_k}(t)$ that retrieves x. For consecutive intervals $\mathcal{I}_k, \mathcal{I}_{k+1}$ in $\mathcal{E}$, let

$$b_{k,k+1} := \begin{cases} (\sup \mathcal{I}_k)- & \text{if } \mathcal{I}_k \text{ is right-open,} \\ \sup \mathcal{I}_k & \text{if } \mathcal{I}_k \text{ is right-closed} \end{cases}$$

and $b_{0,1} := 0$ and $b_{m,m+1} := T$. If for the càdlàg process $X(t) = \int_0^t g\mathrm{d}h$ and partition $(\mathcal{I}_1,\mathcal{I}_2)$ of $[0,T]$ we set $X^{\mathcal{I}_1}(t) = \int_0^t g\mathrm{d}h$ for all $t \in \mathcal{I}_1$ and $X^{\mathcal{I}_2}(t) = X^{\mathcal{I}_1}(b_{1,2}) + \int_{b_{1,2}}^t g\mathrm{d}h$ for all $t \in \mathcal{I}_2$, then if X has a jump at the boundary point $\sup \mathcal{I}_1$, it will be contained in $X^{\mathcal{I}_1}(b_{1,2})$ when $\mathcal{I}_1$ is right-closed and in $\int_{b_{1,2}}^t g\mathrm{d}h$ when $\mathcal{I}_1$ is right-open. If for $y \in D([0,T],\mathbb{R}^n)$ we let $y^{\mathcal{I}_1}(t) := y(t) \in D(\mathcal{I}_1,\mathbb{R}^n)$ and $y^{\mathcal{I}_k}(t) := y(t) - y(b_{k-1,k}) \in D(\mathcal{I}_k,\mathbb{R}^n)$ for $k \geq 2$ be the increment path (not counting jumps at $\inf \mathcal{I}_k$ if $\mathcal{I}_k$ is left-open), then let $\mathrm{rec}(y^{\mathcal{E}})(t) := \sum_k \mathbb{1}_{\mathcal{I}_k}(t) \cdot (\sum_{j<k} y^{\mathcal{I}_j}(b_{j,j+1}) + y^{\mathcal{I}_k}(t))$ be the reconstruction that retrieves y.

Definition 8 (Time-split system of SDEs)**.** Let $\mathcal{D} = (V, W, X_W, f, g, h)$ be a system of causal SDEs on $[0,T]$ and let $\mathcal{E} = \{\mathcal{I}_1,...,\mathcal{I}_m\}$ be a finite partition of $[0,T]$ with temporal ordering. The *time-split system of SDEs* $\mathcal{D}^{\mathcal{E}} := \big(V \times \mathcal{E}, W \times \mathcal{E}, X_{W\times\mathcal{E}}, f_V^{\mathcal{E}}, g_V^{\mathcal{E}}, h_V^{\mathcal{E}}\big)$ is defined with for each $(w,\mathcal{I}_k) \in W \times \mathcal{E}$ the *exogenous increment process* $X_w^{\mathcal{I}_k} \in D(\mathcal{I}_k,\mathbb{R})$ defined as $X_w^{\mathcal{I}_1}(t) := X_w(t)$ for $t \in \mathcal{I}_1$ and $X_w^{\mathcal{I}_k}(t) := X_w(t) - X_w(b_{k-1,k})$ for $t \in \mathcal{I}_k$ for all $k \geq 2$, and for each $(v,\mathcal{I}_k) \in V \times \mathcal{E}$ the causal SDE[6]

$$X_v^{\mathcal{I}_k}(s) = X_v^{\mathcal{I}_{k-1}}(b_{k-1,k}) + f_v^{\mathcal{I}_k}\big(s, X_{\alpha(v)}^{\mathcal{E}}\big) + \int_{b_{k-1,k}}^s g_v^{\mathcal{I}_k}\big(u-, X_{\{v\}\cup\beta(v)}^{\mathcal{E}}\big)\mathrm{d}h_v^{\mathcal{I}_k}\big(u, X_{\gamma(v)}^{\mathcal{E}}\big)$$

for $s \in \mathcal{I}_k$, with

$$\begin{aligned} f_v^{\mathcal{I}_k}(s, x_{\alpha(v)}^{\mathcal{E}}) &:= f_v\big(s, \mathrm{cat}(x_{\alpha(v)\cap V}^{\mathcal{E}}), \mathrm{rec}(x_{\alpha(v)\cap W}^{\mathcal{E}})\big) - f_v\big(b_{k-1,k}, \mathrm{cat}(x_{\alpha(v)\cap V}^{\mathcal{E}}), \mathrm{rec}(x_{\alpha(v)\cap W}^{\mathcal{E}})\big), \\ g_v^{\mathcal{I}_k}(u-, x_{\{v\}\cup\beta(v)}^{\mathcal{E}}) &:= g_v\big(u-, \mathrm{cat}(x_{\{v\}\cup(\beta(v)\cap V)}^{\mathcal{E}}), \mathrm{rec}(x_{\beta(v)\cap W}^{\mathcal{E}})\big), \\ h_v^{\mathcal{I}_k}(u, x_{\gamma(v)}^{\mathcal{E}}) &:= h_v\big(u, \mathrm{rec}(x_{\gamma(v)}^{\mathcal{E}})\big), \end{aligned}$$

and where we set $X_v^{\mathcal{I}_0}(b_{0,1}) := f_v(0, X_{\alpha(v)}^{\mathcal{I}_1})$.

The time-split system is observationally equivalent to the original: concatenating any solution of $\mathcal{D}^{\mathcal{E}}$ yields a solution of $\mathcal{D}$, and conversely the restrictions of any solution of $\mathcal{D}$ form a solution of $\mathcal{D}^{\mathcal{E}}$; see Lemma F.1. Note however that after intervening on a variable $X_v^{\mathcal{I}}$ with $\mathcal{I}$ left-open, the resulting path need not be càdlàg. This imposes an additional condition on the notion of essential unique solvability w.r.t. $O \subseteq V \times \mathcal{E}$, which is stated in terms of input processes $\varphi_S : \mathcal{X}_W \to \mathcal{X}_S$ where $S = V \times \mathcal{E} \setminus O$. Only φ_S should be considered such that for $(v,\mathcal{I}) \in S$ the path $\varphi_{(v,\mathcal{I})}$ has a left limit at $\sup \mathcal{I}$ (which is not automatic when $\mathcal{I}$ is right-open), and if $(v,\mathcal{I}_k),(v,\mathcal{I}_{k+1}) \in S$, the concatenation of $\varphi_{(v,\mathcal{I}_k)}$ and $\varphi_{(v,\mathcal{I}_{k+1})}$ must be right-continuous at $\sup \mathcal{I}_k$. We refer to such φ_S as *admissible*. The causal graph $G(\mathcal{D}^{\mathcal{E}})$ has vertex set $V \times \mathcal{E}$ and edges $(u,\mathcal{I}') \to (v,\mathcal{I})$ derived exactly as in Definition 2. We note three sources of graphical sparsity:

(i) *Adaptedness*: by adaptedness of f_v, g_v, h_v, the values of $f_v^{\mathcal{I}_k}, g_v^{\mathcal{I}_k}, h_v^{\mathcal{I}_k}$ only depend on $X_{V\cup W}^{\mathcal{I}_j}$ with $j \leq k$, so there are no edges into $(v,\mathcal{I}_k)$ from later intervals.

(ii) *Markovianity*: If in the original SDE g_v is Markov and f_v either models the initial condition or is Markov then there are no edges from any $X_u^{\mathcal{I}_j}$ with $u \in V$ and $j < k-1$ to $X_v^{\mathcal{I}_k}$.

[6] The pathwise integration map Ψ is stated for stochastic integrals on $[0,T]$, but can be extended to any interval $\mathcal{I}_k$.

(iii) *Independent increments*: if an exogenous process X_w has independent increments, then any two increment processes $X_w^{\mathcal{I}_j}$ and $X_w^{\mathcal{I}_k}$ with $j \neq k$ are independent. Consequently, if X_v depends on an exogenous process X_w, then in the graph $G(\mathcal{D}^{\mathcal{E}})$ there is no bidirected edge between $X_v^{\mathcal{I}_{k-1}}$ and $X_v^{\mathcal{I}_k}$ because of the common cause X_w.

If $\mathcal{D}$ satisfies Assumption 1 and its exogenous processes have independent increments, then by Lemma F.2 in the Appendix, $\mathcal{D}^{\mathcal{E}}$ is essentially uniquely solvable w.r.t. every $O \subseteq V \times \mathcal{E}$. By Lemma F.3 in the Appendix, every time-split system of causal SDEs $\mathcal{D}^{\mathcal{E}}$ whose exogenous processes have independent increments can be embedded as a system of causal SDEs on $[0, T]$, so the Markov properties (Theorems 6 and 11) and the do-calculus (Theorem 12) apply directly to $\mathcal{D}^{\mathcal{E}}$, with σ- and d-separation read off from the time-split causal graph $G(\mathcal{D}^{\mathcal{E}})$.

Theorem 14 (Markov properties for time-split systems of causal SDEs)**.** *Let $\mathcal{E}$ be a finite partition of $[0, T]$ into intervals with temporal ordering, and let $\mathcal{D}$ be a system of causal SDEs whose exogenous processes X_w ($w \in W$) have independent increments.*

(i) If $\mathcal{D}$ satisfies Assumption 1, then $\mathcal{D}^{\mathcal{E}}$ satisfies the σ-separation Markov property: for all $A, B, C_1, ..., C_r \subseteq V$ and $\mathcal{I}_A, \mathcal{I}_B, \mathcal{I}_{C_1}, ..., \mathcal{I}_{C_r} \subseteq \mathcal{E}$,

$$X_A^{\mathcal{I}_A} \underset{G(\mathcal{D}^{\mathcal{E}})}{\overset{\sigma}{\perp}} X_B^{\mathcal{I}_B} \mid X_{C_1}^{\mathcal{I}_{C_1}}, ..., X_{C_r}^{\mathcal{I}_{C_r}} \implies X_A^{\mathcal{I}_A} \perp\!\!\!\perp X_B^{\mathcal{I}_B} \mid X_{C_1}^{\mathcal{I}_{C_1}}, ..., X_{C_r}^{\mathcal{I}_{C_r}}.$$

(ii) If $\mathcal{D}$ satisfies Assumption 2, then $\mathcal{D}^{\mathcal{E}}$ satisfies the d-separation Markov property: for all $A, B, C_1, ..., C_r \subseteq V$ and $\mathcal{I}_A, \mathcal{I}_B, \mathcal{I}_{C_1}, ..., \mathcal{I}_{C_r} \subseteq \mathcal{E}$,

$$X_A^{\mathcal{I}_A} \underset{G(\mathcal{D}^{\mathcal{E}})}{\overset{d}{\perp}} X_B^{\mathcal{I}_B} \mid X_{C_1}^{\mathcal{I}_{C_1}}, ..., X_{C_r}^{\mathcal{I}_{C_r}} \implies X_A^{\mathcal{I}_A} \perp\!\!\!\perp X_B^{\mathcal{I}_B} \mid X_{C_1}^{\mathcal{I}_{C_1}}, ..., X_{C_r}^{\mathcal{I}_{C_r}}.$$

Theorem 15 (Do-calculus for time-split systems of causal SDEs)**.** *Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes X_w ($w \in W$) have independent increments, and let $\mathcal{E}$ be a finite partition of $[0, T]$ into intervals with temporal ordering. For $A, B, C_1, \ldots, C_r \subseteq V$ and $\mathcal{I}_A, \mathcal{I}_B, \mathcal{I}_{C_1}, \ldots, \mathcal{I}_{C_r} \subseteq \mathcal{E}$ such that the blocks $A^{\mathcal{I}_A}, B^{\mathcal{I}_B}, C_1^{\mathcal{I}_{C_1}}, \ldots, C_r^{\mathcal{I}_{C_r}}$ are pairwise disjoint in $V \times \mathcal{E}$, write $X_C^{\mathcal{I}_C} := (X_{C_1}^{\mathcal{I}_{C_1}}, \ldots, X_{C_r}^{\mathcal{I}_{C_r}})$. Then:*

1. Insertion/deletion of observation. *If $X_A^{\mathcal{I}_A} \perp^{\sigma}_{G(\mathcal{D}^{\mathcal{E}})} X_B^{\mathcal{I}_B} \mid X_C^{\mathcal{I}_C}$, then*

$$\mathbb{P}(X_A^{\mathcal{I}_A} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}) = \mathbb{P}(X_A^{\mathcal{I}_A} \mid X_C^{\mathcal{I}_C}) \quad \mathbb{P}(X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C})\text{-a.s.}$$

2. Action/observation exchange. *If $X_A^{\mathcal{I}_A} \perp^{\sigma}_{G(\mathcal{D}^{\mathcal{E}})_{\mathrm{do}(R_B^{\mathcal{I}_B})}} R_B^{\mathcal{I}_B} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}$ and $\mathbb{P}(X_C^{\mathcal{I}_C} \mid X_B^{\mathcal{I}_B} = x_B) \ll \mathbb{P}(X_C^{\mathcal{I}_C} \mid \mathrm{do}(X_B^{\mathcal{I}_B} = x_B))$ for $\mathbb{P}(X_B^{\mathcal{I}_B})$-a.a. x_B, then*

$$\mathbb{P}(X_A^{\mathcal{I}_A} \mid \mathrm{do}(X_B^{\mathcal{I}_B}), X_C^{\mathcal{I}_C}) = \mathbb{P}(X_A^{\mathcal{I}_A} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}) \quad \mathbb{P}(X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C})\text{-a.s.}$$

3. Insertion/deletion of action. *If $X_A^{\mathcal{I}_A} \perp^{\sigma}_{G(\mathcal{D}^{\mathcal{E}})_{\mathrm{do}(R_B^{\mathcal{I}_B})}} R_B^{\mathcal{I}_B} \mid X_C^{\mathcal{I}_C}$, then for every $x_B \in \mathcal{X}_{B^{\mathcal{I}_B}}$ such that $\mathbb{P}(X_C^{\mathcal{I}_C}) \ll \mathbb{P}(X_C^{\mathcal{I}_C} \mid \mathrm{do}(X_B^{\mathcal{I}_B} = x_B))$,*

$$\mathbb{P}(X_A^{\mathcal{I}_A} \mid \mathrm{do}(X_B^{\mathcal{I}_B} = x_B), X_C^{\mathcal{I}_C}) = \mathbb{P}(X_A^{\mathcal{I}_A} \mid X_C^{\mathcal{I}_C}) \quad \mathbb{P}(X_C^{\mathcal{I}_C})\text{-a.s.}$$

Example 9. *The summary, time-split, and subsampled causal graphs of the repressilator are depicted in Figure 4. In the time-split graph, the parents of $P^{(0,t)}_{GFP}$ are $\{P^0_{GFP}, P^0_{tetR}, P^{(0,t)}_{tetR}\}$. Let $Z := (P^0_{tetR}, P^{(0,t)}_{tetR})$. Since P^0_{GFP} is a source with no other outgoing edges, every backdoor path from $P^{(0,t)}_{GFP}$ to P^T_{GFP} must enter $P^{(0,t)}_{GFP}$ via P^0_{tetR} or $P^{(0,t)}_{tetR}$, so conditioning on Z blocks all such paths. Moreover, Z contains no descendants of $P^{(0,t)}_{GFP}$. Assuming the absolute-continuity conditions of Theorem 15 hold, Rule 3 gives $\mathbb{P}(Z \mid \mathrm{do}(P^{(0,t)}_{GFP})) = \mathbb{P}(Z)$, and Rule 2 gives $\mathbb{P}(P^T_{GFP} \mid \mathrm{do}(P^{(0,t)}_{GFP}), Z) = \mathbb{P}(P^T_{GFP} \mid P^{(0,t)}_{GFP}, Z)$, which combine to yield the backdoor adjustment formula*

$$\mathbb{P}\left(P^T_{GFP} \mid \mathrm{do}(P^{(0,t)}_{GFP})\right) = \int \mathbb{P}\left(P^T_{GFP} \mid P^{(0,t)}_{GFP}, P^0_{tetR}, P^{(0,t)}_{tetR}\right) \mathrm{d}\mathbb{P}(P^0_{tetR}, P^{(0,t)}_{tetR}). \tag{9}$$

Finally, we consider time-split systems where the endogenous variables only consist of time-points.

Definition 9 (Subsampled system of causal SDEs)**.** Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes have independent increments, so that the time-split system $\mathcal{D}^{\mathcal{E}}$ is essentially uniquely solvable w.r.t. every $O \subseteq V \times \mathcal{E}$ by Lemma F.2. Given a set of time-points $\mathcal{S} \subseteq [0, T]$, let $\mathcal{E}'$ be a set of intervals such that $\mathcal{E} := \mathcal{S} \cup \mathcal{E}'$ partitions $[0, T]$. The *subsampled system of causal SDEs* is defined as

$$\mathcal{D}^{\mathcal{S}} := \left(\mathcal{D}^{\mathcal{E}}\right)_{\mathrm{marg}(V \times \mathcal{E}')},$$

the marginalisation of the time-split system $\mathcal{D}^{\mathcal{E}}$ onto the time-point components $V \times \mathcal{S}$. Since $\mathcal{D}^{\mathcal{E}}$ is essentially uniquely solvable w.r.t. every $O \subseteq V \times \mathcal{E}$, Theorem 5 gives that $\mathcal{D}^{\mathcal{S}}$ is again essentially uniquely solvable w.r.t. every $O' \subseteq V \times \mathcal{S}$.

Example 10. *From the Markov property in the summary graph (Figure 4) we infer that $P^{[0,T]}_{lacI} \perp\!\!\!\perp P^{[0,T]}_{GFP} \mid P^{[0,T]}_{tetR}$. If we consider the subsampled time-series, however, we might have $P^{\{0,t,T\}}_{lacI} \not\!\perp\!\!\!\perp P^{\{0,t,T\}}_{GFP} \mid P^{\{0,t,T\}}_{tetR}$.*

Analysis of the information loss of the subsampling operation could be particularly interesting in the light of (im)possibility results for inferring features of $\mathcal{D}$ from the subsampled system. In particular, inferences from subsampled time series need not represent features of the underlying continuous-time system. For example, conditional independence relations which hold in $\mathcal{D}$ might not hold in $\mathcal{D}^{\mathcal{S}}$, as portrayed in the example above; see also Aalen et al. (2016), who show a similar phenomenon for local independence graphs. Another example is that of perfect adaptation: if $A \to B$ but B perfectly adapts between two sampling points, no edge from A to B is present in the graph of the subsampled system (Blom and Mooij, 2023; Weinberger, 2026).

7.1 Granger causality in time-split systems

Having a time-splitting operation at our disposal, we can consider predicting future values of a process from past values of the same and other processes. For discrete-time stochastic processes, Granger (1969, 1980) called A a cause of B if, given the history of all other variables, the history of A contains useful information for predicting B. Florens and Fougere (1996) extended this notion to continuous time. We adopt the following slightly more flexible variant, allowing arbitrary conditioning sets:

Definition 10 (Global Granger non-causation)**.** Let $A, B, C \subseteq V$. We say that X_A does not *globally Granger cause X_B given X_C* if

$$X^{[0,s]}_A \perp\!\!\!\perp X^{(s,t]}_B \mid X^{[0,s]}_C \quad \text{for all } 0 \leq s < t \leq T,$$

equivalently, $X^{[0,s]}_A \perp\!\!\!\perp X^{(s,T]}_B \mid X^{[0,s]}_C$ for all $0 \leq s < T$.

The classical notion of X_A *does not (globally) Granger-cause* X_B is recovered as the special case $C = V \setminus A$. If $\mathcal{D}$ satisfies Assumption 1, global Granger non-causation can be read off from the time-split causal graph via the σ-separation Markov property:

$$X_A^{[0,s]} \underset{G(\mathcal{D}^{\{[0,s],(s,T]\}})}{\overset{\sigma}{\perp}} X_B^{(s,t]} \mid X_C^{[0,s]} \implies X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \mid X_C^{[0,s]}.$$

When $\mathcal{D}$ additionally satisfies Assumption 2, the same conclusion follows from the stronger d-separation Markov property in the time-split graph (Theorem 14(ii)).

The following result relates Granger causation to causation in $G(\mathcal{D})$.

Theorem 16. *Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes X_w ($w \in W$) have independent increments, and let $u, v \in V$.*

(i) Let $\mathcal{D}^{\{[0,s],(s,T]\}}$ be faithful for every $0 \leq s < T$. If there is a directed path from u to v in $G(\mathcal{D})$, then u is a Granger cause of v.

(ii) Let $G(\mathcal{D})$ have no bidirected edges. If u is a Granger cause of v, then there is a directed path from u to v in $G(\mathcal{D})$.

For discrete-time models, a similar link between Granger causation and causation in SCMs has also been remarked by Peters et al. (2013) (Theorem 10.3), White and Lu (2010) and Eichler (2012).

7.1.1 Local independence

Besides the 'global' Granger non-causality, much interest has been shown in a local notion called *local independence* (Schweder, 1970; Aalen, 1978; Didelez, 2008; Mogensen and Hansen, 2020). It has been conjectured that it is equivalent to the conditional independence $X_A^{[0,t)} \perp\!\!\!\perp X_B^t \mid X_C^{[0,t)}$ (Didelez, 2008), but if $B \subseteq C$ and X_B is continuous, then $X_B(t-) = X_B(t)$ a.s. so this conditional independence holds trivially even if there is a local dependence (Christgau et al., 2023). In this section we investigate the relation between Granger non-causality and local independence, and we derive a local independence Markov property based on σ-separation (and d-separation when appropriate) in the time-split graph.

Defining local independence requires some technical background. A *special semimartingale* $X(t)$ adapted to a filtration $\mathcal{F}$ on compact interval $[0, T]$ is a stochastic process that has a Doob-Meyer decomposition $X(t) = \Lambda(t) + M(t)$, where Λ is a $\mathcal{F}$-*predictable* finite variation process with $\Lambda(0) = 0$, and M is a local martingale with respect to $\mathcal{F}$; this decomposition is unique up to indistinguishability. Throughout this section, let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 such that for every $B, C \subseteq V$ the *optional projection* (the unique càdlàg version of) $\mathbb{E}[X_B(t) \mid \mathcal{F}_t^C]$ exists and is a special semimartingale, where $\mathcal{F}^C$ denotes the filtration generated by X_C. A convenient sufficient condition is that $t \mapsto X_B(t)$ is almost surely bounded by some deterministic constant (see Lemma F.4 in the Appendix).

Definition 11 (Local independence)**.** Let $A, B, C \subseteq V$ and consider the following optional projections and their Doob-Meyer decompositions

$$\begin{aligned} \mathbb{E}[X_B(t) \mid \mathcal{F}_t^C] &= \Lambda^C(t) + M^C(t) \\ \mathbb{E}[X_B(t) \mid \mathcal{F}_t^{A,C}] &= \Lambda^{A,C}(t) + M^{A,C}(t). \end{aligned} \tag{10}$$

We say that X_B is *locally independent*[7] from X_A given X_C, written $X_A \not\to X_B \mid X_C$, if $\Lambda^{A,C} = \Lambda^C$ a.s.

[7] Florens and Fougere (1996) refer to this as *weak instantaneous non-causality*, Comte and Renault (1996) call it *local Granger non-causality*.

Local independence is equivalent to $\Lambda^{A,C}$ having a $\mathcal{F}^C$-predictable version, and this version is then necessarily a.s. equal to Λ^C (see Mogensen and Hansen, 2020 Appendix E). Some authors assume that the trajectory $t \mapsto \Lambda^{A,C}(t)$ is absolutely continuous, i.e. $\Lambda^{A,C}(t) = \int_0^t \lambda^{A,C}(s)\mathrm{d}s$ for some $\mathcal{F}^{A,C}$-adapted process $\lambda^{A,C}$ called the *intensity* process, in which case local independence is equivalent to the intensity process $\lambda^{A,C}$ having a $\mathcal{F}^C$-adapted version.

Some authors (e.g. Aalen, 1978; Florens and Fougere, 1996; Comte and Renault, 1996; Didelez, 2008; Christgau et al., 2023) always condition on the history of the target variable, i.e. they consider $B \subseteq C$, in which case we get $X_B(t) = \mathbb{E}[X_B(t) \,|\, \mathcal{F}_t^C] = \mathbb{E}[X_B(t) \,|\, \mathcal{F}_t^{A,C}]$, so (10) reads $X_B = M^{A,C} + \Lambda^{A,C}$ and $X_B = M^C + \Lambda^C$, in which case local independence $X_A \not\rightarrow X_B \,|\, X_C$ is equivalent to $M^{A,C} = M^C$ a.s. If $B \subseteq C$, local independence trivially holds if X_B is almost surely differentiable, since then $X_B(t) = X_B(0) + \int_0^t X_B'(s)\mathrm{d}s$, so $X_B' = \lambda^C = \lambda^{A,C}$ a.s. (Comte and Renault, 1996). Other authors (Mogensen et al., 2018; Mogensen and Hansen, 2020) allow for general conditioning sets C; we follow this convention.

The following result, due to Florens and Fougere (1996) for the special case where $B \subseteq C$, relates Granger non-causality to local independence.

Theorem 17. *Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes X_w ($w \in W$) have independent increments, such that $\sup_{t\in[0,T]} \|X_V(t)\| \leq c$ almost surely for some deterministic $c < \infty$, and let $A, B, C \subseteq V$. If for all $0 < s < t \leq T$ we have $X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \,|\, X_C^{[0,s]}$ in the time-split system $\mathcal{D}^{\{[0,s],(s,t],(t,T]\}}$, then $X_A \not\rightarrow X_B \,|\, X_C$ and $M^C = M^{A,C}$. Conversely, if $B = C$, then $X_A \not\rightarrow X_B \,|\, X_C$ implies $X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \,|\, X_C^{[0,s]}$ for all $0 \leq s < t \leq T$.*

Florens and Fougere (1996) show that equivalence between local independence and Granger non-causality also holds for certain point processes and Markov processes for the case that B is a strict subset of C. As a corollary of Theorem 17, we obtain the following Markov property for local independence in terms of the time-split graph.

Corollary 1 (Markov property for local independence)**.** *Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes X_w ($w \in W$) have independent increments, such that $\sup_{t\in[0,T]} \|X_V(t)\| \leq c$ almost surely for some deterministic $c < \infty$. Let $A, B, C \subseteq V$.*

(i) If $X_A^{[0,s]} \perp^{\sigma}_{G(\mathcal{D}^{\{[0,s],(s,T]\}})} X_B^{(s,T]} \,|\, X_C^{[0,s]}$ for all $0 < s < T$, then $X_A \not\rightarrow X_B \,|\, X_C$.

(ii) If g_v is Markov, f_v either models the initial condition or is Markov, and we have $X_A^{[0,s]} \perp^{\sigma}_{G(\mathcal{D}^{\{[0,s],(s,T]\}})} X_B^{(s,T]} \,|\, X_C^{[0,s]}$ for some $0 < s < T$, then $X_A \not\rightarrow X_B \,|\, X_C$.

(iii) If $\mathcal{D}$ satisfies Assumption 2 and g_v is Markov, and $X_A^{[0,s]} \perp^{d}_{G(\mathcal{D}^{\{[0,s],(s,T]\}})} X_B^{(s,T]} \,|\, X_C^{[0,s]}$ for some $0 < s < T$, then $X_A \not\rightarrow X_B \,|\, X_C$.

Note that in clauses (ii) and (iii), it suffices to verify the separation for a single s since g_v is Markov and f_v either models the initial condition or is Markov, and hence the time-split graph $G(\mathcal{D}^{\{[0,s],(s,T]\}})$ does not depend on $s \in (0,T)$.

Various other Markov properties have been derived for local independence in terms of the causal graph $G(\mathcal{D})$, see e.g. Didelez (2008) for a Markov property for point processes in terms of δ-separation (assuming $B \subseteq C$, excluding latent confounding), Mogensen et al. (2018) for additive noise diffusion processes in terms of μ-separation (not assuming $B \subseteq C$, excluding latent confounding), and Mogensen and Hansen (2022) for Ornstein-Uhlenbeck processes (not assuming $B \subseteq C$, allowing for latent confounding).

8 Discussion

In this work we developed a formal framework that equips systems of stochastic differential equations (SDEs) with causal semantics and lifts the SCM toolkit — causal graphs, graphical Markov properties and the do-calculus — to this setting via a pathwise interpretation of the causal SDEs. We introduced the notion of *essential unique solvability* of a system of causal SDEs, and provided an SCC-wise Lipschitz condition (Assumption 1) and a subclass of additive noise SDE (Assumption 2) that are sufficient for this property to hold. We also introduced a *marginalisation* operation that integrates out latent processes while preserving the observational and interventional distributions on the retained processes, and showed that systems of causal SDEs remain essentially uniquely solvable after marginalisation. Following the acyclification proof strategy of Forré and Mooij (2017); Bongers et al. (2021); Forré and Mooij (2025), we established the σ-separation Markov property for essentially uniquely solvable systems of causal SDEs, and the stronger d-separation Markov property for the class of (cyclic) additive-noise SDEs via total variation convergence of continuous Euler approximations. The σ-separation Markov property implies the do-calculus and with that an explicit causal interpretation of the *causal graph* of a system of causal SDEs. Combining marginalisation with time-split systems yields the *subsampled* system (Definition 9), which models the dynamics observed at a finite set of time-points. Time-split systems give a Markov property for continuous-time Granger non-causality (equivalent to the absence of a directed path in the causal graph under faithfulness and no latent confounding) and an analogous Markov property for local independence. We also argue that time-split systems of causal SDEs provide a natural framework for constraint-based causal discovery from time-series data, as showcased in the following section.

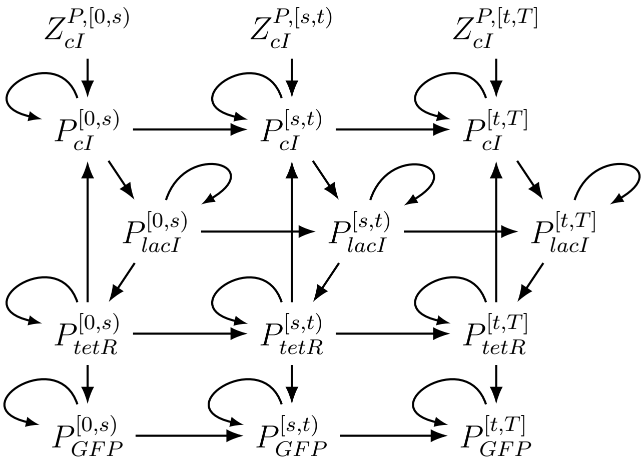
8.1 Constraint-based Causal Discovery

As mentioned earlier, the cornerstone of the discovery of causal relations is experimentation: if $\mathbb{P}(Y \mid \mathrm{do}(X))$ depends on X, then X is a cause of Y. However, in many practical scenarios experimentation is impossible (e.g. expensive or unethical), so one might want to infer causal relations from observational data. Causal discovery algorithms estimate the underlying causal graph of an SCM from observational data by exploiting statistical patterns in the data. *Constraint-based* causal discovery algorithms specifically rely on conditional independence patterns, the Markov property and faithfulness assumption of the SCM.

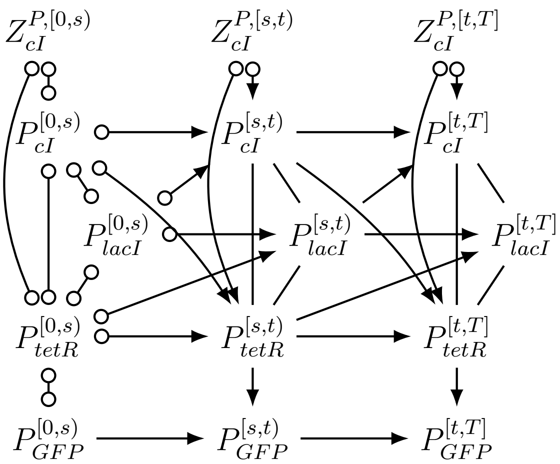
A classical example is the PC algorithm (Spirtes et al., 2000), which assumes that the SCM is acyclic, faithful, and has no latent confounding. To allow for latent confounding, the FCI algorithm has been proposed by Spirtes et al. (1999). It has later been shown to be sound and complete for acyclic graphs (with the d-separation Markov property and faithfulness assumption) by Zhang (2008), and subsequently has been shown to be sound and complete for cyclic SCMs as well with regard to the σ-separation Markov property and faithfulness assumption, in the setting without selection bias (Mooij and Claassen, 2020). FCI does not directly estimate the underlying graph of the SCM, but instead estimates a *Partial Ancestral Graph* (PAG), representing ancestral relations and the Markov equivalence class. We apply FCI to the repressilator in Example 11 below, assuming a conditional independence oracle. Other constraint-based causal discovery algorithms that are known to be sound for (possibly cyclic) systems with the σ-separation Markov property and faithfulness assumption are Local Causal Discovery (Cooper, 1997; Mooij et al., 2020) and Y-structures (Mani, 2006; Mooij et al., 2020).

For cyclic systems with a d-separation Markov property and faithfulness assumption, as in the setting of Theorem 11, the FCI algorithm is not sound. For this scenario (excluding latent confounding and selection bias), Richardson (1996) introduced the CCD algorithm, which was subsequently extended to the CCI algorithm by Strobl (2019) to allow for latent confounding and selection bias. The CCI algorithm outputs a *Maximal Almost Ancestral Graph* (MAAG). A main drawback is that we do not know whether any sparsity in the MAAG represents conditional independencies in the data, since no Markov property has been proven for MAAGs. Until that question is answered, we can only read off the absence and presence of ancestral relations from the MAAG.

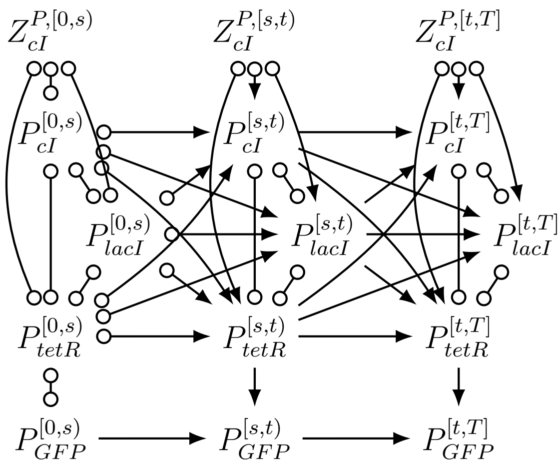
Example 11. *Consider the repressilator $\bar{\mathcal{D}}$ after projection onto $Z^P_{cI}, P_{cI}, P_{lacI}, P_{tetR}$ and P_{GFP}. After time-splitting into the intervals $\mathcal{E} = \{[0,s), [s,t), [t,T]\}$, the time-split causal graph $G(\bar{\mathcal{D}}^{\mathcal{E}})$ is depicted in Figure 6a. When applying CCI to this system with a d-separation oracle (emulating a consistent conditional independence test with infinite data, under the d-separation Markov property and the faithfulness assumption), we obtain the MAAG as depicted in Figure 6b. When applying FCI (assuming no selection bias) to this system with a σ-separation oracle (so under the σ-separation Markov property and the faithfulness assumption), we obtain the PAG as depicted in Figure 6c.*



(a) The ground truth.



(b) The MAAG computed by CCI.



(c) The PAG computed by FCI.

Figure 6: Outputs of the constraint-based causal discovery algorithms CCI and FCI applied to a subset of the variables of the repressilator. The exogenous noise process Z^P_{cI} is included as a parent of P_{cI} in the cycle, for sake of exposing the difference between CCI and FCI.

Other approaches to causal discovery for discrete-time systems include Malinsky and Spirtes (2018); Runge et al. (2019); Reiter et al. (2024); see Assaad et al. (2022) for an overview. Other approaches for causal discovery for systems of SDEs include Guan et al. (2024), Engelke et al.

(2024), Nathaniel et al. (2025) and Brück et al. (2026). Closely related to our framework, Manten et al. (2025) establish an asymmetric Markov property for the conditional independence $X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \mid X_C^{[0,s]}, X_{C\setminus B}^{(s,t]}$, based on σ-separation in a 'lifted graph' constructed from the causal graph $G(\mathcal{D})$; this lifted graph corresponds to the time-split causal graph $G(\mathcal{D}^{\{[0,s],(s,t],(t,T]\}})$ if f_v models the initial condition and g_v is Markov. When Assumption 2 holds, our d-separation Markov property in the time-split graph (Theorem 14(ii)) shows that their Markov property can be strengthened.

8.2 Limitations and open problems

Our framework assumes well-posedness of the underlying SDE system through a global Lipschitz condition, which might be extended to include more general classes of dynamics. We proved the d-separation Markov property for a class of additive-noise SDEs (Theorem 11); we conjecture that the result holds for a larger class of systems of causal SDEs that excludes instantaneous cycles.

A central open direction is proving the d-separation Markov property for marginalised systems. This would boil down to finding a model class like Assumption 2 for which we have a d-separation Markov property and which is itself closed under marginalisation.

Another open problem concerns the d-separation analogue of the do-calculus for cyclic SDEs. Theorem 12 establishes the σ-separation do-calculus for essentially uniquely solvable systems of causal SDEs via an auxiliary system $\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}$ augmented with intervention variables, and on acyclic systems the d-separation analogue follows trivially since d-separation and σ-separation coincide. For cyclic additive-noise SDEs the analogue does not follow from Theorem 11, because the auxiliary system falls outside the additive-noise class. We conjecture that the d-separation do-calculus nevertheless holds for additive-noise SDEs satisfying Assumption 2.

On the causal-discovery side, several open problems are especially relevant for time-split systems of causal SDEs. Immediate ones are the extension of FCI with σ-separation to allow for selection bias, and investigating Markov properties for MAAGs and the completeness of CCI. An especially relevant extension is *tiered FCI* (tFCI) by Andrews et al. (2020), where one specifies a temporal ordering between the variables as background knowledge, excluding causal relations from future to past. Soundness of tFCI has only been proven for acyclic graphs with d-separation Markov property and faithfulness (Andrews et al., 2020) without selection bias; it is unknown whether it is complete in this setting. Moreover, it is unknown whether tiered background knowledge can be used in FCI for cyclic systems with the σ-separation Markov property and faithfulness assumption, with and without selection bias. In the cyclic setting with the d-separation and faithfulness assumption, no extension of CCI has been proposed to incorporate tiered background knowledge. Lastly, it seems reasonable to believe that by using (tiered) FCI/CCI on a time-split system and then mapping back to the 'summary' PAG/MAAG for $\mathcal{D}$, we can identify more edges than straightforward FCI/CCI on $\mathcal{D}$.

For causal discovery, a main bottleneck is consistent conditional independence testing for sample paths: if variables X, Y, Z take values in $D(\mathcal{I}_X, \mathbb{R}^k)$, $D(\mathcal{I}_Y, \mathbb{R}^\ell)$ and $D(\mathcal{I}_Z, \mathbb{R}^m)$ respectively, testing $X \perp\!\!\!\perp Y \mid Z$ is not straightforward. Recent results for functional CI testing are proposed in Lundborg et al. (2022), Laumann et al. (2023) and Manten et al. (2024). Boeken et al. (2026) provide topological criteria for the existence of consistent conditional independence tests that may be useful for analysing this problem.

Beyond causal discovery, future work should explore statistical aspects of causal effect estimation; there is an opportunity to combine this formalism with existing statistical theory in this field. Finally, experimental validation on real-world dynamical systems – such as gene regulatory networks, climate models, and financial systems – will be essential to assess the practical utility of this approach.

Acknowledgements

We thank Claude Code, Sonja Cox and Patrick Forré for helpful remarks and suggestions. Philip Boeken was supported by Booking.com.

A Functional representation of semimartingale SDEs

The pathwise interpretation of systems of causal SDEs (Definition 4) and the solvability results of Section 3 rest on the following functional-representation theorem of Przybyłowicz et al. (2024), which we restate here for completeness. Recall that $x^{\wedge t}(s) := x(s \wedge t)$.

Theorem A.1 (Przybyłowicz et al., 2024, Theorem 3.1)**.** *Let $m, d, r \in \mathbb{N}$ and let $g : [0, T] \times D([0, T], \mathbb{R}^d) \times D([0, T], \mathbb{R}^r) \to \mathbb{R}^{d \times m}$ be a matrix-valued map such that*

(i) $t \mapsto g(t, x, y)$ is càdlàg for all $(x, y) \in D([0, T], \mathbb{R}^d) \times D([0, T], \mathbb{R}^r)$;

(ii) g is $\mathcal{B}([0, T]) \otimes \mathcal{B}(D([0, T], \mathbb{R}^d)) \otimes \mathcal{B}(D([0, T], \mathbb{R}^r))$-measurable;

(iii) $g(t, x, y) = g(t, x^{\wedge t}, y^{\wedge t})$ for all x, y and $t \in [0, T]$;

(iv) g satisfies the linear-growth and Lipschitz conditions (4)–(5) in x given y: there exists a measurable $K : [0, T] \times D([0, T], \mathbb{R}^r) \to (0, \infty)$ with $t \mapsto K(t, y)$ càdlàg, such that for all $x, x_1, x_2 \in D([0, T], \mathbb{R}^d)$, $y \in D([0, T], \mathbb{R}^r)$ and $t \in [0, T]$,

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

Then there exists a map

$$\Psi : D([0, T], \mathbb{R}^d) \times D([0, T], \mathbb{R}^r) \times D([0, T], \mathbb{R}^m) \to D([0, T], \mathbb{R}^d),$$

measurable with respect to the Borel σ-algebras of the Skorokhod topology, such that for every $\mathbb{R}^d$-valued càdlàg adapted process F, every $\mathbb{R}^r$-valued càdlàg adapted process G, and every $\mathbb{R}^m$-valued semimartingale H, the process $X := \Psi(F, G, H)$ satisfies $\mathbb{P}$-almost surely the SDE

$$X_t = F_t + \int_0^t g(s-, X, G) \mathrm{d}H_s, \qquad t \in [0, T].$$

If Y is another solution of the SDE, then $\mathbb{P}(X = Y) = 1$.

By condition (iii) the map Ψ is non-anticipative: $\Psi(F, G, H)^{\wedge t}$ depends on (F, G, H) only through $(F^{\wedge t}, G^{\wedge t}, H^{\wedge t})$, so the induced solution is adapted. The pathwise integral map of Definition 4 arises as the special case $F \equiv 0$ and $g(s-, X, G) = G(s-)$.

Corollary A.1 (Pathwise Itô integral map)**.** *There exists an adapted map*

$$\Psi : D([0, T], \mathbb{R}^m) \times D([0, T], \mathbb{R}^m) \to D([0, T], \mathbb{R}),$$

measurable with respect to the Borel σ-algebras of the Skorokhod topology, such that for every $\mathbb{R}^m$-valued càdlàg adapted process G and every $\mathbb{R}^m$-valued semimartingale H, the process $\Psi(G, H)$ is a version of the Itô integral $\int_0^{\cdot} G(s-) \mathrm{d}H(s)$.

B Proofs of Section 3

Theorem 1. *Let $\mathcal{D}$ be a system of causal SDEs. If $\mathcal{D}$ is essentially uniquely solvable w.r.t. V with solution function $I^{[V]} : \mathcal{X}_W \to \mathcal{X}_V$, then the observational distribution of $\mathcal{D}$ is unique and satisfies*

$$\mathbb{P}(X_V) = \mathbb{P}(I^{[V]}(X_W)).$$

If $L, O, S \subseteq V$ partition V and $\mathcal{D}$ is essentially uniquely solvable w.r.t. $O \subseteq V$ with solution function $I^{[O]} : \mathcal{X}_S \times \mathcal{X}_L \times \mathcal{X}_W \to \mathcal{X}_O$, then for any $x_S \in \mathcal{X}_S$ the intervened system $\mathcal{D}_{\mathrm{do}(X_S=x_S)}$ is essentially uniquely solvable w.r.t. O with solution function $I^{[O]}(x_S, \cdot, \cdot) : \mathcal{X}_L \times \mathcal{X}_W \to \mathcal{X}_O$. If $L = \emptyset$, the interventional distribution satisfies

$$\mathbb{P}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = \mathbb{P}(I^{[O]}(x_S, X_W)),$$

is unique, and $x_S \mapsto \mathbb{P}(X_O \,|\, \mathrm{do}(X_S = x_S))$ is a Markov kernel.

Proof. Let Φ_O be as in Definition 4 and $I^{[V]}$ a solution function of $\mathcal{D}$ w.r.t. V. Since $X_V := I^{[V]}(X_W)$ satisfies $X_V = \Phi_V(X_V, X_W)$ $\mathbb{P}$-a.s., it is an adapted solution of $\mathcal{D}$. Let $\varphi_V : \mathcal{X}_W \to \mathcal{X}_V$ be an adapted measurable map with $\varphi_V(X_W) = \Phi_V(\varphi_V(X_W), X_W)$ $\mathbb{P}$-a.s. By essential unique solvability we have $\varphi_V = I^{[V]}(X_W)$ $\mathbb{P}$-a.s. The law of any adapted solution therefore equals $\mathbb{P}(I^{[V]}(X_W)) = \mathbb{P}_{\mathcal{D}}(\varphi_V(X_W))$, independent of the choice of $I^{[V]}$.

For the interventional case, let $L, O, S \subseteq V$ partition V and $x_S \in \mathcal{X}_S$. Essential unique solvability of $\mathcal{D}$ w.r.t. O provides the fixed-point and uniqueness conditions of Definition 5 for every adapted input process $\varphi_{L\cup S} : \mathcal{X}_W \to \mathcal{X}_{L\cup S}$. Considering instead the map $\varphi^*_{L\cup S} = (\varphi_L, x_S)$ leaves both conditions intact: for each such φ_L the process $X_O := I^{[O]}(x_S, \varphi_L(X_W), X_W)$ is the $\mathbb{P}$-a.s. unique adapted solution of $X_O = \Phi_O(X_O, x_S, \varphi_L(X_W), X_W)$, which is exactly the O-fixed-point equation of $\mathcal{D}_{\mathrm{do}(X_S=x_S)}$. Hence $I^{[O]}(x_S, \cdot, \cdot)$ is a solution function of $\mathcal{D}_{\mathrm{do}(X_S=x_S)}$ w.r.t. O. Taking $L = \emptyset$, the fully intervened system $\mathcal{D}_{\mathrm{do}(X_S=x_S)}$ has the $\mathbb{P}$-a.s. unique adapted solution $I^{[O]}(x_S, X_W)$, and hence

$$\mathbb{P}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = \mathbb{P}(I^{[O]}(x_S, X_W)),$$

independent of the choice of $I^{[O]}$.

The map $I^{[O]} : \mathcal{X}_S \times \mathcal{X}_W \to \mathcal{X}_O$ is jointly measurable by Definition 5. For every Borel set $A \subseteq \mathcal{X}_O$, the indicator $(x_S, x_W) \mapsto \mathbb{1}_A(I^{[O]}(x_S, x_W))$ is jointly measurable, so by Tonelli's theorem the integral

$$x_S \mapsto \mathbb{P}_{\mathcal{D}}\big(X_O \in A \,|\, \mathrm{do}(X_S = x_S)\big) = \int_{\mathcal{X}_W} \mathbb{1}_A\big(I^{[O]}(x_S, x_W)\big) \mathrm{d}\mathbb{P}(X_W)(x_W)$$

is measurable in x_S. Since $A \mapsto \mathbb{P}_{\mathcal{D}}(X_O \in A \mid \mathrm{do}(X_S = x_S))$ is a probability measure on $\mathcal{X}_O$ for each fixed x_S (as a pushforward of $\mathbb{P}(X_W)$), the map $x_S \mapsto \mathbb{P}_{\mathcal{D}}(X_O \,|\, \mathrm{do}(X_S = x_S))$ is a Markov kernel from $\mathcal{X}_S$ to $\mathcal{X}_O$. □

Lemma B.1. *Let $g : [0, T] \times D([0, T], \mathbb{R}^k) \times D([0, T], \mathbb{R}^\ell) \to \mathbb{R}^m$ be measurable, and write $g = (g^{(i)})_{i=1}^m$ with each $g^{(i)} : [0, T] \times D([0, T], \mathbb{R}^k) \times D([0, T], \mathbb{R}^\ell) \to \mathbb{R}$. Then g satisfies the linear-growth condition (4) (resp. the Lipschitz condition (5)) in x given y if and only if each $g^{(i)}$ satisfies the corresponding condition in x given y.*

Proof. '⇒' For each i and every (t, x, y), $|g^{(i)}(t, x, y)| \le \|g(t, x, y)\| \le K(t, y)\big(1+\sup_{0\le s\le t} \|x(s)\|\big)$, so $g^{(i)}$ satisfies (4) with the same K; the Lipschitz analogue is identical.

'⇐' Suppose each $g^{(i)}$ satisfies (4) with measurable càdlàg K_i. Then

$$\|g(t, x, y)\|^2 = \sum_{i=1}^m \big(g^{(i)}(t, x, y)\big)^2 \le \sum_{i=1}^m K_i(t, y)^2 \big(1 + \sup_{0\le s\le t} \|x(s)\|\big)^2,$$

so $\|g(t,x,y)\| \le K(t,y)\big(1+\sup_{0\le s\le t}\|x(s)\|\big)$ with $K(t,y) := \big(\sum_{i=1}^m K_i(t,y)^2\big)^{1/2}$, which is measurable and càdlàg in (t,y) as a finite combination of measurable càdlàg functions. The Lipschitz analogue is identical, with the same K. □

Theorem 2. *Let $\mathcal{D}$ be a system of causal SDEs and let $O \subseteq V$. If for every $v \in O$ we have $\alpha(v) \cap O = \emptyset$ and the integrand g_v satisfies the linear-growth and Lipschitz conditions in $(X_v, X_{\beta(v)\cap O})$ given $X_{\beta(v)\setminus O}$, then $\mathcal{D}$ is essentially uniquely solvable w.r.t. O.*

Proof. Since $\alpha(v) \cap O = \emptyset$ for every $v \in O$, the joint SDE for X_O reads

$$X_O(t) = f_O\big(t, X_{\alpha(O)}\big) + \int_0^t g_O\big(s-, X_O, X_{\beta(O)\setminus O}\big)\mathrm{d}h_O\big(s, X_{\gamma(O)}\big), \tag{11}$$

with $\alpha(O) \subseteq (V \setminus O) \cup W$ (the assumption $\alpha(v) \cap O = \emptyset$), $\beta(O) \setminus O \subseteq (V \setminus O) \cup W$, and $\gamma(O) \subseteq W$ (Definition 1). The componentwise Lipschitz and linear-growth conditions on g_v in $(X_v, X_{\beta(v)\cap O})$ given $X_{\beta(v)\setminus O}$ extend trivially (by ignoring extra coordinates) to componentwise conditions on g_v in X_O given $X_{\beta(O)\setminus O}$, with the same K_v. Lemma B.1 then gives joint Lipschitz and linear-growth conditions on $g_O := (g_v)_{v\in O}$ in X_O given $X_{\beta(O)\setminus O}$, with joint constant $K_O(t,y) := \big(\sum_{v\in O} K_v(t, y_{\beta(v)\setminus O})^2\big)^{1/2}$. Theorem A.1 then yields a measurable map $I^{[O]}$ such that $X_O = I^{[O]}(X_{\alpha(O)}, X_{\beta(O)\setminus O}, X_{\gamma(O)})$ is the $\mathbb{P}$-a.s. unique adapted solution of (11). Since this holds for every X_W-measurable adapted càdlàg process $X_{(\alpha(O)\cup(\beta(O)\setminus O))\cap V}$, $I^{[O]}$ is thus a solution function as in Definition 5, and $\mathcal{D}$ is essentially uniquely solvable w.r.t. O. □

Lemma B.2. *Let $\mathcal{D}$ be a system of causal SDEs and $O \subseteq V$ such that $\mathcal{D}$ is essentially uniquely solvable w.r.t. O. Then there exists an adapted solution function $\tilde{I}^{[O]} : \mathcal{X}_{\mathrm{pa}(O)} \times \mathcal{X}_W \to \mathcal{X}_O$.*

Proof. By essential unique solvability of $\mathcal{D}$ w.r.t. O there exists an adapted solution function $I^{[O]} : \mathcal{X}_S \times \mathcal{X}_W \to \mathcal{X}_O$ where $S := V \setminus O$. Define $L := (V \setminus O) \setminus \mathrm{pa}(O)$, fix any constant $x_L^0 \in \mathcal{X}_L$, and define

$$\tilde{I}^{[O]}(x_{\mathrm{pa}(O)}, x_W) := I^{[O]}(x_{\mathrm{pa}(O)}, x_L^0, x_W),$$

which is measurable and inherits adaptedness from $I^{[O]}$.

Let φ_S be adapted and measurable and set $\varphi_S^0 := (\varphi_{\mathrm{pa}(O)}, x_L^0)$. Since none of f_O, g_O, h_O depend on x_L, Φ_O does not depend on x_L. In particular, for every x_O and x_W,

$$\Phi_O\big(x_O, \varphi_S(x_W), x_W\big) = \Phi_O\big(x_O, \varphi_S^0(x_W), x_W\big).$$

Since $\tilde{I}^{[O]}(\varphi_{\mathrm{pa}(O)}(x_W), x_W) = I^{[O]}(\varphi_S^0(x_W), x_W)$ is an essentially unique fixed point of $x_O = \Phi_O\big(x_O, \varphi_S^0(x_W), x_W\big)$ and thus also for $x_O = \Phi_O\big(x_O, \varphi_S(x_W), x_W\big)$, we obtain the result. □

Lemma B.3 (Composition of SCC solution functions)**.** *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every strongly connected component (SCC) of $G(\mathcal{D})$. Then $\mathcal{D}$ is essentially uniquely solvable w.r.t. V.*

Proof. Let $(C_1, \dots, C_k)$ be a topological ordering of the SCCs of $G(\mathcal{D})$, and for each i let $I^{[C_i]}$ be a solution function of $\mathcal{D}$ w.r.t. C_i. By Lemma B.2 the solution function $I^{[C_i]}$ may be taken to depend only on X_W and the endogenous parents $X_{\mathrm{pa}(C_i)}$ that all lie in $C_1 \cup \dots \cup C_{i-1}$. Writing $I^{[C_i]}(x_{\mathrm{pa}(C_i)}, x_W)$ accordingly, define $I^{[V]} : \mathcal{X}_W \to \mathcal{X}_V$ recursively along the topological order by

$$I^{[V]}_{C_i}(x_W) := I^{[C_i]}\big(I^{[V]}_{\mathrm{pa}(C_i)}(x_W), x_W\big), \qquad i = 1, \dots, k.$$

We verify by induction on the SCCs that $I^{[V]}$ is an essentially unique solution function of $\mathcal{D}$ w.r.t. V. To that end, let $\varphi_V : \mathcal{X}_W \to \mathcal{X}_V$ be adapted with $\varphi_V(X_W) = \Phi_V(\varphi_V(X_W), X_W)$ $\mathbb{P}$-a.s. Since $I^{[V]}_{C_1}(x_W) = I^{[C_1]}(x_W)$ we have

$$I^{[V]}_{C_1}(X_W) = \Phi_{C_1}\big(I^{[V]}_{C_1}(X_W), X_W\big)$$

by the fixed-point property of $I^{[C_1]}$ and $\varphi_{C_1}(X_W) = I^{[V]}_{C_1}(X_W)$ $\mathbb{P}$-a.s. by the essential uniqueness of $I^{[C_1]}$. If for $i \in \{2, ..., k\}$ we have $I^{[V]}_{C_j}(X_W) = \Phi_{C_j}\big(I^{[V]}_{C_j}(X_W), X_W\big)$ and $\varphi_{C_j}(X_W) = I^{[V]}_{C_j}(X_W)$ $\mathbb{P}$-a.s. for all $j < i$, then, since $\mathrm{pa}(C_i) \subseteq C_1 \cup \dots \cup C_{i-1}$ and $I^{[V]}_{C_i}(X_W) = I^{[C_i]}\big(I^{[V]}_{\mathrm{pa}(C_i)}(X_W), X_W\big)$, the fixed-point property of $I^{[C_i]}$ gives

$$I^{[V]}_{C_i}(X_W) = \Phi_{C_i}\big(I^{[V]}_{C_i}(X_W), I^{[V]}_{\mathrm{pa}(C_i)}(X_W), X_W\big) \quad \mathbb{P}\text{-a.s.}$$

Since $\varphi_{C_i}(X_W) = \Phi_{C_i}(\varphi_{C_i}(X_W), \varphi_{\mathrm{pa}(C_i)}(X_W), X_W)$ $\mathbb{P}$-a.s., substituting the essential uniqueness $\varphi_{\mathrm{pa}(C_i)}(X_W) = I^{[V]}_{\mathrm{pa}(C_i)}(X_W)$ from the induction hypothesis, and since we have $I^{[V]}_{C_i}(x_W) = I^{[C_i]}\big(I^{[V]}_{\mathrm{pa}(C_i)}(x_W), x_W\big)$, we obtain by the essential uniqueness of $I^{[C_i]}$ that

$$\varphi_{C_i}(X_W) = \Phi_{C_i}\big(\varphi_{C_i}(X_W), I^{[V]}_{\mathrm{pa}(C_i)}(X_W), X_W\big) = I^{[V]}_{C_i}(X_W) \quad \mathbb{P}\text{-a.s.},$$

which proves the result. $\square$

Theorem 3. *If a system of causal SDEs $\mathcal{D}$ satisfies Assumption 1, then it is essentially uniquely solvable w.r.t. every subset $O \subseteq V$.*

Proof. Fix $O \subseteq V$ and write $S := V \setminus O$. By Definition 5, a solution function of $\mathcal{D}$ w.r.t. O is a single measurable adapted map $I^{[O]} : \mathcal{X}_S \times \mathcal{X}_W \to \mathcal{X}_O$ such that, for every measurable adapted $\varphi_S : \mathcal{X}_W \to \mathcal{X}_S$, the process $I^{[O]}(\varphi_S(X_W), X_W)$ is the essentially-unique adapted solution of

$$x_O = \Phi_O\big(x_O, \varphi_S(X_W), X_W\big).$$

This can equivalently be interpreted as the fixed-point equation of the *intervened system* $\mathcal{D}_{\mathrm{do}(X_S=\varphi_S(X_W))}$ w.r.t. its full endogenous set O: the system of causal SDEs on O obtained by overriding the mechanisms of the variables in S with the adapted process $\varphi_S(X_W)$, leaving each mechanism Φ_v ($v \in O$) unchanged. Thus $\mathcal{D}$ is essentially uniquely solvable w.r.t. O if and only if there is an essentially unique solution function $I^{[O]}$ for $\mathcal{D}_{\mathrm{do}(X_S=\varphi_S(X_W))}$ w.r.t. O for every adapted φ_S; we establish the latter by verifying the conditions of Theorem 2 on each SCC of $G(\mathcal{D}_{\mathrm{do}(X_S=\varphi_S(X_W))})$ and composing via Lemma B.3.

Since φ_S is a function only of the exogenous X_W, overriding X_S by $\varphi_S(X_W)$ adds only exogenous dependence to the O-mechanisms and leaves the directed edges among O exactly those of $G(\mathcal{D})$; hence the SCCs of $G(\mathcal{D}_{\mathrm{do}(X_S=\varphi_S(X_W))})$ are those of the subgraph of $G(\mathcal{D})$ induced on O, each contained in an SCC $\bar{C}$ of $G(\mathcal{D})$. Fix such an SCC $C \subseteq \bar{C}$. By Assumption 1(i) on $\bar{C}$, $\alpha(v) \cap C \subseteq \alpha(v) \cap \bar{C} = \emptyset$ for every $v \in C$; and by Assumption 1(ii) on $\bar{C}$, g_v satisfies the linear-growth and Lipschitz conditions in $(X_v, X_{\beta(v)\cap\bar{C}})$ given $X_{\beta(v)\setminus\bar{C}}$, which implies the conditions in $(X_v, X_{\beta(v)\cap C})$ given $X_{\beta(v)\setminus C}$ by considering the variables $\beta(v) \cap (\bar{C} \setminus C)$ as the given inputs. Hence C satisfies the hypotheses of Theorem 2, and thus $\mathcal{D}_{\mathrm{do}(X_S=\varphi_S(X_W))}$ is essentially uniquely solvable w.r.t. C. Since the resulting solution function $I^{[O]}$ does not depend on φ_S we obtain the result. $\square$

Theorem 4. *If a system of causal SDEs $\mathcal{D}$ satisfies Assumption 2, then it satisfies Assumption 1.*

Proof. For each $v \in V$, writing $\Sigma_{v\cdot}$ for the (constant) v-th row of Σ and $\gamma(v) := \{w \in \gamma : \Sigma_{vw} \neq 0\}$ for the Brownian components driving v, the causal SDE

$$X_v(t) = X^0_{\alpha_v} + \int_0^t \mu_v(s, X_V(s))\mathrm{d}s + \int_0^t \Sigma_{v\cdot}\mathrm{d}X_{\gamma(v)}(s)$$

matches the form $f_v(t, X_{\alpha(v)}) + \int_0^t g_v(s-, X_v, X_{\beta(v)})\mathrm{d}h_v(s, X_{\gamma(v)})$ of (2) with $\alpha(v) := \{\alpha_v\} \subseteq \alpha \subseteq W$ the exogenous initial-condition coordinate, $\beta(v) := V \setminus \{v\}$, $\gamma(v) \subseteq \gamma \subseteq W$ the exogenous

Brownian coordinates driving v, $f_v(t, X_{\alpha(v)}) := X^0_{\alpha_v}$, $g_v(s, X_v, X_{\beta(v)}) := (\mu_v(s, X_V(s)), \Sigma_{v\cdot})$, and $h_v(s, X_{\gamma(v)}) := (s, X_{\gamma(v)}(s))$.

For every $v \in V$, $\alpha(v) \subseteq W$, so v has no endogenous functional parents, and thus $\alpha(v) \cap S = \emptyset$ for every SCC S and every $v \in S$. The drift component $\mu_v(s, X_V(s))$ inherits the linear-growth and Lipschitz conditions from μ by restriction. The constant component $\Sigma_{v\cdot}$ of g_v has no X-dependence and does not contribute to the Lipschitz constant, and adds a term $\|\Sigma_{v\cdot}\|$ to the linear-growth constant. □

C Proofs of Section 4

Theorem 5. *Let $\mathcal{D}$ be a system of causal SDEs, and let $L, O \subseteq V$ be disjoint. If $\mathcal{D}$ is essentially uniquely solvable w.r.t. L and w.r.t. $L \cup O$, then the marginalised system $\mathcal{D}_{\mathrm{marg}(L)}$ is essentially uniquely solvable w.r.t. O. Moreover, for any solution function $I^{[L\cup O]}$ of $\mathcal{D}$ w.r.t. $L \cup O$, its restriction $I^{[L\cup O]}_O$ is a solution function of $\mathcal{D}_{\mathrm{marg}(L)}$ w.r.t. O.*

Proof. Fix the solution function $I^{[L]} : \mathcal{X}_{V\setminus L} \times \mathcal{X}_W \to \mathcal{X}_L$ used in Definition 6, and write $S := V \setminus (L \cup O)$. Since the integrator h_v of each $v \in V \setminus L$ depends only on the exogenous $X_{\gamma(v)}$, marginalising L leaves it unchanged and substitutes $I^{[L]}$ only into f_v, g_v; consequently the pathwise mechanism of $\mathcal{D}_{\mathrm{marg}(L)}$ satisfies

$$\tilde{\Phi}_O(x_O, x_S, x_W) = \Phi_O\big(I^{[L]}(x_S, x_O, x_W), x_O, x_S, x_W\big). \tag{12}$$

Fix a solution function $I^{[L\cup O]}$ of $\mathcal{D}$ w.r.t. $L \cup O$. We will show that $\tilde{I}^{[O]} := I^{[L\cup O]}_O$ is an essentially unique solution function for $\mathcal{D}_{\mathrm{marg}(L)}$. To this end, let $\varphi_S : \mathcal{X}_W \to \mathcal{X}_S$ be measurable and adapted.

By the fixed-point property of $I^{[L\cup O]}$ at φ_S we have

$$I^{[L\cup O]}(\varphi_S(X_W), X_W) = \Phi_{L\cup O}\big(I^{[L\cup O]}(\varphi_S(X_W), X_W), \varphi_S(X_W), X_W\big) \quad \mathbb{P}(X_W)\text{-a.s.}$$

Reading off the L-block and regrouping its O- and S-arguments into the map $\varphi_{S\cup O}(x_W) := (\varphi_S(x_W), \tilde{I}^{[O]}(\varphi_S(x_W), x_W))$ we get that $I^{[L\cup O]}_L(\varphi_S(\cdot), \cdot)$ satisfies the fixed-point property

$$I^{[L\cup O]}_L(\varphi_S(X_W), X_W) = \Phi_L\big(I^{[L\cup O]}_L(\varphi_S(X_W), X_W), \varphi_{S\cup O}(X_W), X_W\big) \quad \mathbb{P}(X_W)\text{-a.s.}$$

Since $\mathcal{D}$ is essentially uniquely solvable w.r.t. L, the uniqueness property of $I^{[L]}$ at $\varphi_{S\cup O}$ then gives

$$I^{[L\cup O]}_L(\varphi_S(x_W), x_W) = I^{[L]}(\varphi_{S\cup O}(x_W), x_W) \quad \mathbb{P}(X_W)\text{-a.s.} \tag{13}$$

Now the O-component of the fixed-point property of $I^{[L\cup O]}$ at φ_S reads, using $\tilde{I}^{[O]} = I^{[L\cup O]}_O$ and $I^{[L\cup O]}(\varphi_S(X_W), X_W) = (I^{[L\cup O]}_L(\varphi_S(X_W), X_W), \tilde{I}^{[O]}(\varphi_S(X_W), X_W))$, as $\mathbb{P}(X_W)$-a.s.:

$$\tilde{I}^{[O]}(\varphi_S(X_W), X_W) = \Phi_O\big((I^{[L\cup O]}_L(\varphi_S(X_W), X_W),\ \tilde{I}^{[O]}(\varphi_S(X_W), X_W)),\ \varphi_S(X_W), X_W\big).$$

By (13), $I^{[L\cup O]}_L(\varphi_S(X_W), X_W) = I^{[L]}(\varphi_{S\cup O}(X_W), X_W) = I^{[L]}((\varphi_S(X_W), \tilde{I}^{[O]}(\varphi_S(X_W), X_W)), X_W)$; substituting this for the L-coordinate, the right-hand side is precisely (12) evaluated at $x_O = \tilde{I}^{[O]}(\varphi_S(X_W), X_W)$ and $x_S = \varphi_S(X_W)$. Hence, $\mathbb{P}(X_W)$-a.s.,

$$\tilde{I}^{[O]}(\varphi_S(x_W), x_W) = \tilde{\Phi}_O\big(\tilde{I}^{[O]}(\varphi_S(x_W), x_W), \varphi_S(x_W), x_W\big).$$

For essential uniqueness, let $\varphi_O : \mathcal{X}_W \to \mathcal{X}_O$ be adapted and measurable with $\varphi_O(x_W) = \tilde{\Phi}_O\big(\varphi_O(x_W), \varphi_S(x_W), x_W\big)$ for $\mathbb{P}(X_W)$-a.a. x_W. Redefine $\varphi_{S\cup O} := (\varphi_S, \varphi_O)$ and set

$$\varphi_{L\cup O}(x_W) := \big(I^{[L]}(\varphi_{S\cup O}(x_W), x_W), \varphi_O(x_W)\big).$$

Then $\varphi_{L\cup O}$ is a fixed point of $x_{L\cup O} = \Phi_{L\cup O}(x_{L\cup O}, \varphi_S(X_W), X_W)$ $\mathbb{P}(X_W)$-a.s., since the L-component satisfies

$$\Phi_L(\varphi_{L\cup O}(x_W), \varphi_S(x_W), x_W) = \Phi_L\big(I^{[L]}(\varphi_{S\cup O}(x_W), x_W), \varphi_{S\cup O}(x_W), x_W\big) = I^{[L]}(\varphi_{S\cup O}(x_W), x_W)$$

by the fixed-point property of $I^{[L]}$ at $\varphi_{S\cup O}$, and the O-component satisfies $\Phi_O(\varphi_{L\cup O}(x_W), \varphi_S(x_W), x_W) = \tilde{\Phi}_O\big(\varphi_O(x_W), \varphi_S(x_W), x_W\big) = \varphi_O(x_W)$, by (12) and the hypothesis. Since $\varphi_{L\cup O}$ is adapted and $\mathcal{D}$ is essentially uniquely solvable w.r.t. $L \cup O$, the essential uniqueness of $I^{[L\cup O]}$ at φ_S gives $\varphi_{L\cup O}(X_W) = I^{[L\cup O]}(\varphi_S(X_W), X_W)$ $\mathbb{P}(X_W)$-a.s. Reading off the O-component, we have

$$\varphi_O(X_W) = I_O^{[L\cup O]}(\varphi_S(X_W), X_W) = \tilde{I}^{[O]}(\varphi_S(X_W), X_W)$$

$\mathbb{P}(X_W)$-a.s. □

Lemma C.1. *Let $\mathcal{D}$ be a system of causal SDEs, and let $L, O, S \subseteq V$ partition V. If $\mathcal{D}$ is essentially uniquely solvable w.r.t. L and $L \cup O$, then*

$$\mathbb{P}_{\mathcal{D}_{\mathrm{marg}(L)}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = \mathbb{P}_{\mathcal{D}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big),$$

and any two marginalisations $\mathcal{D}_{\mathrm{marg}(L)}$ and $\mathcal{D}'_{\mathrm{marg}(L)}$ yield the same distributions.

Proof. Let $I^{[L\cup O]} : \mathcal{X}_S \times \mathcal{X}_W \to \mathcal{X}_{L\cup O}$ be a solution function of $\mathcal{D}$ w.r.t. $L \cup O$, and set $\tilde{I}^{[O]}(x_S, x_W) := I_O^{[L\cup O]}(x_S, x_W)$. By Theorem 5, $\mathcal{D}_{\mathrm{marg}(L)}$ is essentially uniquely solvable w.r.t. O and $\tilde{I}^{[O]}$ is a solution function of $\mathcal{D}_{\mathrm{marg}(L)}$ w.r.t. O. By Theorem 1 the interventional distribution of $\mathcal{D}_{\mathrm{marg}(L)}$ is the pushforward via any solution function. Hence

$$\mathbb{P}_{\mathcal{D}_{\mathrm{marg}(L)}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = \mathbb{P}(\tilde{I}^{[O]}(x_S, X_W)) = \mathbb{P}(I_O^{[L\cup O]}(x_S, X_W)),$$

which by Theorem 1 applied to $\mathcal{D}$ equals $\mathbb{P}_{\mathcal{D}}(X_O \,|\, \mathrm{do}(X_S = x_S))$. □

Lemma C.2 (Sequential marginalisation)**.** *Let $\mathcal{D}$ be a system of causal SDEs, let $L_1, L_2, O, S \subseteq V$ partition V. Assume that $\mathcal{D}$ is essentially uniquely solvable w.r.t. L_1, $L_1 \cup L_2$, and $L_1 \cup L_2 \cup O$. Then $\mathcal{D}_{\mathrm{marg}(L_1\cup L_2)}$ and $(\mathcal{D}_{\mathrm{marg}(L_1)})_{\mathrm{marg}(L_2)}$ are well-defined and essentially uniquely solvable w.r.t. O, and for every $x_S \in \mathcal{X}_S$,*

$$\mathbb{P}_{(\mathcal{D}_{\mathrm{marg}(L_1)})_{\mathrm{marg}(L_2)}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = \mathbb{P}_{\mathcal{D}_{\mathrm{marg}(L_1\cup L_2)}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big).$$

Proof. Let $I^{[L_1\cup L_2\cup O]}$ be a solution function of $\mathcal{D}$ w.r.t. $L_1 \cup L_2 \cup O$. Throughout we apply Theorem 5, whose two hypotheses on a system are essential unique solvability w.r.t. the marginalised set and w.r.t. its union with the retained set. Applied to $\mathcal{D}$ with marginalised set $L_1 \cup L_2$ and retained set O, it gives that $\mathcal{D}_{\mathrm{marg}(L_1\cup L_2)}$ is essentially uniquely solvable w.r.t. O with solution function $I_O^{[L_1\cup L_2\cup O]}$. Applied to $\mathcal{D}$ with marginalised set L_1 and retained set L_2, it gives that $\mathcal{D}_{\mathrm{marg}(L_1)}$ is essentially uniquely solvable w.r.t. L_2, so $(\mathcal{D}_{\mathrm{marg}(L_1)})_{\mathrm{marg}(L_2)}$ is defined; applied to $\mathcal{D}$ with marginalised set L_1 and retained set $L_2 \cup O$, it gives that $I_{L_2\cup O}^{[L_1\cup L_2\cup O]}$ is a solution function of $\mathcal{D}_{\mathrm{marg}(L_1)}$ w.r.t. $L_2 \cup O$. Applied once more to $\mathcal{D}_{\mathrm{marg}(L_1)}$ with marginalised set L_2 and retained set O, it yields $I_O^{[L_1\cup L_2\cup O]}$ as a solution function of $(\mathcal{D}_{\mathrm{marg}(L_1)})_{\mathrm{marg}(L_2)}$ w.r.t. O.

Both $\mathcal{D}_{\mathrm{marg}(L_1\cup L_2)}$ and $(\mathcal{D}_{\mathrm{marg}(L_1)})_{\mathrm{marg}(L_2)}$ therefore admit $I_O^{[L_1\cup L_2\cup O]}$ as a solution function w.r.t. O, and by Theorem 1 the interventional distributions coincide as pushforwards of $\mathbb{P}(X_W)$ under this map. □

Lemma C.3 (Marginalisation commutes with intervention)**.** *Let $\mathcal{D}$ be a system of causal SDEs, let $L, O, S, T \subseteq V$ partition V, and let $x_T \in \mathcal{X}_T$. Assume $\mathcal{D}$ is essentially uniquely solvable w.r.t. L and w.r.t. $L \cup O$. Then $(\mathcal{D}_{\mathrm{marg}(L)})_{\mathrm{do}(X_T=x_T)}$ and $(\mathcal{D}_{\mathrm{do}(X_T=x_T)})_{\mathrm{marg}(L)}$ are well-defined and essentially uniquely solvable w.r.t. O, and for every $x_S \in \mathcal{X}_S$,*

$$\mathbb{P}_{(\mathcal{D}_{\mathrm{marg}(L)})_{\mathrm{do}(X_T=x_T)}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = \mathbb{P}_{(\mathcal{D}_{\mathrm{do}(X_T=x_T)})_{\mathrm{marg}(L)}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big).$$

Proof. Let $I^{[L\cup O]} : \mathcal{X}_{T\cup S} \times \mathcal{X}_W \to \mathcal{X}_{L\cup O}$ be a solution function of $\mathcal{D}$ w.r.t. $L \cup O$.

By Theorem 1, since $\mathcal{D}$ is essentially uniquely solvable w.r.t. L, $\mathcal{D}_{\mathrm{do}(X_T=x_T)}$ is essentially uniquely solvable w.r.t. L, and thus the marginalisation $(\mathcal{D}_{\mathrm{do}(X_T=x_T)})_{\mathrm{marg}(L)}$ is well-defined. Similarly $\mathcal{D}$ and thus also $\mathcal{D}_{\mathrm{do}(X_T=x_T)}$ are essentially uniquely solvable w.r.t. $L \cup O$, where the latter has solution function $I^{[L\cup O]}(x_T, \cdot, \cdot) : \mathcal{X}_S \times \mathcal{X}_W \to \mathcal{X}_{L\cup O}$. By Theorem 5, $(\mathcal{D}_{\mathrm{do}(X_T=x_T)})_{\mathrm{marg}(L)}$ is then essentially uniquely solvable w.r.t. O with solution function $I^{[L\cup O]}_O(x_T, \cdot, \cdot)$.

By Theorem 5, $\mathcal{D}_{\mathrm{marg}(L)}$ is essentially uniquely solvable w.r.t. O with solution function $I^{[L\cup O]}_O : \mathcal{X}_{T\cup S} \times \mathcal{X}_W \to \mathcal{X}_O$. By Theorem 1, the system $(\mathcal{D}_{\mathrm{marg}(L)})_{\mathrm{do}(X_T=x_T)}$ is essentially uniquely solvable w.r.t. O with solution function $I^{[L\cup O]}_O(x_T, \cdot, \cdot) : \mathcal{X}_S \times \mathcal{X}_W \to \mathcal{X}_O$.

Both systems therefore admit $I^{[L\cup O]}_O(x_T, \cdot, \cdot) : \mathcal{X}_S \times \mathcal{X}_W \to \mathcal{X}_O$ as a solution function w.r.t. O, so by Theorem 1 their interventional distributions under $\mathrm{do}(X_S = x_S)$ coincide, both being the pushforward of $\mathbb{P}(X_W)$ under $I^{[L\cup O]}_O(x_T, x_S, \cdot)$. □

D Proofs of Section 5

Theorem D.1. *Let $G = (V, E)$ be a directed graph and let $G^- \subseteq G$ be the directed graph with all self-loops $v \to v$ removed. Let $A, B, C \subseteq V$, then we have*

$$A \overset{d}{\underset{G}{\perp}} B \,|\, C \iff A \overset{d}{\underset{G^-}{\perp}} B \,|\, C$$
$$A \overset{\sigma}{\underset{G}{\perp}} B \,|\, C \iff A \overset{\sigma}{\underset{G^-}{\perp}} B \,|\, C$$

Proof. Since G^- is a subgraph of G, any separation in G immediately implies a separation in G^-.

For the implication $A \not\perp^d_G B \,|\, C \implies A \not\perp^d_{G^-} B \,|\, C$ we follow the proof of Proposition 3.5 in Mogensen and Hansen (2020). Let π be an active walk in G. For any self-loop at v on the walk, let π' be the walk in which this self-loop is removed. If $v \in A$ or $v \in B$, then the walk π' is still active. Otherwise, if v is a non-collider in π', we must have had one of $\ldots \to v \to v \to \ldots$, $\ldots \to v \leftarrow v \to \ldots$, $\ldots \leftarrow v \to v \to \ldots$, $\ldots \leftarrow v \leftarrow v \to \ldots$, $\ldots \leftarrow v \to v \leftarrow \ldots$ or $\ldots \leftarrow v \leftarrow v \leftarrow \ldots$ in π; in each of these the occurrence of v at the tail of the self-loop is a non-collider, so we must have had $v \notin C$ for π to be active, so π' is active. If v is a collider in π', then in π we either had $\ldots \to v \leftarrow v \leftarrow \ldots$ or $\ldots \to v \to v \leftarrow \ldots$, so $v \in \mathrm{Anc}(C)$ for these walks to be open, so π' is open as well. Iterating this for every self-loop on π gives an active walk in G^-, and so we get $A \not\perp^d_{G^-} B \,|\, C$. For σ-separation, the same case analysis applies. The only additional consideration is blockability of non-colliders: a non-collider v is blockable if it has a child on the walk not in the same SCC. Removing self-loops does not affect the SCCs, and therefore the SCCs of G^- are identical to those of G. In particular, blockability of non-colliders is the same in G and G^-, so the argument carries over. □

D.1 Proofs of Section 5.1

We carry out the proof of Theorem 6 via the acyclification strategy of Forré and Mooij (2017); Bongers et al. (2021); Forré and Mooij (2025) adapted to the essentially uniquely solvable setting.

Definition D.1 (Acyclification). Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every SCC C of $G(\mathcal{D})$. An *acyclification* is the SCM $\mathcal{D}^{\mathrm{acy}} = \Big(V, W, \mathcal{X}_V, \mathcal{X}_W, \tilde{\Phi}, \mathbb{P}(X_W)\Big)$ with sample spaces $\mathcal{X}_v = D([0,T], \mathbb{R})$ and, for each SCC C of $G(\mathcal{D})$ and each $v \in C$, structural equation

$$x_v = I_v^{[C]}(x_{\mathrm{pa}(C)}, x_W),$$

where $I^{[C]}$ is an adapted solution function from Lemma B.2.

By construction, the acyclification $\mathcal{D}^{\mathrm{acy}}$ is an acyclic SCM (Pearl, 2009), and hence essentially uniquely solvable with respect to V by recursive substitution of the structural equations. Moreover, it satisfies the d-separation Markov property (Bongers et al., 2021, Theorem 6.3).

Recall from (Forré and Mooij, 2025, Definition 3.5.1) that a *graphical acyclification* $\tilde{G}$ of a graph G on V is for example obtained from G by, for each SCC C, replacing all edges within C and all directed edges $u \to v$ into C (with $u \notin C$, $v \in C$) by bidirected edges between every pair of distinct vertices in C and directed edges $u \to v'$ for every $v' \in C$, respectively. The resulting graph is acyclic on the SCCs of G, and by Forré and Mooij (2025), Proposition 3.5.2, d-separation in $\tilde{G}$ corresponds to σ-separation in G: for all $A, B, C \subseteq V$ we have $A \perp^{\sigma}_{G} B \,|\, C \iff A \perp^{d}_{\tilde{G}} B \,|\, C$.

Lemma D.1. *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every SCC of $G(\mathcal{D})$, and $\mathcal{D}^{\mathrm{acy}}$ an acyclification. Then:*

(i) $\mathbb{P}_{\mathcal{D}}(X_V, X_W) = \mathbb{P}_{\mathcal{D}^{\mathrm{acy}}}(X_V, X_W)$*;*

(ii) $G^+(\mathcal{D}^{\mathrm{acy}}) \subseteq \tilde{G}^+(\mathcal{D})$*, the graphical acyclification of* $G^+(\mathcal{D})$*.*

Proof. Let $C_1 < \cdots < C_k$ be a topological ordering of the DAG of SCCs of $G(\mathcal{D})$.

(i) Define $I^{[V]} : \mathcal{X}_W \to \mathcal{X}_V$ recursively along the topological order of SCCs:

$$I^{[V]}_{C_i}(x_W) := I^{[C_i]}\big(I^{[V]}_{\mathrm{pa}(C_i)}(x_W), x_W\big), \qquad i = 1, \ldots, k.$$

By acyclicity of $G(\mathcal{D}^{\mathrm{acy}})$, this recursive construction yields an adapted solution function of $\mathcal{D}^{\mathrm{acy}}$. As shown in the proof of Lemma B.3 the same $I^{[V]}$ is an essentially unique solution function of $\mathcal{D}$ w.r.t. V. Hence under both $\mathcal{D}$ and $\mathcal{D}^{\mathrm{acy}}$ the endogenous variables are given $\mathbb{P}$-a.s. by the same measurable map $X_V = I^{[V]}(X_W)$ of the common exogenous X_W, so the joint law of (X_V, X_W) is in either case the pushforward of $\mathbb{P}(X_W)$ under $x_W \mapsto (I^{[V]}(x_W), x_W)$; therefore $\mathbb{P}_{\mathcal{D}}(X_V, X_W) = \mathbb{P}_{\mathcal{D}^{\mathrm{acy}}}(X_V, X_W)$.

(ii) Let $v \in V$ and write C for the SCC of v in $G(\mathcal{D})$. By construction, $I_v^{[C]}$ depends only on $x_{\mathrm{pa}(C)}$ and x_W, so any directed edge $u \to v$ in $G^+(\mathcal{D}^{\mathrm{acy}})$ has $u \in \mathrm{pa}(C) \cup W$. If $u \in \mathrm{pa}(C)$, then by definition of $\mathrm{pa}(C)$ there exists $v' \in C$ with $u \to v'$ in $G(\mathcal{D})$, hence $u \to v$ lies in $\tilde{G}^+(\mathcal{D})$. If $u \in W$, then since $I^{[C]}$ is the essentially unique solution of $X_C = \Phi_C(X_C, X_{\mathrm{pa}(C)}, X_W)$, its component $I_v^{[C]}$ can depend on x_u only if some mechanism $\Phi_{v'}$ with $v' \in C$ essentially depends on x_u, that is, $u \to v'$ in $G^+(\mathcal{D})$ for some $v' \in C$; the graphical acyclification therefore puts $u \to v$ in $\tilde{G}^+(\mathcal{D})$. ☐

Theorem 6. *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every SCC of $G(\mathcal{D})$. For all $A, B, C \subseteq V \cup W$ we have*

$$A \underset{G^+(\mathcal{D})}{\overset{\sigma}{\perp}} B \,|\, C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C. \tag{7}$$

Proof. Let $\tilde{G}^+(\mathcal{D})$ be a graphical acyclification of $G^+(\mathcal{D})$. By (Forré and Mooij, 2025, Proposition 3.5.2), $A \perp^{\sigma}_{G^+(\mathcal{D})} B \,|\, C \iff A \perp^{d}_{\tilde{G}^+(\mathcal{D})} B \,|\, C$. By Lemma D.1(ii), $G^+(\mathcal{D}^{\mathrm{acy}}) \subseteq \tilde{G}^+(\mathcal{D})$, and removing edges preserves d-separation so $A \perp^{d}_{\tilde{G}^+(\mathcal{D})} B \,|\, C \implies A \perp^{d}_{G^+(\mathcal{D}^{\mathrm{acy}})} B \,|\, C$. Since $G(\mathcal{D}^{\mathrm{acy}})$

is acyclic, the d-separation global Markov property for acyclic SCMs (Bongers et al., 2021, Theorem 6.3) applies:

$$A \underset{G^+(\mathcal{D}^{\text{acy}})}{\overset{d}{\perp}} B \,|\, C \implies X_A \underset{\mathbb{P}_{\mathcal{D}^{\text{acy}}}}{\perp\!\!\!\perp} X_B \,|\, X_C.$$

By Lemma D.1(i), $\mathcal{D}$ and $\mathcal{D}^{\text{acy}}$ induce the same joint law $\mathbb{P}(X_V, X_W)$, so the conditional independence—which may involve exogenous variables, as $A, B, C \subseteq V \cup W$—transfers to $\mathcal{D}$. $\square$

D.2 Proofs of Section 5.2

Lemma 7. Let $\mathcal{D}$ be the SDE from Assumption 2 and let $\mathcal{M}^{\Delta}_{\mathcal{D}}$ be the Euler SCM (for any $n \in \mathbb{N}$), then for $A, B, C \subseteq V$ we have

$$X_A \underset{G(\mathcal{D})}{\overset{d}{\perp}} X_B \,|\, X_C \implies X^{\Delta}_A \underset{G(\mathcal{M}^{\Delta}_{\mathcal{D}})}{\overset{d}{\perp}} X^{\Delta}_B \,|\, X^{\Delta}_C \implies X^{\Delta}_A \perp\!\!\!\perp X^{\Delta}_B \,|\, X^{\Delta}_C$$

where we write $X^{\Delta}_A := (X^{\Delta}_A(t_0), ..., X^{\Delta}_A(t_n))$.

Proof. Let $A, B, C \subseteq V$ be such that $X^{\Delta}_A \not\perp^d_{G(\mathcal{M}^{\Delta}_{\mathcal{D}})} X^{\Delta}_B \,|\, X^{\Delta}_C$, then we also have a d-connection in the augmented graph $X^{\Delta}_A \not\perp^d_{G^+(\mathcal{M}^{\Delta}_{\mathcal{D}})} X^{\Delta}_B \,|\, X^{\Delta}_C$, so let $\pi^{\Delta} = (v_0^{t_0}, ..., v_n^{t_n})$ be an active path in $G^+(\mathcal{M}^{\Delta}_{\mathcal{D}})$ from X^{Δ}_A to X^{Δ}_B (given X^{Δ}_C), where for ease of notation we write for any component $B_v^{[t_k, t_{k+1}]} = v^{t_{k+1}}$, for $v \in W$. This induces a walk $\pi := (v_0, ..., v_n)$ that is active given C:

- for every collider v on π, there is a t_j such that v^{t_j} is a collider on π^{Δ}, and since $v^{t_j} \in \text{Anc}(X^{\Delta}_C)$, a directed path from v^{t_j} to some c^{t_i} with $c \in C$ in $G^+(\mathcal{M}^{\Delta}_{\mathcal{D}})$ projects onto a directed path from v to c in $G^+(\mathcal{D})$, so $v \in \text{Anc}(C)$;
- for every non-collider v on π, we have
  - $w \to v \to z$ with possibly $w = v$ and possibly $z = v$, and a corresponding non-collider v^{t_j} in $w^{t_{j-1}} \to v^{t_j} \to z^{t_{j+1}}$ we have $v^{t_j} \notin X^{\Delta}_C$ and hence $v \notin C$;
  - $w \leftarrow v \to z$ then either $v \notin V$ so $v \notin C$, or if $v \in V$ then (with possibly either $w = v$ or $z = v$) there is a corresponding non-collider v^{t_j} in $w^{t_{j+1}} \leftarrow v^{t_j} \to z^{t_{j+1}}$ we have $v^{t_j} \notin X^{\Delta}_C$ and hence $v \notin C$;
  - we have $v_0^{t_0} \notin X^{\Delta}_C$ and $v_n^{t_n} \notin X^{\Delta}_C$, and hence $v_0 \notin C$ and $v_n \notin C$.

From π, construct the walk π' in $G^+(\mathcal{D})$ between A and B by replacing every maximal subwalk of the form $v_j \to \cdots \to v_j$ by the single vertex v_j, which retains the boundary edges the subwalk had with the rest of π. Such a subwalk is a directed cycle $v_j \to w_1 \to \cdots \to w_k \to v_j$, visiting v_j at its start and end. Its first occurrence of v_j is the tail of the cycle — incident to the outgoing edge $v_j \to w_1$ — so it is a non-collider in π and $v_j \notin C$; hence if v_j is a non-collider in π', π' stays active there. If v_j is a collider in π' (both boundary edges point into v_j), then its last occurrence is the head of the cycle — incident to the incoming edge $w_k \to v_j$ — and, together with the incoming boundary edge, has two incoming edges, so it is a collider in π, giving $v_j \in \text{Anc}(C)$. The interior cycle vertices are non-colliders on a directed path, hence not in C, so removing them does not affect activeness. (This is the self-loop argument of Theorem D.1 (above), applied to the directed cycle.) Hence π' is active given C, and so $X_A \not\perp^d_{G^+(\mathcal{D})} X_B \,|\, X_C$, or equivalently $X_A \not\perp^d_{G(\mathcal{D})} X_B \,|\, X_C$.

Since $\mathcal{M}^{\Delta}_{\mathcal{D}}$ is acyclic, the d-separation Markov property gives $X^{\Delta}_A \perp\!\!\!\perp X^{\Delta}_B \,|\, X^{\Delta}_C$. $\square$

Lemma D.2. *Let $\mathcal{D}$ satisfy Assumption 2 and fix a grid $t_k = Tk/n$ $(k = 0, \ldots, n)$. For each cell $[t_k, t_{k+1}]$ let $\bar{X}^k_\gamma$ be the Brownian bridge of X_γ,*

$$\bar{X}^k_\gamma(t) := \big(X_\gamma(t) - X_\gamma(t_k)\big) - \tfrac{t-t_k}{t_{k+1}-t_k}\big(X_\gamma(t_{k+1}) - X_\gamma(t_k)\big),$$

and set $\tilde{X}^k_V := \Sigma\bar{X}^k_\gamma$; for $S \subseteq V$ write $\tilde{X}^k_S := (\tilde{X}^k_v)_{v\in S}$ and $\tilde{X}_S := (\tilde{X}^k_S)^{n-1}_{k=0}$. Then the continuous Euler scheme X^n_V of (8) satisfies, for every $S \subseteq V$,

$$\sigma(X^n_S) = \sigma(X^\Delta_S) \vee \sigma(\tilde{X}_S).$$

Moreover, X^Δ_V and $\tilde{X}_V$ are independent.

Proof. Write $\Delta_k X_\gamma := X_\gamma(t_{k+1}) - X_\gamma(t_k)$. At the grid points $X^n_V(t_k) = X^\Delta_V(t_k)$, so on $[t_k, t_{k+1}]$ the continuous Euler scheme (8) reads $X^n_V(t) = X^\Delta_V(t_k) + \mu(t_k, X^\Delta_V(t_k))(t-t_k) + \Sigma\big(X_\gamma(t) - X_\gamma(t_k)\big)$. Substituting the gridpoint recursion $\mu(t_k, X^\Delta_V(t_k))(t_{k+1} - t_k) = X^\Delta_V(t_{k+1}) - X^\Delta_V(t_k) - \Sigma\Delta_k X_\gamma$ and collecting terms,

$$X^n_V(t) = X^\Delta_V(t_k) + \tfrac{t-t_k}{t_{k+1}-t_k}\big(X^\Delta_V(t_{k+1}) - X^\Delta_V(t_k)\big) + \tilde{X}^k_V(t).$$

For $C \subseteq V$, X^n_C is determined by X^Δ_C and $\tilde{X}_C$, and conversely $X^\Delta_C(t_k) = X^n_C(t_k)$ and $\tilde{X}^k_C(t) = X^n_C(t) - X^n_C(t_k) - \tfrac{t-t_k}{t_{k+1}-t_k}\big(X^n_C(t_{k+1}) - X^n_C(t_k)\big)$, so $\sigma(X^n_C) = \sigma(X^\Delta_C, \tilde{X}_C)$.

On each cell the Brownian bridge $\bar{X}^k_\gamma$ and the increment $\Delta_k X_\gamma$ are jointly Gaussian with $\mathrm{Cov}\big(\bar{X}^k_\gamma(t), \Delta_k X_\gamma\big) = (t-t_k)I_d - \tfrac{t-t_k}{t_{k+1}-t_k}(t_{k+1}-t_k)I_d = 0$ for every $t \in [t_k, t_{k+1}]$, hence independent; bridges and increments on disjoint cells are independent by the independence of Brownian increments. Together with the independence of X^0_α and X_γ, the bridges $\tilde{X}_V$ are independent of $\sigma(X^0_\alpha, \{\Delta_k X_\gamma\}_k) \supseteq \sigma(X^\Delta_V)$, i.e. $X^\Delta_V \perp\!\!\!\perp \tilde{X}_V$. □

Lemma D.3. *Let $\mathcal{D}$ satisfy Assumption 2, and let $\tilde{X}_V$ be the Brownian bridges of Lemma D.2. Then for all $A, B, C \subseteq V$,*

$$X_A \underset{G(\mathcal{D})}{\overset{d}{\perp}} X_B \,|\, X_C \implies \tilde{X}_A \perp\!\!\!\perp \tilde{X}_B \,|\, \tilde{X}_C.$$

Proof. Fix an interval k. The components $\bar{X}^k_{\gamma_w}$ $(w \in \gamma)$ of the vector Brownian bridge $\bar{X}^k_\gamma$ are mutually independent and jointly Gaussian (as X_γ has independent components), and $\tilde{X}^k_v = \sum_{w\in\gamma} \Sigma_{vw}\bar{X}^k_{\gamma_w}$. Hence $\tilde{X}^k_V = (\tilde{X}^k_v)_{v\in V}$ is the solution of the acyclic linear structural causal model with mutually independent exogenous variables $(\bar{X}^k_{\gamma_w})_{w\in\gamma}$ and structural equations $\tilde{X}^k_v = \Sigma_{v\cdot}\bar{X}^k_\gamma$, whose latent projection onto V has no directed edges and a bidirected edge $u \leftrightarrow v$ exactly when u and v share a driving component, i.e. $\Sigma_{uw} \neq 0 \neq \Sigma_{vw}$ for some $w \in \gamma$. Since the initial-condition exogenous variables α are in bijection with V and contribute no bidirected edges, these coincide with the bidirected edges of $G(\mathcal{D})$ (Definition 2); as the projection has no further edges, it is a subgraph of $G(\mathcal{D})$. Removing edges preserves d-separation, so $X_A \perp^d_{G(\mathcal{D})} X_B \,|\, X_C$ implies the d-separation of A and B given C in this projection, and the d-separation global Markov property for acyclic SCMs (Bongers et al., 2021, Theorem 6.3) gives $\tilde{X}^k_A \perp\!\!\!\perp \tilde{X}^k_B \,|\, \tilde{X}^k_C$. Since the bridges on distinct intervals are independent, $\tilde{X}_A \perp\!\!\!\perp \tilde{X}_B \,|\, \tilde{X}_C$. □

Theorem 8. *For all $A, B, C \subseteq V$, the continuous Euler scheme satisfies*

$$X_A \underset{G(\mathcal{D})}{\overset{d}{\perp}} X_B \,|\, X_C \implies X^n_A \perp\!\!\!\perp X^n_B \,|\, X^n_C.$$

Proof. By Lemma 7, $X_A \perp^d_{G(\mathcal{D})} X_B \,|\, X_C$ implies $X^\Delta_A \perp\!\!\!\perp X^\Delta_B \,|\, X^\Delta_C$ and by Lemma D.3 we have $\tilde{X}_A \perp\!\!\!\perp \tilde{X}_B \,|\, \tilde{X}_C$. Since X^Δ_V and $\tilde{X}_V$ are independent by Lemma D.2, this combines into the conditional independence $(X^\Delta_A, \tilde{X}_A) \perp\!\!\!\perp (X^\Delta_B, \tilde{X}_B) \,|(X^\Delta_C, \tilde{X}_C)$. By Lemma D.2, $\sigma(X^n_S) = \sigma(X^\Delta_S) \vee \sigma(\tilde{X}_S)$ for each $S \in \{A, B, C\}$, so that X^n_A and X^n_B are measurable functions of $(X^\Delta_A, \tilde{X}_A)$ and $(X^\Delta_B, \tilde{X}_B)$, while conditioning on X^n_C coincides with conditioning on $(X^\Delta_C, \tilde{X}_C)$; hence this gives $X^n_A \perp\!\!\!\perp X^n_B \,|\, X^n_C$. □

Lemma D.4. *Let $\mathcal{D}$ satisfy Assumption 2. Then for any $\delta > 0$, the process $G(t) := \Sigma X_\gamma(t)$ satisfies*

$$\mathbb{E}\left[\exp(3\delta \sup_{0\le s\le T} \|G(s)\|^2)\right] < \infty$$

whenever $6\delta T\|\Sigma\|^2_{\mathrm{op}} < 1$.

Proof. Note that $\|G(s)\| \le \|\Sigma\|_{\mathrm{op}}\|X_\gamma(s)\|$ and $\|X_\gamma(s)\|^2 = \sum_{i=1}^d W_i(s)^2$, where $W_i := X_{\gamma_i}$ are independent standard Brownian motions. Writing $M_i := \sup_{0\le s\le T} |W_i(s)|$, we have $\sup_{0\le s\le T} \|G(s)\|^2 \le \|\Sigma\|^2_{\mathrm{op}} \sum_{i=1}^d M_i^2$, and by independence of the components

$$\mathbb{E}\left[\exp\left(3\delta \sup_{0\le s\le T} \|G(s)\|^2\right)\right] \le \prod_{i=1}^d \mathbb{E}\left[\exp\left(3\delta\|\Sigma\|^2_{\mathrm{op}} M_i^2\right)\right].$$

The event $\{M_i \ge a\}$ is contained in $\{\sup_{0\le s\le T} W_i(s) \ge a\} \cup \{\sup_{0\le s\le T}(-W_i(s)) \ge a\}$, so by a union bound and the symmetry of W_i, the reflection principle $\mathbb{P}(\sup_{0\le s\le T} W_i(s) \ge a) = 2\mathbb{P}(W_i(T) \ge a)$, and the Gaussian tail bound $\mathbb{P}(W_i(T) \ge a) \le \frac{1}{2} e^{-a^2/2T}$,

$$\mathbb{P}(M_i \ge a) \le 2\,\mathbb{P}\Big(\sup_{0\le s\le T} W_i(s) \ge a\Big) = 4\,\mathbb{P}(W_i(T) \ge a) \le 2e^{-a^2/2T}.$$

With $\lambda := 3\delta\|\Sigma\|^2_{\mathrm{op}} > 0$,

$$\mathbb{E}\left[\exp(\lambda M_i^2)\right] = 1 + \int_0^\infty 2\lambda a\, e^{\lambda a^2}\, \mathbb{P}(M_i \ge a)\mathrm{d}a \le 1 + 4\lambda \int_0^\infty a\, e^{-(\frac{1}{2T} - \lambda)a^2} \mathrm{d}a = 1 + \frac{4T\lambda}{1 - 2T\lambda},$$

which is finite whenever $\lambda < 1/(2T)$, i.e. $6\delta T\|\Sigma\|^2_{\mathrm{op}} < 1$. As a finite product of finite factors, $\mathbb{E}[\exp(3\delta \sup_{0\le s\le T} \|G(s)\|^2)]$ is then finite. □

Theorem 9. *Let $\mathcal{D}$ satisfy Assumption 2, then there is a subsequence n_m such that $\mathbb{P}(X^{n_m}_V)$ converges in total variation to $\mathbb{P}(X_V)$.*

Proof. Without loss of generality, assume that the linear growth coefficient $K(t)$ of μ satisfies $1 \le K := \sup_{0\le t\le T} |K(t)| < \infty$. For readability, write $X := X_V$, $X_0 := X^0_\alpha$ and $W := X_\gamma$. Define $M := \|\Sigma^{-1}\|_{\mathrm{op}}$, the process $G(t) := \Sigma W(t)$, and $\theta(s, x) := \Sigma^{-1}\mu(s, x)$ for $x \in \mathbb{R}^d$; for a path or process y we abbreviate $\mu(s, y) := \mu(s, y(s))$ and $\theta(s, y) := \Sigma^{-1}\mu(s, y(s))$.

Step 1: Disintegration with respect to initial condition. By Theorem 2, $X = I^{[V]}(X_0, W)$ where $I^{[V]}$ is the adapted solution function. For any deterministic $x_0 \in \mathbb{R}^d$, the process $X^{x_0} := I^{[V]}(x_0, W)$ is therefore $\mathcal{F}^W$-adapted (and hence so is $\theta(s, X^{x_0})$). We write $\mathbb{P}_{X^{x_0}}$ for the law of the solution with deterministic initial condition x_0. Since X_0 is independent of W, the law with deterministic initial condition is a version of the conditional law, i.e. $\mathbb{P}(X \in \cdot \,|\, X_0 = x_0) = \mathbb{P}(I^{[V]}(x_0, W) \in \cdot) = \mathbb{P}_{X^{x_0}}(\cdot)$, and hence the law of X disintegrates as $\mathbb{P}(X \in \cdot) = \int \mathbb{P}_{X^{x_0}}(\cdot)\mathrm{d}\mathbb{P}_{X_0}(x_0)$. The same holds for the continuous Euler scheme $X^{x_0,n}$ with initial condition x_0.

Step 2: Density of the centred process $\mathbb{P}_{Y^{x_0}}$ with respect to $\mathbb{P}_G$. Fix $x_0 \in \mathbb{R}^d$ and let $Y^{x_0} := X^{x_0} - x_0$, which starts at $Y^{x_0}(0) = 0$; we compute its density with respect to $\mathbb{P}_G$ (the law $\mathbb{P}_{X^{x_0}}$ itself is supported on paths started at x_0 and is mutually singular with $\mathbb{P}_G$ for $x_0 \neq 0$). The SDE for X^{x_0} gives

$$\|X^{x_0}(t)\| \leq \|x_0\| + \int_0^t \|\mu(s, X^{x_0})\|\mathrm{d}s + \|G(t)\| \leq (\|x_0\| + Kt + \sup_{0\leq s\leq t} \|G(s)\|) + K\int_0^t \sup_{0\leq u\leq s} \|X^{x_0}(u)\|\mathrm{d}s,$$

where we used the linear growth condition $\|\mu(s, X^{x_0})\| \leq K(1 + \sup_{0\leq u\leq s} \|X^{x_0}(u)\|)$. Grönwall's inequality then gives $\sup_{0\leq t\leq T} \|X^{x_0}(t)\| \leq (\|x_0\| + KT + \sup_{0\leq s\leq T} \|G(s)\|)e^{KT}$, so for $\delta > 0$ we have

$$\mathbb{E}\left[\exp\left(\delta e^{-2KT} \sup_{0\leq t\leq T} \|X^{x_0}(t)\|^2\right)\right] \leq \exp(3\delta\|x_0\|^2)\exp(3\delta K^2T^2)\mathbb{E}\left[\exp(3\delta \sup_{0\leq s\leq T} \|G(s)\|^2)\right]$$

which is finite for sufficiently small δ by Lemma D.4. Partitioning $[0, T]$ into intervals of width $t_k - t_{k-1} \leq \delta e^{-2KT}/K^2M^2$, we bound

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{1}{2}\int_{t_{k-1}}^{t_k} \|\theta(s, X^{x_0})\|^2\mathrm{d}s\right)\right] &\leq \mathbb{E}\left[\exp\left(\frac{1}{2}M^2K^2(t_k - t_{k-1})(1 + \sup_{0\leq t\leq T} \|X^{x_0}(t)\|)^2\right)\right] \\ &\leq \mathbb{E}\left[\exp\left(\delta e^{-2KT}(1 + \sup_{0\leq t\leq T} \|X^{x_0}(t)\|^2)\right)\right] < \infty\end{aligned}$$

for each k. Since $X^{x_0} = I^{[V]}(x_0, W)$ is $\mathcal{F}^W$-adapted, the piecewise Novikov condition (Karatzas and Shreve, 1988, Section 3.5, Corollary 5.14) gives $\mathbb{E}[Z^{x_0}] = 1$ for the random variable

$$Z^{x_0} := \exp\left(-\int_0^T \theta(s, X^{x_0})\mathrm{d}W(s) - \frac{1}{2}\int_0^T \|\theta(s, X^{x_0})\|^2\mathrm{d}s\right).$$

Defining the process $\xi^{x_0}(t) := \int_0^t \theta(s, X^{x_0})\mathrm{d}s + W(t)$, by Liptser and Shiryaev (2001) Theorem 7.3 the laws $\mathbb{P}_{\xi^{x_0}}$ and $\mathbb{P}_W$ are equivalent with density

$$\frac{\mathrm{d}\mathbb{P}_{\xi^{x_0}}}{\mathrm{d}\mathbb{P}_W}(\omega) = \exp\left(\int_0^T \theta(s, x_0 + \int_0^s \Sigma\mathrm{d}\xi^{x_0}(u)(\omega))\mathrm{d}\xi^{x_0}(s)(\omega) - \frac{1}{2}\int_0^T \|\theta(s, x_0 + \int_0^s \Sigma\mathrm{d}\xi^{x_0}(u)(\omega))\|^2\mathrm{d}s\right),$$

where $\omega \in \Omega$ denotes a point in the underlying probability space and we used that $\theta(s, X^{x_0}) = \theta(s, x_0 + \int_0^s \Sigma\mathrm{d}\xi^{x_0}(u))$ is $\mathcal{F}_T^{\xi^{x_0}}$-measurable. Let $F(W)$ be the solution function of $\int_0^t \Sigma\mathrm{d}W(s)$, with inverse $F^{-1}(G)$ the solution function of $\int_0^t \Sigma^{-1}\mathrm{d}G(s)$. Since $Y^{x_0} = X^{x_0} - x_0 = F(\xi^{x_0})$ and $G = F(W)$, pushing forward through F gives $\mathbb{P}_{Y^{x_0}} \sim \mathbb{P}_G$ with density

$$\frac{\mathrm{d}\mathbb{P}_{Y^{x_0}}}{\mathrm{d}\mathbb{P}_G}(G(\omega)) = \exp\left(\int_0^T \theta(s, x_0 + G(\omega))\mathrm{d}W(s)(\omega) - \frac{1}{2}\int_0^T \|\theta(s, x_0 + G(\omega))\|^2\mathrm{d}s\right). \quad (14)$$

Step 3: Density of $\mathbb{P}_{Y^{x_0,n}}$ with respect to $\mathbb{P}_G$. The continuous Euler scheme $X^{x_0,n}$ satisfies $X^{x_0,n}(t) = x_0 + \int_0^t \mu_n(s, X^{x_0,n}(s))\mathrm{d}s + \int_0^t \Sigma\mathrm{d}W(s)$ with $\mu_n(t, x) = \mu(t_k, x(t_k))$ for $t \in [t_k, t_{k+1})$. Since $t_k \leq t$, the linear growth bound $\|\mu_n(t, x)\| = \|\mu(t_k, x(t_k))\| \leq K(1 + \|x(t_k)\|) \leq K(1 + \sup_{u\leq t} \|x(u)\|)$ holds, and similarly μ_n inherits the Lipschitz constant of μ. Defining $\theta_n(t, x) = \Sigma^{-1}\mu_n(t, x)$ and the centred process $Y^{x_0,n} := X^{x_0,n} - x_0$, the same Grönwall and Novikov arguments as in Step 2 give $\mathbb{P}_{Y^{x_0,n}} \sim \mathbb{P}_G$ with density

$$\frac{\mathrm{d}\mathbb{P}_{Y^{x_0,n}}}{\mathrm{d}\mathbb{P}_G}(G(\omega)) = \exp\left(\int_0^T \theta_n(s, x_0 + G(\omega))\mathrm{d}W(s)(\omega) - \frac{1}{2}\int_0^T \|\theta_n(s, x_0 + G(\omega))\|^2\mathrm{d}s\right). \quad (15)$$

Step 4: Comparing the densities. Comparing (14) and (15), the log-density ratio is

$$\begin{aligned}\log \frac{\mathrm{d}\mathbb{P}_{Y^{x_0,n}}}{\mathrm{d}\mathbb{P}_{Y^{x_0}}}(G(\omega)) = &\int_0^T \big(\theta_n(s, x_0 + G(\omega)) - \theta(s, x_0 + G(\omega))\big)\mathrm{d}W(s)(\omega) \\ &- \frac{1}{2}\int_0^T \big(\|\theta_n(s, x_0 + G(\omega))\|^2 - \|\theta(s, x_0 + G(\omega))\|^2\big)\mathrm{d}s.\end{aligned} \tag{16}$$

Our goal is to show that the densities $\frac{\mathrm{d}\mathbb{P}_{Y^{x_0,n}}}{\mathrm{d}\mathbb{P}_G}$ converge to $\frac{\mathrm{d}\mathbb{P}_{Y^{x_0}}}{\mathrm{d}\mathbb{P}_G}$ almost surely, for almost all x_0. By Scheffé's Theorem, this gives total variation convergence $\mathbb{P}_{Y^{x_0,n}} \xrightarrow{tv} \mathbb{P}_{Y^{x_0}}$ for almost all x_0. Since $X^{x_0,n} = Y^{x_0,n} + x_0$ and $X^{x_0} = Y^{x_0} + x_0$, and total variation distance is invariant under the bi-measurable translation $y \mapsto y + x_0$, this gives $\mathbb{P}_{X^{x_0,n}} \xrightarrow{tv} \mathbb{P}_{X^{x_0}}$ for almost all x_0, after which the disintegration from the beginning of the proof yields the desired total variation convergence $\mathbb{P}_{X^n} \xrightarrow{tv} \mathbb{P}_X$. Since the densities are exponentials, it suffices to show that the log-density ratio (16) converges to 0 for $(\mathbb{P}_{X_0} \otimes \mathbb{P})$-almost all (x_0, ω).

Our strategy is as follows. We bound both terms in (16) in terms of the single quantity $\int_0^T \|\theta_n(s, x_0 + G) - \theta(s, x_0 + G)\|^2\mathrm{d}s$ (Step 5). We then show that this quantity converges to 0 in $L^1(\mathbb{P}_{X_0} \otimes \mathbb{P})$ (Step 6), which by the bounds implies that the log-density ratio converges to 0 in $L^1(\mathbb{P}_{X_0} \otimes \mathbb{P})$ as well. L^1 convergence implies convergence in probability, which guarantees the existence of a subsequence n_m along which the log-density ratio converges to 0 for $(\mathbb{P}_{X_0} \otimes \mathbb{P})$-almost all (x_0, ω), as desired (Step 7).

Step 5: Bounding the log-density ratio. For the stochastic integral in (16), the Itô isometry gives

$$\mathbb{E}\left[\left(\int_0^T \big(\theta_n(s, x_0 + G) - \theta(s, x_0 + G)\big)\mathrm{d}W(s)\right)^2\right] = \mathbb{E}\left[\int_0^T \|\theta_n(s, x_0 + G) - \theta(s, x_0 + G)\|^2\mathrm{d}s\right]$$

and thus $\mathbb{E}[|\int(\theta_n - \theta)\mathrm{d}W(s)|] \le \mathbb{E}[|\int(\theta_n - \theta)\mathrm{d}W(s)|^2]^{1/2} = \mathbb{E}[\int\|\theta_n - \theta\|^2\mathrm{d}s]^{1/2}$ by Jensen's inequality. For the Lebesgue integral in (16), by the reverse triangle inequality $\big|\|\theta_n\|^2 - \|\theta\|^2\big| = (\|\theta_n\| + \|\theta\|)\big|\|\theta_n\| - \|\theta\|\big| \le (2\|\theta\| + \|\theta_n - \theta\|)\|\theta_n - \theta\|$, so combined with Cauchy–Schwarz on the product $\|\theta\|\|\theta_n - \theta\|$ we have

$$\left|\int_0^T \left(\|\theta_n\|^2 - \|\theta\|^2\right)\mathrm{d}s\right| \le 2\left(\int_0^T \|\theta\|^2\mathrm{d}s\right)^{1/2}\left(\int_0^T \|\theta_n - \theta\|^2\mathrm{d}s\right)^{1/2} + \int_0^T \|\theta_n - \theta\|^2\mathrm{d}s,$$

and thus by Cauchy–Schwarz on the product $2\left(\int_0^T \|\theta\|^2\mathrm{d}s\right)^{1/2}\left(\int_0^T \|\theta_n - \theta\|^2\mathrm{d}s\right)^{1/2}$ we get

$$\mathbb{E}\left[\left|\int_0^T \left(\|\theta_n\|^2 - \|\theta\|^2\right)\mathrm{d}s\right|\right] \le L(x_0)\mathbb{E}\left[\int_0^T \|\theta_n - \theta\|^2\mathrm{d}s\right]^{1/2} + \mathbb{E}\left[\int_0^T \|\theta_n - \theta\|^2\mathrm{d}s\right],$$

where $\mathbb{E}[\int_0^T \|\theta\|^2\mathrm{d}s] \le TM^2K^2\mathbb{E}[(1 + \|x_0\| + \sup_u \|G(u)\|)^2] := \frac{1}{4}L(x_0)^2 < \infty$ by linear growth, whose finiteness follows from Doob's maximal inequality (Karatzas and Shreve, 1988, Section 1.3, Theorem 3.8(iv)): $\mathbb{E}[\sup_u \|G(u)\|^2] \le 4\mathbb{E}[\|G(T)\|^2] = 4\mathrm{tr}(\Sigma\Sigma^\top T) < \infty$, where $\Sigma\Sigma^\top T$ is the covariance of $G(T)$. Combining these bounds we obtain

$$\mathbb{E}\left[\left|\log \frac{\mathrm{d}\mathbb{P}_{Y^{x_0,n}}}{\mathrm{d}\mathbb{P}_{Y^{x_0}}}(G)\right|\right] \;\le\; \left(1 + \tfrac{1}{2}L(x_0)\right)\mathbb{E}\left[\int_0^T \|\theta_n - \theta\|^2\mathrm{d}s\right]^{1/2} + \tfrac{1}{2}\mathbb{E}\left[\int_0^T \|\theta_n - \theta\|^2\mathrm{d}s\right]. \tag{17}$$

We aim at showing $L^1(\mathbb{P}_{X_0} \otimes \mathbb{P})$-convergence of the log-density ratio from which we then obtain a subsequence which converges for $(\mathbb{P}_{X_0} \otimes \mathbb{P})$-almost all (x_0, ω). Integrating (17) over X_0 and applying Cauchy–Schwarz over $\mathbb{P}_{X_0}$ to the first term gives

$$\mathbb{E}_{X_0}\left[\mathbb{E}\left[\left|\log \frac{\mathrm{d}\mathbb{P}_{Y^{X_0,n}}}{\mathrm{d}\mathbb{P}_{Y^{X_0}}}(G)\right|\right]\right] \le \mathbb{E}_{X_0}\left[(1+\tfrac{1}{2}L(X_0))^2\right]^{1/2} \mathbb{E}_{X_0}\left[\mathbb{E}\left[\int_0^T \|\theta_n - \theta\|^2 \mathrm{d}s\right]\right]^{1/2} + \tfrac{1}{2}\mathbb{E}_{X_0}\left[\mathbb{E}\left[\int_0^T \|\theta_n - \theta\|^2 \mathrm{d}s\right]\right],$$

where $\mathbb{E}_{X_0}[(1+\frac{1}{2}L(X_0))^2] < \infty$ since $\mathbb{E}[\|X_0\|^2] < \infty$ by assumption. Hence $L^1(\mathbb{P}_{X_0} \otimes \mathbb{P})$-convergence of the log-density ratio reduces to showing

$$\mathbb{E}_{X_0}\left[\mathbb{E}\left[\int_0^T \|\theta_n(s, X_0+G) - \theta(s, X_0+G)\|^2 \mathrm{d}s\right]\right] \to 0. \tag{18}$$

Step 6: $L^1(\mathbb{P}_{X_0} \otimes \mathbb{P} \otimes \lambda)$-convergence of $\|\theta_n - \theta\|^2$. To obtain (18) we show pointwise convergence of $\|\theta_n(s, x_0 + G(\omega)) - \theta(s, x_0 + G(\omega))\|^2 \to 0$ for $\mathbb{P}_{X_0} \otimes \mathbb{P} \otimes \lambda$-almost all (x_0, ω, s), and apply dominated convergence on $(\mathbb{R}^d \times \Omega \times [0,T], \mathbb{P}_{X_0} \otimes \mathbb{P} \otimes \lambda)$ to obtain the result.

Denote with $t_k^n(s)$ the t_k such that $s \in [t_k, t_{k+1})$. Since $\theta_n(s,x) = \Sigma^{-1}\mu(t_k^n(s), x)$ and $\theta(s,x) = \Sigma^{-1}\mu(s,x)$, we have

$$\|\theta_n(s, x_0+G) - \theta(s, x_0+G)\|^2 \le M^2 \|\mu(t_k^n(s), x_0+G) - \mu(s, x_0+G)\|^2. \tag{19}$$

Pointwise convergence: Fix (x_0, ω). By the abbreviation above, $\mu(t_k^n(s), x_0 + G) = \mu(t_k^n(s), x_0 + G(t_k^n(s)))$ and $\mu(s, x_0+G) = \mu(s, x_0 + G(s))$. As $t_k^n(s) \to s$ we have $x_0 + G(t_k^n(s))(\omega) \to x_0 + G(s)(\omega)$ by continuity of G, and hence

$$\|\mu(t_k^n(s), x_0+G) - \mu(s, x_0+G)\| \to 0$$

by joint continuity of μ (continuous in t and Lipschitz, hence continuous, in the state).

Dominating function: By the triangle inequality and linear growth $\|\mu(t, x_0+G)\| \le K(1 + \|x_0\| + \sup_u \|G(u)\|)$, so

$$\|\theta_n(s, x_0+G) - \theta(s, x_0+G)\|^2 \le 4M^2K^2(1 + \|x_0\| + \sup_u \|G(u)\|)^2,$$

uniformly in n and s. Since $\mathbb{E}[\sup_u \|G(u)\|^2] < \infty$ by Doob's maximal inequality and $\mathbb{E}[\|X_0\|^2] < \infty$ by assumption, this dominating function is integrable against $\mathbb{P}_{X_0} \otimes \mathbb{P}$. Dominated convergence on $(\mathbb{R}^d \times \Omega \times [0,T], \mathbb{P}_{X_0} \otimes \mathbb{P} \otimes \lambda)$ gives (18).

Step 7: Conclusion. As argued in Step 5, (18) gives $L^1(\mathbb{P}_{X_0} \otimes \mathbb{P})$-convergence of the log-density ratio, and hence a subsequence n_m along which the log-density ratio converges to 0 for $(\mathbb{P}_{X_0} \otimes \mathbb{P})$-a.a. (x_0, ω). By Fubini's theorem, for $\mathbb{P}_{X_0}$-a.a. x_0 the log-density ratio converges $\mathbb{P}$-a.s. to 0, so the densities $\frac{\mathrm{d}\mathbb{P}_{Y^{x_0,n_m}}}{\mathrm{d}\mathbb{P}_G}$ converge to $\frac{\mathrm{d}\mathbb{P}_{Y^{x_0}}}{\mathrm{d}\mathbb{P}_G}$ $\mathbb{P}_G$-a.s., and by Scheffé's Theorem $\mathbb{P}_{Y^{x_0,n_m}} \overset{tv}{\to} \mathbb{P}_{Y^{x_0}}$; since total variation is invariant under the translation $y \mapsto y + x_0$, also $\mathbb{P}_{X^{x_0,n_m}} \overset{tv}{\to} \mathbb{P}_{X^{x_0}}$. By the disintegration established in Step 1 we have

$$d_{TV}(\mathbb{P}_{X^{n_m}}, \mathbb{P}_X) \le \int d_{TV}(\mathbb{P}_{X^{x_0,n_m}}, \mathbb{P}_{X^{x_0}}) \mathrm{d}\mathbb{P}_{X_0}(x_0) \to 0,$$

where the convergence to 0 follows by dominated convergence (since $d_{TV} \le 1$). $\square$

E Proofs of Section 6

Rules 2 and 3 of the do-calculus (Theorem 12) involve σ-separation conditions in a graph augmented with intervention nodes R_v (Definition 7). One approach to formalise this is to introduce a separate framework of SCMs with free input nodes — the iSCM framework of Forré and Mooij (2025) — and establish the Markov property for that framework. Instead, we model the intervention variables as genuine endogenous variables in an extended system of causal SDEs with an auxiliary exogenous distribution ν. This allows us to apply the σ-separation Markov property (Theorem 6) directly to the extended system, without having to re-prove it for a different framework.

Definition E.1 (System of causal SDEs with intervention variables)**.** Let $\mathcal{D}$ be a system of causal SDEs, $B \subseteq V$, and ν a probability measure on $\mathcal{X}_{E_B} := \prod_{v\in B}(\mathcal{X}_v \cup \{\star\})$. The *system with intervention variables* $\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}$ has:

- endogenous variables $V \cup R_B$, where each indicator X_{R_v} has sample space $\mathcal{X}_{R_v} := \mathcal{X}_v \cup \{\star\}$;
- exogenous variables $W \cup E_B$, where each X_{E_v} has sample space $\mathcal{X}_{E_v} := \mathcal{X}_v \cup \{\star\}$, with joint distribution $\mathbb{P}(X_W) \otimes \nu$;
- for each $v \in B$, the equation $X_{R_v} = X_{E_v}$ for the intervention variable, and for each $v \in V$ the modified causal SDE

$$X_v(t) = \begin{cases} X_{R_v}(t) & \text{if } v \in B \text{ and } X_{R_v} \in \mathcal{X}_v, \\ f_v(t, X_{\alpha(v)}) + \int_0^t g_v(s-, X_v, X_{\beta(v)})\mathrm{d}h_v(s, X_{\gamma(v)}) & \text{otherwise.} \end{cases}$$

For non-product ν the intervention variables X_{R_v} ($v \in B$) are dependent. For the causal graph and the Markov properties we regard their exogenous parents as a single node E_B with law ν, each X_{R_v} reading its coordinate; this adds the bidirected edges $R_u \leftrightarrow R_v$ ($u \neq v \in B$) to $G(\mathcal{D}_{\mathrm{do}(R_B\sim\nu)})$ and makes $\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}$ a system of causal SDEs and thus equipped with a σ-separation Markov property.[8]

For any partition $B = B_1 \cup B_2$ and $x_{B_2} \in \mathcal{X}_{B_2}$, we will abuse notation and write $\mathcal{D}_{\mathrm{do}(X_{R_{B_1}}=\star, X_{R_{B_2}}=x_{B_2})}$ for $(\mathcal{D}_{\mathrm{do}(R_B\sim\nu)})_{\mathrm{do}(X_{R_{B_1}}=\star, X_{R_{B_2}}=x_{B_2})}$.

Before proving the lemma we record a localisation property: on an event decided at time 0, essential unique solvability pins down the solution on that event alone.

Lemma E.1 (Localisation of essential unique solvability)**.** *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. $O \subseteq V$, with solution function $I^{[O]}$ and $S := V \setminus O$. Fix an adapted input $\varphi_S : \mathcal{X}_W \to \mathcal{X}_S$ and a measurable set $G \subseteq \mathcal{X}_W$ such that $\mathbb{1}_G(X_W) = \mathbb{1}_G(X_W^{\wedge 0})$. If an adapted measurable map $\varphi_O : \mathcal{X}_W \to \mathcal{X}_O$ satisfies $\varphi_O(X_W) = \Phi_O(\varphi_O(X_W), \varphi_S(X_W), X_W)$ $\mathbb{P}$-a.s. on G, then $\varphi_O(X_W) = I^{[O]}(\varphi_S(X_W), X_W)$ $\mathbb{P}$-a.s. on G.*

Proof. Write $I := I^{[O]}(\varphi_S(X_W), X_W)$ and set $\tilde{\varphi}_O := \varphi_O \mathbb{1}_G + I\,\mathbb{1}_{G^c}$. Since I, φ_O and $\mathbb{1}_G$ are adapted (the latter since $\mathbb{1}_G(X_W) = \mathbb{1}_G(X_W^{\wedge 0})$), $\tilde{\varphi}_O$ is adapted as well. By construction we then have

$$\Phi_O(\tilde{\varphi}_O(X_W), \varphi_S(X_W), X_W) = \begin{cases} \Phi_O(\varphi_O(X_W), \varphi_S(X_W), X_W) & \text{on } G \\ \Phi_O(I, \varphi_S(X_W), X_W) & \text{on } G^c. \end{cases}$$

Since both right-hand sides are fixed points on their respective events — the first by hypothesis, the second because I is a fixed point in general — we have that $\tilde{\varphi}_O(X_W) = \Phi_O(\tilde{\varphi}_O(X_W), \varphi_S(X_W), X_W)$ $\mathbb{P}$-a.s. The essential uniqueness of I then gives $\tilde{\varphi}_O = I$ $\mathbb{P}$-a.s., and thus $\varphi_O = I$ $\mathbb{P}$-a.s. on G. ☐

[8] Treating E_B as a single node makes it a higher-dimensional exogenous process, a relaxation of the one-dimensional exogenous process convention of Definition 1.

Lemma E.2. *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every $O \subseteq V$, let $B \subseteq V$, and let ν be a probability measure on $\mathcal{X}_{E_B}$.*

(i) $\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}$ is essentially uniquely solvable w.r.t. every SCC of $G(\mathcal{D}_{\mathrm{do}(R_B\sim\nu)})$.[9]

(ii) We have ν-a.s.

$$\mathbb{P}_{\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}}(X_V \mid \mathrm{do}(X_{R_B} = x_{R_B})) = \mathbb{P}_{\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}}(X_V \mid X_{R_B} = x_{R_B}).$$

(iii) For any partition $B = B_1 \cup B_2$ and $x_{B_2} \in \mathcal{X}_{B_2}$, we have

$$\mathbb{P}_{\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}}(X_V \mid \mathrm{do}(X_{R_{B_1}} = \star, X_{R_{B_2}} = x_{B_2})) = \mathbb{P}_{\mathcal{D}}(X_V \mid \mathrm{do}(X_{B_2} = x_{B_2})).$$

Proof. **(i)** We verify essential unique solvability w.r.t. each SCC C of $G(\mathcal{D})$. The singletons are immediate: for $O = \{R_v\}$ the mechanism is $X_{R_v} = X_{E_v}$, so $\tilde{I}^{[R_v]} := x_{E_v}$ is the unique adapted solution function. It remains to treat an SCC C of $G(\mathcal{D})$.

So fix an SCC $C \subseteq V$ of $G(\mathcal{D})$, and let Φ denote the pathwise mechanism of the augmented system $\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}$. The complement of C in $V \cup R_B$ splits into the endogenous inputs $S := V \setminus C$ and the regime indicators R_B; accordingly fix adapted measurable maps $\varphi_S : \mathcal{X}_W \times \mathcal{X}_{E_B} \to \mathcal{X}_S$ and $\varphi_{R_B} : \mathcal{X}_W \times \mathcal{X}_{E_B} \to \mathcal{X}_{R_B}$. We must exhibit a measurable adapted map $\tilde{I}^{[C]} : \mathcal{X}_S \times \mathcal{X}_{R_B} \times \mathcal{X}_W \times \mathcal{X}_{E_B} \to \mathcal{X}_C$ that is the following fixed point:

$$\tilde{I}^{[C]}\big(\varphi_S, \varphi_{R_B}, X_W, X_{E_B}\big) = \Phi_C\big(\tilde{I}^{[C]}(\varphi_S, \varphi_{R_B}, X_W, X_{E_B}),\ \varphi_S,\ \varphi_{R_B},\ X_W, X_{E_B}\big)$$

$(\mathbb{P}(X_W) \otimes \nu)$-a.s. (evaluating φ_S, φ_{R_B} both at (X_W, X_{E_B})), and the a.s.-unique such fixed point among adapted processes: every adapted φ_C with $\varphi_C = \Phi_C(\varphi_C, \varphi_S, \varphi_{R_B}, \cdot)$ $(\mathbb{P}(X_W) \otimes \nu)$-a.s. equals $\tilde{I}^{[C]}(\varphi_S, \varphi_{R_B}, \cdot)$ $(\mathbb{P}(X_W) \otimes \nu)$-a.s.

A regime configuration $x_{R_B} \in \mathcal{X}_{R_B}$ specifies for each $w \in B$ whether X_{R_w} is idle ($x_{R_w} = \star$) or set to a value $x_{R_w} \in \mathcal{X}_w$. Only the coordinates $w \in C \cap B$ act on the C-block, so for each partition $\pi = B_1^\pi \cup B_2^\pi$ of $C \cap B$, let

$$F_\pi := \big\{x_{R_B} = (x_{R_{B_1^\pi}}, x_{R_{B_2^\pi}}, x_{R_{B\setminus C}}) \in \mathcal{X}_{R_B} : x_{R_{B_1^\pi}} = \star \text{ and } x_{R_{B_2^\pi}} \neq \star\big\}$$

be the cell of regime configurations on which exactly the variables in B_2^π are intervened on; the $2^{|C\cap B|}$ Borel sets $\{F_\pi\}$ partition $\mathcal{X}_{R_B}$. In particular, for $x_{R_B} \in F_\pi$, all variables in $C \cap B_2^\pi$ are intervened on, and all variables in $C^\pi := C \setminus B_2^\pi$ are governed by the original mechanism Φ_{C^π} of $\mathcal{D}$. Let $I^{[C^\pi]}$ be a solution function of $\mathcal{D}$ w.r.t. C^π, and define $\tilde{I}^{[C]} : \mathcal{X}_S \times \mathcal{X}_{R_B} \times \mathcal{X}_W \times \mathcal{X}_{E_B} \to \mathcal{X}_C$ on each cell F_π by matching this mechanism,

$$\tilde{I}_v^{[C]}(x_S, x_{R_B}, x_W)\big|_{F_\pi} := \begin{cases} x_{R_v} & v \in C \cap B_2^\pi, \\ I_v^{[C^\pi]}(x_S, x_{R_{B_2^\pi}}, x_W) & v \in C^\pi, \end{cases}$$

constant in x_{E_B}, and define the adapted and measurable map $\tilde{I}^{[C]} := \sum_\pi \mathbb{1}_{F_\pi}(x_{R_B})\, \tilde{I}^{[C]}\big|_{F_\pi}$.

Composing $\tilde{I}^{[C]}$ with φ_{R_B} shifts the choice of which variables in R_B are intervened on to the realised exogenous values (x_W, x_{E_B}). To analyse the fixed-point and uniqueness properties of $\tilde{I}^{[C]}$, define

$$G_\pi := \big\{(x_W, x_{E_B}) : \varphi_{R_B}(x_W, x_{E_B}) \in F_\pi\big\},$$

such that we can express

$$\tilde{I}_v^{[C]}(\varphi_S, \varphi_{R_B}, x_W)\big|_{G_\pi} := \begin{cases} \varphi_{R_v} & v \in C \cap B_2^\pi, \\ I_v^{[C^\pi]}(\varphi_S, \varphi_{R_{B_2^\pi}}, x_W) & v \in C^\pi. \end{cases}$$

[9] This can be strengthened to being essentially uniquely solvable w.r.t. every $O \subseteq V \cup R_B$, but since this is more complicated and not necessary for our purposes, we do not prove this stronger result.

We analyse the fixed-point and uniqueness properties on a single cell G_π. For each x_{E_B} write $G_\pi(x_{E_B}) := \{x_W : (x_W, x_{E_B}) \in G_\pi\} \subseteq \mathcal{X}_W$. For each $v \in C^\pi$, by the fixed-point property of $I_v^{[C^\pi]}$ the process $I_v^{[C^\pi]}(\varphi_{S\cup R_{B_2^\pi}}(\cdot, x_{E_B}), \cdot)$ is a fixed point of $\Phi_v(\cdot, \varphi_{S\cup R_{B_2^\pi}}(\cdot, x_{E_B}), \cdot)$ on $G_\pi(x_{E_B})$ outside of some $\mathbb{P}(X_W)$-null set, and for every other solution $\varphi_v : \mathcal{X}_W \times \mathcal{X}_{E_B} \to \mathcal{X}_v$ we have by Lemma E.1 that $\varphi_v = I_v^{[C^\pi]}$ $\mathbb{P}(X_W)$-a.s. on $G_\pi(x_{E_B})$. The set of $(x_W, x_{E_B}) \in G_\pi$ where these properties fail is Borel with each section $G_\pi(x_{E_B})$ being $\mathbb{P}(X_W)$-null, hence is $(\mathbb{P}(X_W) \otimes \nu)$-null by Fubini. For each $v \in C \cap B_2^\pi$, the mechanism Φ_v in $\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}$ does not depend on X_C: it sets $X_v = X_{R_v}$ (Definition E.1), which on G_π equals $\varphi_{R_v} \in \mathcal{X}_v$. The v-component of the fixed-point equation is thus $X_v = \varphi_{R_v}$, whose unique adapted solution is $\tilde{I}_v^{[C]} = \varphi_{R_v}$ (on the whole of G_π).

Since $\{G_\pi\}_\pi$ partitions $\mathcal{X}_W \times \mathcal{X}_{E_B}$, $\tilde{I}^{[C]}(\varphi_S(\cdot), \varphi_{R_B}(\cdot), \cdot)$ is $(\mathbb{P}(X_W) \otimes \nu)$-a.s. a fixed point of Φ_C, and any adapted fixed point φ_C equals it $(\mathbb{P}(X_W) \otimes \nu)$-a.s. Since $(\varphi_S, \varphi_{R_B})$ was an arbitrary adapted input, $\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}$ is essentially uniquely solvable w.r.t. C.

(ii) For a regime $x_{R_B} \in \mathcal{X}_{R_B}$, let $B_2 := \{v \in B : x_{R_v} \neq \star\}$ be its intervened coordinates and $x_{B_2} := (x_{R_v})_{v\in B_2} \in \mathcal{X}_{B_2}$ their values. By $\mathcal{D}$'s essential unique solvability and Theorem 1 (intervening on $B_2 \subseteq V$), the system $\mathcal{D}_{\mathrm{do}(X_{B_2}=x_{B_2})}$ is essentially uniquely solvable w.r.t. V; write $h(x_W, x_{R_B}) \in \mathcal{X}_V$ for its V-solution, giving a measurable map $h : \mathcal{X}_W \times \mathcal{X}_{R_B} \to \mathcal{X}_V$. By part (i) and Lemma B.3, $\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}$ is essentially uniquely solvable w.r.t. $V \cup R_B$, so it admits the solution function $I^{[V\cup R_B]} : \mathcal{X}_W \times \mathcal{X}_{E_B} \to \mathcal{X}_V \times \mathcal{X}_{R_B}$. By its essential unique solvability we have for each x_{E_B} that $I_V^{[V\cup R_B]}(x_W, x_{E_B}) = h(x_W, x_{E_B})$ holds $\mathbb{P}(X_W)$-a.s.

For each $v \in V$ the causal SDE for v in $\mathcal{D}_{\mathrm{do}(X_{R_B}=x_{R_B})}$ coincides with the causal SDEs for v in $\mathcal{D}_{\mathrm{do}(X_{B_2}=x_{B_2})}$. Hence, for every x_{R_B},

$$\mathbb{P}_{\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}}(X_V \,|\, \mathrm{do}(X_{R_B} = x_{R_B})) = \mathbb{P}\big(h(X_W, x_{R_B})\big). \tag{20}$$

As $X_V = h(X_W, X_{R_B})$ with $X_{R_B} = X_{E_B}$ a.s. and $X_W \perp\!\!\!\perp X_{E_B}$ under $\mathbb{P}(X_W) \otimes \nu$, we have for ν-a.a. x_{R_B}

$$\mathbb{P}_{\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}}(X_V \,|\, X_{R_B} = x_{R_B}) = \mathbb{P}\big(h(X_W, x_{R_B}) \,|\, X_{E_B} = x_{R_B}\big) = \mathbb{P}\big(h(X_W, x_{R_B})\big).$$

The interventional and conditional distributions therefore agree ν-a.s.

(iii) Taking $x_{R_B} = (\star, x_{B_2})$ in (20) gives

$$\mathbb{P}_{\mathcal{D}_{\mathrm{do}(R_B\sim\nu)}}(X_V \,|\, \mathrm{do}(X_{R_{B_1}} = \star, X_{R_{B_2}} = x_{B_2})) = \mathbb{P}\big(h(X_W, (\star, x_{B_2}))\big).$$

By definition $h(\cdot, (\star, x_{B_2}))$ is the V-solution of $\mathcal{D}_{\mathrm{do}(X_{B_2}=x_{B_2})}$, so the right-hand side equals $\mathbb{P}_{\mathcal{D}}(X_V \,|\, \mathrm{do}(X_{B_2} = x_{B_2}))$. □

Theorem 12. *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every $O \subseteq V$, and let $A, B, C \subseteq V$ be pairwise disjoint. Then:*

1. Insertion/deletion of observation. *If $A \perp^\sigma_{G(\mathcal{D})} B \,|\, C$, then*

$$\mathbb{P}(X_A \,|\, X_B, X_C) = \mathbb{P}(X_A \,|\, X_C) \quad \mathbb{P}(X_B, X_C)\text{-a.s.}$$

2. Action/observation exchange. *If $A \perp^\sigma_{G(\mathcal{D})_{\mathrm{do}(R_B)}} R_B \,|\, B \cup C$ and $\mathbb{P}(X_C \mid X_B = x_B) \ll \mathbb{P}(X_C \mid \mathrm{do}(X_B = x_B))$ for $\mathbb{P}(X_B)$-a.a. x_B, then*

$$\mathbb{P}(X_A \,|\, \mathrm{do}(X_B), X_C) = \mathbb{P}(X_A \,|\, X_B, X_C) \quad \mathbb{P}(X_B, X_C)\text{-a.s.}$$

3. Insertion/deletion of action. *If $A \perp^\sigma_{G(\mathcal{D})_{\mathrm{do}(R_B)}} R_B \,|\, C$, then for every $x_B \in \mathcal{X}_B$ such that $\mathbb{P}(X_C) \ll \mathbb{P}(X_C \mid \mathrm{do}(X_B = x_B))$,*

$$\mathbb{P}(X_A \,|\, \mathrm{do}(X_B = x_B), X_C) = \mathbb{P}(X_A \,|\, X_C) \quad \mathbb{P}(X_C)\text{-a.s.}$$

Proof. The proof follows (Forré and Mooij, 2025, Theorem 5.1.2).

Rule 1. By Theorem 6, the σ-separation $A \perp^{\sigma}_{G(\mathcal{D})} B \mid C$ gives $X_A \perp\!\!\!\perp X_B \mid X_C$ under $\mathbb{P}_{\mathcal{D}}$. By Kallenberg (2021), Theorem 8.9, this gives $\mathbb{P}_{\mathcal{D}}(X_A \mid X_B, X_C) = \mathbb{P}_{\mathcal{D}}(X_A \mid X_C)$ $\mathbb{P}_{\mathcal{D}}(X_B, X_C)$-a.s.

Rule 2. Define $\nu := \frac{1}{2}\delta_{(\star,\dots,\star)} + \frac{1}{2}\mathbb{P}_{\mathcal{D}}(X_B)$ on $\mathcal{X}_{R_B}$. Write $\mathcal{D}_\nu := \mathcal{D}_{\mathrm{do}(R_B \sim \nu)}$ and let $\tilde{G}(\mathcal{D})_{\mathrm{do}(R_B)}$ be the graph $G(\mathcal{D})_{\mathrm{do}(R_B)}$ appended with bidirected edges $R_u \leftrightarrow R_v$ for all $u \neq v \in B$. Then we also have the σ-separation $A \perp^{\sigma}_{\tilde{G}(\mathcal{D})_{\mathrm{do}(R_B)}} R_B \mid B \cup C$, since compared to $G(\mathcal{D})_{\mathrm{do}(R_B)}$, adding bidirected edges between the intervention variables does not open up a walk between A and R_B (conditioned on $B \cup C$). Since $G(\mathcal{D}_\nu)$ is a subgraph of $\tilde{G}(\mathcal{D})_{\mathrm{do}(R_B)}$ we also have $A \perp^{\sigma}_{G(\mathcal{D}_\nu)} R_B \mid B \cup C$. By Lemma E.2(i) $\mathcal{D}_\nu$ is essentially uniquely solvable with respect to every strongly connected component of $G(\mathcal{D}_\nu)$, so the σ-separation Markov property holds and we have under $\mathbb{P}_{\mathcal{D}_\nu}$ the conditional independence $X_A \perp\!\!\!\perp X_{R_B} \mid X_B, X_C$, which gives the factorisation

$$\mathbb{P}_{\mathcal{D}_\nu}(X_A, X_B, X_C, X_{R_B}) = \mathbb{Q}(X_A \mid X_B, X_C) \otimes \mathbb{P}_{\mathcal{D}_\nu}(X_B, X_C, X_{R_B}),$$

for some Markov kernel $\mathbb{Q} : \mathcal{X}_B \times \mathcal{X}_C \to \mathcal{P}(\mathcal{X}_A)$. By conditioning the factorisation above on X_{R_B} and applying Lemma E.2(ii) we get

$$\mathbb{P}_{\mathcal{D}_\nu}(X_A, X_B, X_C \mid \mathrm{do}(X_{R_B})) = \mathbb{Q}(X_A \mid X_B, X_C) \otimes \mathbb{P}_{\mathcal{D}_\nu}(X_B, X_C \mid \mathrm{do}(X_{R_B})) \tag{21}$$

ν-a.s. Under $\mathrm{do}(X_{R_B} = \star)$, Lemma E.2(iii) and (21) give

$$\mathbb{P}_{\mathcal{D}}(X_A, X_B, X_C) = \mathbb{Q}(X_A \mid X_B, X_C) \otimes \mathbb{P}_{\mathcal{D}}(X_B, X_C),$$

and hence $\mathbb{Q}(X_A \mid X_B, X_C) = \mathbb{P}_{\mathcal{D}}(X_A \mid X_B, X_C)$ holds $\mathbb{P}_{\mathcal{D}}(X_B, X_C)$-a.s. Under $\mathrm{do}(X_{R_B} = x_B)$ for $x_B \in \mathcal{X}_B$, Lemma E.2(iii) and (21) give, for $\mathbb{P}_{\mathcal{D}}(X_B)$-a.a. x_B that

$$\mathbb{P}_{\mathcal{D}}(X_A, X_C \mid \mathrm{do}(X_B = x_B)) = \mathbb{Q}(X_A \mid X_B = x_B, X_C) \otimes \mathbb{P}_{\mathcal{D}}(X_C \mid \mathrm{do}(X_B = x_B)) \tag{22}$$

and hence $\mathbb{Q}(X_A \mid X_B, X_C) = \mathbb{P}_{\mathcal{D}}(X_A \mid \mathrm{do}(X_B = x_B), X_C)$ holds for $\mathbb{P}_{\mathcal{D}}(X_B)$-almost all x_B, $\mathbb{P}_{\mathcal{D}}(X_C \mid \mathrm{do}(X_B = x_B))$-a.s. If for $\mathbb{P}_{\mathcal{D}}(X_B)$-a.a. x_B we also have $\mathbb{P}_{\mathcal{D}}(X_C \mid X_B = x_B) \ll \mathbb{P}_{\mathcal{D}}(X_C \mid \mathrm{do}(X_B = x_B))$, then $\mathbb{Q}(X_A \mid X_B, X_C) = \mathbb{P}_{\mathcal{D}}(X_A \mid \mathrm{do}(X_B), X_C)$ holds $\mathbb{P}_{\mathcal{D}}(X_B, X_C)$-a.s. as well, in which case $\mathbb{P}_{\mathcal{D}}(X_A \mid X_B, X_C) = \mathbb{P}_{\mathcal{D}}(X_A \mid \mathrm{do}(X_B), X_C)$ holds $\mathbb{P}_{\mathcal{D}}(X_B, X_C)$-a.s.

Rule 3. Let $x_B \in \mathcal{X}_B$ and define $\nu := \bigotimes_{v \in B} \left(\frac{1}{2}\delta_\star + \frac{1}{2}\delta_{x_v}\right)$ on $\mathcal{X}_{R_B}$. Since ν is a product measure we have that $G(\mathcal{D}_\nu)$ is a subgraph of $G(\mathcal{D})_{\mathrm{do}(R_B)}$, so $A \perp^{\sigma}_{G(\mathcal{D}_\nu)} R_B \mid C$. By the σ-separation Markov property for $\mathcal{D}_\nu$ we have the conditional independence $X_A \perp\!\!\!\perp X_{R_B} \mid X_C$, giving the factorisation

$$\mathbb{P}_{\mathcal{D}_\nu}(X_A, X_C, X_{R_B}) = \mathbb{Q}(X_A \mid X_C) \otimes \mathbb{P}_{\mathcal{D}_\nu}(X_C, X_{R_B}),$$

for some Markov kernel $\mathbb{Q} : \mathcal{X}_C \to \mathcal{P}(\mathcal{X}_A)$, and similarly to Rule 2 this yields

$$\mathbb{P}_{\mathcal{D}_\nu}(X_A, X_C \mid \mathrm{do}(X_{R_B})) = \mathbb{Q}(X_A \mid X_C) \otimes \mathbb{P}_{\mathcal{D}_\nu}(X_C \mid \mathrm{do}(X_{R_B})) \quad \nu\text{-a.s.}$$

Under $\mathrm{do}(X_{R_B} = \star)$, Lemma E.2(iii) gives $\mathbb{Q}(X_A \mid X_C) = \mathbb{P}_{\mathcal{D}}(X_A \mid X_C)$ $\mathbb{P}_{\mathcal{D}}(X_C)$-a.s. Since ν assigns positive mass to x_B, we have by Lemma E.2(iii) under $\mathrm{do}(X_{R_B} = x_B)$ that

$$\mathbb{Q}(X_A \mid X_C) = \mathbb{P}_{\mathcal{D}}(X_A \mid \mathrm{do}(X_B = x_B), X_C) \quad \mathbb{P}_{\mathcal{D}}(X_C \mid \mathrm{do}(X_B = x_B))\text{-a.s.}$$

If $\mathbb{P}_{\mathcal{D}}(X_C) \ll \mathbb{P}_{\mathcal{D}}(X_C \mid \mathrm{do}(X_B = x_B))$ then this last equality holds $\mathbb{P}_{\mathcal{D}}(X_C)$-a.s., and combining with the previous equality yields $\mathbb{P}_{\mathcal{D}}(X_A \mid \mathrm{do}(X_B = x_B), X_C) = \mathbb{P}_{\mathcal{D}}(X_A \mid X_C)$ $\mathbb{P}_{\mathcal{D}}(X_C)$-a.s. ☐

Theorem 13. *Let $\mathcal{D}$ be a system of causal SDEs that is essentially uniquely solvable w.r.t. every $O \subseteq V$ and let $u, v \in V$. If there is no directed path from u to v in $G(\mathcal{D})$, then $\mathbb{P}(X_v \mid \mathrm{do}(X_u)) = \mathbb{P}(X_v)$.*

Proof. If there is no directed path from u to v in $G(\mathcal{D})$, then there is no directed path from u to v in $G(\mathcal{D})_{\mathrm{do}(R_u)}$ (since R_u has only the edge $R_u \to u$), and hence $v \perp^{\sigma}_{G(\mathcal{D})_{\mathrm{do}(R_u)}} R_u$. Applying Rule 3 of Theorem 12 with $A = \{v\}$, $B = \{u\}$, $C = \emptyset$ (the absolute continuity condition is vacuous since $C = \emptyset$) gives $\mathbb{P}(X_v \mid \mathrm{do}(X_u)) = \mathbb{P}(X_v)$. ☐

F Proofs of Section 7

Define the *restriction map* $\mathrm{res}^{\mathcal{I}} : D([0,T],\mathbb{R}^n) \to D(\mathcal{I},\mathbb{R}^n)$ as $\mathrm{res}^{\mathcal{I}}(x) := x|_{\mathcal{I}}$, and extend as $\mathrm{res}^{\mathcal{E}} = (\mathrm{res}^{\mathcal{I}_1}, \ldots, \mathrm{res}^{\mathcal{I}_m})$.

Lemma F.1. *Let $\mathcal{D} = (V, W, X_W, f, g, h)$ be a system of causal SDEs, $\mathcal{E}$ a finite partition of $[0,T]$ into intervals and $\mathcal{D}^{\mathcal{E}}$ the time-split SDE. Then:*

(i) If X_V is a solution of $\mathcal{D}$, then $X_V^{\mathcal{E}} := \mathrm{res}^{\mathcal{E}}(X_V)$ is a solution of $\mathcal{D}^{\mathcal{E}}$.

(ii) Conversely, if $X_V^{\mathcal{E}}$ is a solution of $\mathcal{D}^{\mathcal{E}}$, then $X_V := \mathrm{cat}(X_V^{\mathcal{E}})$ is a solution of $\mathcal{D}$.

Proof. *(i)* Let X_V solve $\mathcal{D}$ and set $X_V^{\mathcal{E}} := \mathrm{res}^{\mathcal{E}}(X_V)$; let $X_W^{\mathcal{E}}$ denote the corresponding tuple of exogenous increment processes from Definition 8. Fix $v \in V$, $\mathcal{I}_k \in \mathcal{E}$, $s \in \mathcal{I}_k$. Evaluating the SDE at s and at $b_{k-1,k}$ gives

$$X_v(s) = X_v(b_{k-1,k}) + \big(f_v(s, X_{\alpha(v)}) - f_v(b_{k-1,k}, X_{\alpha(v)})\big) + \int_{b_{k-1,k}}^{s} g_v(u-, X_v, X_{\beta(v)})\mathrm{d}h_v(u, X_{\gamma(v)}).$$

We have $X_v(b_{k-1,k}) = X_v^{\mathcal{I}_{k-1}}(b_{k-1,k})$, which for $k=1$ equals $X_v(0) = f_v(0, X_{\alpha(v)}^{\mathcal{I}_1})$ by Definition 8. Under the reconstructions $X_{\alpha(v)} = (\mathrm{cat}(X_{\alpha(v)\cap V}^{\mathcal{E}}), \mathrm{rec}(X_{\alpha(v)\cap W}^{\mathcal{E}}))$ and similarly for $\{v\} \cup \beta(v)$ and $\gamma(v)$, writing out the definitions of $f_v^{\mathcal{I}_k}, g_v^{\mathcal{I}_k}, h_v^{\mathcal{I}_k}$ yields the time-split SDE for $X_v^{\mathcal{I}_k}$.

(ii) Let $X_V^{\mathcal{E}}$ solve $\mathcal{D}^{\mathcal{E}}$ and set $X_V := \mathrm{cat}(X_V^{\mathcal{E}})$, $X_W := \mathrm{rec}(X_W^{\mathcal{E}})$. Under these reconstructions, $f_v^{\mathcal{I}_k}, g_v^{\mathcal{I}_k}, h_v^{\mathcal{I}_k}$ evaluate via the original f_v, g_v, h_v as in Definition 8. We show by induction on k that $X_v(s) = f_v(s, X_{\alpha(v)}) + \int_0^s g_v(u-, X_v, X_{\beta(v)})\mathrm{d}h_v(u, X_{\gamma(v)})$ for all $s \in \mathcal{I}_k$.

Base ($k=1$). The time-split SDE on $\mathcal{I}_1$ straightforwardly reduces to $X_v^{\mathcal{I}_1}(s) = f_v(s, X_{\alpha(v)}) + \int_0^s g_v(u-, X_v, X_{\beta(v)})\mathrm{d}h_v(u, X_{\gamma(v)})$.

Inductive step ($k \geq 2$). The time-split SDE on $\mathcal{I}_k$ reads

$$X_v^{\mathcal{I}_k}(s) = X_v^{\mathcal{I}_{k-1}}(b_{k-1,k}) + \big(f_v(s, X_{\alpha(v)}) - f_v(b_{k-1,k}, X_{\alpha(v)})\big) + \int_{b_{k-1,k}}^{s} g_v(u-, X_v, X_{\beta(v)})\mathrm{d}h_v(u, X_{\gamma(v)}).$$

We have $X_v^{\mathcal{I}_{k-1}}(b_{k-1,k}) = X_v(b_{k-1,k})$, which by the induction hypothesis equals $f_v(b_{k-1,k}, X_{\alpha(v)}) + \int_0^{b_{k-1,k}} g_v(u-, X_v, X_{\beta(v)})\mathrm{d}h_v(u, X_{\gamma(v)})$. Substituting and combining the integrals gives $X_v(s) = f_v(s, X_{\alpha(v)}) + \int_0^s g_v(u-, X_v, X_{\beta(v)})\mathrm{d}h_v(u, X_{\gamma(v)})$. $\square$

Lemma F.2 (Solvability of the time-split system)**.** *Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes have independent increments, and let $\mathcal{E}$ be a finite partition of $[0,T]$ into intervals. Then the time-split system $\mathcal{D}^{\mathcal{E}}$ is essentially uniquely solvable w.r.t. every $O \subseteq V \times \mathcal{E}$.*

Proof. By adaptedness, $G(\mathcal{D}^{\mathcal{E}})$ has no edges from later to earlier intervals, so every SCC of $G(\mathcal{D}^{\mathcal{E}})$ is of the form $S \times \{\mathcal{I}_k\}$ for some SCC S of $G(\mathcal{D})$. On each such SCC, the functions $f_v^{\mathcal{I}_k}, g_v^{\mathcal{I}_k}, h_v^{\mathcal{I}_k}$ are compositions of f_v, g_v, h_v with cat and rec, and the linear-growth and Lipschitz bounds transfer pointwise since $\|\mathrm{cat}(x^{\mathcal{E}})\|_\infty \leq \max_k \|x^{\mathcal{I}_k}\|_\infty$ and $\|\mathrm{rec}(y^{\mathcal{E}})\|_\infty \leq \sum_k \|y^{\mathcal{I}_k}\|_\infty \leq |\mathcal{E}| \max_k \|y^{\mathcal{I}_k}\|_\infty$. Hence $\mathcal{D}^{\mathcal{E}}$ satisfies Assumption 1, and by Theorem 3 it is essentially uniquely solvable w.r.t. every $O \subseteq V \times \mathcal{E}$. $\square$

For any subset $O \subseteq (V \cup W) \times \mathcal{E}$ and tuple of processes $\bar{x}_O = (\bar{x}_u^{\mathcal{I}})_{(u,\mathcal{I}) \in O} \in D([0,T],\mathbb{R})^{|O|}$, let the restriction map $\mathrm{res}(\bar{x}_O) = (\mathrm{res}^{\mathcal{I}}(\bar{x}_u^{\mathcal{I}}))_{(u,\mathcal{I}) \in O}$ be applied element-wise. As a main tool for embedding each variable $x_v^{\mathcal{I}_k} \in D(\mathcal{I}_k,\mathbb{R})$ of $\mathcal{D}^{\mathcal{E}}$ in $D([0,T],\mathbb{R})$, consider the following map:

$$\mathrm{ext}(x_v^{\mathcal{I}_k})(t) := \begin{cases} 0 & \text{if } t < \inf \mathcal{I}_k \\ \lim_{s \downarrow \inf \mathcal{I}_k} x_v^{\mathcal{I}_k}(s) & \text{if } t = \inf \mathcal{I}_k \text{ and } t \notin \mathcal{I}_k \\ x_v^{\mathcal{I}_k}(t) & \text{if } t \in \mathcal{I}_k \\ x_v^{\mathcal{I}_k}(b_{k,k+1}) & \text{if } t \geq \sup \mathcal{I}_k, \end{cases}$$

extended componentwise. Note that $\mathrm{res} \circ \mathrm{ext} = \mathrm{id}$.

Definition F.1 (Embedded time-split system)**.** Given $\mathcal{D}$ such that each exogenous process X_w has independent increments, let $\mathcal{D}^{\mathcal{E}} = \big(V \times \mathcal{E}, W \times \mathcal{E}, X_{W\times\mathcal{E}}, f_V^{\mathcal{E}}, g_V^{\mathcal{E}}, h_V^{\mathcal{E}}\big)$ be the time-split system of causal SDEs (Definition 8) for some finite partition $\mathcal{E}$ of $[0,T]$ into intervals, and let $\Phi_v^{\mathcal{I}_k}$ denote its pathwise mechanism. We define the *embedded time-split system* $\bar{\mathcal{D}}^{\mathcal{E}} := \big(V \times \mathcal{E}, W \times \mathcal{E}, \bar{\mathcal{X}}_{V\times\mathcal{E}}, \bar{\mathcal{X}}_{W\times\mathcal{E}}, \bar{\Phi}_V^{\mathcal{E}}, \bar{\mathbb{P}}\big)$ on $[0,T]$ as follows:

- Sample spaces $\bar{\mathcal{X}}_{V\times\mathcal{E}} := D([0,T],\mathbb{R})^{|V\times\mathcal{E}|}$ and $\bar{\mathcal{X}}_{W\times\mathcal{E}} := D([0,T],\mathbb{R})^{|W\times\mathcal{E}|}$.
- For each $(v,\mathcal{I}_k) \in V \times \mathcal{E}$, the embedded causal SDE $\bar{X}_v^{\mathcal{I}_k}(t) = \bar{\Phi}_v^{\mathcal{I}_k}(\bar{x}_V^{\mathcal{E}}, \bar{x}_W^{\mathcal{E}})(t)$ with
$$\bar{\Phi}_v^{\mathcal{I}_k}(\bar{x}_V^{\mathcal{E}}, \bar{x}_W^{\mathcal{E}}) := \mathrm{ext}\big(\Phi_v^{\mathcal{I}_k}(\mathrm{res}(\bar{x}_V^{\mathcal{E}}), \mathrm{res}(\bar{x}_W^{\mathcal{E}}))\big).$$
- As exogenous distribution $\mathbb{P}(\bar{X}_W^{\mathcal{E}}) := \bigotimes_{(w,\mathcal{I})\in W\times\mathcal{E}} \mathbb{P}(\bar{X}_w^{\mathcal{I}})$ the distribution of the random variable
$$\bar{X}_w^{\mathcal{I}_k} := \mathrm{ext}(X_w^{\mathcal{I}_k})$$
where $X_W^{\mathcal{E}}$ is any random variable with distribution $\mathbb{P}(X_W^{\mathcal{E}})$.

Lemma F.3. *Let $\mathcal{D}^{\mathcal{E}}$ be the time-split system of causal SDEs from Definition 8 with exogenous processes X_w ($w \in W$) having independent increments, then:*

(i) If $\mathcal{D}^{\mathcal{E}}$ is essentially uniquely solvable w.r.t. $O \subseteq V \times \mathcal{E}$, then so is $\bar{\mathcal{D}}^{\mathcal{E}}$.

(ii) $G(\mathcal{D}^{\mathcal{E}}) = G(\bar{\mathcal{D}}^{\mathcal{E}})$.

(iii) For every $S \subseteq V \times \mathcal{E}$, with $O := (V \times \mathcal{E}) \setminus S$, and every $x_S \in \mathcal{X}_S$,
$$\mathbb{P}_{\mathcal{D}^{\mathcal{E}}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = \mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}\big(\mathrm{res}(\bar{X}_O) \,|\, \mathrm{do}(\bar{X}_S = \mathrm{ext}(x_S))\big).$$

Proof. (i) Fix $O \subseteq V \times \mathcal{E}$ and write $S := (V \times \mathcal{E}) \setminus O$. By essential unique solvability of $\mathcal{D}^{\mathcal{E}}$ w.r.t. O, there is an adapted measurable solution function $I^{[O]} : \mathcal{X}_S \times \mathcal{X}_{W\times\mathcal{E}} \to \mathcal{X}_O$, that satisfies for every admissible adapted measurable $\varphi_S : \mathcal{X}_{W\times\mathcal{E}} \to \mathcal{X}_S$ the fixed-point property
$$I^{[O]}(\varphi_S(X_W^{\mathcal{E}}), X_W^{\mathcal{E}}) = \Phi_O\big(I^{[O]}(\varphi_S(X_W^{\mathcal{E}}), X_W^{\mathcal{E}}), \varphi_S(X_W^{\mathcal{E}}), X_W^{\mathcal{E}}\big) \tag{23}$$
$\mathbb{P}$-a.s., and is essentially unique. Define $\bar{I}^{[O]} : \bar{\mathcal{X}}_S \times \bar{\mathcal{X}}_{W\times\mathcal{E}} \to \bar{\mathcal{X}}_O$ as the componentwise extension of the time-split solution to $[0,T]$,
$$\bar{I}^{[O]}(\bar{x}_S, \bar{x}_W^{\mathcal{E}}) := \mathrm{ext}\big(I^{[O]}(\mathrm{res}(\bar{x}_S), \mathrm{res}(\bar{x}_W^{\mathcal{E}}))\big),$$
which is adapted and measurable. Fix an admissible, adapted measurable $\bar{\varphi}_S : \bar{\mathcal{X}}_{W\times\mathcal{E}} \to \bar{\mathcal{X}}_S$ and let $\varphi_S := \mathrm{res} \circ \bar{\varphi}_S \circ \mathrm{ext} : \mathcal{X}_{W\times\mathcal{E}} \to \mathcal{X}_S$, which is admissible, adapted and measurable as well. Write $X_W^{\mathcal{E}} := \mathrm{res}(\bar{X}_W^{\mathcal{E}})$, so that $\mathrm{res}(\bar{\varphi}_S(\bar{X}_W^{\mathcal{E}})) = \varphi_S(x_W^{\mathcal{E}})$. Using $\bar{\Phi}_O = \mathrm{ext} \circ \Phi_O \circ \mathrm{res}$, $\mathrm{res} \circ \mathrm{ext} = \mathrm{id}$, and (23),
$$\begin{aligned}\bar{\Phi}_O\big(\bar{I}^{[O]}(\bar{\varphi}_S(\bar{X}_W^{\mathcal{E}}), \bar{X}_W^{\mathcal{E}}), \bar{\varphi}_S(\bar{X}_W^{\mathcal{E}}), \bar{X}_W^{\mathcal{E}}\big) &= \mathrm{ext}\big(\Phi_O(I^{[O]}(\varphi_S(x_W^{\mathcal{E}}), x_W^{\mathcal{E}}), \varphi_S(x_W^{\mathcal{E}}), x_W^{\mathcal{E}})\big) \\ &= \mathrm{ext}\big(I^{[O]}(\varphi_S(x_W^{\mathcal{E}}), x_W^{\mathcal{E}})\big) = \bar{I}^{[O]}(\bar{\varphi}_S(\bar{X}_W^{\mathcal{E}}), \bar{X}_W^{\mathcal{E}})\end{aligned}$$
$\bar{\mathbb{P}}$-a.s., so $\bar{I}^{[O]}(\bar{\varphi}_S(\bar{X}_W^{\mathcal{E}}), \bar{X}_W^{\mathcal{E}})$ is a fixed point of $\bar{\Phi}_O$.

Let $\bar{\varphi}_O : \bar{\mathcal{X}}_{W\times\mathcal{E}} \to \bar{\mathcal{X}}_O$ be a measurable adapted function with $\bar{\varphi}_O(\bar{X}_W^{\mathcal{E}}) = \bar{\Phi}_O(\bar{\varphi}_O(\bar{X}_W^{\mathcal{E}}), \bar{\varphi}_S(\bar{X}_W^{\mathcal{E}}), \bar{X}_W^{\mathcal{E}})$ $\bar{\mathbb{P}}$-a.s., and set $\varphi_O := \mathrm{res} \circ \bar{\varphi}_O \circ \mathrm{ext}$, so that $\mathrm{res}(\bar{\varphi}_O(\bar{X}_W^{\mathcal{E}})) = \varphi_O(x_W^{\mathcal{E}})$ $\bar{\mathbb{P}}$-a.s. Using $\bar{\Phi}_O = \mathrm{ext} \circ \Phi_O \circ \mathrm{res}$ and applying res (with $\mathrm{res} \circ \mathrm{ext} = \mathrm{id}$) gives $\varphi_O(x_W^{\mathcal{E}}) = \Phi_O(\varphi_O(x_W^{\mathcal{E}}), \varphi_S(x_W^{\mathcal{E}}), x_W^{\mathcal{E}})$; since

$\bar{\mathbb{P}}(\mathrm{res}(\bar{X}^{\mathcal{E}}_W)) = \mathbb{P}(X^{\mathcal{E}}_W)$ this holds $\mathbb{P}$-a.s., so by the essential uniqueness of $I^{[O]}$, $\varphi_O(x^{\mathcal{E}}_W) = I^{[O]}(\varphi_S(x^{\mathcal{E}}_W), x^{\mathcal{E}}_W)$ $\mathbb{P}$-a.s. Substituting back and using (23),

$$\bar{\varphi}_O(\bar{X}^{\mathcal{E}}_W) = \mathrm{ext}\big(\Phi_O(I^{[O]}(\varphi_S(x^{\mathcal{E}}_W), x^{\mathcal{E}}_W), \varphi_S(x^{\mathcal{E}}_W), x^{\mathcal{E}}_W)\big) = \mathrm{ext}\big(I^{[O]}(\varphi_S(x^{\mathcal{E}}_W), x^{\mathcal{E}}_W)\big) = \bar{I}^{[O]}(\bar{\varphi}_S(\bar{X}^{\mathcal{E}}_W), \bar{X}^{\mathcal{E}}_W)$$

$\bar{\mathbb{P}}$-a.s.

(ii) By Definition F.1, $\bar{\Phi}^{\mathcal{I}_k}_v = \mathrm{ext} \circ \Phi^{\mathcal{I}_k}_v \circ \mathrm{res}$, and ext, res are deterministic with $\mathrm{res}^{\mathcal{I}_j}(\bar{x}^{\mathcal{I}_j}_u) = x^{\mathcal{I}_j}_u$; hence $\bar{\Phi}^{\mathcal{I}_k}_v$ essentially depends on $\bar{x}^{\mathcal{I}_j}_u$ if and only if $\Phi^{\mathcal{I}_k}_v$ essentially depends on $x^{\mathcal{I}_j}_u$, so the directed edges into $(v, \mathcal{I}_k)$ coincide in $G(\bar{\mathcal{D}}^{\mathcal{E}})$ and $G(\mathcal{D}^{\mathcal{E}})$ — including the initial-value edge $(v, \mathcal{I}_{k-1}) \to (v, \mathcal{I}_k)$, carried by the term $X^{\mathcal{I}_{k-1}}_v(b_{k-1,k})$ inside $\Phi^{\mathcal{I}_k}_v$ (Definition 8). Bidirected edges match via the product factorisation $\bar{\mathbb{P}} = \bigotimes_{(w,\mathcal{I})} \mathbb{P}(\bar{X}^{\mathcal{I}}_w)$. Hence $G(\bar{\mathcal{D}}^{\mathcal{E}}) = G(\mathcal{D}^{\mathcal{E}})$.

(iii) Fix a solution function $I^{[O]} : \mathcal{X}_S \times \mathcal{X}_{W \times \mathcal{E}} \to \mathcal{X}_O$ of $\mathcal{D}^{\mathcal{E}}$ w.r.t. O, and let $\bar{I}^{[O]} = \mathrm{ext} \circ I^{[O]} \circ \mathrm{res}$ be the solution function of $\bar{\mathcal{D}}^{\mathcal{E}}$ from part (i). Since $\mathrm{res} \circ \mathrm{ext} = \mathrm{id}$,

$$\mathrm{res}\big(\bar{I}^{[O]}(\bar{x}_S, \bar{x}^{\mathcal{E}}_W)\big) = I^{[O]}\big(\mathrm{res}(\bar{x}_S), \mathrm{res}(\bar{x}^{\mathcal{E}}_W)\big). \tag{24}$$

By part (i), $\bar{\mathcal{D}}^{\mathcal{E}}$ is essentially uniquely solvable, so its interventional distributions are well-defined as pushforwards under the (intervened) solution function. Applied to $\mathcal{D}^{\mathcal{E}}$ and to $\bar{\mathcal{D}}^{\mathcal{E}}$, this gives

$$\mathbb{P}_{\mathcal{D}^{\mathcal{E}}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big) = I^{[O]}(x_S, \cdot)_* \mathbb{P}(X^{\mathcal{E}}_W),$$
$$\mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}\big(\bar{X}_O \,|\, \mathrm{do}(\bar{X}_S = \mathrm{ext}(x_S))\big) = \bar{I}^{[O]}(\mathrm{ext}(x_S), \cdot)_* \bar{\mathbb{P}}(\bar{X}^{\mathcal{E}}_W).$$

The law of $\mathrm{res}(\bar{X}_O)$ under the embedded interventional distribution is the pushforward of $\bar{\mathbb{P}}(\bar{X}^{\mathcal{E}}_W)$ under

$$\bar{x}^{\mathcal{E}}_W \mapsto \mathrm{res}\big(\bar{I}^{[O]}(\mathrm{ext}(x_S), \bar{x}^{\mathcal{E}}_W)\big) \overset{(24)}{=} I^{[O]}\big(\mathrm{res}(\mathrm{ext}(x_S)), \mathrm{res}(\bar{x}^{\mathcal{E}}_W)\big) = I^{[O]}\big(x_S, \mathrm{res}(\bar{x}^{\mathcal{E}}_W)\big),$$

where the final equality uses $\mathrm{res} \circ \mathrm{ext} = \mathrm{id}$. By Definition F.1, $\bar{\mathbb{P}}(\mathrm{res}(\bar{X}^{\mathcal{E}}_W)) = \mathbb{P}(X^{\mathcal{E}}_W)$, so this pushforward equals $I^{[O]}(x_S, \cdot)_* \mathbb{P}(X^{\mathcal{E}}_W) = \mathbb{P}_{\mathcal{D}^{\mathcal{E}}}\big(X_O \,|\, \mathrm{do}(X_S = x_S)\big)$ as required. □

Theorem 14 (Markov properties for time-split systems of causal SDEs)**.** *Let $\mathcal{E}$ be a finite partition of $[0,T]$ into intervals with temporal ordering, and let $\mathcal{D}$ be a system of causal SDEs whose exogenous processes X_w ($w \in W$) have independent increments.*

(i) If $\mathcal{D}$ satisfies Assumption 1, then $\mathcal{D}^{\mathcal{E}}$ satisfies the σ-separation Markov property: for all $A, B, C_1, ..., C_r \subseteq V$ and $\mathcal{I}_A, \mathcal{I}_B, \mathcal{I}_{C_1}, ..., \mathcal{I}_{C_r} \subseteq \mathcal{E}$,

$$X^{\mathcal{I}_A}_A \underset{G(\mathcal{D}^{\mathcal{E}})}{\overset{\sigma}{\perp}} X^{\mathcal{I}_B}_B \,|\, X^{\mathcal{I}_{C_1}}_{C_1}, ..., X^{\mathcal{I}_{C_r}}_{C_r} \implies X^{\mathcal{I}_A}_A \perp\!\!\!\perp X^{\mathcal{I}_B}_B \,|\, X^{\mathcal{I}_{C_1}}_{C_1}, ..., X^{\mathcal{I}_{C_r}}_{C_r}.$$

(ii) If $\mathcal{D}$ satisfies Assumption 2, then $\mathcal{D}^{\mathcal{E}}$ satisfies the d-separation Markov property: for all $A, B, C_1, ..., C_r \subseteq V$ and $\mathcal{I}_A, \mathcal{I}_B, \mathcal{I}_{C_1}, ..., \mathcal{I}_{C_r} \subseteq \mathcal{E}$,

$$X^{\mathcal{I}_A}_A \underset{G(\mathcal{D}^{\mathcal{E}})}{\overset{d}{\perp}} X^{\mathcal{I}_B}_B \,|\, X^{\mathcal{I}_{C_1}}_{C_1}, ..., X^{\mathcal{I}_{C_r}}_{C_r} \implies X^{\mathcal{I}_A}_A \perp\!\!\!\perp X^{\mathcal{I}_B}_B \,|\, X^{\mathcal{I}_{C_1}}_{C_1}, ..., X^{\mathcal{I}_{C_r}}_{C_r}.$$

Proof. (i) By Lemma F.2, $\mathcal{D}^{\mathcal{E}}$ is essentially uniquely solvable w.r.t. every $O \subseteq V \times \mathcal{E}$. By Lemma F.3(i), the embedded time-split system $\bar{\mathcal{D}}^{\mathcal{E}}$ is also essentially uniquely solvable as a system of causal SDEs on $[0,T]$, so the σ-separation Markov property applies to it. Suppose $X^{\mathcal{I}_A}_A \perp^{\sigma}_{G(\mathcal{D}^{\mathcal{E}})} X^{\mathcal{I}_B}_B \,|\, X^{\mathcal{I}_{C_1}}_{C_1}, \ldots, X^{\mathcal{I}_{C_r}}_{C_r}$. By Lemma F.3(ii), $G(\bar{\mathcal{D}}^{\mathcal{E}}) = G(\mathcal{D}^{\mathcal{E}})$, so the same σ-separation holds in $G(\bar{\mathcal{D}}^{\mathcal{E}})$ between the corresponding embedded variables: $\bar{X}^{\mathcal{I}_A}_A \perp^{\sigma}_{G(\bar{\mathcal{D}}^{\mathcal{E}})} \bar{X}^{\mathcal{I}_B}_B \,|\, \bar{X}^{\mathcal{I}_{C_1}}_{C_1}, \ldots, \bar{X}^{\mathcal{I}_{C_r}}_{C_r}$. The σ-separation Markov property in $\bar{\mathcal{D}}^{\mathcal{E}}$ then gives $\bar{X}^{\mathcal{I}_A}_A \perp\!\!\!\perp \bar{X}^{\mathcal{I}_B}_B \,|\, \bar{X}^{\mathcal{I}_{C_1}}_{C_1}, \ldots, \bar{X}^{\mathcal{I}_{C_r}}_{C_r}$ under $\bar{\mathbb{P}}$. By

Lemma F.3(iii) with $S = \emptyset$, $\bar{\mathbb{P}}(\mathrm{res}(\bar{X}_V)) = \mathbb{P}(X_V)$, and hence $X_A^{\mathcal{I}_A} \perp\!\!\!\perp X_B^{\mathcal{I}_B} \mid X_{C_1}^{\mathcal{I}_{C_1}}, \ldots, X_{C_r}^{\mathcal{I}_{C_r}}$ under $\mathbb{P}$.

(ii) Let $E_{\mathcal{E}} := \bigcup_{\mathcal{I} \in \mathcal{E}} \{\inf \mathcal{I}, \sup \mathcal{I}\}$ be the set of end-points of the partition $\mathcal{E}$, and consider the Euler scheme X_V^{Δ} on a grid containing $E_{\mathcal{E}}$. Write $X_C^{\mathcal{I}_C} := (X_{C_1}^{\mathcal{I}_{C_1}}, \ldots, X_{C_r}^{\mathcal{I}_{C_r}})$ for brevity, and similarly $(X_C^{\Delta})^{\mathcal{I}_C}$ and $(X_C^{n})^{\mathcal{I}_C}$ for the Euler-grid and continuous-Euler analogues. Suppose $X_A^{\mathcal{I}_A} \perp^d X_B^{\mathcal{I}_B} \mid X_C^{\mathcal{I}_C}$ in the time-split graph $G(\mathcal{D}^{\mathcal{E}})$. By the same argument as Lemma 7 adapted to the time-split graph (collapsing time-points within each interval to its $(v, \mathcal{I})$-vertex respects interval membership since the Euler grid contains $E_{\mathcal{E}}$), this gives $(X_A^{\Delta})^{\mathcal{I}_A} \perp^d (X_B^{\Delta})^{\mathcal{I}_B} \mid (X_C^{\Delta})^{\mathcal{I}_C}$ in the Euler graph $G(\mathcal{M}_{\mathcal{D}}^{\Delta})$, where we write $(X_A^{\Delta})^{\mathcal{I}_A} = (X_A^{\Delta}(t_k) : t_k \in \mathcal{I}_A)$. Since the Euler graph is acyclic, the d-separation Markov property gives $(X_A^{\Delta})^{\mathcal{I}_A} \perp\!\!\!\perp (X_B^{\Delta})^{\mathcal{I}_B} \mid (X_C^{\Delta})^{\mathcal{I}_C}$. Let X_V^n be the continuous Euler scheme from (8). By the argument in the proof of Theorem 8 applied to the interval-restricted blocks — the gridpoints (functions of the increments) being independent of the Brownian bridges, with the gridpoint conditional independence above and the bridge conditional independence from the shared-noise structure — the continuous Euler scheme satisfies $(X_A^n)^{\mathcal{I}_A} \perp\!\!\!\perp (X_B^n)^{\mathcal{I}_B} \mid (X_C^n)^{\mathcal{I}_C}$. Let $\mathrm{proj}(X_V) := (X_A^{\mathcal{I}_A}, X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C})$, then

$$\begin{aligned} d_{TV}(\mathbb{P}(\mathrm{proj}(X_V^n)), \mathbb{P}(\mathrm{proj}(X_V))) &= \sup_{A \in \sigma(\mathrm{proj})} |\mathbb{P}(X_V^n \in A) - \mathbb{P}(X_V \in A)| \\ &\leq \sup_{A \in \mathcal{B}(C([0,T], \mathbb{R}^d))} |\mathbb{P}(X_V^n \in A) - \mathbb{P}(X_V \in A)| \\ &= d_{TV}(\mathbb{P}(X_V^n), \mathbb{P}(X_V)). \end{aligned}$$

By Theorem 9 there exists a subsequence n_m such that the right-hand side converges to 0, and by Theorem 10 we conclude that $X_A^{\mathcal{I}_A} \perp\!\!\!\perp X_B^{\mathcal{I}_B} \mid X_C^{\mathcal{I}_C}$. □

Theorem 15 (Do-calculus for time-split systems of causal SDEs)**.** *Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes X_w ($w \in W$) have independent increments, and let $\mathcal{E}$ be a finite partition of $[0, T]$ into intervals with temporal ordering. For $A, B, C_1, \ldots, C_r \subseteq V$ and $\mathcal{I}_A, \mathcal{I}_B, \mathcal{I}_{C_1}, \ldots, \mathcal{I}_{C_r} \subseteq \mathcal{E}$ such that the blocks $A^{\mathcal{I}_A}, B^{\mathcal{I}_B}, C_1^{\mathcal{I}_{C_1}}, \ldots, C_r^{\mathcal{I}_{C_r}}$ are pairwise disjoint in $V \times \mathcal{E}$, write $X_C^{\mathcal{I}_C} := (X_{C_1}^{\mathcal{I}_{C_1}}, \ldots, X_{C_r}^{\mathcal{I}_{C_r}})$. Then:*

1. Insertion/deletion of observation. *If $X_A^{\mathcal{I}_A} \perp^{\sigma}_{G(\mathcal{D}^{\mathcal{E}})} X_B^{\mathcal{I}_B} \mid X_C^{\mathcal{I}_C}$, then*

$$\mathbb{P}(X_A^{\mathcal{I}_A} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}) = \mathbb{P}(X_A^{\mathcal{I}_A} \mid X_C^{\mathcal{I}_C}) \quad \mathbb{P}(X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C})\text{-a.s.}$$

2. Action/observation exchange. *If $X_A^{\mathcal{I}_A} \perp^{\sigma}_{G(\mathcal{D}^{\mathcal{E}})_{\mathrm{do}(R_B^{\mathcal{I}_B})}} R_B^{\mathcal{I}_B} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}$ and $\mathbb{P}(X_C^{\mathcal{I}_C} \mid X_B^{\mathcal{I}_B} = x_B) \ll \mathbb{P}(X_C^{\mathcal{I}_C} \mid \mathrm{do}(X_B^{\mathcal{I}_B} = x_B))$ for $\mathbb{P}(X_B^{\mathcal{I}_B})$-a.a. x_B, then*

$$\mathbb{P}(X_A^{\mathcal{I}_A} \mid \mathrm{do}(X_B^{\mathcal{I}_B}), X_C^{\mathcal{I}_C}) = \mathbb{P}(X_A^{\mathcal{I}_A} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}) \quad \mathbb{P}(X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C})\text{-a.s.}$$

3. Insertion/deletion of action. *If $X_A^{\mathcal{I}_A} \perp^{\sigma}_{G(\mathcal{D}^{\mathcal{E}})_{\mathrm{do}(R_B^{\mathcal{I}_B})}} R_B^{\mathcal{I}_B} \mid X_C^{\mathcal{I}_C}$, then for every $x_B \in \mathcal{X}_{B^{\mathcal{I}_B}}$ such that $\mathbb{P}(X_C^{\mathcal{I}_C}) \ll \mathbb{P}(X_C^{\mathcal{I}_C} \mid \mathrm{do}(X_B^{\mathcal{I}_B} = x_B))$,*

$$\mathbb{P}(X_A^{\mathcal{I}_A} \mid \mathrm{do}(X_B^{\mathcal{I}_B} = x_B), X_C^{\mathcal{I}_C}) = \mathbb{P}(X_A^{\mathcal{I}_A} \mid X_C^{\mathcal{I}_C}) \quad \mathbb{P}(X_C^{\mathcal{I}_C})\text{-a.s.}$$

Proof. We prove Rule 2 explicitly; Rules 1 and 3 follow from similar arguments. Suppose that for $\mathcal{D}^{\mathcal{E}}$ we have $X_A^{\mathcal{I}_A} \perp^{\sigma}_{G(\mathcal{D}^{\mathcal{E}})_{\mathrm{do}(R_B^{\mathcal{I}_B})}} R_B^{\mathcal{I}_B} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}$, and $\mathbb{P}_{\mathcal{D}^{\mathcal{E}}}(X_C^{\mathcal{I}_C} \mid X_B^{\mathcal{I}_B} = x_B) \ll \mathbb{P}_{\mathcal{D}^{\mathcal{E}}}(X_C^{\mathcal{I}_C} \mid \mathrm{do}(X_B^{\mathcal{I}_B} = x_B))$ for $\mathbb{P}_{\mathcal{D}^{\mathcal{E}}}(X_B^{\mathcal{I}_B})$-a.a. x_B. By Lemma F.3(ii), $G(\bar{\mathcal{D}}^{\mathcal{E}}) = G(\mathcal{D}^{\mathcal{E}})$, and so $X_A^{\mathcal{I}_A} \perp^\sigma_{G(\bar{\mathcal{D}}^{\mathcal{E}})_{\mathrm{do}(R_B^{\mathcal{I}_B})}} R_B^{\mathcal{I}_B} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}$ as well. By Lemma F.3(iii), for every $S \subseteq V \times \mathcal{E}$ (write $O := (V \times \mathcal{E}) \setminus S$) and every $x_S \in \mathcal{X}_S$,

$$\mathbb{P}_{\mathcal{D}^{\mathcal{E}}}\big(X_O \mid \mathrm{do}(X_S = x_S)\big) = \mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}\big(\mathrm{res}(\bar{X}_O) \mid \mathrm{do}(\bar{X}_S = \mathrm{ext}(x_S))\big), \tag{25}$$

and similarly:

$$\mathbb{P}_{\mathcal{D}^{\mathcal{E}}}\big(\mathrm{ext}(X_O) \mid \mathrm{do}(X_S = x_S)\big) = \mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}\big(\bar{X}_O \mid \mathrm{do}(\bar{X}_S = \mathrm{ext}(x_S))\big). \tag{26}$$

Here (26) follows by applying ext to (25): by Lemma F.3(i) the intervened solution of $\bar{\mathcal{D}}^{\mathcal{E}}$ is $\bar{X}_O = \bar{I}^{[O]}(\mathrm{ext}(x_S), \bar{X}_W^{\mathcal{E}}) = \mathrm{ext}\big(I^{[O]}(x_S, X_W^{\mathcal{E}})\big)$, which lies in the image of ext, so $\bar{X}_O = \mathrm{ext}(\mathrm{res}(\bar{X}_O))$. Since absolute continuity is preserved under push-forwards, (26) gives $\mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}(\bar{X}_C^{\mathcal{I}_C} \mid \bar{X}_B^{\mathcal{I}_B} = \bar{x}_B) \ll \mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}(\bar{X}_C^{\mathcal{I}_C} \mid \mathrm{do}(\bar{X}_B^{\mathcal{I}_B} = \bar{x}_B))$ for $\mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}(\bar{X}_B^{\mathcal{I}_B})$-almost all $\bar{x}_B$. The embedded time-split system $\bar{\mathcal{D}}^{\mathcal{E}}$ is a system of causal SDEs on $[0, T]$; since $\mathcal{D}^{\mathcal{E}}$ is essentially uniquely solvable w.r.t. every $O \subseteq V \times \mathcal{E}$ (Lemma F.2), so is $\bar{\mathcal{D}}^{\mathcal{E}}$ by Lemma F.3(i), and Theorem 12 applies to $\bar{\mathcal{D}}^{\mathcal{E}}$. Applying Rule 2 to $\bar{\mathcal{D}}^{\mathcal{E}}$ then gives

$$\mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}\big(\bar{X}_A^{\mathcal{I}_A} \mid \bar{X}_B^{\mathcal{I}_B}, \bar{X}_C^{\mathcal{I}_C}\big) = \mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}\big(\bar{X}_A^{\mathcal{I}_A} \mid \mathrm{do}(\bar{X}_B^{\mathcal{I}_B}), \bar{X}_C^{\mathcal{I}_C}\big) \quad \mathbb{P}_{\bar{\mathcal{D}}^{\mathcal{E}}}(\bar{X}_B^{\mathcal{I}_B}, \bar{X}_C^{\mathcal{I}_C})\text{-a.s.}$$

Componentwise pushforward by res (using (25)) yields $\mathbb{P}_{\mathcal{D}^{\mathcal{E}}}(X_A^{\mathcal{I}_A} \mid X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C}) = \mathbb{P}_{\mathcal{D}^{\mathcal{E}}}(X_A^{\mathcal{I}_A} \mid \mathrm{do}(X_B^{\mathcal{I}_B}), X_C^{\mathcal{I}_C})$ $\mathbb{P}_{\mathcal{D}^{\mathcal{E}}}(X_B^{\mathcal{I}_B}, X_C^{\mathcal{I}_C})$-a.s., establishing Rule 2 for $\mathcal{D}^{\mathcal{E}}$. □

F.1 Proofs of Section 7.1

Theorem 16. *Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes X_w ($w \in W$) have independent increments, and let $u, v \in V$.*

(i) Let $\mathcal{D}^{\{[0,s],(s,T]\}}$ be faithful for every $0 \le s < T$. If there is a directed path from u to v in $G(\mathcal{D})$, then u is a Granger cause of v.

(ii) Let $G(\mathcal{D})$ have no bidirected edges. If u is a Granger cause of v, then there is a directed path from u to v in $G(\mathcal{D})$.

Proof. *(i)* Let $u = z_0 \to z_1 \to \cdots \to z_n = v$ be a directed path in $G(\mathcal{D})$, and fix any $0 \le s < T$. By Definition 8, the causal SDE for $X_u^{(s,T]}$ contains the initial-value term $X_u^{[0,s]}(b)$, giving the self-edge $u^{[0,s]} \to u^{(s,T]}$ in $G(\mathcal{D}^{\{[0,s],(s,T]\}})$. Within $(s, T]$, the causal SDE for each $X_{z_{i+1}}^{(s,T]}$ has parent set $\alpha(z_{i+1}) \cup \beta(z_{i+1}) \cup \gamma(z_{i+1})$, inheriting the edge $z_i^{(s,T]} \to z_{i+1}^{(s,T]}$ from $z_i \to z_{i+1}$ in $G(\mathcal{D})$. Composing these yields the walk

$$u^{[0,s]} \to u^{(s,T]} \to z_1^{(s,T]} \to \cdots \to z_{n-1}^{(s,T]} \to v^{(s,T]}$$

in $G(\mathcal{D}^{\{[0,s],(s,T]\}})$. Every non-endpoint node lies in $(s, T]$ and so is not in the conditioning set $X_{V\setminus\{u\}}^{[0,s]}$, so the walk is σ-active. By faithfulness of $\mathcal{D}^{\{[0,s],(s,T]\}}$, $X_u^{[0,s]} \not\perp\!\!\!\perp X_v^{(s,T]} \mid X_{V\setminus\{u\}}^{[0,s]}$, so u is a Granger cause of v.

(ii) We prove the contrapositive. Suppose there is no directed path from u to v in $G(\mathcal{D})$, and there is no latent confounding (so $G(\mathcal{D}^{\{[0,s],(s,T]\}})$ contains no bidirected edges). Fix $0 \le s < T$ and write $C := X_{V\setminus\{u\}}^{[0,s]}$. Let π be any walk from $u^{[0,s]}$ to $v^{(s,T]}$; we show that it is σ-blocked given C. Let $w^{[0,s]}$ be the *last* node of π lying in $[0, s]$, so all subsequent nodes lie in $(s, T]$. By adaptedness $G(\mathcal{D}^{\{[0,s],(s,T]\}})$ has no edges from $(s, T]$ to $[0, s]$, so the edge from $w^{[0,s]}$ to the next node is of the form $w^{[0,s]} \to w'^{(s,T]}$; hence $w^{[0,s]}$ is a non-collider that points to $w'^{(s,T]}$, which lies in a different interval and thus a different strongly connected component (SCCs of the time-split graph never span intervals), so $w^{[0,s]}$ is blockable. If $w \neq u$, then $w^{[0,s]} \in C$ is a blockable

non-collider, so π is σ-blocked. If $w = u$, then every node of π after $u^{[0,s]}$ lies in $(s,T]$. Suppose π were σ-open; since $C \subseteq [0,s]$ and no $(s,T]$-node is an ancestor of a $[0,s]$-node, any collider on the portion after $u^{[0,s]}$ would lie outside $\mathrm{Anc}(C)$ and block π, so this portion is collider-free. As its first edge $u^{[0,s]} \to w'^{(s,T]}$ is outgoing, it is then a directed path $u^{[0,s]} \to w'^{(s,T]} \to \cdots \to v^{(s,T]}$, giving a directed path from u to v in $G(\mathcal{D})$, contradicting the assumption; hence π is σ-blocked. Therefore every walk from $u^{[0,s]}$ to $v^{(s,T]}$ is σ-blocked given C, and by the σ-separation Markov property (Theorem 14(i)), $X_u^{[0,s]} \perp\!\!\!\perp X_v^{(s,T]} \,|\, X_{V\setminus\{u\}}^{[0,s]}$ for every $0 \leq s < T$. Hence u is not a Granger cause of v. $\square$

F.1.1 Proofs of Section 7.1.1

Lemma F.4. *If X_B is a semimartingale with respect to $\mathcal{F}^V$ such that $\sup_{t\in[0,T]} \|X_B(t)\| \leq c$ almost surely for some deterministic $c < \infty$, then for all $C \subseteq V$, the optional projection $\mathbb{E}[X_B(t) \,|\, \mathcal{F}_t^C]$ exists and is a special semimartingale.*

Proof. If X_B is bounded, then the optional projection $\mathbb{E}[X_B(t) \,|\, \mathcal{F}_t^C]$ exists and is bounded (see Protter (2005) VI.2 at the definition of optional projection).

Since X_B is bounded, the process $|X_B(t)|$ is bounded as well, hence the optional projection $\mathbb{E}[|X_B(t)| \,|\, \mathcal{F}_t^C]$ exists and is bounded. Since X_B is also a semimartingale with respect to $\mathcal{F}^V$, it follows that the optional projection $\mathbb{E}[X_B(t) \,|\, \mathcal{F}_t^C]$ is a bounded semimartingale (Föllmer and Protter (2011), Theorem 3.5).

Every bounded process has bounded jumps and every semimartingale with bounded jumps is a special semimartingale (Protter (2005), III.7.35), so the bounded semimartingale $\mathbb{E}[X_B(t) \,|\, \mathcal{F}_t^C]$ is a special semimartingale. $\square$

In the original work of Florens and Fougere (1996), it is claimed for the setting where $B \subseteq C$ that if $X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \,|\, X_C^{[0,s]}$, then in the Doob-Meyer decomposition $X_B = \Lambda^C + M^C$, the $\mathcal{F}_t^C$-local martingale M^C is a $\mathcal{F}_t^{A,C}$-local martingale, in which case we also have $\Lambda^C = \Lambda^{A,C}$ and hence local independence. That M^C is a $\mathcal{F}_t^{A,C}$-local martingale should follow from their Corollary 2.1, saying that if $X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \,|\, X_C^{[0,s]}$, then any $\mathcal{F}_t^B$-adapted $\mathcal{F}_t^C$-local martingale is an $\mathcal{F}_t^{A,C}$-local martingale. However, the $\mathcal{F}_t^C$-local martingale M^C is not necessarily $\mathcal{F}_t^B$-adapted, so one cannot apply Corollary 2.1 to obtain that M^C is also a $\mathcal{F}_t^{A,C}$-local martingale. The following proof corrects this error and extends the theorem to the setting where B is not contained in C.

Theorem 17. *Let $\mathcal{D}$ be a system of causal SDEs satisfying Assumption 1 whose exogenous processes X_w $(w \in W)$ have independent increments, such that $\sup_{t\in[0,T]} \|X_V(t)\| \leq c$ almost surely for some deterministic $c < \infty$, and let $A, B, C \subseteq V$. If for all $0 < s < t \leq T$ we have $X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \,|\, X_C^{[0,s]}$ in the time-split system $\mathcal{D}^{\{[0,s],(s,t],(t,T]\}}$, then $X_A \not\to X_B \,|\, X_C$ and $M^C = M^{A,C}$. Conversely, if $B = C$, then $X_A \not\to X_B \,|\, X_C$ implies $X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \,|\, X_C^{[0,s]}$ for all $0 \leq s < t \leq T$.*

Proof. We must show that $\Lambda_t^C = \Lambda_t^{A,C}$ a.s. – we will derive this property via the definition of these processes as given in the proof of the Doob-Meyer decomposition from Kallenberg (2021), Theorem 10.5. Let $X_B^C(t) := \mathbb{E}[X_B(t) \,|\, \mathcal{F}_t^C]$ and $X_B^{A,C}(t) := \mathbb{E}[X_B(t) \,|\, \mathcal{F}_t^{A,C}]$. Consider the points $t_j^n := jT/2^n$ that define the dyadic partition of $[0,T]$. Define

$$(\alpha_T^C)^n := \sum_{j=1}^{2^n} \mathbb{E}[X_B^C(t_j^n) - X_B^C(t_{j-1}^n) \,|\, \mathcal{F}_{t_{j-1}^n}^C], \tag{27}$$

then $\{(\alpha_T^C)^n : n \geq 1\}$ is uniformly integrable, so by the Dunford-Pettis criterion there is a subsequence n_k such that $(\alpha_T^C)^{n_k}$ has a weak limit $\alpha_T^C \in L^1(\mathcal{F}_T^C)$. The Doob-Meyer decomposition is then given by

$$\Lambda_t^C := X_B^C(t) - \mathbb{E}[X_B^C(T) - \alpha_T^C \mid \mathcal{F}_t^C] \tag{28}$$

and $M_t^C := X_B^C(t) - \Lambda_t^C$. That M_t^C is a martingale is immediate, and see Kallenberg (2021) for a proof that Λ^C is a predictable finite-variation process. We obtain the Doob-Meyer decomposition of $X_B^{A,C}$ with respect to $\mathcal{F}_t^{A,C}$ similarly, via the last element $\alpha_T^{A,C} \in \mathcal{F}_T^{A,C}$ and defining $\Lambda^{A,C}$ in terms of $\alpha_T^{A,C}$. Since X_B is càdlàg, $X_B(s) = \lim_{\epsilon \downarrow 0} X_B(s+\epsilon)$, so $X_B(s)$ is measurable with respect to $\sigma(X_B^{(s,t]})$, and hence, for $0 < s < t$, the assumed conditional independence $X_A^{[0,s]} \perp\!\!\!\perp X_B^{(s,t]} \mid X_C^{[0,s]}$ implies $X_A^{[0,s]} \perp\!\!\!\perp X_B(s) \mid X_C^{[0,s]}$ and thus $X_B^C(s) = \mathbb{E}[X_B(s) \mid \mathcal{F}_s^C] = \mathbb{E}[X_B(s) \mid \mathcal{F}_s^{A,C}] = X_B^{A,C}(s)$. Since the optional projections are càdlàg and the filtrations right-continuous, letting $s \downarrow 0$ extends this identity to $s = 0$ and yields $\mathbb{E}[X_B(r) \mid \mathcal{F}_0^C] = \mathbb{E}[X_B(r) \mid \mathcal{F}_0^{A,C}]$ for every $r \in (0, T]$; hence $X_B^C = X_B^{A,C}$. We then see from (27) (and the tower property) that $(\alpha_T^C)^n = (\alpha_T^{A,C})^n$, and hence also $\alpha_T^C = \alpha_T^{A,C}$. We also have for any $t \in [0, T]$ and i such that $t_{i-1}^n \leq t < t_i^n$ that

$$\mathbb{E}[(\alpha_T^C)^n \mid \mathcal{F}_t^C] = \sum_{j=1}^{i} \mathbb{E}[X_B^C(t_j^n) - X_B^C(t_{j-1}^n) \mid \mathcal{F}_{t_{j-1}^n}^C] + \sum_{j=i+1}^{2^n} \mathbb{E}[X_B^C(t_j^n) - X_B^C(t_{j-1}^n) \mid \mathcal{F}_t^C],$$

which by the assumed conditional independence and the tower property is equal to $\mathbb{E}[(\alpha_T^C)^n \mid \mathcal{F}_t^{A,C}]$. Since the map $(\alpha_T^C)^n \mapsto \mathbb{E}[(\alpha_T^C)^n \mid \mathcal{F}_t^C]$ is continuous from the weak topology $\sigma(L^1(\mathcal{F}_T^C), L^\infty(\mathcal{F}_T^C))$ to $\sigma(L^1(\mathcal{F}_t^C), L^\infty(\mathcal{F}_t^C))$ (see e.g. the discussion in Dellacherie and Meyer (1978), II.42), we get that $\mathbb{E}[(\alpha_T^C)^{n_k} \mid \mathcal{F}_t^C]$ converges weakly to $\mathbb{E}[\alpha_T^C \mid \mathcal{F}_t^C]$, giving $\mathbb{E}[\alpha_T^C \mid \mathcal{F}_t^C] = \mathbb{E}[\alpha_T^C \mid \mathcal{F}_t^{A,C}]$. By the independence assumption we also have $\mathbb{E}[X_B(T) \mid \mathcal{F}_t^C] = \mathbb{E}[X_B(T) \mid \mathcal{F}_t^{A,C}]$, so from (28) we obtain $\Lambda^C = \Lambda^{A,C}$. Since $X_B^C = X_B^{A,C}$ this also gives $M^C = M^{A,C}$.

If $B = C$, the remaining implication follows directly from Brémaud and Yor (1978) Theorem 3, as also remarked by Florens and Fougere (1996). □

References

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

Aalen, O., Røysland, K., Gran, J., Kouyos, R., and Lange, T. (2016). Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Statistical Methods in Medical Research*, 25(5):2294–2314.

Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). *Survival and Event History Analysis: A Process Point of View*. Statistics for Biology and Health. Springer, New York.

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.

Assaad, C., Devijver, E., and Gaussier, E. (2022). Survey and Evaluation of Causal Discovery Methods for Time Series. *Journal of Artificial Intelligence Research*, 73:767–819.

Blom, T. and Mooij, J. (2023). Causality and independence in perfectly adapted dynamical systems. *Journal of Causal Inference*, 11(1).

Boeken, P. and Mooij, J. M. (2024). Dynamic Structural Causal Models.

Boeken, P., Skapinakis, E., Genin, K., and Mooij, J. M. (2026). Topological Criteria for Hypothesis Testing with Finite-Precision Measurements.

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. *The Annals of Statistics*, 49(5).

Bowsher, C. G. (2010). Stochastic kinetic models: Dynamic independence, modularity and graphs. *The Annals of Statistics*, 38(4):2242–2281.

Brémaud, P. and Yor, M. (1978). Changes of filtrations and of probability measures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 45(4):269–295.

Brück, F., Engelke, S., and Volgushev, S. (2026). Graph structure learning for stable processes.

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

Comte, F. and Renault, E. (1996). Noncausality in Continuous Time Models. *Econometric Theory*, 12(2):215–256.

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

Dawid, P. (2021). Decision-theoretic foundations for statistical causality. *Journal of Causal Inference*, 9(1):39–77.

Dellacherie, C. and Meyer, P. (1978). *Probabilities and Potential*. Number 29 in North-Holland Mathematics Studies. Hermann [u.a.], Paris.

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.

Eichler, M. (2012). Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153(1):233–268.

Eichler, M. and Didelez, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Analysis*, 16(1):3–32.

Elowitz, M. B. and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338.

Engelke, S., Ivanovs, J., and Thøstesen, J. (2024). L\'evy graphical models.

Ferreira, S. and Assaad, C. (2024). Identifying macro conditional independencies and macro total effects in summary causal graphs with latent confounding.

Florens, J. and Fougere, D. (1996). Noncausality in Continuous Time. *Econometrica*, 64(5):1195–1212.

Föllmer, H. and Protter, P. (2011). Local martingales and filtration shrinkage. *ESAIM: Probability and Statistics*, 15:S25–S38.

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. (2025). A Mathematical Introduction to Causality.

Gill, R. and Robins, J. (2001). Causal Inference for Complex Longitudinal Data: The Continuous Case. *The Annals of Statistics*, 29(6):1785–1811.

Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438.

Granger, C. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352.

Guan, V., Janssen, J., Rahmani, H., Warren, A., Zhang, S., Robeva, E., and Schiebinger, G. (2024). Identifying Drift, Diffusion, and Causal Structure from Temporal Snapshots.

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

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

Kallenberg, O. (1996). On the existence of universal functional solutions to classical SDE's. *The Annals of Probability*, 24(1).

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

Karandikar, R. L. (1995). On pathwise stochastic integration. *Stochastic Processes and their Applications*, 57(1):11–18.

Karatzas, I. and Shreve, S. (1988). *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer US, New York, NY.

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.

Lauritzen, S. (2024). Total variation convergence preserves conditional independence. *Statistics & Probability Letters*, 214:110200.

Liptser, R. and Shiryaev, A. (2001). *Statistics of Random Processes: I: General Theory*. Number 5 in Applications of Mathematics. Springer, Berlin, Germany ; Heidelberg, Germany.

Lok, J. (2008). Statistical Modeling of Causal Effects in Continuous Time. *The Annals of Statistics*, 36(3):1464–1507.

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., 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.

Manten, G., Casolo, C., Mogensen, S., and Kilbertus, N. (2025). An Asymmetric Independence Model for Causal Discovery on Path Spaces.

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

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

Mogensen, S., Malinsky, D., and Hansen, N. (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., Janzing, D., and Schölkopf, B. (2013). From Ordinary Differential Equations to Structural Causal Models: The deterministic case.

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

Nathaniel, J., Roesch, C., Buch, J., DeSantis, D., Rupe, A., Lamb, K., and Gentine, P. (2025). Deep Koopman operator framework for causal discovery in nonlinear dynamical systems.

Neal, R. M. (2000). On Deducing Conditional Independence from d-Separation in Causal Graphs with Feedback (Research Note). *Journal of Artificial Intelligence Research*, 12:87–91.

Niemiro, W. (2024). Causal graphs, composable stochastic processes and conditional independence. *Applicationes Mathematicae*, pages 1–22.

Pearl, J. (1993). Comment: Graphical Models, Causality and Intervention. *Statistical Science*, 8(3):266–269.

Pearl, J. (1995). Causal Diagrams for Empirical Research. *Biometrika*, 82(4):669–688.

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

Pearl, J. and Dechter, R. (1996). Identifying independencies in causal graphs with feedback. In *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, UAI'96, pages 420–426, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Peters, J., Bauer, S., and Pfister, N. (2020). Causal models for dynamical systems.

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. (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. (2024). A Skorohod measurable universal functional representation of solutions to semimartingale SDEs. *Stochastic Analysis and Applications*, 42(6):1137–1155.

Reisach, A. G., Suárez, A., Weichwald, S., and Chambaz, A. (2025). The Case for Time in Causal DAGs.

Reiter, N., Gerhardus, A., Wahl, J., and Runge, J. (2024). Causal Inference on Process Graphs, Part I: The Structural Equation Process Representation.

Richardson, T. (1996). A discovery algorithm for directed cyclic graphs. In *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, UAI'96, pages 454–461, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

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., Bongers, S., Schölkopf, B., and Mooij, J. (2018). From Deterministic ODEs to Dynamic Structural Causal Models. In *Conference on Uncertainty in Artificial Intelligence*.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11):eaau4996.

Ryalen, P. C., Stensrud, M. J., and Røysland, K. (2026). On causal inference with marked point process data.

Rytgaard, H., Gerds, T., and Van Der Laan, M. (2022). Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes. *The Annals of Statistics*, 50(5).

Scheffé, H. (1947). A Useful Convergence Theorem for Probability Distributions. *The Annals of Mathematical Statistics*, 18(3):434–438.

Schweder, T. (1970). Composable Markov Processes. *Journal of Applied Probability*, 7(2):400–410.

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

Spirtes, P. (1994). Conditional independence in directed cyclic graphical models for feedback. *Technical Report CMU-PHIL-54, Carnegie Mellon University*.

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

Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction, and Search*, volume 81 of *Lecture Notes in Statistics*. Springer, New York, NY.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, Mass, 2nd ed edition.

Spirtes, P., Meek, C., and Richardson, T. (1999). An Algorithm for Causal Inference in the Presence of Latent Variables and Selection Bias. In *Computation, Causation and Discovery*, pages 211–252. The MIT Press.

Strobl, E. (2019). A constraint-based algorithm for causal discovery with cycles, latent variables and selection bias. *International Journal of Data Science and Analytics*, 8(1):33–56.

Weinberger, N. (2026). Homeostasis and causal control. *Biology & Philosophy*, 41(2):18.

White, H. and Lu, X. (2010). Granger Causality and Dynamic Structural Systems. *Journal of Financial Econometrics*, 8(2):193–243.

Yamada, T. and Watanabe, S. (1971). On the uniqueness of solutions of stochastic differential equations. *Kyoto Journal of Mathematics*, 11(1).

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.