Proceedings of Machine Learning Research 275:1–26, 2025 4th Conference on Causal Learning and Reasoning

An Asymmetric Independence Model for Causal Discovery on Path Spaces

Georg Manten GEORG.MANTEN@TUM.DE
Technical University of Munich, Germany
Helmholtz Munich, Germany
Munich Center for Machine Learning (MCML), Germany

Cecilia Casolo CECILIA.CASOLO@TUM.DE
Technical University of Munich, Germany
Helmholtz Munich, Germany
Munich Center for Machine Learning (MCML), Germany

Søren Wengel Mogensen† SWM.FI@CBS.DK
Department of Finance, Copenhagen Business School, Denmark

Niki Kilbertus† NIKI.KILBERTUS@TUM.DE
Technical University of Munich, Germany
Helmholtz Munich, Germany
Munich Center for Machine Learning (MCML), Germany

Editors: Biwei Huang and Mathias Drton

Abstract

We develop the theory linking 'E-separation' in directed mixed graphs (DMGs) with conditional independence relations among coordinate processes in stochastic differential equations (SDEs), where causal relationships are determined by "which variables enter the governing equation of which other variables". We prove a global Markov property for cyclic SDEs, which naturally extends to partially observed cyclic SDEs, because our asymmetric independence model is closed under marginalization. We then characterize the class of graphs that encode the same set of independence relations, yielding a result analogous to the seminal 'same skeleton and v-structures' result for directed acyclic graphs (DAGs). In the fully observed case, we show that each such equivalence class of graphs has a greatest element as a parsimonious representation and develop algorithms to identify this greatest element from data. We conjecture that a greatest element also exists under partial observations, which we verify computationally for graphs with up to four nodes.

Keywords: causal discovery, stochastic processes, path space, E-separation, graphical models, conditional independence, independence model

1. Introduction and Related Work

Discovering causal relationships from observational data holds great promise across many domains: in biology one may aim at inferring which genes regulate which other genes from single cell transcriptomics; in medical settings, one strives to understand the causal interactions between different diseases and symptoms; in finance, successful trading strategies require an understanding of causal drivers; in complex engineered systems, uncovering the underlying causal processes is crucial for effective predictive maintenance. Causal relationships between different variables are often described by graphs. The predominant graphical framework for causal discovery (or causal structure

† These authors contributed equally.

learning) are (variants of) Structural Causal Models (SCM) (Pearl, 2009; Peters et al., 2017), which can be considered a description of the underlying data generating mechanism. SCMs encode causal assumptions in the form of Directed Acyclic Graphs (DAGs) and leverage information encoded in the DAG about the observed distribution in terms of conditional independencies for causal inferences (Pearl, 1995; Lauritzen, 2001). Learning causal structures from observational data within this framework is a high-impact endeavor that has received considerable attention over the recent decades (Spirtes et al., 2000, 1995; Zheng et al., 2018; Chickering, 2002; Vowels et al., 2022).

Most existing works consider 'static models', where independence relations are symmetric. In this work, we focus on dynamic systems that evolve in continuous time. Thus, instead of assuming a static joint observational distribution, we consider as observations multi-variate stochastic processes. For these systems, exploiting the direction of time, i.e., the fact that causal influences can never go from the future to the past, can provide additional valuable information. Leveraging this information requires an asymmetric independence notion that captures the fundamental asymmetry between past and future.

One natural candidate to model causal relations among multivariate stochastic processes is the notion of local independence, which evaluates the dependence of the present on the past among the different coordinate processes. Local independence captures arbitrarily 'fast' interactions between coordinate processes (Schweder, 1970; Mogensen et al., 2018). Various theoretically oriented works have proposed (conditional) local independence to infer (partial) causal graphs in, e.g., point process models (Didelez, 2008; Meek, 2014; Mogensen et al., 2018; Mogensen and Hansen, 2020). Unfortunately, as of now local independence cannot be tested in practice with only one exception: Christgau et al. (2023) propose a practical test specifically for counting processes. There exists no statistical test of local independence for other classes of stochastic processes such as diffusions.

Another common approach to causal discovery in dynamical settings is to extend the methodology from the static case by assuming stochastic processes to be observed on a fixed regular time grid. One can then 'unroll' the causal graph in time and assume a discrete (auto-regressive) law governing the observed data (Runge, 2018; Runge et al., 2019; Runge, 2020; Runge et al., 2023). The 'discrete regular time grid' assumption fundamentally limits these approaches: **(i)** They require observations of all coordinate processes at fixed, evenly spaced time points, disallowing for irregular sampling or missing observations. **(ii)** All inferences critically depend on the sampling frequency. **(iii)** Methods are typically sensitive to the unknown maximal interaction time (i.e., the lag), which has to be chosen heuristically. Practically, these approaches also suffer from requiring large numbers of (symmetric) conditional independence tests for causal discovery (Shah and Peters, 2020; Lundborg et al., 2022).

Recently, continuous time systems have received attention in approaches that examine dependence on entire path-segments. Laumann et al. (2023) developed a test for conditional independence tailored to functional data, including path-valued random variables. Their approach treats these variables as standard random variables, ignoring the temporal nature, and utilizes established results from static structural causal models (SCMs). They apply the resulting global Markov property for acyclic graphs to perform causal discovery using traditional algorithms (e.g., the PC algorithm (Spirtes et al., 2000)) by testing functional data on the full intervals. Instead, Manten et al. (2024) recently introduced both a conditional independence test and a Markov property that explicitly accounts for time. By partitioning the time interval into a past and a future segment, this approach enables more informative causal discovery. However, their approach is still restricted to *acyclic* SDE models. Boeken and Mooij (2024) apply the Markov property found by Mooij and Claassen

(2023) to path-valued random variables using *symmetric* σ-separation to model cyclic dependencies. While allowing for cyclic dependencies – arguably one of the key characteristics of dynamical systems – this approach again ignores the direction of time, leading to weaker results than can be achieved by exploiting temporal order.

In this paper, we develop a graphical framework that offers several contributions: **(i)** it allows for cycles (unlike Manten et al., 2024; Laumann et al., 2023); **(ii)** it leverages the direction of time (unlike Boeken and Mooij, 2024; Laumann et al., 2023); **(iii)** it is practically testable (unlike local independence methods Mogensen et al., 2018); and **(iv)** it can handle partial observations, i.e., still provides (partial) results in the presence of unobserved confounding. The last point necessitates a constraint-based approach, which we focus on in this work. We achieve **(i)-(iv)** by a novel, more informative asymmetric version of σ-separation, called E-separation. Our criterion can be practically tested in partially observed cyclic SDE models using the conditional independence tests developed by Manten et al. (2024). For empirical validation of our theoretical contributions, we provide two example experiments in Appendix B. In summary, our framework is strictly more comprehensive than existing methods to constraint-based causal discovery in continuous time dynamical systems along criteria **(i)-(iv)**.

2. Stochastic Differential Equations, Graphs, and Separation

Data generating process. We focus on processes arising from stochastic differential equations (SDEs). SDEs are often used to model a variety of systems in physics, health, finance, and beyond, and also allow for a natural causal interpretation between the different coordinate processes. Following Manten et al. (2024), we assume a stationary, path-dependent 'SDE model'

$$\begin{cases} \mathrm{d}X_t^k = \mu^k(X_{[0,t]})\mathrm{d}t + \sigma^k(X_{[0,t]})\mathrm{d}W_t^k, \\ X_0^k = x_0^k \qquad \text{for } k \in [d] := \{1,\dots,d\}\,, \end{cases} \tag{1}$$

with potentially multi-dimensional $X_t^k \in \mathbb{R}^{n_k}$, $n_k \in \mathbb{N}_{>0}$, Brownian motions $W_t^k \in \mathbb{R}^{m_k}$, $m_k \in \mathbb{N}_{>0}$ and drift $\mu^k : C^0([0,1], \mathbb{R}^n) \to \mathbb{R}^{n_k}$ and diffusion $\mu^k : \mathcal{X}' \to \mathbb{R}^{n_k \times m_k}$ being functions of solution paths of the SDE up to time t. We assume the initial conditions x_0^k and Brownian motions are jointly independent. The structure of Equation (1), where each coordinate process influences the differential of others, naturally defines a directed "dependency graph" $\mathcal{G} = (V, \mathcal{E}_d)$ with vertices $V = [d] := \{1,\dots,d\}$ representing the individual coordinate processes. In this graph, i is a parent of j ($i \in \mathrm{pa}_j^{\mathcal{G}}$) when either μ^j or σ^j is not constant in the i-th argument. Importantly, unlike Manten et al. (2024), we do not impose acyclicity on this graph and allow for partial observations.

Basic graph terminology. We follow common notation and terminology from Mogensen and Hansen (2020); Forré and Mooij (2023); Peters et al. (2017) and provide a concise summary of the required concepts here.

Definition 2.1 (Directed (Mixed) Graph, D(M)G) *A directed mixed graph (DMG) is a triple $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ with a set of nodes $V \cong [d]$, $d \in \mathbb{N}$ and two sets of edges, the directed edges $\mathcal{E}_d \subset V \times V =: \mathcal{E}_d(V)$ and the bidirected edges $\mathcal{E}_{bi} \subset ((V \times V))/\!\sim\, =: \mathcal{E}_{bi}(V)$ with equivalence relation $(v_1, v_2) \sim (v_3, v_4) :\Leftrightarrow ((v_1 = v_3) \wedge (v_2 = v_4)) \vee ((v_1 = v_4) \wedge (v_2 = v_3))$ and equivalence classes denoted by $[(j,k)]_{bi}$. $\mathcal{G}$ is called directed graph (DG), if $\mathcal{E}_{bi} = \emptyset$. The induced subgraph on $A \subset V$ by $\mathcal{G}$ is the DMG $\mathcal{G}\big|_A = (A, \mathcal{E}_d \cap \mathcal{E}_d(A), \mathcal{E}_{bi} \cap \mathcal{E}_{bi}(A))$.*

Throughout this paper, $\mathcal{G}$ will denote a DG or DMG, $V \cong [d]$ its node set (corresponding to coordinate processes) and $\mathcal{E}_d$, $\mathcal{E}_{bi}$ its edges. For simplicity, we sometimes write $v \in \mathcal{G}$ for $v \in V$, $v_1 \to v_2 \in \mathcal{G}$ for $(v_1, v_2) \in \mathcal{E}_d$ (we then say that the edge has a *tail* at v_1 and a *head* at v_2, also referred to as *edge marks*), and $v_1 \leftrightarrow v_2 \in \mathcal{G}$ for $[(v_1, v_2)]_{bi} \in \mathcal{E}_{bi}$. Moreover, $v_1 \circ\!\!\to v_2 \in \mathcal{G}$ symbolizes either $v_1 \to v_2 \in \mathcal{G}$ or $v_1 \leftrightarrow v_2 \in \mathcal{G}$ and $v_1 \sim v_2 \in \mathcal{G}$ or $v_1 \circ\!\!-\!\!\circ v_2 \in \mathcal{G}$ that $v_1 \to v_2 \in \mathcal{G}$ or $v_1 \leftarrow v_2 \in \mathcal{G}$ or $v_1 \leftrightarrow v_2 \in \mathcal{G}$, meaning that v_1 and v_2 are *adjacent* in $\mathcal{G}$. The circle $\circ$ is like a placeholder for either an 'arrowhead' or a 'tail'. For nodes $v, w \in \mathcal{G}$, a *walk from v to w in $\mathcal{G}$*, is a finite sequence $\{(v_k, e_k)\}_{k\in[n]}$ such that $v_0 := v$, $v_{n+1} := w$ and $e_k \in \{(v_k, v_{k+1}), (v_{k+1}, v_k), [v_k, v_{k+1}]_{bi}\}$ $\forall k \in [n]$ and often denoted $v = v_0 \overset{e_0}{\sim} v_1 \overset{e_1}{\sim} \dots \overset{e_n}{\sim} v_{n+1} = w$. We denote by π^{-1} the *inverse walk*, $w = v_{n+1} \overset{e_n}{\sim} v_n \overset{e_{n-1}}{\sim} \dots \overset{e_0}{\sim} v_0 = v$. The *trivial walk* (from v to v), $\emptyset$ is the walk consisting of only a single node $v \in \mathcal{G}$ and no edges. We call the walk $\pi : v \sim \dots \sim w$ *bidirected* (*directed*) if $e_k = [v_k, v_{k+1}]_{bi} \in \mathcal{E}_{bi}$ ($e_k = (v_k, v_{k+1}) \in \mathcal{E}_d$) for all $k \in [n]$. A walk $v = v_0 \sim \dots \sim v_{n+1} = w$ is called *path from v to w in $\mathcal{G}$*, if $|\{v_0, \dots, v_n\}| = n + 1$, i.e., no node besides the endpoint occurs more than once. Definitions of (bi)directedness for walks carry over directly from the corresponding definitions for paths.

Walks $\pi_1 = v \overset{e_0^1}{\sim} v_1^1 \overset{e_1^1}{\sim} \dots \overset{e_n^1}{\sim} w$ and $\pi_2 = v \overset{e_0^2}{\sim} v_1^2 \overset{e_1^2}{\sim} \dots \overset{e_m^2}{\sim} w$ are *endpoint-identical*, if the marks of e_0^1, e_0^2 at v and of e_n^2, e_m^2 at w agree. A DMG $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ is *acyclic*, if there exists no non-trivial directed walk $v \to \dots \to v$ for all $v \in V$ $\mathcal{G}$, making it an *acyclic directed mixed graph (ADMG)* and an *directed acyclic graph (DAG)* if in addition $\mathcal{E}_{bi} = \emptyset$.

Let $\mathcal{G}_1 = (V, \mathcal{E}_d^1, \mathcal{E}_{bi}^1)$, $\mathcal{G}_2 = (V, \mathcal{E}_d^2, \mathcal{E}_{bi}^2)$ be DMGs. We call $\mathcal{G}_1$ *subgraph* of $\mathcal{G}_2$ ($\mathcal{G}_2$ *supergraph* of $\mathcal{G}_1$) and write $\mathcal{G}_1 \subseteq \mathcal{G}_2$, if $\mathcal{E}_d^1 \subseteq \mathcal{E}_d^2$ and $\mathcal{E}_{bi}^1 \subseteq \mathcal{E}_{bi}^2$. Note: $\subseteq$ is a partial order over the set of DMGs over nodes V. Furthermore, we define the *complete DG (DMG)* to be the graph with all possible edges, $(V, \mathcal{E}_d = V \times V)$ ($(V, \mathcal{E}_d = V \times V, \mathcal{E}_{bi} = (V \times V)/\!\sim)$). Finally, for a node v in a DMG $\mathcal{G}$ we define its *parents* $\mathrm{pa}_v^{\mathcal{G}} := \{w \in \mathcal{G} \mid w \to v \in \mathcal{G}\}$, *children* $\mathrm{ch}_v^{\mathcal{G}} := \{w \in \mathcal{G} \mid v \to w \in \mathcal{G}\}$, *siblings* $\mathrm{sib}_v^{\mathcal{G}} := \{w \in \mathcal{G} \mid w \leftrightarrow v \in \mathcal{G}\}$, *ancestors* $\mathrm{an}_v^{\mathcal{G}} := \{w \in \mathcal{G} \mid \exists \text{ directed path } w \to \dots \to v \in \mathcal{G}\}$, *descendants* $\mathrm{de}_v^{\mathcal{G}} := \{w \in \mathcal{G} \mid \exists \text{ directed path } v \to \dots \to w \in \mathcal{G}\}$, and *non-descendants* $\mathrm{nd}_v^{\mathcal{G}} := V \setminus \mathrm{de}_v^{\mathcal{G}}$. All notions extend to sets of nodes by unions, e.g., $\mathrm{pa}_A^{\mathcal{G}} := \bigcup_{v\in A} \mathrm{pa}_v^{\mathcal{G}}$, with $\mathrm{nd}_A^{\mathcal{G}} := V \setminus \mathrm{de}_A^{\mathcal{G}}$ for $A \subseteq V$. Note, that $v \in \mathrm{an}_v^{\mathcal{G}}, A \subseteq \mathrm{an}_A^{\mathcal{G}}, v \in \mathrm{de}_v^{\mathcal{G}}, A \subseteq \mathrm{de}_A^{\mathcal{G}}$ because we allow for trivial paths.

When dealing with cycles, we also need the notion of the *strongly connected component of v in $\mathcal{G}$*, denoted by $\mathrm{scc}_v^{\mathcal{G}} := (\mathrm{an}_v^{\mathcal{G}} \cap \mathrm{de}_v^{\mathcal{G}})$, having $v \in \mathrm{scc}_v^{\mathcal{G}}$. The *set of strongly connected components* of $\mathcal{G}$ is indicated by $\mathbf{S}(\mathcal{G}) := \{A \subseteq V \mid \exists j \in V : A = \mathrm{scc}_j^{\mathcal{G}}\}$ and it also defines an equivalence relation on V given by $v \sim_1 w :\Leftrightarrow w \in \mathrm{scc}_v^{\mathcal{G}}$ such that the equivalence classes partition the node set V. The *DAG of strongly connected components* for a DG $\mathcal{G} = (V, \mathcal{E}_d)$ is denoted by $\mathbf{S}(\mathcal{G}) := (V/\!\sim_1, (\mathcal{E}_d \setminus \Delta_V)/\!\sim_2)$ with the equivalence relation $(v \to w) \sim_2 (v', w') :\Leftrightarrow (v \sim_1 v') \wedge (w \sim_1 w')$.

σ-/d-Separation. To make this paper self-contained, we proceed by briefly recalling the relevant existing separation notions. Given a DMG $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$, we call a node v_k (or rather the position $k \in \{0, \dots, n+1\}$) on a walk π a *non-collider on π*, if it is an end-point ($k \in \{0, n+1\}$), in a left-chain ($v_{k-1} \leftarrow v_k \leftarrow\!\!\circ v_{k+1}$), in a right-chain ($v_{k-1} \circ\!\!\to v_k \to v_{k+1}$), or in a fork ($v_{k-1} \leftarrow v_k \to v_{k+1}$). It is a *collider on π*, if $v_{k-1} \circ\!\!\to v_k \leftarrow\!\!\circ v_{k+1}$. The set of colliders/non-colliders of π is denoted coll_π/ncoll_π. We further call a non-collider v_k on walk π an *unblockable non-collider on π* if $k \notin \{0, n\}$ and it is in a left-chain ($v_{k-1} \leftarrow v_k \leftarrow\!\!\circ v_{k+1} \ \wedge \ v_{k-1} \in \mathrm{scc}_{v_k}^{\mathcal{G}}$), it is in a right-chain ($v_{k-1} \circ\!\!\to v_k \to v_{k+1} \ \wedge \ v_{k+1} \in \mathrm{scc}_{v_k}^{\mathcal{G}}$), or it is in a fork ($v_{k-1} \leftarrow v_k \to v_{k+1} \ \wedge \ (v_{k-1} \in$

$\mathrm{scc}^{\mathcal{G}}_{v_k} \wedge v_{k+1} \in \mathrm{scc}^{\mathcal{G}}_{v_k}$)). Otherwise, we call v_k a *blockable non-collider on* π. Similarly, we denote the set of blockable/unblockable non-colliders of π by $\mathrm{ncoll}_{b,\pi}/\mathrm{ncoll}_{ub,\pi}$.

With this setup, we can define *d-separation* for acyclic graphs (e.g., (Pearl, 1985) for DAGs). *d-separation:* A walk $\pi = v_0 \overset{e_0}{\sim} \dots \overset{e_n}{\sim} v_{n+1}$ is called *d-open given C or d-C-open* for $C \subseteq V$, if $\mathrm{coll}_\pi \subseteq \mathrm{an}^{\mathcal{G}}_C$ and $\mathrm{ncoll}_\pi \cap C = \emptyset$. Otherwise it is called *d-blocked given C or d-C-blocked*, meaning that $\mathrm{coll}_\pi \not\subseteq \mathrm{an}^{\mathcal{G}}_C$ or $\mathrm{ncoll}_\pi \cap C \neq \emptyset$. If each walk $\pi = a \sim \dots \sim b$ between sets $A, B \subseteq V$ is d-C-blocked, we call *A d-separated from B given C*, ($A \perp\!\!\!\perp^{\mathcal{G}}_d B \mid C$); otherwise we write $A \not\!\perp\!\!\!\perp^{\mathcal{G}}_d B \mid C$.

A more suitable extension to cyclic settings is the notion of σ-*separation*, introduced by Forré and Mooij (2017). σ-*separation:* A walk $\pi = v_0 \overset{e_0}{\sim} \dots \overset{e_n}{\sim} v_{n+1}$ is called *σ-open given C or σ-C-open* for $C \subseteq V$, if $\mathrm{coll}_\pi \subseteq \mathrm{an}^{\mathcal{G}}_C$ and $\mathrm{ncoll}_{b,\pi} \cap C = \emptyset$. Otherwise it is called *σ-blocked given C or σ-C-blocked*, meaning that $\mathrm{coll}_\pi \not\subseteq \mathrm{an}^{\mathcal{G}}_C$ or $\mathrm{ncoll}_{b,\pi} \cap C \neq \emptyset$. We analogously extend this notion to sets to define *A σ-separated from B given C*, ($A \perp\!\!\!\perp^{\mathcal{G}}_\sigma B \mid C$) and otherwise write $A \not\!\perp\!\!\!\perp^{\mathcal{G}}_\sigma B \mid C$.

3. A Dynamic Global Markov Property for DMGs on Path-Space

3.1. Lifted Dependency Graph and Separation

We now define an extension of the *lifted dependency graph* introduced in Manten et al. (2024), which enables us to leverage the direction of time.

Lifted dependency graph. For a DMG $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$, the *lifted dependency graph* $\tilde{\mathcal{G}}$ is the DMG $\tilde{\mathcal{G}} = (\tilde{V}, \tilde{\mathcal{E}}_d, \tilde{\mathcal{E}}_{bi})$, where $\tilde{V} := V_0 \sqcup V_1$ is a disjoint union of two copies of V, with nodes subscripted by $0/1$ to indicate set membership and edges $\tilde{\mathcal{E}}_d := \{(u_0, v_0), (u_0, v_1), (u_1, v_1) \ : \ (u, v) \in \mathcal{E}_d\}$ and $\tilde{\mathcal{E}}_{bi} := \{[(u_0, v_0)], [(u_0, v_1)], [(u_1, v_0)], [(u_1, v_1)] \ : \ [(u, v)] \in \mathcal{E}_{bi}\}$.

Next we introduce a graphical separation criterion on the original DMG $\mathcal{G}$ in terms of σ-separation and its lifted dependency graph.

E-Separation. For a DMG $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$, B is *E-separated* from A given C in $\mathcal{G}$ if

$$A_0 \perp\!\!\!\perp^{\tilde{\mathcal{G}}}_\sigma B_1 \mid C_0, C_1 \setminus B_1,$$

where $A, B, C \subset V$, $A_0 = \{a_0 : a \in A\} \subseteq \tilde{V}$ and $B_1 = \{b_1 : b \in B\} \subseteq \tilde{V}$, $C_0 := \{c_0 : c \in C\} \subseteq \tilde{V}$, $C_1 := \{c_1 : c \in C\} \subseteq \tilde{V}$ in the lifted dependency graph $\tilde{\mathcal{G}}$ of $\mathcal{G}$. We denote this E-separation by $A \perp\!\!\!\perp^{\mathcal{G}}_E B \mid C$. An example DMG $\mathcal{G}$ and its lifted dependency graph are shown in Fig. 1, illustrating how time-ordering can be modeled graphically. The asymmetric notion implies that X^3 cannot be E-separated from X^1 due to the edge $X^1_0 \to X^3_0$, however, X^1 can be E-separated from X^3 as each walk is blocked by the empty set. Moreover X^2 and X^4 can not be E-separated from X^1 due to the edge $X^1_0 \to X^2_1$ directly feeding into the future strongly connected component $\{X^2_1, X^4_1\}$. Therefore, X^1 and X^4 are not separable, even though there is no directed edge $X^1 \to X^4$.

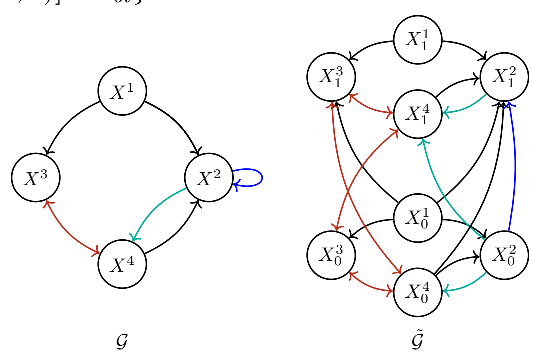

Figure 1: The lifted dependency graph $\tilde{\mathcal{G}}$ for the DMG $\mathcal{G}$.

3.2. Asymmetric Graphoids

Analogous to properties found in the literature on conditional independence models (see, e.g., Lauritzen et al., 1990), we can also define *abstract independence models* over a finite set V, generically denoted here by $\mathcal{I} \subseteq \mathcal{P}(V)^3$, which are sets of triples $(A, B, C) \subseteq V$ that admit for certain properties and encode important relations. For example, if (A, B, C) is inside $\mathcal{I}$, a set A is 'independent' from set B given set C (or 'separated' when referring to graphical criteria). Unlike traditional frameworks such as the notion of σ-separation, which is symmetric in A and B, the unidirectionality of time in our context necessitates an asymmetric independence model $\mathcal{I} \subseteq \mathcal{P}(V)^3$, whose properties are sometimes referred to as *asymmetric (semi) graphoid properties* (see Didelez (2008); Mogensen and Hansen (2020)):

(LR) *Left redundancy*: $(A, B, A) \in \mathcal{I}$

(RR) *Right redundancy*: $(A, B, B) \in \mathcal{I}$

(LD) *Left decomposition*: $(A, B, C) \in \mathcal{I}$, $D \subset A \Rightarrow (D, B, C) \in \mathcal{I}$

(RD) *Right decomposition*: $(A, B, C) \in \mathcal{I}$, $D \subset B \Rightarrow (A, D, C) \in \mathcal{I}$

(LWU) *Left weak union*: $(A, B, C) \in \mathcal{I}$, $D \subset A \Rightarrow (A, B, C \cup D) \in \mathcal{I}$

(RWU) *Right weak union*: $(A, B, C) \in \mathcal{I}$, $D \subset B \Rightarrow (A, B, C \cup D) \in \mathcal{I}$

(LC) *Left contraction*: $(A, B, C) \in \mathcal{I}$, $(D, B, A \cup C) \in \mathcal{I} \Rightarrow (A \cup D, B, C) \in \mathcal{I}$

(RC) *Right contraction*: $(A, B, C) \in \mathcal{I}$, $(A, D, B \cup C) \in \mathcal{I} \Rightarrow (A, B \cup D, C) \in \mathcal{I}$

(LI) *Left intersection*: $(A, B, C) \in \mathcal{I}$, $(C, B, A) \in \mathcal{I} \Rightarrow (A \cup C, B, A \cap C) \in \mathcal{I}$

(RR) *Right intersection*: $(A, B, C) \in \mathcal{I}$, $(A, C, B) \in \mathcal{I} \Rightarrow (A, B \cup C, B \cap C) \in \mathcal{I}$.

(LCo) *Left composition*: $(A, B, C) \in \mathcal{I} \Leftrightarrow (\{a\}, B, C) \in \mathcal{I}\ \forall\, a \in A$

(RCo) *Right composition*: $(A, B, C) \in \mathcal{I} \Leftrightarrow (A, \{b\}, C) \in \mathcal{I}\ \forall\, b \in B$

Proposition 3.1 (Ternary relation defined by $\perp\!\!\!\perp_E^{\mathcal{G}}$) *Let $\mathcal{G} = (V, \mathcal{E}_d)$ be a DG. Then $\perp\!\!\!\perp_E^{\mathcal{G}}$ defines the following ternary relation on the node set V,*

$$\mathcal{I}_E := \mathcal{I}_E^{\mathcal{G}} := \{(A, B, C) \in \mathcal{P}(V)^3 \ : \ A \perp\!\!\!\perp_E^{\mathcal{G}} B \mid C\}$$

which satisfies (LR), (LD), (RD), (LC), (LCo), (RCo).

We defer all proofs to Appendix A due to space restrictions. The proof of Proposition 3.1 is in Appendix A.1. Note that right redundancy (RR) does not hold, e.g., consider the graph $\mathcal{G} = (V = \{a, b\}, \mathcal{E}_d = \{(a, b)\})$. In addition, (LWU) does not hold with $\mathcal{G} = (V = \{a, b, d, e\}, \mathcal{E}_d = \{(a, e), (b, e), (e, d)\})$ yielding a counterexample for $A = \{a\}$, $B = \{b\}$, $D = \{d\}$ and $C = \emptyset$. For causal discovery, the notion of 'separability' is important, which we define for abstract independence models.

Definition 3.2 (Separability) *Let $\mathcal{I} \subseteq \mathcal{P}(V)^3$ be an independence model over V and $a, b \in V$. We call b separable from a if there exists a $C \subseteq V \setminus \{a\}$ such that $(\{a\}, \{b\}, C) \in \mathcal{I}$ and otherwise inseparable. We denote the set of nodes inseparable from a node b by*

$$u(b, \mathcal{I}_E^{\mathcal{G}}) := \left\{a \in V \ : \ (\{a\}, \{b\}, C) \notin \mathcal{I}_E^{\mathcal{G}},\ C \subseteq V \setminus \{a\}\right\} .$$

We adapt a result by Mogensen and Hansen (2020, Prop. 3.5) to relate paths and walks.

Lemma 3.3 *Let $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ be a DMG, $\tilde{G}$ its lifted dependency graph, $a, b \in V$ and $C \subseteq V \setminus \{a\}$. Then it holds: If there exists a $C_0 \sqcup C_1 \setminus \{b_1\}$-$\sigma$-open walk $\pi : a_0 \sim \ldots \sim b_1$ in $\tilde{G}$, there exists $C_0 \sqcup C_1 \setminus \{b_1\}$-$\sigma$-open path $\tilde{\pi} : a_0 \sim \ldots \sim b_1$ in $\tilde{\mathcal{G}}$ consisting of edges from π.*

The proof can be found in Appendix A.2.

3.3. Conditional Independence

In this section, we introduce our independence concept for stochastic dynamical systems based on an asymmetric notion of conditional independence that also takes into account time. We then show that it satisfies the global Markov property with respect to a DG using E-separation.

Definition 3.4 (Future-extended h-locally CI) *Let $\{X_t^i\}_{i\in[d]}$ be the coordinate processes of a solution of Eq. (1) for $t \in [0,1]$, $[0,s], [s, s+h] \subseteq [0,1]$ be subintervals for $s, h > 0$ and $A, B, C \subseteq [d]$. We say that X^A is future-extended h-locally conditionally independent (CI) of X^B given X^C at s, ($X^A \perp\!\!\!\perp_{s,h}^{+} X^B \mid X^C$) if*

$$X_{[0,s]}^{A} \perp\!\!\!\perp X_{[s,s+h]}^{B} \mid X_{[0,s]}^{C}, X_{[0,s+h]}^{C\setminus B} \tag{2}$$

Remark *By the CI statement (2), we refer to the independence of increments rather than the independence of consecutive path segments. By construction, path-valued random variables $\omega \mapsto ([a,b] \ni t \mapsto X_t^i(\omega) - X_a^i(\omega))$, $i \in [d]$, do not depend on the initial conditions and only on subsequent increments. For brevity, we denote these paths by X_t^i. We note that signature kernel–based CI-tests implicitly address this nuance (see, e.g., Manten et al. (2024)) as the signature transform is translation invariant ($S(X(t))_{a,b} = S(X(t) - X(a))_{a,b}$).*

Definition 3.4 differs from the one by Manten et al. (2024) in that B is not necessarily inside the conditioning set C. We now embed $\perp\!\!\!\perp_{s,h}^{+}$ into the context of asymmetric independence models.

Proposition 3.5 ($\perp\!\!\!\perp_{s,h}^{+}$ as a ternary relation) *Under the conditions of Definition 3.4, $\perp\!\!\!\perp_{s,h}^{+}$ defines the following ternary relation on V,*

$$\mathcal{I}_{s,h} = \{(A, B, C) \in \mathcal{P}(V)^3 \,:\, X^A \perp\!\!\!\perp_{s,h}^{+} X^B \mid X^C\}$$

which satisfies (LR) and (LD).

Because Propositions 3.1 and 3.5 show that both independence models satisfy (LR) and (LD), we assume from now on that $A \cap C = \emptyset$ in statements of the form $(A, B, C) \in \mathcal{I}_E$ ($\mathcal{I}_{s,h}$ respectively).

Proposition 3.6 (Global Markov property for E-separation and $\perp\!\!\!\perp_{s,h}^{+}$) *Let $\{X_t^i\}_{i\in[d]}$ be the coordinate processes of a solution of Eq. (1) for $t \in [0,1]$, $\mathcal{G} = (V \cong [d], \mathcal{E}_d)$ the adjacency graph defined by Eq. (1), $A, B, C \subseteq V$. Then*

$$A \perp\!\!\!\perp_E^{\mathcal{G}} B \mid C \;\Rightarrow\; X^A \perp\!\!\!\perp_{s,h}^{+} X^B \mid X^C. \tag{3}$$

The proof relies on an application of Forré and Mooij (2018, Theorem B.2) based on the concept of *acyclification* and is worked out in detail in Appendix A.3.

If the summary graph $\mathcal{G} = (V, \mathcal{E}_d)$ is acyclic except for self-loops, the lifted dependency graph $\tilde{\mathcal{G}}$ becomes a DAG, hence all non-colliders are blockable and E-separation reduces to $A_0 \perp\!\!\!\perp_d^{\tilde{\mathcal{G}}} B_1 \mid$

$C_0, C_1 \setminus B_1$. Hence, the Markov property found by Manten et al. (2024) is a simple special case of ours. This special case can be used in causal discovery algorithms that identify the full graph in the acyclic case (including self-loops) by testing $\perp\!\!\!\perp^+_{s,h}$ (under a faithfulness assumption). However, when allowing for cycles beyond self-loops, i.e., there are strongly connected components of size at least 2, self-loops cannot be distinguished anymore from data alone and multiple ground truth graphs are possible. As a result, a maximally informative graphical representation is desirable to capture all that can be consistently inferred about the dependency graph from observational data.

3.4. A Characterization of Markov Equivalence in E-Separation DGs

Several graphs may encode the same set of independence triples under some graphical separation criterion. When this occurs, the graphs are referred to as being *Markov equivalent* and one typically aims at characterizing and finding a useful representation of the entire Markov equivalence class. In the case of DAGs under d-separation, a common approach is to use a CPDAG to represent a Markov equivalence class of DAGs, despite the CPDAG itself not being a DAG (Pearl, 2009). In DGs with σ-separation, there need not exist a greatest element within an equivalence class of DGs (see Appendix C for an example). This section will demonstrate that for DGs under E-separation, each equivalence class contains a greatest element, which serves as an informative representative of the class.

Definition 3.7 (Markov equivalence) *Let $\mathcal{G}^1 = (V, \mathcal{E}_d^1, \mathcal{E}_{bi}^1)$, $\mathcal{G}^2 = (V, \mathcal{E}_d^2, \mathcal{E}_{bi}^2)$ be two DMGs over a common node set V. We say that $\mathcal{G}^1$ and $\mathcal{G}^2$ are (E-separation) Markov equivalent if $\mathcal{I}_E^{\mathcal{G}^1} = \mathcal{I}_E^{\mathcal{G}^2}$. E-separation Markov equivalence induces an equivalence relation on the set of DMGs over V, and we denote the (Markov) equivalence class of $\mathcal{G}^1$ by $[\mathcal{G}^1]_E$.*

Definition 3.8 (Maximal DMGs) *We call a DMG $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ maximal, if $\mathcal{G}$ is complete or for all $e \in (V \times V) \setminus \mathcal{E}_d$ we have $\mathcal{I}_E^{\mathcal{G} \cup \{e\}} \neq \mathcal{I}_E^{\mathcal{G}}$, where $\mathcal{G} \cup \{e\} := (V, \mathcal{E}_d \cup \{e\})$.*

Definition 3.9 (Greatest element) *We say that $\mathcal{G}_0$ is the greatest element of an equivalence class $[\mathcal{G}_1]_E$ if $\mathcal{G}_0 \in [\mathcal{G}_1]_E$ and $\mathcal{G}_2 \subseteq \mathcal{G}_0$ for all $\mathcal{G}_2 \in [\mathcal{G}_1]_E$.*

To characterize Markov equivalence classes, we first show that a strongly connected component can be shielded from the outside given itself and its parents. Note that $\mathrm{scc}_v^{\mathcal{G}} \subseteq \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_v^{\mathcal{G}}}$ if $|\mathrm{scc}_v^{\mathcal{G}}| \geq 2$.

Proposition 3.10 *Let $\mathcal{G} = (V, \mathcal{E}_d)$ be a DG and let $v \in V$. Then $(V \setminus \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_v^{\mathcal{G}}}, \mathrm{scc}_v^{\mathcal{G}}, \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_v^{\mathcal{G}}}) \in \mathcal{I}_E^{\mathcal{G}}$. Moreover, we have:*

(i) If $w \in V$, $w \neq v$ such that $\mathrm{scc}_v^{\mathcal{G}} \neq \mathrm{scc}_w^{\mathcal{G}}$, then either $(\{v\}, \{w\}, \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_w^{\mathcal{G}}}) \in \mathcal{I}_E^{\mathcal{G}}$ or $(\{w\}, \{v\}, \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_v^{\mathcal{G}}}) \in \mathcal{I}_E^{\mathcal{G}}$.

(ii) If $\{v\} = \mathrm{scc}_v^{\mathcal{G}}$, meaning v's strongly connected component is a singleton, then $(v, v) \notin \mathcal{E}_d$ if and only if $(\{v\}, \{v\}, \mathrm{pa}^{\mathcal{G}}) \in \mathcal{I}_E^{\mathcal{G}}$.

The proof can be found in Appendix A.5. Next, we characterize which types of edges can be added to a DG $\mathcal{G}$ and their impact on the underlying independence model $\mathcal{I}_E^{\mathcal{G}}$.

Lemma 3.11 *Let $\mathcal{G} = (V, \mathcal{E}_d)$ be a directed graph, $i, j \in V$, $(i, j) \notin \mathcal{E}_d$ and denote $\mathcal{G}' = (V, \mathcal{E}_d \cup \{(i, j)\})$. We then have for $i \neq j$:*

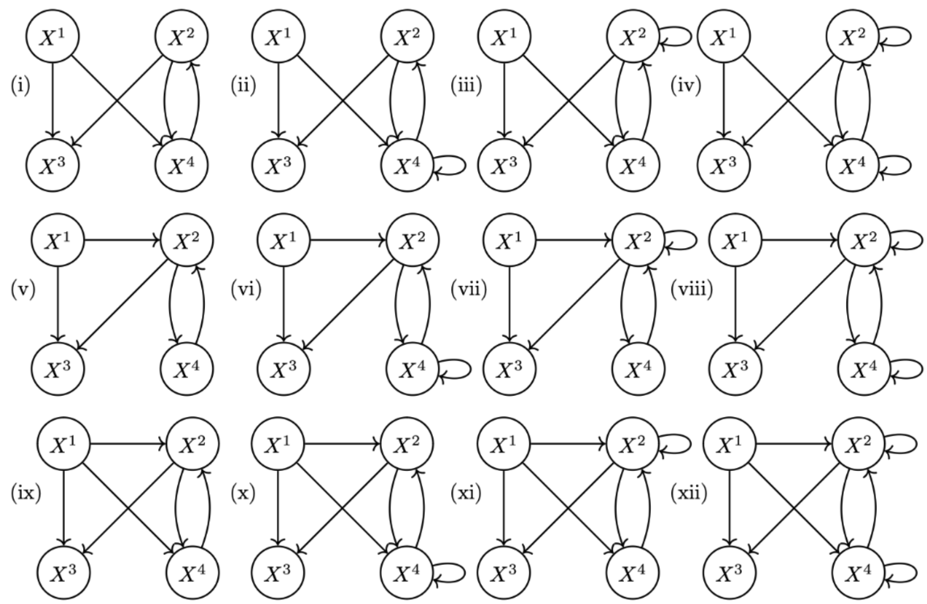

Figure 2: DGs (i)-(xii) are the elements of one Markov equivalence and (xii) is its greatest element.

(i) If $i \in \mathrm{scc}_j^{\mathcal{G}}$, *then* $\mathcal{I}_E^{\mathcal{G}} = \mathcal{I}_E^{\mathcal{G}'}$.

(ii) If $i \notin \mathrm{scc}_j^{\mathcal{G}}$ *and there is a* $j' \in \mathrm{scc}_j^{\mathcal{G}}$ *such that* $(i, j') \in \mathcal{E}_d$ *then* $\mathcal{I}_E^{\mathcal{G}} = \mathcal{I}_E^{\mathcal{G}'}$.

(iii) If $i \notin \mathrm{scc}_j^{\mathcal{G}}$ *and there is a* $j' \in \mathrm{scc}_j^{\mathcal{G}}$ *such that* $(i, j') \in \mathcal{E}_d$ *then* $\mathcal{I}_E^{\mathcal{G}} \neq \mathcal{I}_E^{\mathcal{G}'}$.

For $i = j$ *we have:*

(iv) If $|\mathrm{scc}_j^{\mathcal{G}}| = 1$, *then* $\mathcal{I}_E^{\mathcal{G}} \neq \mathcal{I}_E^{\mathcal{G}'}$.

The proof is in Appendix A.5. It works by showing that an open path in one graph can be used to construct an open path in the other graph. Lemma 3.11 explicitly states which edges can be added or removed from a graph to reach all other graphs within the same Markov equivalence class. It also implies that when $\mathcal{G}$ is an acyclic graph with potential self-loops, its Markov equivalence class consists solely of $\mathcal{G}$ itself, enabling identification of the full graph via causal discovery.

Figure 2 shows 12 DGs (i)-(xii) over 4 nodes X^1, X^2, X^3, X^4, all having the same set of irrelevance relations and with graph (xii) being the greatest element. This example illustrates the characterization in Lemma 3.11. Since $(\{X^1\}, \{X^1\}, \emptyset) \in \mathcal{I}_E$, none of the graphs can contain the self loop $(X^1, X^1) \in \mathcal{E}_d$, however no changes to $\mathcal{I}_E$ occur when adding self-loops to nodes of $\mathrm{scc}_{X^2}^{\mathcal{G}} = \{X^2, X^4\}$. Moreover there always exists an edge going from $\mathrm{scc}_{X^1}^{\mathcal{G}}$ into $\mathrm{scc}_{X^2}^{\mathcal{G}} = \{X^2, X^4\}$, but all options are possible. Finally, the edge $X^4 \to X^3$ can be excluded even though $X^2 \to X^4$ is always present. With Lemma 3.11 we can now prove that each Markov equivalence class of DGs has a greatest element.

Theorem 3.12 (Greatest Markov equivalent DG) *Let* $\mathcal{G} = (V, \mathcal{E}_d)$ *be a DG. The Markov equivalence class of* $\mathcal{G}$, $[\mathcal{G}]_E$, *contains a greatest element with respect to the partial order* $\subseteq$. *That is, there exists a* $\hat{\mathcal{G}} \in [\mathcal{G}]_E$ *such that* $\mathcal{G}' \subseteq \hat{\mathcal{G}}$ *for all* $\mathcal{G}' \in [\mathcal{G}]_E$.

Finally, let $\mathcal{G}_0, \mathcal{G}_1, \ldots, \mathcal{G}_m$, $\mathcal{G}_i = (V, \mathcal{E}_i)$ *be a finite sequence of DGs such that* $\mathcal{G}_0 = \hat{\mathcal{G}}$ *and* $\mathcal{E}_{i+1} = \mathcal{E}_i \setminus \{(v_i, w_i)\}$. *Then for any of the following cases* *(i)* $v_i \in \mathrm{scc}_{w_i}^{\mathcal{G}_i}$, $v_i \neq w_i$ *and* $\mathrm{scc}_{w_{i+1}}^{\mathcal{G}_i} \in \mathbf{S}[\mathcal{G}_{i+1}]$, *meaning the strongly connected component remains strongly connected after removing the edge;*

(ii) $v_i \notin \mathrm{scc}_{w_i}^{\mathcal{G}_i}$ *and* $|\{(v_i, w') \in \mathcal{E}_d^i \,:\, w' \in \mathrm{scc}_{w_i}^{\mathcal{G}_i}\}| \geq 2$*;*

(iii) $v_i = w_i$ *and* $|\mathrm{scc}_{w_i}^{\mathcal{G}_i}| \geq 2$

we have that $\mathcal{I}_E^{\mathcal{G}_i} = \mathcal{I}_E^{\hat{\mathcal{G}}} = \mathcal{I}_E^{\mathcal{G}}$ *for all* $i \in [m]$.

The proof is in Appendix A.5. DMGs have also been used as representations of so-called local independence in multivariate stochastic process using a graphical separation criterion known as μ-separation. Markov equivalence classes of DMGs under μ-separation, as well as so-called weak equivalence classes, also have a greatest element (Mogensen and Hansen, 2020; Mogensen, 2025).

In analogy to the DAG case, where two DAGs are Markov equivalent if and only if they have the same skeleton and the same v-structures (see, e.g., Lauritzen (2019, Theorem 2.62)), we can now establish the following graphical characterization result, which can be shown using Lemma 3.11.

Theorem 3.13 (Characterization of Markov Equivalent DGs) *Let* $\mathcal{G}_1 = (V, \mathcal{E}_d^1)$ *and* $\mathcal{G}_2 = (V, \mathcal{E}_d^2)$ *be DGs over* V. *Then* $\mathcal{I}_E^{\mathcal{G}_1} = \mathcal{I}_E^{\mathcal{G}_2}$ *is equivalent to the following three properties:*

(i) $\mathbf{S}(\mathcal{G}_1) = \mathbf{S}(\mathcal{G}_2)$, *i.e., the strongly connected components coincide;*

(ii) for $[v] \in \mathbf{S}(\mathcal{G}_1)$ *with* $|\mathrm{scc}_v^{\mathcal{G}_1}| = 1$

$$\mathrm{pa}_{\mathrm{scc}_v^{\mathcal{G}_1}} = \mathrm{pa}_{\mathrm{scc}_v^{\mathcal{G}_2}} \,;$$

(iii) for $[v] \in \mathbf{S}(\mathcal{G}_1)$ *with* $|\mathrm{scc}_v^{\mathcal{G}_1}| \geq 2$

$$\mathrm{pa}_{\mathrm{scc}_v^{\mathcal{G}_1}} \setminus \mathrm{scc}_v^{\mathcal{G}_1} = \mathrm{pa}_{\mathrm{scc}_v^{\mathcal{G}_2}} \setminus \mathrm{scc}_v^{\mathcal{G}_2} \,;$$

3.5. E-Separation in the Presence of Latent Variables

Latent (or unobserved) variables are typically represented in graphical frameworks by mixed edges, e.g., bi-directed edges. The main purpose of graphs with mixed edges is to graphically represent independencies in marginalized distributions (where the latents have been marginalized out). Accordingly, one can obtain graphs with mixed edges via a 'marginalization' operation on the DG that still contains the latent variables. For example, ADMGs are often obtained via a latent projection of a DAG and represents (conditional) independencies in the marginal joint distribution over observed variables.

In this section, we show that the global Markov property, which we have shown for DGs in Proposition 3.6, with respect to a DMG obtained via latent projection from a larger DG is inherited from the DG. To keep this paper self-contained, we briefly recap the notion of *latent projection* and *marginal independence model*.

Definition 3.14 (The latent projection) *Let* $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ *be a DMG,* $V = V_{obs} \,\dot\cup\, V_{lat}$. *The* latent projection of $\mathcal{G}$ on V_{obs} $:= \mathcal{G}' = (V_{obs}, \mathcal{E}_d', \mathcal{E}_{bi}')$ *with*

$$\mathcal{E}_d' := \big\{(v_1, v_2) \mid v_1, v_2 \in V_{obs} \text{ and there exists a non-trivial walk } \pi = \{(w_i, e_i)\}_{i\in[n]} = v_1 \to \ldots \to v_2 \text{ with } w_i \in V_{lat} \text{ for all } i \in [n]\big\}\,,$$

$$\mathcal{E}_{bi}' := \big\{[(v_1, v_2)] \mid v_1, v_2 \in V_{obs} \text{ and there exists a non-trivial walk } \pi = \{(w_i, e_i)\}_{i\in[n]} = v_1 \leftarrow\!\circ \ldots \circ\!\!\to v_2 \text{ with } w_i \in V_{lat} \text{ for all } i \in [n] \text{ such that } \mathrm{coll}_\pi = \emptyset\big\}\,.$$

Definition 3.15 (Marginal independence model) *Assume $\mathcal{I} \subseteq \mathcal{P}(V)^3$ is an abstract independence model over V. The marginal independence model of $\mathcal{I}$ over $V_{obs} \subseteq V$ is defined as*

$$\mathcal{I}\big|_{V_{obs}} := \left\{(A,B,C) \in \mathcal{I} \mid A,B,C \subseteq V_{obs}\right\}.$$

Analogous to other graphical frameworks (e.g., Mogensen and Hansen, 2020, Theorem 3.12) we can show that a DMG obtained by marginalization of a larger directed graph has the same independence relations among the observed ('not-marginalized-out') nodes than the original graph with respect to general E-separation.

Proposition 3.16 *Let $\mathcal{G} = (V, \mathcal{E}_d)$ be a directed graph, $V = V_{obs} \cup V_{lat}$, $A, B, C \subseteq V_{obs}$ (we assume without loss of generality that $A \cap C = \emptyset$) and $\mathcal{G}' = (V_{obs}, \mathcal{E}'_d, \mathcal{E}'_{bi})$ the latent projection of $\mathcal{G}$. Then*

$$(A,B,C) \in \mathcal{I}_E^{\mathcal{G}} \Leftrightarrow (A,B,C) \in \mathcal{I}_E^{\mathcal{G}'}.$$

The proof is in Appendix A.4. With this we can then extend the global Markov property from Proposition 3.6 to the partially observed setting.

Proposition 3.17 (Global Markov property for Latent Models) *Let $V \cong [d], d \in \mathbb{N}$, $\{X_t^i\}_{i \in V}$ the coordinate processes of a solution of Eq. (1) for $t \in [0,1]$, $\mathcal{G} = (V, \mathcal{E}_d)$ the adjacency graph defined by Eq. (1), $V = V_{obs} \cup V_{lat}$ a partition of V into observed and latent nodes and $\mathcal{G}' = (V_{obs}, \mathcal{E}'_d, \mathcal{E}'_{bi})$ the latent projection of $\mathcal{G}$ on V_{obs}. Then for $A, B, C \subseteq V_{obs}$ it holds*

$$A \perp\!\!\!\perp_E^{\mathcal{G}'} B \mid C \Rightarrow X^A \perp\!\!\!\perp_{s,h}^{+} X^B \mid X^C. \tag{4}$$

3.6. Properties of DMGs with respect to E-Separation

In this section, we investigate the properties of Markov equivalence classes of general DMGs with respect to the graphical separation criterion $\perp\!\!\!\perp_E$. We once more align our methodology with the approach outlined by Mogensen and Hansen (2020) for μ-separation.

The separability of pairs of nodes and their subsequent inclusion in the independence model is reflected by the presence of so-called *inducing walks* or *paths* in various graph classes and their graphical separation criteria (see, e.g., Zhang (2008) for ancestral graphs, Forré and Mooij (2023, Chapter 11) for CDMGs and σ-separation, or Mogensen and Hansen (2020) for DMGs and μ-separation). Since E-separation is an asymmetric version of σ-separation, we also have to define an altered, asymmetric version of inducing paths first. Figure 3 highlights that this definition has to incorporate that the last edge before entering a strongly connected component $\mathrm{scc}_w^{\mathcal{G}}$ has to point into the strongly connected component.

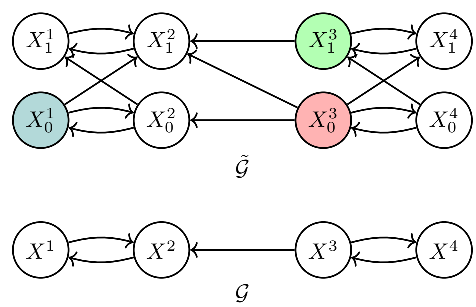

Figure 3: An inducing path connects X^1 and X^3 in $\mathcal{G}$, and X^3 (red) E-separates X^3 (green) from X^1 (teal).

Definition 3.18 ((Asymmetric) inducing walks) *Let $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ be a DMG and $v, w \in V$. An (asymmetric) inducing path from v to w in $\mathcal{G}$ is a non-trivial path $v = v_0 \overset{e_0}{\sim} v_1 \overset{e_1}{\sim} \dots \overset{e_{n-1}}{\sim} v_n \overset{e_n}{\sim} v_{n+1} = w$ such that $\mathrm{coll}_\pi \subseteq \mathrm{an}_{\{v,w\}}^{\mathcal{G}}$, each $v_i \in \mathrm{ncoll}_\pi \setminus \{v_0\}$ is unblockable, and the last edge before entering $\mathrm{scc}_w^{\mathcal{G}}$ has to be into the $\mathrm{scc}_w^{\mathcal{G}}$.*

We can now show the first part of the relation between separability and inducing paths.

Proposition 3.19 *Let $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ be a DMG. If there exists an (asymmetric) inducing path $\nu : v = v^0 \overset{e_0}{\sim} \ldots \overset{e_n}{\sim} w$ in $\mathcal{G}$ from v to w, then $(\{v\}, \{w\}, C) \notin \mathcal{I}_E$ for all $C \subseteq V \setminus \{v\}$, meaning that w is inseparable from v in $\mathcal{G}$.*

A direct consequence is the following 'symmetry' relation of inducing paths to nodes within a strongly connected component.

Lemma 3.20 *Let $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ be a DMG. If there exists an (asymmetric) inducing path $\nu : v = v^0 \overset{e_0}{\sim} \ldots \overset{e_n}{\sim} w$ in $\mathcal{G}$ from v to w and $u \in \mathrm{scc}_w^{\mathcal{G}}$ then there also exists an (asymmetric) inducing path from v to u.*

With a similar approach as in Lemma 3.11, we can now show the following characterizations of the independence model, which states that two nodes from different strongly connected components, that are connected by a bidirected edge can not be separated.

Lemma 3.21 *Let $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ be a DMG and $v, w \in V$ nodes such that $w \notin \mathrm{scc}_v^{\mathcal{G}}$. If there exists $v', w' \in V$ such that $v' \in \mathrm{scc}_v^{\mathcal{G}}$, $w' \in \mathrm{scc}_w^{\mathcal{G}}$ and $[(v', w')] \in \mathcal{E}_{bi}$ then*

$$(\{\hat{v}\}, \{\hat{w}\}, C) \notin \mathcal{I}_E \qquad \forall \hat{v}, \hat{w} \in \mathrm{scc}_v^{\mathcal{G}} \cup \mathrm{scc}_w^{\mathcal{G}}, C \subseteq V \setminus \{\hat{v}\} \,. \tag{5}$$

This follows from the existence of the path $\hat{v}_0 \leftarrow \ldots \leftarrow v'_0 \leftrightarrow w'_1 \rightarrow \ldots \rightarrow \hat{w}_0$ in $\tilde{\mathcal{G}}$. Finally, in Appendix A.7, we prove that one can and arbitrary many bidirected edges between strongly connected components that are already connected by such a bidirected edge without altering the independence model.

Proposition 3.22 *Let $\mathcal{G} = (V, \mathcal{E}_d, \mathcal{E}_{bi})$ be a DMG, $i, j \in V$, $i \neq j$, $[(i,j)] \notin \mathcal{E}_{bi}$ and denote $\mathcal{G}' = (V, \mathcal{E}_d, \mathcal{E}_{bi} \cup \{[(i,j)]\})$. Then, if $i \notin \mathrm{scc}_j^{\mathcal{G}}$ and there exist $i' \in \mathrm{scc}_i^{\mathcal{G}}, j' \in \mathrm{scc}_j^{\mathcal{G}}$ such that $[(i', j')] \in \mathcal{E}_{bi}$, we have that $\mathcal{I}_E^{\mathcal{G}} = \mathcal{I}_E^{\mathcal{G}'}$.*

A Conjecture about the greatest equivalent DMG. We conjecture that each E-separation Markov equivalence class of directed mixed graphs (DMGs) contains a greatest element, analogous to the findings for directed graphs (DGs) in Theorem 3.12 and to μ-separation DMGs (Mogensen and Hansen, 2022, Theorem 5.9). This conjecture is supported by experiments conducted on DMGs up to $d = 4$, which we describe in the following section.

4. Experiments

For initial insights into whether each E-separation Markov equivalence class of DMGs contains a greatest element, we computationally tested whether this is true exhaustively for all DMGs up to $d = 4$ nodes. After procedurally generating all possible DGs and DMGs with $d \in \{2, 3, 4\}$ nodes, we calculate all separation triples for each one. Among all graphs with the same separation triples (i.e., within each Markov equivalence class), we then exhaustively search for a greatest element. For $d = 4$, we searched over 65536 DGs and 4194304 DMGs, respectively. Due to left- and right composability of $\mathcal{I}_E$ (see Proposition 3.1), we only need to consider single nodes (instead of sets of nodes) in the first and second argument, which helped speeding up the computations. Our hypothesis holds true for up to 4 nodes and we hypothesize that it holds for all d, i.e., a greatest element exists not only for DGs, but also for all DMGs. Moreover, in Appendix B, we provide empirical evidence that the Markov equivalence class of SDE models of the form Eq. (1) can be estimated from data, serving as a proof-of-concept to validate our theoretical findings.

5. Discussion and Conclusion

In this work, we have extended the causal discovery framework of Manten et al. (2024) to cyclic and partially observed SDE-models. We introduce E-separation by partitioning the time interval into past and future segments to obtain an asymmetric separation criterion that is sensitive to the direction of time. E-separation thus provides an extension of σ-separation. Our framework accommodates not only cyclic dependencies—arguably ubiquitous in dynamical settings—but also partial observations. This facilitates causal discovery in realistic, continuous-time models. In the fully observed setting, we characterized the Markov equivalence class explicitly and showed the existence of a greatest element—a parsimonious representation of the Markov equivalence class that is discoverable from observational data (Appendix B). Furthermore, based on computational verification for DMGs with up to four nodes, we conjecture that E-separation Markov equivalence classes of DMGs also contain a greatest element.

Acknowledgments

This work is supported by the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research. This work was supported by the Helmholtz Association's Initiative and Networking Fund on the HAICORE@FZJ partition. We thank Emilio Ferrucci and Cristopher Salvi for useful discussions throughout the project. We also thank Emilio for being the namesake of E-separation.

References

Philip Boeken and Joris M. Mooij. Dynamic structural causal models, 2024. URL https://arxiv.org/abs/2406.01161.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.

Alexander Mangulad Christgau, Lasse Petersen, and Niels Richard Hansen. Nonparametric conditional local independence testing. *The Annals of Statistics*, 51(5):2116 – 2144, 2023. doi: 10.1214/23-AOS2323. URL https://doi.org/10.1214/23-AOS2323.

Vanessa Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):245–264, 2008.

Patrick Forré and Joris M. Mooij. Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. In Amir Globerson and Ricardo Silva, editors, *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018*, pages 269–278. AUAI Press, 2018. URL http://auai.org/uai2018/proceedings/papers/117.pdf.

Patrick Forré and Joris M. Mooij. Markov properties for graphical models with cycles and latent variables, 2017. URL https://arxiv.org/abs/1710.08775.

Patrick Forré and Joris M. Mooij. A mathematical introduction to causality: Lecture notes, September 2023. URL https://staff.fnwi.uva.nl/j.m.mooij/articles/causality_lecture_notes_2023.pdf. Accessed: 2024-09-12.

Felix Laumann, Julius Von Kügelgen, Junhyung Park, Bernhard Schölkopf, and Mauricio Barahona. Kernel-based independence tests for causal structure learning on functional data. *Entropy*, 25(12): 1597, 2023.

Steffen L Lauritzen. Causal inference from graphical models. *Monographs on Statistics and Applied Probability*, 87:63–108, 2001.

Steffen L. Lauritzen. *Lectures on Graphical Models, 3rd edition*. Department of Mathematical Sciences, Faculty of Science, University of Copenhagen, 2019. ISBN 978-87-70787-53-6.

Steffen L Lauritzen, A Philip Dawid, Birgitte N Larsen, and H-G Leimer. Independence properties of directed markov fields. *Networks*, 20(5):491–505, 1990.

Anton Rask Lundborg, Rajen D Shah, and Jonas Peters. Conditional independence testing in hilbert spaces with applications to functional data analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):1821–1850, 2022.

Georg Manten, Cecilia Casolo, Emilio Ferrucci, Søren Wengel Mogensen, Cristopher Salvi, and Niki Kilbertus. Signature kernel conditional independence tests in causal discovery for stochastic processes, 2024. URL https://arxiv.org/abs/2402.18477.

Christopher Meek. Toward learning graphical and causal process models. In *CI@ UAI*, pages 43–48, 2014.

Søren Wengel Mogensen. Weak equivalence of local independence graphs. *Bernoulli*, 2025. (to appear).

Søren Wengel Mogensen and Niels Richard Hansen. Markov equivalence of marginalized local independence graphs. *The Annals of Statistics*, 48(1):539 – 559, 2020.

Søren Wengel Mogensen and Niels Richard Hansen. Graphical modeling of stochastic processes driven by correlated noise. *Bernoulli*, 28(4):3023 – 3050, 2022.

Søren Wengel Mogensen, Daniel Malinsky, and Niels Richard Hansen. Causal learning for partially observed stochastic dynamical systems. In *Conference on Uncertainty in Artificial Intelligence,*, pages 350–360, 2018.

Joris M. Mooij and Tom Claassen. Constraint-based causal discovery using partial ancestral graphs in the presence of cycles, 2023.

Judea Pearl. A constraint - propagation approach to probabilistic reasoning. In *Proceedings of the First Conference on Uncertainty in Artificial Intelligence*, UAI'85, page 31–42, Arlington, Virginia, USA, 1985. AUAI Press. ISBN 0444700587.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. ISBN 0-521-77593-0. doi: 10.1017/CBO9781107590120. URL https://doi.org/10.1017/CBO9781107590120. Itô calculus, Reprint of the second (1994) edition.

Jakob Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075310, 2018.

Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence*, pages 1388–1397. PMLR, 2020.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science advances*, 5 (11):eaau4996, 2019.

Jakob Runge, Andreas Gerhardus, Gherardo Varando, Veronika Eyring, and Gustau Camps-Valls. Causal inference for time series. *Nature Reviews Earth & Environment*, 4(7):487–505, 2023.

Tore Schweder. Composable markov processes. *Journal of applied probability*, 7(2):400–410, 1970.

Rajen D Shah and Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. *The Annals of Statistics*, 48(3):1514–1538, 2020.

Peter Spirtes, Christopher Meek, and Thomas Richardson. Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 499–506, 1995.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

Matthew J Vowels, Necati Cihan Camgoz, and Richard Bowden. D'ya like dags? a survey on structure learning and causal discovery. *ACM Computing Surveys*, 55(4):1–36, 2022.

Jiji Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896, 2008. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2008.08.001. URL https://www.sciencedirect.com/science/article/pii/S0004370208001008.

Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. Dags with no tears: Continuous optimization for structure learning. *Advances in neural information processing systems*, 31, 2018.

Appendix A. Proofs

A.1. Proofs of Asymmetric Graphoid Properties Proposition 3.1 and Proposition 3.5

Proof [Proof of Proposition 3.1]

(LR): Let $A, B \subseteq V$ and $\pi = a_0 \sim \ldots \sim b^1$ a walk in $\tilde{\mathcal{G}}$, $a \in A$, $b \in B$. As conditioning on an endpoint σ-blocks the walk, clearly π is σ-$(A_0 \cup (A_1 \setminus B_1))$-blocked.

(LD): Assume $(A, B, C) = (A' \cup D, B, C) \in \mathcal{I}_E^{\mathcal{G}}$ with $A' := A \setminus D$, $d \in D$, $b \in B$ and let $\pi = d_0 \sim \ldots \sim b^1$ be a walk in $\tilde{\mathcal{G}}$. Then π is also a walk from A_0 to B_1. By assumption it is σ-$(C_0 \cup (C_1 \setminus B_1))$-blocked. Thus $(D, B, C) \in \mathcal{I}_E^{\mathcal{G}}$.

(LC): We have to show that $(A, B, C) \in \mathcal{I}_E^{\mathcal{G}} \wedge (D, B, A \cup C) \in \mathcal{I}_E^{\mathcal{G}}$ implies $(A \cup D, B, C) \in \mathcal{I}_E^{\mathcal{G}}$. Assume instead $(A \cup D, B, C) \notin \mathcal{I}_E^{\mathcal{G}}$; let $\pi = a_0 \sim \ldots \sim b_1$, $b \in B$, $a \in A \cup D$ σ-$(C_0 \cup (C_1 \setminus B_1))$-open shortest path. This means the colliders satisfy $\mathrm{coll}_\pi \subseteq \mathrm{an}^{\tilde{\mathcal{G}}}_{(C_0 \cup (C_1 \setminus B_1))}$ and for the blockable noncolliders it holds $\mathrm{ncoll}_{b,\pi} \cap (C_0 \cup (C_1 \setminus B_1)) = \emptyset$. Moreover by assumption $a_0 \in A_0 \cup D_0$, $b_1 \in B_1$ the only elements of $A_0 \cup D_0$ and B_1 on π, respectively.

Case $a_0 \in D_0$, $a_0 \in A_0$: Then we have a path $\pi : a_0 \sim \ldots \sim b_1$ from D_0 to B_1 such that $\mathrm{coll}_\pi \subseteq \mathrm{an}^{\tilde{\mathcal{G}}}_{C_0 \cup (C_1 \setminus B_1)} \subseteq \mathrm{an}^{\tilde{\mathcal{G}}}_{A_0 \cup C_0 \cup (A_1 \cup C_1 \setminus B_1)}$ (meaning all colliders are still open) and we therefore only have to care about the blockable noncolliders. Note that $\mathrm{ncoll}_{b,\pi} \cap (C_0 \cup (C_1 \setminus B_1)) = \emptyset$. Furthermore we can assume $\mathrm{ncoll}_{b,\pi} \cap (A_0 \cup A_1 \setminus B : 1) = \emptyset$. This is because by assumption $\pi \cap A_0 = \emptyset$ and if there exists an $\tilde{a}_1 \in A_1$ which is a blockable non-collider on π, meaning there exists a $k \in \mathbb{N}$ such that $\overset{e_k}{\leftarrow\!\circ}\ \tilde{a}_1\ \overset{e_{k+1}}{\circ\!\rightarrow}$ (with possibly other orientations). If $e_{k+1} = \circ\!\rightarrow$ then we can immediately construct an open path from A_0 to B_1 and if e_k and e_{k+1} are of the form $\leftarrow$, there has to exists an $r < k$ and a collider at v^k which is by definition open as well as all non-colliders on the subpath $v_1^r \leftarrow v_1^{r+1} \leftarrow \ldots \leftarrow \tilde{a}_1$. But then we can simply construct the open path $a_0 \rightarrow \ldots \rightarrow v_0^{r+1} \rightarrow v_1^r \leftarrow v_1^{r+1} \leftarrow \ldots \leftarrow \tilde{a}_1 \sim \ldots \sim b_1$ contradicting $(A, B, C) \in \mathcal{I}_E^{\mathcal{G}}$. Thus we can assume also $\mathrm{ncoll}_{b,\pi} \cap (A_0 \cup A_1 \setminus B : 1) = \emptyset$.

But since now all blockable noncolliders are also open, making π σ-$(A_0 \cup C_0 \cup ((A_1 \cup C_1) \setminus B_1)$-open, we obtain a contradiction to $(D, B, A \cup C) \in \mathcal{I}_E^{\mathcal{G}}$.

Case $a_0 \in A_0$: Then we have a path $\pi : a_0 \sim \ldots \sim b_1$ from D_0 to B_1 such that $\mathrm{coll}_\pi \subseteq \mathrm{an}^{\tilde{\mathcal{G}}}_{C_0 \cup (C_1 \setminus B_1)}$ and also $\mathrm{ncoll}_{b,\pi} \cap (C_0 \cup (C_1 \setminus B_1)) = \emptyset$. We therefore obtain a contradiction to $(A, B, C) \in \mathcal{I}_E^{\mathcal{G}}$.

(LCo): This is obvious as E-separation is defined in terms of σ-separation, which is defined node-wise.

(RCo): Holds with the same justification as (LCo).

■

To prove Proposition 3.5, we first recall the following two concepts, taken from Lauritzen (2019).

Definition A.1 (Conditional independence) *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$ and $\mathcal{H}$ a sub-σ-algebra of $\mathcal{F}$. The two sets $\mathcal{A}$ and $\mathcal{B}$ are conditional independent given σ-algebra $\mathcal{H}$, $\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{H}$:⇔*

$$\begin{aligned}\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{H} \ &:\Leftrightarrow A \perp\!\!\!\perp B \mid \mathcal{H} \quad \forall A \in \mathcal{A}, B \in \mathcal{B} \\ &\Leftrightarrow P(A \cap B \mid \mathcal{H}) \overset{a.s.}{=} P(A \mid \mathcal{H}) P(B \mid \mathcal{H})\end{aligned}$$

where $P(A \mid \mathcal{H}) = \mathbb{E}[\chi_A \mid \mathcal{H}]$.

An equivalent definition of the above conditional independence is given by the following theorem (Lauritzen (2019, Theorem 2.10)):

Theorem A.2 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space,* $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$ *and* $\mathcal{H}$ *a sub-*σ*-algebra of* $\mathcal{F}$*. Then:*

$$\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{H} \;\Leftrightarrow\; P(B \mid \mathcal{A} \vee \mathcal{H}) = P(B \mid \mathcal{H}) \quad B \in \mathcal{B}$$

Proof [Proof of Proposition 3.5] The stated properties are all direct consequences of conditional independence (of σ-algebras). We use the notation:

$$\mathcal{F}_s(X^A) := \sigma\left(\{X^a_{t'} \;:\; t' \le s, a \in A\}\right)$$
$$\mathcal{F}_{s,h}(X^A) := \sigma\left(\{X^a_{t'} \;:\; s \le t' \le s+h, a \in A\}\right)$$

and $P(A \mid \mathcal{H}) := \mathbb{E}[\chi_A \mid \mathcal{H}]$ where $\mathcal{H}$ a sub-σ-algebra.

(LR): We have to show that $(A, B, A) \in \mathcal{I}_{s,h}$.
This holds as

$$P(\tilde{B} \mid \underbrace{\mathcal{F}^A_s \vee \mathcal{F}^A_{s+h}}_{=\mathcal{F}^A_{s+h}}) = P(\tilde{B} \mid \mathcal{F}^A_{s+h}) \qquad \forall \tilde{B} \in \mathcal{F}^B_{s,h}$$

(LD): We have to show that $(A, B, C) \in \mathcal{I}_{s,h}$, $D \subseteq A$ implies $(D, B, C) \in \mathcal{I}_{s,h}$
By assumption

$$P(\tilde{A} \cap \tilde{B} \mid \mathcal{F}^C_{s+h}) \overset{a.s.}{=} P(\tilde{A} \mid \mathcal{F}^C_{s+h}) P(\tilde{B} \mid \mathcal{F}^C_{s+h}) \quad \tilde{A} \in \mathcal{F}^A_s, \tilde{B} \in \mathcal{F}^B_s$$

and since $\mathcal{F}^A_s \subseteq \mathcal{F}^D_s$ the statement follows.

■

A.2. Basic Properties of the Lifted Dependency Graph

Proof [Proof of Lemma 3.3] Let $\pi : a_0 \overset{e_0}{\sim} v^1_{i_1} \overset{e_0}{\sim} \ldots \overset{e_n}{\sim} v^{n+1}_{i_{n+1}} = b_1$ be $C_0 \sqcup (C_1 \setminus \{b_1\})$-$\sigma$-open walk. and assume that a node appears multiple, thus we take the smallest $k \in [n]$ such that

$$\ell_k = \max\{\{\ell \in [\![k+1, n+1]\!] : v^k_{i_k} = v^\ell_{i_\ell}\} > k$$

If $v^k_{i_k} = b_1$ we can just use the subwalk $\tilde{\pi} : a_0 \overset{e_0}{v^1_{i_1}} \overset{e_1}{\sim} \ldots \overset{e_{k-1}}{\sim} v^k_{i_k} = b_1$. We can therefore assume $k < \ell_k < n+1$ and we consider the walk

$$\tilde{\pi} : a_0 \overset{e_0}{\sim} v^1_{i_1} \overset{e_1}{\sim} \ldots \overset{e_{k-1}}{\sim} v^k_{i_k} \overset{e_{\ell+1}}{\sim} v^{\ell_k+1}_{i_{\ell_l+1}} \sim \ldots \sim b_1$$

and we have to show now that it is still $C_0 \sqcup C_1 \setminus \{b_1\}$-$\sigma$-open.

Case $\overset{e_{k-1}}{\circ\!\!\rightarrow} v^k_{i_k} \overset{e_{\ell_k}}{\leftarrow\!\!\circ}$: if $\overset{e_k}{\leftarrow\!\!\circ}$ or $\overset{e_{\ell_k}}{\circ\!\!\rightarrow}$ then we immediately have that $v^k_{i_k} \in \mathrm{an}^{\mathcal{G}}_{C_0 \sqcup C_1 \setminus \{b_1\}}$ hence the collider and thus $\tilde{\pi}$ are open.

If we instead have $\overset{e_k}{\rightarrow}$ or $\overset{e_{\ell-1}}{\leftarrow}$ then there exists

$$\hat{k} = \max\{\hat{k}' \geq k : \overset{e_{\hat{k}'-1}}{\rightarrow} v^{\hat{k}'}_{i_{\hat{k}'}} \leftarrow\!\circ\} \leq \ell_k - 1$$

and again we have $\mathrm{an}^{\tilde{\mathcal{G}}}_{C_0 \sqcup C_1 \setminus \{b_1\}}$ as the collider $v^{\hat{k}}_{i_{\hat{k}}}$ needs to be open.

Case $v^k_{i_k} \in \mathrm{ncoll}^{\tilde{\mathcal{G}}}_{b,\tilde{\pi}}$*: (* $\overset{e_{k-1}}{\leftarrow} v^k_{i_k} \overset{e_{\ell_k}}{\circ\!\!-\!\!\circ}$ *and* $v^{k-1}_{i_{k-1}} \notin \mathrm{scc}^{\tilde{\mathcal{G}}}_{v^k_{i_k}}$ *) or (* $\overset{e_{k-1}}{\circ\!\!-\!\!\circ} v^k_{i_k} \overset{e_{\ell_k}}{\rightarrow}$ *and* $v^{\ell_k}_{i_{\ell_k}} \notin \mathrm{scc}^{\tilde{\mathcal{G}}}_{v^k_{i_k}}$ *):* We consider the first case, the second follows analog. Then $v^k_{i_k}$ is already a blockable non-collider on π with $v^{k-1}_{i_{k-1}} \notin \mathrm{scc}^{\tilde{\mathcal{G}}}_{v^k_{i_k}}$ thus $v^k_{i_k} \notin C_0 \sqcup C_1 \setminus \{b_1\}$. ■

A.3. Proof of the global Markov Property Proposition 3.6

Before giving the proof, we first need to define the concept of *acyclification* (for an overview, see e.g., Forré and Mooij (2023)). For our purposes, we use the following definition from Forré and Mooij (2017):

Definition A.3 (*Acyclification*) *Let $\mathcal{G} = (V, \mathcal{E}_d)$ be a directed graph. A DAG $\mathcal{G}' = (V, \mathcal{E}'_d)$ is called an* acyclification of $\mathcal{G}$:⇔

$$v \to w \in \mathcal{E}'_d \;\Leftrightarrow\; v \notin \mathrm{scc}^{\mathcal{G}}_w \;\wedge\; \exists w' \in \mathrm{scc}^{\mathcal{G}}_w : \; v \to w' \in \mathcal{E}_d \tag{6}$$

This means that we include a directed edge from v to every node w in a different strongly connected component, $\mathrm{scc}^{\mathcal{G}}_w$, if v has an outgoing edge into $\mathrm{scc}^{\mathcal{G}}_w$, and we remove all edges within each strongly connected component.

Remark A.4 *Note that there are other definitions of acyclifications (e.g., in Forré and Mooij (2023)) which also fully connect nodes within each strongly connected component in such a way that one still obtains a DAG. Such a DAG is in general not unique. In this paper, we use the minimal version above and denote the acyclification of a DG $\mathcal{G} = (V, \mathcal{E}_d)$ by $\mathcal{G}^{acy} = (V, \mathcal{E}^{acy}_d)$.*

The reason for using acyclifications is the following relation between σ- and d-separation (Forré and Mooij, 2017, Theorem 2.8.2) (note that we only state the result for directed graphs):

Theorem A.5 *Let $\mathcal{G} = (V, \mathcal{E}_d)$ be a DG, and let $\mathcal{G}' = (V, \mathcal{E}'_d)$ be a DG with $\mathcal{E}'_d \subseteq \mathcal{E}^{acy}_d \cup \mathcal{E}^{scc}_d$ where $\mathcal{E}^{scc}_d := \{v \to w \mid v \in V, w \in \mathrm{scc}^{\mathcal{G}}_v\}$. Then for $A, B, C \subseteq V$:*

$$A \perp\!\!\!\perp^{\mathcal{G}}_{\sigma} B \mid C \;\Rightarrow\; A \perp\!\!\!\perp^{\mathcal{G}'}_{d} B \mid C$$

We now come to the proof of the global Markov property:

Proof [Proof of Proposition 3.6] The proof is a direct application of Forré and Mooij (2018, Theorem B.2), based on the concept of *acyclification* (for each DMG with σ-separation, there exists an ADMG that encodes the same d-separation triple) and the fact that each coordinate process on the intervals $[0, s]$ (and $[s, s+h]$) is determined by the Borel-measurable functions F^j_0 defined as

$$\begin{cases} \mathbb{R}^{\dim(\mathrm{scc}^{\mathcal{G}}_j)} \times C^0\left([0,s], \mathbb{R}^{\dim^W(\mathrm{scc}^{\mathcal{G}}_j)}\right) \times C^0\left([0,s], \mathbb{R}^{\dim(\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}^{\mathcal{G}}_j} \setminus \mathrm{scc}^{\mathcal{G}}_j)}\right) \to C^0\left([0,s], \mathbb{R}^{n_j}\right) \\ (X^{\mathrm{scc}^{\mathcal{G}}_j}_0, W^{\mathrm{scc}^{\mathcal{G}}_j}_{[0,s]}, X^{\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}^{\mathcal{G}}_j} \setminus \mathrm{scc}^{\mathcal{G}}_j}_{[0,s]}) \mapsto X^j_{[0,s]} \end{cases}$$

respectively F_1^j being defined as

$$\begin{cases} C^0\left([s,s+h],\mathbb{R}^{\dim^W(\mathrm{scc}_j^{\mathcal{G}})}\right)\times C^0\left([0,s],\mathbb{R}^{\dim(\mathrm{scc}_j^{\mathcal{G}})}\right)\times \\ C^0\left([0,s+h],\mathbb{R}^{\dim(\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_j^{\mathcal{G}}}\setminus\mathrm{scc}_j^{\mathcal{G}})}\right)\to C^0\left([s,s+h],\mathbb{R}^{n_j}\right) \\ \left((W_{s+t}^{\mathrm{scc}_j^{\mathcal{G}}}-W_s^{\mathrm{scc}_j^{\mathcal{G}}})_{0\le t\le h},X_{[0,s]}^{\mathrm{scc}_j^{\mathcal{G}}},X_{[0,s+h]}^{\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_j^{\mathcal{G}}}\setminus\mathrm{scc}_j^{\mathcal{G}}}\right)\mapsto X^j_{[s,s+h]}\end{cases}$$

for each $j\in\mathrm{scc}_k^{\mathcal{G}}=\mathrm{scc}_j^{\mathcal{G}}$ where X_0^j and, respectively, $X_0^{\mathrm{scc}_j^{\mathcal{G}}})$ are the initial conditions, which are independent of the Brownian path segments $W_{[0,s]}^{\mathrm{scc}_j^{\mathcal{G}}},W_{[s,s+h]}^{\mathrm{scc}_j^{\mathcal{G}}}$.

Note that the above definitions for the mappings are not defined over the entire space of continuous functions on the interval and it also suffices to define them on a measurable set of paths that includes all solutions to the SDEs over the respective intervals as for instance F_1^j is defined as

$$\begin{aligned} &F_1^j((W_{s+t}^{\mathrm{scc}_j^{\mathcal{G}}}-W_s^{\mathrm{scc}_j^{\mathcal{G}}})_{0\le t\le h},X_{[0,s]}^{\mathrm{scc}_j^{\mathcal{G}}},X_{[0,s+h]}^{\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_j^{\mathcal{G}}}\setminus\mathrm{scc}_j^{\mathcal{G}}})=X^j_{[s,s+h]}=\text{solution of } X_t^j\\ &=\int_s^t\mu^j(X_{[0,s]}^{\mathrm{scc}_j^{\mathcal{G}}}*X_{[s,t']}^{\mathrm{scc}_j^{\mathcal{G}}},X_{[0,s]}^{\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_j^{\mathcal{G}}}\setminus\mathrm{scc}_j^{\mathcal{G}}}*X_{[s,t']}^{\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_j^{\mathcal{G}}}\setminus\mathrm{scc}_j^{\mathcal{G}}})dt'\\ &+\int_s^t\sigma^j(X_{[0,s]}^{\mathrm{scc}_j^{\mathcal{G}}}*X_{[s,t']}^{\mathrm{scc}_j^{\mathcal{G}}},X_{[0,s]}^{\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_j^{\mathcal{G}}}\setminus\mathrm{scc}_j^{\mathcal{G}}}*X_{[s,t']}^{\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_j^{\mathcal{G}}}\setminus\mathrm{scc}_j^{\mathcal{G}}})dW_{t'}^{\dim_j^W}\end{aligned}$$

where $*$ denotes the path concatenation of the solution on the two intervals and we have removed the dependence on the initial condition X_s^j. By Rogers and Williams (2000, Theorem 10.4), F_0^j and F_1^j are well-defined and measurable. We can now adapt the technique used in Forré and Mooij (2018, Theorem B.2) to our setting. The proof works by induction on the number of strongly connected components $|\mathbf{S}(\mathcal{G})|=N$.

(IB): Let $V=\mathrm{scc}_v^{\mathcal{G}}$, $v\in V$. (the case $N=1$)
Case $((v,v)\in\mathcal{E}_d\wedge|\mathrm{scc}_v^{\mathcal{G}}|=1)\vee|\mathrm{scc}_v^{\mathcal{G}}|\ge 2$: For $a,b\in V$ there exists a directed walk $a_0\to v_1^1\to v_1^2\to\ldots\to b_1$ that is only $\sigma-C$-blocked if $a_0\in C$ ('left-redundancy') hence clear the global MP holds. In addition we have the acyclification of $\tilde{\mathcal{G}}$ given by $\tilde{\mathcal{G}}^{acy}=(V_0\sqcup V_1,\tilde{\mathcal{E}}_d^{acy})$ DAG with $\tilde{\mathcal{E}}_d^{acy}=\{(i_0,j_1)\mid\forall i,j\in V\}$ and most importantly, satisfying the following global MP with respect to d-separation:

$$A_0\perp\!\!\!\perp_d^{\tilde{\mathcal{G}}^{acy}}B_1\mid C_0,C_1\backslash B_1\;\Rightarrow\;X_{[0,s]}^A\perp\!\!\!\perp X_{[s,s+h]}^B\mid X_{[0,s]}^C,X_{[s,s+h]}^{C\backslash B}\tag{7}$$

for $A,B,C\subseteq V$ and by Theorem A.5:

$$A_0\perp\!\!\!\perp_\sigma^{\tilde{\mathcal{G}}}B_1\mid C_0,C_1\backslash B_1\Rightarrow A_0\perp\!\!\!\perp_d^{\tilde{\mathcal{G}}^{acy}}B_1\mid C_0,C_1\backslash B_1\tag{8}$$

Case $\mathcal{G}=(\{v\},\emptyset)$: This case is obvious.

(IH): Assume for a DG $\mathcal{G}$ with $|\mathbf{S}(\mathcal{G})|=N$, we have given an acyclification for $\tilde{\mathcal{G}}$ for which the global Markov property Eq. (7) holds.

(IS): Assume now $|\mathbf{S}(\mathcal{G})| = N+1$ and let $[v] = \mathrm{scc}_v^{\mathcal{G}} \in \mathbf{S}(\mathcal{G})$ a terminal strongly connected component for some $v \in V$. Denoting $\mathcal{G}' := \mathcal{G}\big|_{V\setminus \mathrm{scc}_v^{\mathcal{G}}}$, by (IH) we have the acyclification of the lifted dependency graph $\tilde{\mathcal{G}}'$, $\left(\tilde{\mathcal{G}}'\right)^{acy} = (V_0' \sqcup V_1', (\tilde{\mathcal{E}}_d')^{acy})$ where $V' = V \setminus \mathrm{scc}_v^{\mathcal{G}}$. We assume $|\mathrm{scc}_v^{\mathcal{G}}| \geq 2$, the case $|\mathrm{scc}_v^{\mathcal{G}}| = 1$ with or without $(v,v) \in \mathcal{E}_d$ works with a similar argument.

Then the acyclification $\tilde{\mathcal{G}}^{acy}$ of $\tilde{\mathcal{G}}$ is given by $\tilde{\mathcal{G}}^{acy} = (V_0 \sqcup V_1, \tilde{\mathcal{E}_d}^{acy})$ with edges

$$\begin{aligned}\tilde{\mathcal{E}_d}^{acy} =& (\tilde{\mathcal{E}}_d')^{acy} \cup \{(i_0, j_1) \mid i, j \in \mathrm{scc}_v^{\mathcal{G}}\} \\ & \cup \{(i_0, j_0), (i_0, j_1), (i_1, j_1) \mid j \in \mathrm{scc}_v^{\mathcal{G}}, i \in \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_v^{\mathcal{G}}} \setminus \mathrm{scc}_v^{\mathcal{G}}\}\end{aligned}$$

Given an arbitrary enumeration of the nodes in the strongly connected component, $\mathrm{scc}_v^{\mathcal{G}} = \{v^1, \ldots, v^r\}$, we can now add each node in the order $v_0^1, \ldots, v_0^r, v_1^1, \ldots, v_1^r$ to $\tilde{\mathcal{G}}'$: So, starting from the DAG $\tilde{\mathcal{G}}'$, we obtain a sequence of DAGs:

$$\hat{\mathcal{G}}_1 := \tilde{\mathcal{G}}^{acy}\Big|_{V_0' \sqcup V_1' \cup \{v_0^1\}}, \hat{\mathcal{G}}_2 := \tilde{\mathcal{G}}^{acy}\Big|_{V_0' \sqcup V_1' \cup \{v_0^1, v_0^2\}} \ldots, \hat{\mathcal{G}}_{r+1} := \tilde{\mathcal{G}}^{acy}\Big|_{V_0' \sqcup V_1' \cup \{v_0^1, \ldots v_0^r, v_1^1\}}, \ldots, \hat{\mathcal{G}}_{2r} := \tilde{\mathcal{G}}^{acy}.$$

We can then add nodes inductively. First add the 'past-nodes' v_0^i, $i \in [r]$. Since $X_{[0,s]}^{v^i}$ can be written as a function of $\{X_{[0,s]}^{v'}\}_{v' \in \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_v^{\mathcal{G}}} \setminus \mathrm{scc}_v^{\mathcal{G}}}$ and the independent noises $\{W_{[0,s]}^{v'}\}_{v' \in \mathrm{scc}_v^{\mathcal{G}}}$ (for example by Picard's successive approximation method), we can apply Forré and Mooij (2018, Lemma B.1) to establish

$$A_0 \perp\!\!\!\perp_d^{\hat{\mathcal{G}}_i} B_1 \mid C_0, C_1 \backslash B_1 \ \Rightarrow\ X_{[0,s]}^A \perp\!\!\!\perp X_{[s,s+h]}^B \mid X_{[0,s]}^C, X_{[s,s+h]}^{C\backslash B} \tag{9}$$

for $A, B, C \subseteq V$ with $B \cap \mathrm{scc}_v^{\mathcal{G}} = \emptyset$. Adding the 'future-nodes' v_1^i, $i \in [r]$ in similar fashion gives a global Markov property with respect to to the acyclification $\tilde{\mathcal{G}}^{acy}$,

$$A_0 \perp\!\!\!\perp_d^{\tilde{\mathcal{G}}^{acy}} B_1 \mid C_0, C_1 \backslash B_1 \ \Rightarrow\ X_{[0,s]}^A \perp\!\!\!\perp X_{[s,s+h]}^B \mid X_{[0,s]}^C, X_{[s,s+h]}^{C\backslash B}, \quad A, B, C \subseteq V.$$

Theorem A.5 then establishes Eq. (3) and concludes the proof.

■

A.4. Proof of Proposition 3.17 for the latent Markov Property

Proof [Proof of Proposition 3.17] The proof works in both directions via contraposition. *'⇐':* Let $\pi : a_0 = v_0^0 \overset{\hat{e}_0}{\sim} v_{\ell_1}^1 \overset{\hat{e}_1}{\sim} \ldots \overset{\hat{e}_n}{\sim} v_1^{n+1} = b_1$ be a σ-$C_0 \cup (C_1 \setminus B_1)$-open walk in $\tilde{\mathcal{G}}$, hence $\mathrm{coll}_\pi \subseteq \mathrm{an}^{\tilde{\mathcal{G}}}_{C_0 \cup (C_1 \setminus B_1)}$ and $\mathrm{ncoll}_{b,\pi} \cap (C_0 \cup (C_1 \setminus B_1)) = \emptyset$. Then for each collider $v_{\ell_k}^k \in \mathrm{coll}_\pi$ there exists a $c^k \in (C_0 \cup (C_1 \setminus B_1))$ such that $v_{\ell_k}^k \in \mathrm{an}^{\tilde{\mathcal{G}}}_{c^l}$, therefore there exists a directed path $v_{\ell_k}^k \to \ldots \to c^k$ of minimal length, implying that there is no other element of $(C_0 \cup (C_1 \setminus B_1))$ on it. By inserting the subwalk $v_{\ell_k}^k \to \ldots \to c^k \leftarrow \ldots \leftarrow v_{\ell_k}^k$ into π for each collider $v_{\ell_k}^k$, we obtain a walk $\bar{\pi} : a_0 = u_0^0 \overset{e_0}{\sim} u_{l_1}^1 \overset{e_1}{\sim} \ldots \overset{e_m}{\sim} u_1^{m+1} = b_1$ such that $\mathrm{coll}_{\bar{\pi}} \subseteq (C_0 \cup (C_1 \setminus B_1))$ and $\mathrm{ncoll}_{b,\pi} \cap (C_0 \cup (C_1 \setminus B_1)) = \emptyset$. Then, for each $u_{l_i}^i$ which is also in $V_0^{lat} \sqcup V_1^{lat}$, there exists $k_i, \hat{k}_i \in \mathbb{N}$

and a subwalk segment $u^{i-k_i}_{l_{i-k_i}} \sim \ldots \sim u^{i+\hat{k}_i}_{l_{i+\hat{k}_i}}$ of $\bar{\pi}$ without colliders and only the endpoints are in $V_0^{obs} \sqcup V_1^{obs}$. As there are no colliders on these segments, there exists edges $u^{i-k_i}_{l_{i-k_i}} \circ\!\!-\!\!\circ\, u^{i+\hat{k}_i}_{l_{i+\hat{k}_i}}$ in $\tilde{\mathcal{G}}'$ with the same endpoint marks. Hence by replacing the segments in question by those edges, we obtain a σ-$C_0 \cup (C_1 \setminus B_1)$-open walk π' in $\tilde{\mathcal{G}}'$. *'⇒':* $\pi' : a_0 = v_0^0 \overset{\hat{e}_0}{\sim} v^1_{\ell_1} \overset{\hat{e}_1}{\sim} \ldots \overset{\hat{e}_n}{\sim} v_1^{n+1} = b_1$ be a σ-$C_0 \cup (C_1 \setminus B_1)$-open walk in $\tilde{\mathcal{G}}'$. For each $e_i \notin \tilde{\mathcal{G}}$ there exists an endpoint mark identical subwalk without colliders with nodes in $V_0^{lat} \sqcup V_1^{lat}$. By replacing these edges by those corresponding line-segments and as $C_0 \cup (C_1 \setminus B_1) \subseteq V_0^{obs} \sqcup V_1^{obs}$, we obtain a σ-$C_0 \cup (C_1 \setminus B_1)$-open walk π in $\tilde{\mathcal{G}}$.
■

A.5. Proofs to establish the Markov equivalence class of DGs

Proof [Proof of Proposition 3.10] We have to show that each path (or walk) $\pi : u_0 \sim v^1_{\ell_1} \sim \ldots \sim v^n_{\ell_n} \sim v_1$ from an $u \notin \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}^{\mathcal{G}}_v}$ is σ-$C_0 \cup C_1 \setminus \{v_1\}$-blocked where $\ell_k \in \{0, 1\}$ for each $k \in [n]$ and $C := \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}^{\mathcal{G}}_v}$. We denote by $r := \max\{k \in [n] \ : \ v^k \notin \mathrm{scc}^{\mathcal{G}}_v\}$ the last index before entering the strongly connected component of v. We therefore have to consider the following three cases:
Case 'The walk enters $\mathrm{scc}^{\mathcal{G}}_v$ through its parents:' Then we have $v^r \in \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}^{\mathcal{G}}_v} \setminus \mathrm{scc}^{\mathcal{G}}_j$, $v^{r+1} \in \mathrm{scc}^{\mathcal{G}}_j$ and $v^r_{\ell_r} \in \mathrm{ncoll}^{\tilde{\mathcal{G}}}_{b,\pi}$ a blockable non-collider on the walk π which is conditioned upon. Hence π is σ $(C_0 \cup (C_1 \setminus \{v_1\}))$ blocked.
Case 'The walk enters through the children of $\mathrm{scc}^{\mathcal{G}}_v$ but its past:' This means we have $v^{r+1} \in \mathrm{scc}^{\mathcal{G}}_v$, $\ell_{r+1} = 0$ such that $v_0^{r+1} \to v^r_{\ell_r}$. Then again $v_0^{r+1} \in \mathrm{ncoll}^{\tilde{\mathcal{G}}}_{b,\pi}$ a blockable collider on the walk π which is conditioned upon. Hence π is σ $(C_0 \cup (C_1 \setminus \{v_1\}))$ blocked.
Case 'The walk enters through the children of $\mathrm{scc}^{\mathcal{G}}_v$ but its future:' This means $v^{r+1} \in \mathrm{scc}^{\mathcal{G}}_v$, $v^r \in \mathrm{ch}^{\mathcal{G}}_{\mathrm{scc}^{\mathcal{G}}_j} \setminus \mathrm{scc}^{\mathcal{G}}_j$ and $v_1^{r+1} \to v_1^r$ the taken edge. Denoting $s = \max\{k \in [n] \ : \ \ell_k \neq 1\}$ the largest index on the path before being entirely in the future node components, we can assume $s < r$. (Otherwise the path π would already be blocked since we condition on the past $\mathrm{scc}^{\tilde{\mathcal{G}}}_{v_0}$). In the case that $v^s = \mathrm{de}^{\mathcal{G}}_v$, then there is a collider on the walk because π contains edges $v_0^s \to v_1^{s+1}$ and $v_1^r \leftarrow v_1^{r+1}$ and this collider is not opened as we are in the descendants of the conditioning set; thus π is blocked. If $v^s \notin \mathrm{de}^{\mathcal{G}}_v$, then there $\exists s' > s$ such that $v^{s'} \notin \mathrm{de}^{\mathcal{G}}_v$ but $v^{s'+1} \in \mathrm{de}^{\mathcal{G}}_v$, hence we have the edge $v^{s'} \to v^{s'+1}$ and again, since $v_1^r \leftarrow v_1^{r+1}$, π contains collider that is not open as it is in the descendants of the conditioning set; hence π is blocked. This concludes the first statement.

The other two statement are immediate consequences of it:

(i) Assume $v, w \in V$, $v \neq w$ and $\mathrm{scc}^{\mathcal{G}}_v \neq \mathrm{scc}^{\mathcal{G}}_w$ (meaning they are disjoint as it is an equivalence relation). If $v \notin \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}w} \setminus \mathrm{scc}^{\mathcal{G}}_w$, then the statement follows from the preceding statement. If however $v \in \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}w} \setminus \mathrm{scc}^{\mathcal{G}}_w$ in the parents of the strongly connected component of w, the w can be E-separated from v by $\mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}^{\mathcal{G}}_w}$ as otherwise w and v would be in the same strongly connected component.

(ii) This works with a similar argument then the first since each path $v_0 \sim \ldots \sim v_1$ can be blocked by the parental set (which does not contain v itself).

■

Proof [Proof of Lemma 3.11]

(i) Let $i \in \mathrm{scc}_j^{\mathcal{G}} \wedge i \neq j$. Then obviously $|\mathrm{scc}_j^{\mathcal{G}}| \geq 2$.
Assume the contradiction, meaning $\exists\, A, B, C \subseteq V$: $(A, B, C) \in \mathcal{I}_E^{\mathcal{G}}$ but $(A, B, C) \notin \mathcal{I}_E^{\mathcal{G}'}$. Then $\exists\ \sigma - (C_0 \cup C_1 \setminus B_1)$-open walk $\pi' = a_0 \sim v_{k_1}^1 \sim \ldots \sim v_{k_n}^n \sim b^1$ in $\tilde{\mathcal{G}}'$ for $a \in A$ and $b \in B$. If $(i_0, j_1), (i_0, j_0), (i_1, j_1) \in \pi'$, then we can replace each of these edges by the respective corresponding, endpoint-identical, directed path segments $i_0 \to j_1' \to \ldots \to j_1$ (with $j_1' \to \ldots \to j_1$ in $\mathrm{scc}_{j_1}^{\tilde{\mathcal{G}}}$), $i_0 \to \ldots \to j_0$ (which is inside $\mathrm{scc}_{i_0}^{\tilde{\mathcal{G}}}$), $i_1 \to \ldots \to j_1$ (inside $\mathrm{scc}_{i_1}^{\tilde{\mathcal{G}}}$) and only the first path-segment $i_0 \to j_1' \to \ldots \to j_1$ has a blockable non-collider namely i_0. But $i_0 \notin C_0$ as otherwise π' would already be blocked.

Hence we obtain a walk $\pi : a_0 \sim w_{\ell_1}^1 \sim \ldots \sim w_{\ell_m}^m \sim b_1^1$ in $\tilde{\mathcal{G}}$ by replacing the edges $(i_0, j_1), (i_0, j_0), (i_0, j_1)$ on π' with their above mentioned edge-identical counterparts. As shown above, π has the same set of blockable non-colliders as we only added the above mentioned segments and also the same collider nodes as the added path-segments are directed and direction-preserving, thus giving $\mathrm{coll}_\pi^{\tilde{\mathcal{G}}} = \mathrm{coll}_{\pi'}^{\tilde{\mathcal{G}}'}$. However as π' is open, each collider c_k with $\mathrm{an}_{c_k}^{\tilde{\mathcal{G}}'} \cap (C_0 \cup C_1 \setminus B_1)$ and thus we can also establish the same ancestral relation in $\tilde{\mathcal{G}}$ by replacing edges $(i_0, j_1), (i_0, j_0), (i_1, j_1)$ by directed path-segments $i_0 \to j_1' \to \ldots \to j_1$, $i_0 \to \ldots \to j_0, i_1 \to \ldots \to j_1$. Hence π is $\sigma - (C_0 \cup C_1 \setminus B_1)$-open walk in $\tilde{\mathcal{G}}$ as well ↯

(ii) Let $i \notin \mathrm{scc}_j^{\mathcal{G}}$ and assume there exists a $j' \in \mathrm{scc}_j^{\mathcal{G}}$ such that $(i, j') \in \mathcal{E}_d$. Assume first $\exists$ $A, B, C \subseteq V$ such that $(A, B, C) \in \mathcal{I}_E^{\mathcal{G}}$ but $(A, B, C) \notin \mathcal{I}_E^{\mathcal{G}'}$. Then there exists a shortest σ-$(C_0 \cup C_1 \setminus B_1)$-open walk $\pi' = a_0 \sim v_{k_1}^1 \sim \ldots \sim v_{k_n}^n \sim b^1$ in $\tilde{\mathcal{G}}'$ from $a \in A$ to $b \in B$. If $(i_0, j_1), (i_0, j_0), (i_1, j_1) \in \pi'$ we can again replace those by the edge-point identical, directed path-segments $i_0 \to j_1' \to \ldots \to j_1$ (with $j_1' \to \ldots \to j_1$ in $\mathrm{scc}_{j_1}^{\tilde{\mathcal{G}}}$), $i_0 \to j_0' \to \ldots \to j_0$ (with $j_0' \to \ldots \to j_0$ in $\mathrm{scc}_{j_0}^{\tilde{\mathcal{G}}}$), $i_1 \to j_1' \ldots \to j_1$ (with $j_1' \ldots \to j_1$ in $\mathrm{scc}_{j_1}^{\tilde{\mathcal{G}}}$) with the only blockable non-collider being i_0 and i_1, respectively. But $i_0, i_1 \notin C_0 \cup (C_1 \setminus B_1)$ (as otherwise π' already not open in $\tilde{\mathcal{G}}'$) and also $i_1 \notin B_1$ as one would otherwise have a shorter $\sigma - (C_0 \cup C_1 \setminus B_1)$-open walk from A_0 to B_1. Hence we obtain a walk $\pi : a_0 \sim w_{\ell_1}^1 \sim \ldots \sim w_{\ell_m}^m \sim b_1^1$ with the same set of blockable non-colliders as we only added the above mentioned segments and keep the same collider nodes. Since $\mathrm{coll}_\pi^{\tilde{\mathcal{G}}} = \mathrm{coll}_{\pi'}^{\tilde{\mathcal{G}}'}$ as for each collider c_k with $\mathrm{an}{c_k}^{\tilde{\mathcal{G}}'} \cap (C_0 \cup C_1 \setminus B_1)$ we can also establish the same ancestral relation in $\tilde{\mathcal{G}}$ by replacing edges $(i_0, j_1), (i_0, j_0), (i_1, j_1)$ by the above, directed path-segments $i_0 \to j_1' \to \ldots \to j_1$, $i_0 \to \ldots \to j_0, i_1 \to \ldots \to j_1$. Hence π is $\sigma - (C_0 \cup C_1 \setminus B_1)$-open walk in $\tilde{\mathcal{G}}$ ↯

(iii) This follows immediately from Proposition 3.10.

■

Proof [Proof of Theorem 3.12] Let $\mathcal{G} = (V, \mathcal{E}_d)$ and $\mathcal{G}' = (V, \mathcal{E}_d')$ be two DGs such that $\mathcal{I}_E^{\mathcal{G}} = \mathcal{I}_E^{\mathcal{G}'}$. Then first of all, the set of strongly connected components in both graphs have to be the same. If otherwise there are $v \neq w \in V$ and $w \in \mathrm{scc}_v^{\mathcal{G}}$ but $w \notin \mathrm{scc}_v^{\mathcal{G}'}$, then the two nodes are inseparable in $\mathcal{G}$ but not in $\mathcal{G}'$ (either $(\{v\}, \{w\}, \mathrm{pa}_{\mathrm{scc}_w^{\mathcal{G}'}}^{\mathcal{G}'}) \in \mathcal{I}_E^{\mathcal{G}'}$ or $(\{w\}, \{v\}, \mathrm{pa}_{\mathrm{scc}_v^{\mathcal{G}'}}^{\mathcal{G}'}) \in \mathcal{I}_E^{\mathcal{G}'}$). We can therefore use the notion $\mathrm{scc}_v^{\mathcal{G}}$ and $\mathrm{scc}_v^{\mathcal{G}'}$ interchangeably. Moreover, with the same argumentation, for each singleton $\{v\} \in \boldsymbol{S}_{\mathcal{G}} = \boldsymbol{S}_{\mathcal{G}'}$ the statement '$(v, v) \in \mathcal{E}_d \Leftrightarrow (v, v) \in \mathcal{E}_d'$' has to hold. In addition, the DAGs of strongly connected components for both graphs have to coincide. Otherwise, if this would not be the case, there are $v, w \in V$: $w \notin \mathrm{scc}_v^{\mathcal{G}}$ and there exists $v' \in \mathrm{scc}_v^{\mathcal{G}}$, $w' \in \mathrm{scc}_w^{\mathcal{G}}$

such that $(v', w') \in \mathcal{E}_d$ but for no pair of nodes $v'' \in \mathrm{scc}_v^{\mathcal{G}}$, $w'' \in \mathrm{scc}_w^{\mathcal{G}}$ we have $(v'', w'') \in \mathcal{E}_d'$. But the this means that $(\{v\}, \{w\}, \mathrm{pa}^{\mathcal{G}'}_{\mathrm{scc}_w^{\mathcal{G}'}}) \in \mathcal{I}_E^{\mathcal{G}'}$ but not in $\mathcal{I}_E^{\mathcal{G}}$ as $\mathrm{pa}^{\mathcal{G}'}_{\mathrm{scc}_w^{\mathcal{G}'}}$ does not intersect $\mathrm{scc}_v^{\mathcal{G}}$. Finally, if for $v \neq w$ there is the edge $[v] \to [w]$ in $\boldsymbol{S}_{\mathcal{G}}$ (hence also in $\boldsymbol{S}_{\mathcal{G}'}$), then the nodes, from which edges are outgoing from $\mathrm{scc}_v^{\mathcal{G}}$ to $\mathrm{scc}_w^{\mathcal{G}}$ have to coincide. If this would not be the case, then without loss of generality there exists a $v' \in (\mathrm{pa}^{\mathcal{G}'}_{\mathrm{scc}_w^{\mathcal{G}}} \cap \mathrm{scc}_v^{\mathcal{G}})$ such that $v' \notin (\mathrm{pa}^{\mathcal{G}'}_{\mathrm{scc}_w^{\mathcal{G}}} \cap \mathrm{scc}_v^{\mathcal{G}})$. But then $(\{v'\}, \{w\}, \mathrm{pa}^{\mathcal{G}}_{\mathrm{scc}_w^{\mathcal{G}}}) \in \mathcal{I}_E^{\mathcal{G}}$ but not in $\mathcal{I}_E^{\mathcal{G}'}$. This proves the above statement. Together with Lemma 3.11, the existence of a greatest element can be concluded. The remaining statements are trivial. ∎

A.6. Proofs about Inducing Paths and the Markov equivalence class of DMGs

Proof [Proof of Proposition 3.19] In the case that $v \in \mathrm{scc}_w^{\mathcal{G}}$ or $v^1 \mathrm{scc}_w^{\mathcal{G}}$ (meaning there exists $\bar{\nu} : v_0 \to w_1$ respectively $\bar{\nu} : v_0 \circ\!\!\to v_1^1 \to \ldots \to w^1$) the statement is obvious. We can also assume that there have to be at least three scc's involved.

We furthermore can assume that ν has minimal length. Let now $C \subseteq V \setminus \{v\}$, we want to show that $(\{v\}, \{w\}, C) \notin \mathcal{I}_E$ meaning there exists a σ-$C_0 \cup (C_1 \setminus \{w_1\})$-open path.

Therefore consider the following set:

$$S_C = \big\{\pi : v_0 \sim \ldots \sim w_1 \text{ walk in } \tilde{\mathcal{G}} : \mathrm{coll}_\pi \subseteq \mathrm{an}^{\tilde{\mathcal{G}}}_{\{v_0,w_1\}\cup C_0 \cup (C_1 \setminus \{w_1\})} \wedge \\ \mathrm{ncoll}_{b,\pi} \cap (C_0 \cup (C_1 \setminus \{w_1\})\big\}$$

Note the relation $k \in \mathrm{an}_v^{\mathcal{G}}$ implies $k_0 \in \mathrm{an}_{v_0}^{\tilde{\mathcal{G}}}$ and $k_1 \in \mathrm{an}_{v_1}^{\tilde{\mathcal{G}}}$ and also $k_0 \in \mathrm{an}_{v_1}^{\tilde{\mathcal{G}}}$.
The set is not empty as from the inducing path $\nu : v = v^0 \overset{e_0}{\sim} \ldots \overset{e_n}{\sim} w$, with $\hat{k} = \max\{k \in [n] \,:\,$ we can construct the following path $\hat{\nu} : v_0 \overset{\hat{e_0}}{\sim} v^1_{\ell_1} \overset{\hat{e_1}}{\sim} \ldots v^{\hat{k}}_{\ell\hat{k}} \overset{e\hat{k}}{\circ\!\!\to} v_1^{\hat{k}+1} \sim \ldots \sim w^1$ with the following properties:

- for each collider $v^k_{\ell_k} \in \mathrm{coll}_{\hat{\nu}}$ in holds $v^k_{\ell_k} \in \mathrm{an}^{\tilde{\mathcal{G}}}_{\{v_0,w_1\}}$ since by assumption and the remark above, $v^k \in \mathrm{an}^{\mathcal{G}}_{\{v,w\}} \setminus \{v, w\}$.
- each non-collider $v^k_{\ell_k} \in \mathrm{ncoll}_{\hat{\nu}}$ is unblockable

Consider now the path $\hat{\omega} \in S_C$ with a minimal number of colliders, and we denote him $v_0 = u_0^0 \sim \ldots \sim u^k_{\ell_k} \sim \ldots \sim u^{m+1}_{\ell_{m+1}} = w^1$.

We want to show that $\mathrm{coll}_{\hat{\omega}} \subseteq \mathrm{an}^{\tilde{\mathcal{G}}}_{C_0 \cup C_1 \setminus \{w_1\}}$, meaning that in this case we have found a σ-$C_0 \cup (C_1 \setminus \{w_1\})$-open path.

By definition, all colliders are in $\mathrm{an}^{\tilde{\mathcal{G}}}_{\{v_0,w_1\}\cup C_0 \cup (C_1 \setminus \{w_1\})}$, therefore have to following case-distinction: If there would exist a collider $u_0^k \in \mathrm{an}^{\tilde{\mathcal{G}}}_{v_0} \setminus \mathrm{an}^{\tilde{\mathcal{G}}}_{C_0 \cup C_1 \setminus \{w_1\}}$ (we assume k to be maximal with this property) then this means there exists a directed path $v_0^k \to \ldots \to v_0$ that does not intersect C_0 and we can construct a path $v_0 \leftarrow \ldots \leftarrow \underbrace{u_0^k \sim \ldots \sim w^1}_{\text{rest of } \hat{\omega}}$. ∎

A.7. On the E-Separation Independence Model of DMGs

Proof [Proof of Proposition 3.22] Let $i \notin \mathrm{scc}_j^{\mathcal{G}}$, assume there exists an $\exists i' \in \mathrm{scc}_i^{\mathcal{G}}$ and a $j' \in \mathrm{scc}_j^{\mathcal{G}}$ such that $i' \leftrightarrow j' \in \mathcal{E}_b i$ but $i \leftrightarrow j \notin \mathcal{E}_{bi}$. Assume first $\exists\, A, B, C \subseteq V$ such that $(A, B, C) \in \mathcal{I}_E^{\mathcal{G}}$ but $(A, B, C) \notin \mathcal{I}_E^{\mathcal{G}'}$.

Then there exists a shortest σ-$(C_0 \cup C_1 \setminus B_1)$-open walk $\pi' = a_0 \sim v_{k_1}^1 \sim \ldots \sim v_{k_n}^n \sim b^1$ in $\tilde{\mathcal{G}}'$ from $a \in A$ to $b \in B$ and since bidirected edges do not change ancestral relations, one of the edges $i_0 \leftrightarrow j_1$, $i_0 \leftrightarrow j_0$, $i_1 \leftrightarrow j_1$ or $i_1 \leftrightarrow j_0$ is present on this path. These however can be replaced by their end-point and direction-preserving counterparts:

$$\begin{aligned}
&i_0 \leftarrow \ldots \leftarrow i_0' \leftrightarrow j_1' \rightarrow \ldots \rightarrow j_1\\
&i_0 \leftarrow \ldots \leftarrow i_0' \leftrightarrow j_0' \rightarrow \ldots \rightarrow j_0\\
&i_1 \leftarrow \ldots \leftarrow i_1' \leftrightarrow j_1' \rightarrow \ldots \rightarrow j_1\\
&i_1 \leftarrow \ldots \leftarrow i_1' \leftrightarrow j_0' \rightarrow \ldots \rightarrow j_0
\end{aligned}$$

and all nodes within are unblockable non-colliders such that we can construct a σ-$(C_0 \cup C_1 \setminus B_1)$-open walk $\pi : a_0 \sim \ldots \sim b_1$, which is a contradiction to the above assumption $\lightning$ ■

Appendix B. Causal Discovery in the fully observed SDE Model

In this section, we empirically show, that when applying the causal discovery algorithm introduced in Manten et al. (2024), to SDEs with cyclic adjacencies, we can reliably estimate the Markov equivalence class directly from data, demonstrating the real world applicability as well as practically verifying the theoretical findings in Section 3.4. Since the test and algorithm were developed and thoroughly analyzed in another paper, we omit large-scale experiments and comparisons with other methods here, focusing instead on a proof-of-concept demonstration.

B.1. The Algorithm

The algorithm is only slightly altered in that we now use the right-decomposable notion of E-separation:

Algorithm 1 Causal discovery for SDEs.

```
Input: DG 𝒢 = (V ≅ [d], ℰ_d), CIT ⫫⁺_{s,h}
Ṽ ← {k_0, k_1 | k ∈ V} ℰ̃_d ← {i_0 → j_0, i_1 → j_1, i_0 → j_1 | (i, j) ∈ ℰ_d}
for c ∈ {0, ..., d − 1} do
    for (i, j) ∈ V do
        for K ⊆ V \ {i}, |K| = c, such that (k_0 → j_1) ∈ ℰ̃_d for k ∈ K do
            if X^i ⫫⁺_{s,h} X^j | X^K then
                ℰ̃_d ← ℰ̃_d \ {i_0 → j_0, i_1 → j_1, i_0 → j_1}
𝒢 = (V, ℰ_d) ← collapse(Ṽ, ℰ̃_d)
return 𝒢
```

B.2. Recovering the Markov Equivalence Class from Data

To demonstrate the practical efficacy of the Algorithm 1 and verify the theoretical results presented in Section 3.4, we apply it by drawing 50 sets of parameters for each of the two 3-dimensional linear SDEs:

$$\mathrm{d}\begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} = \left(\begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \right) \mathrm{d}t + \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \mathrm{d} \begin{pmatrix} W_t^1 \\ W_t^2 \\ W_t^3 \end{pmatrix}. \tag{10}$$

and

$$\mathrm{d}\begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} = \left(\begin{pmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \right) \mathrm{d}t + \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \mathrm{d} \begin{pmatrix} W_t^1 \\ W_t^2 \\ W_t^3 \end{pmatrix}. \tag{11}$$

with different adjacency-structures and parameters drawn as $a_{ij} \sim \mathcal{U}((-1.5, 1] \cup [1, 1.5))$ for $i \neq j$, $a_{ii} \sim \mathcal{U}([-0.5, 0.5])$, $c_i \sim \mathcal{U}([0, 0.1))$, $d_i \sim \mathcal{U}([0.3, 0.5))$. From each SDE, we draw 400 sample-paths and try to discovery the underlying causal relationships using algorithm Algorithm 1. The result, shown in Fig. 4, display the predicted probabilities next to each edge, confirming the theoretical expectations: $\mathcal{G}_1$ closely corresponds to the equivalence class for the adjacency structure (considered as a DG) of Eq. (10), while $\mathcal{G}_2$ corresponds to that in Eq. (11).

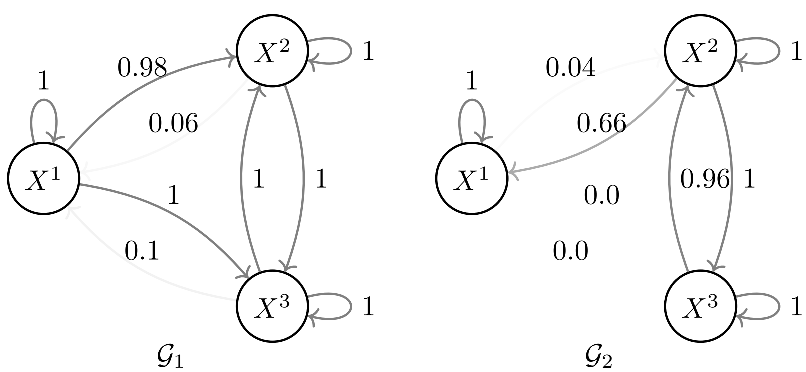

Figure 4: Results from applying the algorithm to two 3-dimensional linear SDEs, showing predicted probabilities for each edge and confirming the expected equivalence classes.

Appendix C. Counterexample for greatest element for σ-Separation

As can be seen in Fig. 5, the graphs $\mathcal{G}_1$-$\mathcal{G}_5$ are all in the same equivalence class with respect to their induced graphoid $\mathcal{I}_\sigma$, in which adjacent nodes X^1 and X^2 and X^2 and X^3 cannot be separated from each other except by conditioning on the respective endpoints and X^1 and X^3 can be separated from each other by conditioning on X^2. However their supremum with respect to the partial subset ordering on $\mathcal{E}_d$, $\mathcal{G}_6$ does not allow for a separation between X^1 and X^3 at all.

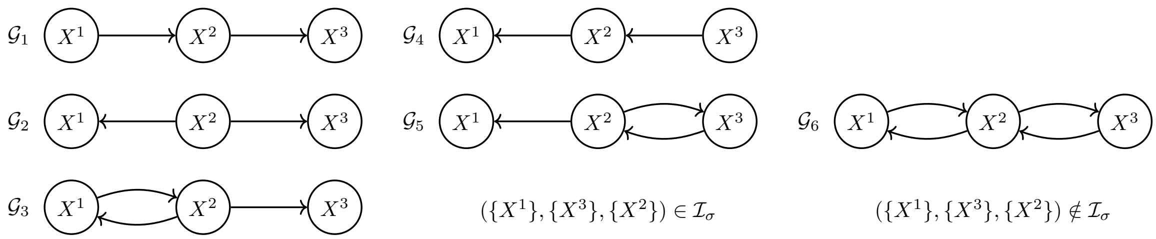

Figure 5: The graphs $\mathcal{G}_1$-$\mathcal{G}_5$ all have the same independence model, however in their supremum with respect to partial subset ordering, $\mathcal{G}_6$, X^1 and X^3 cannot be separated by X^2.