Bernoulli **28**(4), 2022, 3023–3050
https://doi.org/10.3150/21-BEJ1446

Graphical modeling of stochastic processes driven by correlated noise

SØREN WENGEL MOGENSEN[1,2,3,a] and NIELS RICHARD HANSEN[1,b]

[1]*Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark.*
[a]*soren.wengel_mogensen@control.lth.se,* [b]*niels.r.hansen@math.ku.dk*
[2]*Section for Cognitive Systems, Technical University of Denmark, Kgs. Lyngby, Denmark*
[3]*Current affiliation: Department of Automatic Control, Lund University, Lund, Sweden*

We study a class of graphs that represent local independence structures in stochastic processes allowing for correlated noise processes. Several graphs may encode the same local independencies and we characterize such equivalence classes of graphs. In the worst case, the number of conditions in our characterizations grows superpolynomially as a function of the size of the node set in the graph. We show that deciding Markov equivalence of graphs from this class is coNP-complete which suggests that our characterizations cannot be improved upon substantially. We prove a global Markov property in the case of a multivariate Ornstein-Uhlenbeck process which is driven by correlated Brownian motions.

Keywords: Graphical models; stochastic processes; local independence; Markov equivalence; Ornstein–Uhlenbeck processes

1. Introduction

Graphical modeling studies how to relate graphs to properties of probability distributions [31]. There is a rich literature on graphical modeling of distributions of multivariate random variables [34], in particular on graphs as representations of conditional independencies. In stochastic processes, local independence can be used as a concept analogous to conditional independence and several papers use graphs to encode local independencies [3,14,15,37,39,50]. Didelez [13, 15] studies graphical modeling of local independence of multivariate point processes. Mogensen et al. [39] also consider diffusions. This previous work only models direct influence between coordinate processes in a multivariate stochastic process. We consider the more general case in which the noise processes driving the continuous-time stochastic process may be correlated. Eichler [17, 19, 20],Eichler and Didelez [21] study this in the time series case (i.e., stochastic processes indexed by discrete time).

A specific local independence structure can be represented by several different graphs, and the characterization of such Markov equivalence classes is an important question in graphical modeling. We study these equivalence classes and characterize them. Our characterizations are computationally demanding as they may involve exponentially many conditions (as a function of the number of nodes in the graphs). We prove that deciding Markov equivalence in this class of graphs is coNP-hard, and therefore one would not expect to find a characterization which allows the problem of Markov equivalence to be decided in polynomial time.

Markov properties are central in graphical modeling as they allow us to deduce independence from graphs. The graphical results in this paper apply to various classes of stochastic processes for which it is possible to show a so-called global Markov property. As an example, we study systems of linear stochastic differential equations (SDEs), and in particular Ornstein-Uhlenbeck processes. Such models have been used in numerous fields such as psychology [25], neuroscience [16,43,53], finance [6,51,58], biology [5], and survival analysis [2,32]. In this paper, we show that multivariate Ornstein-Uhlenbeck pro-

cesses with correlated driving Brownian motions satisfy a global Markov property with respect to certain graphs. Previous work in continuous-time models considers independent noise processes only and the present work extends this framework to cases where the driving processes are correlated. We emphasize that the global Markov property proven in this paper only applies to Ornstein-Uhlenbeck processes. In the case of uncorrelated noise, analogous results apply to quite general classes of continuous-time stochastic processes [15,39] and for this reason we expect that more general continuous-time versions of the global Markov property are possible, also in the presence of correlated noise. To our knowledge, the global Markov property in this paper is the first such result allowing correlated noise in continuous-time models. It is analogous to results in time series models with correlated noise processes [17,19–21]. The graphical and algorithmic results we present also apply to these time series models. They also apply to more general continuous-time processes if a similar global Markov property can be shown.

Many results and ideas in this paper are reminiscent of classical probabilistic graphical models such as Bayesian networks and chain graph models [34]. In these models, graphs are used as representations of conditional independencies of a multivariate random vector. Conditional independence is a *symmetric* ternary relation in the sense that if X is conditionally independent of Y given Z then Y is also conditionally independent of X given Z. When considering graphical modeling of stochastic processes, we can part ways with the symmetry and obtain a more fine-grained representation of independence and for this reason we study representations of local independence. The asymmetry of local independence also leads to different notions of graphical separation and therefore a different graph-theoretic framework is needed than in the symmetric case. This is also evident in earlier work on graphical models of local independence [13,15,18,37].

1.1. Overview and organization

The paper consists of two parts. In the first part (Section 2), we describe local independence for Itô processes. The basic definitions relating to local independence and local independence graphs apply to a wide range of stochastic processes and for this reason we start from this general class. In Subsection 2.2, we state the global Markov property which relates the graphical representations used in this paper to local independence statements. In Subsection 2.3, we consider the smaller class of Itô diffusions. First, we show that in a certain subclass of Itô diffusions one can obtain a very simple relation between the local independence structure of a stochastic process and the conditional independence structure of its equilibrium distribution. However, in general local independence structure cannot be deduced from conditional independence in the equilibrium distribution and we show this with an example using the class of Ornstein-Uhlenbeck processes. We finish this section by proving that the global Markov property holds in Ornstein-Uhlenbeck processes driven by correlated noise.

The second part of the paper (Sections 3, 4, 5) provides results on *directed correlation graphs* (cDGs) – the class of graphs that we use to represent local independencies in a stochastic process with correlated noise. Section 3 defines cDGs and introduces the fundamental graphical definitions we need. Section 4 gives a characterization of the cDGs that encode the same independencies. This directly leads to an algorithm for checking equivalence of cDGs. This algorithm runs in exponential time (in the number of nodes in the graphs). In Section 5 we state another characterization of Markov equivalence and we prove that deciding Markov equivalence of cDGs is coNP-complete.

The two parts of the paper are connected by the global Markov property which allows us to infer local independencies from properties of a cDG. Using the global Markov property, e.g., in the class of Ornstein-Uhlenbeck processes with correlated noise, the results of the paper allow us to reason about the local independence structure of a stochastic process on the basis of a graph which represents both

direct influence between coordinate processes and the correlation structure of the noise processes. The graphical results of the second part describe for which processes the global Markov property implies the same local independence structure. This enables structure learning based on local independence, i.e., recovering the graphical structure from observation of the process.

Proofs that were omitted from the main text can be found in the supplementary material [38].

2. Local independence

Before diving into a formal introduction, we will consider a motivating example.

Example 1. Consider the three-dimensional Ornstein-Uhlenbeck process which solves the following stochastic differential equation

$$\mathrm{d}\begin{pmatrix} X_t^\alpha \\ X_t^\beta \\ X_t^\gamma \end{pmatrix} = \underbrace{\begin{pmatrix} M_{\alpha\alpha} & 0 & 0 \\ M_{\beta\alpha} & M_{\beta\beta} & 0 \\ 0 & 0 & M_{\gamma\gamma} \end{pmatrix}}_{=M} \begin{pmatrix} X_t^\alpha \\ X_t^\beta \\ X_t^\gamma \end{pmatrix} \mathrm{d}t + \underbrace{\begin{pmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_\beta & 0 & \rho_\beta \\ 0 & 0 & \sigma_\gamma & \rho_\gamma \end{pmatrix}}_{=\sigma_0} \mathrm{d}\begin{pmatrix} W_t^1 \\ W_t^2 \\ W_t^3 \\ W_t^4 \end{pmatrix}$$

where $(W_t^1, W_t^2, W_t^3, W_t^4)^T$ is a standard four-dimensional Brownian motion. In this example, all entries in the matrices M and σ_0 above that are not explicitly 0 are assumed nonzero.

The interpretation of the stochastic differential equation via the Euler-Maruyama scheme yields the update equation

$$\begin{aligned} \tilde{X}_{t+\Delta}^\alpha &= \tilde{X}_t^\alpha + \Delta M_{\alpha\alpha}\tilde{X}_t^\alpha + \sqrt{\Delta}\sigma_\alpha \varepsilon_t^1 \\ \tilde{X}_{t+\Delta}^\beta &= \tilde{X}_t^\beta + \Delta(M_{\beta\alpha}\tilde{X}_t^\alpha + M_{\beta\beta}\tilde{X}_t^\beta) + \sqrt{\Delta}\left(\sigma_\beta \varepsilon_t^2 + \rho_\beta \varepsilon_t^4\right) \\ \tilde{X}_{t+\Delta}^\gamma &= \tilde{X}_t^\gamma + \Delta M_{\gamma\gamma}\tilde{X}_t^\gamma + \sqrt{\Delta}\left(\sigma_\gamma \varepsilon_t^3 + \rho_\gamma \varepsilon_t^4\right) \end{aligned}$$

where $\varepsilon_t \sim \mathcal{N}(0, I)$. The Euler-Maruyama scheme evaluated in $t = n\Delta$ for $n \in \mathbb{N}_0$ gives a process, $(\tilde{X}_{n\Delta})_{n\geq 0}$, which, as $\Delta \to 0$, converges to the Ornstein-Uhlenbeck process, $(X_t)_{t\geq 0}$, solving the stochastic differential equation. From the update equations we see that the infinitesimal increment of each coordinate depends on the value of that coordinate, and in addition the β-coordinate increment depends on the α-coordinate (because $M_{\beta\alpha} \neq 0$). Moreover, the increments for coordinates β and γ are correlated as they share the noise variable ε_t^4. Figure 1 (left) provides a graphical representation with arrows readily read from the drift matrix, M, and the diffusion matrix, $\sigma_0\sigma_0^T$. The 'unrolled' graph (Figure 1, right) is a *directed acyclic graph* (DAG) which corresponds to the Euler-Maruyama scheme and provides a discrete-time representation of the dynamics.

A central purpose of this paper is to clarify the mathematical interpretation of *local independence graphs* with blunt edges such as the one in Figure 1 (left), and our results include a characterization of all graphs with equivalent mathematical content. As showcased in the example above, we allow a nondiagonal $\sigma_0\sigma_0^T$ which is a novelty in graphical modeling of continuous-time stochastic processes.

2.1. Itô processes and local independence graphs

We will for the purpose of this paper focus on vector-valued, continuous-time stochastic processes with continuous sample paths. Thus let $X = (X_t)_{t\in\mathcal{T}}$ denote such an n-dimensional process with time index

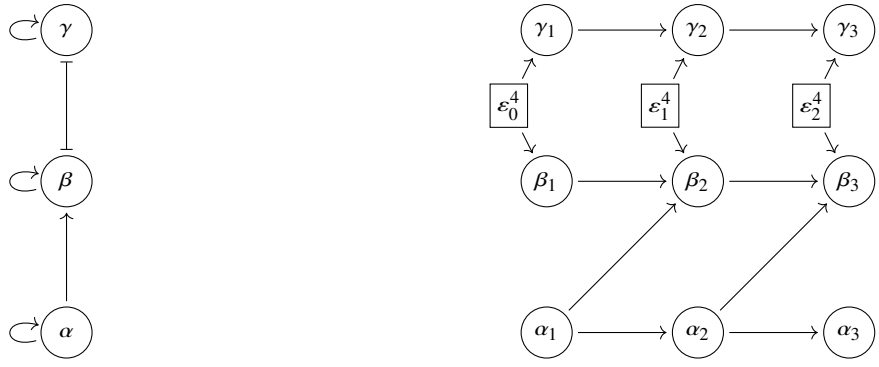

Figure 1. A local independence graph (left) and an 'unrolled' graph (right) where time is made explicit (see Example 1). The two graphs represent the same local independence structure of a stochastic process, X. A node δ for $\delta \in \{\alpha, \beta, \gamma\}$ represents the increments of the X_t^δ-process at time t. On the right, the ε^4-process is a 'white noise' process which creates dependence between X_t^β and X_t^γ. In the 'rolled' version of the graph (left) this is represented by a *blunt* edge, $\beta \mapsto \gamma$. When unrolling a local independence graph to obtain a graphical representation in terms of lagged variables, one could also choose to include $\alpha_s \to \beta_t$ in the unrolled graph for all $s < t$ if $\alpha \to \beta$ in the local independence graph (see also [12,27,55] and [37, supplementary material]).

$t \in \mathcal{T} \subseteq \mathbb{R}$ and with $X_t = (X_t^\alpha)_{\alpha \in [n]} \in \mathbb{R}^n$ being a real-valued vector indexed by $[n] = \{1, \ldots, n\}$. The time index set, $\mathcal{T}$, will in practice be of the forms $[0,T]$, $[0,\infty)$, or $\mathbb{R}$, however, we will in general just assume that $\mathcal{T}$ is an interval containing 0.

We use *local independence* [1,9,15,52] to give a mathematically precise definition of what it means for the historical evolution of one coordinate, $\alpha \in [n]$, to *not* be predictive of the infinitesimal increment of another coordinate, $\beta \in [n]$, given the historical evolution of a set of coordinates, $C \subseteq [n]$. As such, it is a continuous-time version of Granger causality [see, e.g., 24], and its formulation is directly related to filtration problems for stochastic processes. In a statistical context, local independence allows us to express simplifying structural constraints that are directly useful for forecasting and such constraints are also useful for causal structure learning.

The process X is defined on the probability space $(\Omega, \mathcal{F}, P)$ and we let $\sigma(X_s^\delta; s \leq t, \delta \in D) \subseteq \mathcal{F}$ denote the σ-algebra on Ω generated by X_s^δ for all $s \in \mathcal{T}$ up to time t and all $\delta \in D$. For technical reasons, we define $\mathcal{F}_t^D$ to be the P-completion of the σ-algebra

$$\bigcap_{t' > t} \sigma(X_s^\delta; s \leq t', \delta \in D),$$

so that $(\mathcal{F}_t^D)_{t \in \mathcal{T}}$ is a complete, right-continuous filtration for all $D \subseteq [n]$. We will let $\mathcal{F}_t = \mathcal{F}_t^{[n]}$ denote the filtration generated by all coordinates of the process. Within this setup we will restrict our attention to Itô processes with continuous drift and constant diffusion coefficient.

Definition 2 (Regular Itô process). We say that X is a *regular Itô process* if there exists a continuous, $\mathcal{F}_t$-adapted process, λ, with values in $\mathbb{R}^n$ and an $n \times n$ invertible matrix σ such that

$$W_t = \sigma^{-1} \left(X_t - X_0 - \int_0^t \lambda_s \mathrm{d}s \right)$$

is an $\mathcal{F}_t$-adapted standard Brownian motion.

One reason for the interest in the general class of Itô processes is the fact that they are closed under marginalization. That is, the marginal of an Itô process when marginalizing over a set of coordinate processes is again an Itô process which follows from Theorem VI.8.4 in [49]. A regular Itô process is sometimes written in differential form as

$$\mathrm{d}X_t = \lambda_t \mathrm{d}t + \sigma \, \mathrm{d}W_t. \tag{1}$$

Here λ_t is known as the drift of the process and σ as the (constant) diffusion coefficient (an $n \times n$ matrix). We define the *diffusion matrix* for a regular Itô process as the positive definite matrix

$$\Sigma = \sigma\sigma^T. \tag{2}$$

If we consider the more general case where there may be more noise processes than observed processes, we can define the process X_t, as in Example 1, as the solution of the stochastic differential equation

$$\mathrm{d}X_t = \lambda_t \mathrm{d}t + \sigma_0 \, \mathrm{d}W_t \tag{3}$$

for an m-dimensional standard Brownian motion W and with the diffusion coefficient σ_0 an $n \times m$ matrix. If σ_0 has rank n, such a solution is also a regular Itô process with diffusion matrix $\Sigma = \sigma_0\sigma_0^T$. Indeed, we can take $\sigma = (\sigma_0\sigma_0^T)^{1/2}$ in Definition 2. Observe also that for any regular Itô process,

$$X_t - X_0 - \int_0^t \lambda_s \mathrm{d}s = \sigma W_t$$

is an $\mathcal{F}_t$-martingale and $\int_0^t \lambda_s \mathrm{d}s$ is the compensator of X_t in its Doob-Meyer decomposition.

Definition 3. Let X be a regular Itô process with drift λ, and let $A, B, C \subseteq [n]$. We say that B is locally independent of A given C, and write $A \not\to B \mid C$, if for all $\beta \in B$ the process

$$t \mapsto E\left(\lambda_t^\beta \mid \mathcal{F}_t^C\right)$$

is a version of

$$t \mapsto E\left(\lambda_t^\beta \mid \mathcal{F}_t^{C \cup A}\right).$$

We note that local independence is *asymmetric* in the sense that B being locally independent of A given C does not imply that A is locally independent of B given C. Let $\alpha, \beta \in [n]$. It follows immediately from the definition that $\alpha \not\to \beta \mid [n] \setminus \{\alpha\}$ if λ_t^β is $\mathcal{F}_t^{[n]\setminus\{\alpha\}}$-measurable. That is, if λ_t^β does not depend on the sample path of the α-coordinate.

We define a *local independence graph* below and this generalizes the definitions of Didelez [15] and Mogensen and Hansen [37] in the context of continuous-time stochastic processes to allow a nondiagonal diffusion matrix, Σ. Eichler [17] gives a related definition in the case of time series (discrete time) with correlated noise and uses the term *path diagram* (see also Definition 6). Moreover, local independence graphs can be seen as abstract generalizations of continuous-time Bayesian networks. In these models, a multivariate continuous-time Markov process in a finite state space is represented by a graph in which edges indicate how transition intensities of a coordinate process depend on the states of other coordinate processes [40,41]. Didelez [14] describes the connection between this model class and local independence.

Definition 4 (Local independence graph). Consider a regular Itô process with diffusion matrix Σ. A *local independence graph* is a graph, $\mathcal{D}$, with nodes $[n]$ such that for all $\alpha,\beta \in [n]$

$$\alpha \not\to_{\mathcal{D}} \beta \;\Rightarrow\; \alpha \not\to \beta \,|\, [n] \setminus \{\alpha\}$$

and such that for $\alpha \neq \beta$

$$\alpha \not\dashv_{\mathcal{D}} \beta \;\Rightarrow\; \Sigma_{\alpha\beta} = 0$$

where $\to_{\mathcal{D}}$ denotes a directed edge in $\mathcal{D}$ and $\mapsto_{\mathcal{D}}$ denotes a blunt edge.

A local independence graph can be inferred directly from λ and Σ, see also Definition 6 below.

2.2. The global Markov property

Graphical representations of local independence are mostly of interest when they can be used to infer nontrivial results about additional local independencies. This is the case when a *global Markov property* holds (its definition uses the concept of μ-separation which will be defined in Section 3).

Definition 5 (The global Markov property). Let X be a regular Itô process for which the coordinate processes are indexed by $[n]$ and let $\mathcal{D}$ be a local independence graph for X (Definition 4). We say that X *satisfies the global Markov property with respect to* $\mathcal{D}$ if for all $A,B,C \subseteq [n]$ it holds that if B is μ-separated by A given C in $\mathcal{D}$, then B is locally independent of A given C in the distribution of X.

We note that μ-separation is a property of the graph which means that when the global Markov property holds, we can read local independencies from the graph alone. Mogensen et al. [39] show that regular Itô processes with a diagonal σ satisfy the global Markov property with respect to their local independence graphs – assuming certain integrability constraints are satisfied – and one can read local independencies from the graph using a straightforward algorithm. This allows us to answer a filtration question: for $D \subseteq [n]$ and $\beta \in [n]$, which coordinates in D does

$$E\left(\lambda_t^\beta \,|\, \mathcal{F}_t^D\right)$$

depend upon? We conjecture that the global Markov property holds for nondiagonal σ in a broad class of processes, but this cannot be shown using the same techniques as in [39]. We do, however, show in Theorem 10 that for a particular class of Itô diffusions the global Markov property does in fact hold for the *canonical* local independence graph that will be defined below. The proof uses an explicit representation of the conditional expectation processes within this particular class of processes and therefore does not generalize to all Itô processes. Global Markov properties have also been proven in (discrete-time) time series models [22]. These results and the analogy to Markov properties in DAG-based models suggest that generalization of Theorem 10 is possible, even if other methods of proof are needed.

The global Markov property can be seen to be somewhat similar to that of chain graphs under the MVR interpretation [11,56] (see also [30]). In this sense, one can think of Theorem 10 as analogous to extending the global Markov property from DAGs to chain graphs.

2.3. Itô diffusions

A regular Itô diffusion is a regular Itô process such that the drift is of the form

$$\lambda_t = \lambda(X_t)$$

for a continuous function $\lambda : \mathbb{R}^n \to \mathbb{R}^n$. In differential form,

$$\mathrm{d}X_t = \lambda(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Itô diffusions with a constant diffusion coefficient are particularly interesting examples of Itô processes. They are Markov processes, but they are not closed under marginalization and we need to consider the larger class of Itô processes to obtain a class which is closed under marginalization.

The following definition introduces a *canonical local independence graph* which we will use to show a global Markov property. To our knowledge, this is a novel definition in this context, though it is very similar in spirit to path diagrams [61,62] and other mixed graph representations of multivariate random variables. Analogous definitions can also be found in discrete-time processes [17].

Definition 6 (Canonical local independence graph). Let X be a regular Itô diffusion with a continuously differentiable drift $\lambda : \mathbb{R}^n \to \mathbb{R}^n$ and diffusion matrix Σ. The canonical local independence graph is the graph, $\mathcal{D}$, with nodes $[n]$ such that for all $\alpha, \beta \in [n]$

$$\partial_\alpha \lambda_\beta \neq 0 \quad \Leftrightarrow \quad \alpha \to_{\mathcal{D}} \beta$$

and such that for $\alpha \neq \beta$

$$\Sigma_{\alpha\beta} \neq 0 \quad \Leftrightarrow \quad \alpha \mapsto_{\mathcal{D}} \beta.$$

As $\partial_\alpha \lambda_\beta = 0$ implies that $\lambda_t^\beta = \lambda_\beta((X_t^\delta)_{\delta \in [n] \setminus \{\alpha\}})$ is $\mathcal{F}_t^{[n]\setminus\{\alpha\}}$-measurable, the following result is an immediate consequence of Definitions 4 and 6.

Proposition 7. *The canonical local independence graph is a local independence graph.*

Definition 6 gives a simple operational procedure for determining the canonical local independence graph for a regular Itô diffusion directly from λ and Σ, though it is not guaranteed to be neither a unique nor a minimal local independence graph. If, for instance, $X = (X^\alpha, X^\beta, X^\gamma)$ is a process with $\lambda_\beta = X^\gamma X^\alpha$ then $\alpha \to \beta$ in the canonical local independence graph, but if X satisfies the constraint $X^\alpha + X^\gamma = 0$, then $\alpha \not\to \beta \,|\, \{\beta, \gamma\}$. This follows from $\lambda_\beta = X^\gamma X^\alpha = -(X^\gamma)^2$ being $\mathcal{F}_t^{\{\beta,\gamma\}}$-adapted. It is thus possible that λ_β has a functional form that appears to depend on the coordinate α, while actually $\alpha \not\to \beta \,|\, [n] \setminus \{\alpha\}$.

Example 8 (Smoluchowski diffusion). In this example we link the notion of local independence and the local independence graph to classical undirected graphical models (see, e.g., [31]) for a special class of diffusions that are widely studied in equilibrium statistical physics. A *Smoluchowski diffusion* is a regular Itô diffusion with

$$\lambda(x) = -\nabla V(x)$$

for a continuously differentiable function $V : \mathbb{R}^n \to \mathbb{R}$ and $\sigma = \sqrt{2\tau} I$ for a constant $\tau > 0$. Thus the diffusion matrix $\Sigma = 2\tau I$ is diagonal. The function V is called the potential and τ is called a temperature

parameter. Since the drift is a gradient, the dynamics of a Smoluchowski diffusion are a gradient flow perturbed by white noise. If $V(x) \to \infty$ for $\|x\| \to \infty$ and

$$Z = \int e^{-\frac{1}{\tau}V(x)} \mathrm{d}x < \infty,$$

the diffusion has the Gibbs measure with density

$$\pi(x) = \frac{1}{Z} e^{-\frac{1}{\tau}V(x)}$$

as equilibrium distribution, see Proposition 4.2 in [42]. When V is twice differentiable, Definition 6 gives the canonical local independence graph, $\mathcal{D}$, with arrows $\alpha \to_{\mathcal{D}} \beta$ whenever $\partial_\alpha \lambda_\beta = -\partial_\alpha \partial_\beta V \neq 0$. Since

$$\partial_\alpha \lambda_\beta = -\partial_\alpha \partial_\beta V = -\partial_\beta \partial_\alpha V = \partial_\beta \lambda_\alpha$$

the graph $\mathcal{D}$ enjoys the symmetry property that $\alpha \to_{\mathcal{D}} \beta$ if and only if $\beta \to_{\mathcal{D}} \alpha$. We denote by $\mathcal{G}$ the undirected version of $\mathcal{D}$, i.e., $\alpha -_{\mathcal{G}} \beta$ if and only if $\alpha \to_{\mathcal{D}} \beta$ if and only if $\beta \to_{\mathcal{D}} \alpha$ (see Figure 2). For any $\alpha, \beta \in [n]$ with $\alpha \not\to_{\mathcal{G}} \beta$ it follows from $\partial_\alpha \partial_\beta V = \partial_\beta \partial_\alpha V = 0$ that

$$V(x) = V_1(x_\alpha, x_{-\{\alpha,\beta\}}) + V_2(x_\beta, x_{-\{\alpha,\beta\}})$$

where $x_{-\{\alpha,\beta\}}$ denotes the vector x with coordinates x_α and x_β removed. From this decomposition of V we see that π has the pairwise Markov property with respect to $\mathcal{G}$, and it follows from the Hammersley-Clifford theorem that π factorizes according to $\mathcal{G}$. That is, the potential has the following additive decomposition

$$V(x) = \sum_{c \in \mathcal{C}(\mathcal{G})} V_c(x_c)$$

where $\mathcal{C}(\mathcal{G})$ denotes the cliques of $\mathcal{G}$. This establishes a correspondence between local independencies for a Smoluchowski diffusion and Markov properties of its equilibrium distribution.

We emphasize that the link in Example 8 between local independencies representing structural constraints on the dynamics on the one side and Markov properties of an equilibrium distribution on the other side is a consequence of the symmetry of the drift of Smoluchowski diffusions combined with the

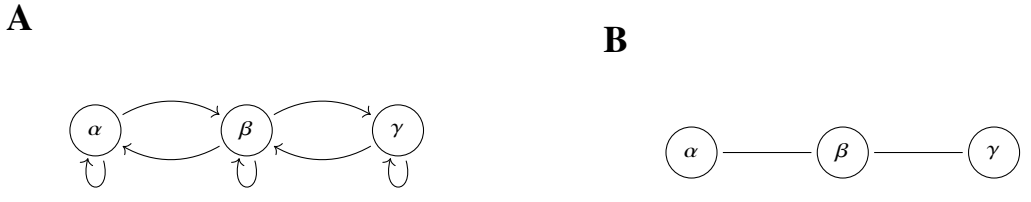

Figure 2. **A**: Canonical local independence graph of a Smoluchowski diffusion such that $\partial_\alpha \partial_\gamma V = \partial_\gamma \partial_\alpha V = 0$ (see Example 8). **B**: The equilibrium distribution of a Smoluchowski diffusion with canonical local independence graph as in **A** satisfies a Markov property with respect to this undirected graph. This would not be true of a general Ornstein-Uhlenbeck process with local independence graph as in **A** (see Example 9). This is an example of the fact that the conditional independencies of the equilibrium distribution need not reflect the local independencies of a stochastic process.

diffusion matrix being a scalar multiple of the identity matrix. For diffusions with a non-gradient drift or with a more complicated diffusion matrix the equilibrium distribution may have no conditional independencies even though there are strong structural constraints on the dynamics of the process which can be expressed in terms of a sparse local independence graph. A simple process which can illustrate this is the Ornstein-Uhlenbeck process.

Example 9 (Ornstein-Uhlenbeck process). A regular Itô diffusion with drift

$$\lambda(x) = M(x-\mu)$$

for an $n \times n$ matrix M and an n-dimensional vector μ is called a regular *Ornstein-Uhlenbeck process*. With $\mathcal{D}$ its canonical local independence graph, $\alpha \to_{\mathcal{D}} \beta$ whenever $M_{\beta\alpha} \neq 0$, and $\alpha \not\to \beta \mid [n] \setminus \{\alpha\}$ if $M_{\beta\alpha} = 0$. If M is a *stable* matrix (all eigenvalues have negative real parts), then the Ornstein-Uhlenbeck process has an invariant Gaussian distribution $\mathcal{N}(\mu, \Gamma_\infty)$ where Γ_∞ solves the Lyapunov equation,

$$M\Gamma_\infty + \Gamma_\infty M^T + \Sigma = 0,$$

see Proposition 3.5 in [42] or Theorem 2.12 in [28].

If M is also symmetric, then λ is a gradient, and if $\Sigma = 2\tau I$ we see that the solution of the Lyapunov equation is $\Gamma_\infty = -\tau M^{-1}$, and λ is the negative gradient of the quadratic potential

$$V(x) = -\frac{1}{2}(x-\mu)^T M(x-\mu) = \frac{\tau}{2}(x-\mu)^T \Gamma_\infty^{-1}(x-\mu).$$

Thus the equilibrium distribution is in a Gaussian graphical model represented by an undirected graph $\mathcal{G}$ in which the edges are determined by the non-zero entries of $\Gamma_\infty^{-1} = -\frac{1}{\tau}M$. For this Smoluchowski diffusion we see very explicitly that the edge $\alpha - \beta$ is in $\mathcal{G}$ if and only if both $\alpha \to \beta$ and $\beta \to \alpha$ are in $\mathcal{D}$. However, it is not difficult to find an asymmetric but stable matrix M such that Γ_∞^{-1} is a dense matrix, even if $\Sigma = I$, and the canonical local independence graph cannot in general be determined from Markov properties of the invariant distribution.

For a general M and general Σ, and with $D \subseteq [n]$, we see that

$$\begin{aligned} E\left(\lambda_t^\beta \mid \mathcal{F}_t^D\right) &= \sum_{\delta \in V} M_{\beta\delta}\left(E\left(X_t^\delta \mid \mathcal{F}_t^D\right) - \mu_\delta\right) \\ &= \sum_{\delta \in \mathrm{pa}(\beta)} M_{\beta\delta}\left(E\left(X_t^\delta \mid \mathcal{F}_t^D\right) - \mu_\delta\right), \end{aligned}$$

where $\mathrm{pa}(\beta) = \{\delta \mid M_{\beta\delta} \neq 0\}$ denotes the set of parents of β in $\mathcal{D}$. Thus determining if $\alpha \not\to \beta \mid C$ (Definition 3) amounts to determining if

$$E\left(X_t^\delta \mid \mathcal{F}_t^C\right)$$

are versions of

$$E\left(X_t^\delta \mid \mathcal{F}_t^{C\cup\{\alpha\}}\right)$$

for $\delta \in \mathrm{pa}(\beta)$. In words, if we can predict the values of all the processes X_t^δ, for $\delta \in \mathrm{pa}(\beta)$, that enter into the drift of coordinate β just as well from the C-histories as we can from the $C \cup \{\alpha\}$-histories, then β is locally independent of α given C.

As a final component of this section, we state a result showing that an Ornstein-Uhlenbeck process, X, satisfies a global Markov property with respect to a cDG, $\mathcal{D}$, when $\mathcal{D}$ is the canonical local independence graph of X (Definition 6). As we identify the coordinate processes of X with nodes in $\mathcal{D}$, we use $[n]$ to denote both the node set of $\mathcal{D}$ and the index set of the coordinate processes of X. In the case of a diagonal Σ, the global Markov property for Itô processes was shown in [39] under some regularity conditions, and we extend this to the case of nondiagonal Σ, i.e., allowing correlated driving Brownian motions, for Ornstein-Uhlenbeck processes.

We assume that the Ornstein-Uhlenbeck process, $X = (X_t)_{t\geq 0}$, is started at time zero and before stating the global Markov property, we will describe a condition on the initial distribution, that is, the distribution of X_0. Let $\mathcal{D}$ be a cDG. We say that the distribution of X_0 is *compatible with* $\mathcal{D}$ if for all disjoint $A, B, C \subseteq [n]$ it holds that $A \perp_m B \mid C$ in $\mathcal{D}$ (see a definition of m-separation in the supplementary material [38]) implies that X_0^A and X_0^B are conditionally independent given X_0^C. This is a quite natural assumption on the distribution of X_0 in the following sense. From $\mathcal{D}$, we can construct a *directed mixed graph* (DMG), $\mathcal{G}$, by simply replacing blunt edges in $\mathcal{D}$ with bidirected edges. Such a DMG has been used as a graphical representation of a *linear structural equation model* [30],

$$X_0 = \mu + BX_0 + \varepsilon,$$

where the directed edges of $\mathcal{G}$ represent the nonzero entries of the matrix B and the bidirected edges of $\mathcal{G}$ represent a nonzero correlation in the covariance of the Gaussian noise term ε. If X_0 is generated from a linear structural equation model with sparsity represented by $\mathcal{G}$ (the DMG corresponding to $\mathcal{D}$), then it follows from global Markov properties in this model class that the distribution of X_0 is compatible with $\mathcal{D}$ [30,57].

Theorem 10. *Let $X = (X_t)_{t\geq 0}$ be a regular Ornstein-Uhlenbeck process, let $\mathcal{D}$ be its canonical local independence graph (Definition 6), and let $A, B, C \subseteq [n]$. Assume that X_0 is a (non-degenerate) multivariate Gaussian vector and that X_0 is independent of the Brownian motion driving the Ornstein-Uhlenbeck process. Assume furthermore that the distribution of X_0 is compatible with $\mathcal{D}$. Then X satisfies the global Markov property with respect to $\mathcal{D}$.*

The result allows us to infer sparsity in the dependence structure in the evolution of the process from structural sparsity encoded by a cDG and μ-separation. The proof of Theorem 10 is found in the supplementary material [38] and it uses a set of equations describing the conditional mean processes, $t \mapsto \mathrm{E}\left[X_t^U \mid \mathcal{F}_t^W\right]$, $[n] = U \dot\cup W$, see [33]. From these somewhat explicit representations, one can reason about the measurability of the conditional mean processes. If every local independence in X implies a μ-separation in $\mathcal{D}$, we say that X is *faithful* to $\mathcal{D}$. One could possibly use similar techniques to show faithfulness results in the class of Ornstein-Uhlenbeck processes.

The following sections of the paper will develop the graph theory needed to answer questions about local independence via properties of a local independence graph. This theory can be applied as long as the processes considered have the global Markov property.

3. Directed correlation graphs

A *graph* is a pair (V, E) where V is a set of nodes and E is a set of edges. Each node represents a coordinate process and therefore we will let $V = \{1, 2 \ldots, n\} = [n]$ when we model a stochastic process $X = (X_t)_{t\in\mathcal{T}}$ such that $X_t = (X_t^\alpha)_{\alpha\in[n]} \in \mathbb{R}^n$. Every edge is between a pair of nodes. Edges can be of different types. In this paper, we will consider *directed* edges, $\to$, *bidirected* edges, $\leftrightarrow$, and *blunt* edges,

$\mapsto$. Let $\alpha, \beta \in V$. Note that $\alpha \to \beta$ and $\beta \to \alpha$ are different edges. We do not distinguish between $\alpha \leftrightarrow \beta$ and $\beta \leftrightarrow \alpha$, nor between $\alpha \mapsto \beta$ and $\beta \mapsto \alpha$. We allow directed and bidirected loops (self-edges), $\alpha \to \alpha$ and $\alpha \leftrightarrow \alpha$, but not blunt loops, $\alpha \mapsto \alpha$. If the edge $\alpha \to \beta$ is in a graph, then we say that α is a *parent* of β and write $\alpha \in \mathrm{pa}(\beta)$. If α and β are joined by a blunt edge, $\alpha \mapsto \beta$, then we say that they are *spouses*. We use $\alpha \sim \beta$ to symbolize a generic edge between $\alpha \in V$ and $\beta \in V$ of any of these three types. We say that α and β are *adjacent* in the graph $\mathcal{D}$ if $\alpha \sim \beta$ in $\mathcal{D}$. We use the notation $\alpha \sim_{\mathcal{D}} \beta$ to highlight that the edge is in $\mathcal{D}$ and we use subscript $\sim_j$ to identify edges, $j \in \mathbb{N}$. We use $\alpha \ast\!\!\to \beta$ to symbolize that either $\alpha \to \beta$ or $\alpha \leftrightarrow \beta$.

Definition 11. Let $\mathcal{D} = (V, E)$ be a graph. We say that $\mathcal{D}$ is a *directed graph* (DG) if every edge is directed. We say that $\mathcal{D}$ is a *directed correlation graph* (cDG) if every edge is directed or blunt. We say that $\mathcal{D}$ is a *directed mixed graph* (DMG) if every edge is directed or bidirected.

One should note that a bidirected edge $\alpha \leftrightarrow \beta$ is not the same as the combination of the directed edges $\alpha \to \beta$ and $\alpha \leftarrow \beta$. Therefore, between a pair of nodes, α and β, in a DMG there may be any subset of edges $\{\alpha \to \beta, \alpha \leftarrow \beta, \alpha \leftrightarrow \beta\}$. See Figure 3 for an example cDG and an example DMG.

The class of DMGs is studied by Mogensen and Hansen [37],Mogensen et al. [39]. Eichler [17, 19] studies classes of graphs similar to cDGs as well as a class of graphs which contains both the DMGs and the cDGs as subclasses. Varando and Hansen [59] use cDGs as representations of sparse parametrizations of equilibrium covariance matrices of stochastic processes in the context of structure learning. This paper is mostly concerned with the class of cDGs, however, we mention the DMGs for two reasons: 1) to compare with the cDGs and demonstrate their differences, and 2) to show that the concept of μ-separation can be applied to both classes of graphs, and therefore also to a superclass of graphs containing both the DMGs and the cDGs. In a cDG, a directed edge corresponds to a direct dependence in the drift of the process while a blunt edge represents a correlation in the driving Brownian motions (Definition 4). In a DMG, a directed edge has the same interpretation as in a cDG, however, a bidirected edge corresponds to a dependence arising from partial observation, i.e., marginalization. Correlated driving Brownian motions and marginalization create different local independence structures, hence the distinction between DMGs and cDGs.

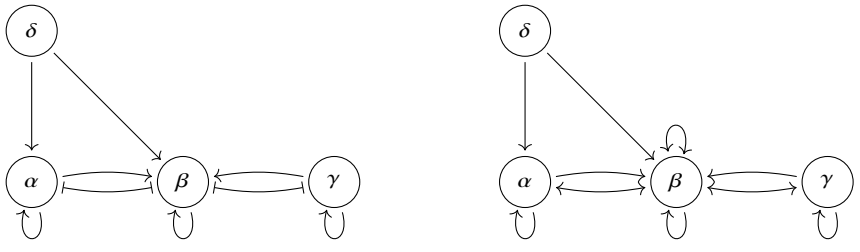

Figure 3. Example cDG (left) and example DMG (right). The blunt edges in a cDG correspond to correlated driving processes which is different from the bidirected edges of a DMG as those correspond to marginalization, i.e., unobserved processes. The notion of μ-separation can be applied to both classes of graphs. Left: cDG on nodes $V = \{\alpha, \beta, \gamma, \delta\}$. γ is μ-separated (Definition 13) from δ by α as $\beta \notin \mathrm{an}(\alpha)$ is a collider on any walk from δ to γ. On the other hand, α is not μ-separated from β given $\varnothing$ as, e.g., $\beta \mapsto \alpha \to \alpha$ is μ-connecting given $\varnothing$. The same walk is not μ-connecting from β to α given α, however, $\beta \leftarrow \delta \to \alpha$ is μ-connecting from β to α given α. We see that α is μ-separated from β given $\{\alpha, \delta\}$. Right: bidirected edges have heads at both ends and this means that $\beta \leftrightarrow \alpha$ is μ-connecting from β to α given any subset of $V \setminus \{\beta\}$. In particular, α is not μ-separated from β given $\{\alpha, \delta\}$. This is not true in the cDG (left).

A *walk*, ω, is an ordered, alternating sequence of nodes (γ_i) and edges ($\sim_j$) such that each edge, $\sim_i$, is between γ_i and γ_{i+1},

$$\gamma_1 \sim_1 \gamma_2 \sim_2 \cdots \sim_k \gamma_{k+1}.$$

For each directed edge, its orientation is also known as otherwise $\alpha \to \alpha$ and $\alpha \leftarrow \alpha$ would be indistinguishable. We say that γ_1 and γ_{k+1} are *endpoint nodes*, and we say that the walk is *from* γ_1 *to* γ_{k+1}. For later purposes, orientation of the walk is essential. We let ω^{-1} denote the walk obtained by traversing the nodes and edges of ω in reverse order. At times, we will also say that a walk, ω, is *between* γ_1 and γ_{k+1}, but only when its orientation does not matter in which case we essentially identify ω with ω^{-1}. We say that a walk is *trivial* if it has no edges and therefore only a single node, and otherwise we say that it is *nontrivial*. Consider a walk as above. We say that a non-endpoint node, γ_i, $i \notin \{1, k+1\}$, is a *collider* if the subwalk

$$\gamma_{i-1} \sim_{i-1} \gamma_i \sim_i \gamma_{i+1}$$

is of one of the following types

$$\gamma_{i-1} *\!\!\to \gamma_i \leftarrow\!\!* \gamma_{i+1},$$
$$\gamma_{i-1} *\!\!\to \gamma_i \mapsto \gamma_{i+1},$$
$$\gamma_{i-1} \mapsto \gamma_i \leftarrow\!\!* \gamma_{i+1},$$
$$\gamma_{i-1} \mapsto \gamma_i \mapsto \gamma_{i+1},$$

and otherwise we say that it is a *noncollider*. This means that the property of being a collider or a noncollider is relative to a walk and, seeing that nodes may be repeated on a walk, it is actually a property of an instance of a node on a specific walk. Note that endpoint nodes are neither colliders nor noncolliders. We say that α and β are *collider connected* if there exists a (nontrivial) walk from α to β such that every non-endpoint node is a collider.

We say that $\alpha *\!\!\to \beta$ has a *head* at β, and that $\alpha \to \beta$ has a *tail* at α. We say that $\alpha \mapsto \beta$ has a *stump* at α. We say that edges $\alpha \mapsto \beta$ and $\alpha *\!\!\to \beta$ have a *neck* at β. It follows that γ_i above is a collider if and only if both adjacent edges have a neck at γ_i. A *path* is a walk such that every node occurs at most once. We say that a path from α to β is *directed* if every edge on the path is directed and pointing towards β. If there is a directed path from α to β, then we say that α is an *ancestor* of β and that β is a *descendant* of α. We let $\mathrm{an}(\beta)$ denote the set of ancestors of β, and for $C \subseteq V$, we define $\mathrm{an}(C) = \cup_{\gamma \in C} \mathrm{an}(\gamma)$. We let $\mathrm{an}(\gamma_1, \ldots, \gamma_k)$ denote $\mathrm{an}(\{\gamma_1, \ldots, \gamma_k\})$. Note that $C \subseteq \mathrm{an}(C)$. A *cycle* is a path $\alpha \sim \ldots \sim \beta$ composed with an edge $\beta \sim \alpha$. If the path from α to β is directed and the edge is $\beta \to \alpha$, then we say that the cycle is *directed*. A DG without any directed cycles is said to be a *directed acyclic graph* (DAG).

When $\mathcal{D} = (V, E)$ is a graph and $\bar{V} \subseteq V$, we let $\mathcal{D}_{\bar{V}}$ denote the *induced graph* on nodes $\bar{V}$, i.e., $\mathcal{D}_{\bar{V}} = (\bar{V}, \bar{E})$,

$$\bar{E} = \{e \in E : e \text{ is between } \alpha, \beta \in \bar{V}\}.$$

We will use *μ-connecting walks* and *μ-separation* to encode independence structures using cDGs. These concepts were introduced in [37,39] and they generalize the notion of δ-separation [13,15]. μ-separation was originally used in DMGs, though the adaptation to cDGs is straightforward.

Definition 12 (μ-connecting walk [37]). Consider a nontrivial walk, ω,

$$\alpha \sim_1 \gamma_2 \sim_2 \cdots \sim_{k-1} \gamma_k \sim_k \beta$$

and a set $C \subseteq V$. We say that ω is *μ-connecting* from α to β given C if $\alpha \notin C$, every collider on ω is in $\mathrm{an}(C)$, no noncollider is in C, and $\sim_k$ has a head at β.

It is essential that the above definition uses walks, and not only paths. As an example consider $\alpha \mapsto \beta \leftarrow \gamma$. In this graph, there is no μ-connecting path from α to β given β, but there is a μ-connecting walk.

Definition 13 (μ-separation [37]). Let $\mathcal{D} = (V, E)$ be a cDG or a DMG and let $A, B, C \subseteq V$. We say that B is *μ-separated* from A given (or by) C in $\mathcal{D}$ if there is no μ-connecting walk from any $\alpha \in A$ to any $\beta \in B$ given C and we denote this by $A \perp_\mu B \mid C\ [\mathcal{D}]$, or just $A \perp_\mu B \mid C$.

When sets A, B, or C above are singletons, e.g., $A = \{\alpha\}$, we write α instead of $\{\alpha\}$ in the context of μ-separation. We say that the set C in the definition of μ-separation is a *conditioning set*. Mogensen and Hansen [37] introduced μ-separation as a generalization of δ-separation [13,15], however, only in DMGs, and not in cDGs. As in other classes of graphs, one can decide μ-separation in cDGs by using an auxiliary undirected graph, known as a *moral* or *augmented graph*. This is described in the supplementary material [38]. We will mostly use the above *walk-based* definition, though at times we use or mention the definition using augmented graphs when this definition is particularly useful.

The following definitions are often applied in the literature to different classes of graphs and separation criteria. When $\mathcal{D} = (V, E)$ is a cDG or DMG, we define its *independence model* (or *separation model*), $\mathcal{I}(\mathcal{D})$, as the collection of μ-separations that hold, i.e.,

$$\mathcal{I}(\mathcal{D}) = \{(A, B, C) : A, B, C \subseteq V,\ A \perp_\mu B \mid C\ [\mathcal{D}]\}.$$

Definition 14 (Markov equivalence). Let $\mathcal{D}_1 = (V, E_1)$ be a cDG or a DMG and let $\mathcal{D}_2 = (V, E_2)$ be a cDG or a DMG. We say that $\mathcal{D}_1$ and $\mathcal{D}_2$ are *Markov equivalent* if $\mathcal{I}(\mathcal{D}_1) = \mathcal{I}(\mathcal{D}_2)$.

For any finite set V, Markov equivalence is an equivalence relation on a set of graphs with node set V. When $\mathcal{D}$ is a cDG or a DMG, we let $[\mathcal{D}]$ denote the Markov equivalence class of $\mathcal{D}$ *restricted to its own class of graphs*. That is, if $\mathcal{D}$ is a cDG, then $[\mathcal{D}]$ denotes the set of Markov equivalent cDGs. If $\mathcal{D}$ is a DMG, then $[\mathcal{D}]$ denotes the set of Markov equivalent DMGs. For a cDG or DMG, $\mathcal{D} = (V, E)$, and a directed, blunt, or bidirected edge, e, between $\alpha \in V$ and $\beta \in V$, we use $\mathcal{D} + e$ to denote the graph $(V, E \cup \{e\})$.

For graphs $\mathcal{D}_1 = (V, E_1)$ and $\mathcal{D}_2 = (V, E_2)$, we write $\mathcal{D}_1 \subseteq \mathcal{D}_2$ if $E_1 \subseteq E_2$. Graphical separation criteria (including μ-separation) are most often monotone in the sense that if $\mathcal{D}_1 \subseteq \mathcal{D}_2$, then $\mathcal{I}(\mathcal{D}_2) \subseteq \mathcal{I}(\mathcal{D}_1)$. In this case, we define the notion of a *maximal* graph.

Definition 15 (Maximality). Let $\mathcal{D} = (V, E)$ be a cDG (DMG). We say that $\mathcal{D}$ is *maximal* if no directed or blunt (directed or bidirected) edge can be added Markov equivalently, i.e., if for every directed or blunt (directed or bidirected) edge, e, such that $e \notin E$, it holds that $\mathcal{D}$ and $\mathcal{D} + e$ are not Markov equivalent.

Remark 16. Eichler [17],Eichler and Didelez [22], and Eichler [18] describe graphs that represent local independence (or Granger non-causality) in time series in the presence of correlated noise processes.

In those papers, undirected (dashed or solid) edges, $-$, are used to represent the correlations among the noise variables, while we use blunt edges, $\mapsto$. Using an undirected edge could suggest that the edge acts like an edge with tails in both ends which is not the case. A blunt edge does also not act like a bidirected edge in a DMG, and this warrants the usage of an edge with a third kind of mark.

Notational clarity and simplicity become even more important when considering graphical marginalizations of cDGs. When marginalizing a cDG one needs to consider edges that, when composed with other edges, act like a blunt edge in one end and like a directed edge in the other (see also [19]). This can naturally be visualized by the edge $\mapsto$. For example, the graph $\alpha \to \beta \mapsto \gamma \to \delta$ leads to the graph $\alpha \to \beta \mapsto \delta$ when γ is marginalized away. We will not in this paper pursue this larger class of graphs, but our choice of the blunt edge, $\mapsto$, to represent correlations among the noise variables was made so that it extends naturally to marginalized cDGs.

4. Markov equivalence of directed correlation graphs

Different cDGs can encode the same separation model and in this section we will describe the Markov equivalence classes of cDGs. This is essential as it allows us to understand which graphical structures represent the same local independencies. This understanding is needed if we want to learn graphical representations from tests of local independence in observed data. We begin the section by noting a strong link between the independence model of a cDG and its directed edges.

Proposition 17. *Let $\mathcal{D} = (V, E)$ be a cDG. Then $\alpha \to_{\mathcal{D}} \beta$ if and only if $\alpha \not\perp_{\mu} \beta \,|\, V \setminus \{\alpha\}$ does not hold.*

The proposition can be found in [37] in the case of DGs. It implies that if $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent cDGs, then they have the same directed edges, and therefore $\mathrm{an}_{\mathcal{D}_1}(C) = \mathrm{an}_{\mathcal{D}_2}(C)$ for all node sets C. We will often omit the subscript when it is clear from the context from which graph(s) the ancestry should be read.

Proof. If the edge is in the graph, it is μ-connecting given any subset of V that does not contain α, in particular given $V \setminus \{\alpha\}$. On the other hand, assume $\alpha \to \beta$ is not in the graph. Any μ-connecting walk from α to β must have a head at β,

$$\alpha \sim \ldots \sim \gamma \to \beta.$$

We must have that $\gamma \neq \alpha$, and it follows that γ is in the conditioning set, i.e., the walk is closed. □

In graphs that represent conditional independence in multivariate distributions, such as ancestral graphs and acyclic directed mixed graphs, one can use *inducing paths* to characterize which nodes cannot be separated by any conditioning set [45,60]. In DMGs, inducing paths can be defined similarly [37]. In cDGs, we define both inducing paths and *weak inducing paths*. We say that a path is a *collider path* if every non-endpoint node on the path is a collider. If $\alpha \neq \beta$, then $\alpha \to \beta$ and $\alpha \mapsto \beta$ are both collider paths.

Definition 18 (Inducing path (strong)). A (nontrivial) collider path from α to β is a *(strong) inducing path* if the final edge has a head at β and every non-endpoint node is an ancestor of α or of β.

Mogensen and Hansen [37] also allow cycles in the definition of inducing paths. In the following, we assume that $\alpha \to \alpha$ for all $\alpha \in V$ and therefore this would be an unnecessary complication. We

see immediately that in a cDG, the only inducing path is a directed edge. However, we include this definition to conform with the terminology in DMGs where more elaborate inducing paths exist. In this paper, we drop one of the conditions from Definition 18 to obtain a graphical structure which is more interesting in cDGs, a *weak inducing path*.

Definition 19 (Weak inducing path). A (nontrivial) collider path between α and β is a *weak inducing path* if every non-endpoint node is an ancestor of α or of β.

We note that a strong inducing path is also a weak inducing path. Furthermore, if there is a weak inducing path from α to β, there is also one from β to α, and this justifies saying that a weak inducing path is *between* α and β in Definition 19. Also note that a *weak* inducing path is most often called an inducing path in the literature on acyclic graphs. When we just say *inducing path*, we mean a strong inducing path.

If $\mathcal{D}$ is a cDG such that $\alpha \to_{\mathcal{D}} \alpha$ for all $\alpha \in V$, then we say that $\mathcal{D}$ *contains every loop*. From this point on, we will assume that the cDGs we consider all contain every loop.

Proposition 20. *Let $\mathcal{D} = (V, E)$ be a cDG such that $\alpha \to \alpha$ for all $\alpha \in V$. There is a weak inducing path between α and β if and only if there is no $C \subseteq V \setminus \{\alpha, \beta\}$ such that $\alpha \perp_\mu \beta \mid C$.*

Mogensen and Hansen [37] show a similar result in the case of strong inducing paths in DMGs. We say that β is *inseparable* from α if there is no $C \subseteq V \setminus \{\alpha\}$ such that β is μ-separated from α by C.

Example 21. Mogensen and Hansen [37] use μ-separation in *directed mixed graphs* (DMGs) to represent local independence models. It is natural to ask if the independence models of cDGs can be represented by DMGs. The answer is no and to show this we consider the cDG in Figure 4**A**. We ask

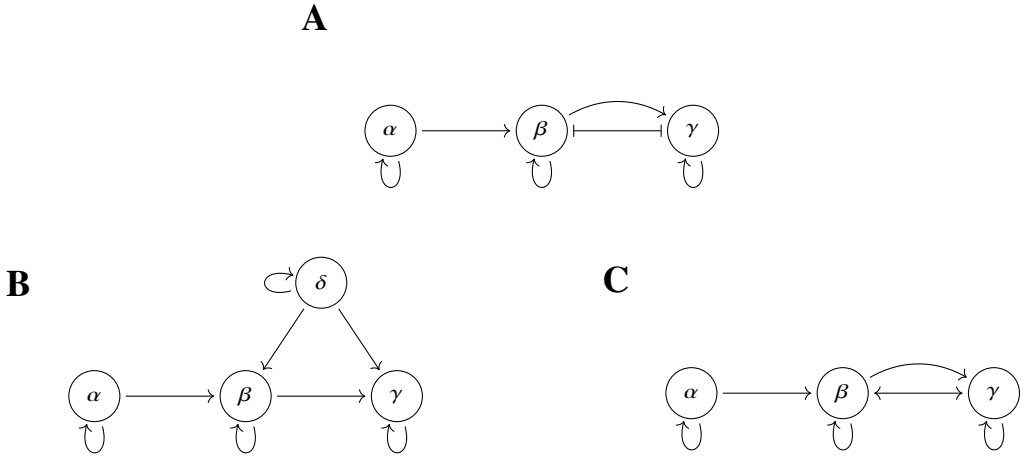

Figure 4. **A**: A cDG, $\mathcal{D}$, on nodes $V = \{\alpha, \beta, \gamma\}$ such that the separation model $\mathcal{I}(\mathcal{D})$ cannot be represented by a DMG on nodes V (see Example 21). **B**: A *directed graph* (DG). **C**: When the δ-node (δ-process) is unobserved in **B**, the DG *marginalizes* to the *directed mixed graph* (DMG) in **C** in the sense that the local independencies over the observed set of coordinate processes, $\{\alpha, \beta, \gamma\}$, implied by μ-separation are the same in graphs **B** and **C** [37]. Bidirected loops are omitted from the visualization of the DMG.

then if there exists a DMG on the same node set which has the same set of μ-separations as this cDG. In the cDG, we see that the node γ is separable from α and vice versa, i.e., there can be no edge between the two in the DMG. The node γ is not separated from α given $\{\beta\}$, and therefore β must be a collider on a path between the two. However, then there is a head at β on an edge from γ and therefore β is inseparable from γ which is a contradiction. This shows that the independence model of the cDG in Figure 4**A** cannot be represented by any DMG on the same node set. It follows that the set of separation models of cDGs on some node set is not in general a subset of the separation models of DMGs on the same node set. Similarly, one can find DMGs that are not Markov equivalent with any cDG.

The DMGs represent local independence in partially observed stochastic processes (some coordinate processes may be unobserved) through a global Markov property using μ-separation. The bidirected edges represent unobserved 'confounder' processes in the DMGs, i.e., confounder processes that have been marginalized away (see Figure 4, graphs **B** and **C**, for an example). Both DMGs and cDGs represent local independence through their respective global Markov properties and μ-separation. We note that the local independence models corresponding to marginalization and correlated noise are different when represented by DMGs and cDGs, respectively, as there exist DMGs that are not Markov equivalent with any cDG and vice versa.

DGs constitute a subclass of cDGs and within the class of DGs every Markov equivalence class is a singleton, i.e., two DGs are Markov equivalent if and only if they are equal.

Proposition 22 (Markov equivalence of DGs [37]). *Let $\mathcal{D}_1 = (V, E_1)$ and $\mathcal{D}_2 = (V, E_2)$ be DGs. Then $\mathcal{D}_1 \in [\mathcal{D}_2]$ if and only if $\mathcal{D}_1 = \mathcal{D}_2$.*

Proposition 22 does not hold in general when $\mathcal{D}_1$ and $\mathcal{D}_2$ are cDGs. As an example, consider a graph on nodes $\{\alpha, \beta\}$ such that $\alpha \to \beta$ and $\beta \to \alpha$. This graph is Markov equivalent with the graph where $\alpha \mapsto \beta$ is added. The next result is an immediate consequence of Proposition 17 and shows that Markov equivalent cDGs always have the same directed edges.

Corollary 23. *Let $\mathcal{D}_1 = (V, E_1)$ and $\mathcal{D}_2 = (V, E_2)$ be cDGs. If they are Markov equivalent, then for all $\alpha, \beta \in V$ it holds that $\alpha \to_{\mathcal{D}_1} \beta$ if and only if $\alpha \to_{\mathcal{D}_2} \beta$.*

We say that a graph, $\mathcal{D}$, is a *greatest* element of its equivalence class, $[\mathcal{D}]$, if it is a supergraph of all members of the class, i.e., $\tilde{\mathcal{D}} \subseteq \mathcal{D}$ for all $\tilde{\mathcal{D}} \in [\mathcal{D}]$. We say that $\mathcal{D}$ is a *least* element if $\mathcal{D} \subseteq \tilde{\mathcal{D}}$ for all $\tilde{\mathcal{D}} \in [\mathcal{D}]$. Mogensen and Hansen [37] show the below result on Markov equivalence.

Theorem 24 (Greatest Markov equivalent DMG [37]). *Let $\mathcal{G}$ be a directed mixed graph. Then $[\mathcal{G}]$ has a greatest element (within the class of DMGs), i.e., there exists $\bar{\mathcal{G}} \in [\mathcal{G}]$ such that $\bar{\mathcal{G}}$ is a supergraph of all Markov equivalent DMGs.*

The theorem provides a concise and intuitive way to understand sets of Markov equivalent DMGs. If $\mathcal{G}$ is a DMG, then we can visualize $[\mathcal{G}]$ by drawing its greatest element and simply showing which edges are in every DMG in $[\mathcal{G}]$ and which are only in some DMGs in $[\mathcal{G}]$. cDGs represent local independencies allowing for correlation in the driving noise processes and one can ask if the same result on Markov equivalence holds in this class of graphs. The answer is in the negative as illustrated by the following example.

Example 25. Consider the graph, $\mathcal{D}$, to the left on the first row of Figure 5. The edge $\alpha \mapsto \gamma$ can be added Markov equivalently and the edge $\beta \mapsto \gamma$ can be added Markov equivalently (center and right

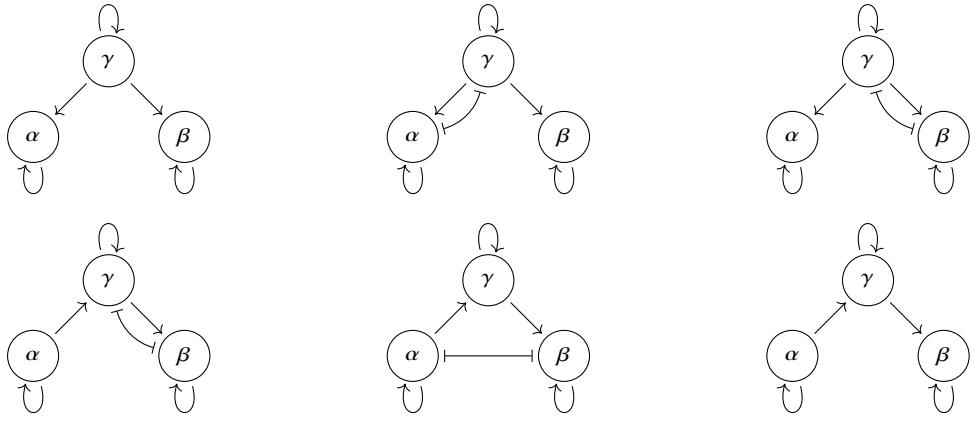

Figure 5. First row: An equivalence class illustrating that a greatest element need not exist (see Example 25). Second row: The left and center graphs are Markov equivalent. The graph on the right is the largest graph which is a subgraph of both of them, and this graph is not Markov equivalent, i.e., the Markov equivalence class of the left and center graphs does not have a least element. Theorem 31 gives a characterization of Markov equivalence of cDGs.

graphs), but adding them both results in a graph which is no longer Markov equivalent with $\mathcal{D}$. This shows that the equivalence class of $\mathcal{D}$ does not contain a greatest element. Figure 5 also gives an example showing that an equivalence class of cDGs does not necessarily contain a least element.

4.1. A characterization of Markov equivalence of cDGs

When we have a global Markov property, such as the one in Theorem 10, the μ-separations of a cDG imply local independencies in the distribution of the stochastic process. We saw in Figure 5 that different cDGs may represent the same μ-separations and it is therefore important to understand which cDGs are equivalent in terms of the μ-separations that they entail, that is, are Markov equivalent. The central result of this section is a characterization of Markov equivalence of cDGs. We introduce the notion of *collider equivalence* of graphs as a first step in stating this result.

Definition 26. Let $\mathcal{D}_1 = (V, E_1)$, $\mathcal{D}_2 = (V, E_2)$ be cDGs with the same directed edges, and let ω be a (nontrivial) collider path in $\mathcal{D}_1$,

$$\alpha \sim \gamma_1 \sim \ldots \sim \gamma_{k_1} \sim \beta.$$

We say that ω is *covered* in $\mathcal{D}_2$ if there exists a (nontrivial) collider path in $\mathcal{D}_2$

$$\alpha \sim \bar{\gamma}_1 \sim \ldots \sim \bar{\gamma}_{k_2} \sim \beta$$

such that for each $\bar{\gamma}_j$ we have $\bar{\gamma}_j \in \mathrm{an}(\alpha, \beta)$ or $\bar{\gamma}_j \in \cup_i \mathrm{an}(\gamma_i)$.

In the above definition $\{\gamma_j\}$ or $\{\bar{\gamma}_j\}$ may be the empty set, corresponding to α and β being adjacent, $\alpha \sim \beta$. One should also note that a single edge, $\alpha \sim \beta$, constitutes a collider path between α and β (when

$\alpha \neq \beta$) and that a single edge covers any collider path between α and β as it has no non-endpoint nodes. When $\mathcal{D}_1$ and $\mathcal{D}_2$ have the same directed edges it holds that $\mathrm{an}_{\mathcal{D}_1}(C) = \mathrm{an}_{\mathcal{D}_2}(C)$ for all $C \subseteq V$ and therefore one can read the ancestry of α, β, and $\{\gamma_i\}$ in either of the graphs in the above definition.

Definition 27 (Collider equivalence). Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be cDGs on the same node set and with the same directed edges. We say that $\mathcal{D}_1$ and $\mathcal{D}_2$ are *collider equivalent* if every (nontrivial) collider path in $\mathcal{D}_1$ is covered in $\mathcal{D}_2$ and every (nontrivial) collider path in $\mathcal{D}_2$ is covered in $\mathcal{D}_1$.

In the context of collider equivalence, the convention that every node is an ancestor of itself, i.e., $\gamma \in \mathrm{an}(\gamma)$ for all $\gamma \in V$, is important. From this convention, it follows immediately that every cDG is collider equivalent with itself. However, this would not necessarily be the case without this convention.

We do not need to consider walks in the above definitions (only paths) as we assume that all directed loops are included and therefore all nodes are collider connected to themselves by assumption. If there is a collider walk between α and β ($\alpha \neq \beta$), then there is also a collider path. Furthermore, if a collider walk between α and β ($\alpha \neq \beta$) in $\mathcal{D}_1$ is covered by a collider walk in $\mathcal{D}_2$, then it is also covered by a collider path, and we see that one would obtain an equivalent definition by using collider walks instead of collider paths in Definitions 26 and 27.

Remark 28. Collider equivalence implies that two graphs have the same weak inducing paths in the following sense. Assume ω is a weak inducing path between α and β in $\mathcal{D}_1$, and that $\mathcal{D}_1$ and $\mathcal{D}_2$ are collider equivalent and have the same directed edges. In $\mathcal{D}_2$, there exists a collider path, $\bar{\omega}$, such that every non-endpoint node is an ancestor of a node on ω, i.e., an ancestor of $\{\alpha,\beta\}$ using the fact that ω is a weak inducing path. This means that $\bar{\omega}$ is a weak inducing path in $\mathcal{D}_2$.

Lemma 29. *Let $\mathcal{D}_1 = (V, E_1)$, $\mathcal{D}_2 = (V, E_2)$ be cDGs that contain every loop. If $\mathcal{D}_1$ and $\mathcal{D}_2$ are not collider equivalent, then they are not Markov equivalent.*

Proposition 30. *Assume $\alpha,\beta \notin C$. If ω is a collider path between α and β such that every collider is in $an(\{\alpha,\beta\} \cup C)$, then there is a walk between α and β such that no noncollider is in C and every collider is in $an(C)$.*

A more general version of Proposition 30 was shown by Richardson [44] using a similar proof and m-separation (a definition of m-separation is given in the supplementary material [38]).

Theorem 31 (Markov equivalence of cDGs). *Let $\mathcal{D}_1 = (V, E_1)$ and $\mathcal{D}_2 = (V, E_2)$ be cDGs that contain every loop. The graphs $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent if and only if they have the same directed edges and are collider equivalent.*

We give a direct proof of this theorem. One can also use the augmentation criterion in the supplementary material [38] to show this result.

Proof. Assume first that $\mathcal{D}_1$ and $\mathcal{D}_2$ have the same directed edges and are collider equivalent. Then $\mathrm{an}_{\mathcal{D}_1}(C) = \mathrm{an}_{\mathcal{D}_2}(C)$ for all $C \subseteq V$ so we will omit the subscript and write simply $\mathrm{an}(C)$. Let ω denote a μ-connecting walk from α to β given C in $\mathcal{D}_1$. We will argue that we can also find a μ-connecting walk in $\mathcal{D}_2$. We say that a nontrivial subwalk of ω is a *maximal collider segment* if all its non-endpoint nodes are colliders on ω, its endpoint nodes are not colliders, and it contains at least one blunt edge (note that on a general walk this should be read as *instances* of these nodes and edges as nodes and edges may be repeated on a walk). We can partition ω into a sequence of subwalks such that every

Figure 6. The two cDGs constitute a Markov equivalence class, and they are both seen to be maximal. However, they do not have the same adjacencies. A similar phenomenon can occur in DGs (without loops) under d-separation [46,48].

subwalk is either a maximal collider segment, or a subwalk consisting of directed edges only. We note that maximal collider segments may be adjacent, i.e., share an endpoint. Every segment of ω that consists of directed edges only is also present in $\mathcal{D}_2$. Consider a maximal collider segment between δ and ε. This is necessarily a collider walk in $\mathcal{D}_1$. If $\delta \neq \varepsilon$, there exists a collider path between δ and ε in $\mathcal{D}_1$, and therefore a covering collider path, ρ, in $\mathcal{D}_2$ using collider equivalence. The final edge of ω must be directed and point towards the final instance of β and therefore β is not in a maximal collider segment, and δ and ε are not the final node on ω. If $\delta = \varepsilon$, then we can remove this segment from ω and obtain a connecting walk as the final β is not an endpoint of a maximal collider segment. Assume $\delta \neq \varepsilon$. δ and ε are noncolliders on ω, or endpoint nodes on ω. If $\delta = \alpha$ or $\varepsilon = \alpha$, then they are not in C. In either case, we see that $\delta, \varepsilon \notin C$. We will now find an open (given C) walk between δ and ε using ρ. We know that ρ is a collider path and that every non-endpoint node on ρ is an ancestor of $\{\delta, \varepsilon\}$ or of a collider in the original maximal collider segment, and therefore of C. It follows from Proposition 30 that we can find a walk between δ and ε such that no noncollider is in C and every collider is in $\mathrm{an}(C)$. We create a walk from α to β in $\mathcal{D}_2$ by simply substituting each maximal collider segment with the corresponding open walk. This walk is open in any node which is not an endpoint of a maximal collider segment. If an endpoint of a maximal collider segment changes collider status on this new walk, then it must be a noncollider on ω and a parent of a node in $\mathrm{an}(C)$, i.e., also in $\mathrm{an}(C)$ itself. Finally, we note that the last segment (into β) is not a maximal collider segment and therefore still has a head into β.

On the other hand, if they do not have the same directed edges, it follows from Proposition 17 that they are not Markov equivalent. If they are not collider equivalent, it follows from Lemma 29 that they are not Markov equivalent. □

In the case of *directed acyclic graphs* it holds that Markov equivalent graphs have the same adjacencies, however, this is not true in the case of cDGs, and in fact, it is also not true among maximal cDGs (Definition 15) as seen in Figure 6.

Proposition 32. *Let $\mathcal{D} = (V, E)$ be a cDG, and let $\alpha, \beta \in V$. Let e denote a blunt edge between α and β. If α and β are connected by a weak inducing path consisting of blunt edges only, then $\mathcal{D} + e \in [\mathcal{D}]$.*

4.2. Markov equivalent permutation of nodes

The example in Figure 6 shows a characteristic of some Markov equivalent cDGs. In the example, one can obtain one graph from the other by a permutation of the endpoints of blunt edges within the set

$\{\gamma,\delta\}$. In this section, we formulate sufficient conditions for a cDG to be Markov equivalent with a *permutation graph*. While the formal definition of a permutation graph is new to our knowledge, the basic idea is also found in earlier work on Markov equivalence of DGs under d-separation [46,48].

Definition 33 (Cyclic set). We say that $S \subseteq V$ is a *cyclic set* if for every $(\alpha,\beta) \in S \times S$, it holds that $\alpha \in \mathrm{an}(\beta)$.

The following is a formal definition of a *permutation graph* as illustrated in the example of Figure 6.

Definition 34 (Permutation graph). Let $\mathcal{D} = (V,E)$ be a cDG and let ρ be a permutation of the node set, V. We define $\mathcal{P}_\rho(\mathcal{D})$ as the cDG on nodes V such that

$$\alpha \to_{\mathcal{P}_\rho(\mathcal{D})} \beta \qquad \text{if and only if } \alpha \to_{\mathcal{D}} \beta,$$
$$\rho(\alpha) \mapsto_{\mathcal{P}_\rho(\mathcal{D})} \rho(\beta) \text{ if and only if } \alpha \mapsto_{\mathcal{D}} \beta.$$

Proposition 35. *Let $\mathcal{D} = (V,E)$ be a cDG which contains every loop and let $S \subseteq V$. Let ρ be a permutation of V such that $\rho(\alpha) = \alpha$ for all $\alpha \notin S$. If $\mathrm{pa}(\beta) = \mathrm{pa}(\gamma)$ for all $\beta,\gamma \in S$, then $\mathcal{P}_\rho(\mathcal{D}) \in [\mathcal{D}]$.*

Note that $\mathrm{pa}(\beta) = \mathrm{pa}(\gamma)$ implies that $\beta \to_{\mathcal{D}} \gamma$ as $\beta \in \mathrm{pa}(\beta)$ for all β. Furthermore, the fact that $\beta \to_{\mathcal{D}} \gamma$ for all $\beta,\gamma \in S$ implies that S is a cyclic set.

Figure 6 shows two graphs that are Markov equivalent by Proposition 35. In some graphs one can find permutations, not fulfilling the assumptions of Proposition 35, that generate Markov equivalent graphs, and this proposition is therefore not a necessary condition for Markov equivalence under permutation of blunt edges. One example is in the first row of Figure 7. The center and right graphs are Markov equivalent and one is generated from the other by permuting the blunt edges of β and γ, however, the conditions of Proposition 35 are not fulfilled.

5. Deciding Markov equivalence

In this section, we will consider the problem of deciding Markov equivalence algorithmically. That is, given two cDGs on the same node set, how can we decide if they are Markov equivalent or not? A possible starting point is Theorem 31. While it is computationally easy to check whether the directed edges of two cDGs are the same (quadratic in the number of nodes in their mutual node set), collider equivalence could be hard as there may be exponentially many collider paths in a cDG. In this section, we give a different characterization of Markov equivalence (Theorem 36) which proves the correctness of a simple algorithm (Algorithm 1) for deciding Markov equivalence of two cDGs. This algorithm avoids checking each collider path explicitly. However, in the worst case it also has a superpolynomial runtime which is to be expected due to the complexity result in Theorem 42.

The *directed part* of a cDG, $\mathbb{D}(\mathcal{D}) = (V,F)$, is the DG on nodes V such that $\alpha \to_{\mathbb{D}(\mathcal{D})} \beta$ if and only if $\alpha \to_{\mathcal{D}} \beta$. The *blunt part* of a cDG, $\mathbb{U}(\mathcal{D})$, is the cDG obtained by removing all directed edges. Let $\mathcal{G} = (V,E)$ be a graph with only blunt edges. The *connected components* of $\mathcal{G}$ are the disjoint sets, $V_1,\ldots,V_l$, such that $\cup_i V_i = V$ and such that $\alpha,\beta \in V_i$ if and only if there is a path in $\mathcal{G}$ connecting α and β. The connected components can be found in time which is proportional to $\max(|V|,|E|)$ [26]. The *blunt components* of $\mathcal{D}$ are the connected components of $\mathbb{U}(\mathcal{D})$. We say that $\mathcal{D}_1 = (V,E_1)$ and $\mathcal{D}_2 = (V,E_2)$ have *the same collider connections* if it holds for all $\alpha \in V$ and $\beta \in V$ that α and β are collider connected

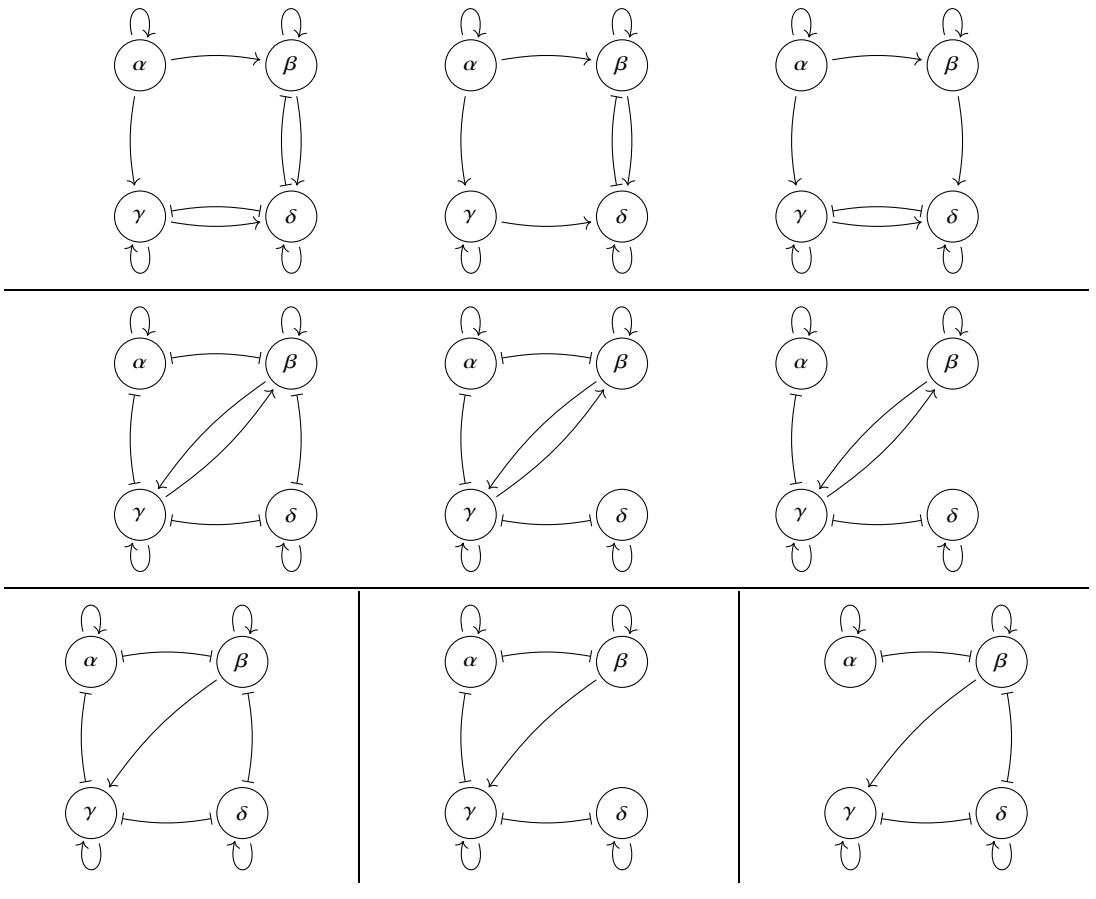

Figure 7. Markov equivalence in cDGs. Graphs from different Markov equivalence classes are separated by line segments. First row: These are three members of a Markov equivalence class of size 21. The only restriction on 2^5 combinations of blunt edges (all but $\beta \mapsto \gamma$ can be present) is the fact that we cannot have both $\alpha \mapsto \beta$ and $\alpha \mapsto \gamma$ present and that either (α,δ), (β,δ), or (γ,δ) are spouses as otherwise there would not be a weak inducing path between α and δ. Second row: These graphs are Markov equivalent. The collider path $\alpha \mapsto \beta \mapsto \delta$ in the first graph is covered in the two others by the walk $\alpha \mapsto \gamma \mapsto \delta$ as $\gamma \in \mathrm{an}(\beta)$. The edge $\beta \mapsto \delta$ is covered by the weak inducing path $\delta \mapsto \gamma \leftarrow \beta$ in the center and right graphs of the row. The equivalence class of these graphs has cardinality 16 which is every combination of blunt edges (excluding $\alpha \mapsto \delta$ which cannot be in a Markov equivalent graph) that makes the graph connected via blunt edges as well as two more (one of which is the rightmost graph of this row). Third row: The first graph is not collider equivalent with the following two: the collider path $\alpha \mapsto \beta \mapsto \delta$ is not covered by any collider path in the second graph; the collider path $\alpha \mapsto \gamma$ is not covered by any collider path in the third.

in $\mathcal{D}_1$ if and only if they are collider connected in $\mathcal{D}_2$. We say that a subset of nodes, A, is *ancestral* if $A = \mathrm{an}(A)$. We will only consider cDGs that contain every loop.

We start from the following result which is seen to be a reformulation of the augmentation criterion in the supplementary material [38].

Algorithm 1 Markov equivalence

Require: cDGs, $D_1 = (V, E_1), D_2 = (V, E_2)$
  if $\mathbb{D}(\mathcal{D}_1) \neq \mathbb{D}(\mathcal{D}_2)$ **then**
    return FALSE
  end if
  for $\mathbf{A} \in \mathbb{A}(\mathcal{C}(\mathcal{D}_1))$ **do**
    Define $A = \bigcup_{C \in \mathbf{A}} C$
    if $(\mathcal{D}_1)_A$ and $(\mathcal{D}_2)_A$ do not have the same collider connections **then**
      return FALSE
    end if
  end for
  return TRUE

Theorem 36. *Let $\mathcal{D}_1 = (V, E_1)$ and $\mathcal{D}_2 = (V, E_2)$ be cDGs (both containing every loop) such that $\mathbb{D}(\mathcal{D}_1) = \mathbb{D}(\mathcal{D}_2)$. $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent if and only if for every ancestral set it holds that $(\mathcal{D}_1)_A$ and $(\mathcal{D}_2)_A$ have the same collider connections.*

The above theorem can easily be turned into an algorithm for deciding if two cDGs are Markov equivalent (Algorithm 1). However, there may be exponentially many ancestral sets in a cDG. For instance, in the case where the only directed edges are loops all subsets of V are ancestral and therefore the algorithm would need to compare collider connections in 2^n pairs of graphs where n is the number of nodes in the graphs (of course, one could omit the empty set and singletons).

5.1. An algorithm for deciding equivalence

In the algorithm based on Theorem 36 we will use the *condensation* of a cDG. This is not needed, but does provide a convenient representation of the ancestor relations between nodes in a cyclic graph. Let $\mathcal{D} = (V, E)$ be a cDG. We say that $\alpha, \beta \in V$ are *strongly connected* if there exists a directed path from α to β and a directed path from β to α, allowing trivial paths. Equivalently, α and β are strongly connected if and only if $\alpha \in \mathrm{an}(\beta)$ and $\beta \in \mathrm{an}(\alpha)$. This is an equivalence relation on the node set of a cDG and we say that the equivalence classes are the *strongly connected components* of the graph. The definition of strong connectivity is often used in DGs [10]. We simply use a straightforward generalization to the class of cDGs such that the directed part of the cDG determines strong connectivity. The strongly connected components are also the maximal cyclic sets (Definition 33).

The *condensation* of $\mathcal{D}$ (also known as the *acyclic component graph* of $\mathcal{D}$) is the directed acyclic graph obtained by contracting each strongly connected component to a single vertex. That is, if $C_1, \ldots, C_m$ are the strongly connected components of $\mathcal{D}$ ($C_i \subseteq V$ for all i), then the condensation of $\mathcal{D}$ has node set $\mathbb{C} = \{C_1, \ldots, C_m\}$ and $C_i \to C_j$ if $i \neq j$ and there exists $\alpha \in C_i, \beta \in C_j$ such that $\alpha \to_{\mathcal{D}} \beta$ [10]. We denote the condensation of $\mathcal{D}$ by $\mathcal{C}(\mathcal{D})$. We also define the *completed condensation* of $\mathcal{D}$, $\bar{\mathcal{C}}(\mathcal{D})$, which is the graph on nodes $\mathbb{C} \cup \{\varnothing\}$ such that $\bar{\mathcal{C}}(\mathcal{D})_{\mathbb{C}} = \mathcal{C}(\mathcal{D})$ and such that $\varnothing$ is a parent of every other node and a child of none. The condensation and the completed condensation are both DAGs. When $\mathcal{D}$ has d directed edges that are not loops, then strongly connected components can be found in linear time, that is, $O(n + d)$ where $n = |V|$ [10].

In the following, we will be considering sets of nodes in $\mathcal{D}$, i.e., subsets of V, as well as sets of nodes in $\mathcal{C}(\mathcal{D})$, that is, subsets of $\mathbb{C}$. We write the former as capital letters, A, B, C. We write the latter as capital letters in bold font, $\mathbf{A}, \mathbf{B}, \mathbf{C}$, to emphasize that they are subsets of $\mathbb{C}$, not of V.

Proposition 37. *The ancestral sets in $\mathcal{D}$ are exactly the sets of the form $\bigcup_{C\in\mathbf{A}} C$ for an ancestral set, $\mathbf{A}$, in $\mathcal{C}(\mathcal{D})$.*

The above proposition shows that we can consider the condensation when finding ancestral sets in a cDG. We let $\mathbb{A}(\mathcal{D})$ denote the set of ancestral sets in $\mathcal{D}$. The correctness of Algorithm 1 follows from Theorem 36 and Proposition 37. The algorithm considers ancestral sets in the condensation, however, a version using ancestral sets directly in $\mathcal{D}_1$ is of course also possible and this essentially gives an equivalent algorithm. In the algorithm, one can decide collider connectivity by noting that α and β are collider connected in a cDG, $\mathcal{D}$, if and only if there exists a blunt component, U, such that $\alpha \in \mathrm{pa}_{\mathcal{D}}(U)$ and $\beta \in \mathrm{pa}_{\mathcal{D}}(U)$, using that the graphs contain every loop.

5.2. Virtual collider tripaths

This section describes a graphical structure that we will call a *virtual collider tripath*. We will use these to give a necessary condition for Markov equivalence which is computationally easy to check. However, this is only a necessary condition and not sufficient for Markov equivalence. The results in this subsection therefore lead to an algorithm for checking Markov equivalence which is computationally feasible, but is only an approximation in the sense that it will not always correctly distinguish between graphs that are not Markov equivalent. In the next section we see that the problem of deciding Markov equivalence of cDGs is coNP-complete and therefore we should not expect to find an algorithm which is always correct and also computationally efficient.

Definition 38 (Virtual collider tripath). Let $\alpha,\beta \in V$ and let C be a node in $\bar{\mathcal{C}}(\mathcal{D})$, i.e., C is a strongly connected component or the empty set. We say that (α,β,C) is a *virtual collider tripath* if there exists a (nontrivial) collider path $\alpha \sim \gamma_1 \sim \ldots \gamma_m \sim \beta$ such that $\gamma_i \in \mathrm{an}(\{\alpha,\beta\} \cup C)$ for all $i = 1,\ldots,m$.

Note that if $\alpha = \beta$, then there is no path fulfilling the requirements of Definition 38, hence (α,α,C) is not a virtual collider tripath for any C. Richardson [47] describes *virtual adjacencies* in DGs equipped with d-separation. Those are structures that in terms of separation act as adjacencies. The idea behind virtual collider tripaths is similar; for a fixed pair of nodes, α and β, a virtual collider tripath, (α,β,C), acts as if there exists $\gamma \in C$ such that $\alpha \sim \gamma \sim \beta$ is a collider walk. Note also that if α and β are adjacent, then (α,β,C) is a virtual collider tripath for any strongly connected component C. Finally, note that there are no restrictions on whether or not α, β, or both are elements in the set $C \subseteq V$.

Definition 39 (Maximal virtual collider tripath). We say that a virtual collider tripath, (α,β,C), is *maximal* if there is no $\tilde{C} \neq C$ such that $(\alpha,\beta,\tilde{C})$ is a virtual collider tripath and $\tilde{C}$ is an ancestor of C in $\bar{\mathcal{C}}(\mathcal{D})$.

We say that two cDGs, $\mathcal{D}_1$ and $\mathcal{D}_2$, have the same (maximal) virtual collider tripaths if it holds that (α,β,C) is a (maximal) virtual collider tripath in $\mathcal{D}_1$ if and only if (α,β,C) is a (maximal) virtual collider tripath in $\mathcal{D}_2$.

Proposition 40. *Let C be a strongly connected component or the empty set. If (α,β,C) is not a virtual collider tripath and $\alpha \neq \beta$, then β and α are m-separated by $\mathrm{an}(\{\alpha,\beta\} \cup C) \setminus \{\alpha,\beta\}$.*

A definition of m-separation can be found in the supplementary material [38]. The next theorem gives a necessary condition for Markov equivalence of cDGs.

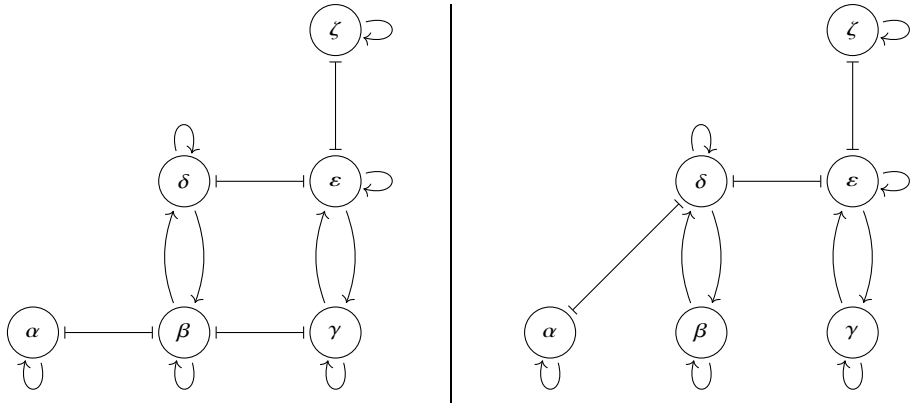

Figure 8. These cDGs on nodes $\{\alpha,\beta,\gamma,\delta,\varepsilon,\zeta\}$ have the same maximal virtual collider tripaths, however, they disagree on whether ζ is μ-separated from α by $\{\beta,\gamma,\delta,\varepsilon\}$.

Theorem 41. *Let $\mathcal{D}_1 = (V, E_1)$ and $\mathcal{D}_2 = (V, E_2)$ be cDGs containing every loop. If they are Markov equivalent, then they have the same directed edges and the same maximal virtual collider tripaths.*

The example in Figure 8 shows that having the same directed edges and the same maximal virtual collider tripaths is not a sufficient condition for Markov equivalence.

5.3. Complexity of deciding Markov equivalence

We have given two characterizations of Markov equivalence of cDGs and argued that they both use exponentially many conditions in the worst case. In this section, we prove that this, most likely, cannot be circumvented.

coNP is the class of decision problems for which a no-instance can be verified using a polynomial-length counterexample in polynomial time and a problem is in coNP if and only if its complement is in NP. If a problem is as hard as any problem in coNP, then we say that the problem is coNP-*hard*. If a problem is coNP-hard and also in coNP, we say that it is coNP-*complete* [23,54]. Various inference problems in graphical models are known to be computationally hard [7,8,29,35]. On the other hand, there exist polynomial-time algorithms for deciding Markov equivalence in several classes of graphs, e.g., maximal ancestral graphs [4] and DGs under d-separation [48]. This is different in cDGs under μ-separation.

Theorem 42. *Deciding Markov equivalence of cDGs is coNP-complete.*

The complexity result implies that, unless P = coNP (which is commonly believed to not be the case), one cannot find a characterization of Markov equivalence of cDGs which allows us to decide equivalence of two cDGs in polynomial time as a function of the size of the graphs.

6. Conclusion

We have studied graphs that represent independence structures in stochastic processes that are driven by correlated noise processes. We have characterized their equivalence classes in two ways and proven

that deciding equivalence is coNP-complete. The characterizations of Markov equivalence do, however, suggest subclasses of cDGs in which deciding Markov equivalence is feasible, e.g., in cDGs with blunt components of bounded size, or in cDGs with other restrictions on blunt paths (Mogensen [36, Chapter 4] provides an example).

We have also shown a global Markov property in the case of Ornstein-Uhlenbeck processes driven by correlated Brownian motions. It is an open question if and how this can be extended to other or larger classes of continuous-time stochastic processes.

The definition of a canonical local independence graph which was used to prove the global Markov property resembles that of a causal graph. In time series models (discrete time) a causal semantics can be defined rigorously [21] and by extending the material in the present paper one may possibly provide a causal semantics in the continuous-time framework.

Acknowledgments

We thank two anonymous referees for their comments and suggestions.

Funding

This work was supported by VILLUM FONDEN (research grant 13358).

Supplementary Material

Proofs and an augmentation criterion (DOI: 10.3150/21-BEJ1446SUPP; .pdf). The supplementary material contains proofs that were omitted from the main text. It also contains an augmentation criterion for deciding μ-separation in cDGs.

References

[1] Aalen, O.O. (1987). Dynamic modelling and causality. *Scand. Actuar. J.* **3-4** 177–190. MR0943579 https://doi.org/10.1016/j.rser.2011.04.029

[2] Aalen, O.O. and Gjessing, H.K. (2004). Survival models based on the Ornstein–Uhlenbeck process. *Lifetime Data Anal.* **10** 407–423. MR2125423 https://doi.org/10.1007/s10985-004-4775-9

[3] Aalen, O.O., Røysland, K., Gran, J.M. and Ledergerber, B. (2012). Causality, mediation and time: A dynamic viewpoint. *J. Roy. Statist. Soc. Ser. A* **175** 831–861. MR2993496 https://doi.org/10.1111/j.1467-985X.2011.01030.x

[4] Ali, R.A., Richardson, T.S. and Spirtes, P. (2009). Markov equivalence for ancestral graphs. *Ann. Statist.* **37** 2808–2837. MR2541448 https://doi.org/10.1214/08-AOS626

[5] Bartoszek, K., Glémin, S., Kaj, I. and Lascoux, M. (2017). Using the Ornstein–Uhlenbeck process to model the evolution of interacting populations. *J. Theoret. Biol.* **429** 35–45. MR3672925 https://doi.org/10.1016/j.jtbi.2017.06.011

[6] Bormetti, G., Cazzola, V. and Delpini, D. (2010). Option pricing under Ornstein-Uhlenbeck stochastic volatility: A linear model. *Int. J. Theor. Appl. Finance* **13** 1047–1063. MR2738761 https://doi.org/10.1142/S0219024910006108

[7] Chandrasekaran, V., Srebro, N. and Harsha, P. (2008). Complexity of inference in graphical models. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*.

[8] Chickering, D.M., Heckerman, D. and Meek, C. (2004). Large-sample learning of Bayesian networks is NP-hard. *J. Mach. Learn. Res.* **5** 1287–1330. MR2248018

[9] Commenges, D. and Gégout-Petit, A. (2009). A general dynamical statistical model with causal interpretation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 719–736. MR2749916 https://doi.org/10.1111/j.1467-9868.2009.00703.x

[10] Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2009). *Introduction to Algorithms*, 3rd ed. Cambridge, MA: MIT Press. MR2572804

[11] Cox, D.R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statist. Sci.* **8** 204–218. MR1243593

[12] Danks, D. and Plis, S. (2013). Learning causal structure from undersampled time series. In *JMLR: Workshop and Conference Proceedings* **10** 1–10.

[13] Didelez, V. (2000). Graphical Models for Event History Analysis based on Local Independence Ph.D. thesis, Universität Dortmund.

[14] Didelez, V. (2006). Graphical models for composable finite Markov processes. *Scand. J. Stat.* **34** 169–185. MR2325249 https://doi.org/10.1111/j.1467-9469.2006.00528.x

[15] Didelez, V. (2008). Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 245–264. MR2412641 https://doi.org/10.1111/j.1467-9868.2007.00634.x

[16] Ditlevsen, S. and Lansky, P. (2005). Estimation of the input parameters in the Ornstein-Uhlenbeck neuronal model. *Phys. Rev. E (3)* **71** 011907. MR2139313 https://doi.org/10.1103/PhysRevE.71.011907

[17] Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econometrics* **137** 334–353. MR2354948 https://doi.org/10.1016/j.jeconom.2005.06.032

[18] Eichler, M. (2012). Graphical modelling of multivariate time series. *Probab. Theory Related Fields* **153** 233–268. MR2925574 https://doi.org/10.1007/s00440-011-0345-8

[19] Eichler, M. (2012). Causal inference in time series analysis. In *Causality: Statistical Perspectives and Applications* (C. Berzuini, P. Dawid and L. Bernardinelli, eds.) 327–354. New York: Wiley.

[20] Eichler, M. (2013). Causal inference with multiple time series: Principles and problems. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **371** 20110613. MR3081173 https://doi.org/10.1098/rsta.2011.0613

[21] Eichler, M. and Didelez, V. (2007). Causal reasoning in graphical time series models. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*.

[22] Eichler, M. and Didelez, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* **16** 3–32. MR2575937 https://doi.org/10.1007/s10985-009-9143-3

[23] Garey, M.R. and Johnson, D.S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. *A Series of Books in the Mathematical Sciences*. San Francisco, CA: W. H. Freeman and Co. MR0519066

[24] Granger, C.W.J. and Newbold, P. (1986). *Forecasting Economic Time Series*, 2nd ed. San Diego: Academic Press.

[25] Heath, R.A. (2000). The Ornstein-Uhlenbeck model for decision time in cognitive tasks: An example of control of nonlinear network dynamics. *Psychol. Res.* **63** 183–191.

[26] Hopcroft, J. and Tarjan, R. (1973). Algorithm 447: Efficient algorithms for graph manipulation. *Commun. ACM* **16** 372–378.

[27] Hyttinen, A., Plis, S., Järvisalo, M., Eberhardt, F. and Danks, D. (2016). Causal discovery from subsampled time series data by constraint optimization. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models (PGM)*.

[28] Jacobsen, M. (1993). A brief account of the theory of homogeneous Gaussian diffusions in finite dimensions. In *Frontiers in Pure and Applied Probability* (H. Niemi et al., eds.) **1** 86–94.

[29] Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR2778120

[30] Koster, J.T.A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scand. J. Stat.* **26** 413–431. MR1712043 https://doi.org/10.1111/1467-9469.00157

[31] Lauritzen, S. (1996). *Graphical Models*. Oxford: Clarendon Press.

[32] Lee, M.-L.T. and Whitmore, G.A. (2006). Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. *Statist. Sci.* **21** 501–513. MR2380714 https://doi.org/10.1214/088342306000000330

[33] Liptser, R.S. and Shiryayev, A.N. (1977). *Statistics of Random Processes. I. General Theory*. *Applications of Mathematics, Vol. 5*. New York: Springer. MR0474486
[34] Maathuis, M., Drton, M., Lauritzen, S. and Wainwright, M., eds. (2019). *Handbook of Graphical Models*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. Boca Raton, FL: CRC Press. MR3889064
[35] Meek, C. (2001). Finding a path is harder than finding a tree. *J. Artificial Intelligence Res.* **15** 383–389. MR1884083 https://doi.org/10.1613/jair.914
[36] Mogensen, S.W. (2020). Graphical modeling in dynamical systems Ph.D. thesis, Univ. Copenhagen.
[37] Mogensen, S.W. and Hansen, N.R. (2020). Markov equivalence of marginalized local independence graphs. *Ann. Statist.* **48** 539–559. MR4065173 https://doi.org/10.1214/19-AOS1821
[38] Mogensen, S.W. and Hansen, N.R. Supplement to "Graphical modeling of stochastic processes driven by correlated noise." https://doi.org/10.3150/21-BEJ1446SUPP
[39] Mogensen, S.W., Malinsky, D. and Hansen, N.R. (2018). Causal learning for partially observed stochastic dynamical systems. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[40] Nodelman, U., Shelton, C.R. and Koller, D. (2002). Continuous time Bayesian networks. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[41] Nodelman, U., Shelton, C.R. and Koller, D. (2003). Learning continuous time Bayesian networks. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[42] Pavliotis, G.A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. *Texts in Applied Mathematics* **60**. New York: Springer. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7
[43] Ricciardi, L.M. and Sacerdote, L. (1979). The Ornstein-Uhlenbeck process as a model for neuronal activity. *Biol. Cybernet.* **35** 1–9.
[44] Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.* **30** 145–157. MR1963898 https://doi.org/10.1111/1467-9469.00323
[45] Richardson, T. and Spirtes, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015
[46] Richardson, T.S. (1996). A discovery algorithm for directed cyclic graphs. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[47] Richardson, T.S. (1996). A polynomial-time algorithm for deciding Markov equivalence of directed cyclic graphical models. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[48] Richardson, T.S. (1997). A characterization of Markov equivalence for directed cyclic graphs. *Internat. J. Approx. Reason.* **17** 107–162.
[49] Rogers, L.C.G. and Williams, D. (2000). *Diffusions, Markov Processes, and Martingales. Vol. 2. Itô Calculus*. *Cambridge Mathematical Library*. Cambridge: Cambridge Univ. Press. MR1780932 https://doi.org/10.1017/CBO9781107590120
[50] Røysland, K. (2012). Counterfactual analyses with graphical models based on local independence. *Ann. Statist.* **40** 2162–2194. MR3059080 https://doi.org/10.1214/12-AOS1031
[51] Schöbel, R. and Zhu, J. (1999). Stochastic volatility with an Ornstein-Uhlenbeck process: An extension. *Eur. Finance Rev.* **3** 23–46.
[52] Schweder, T. (1970). Composable Markov processes. *J. Appl. Probab.* **7** 400–410. MR0264755 https://doi.org/10.2307/3211973
[53] Shimokawa, T., Pakdaman, K., Takahata, T., Tanabe, S. and Sato, S. (2000). A first-passage-time analysis of the periodically forced noisy leaky integrate-and-fire model. *Biol. Cybernet.* **83** 327–340.
[54] Sipser, M. (2013). *Introduction to the Theory of Computation*, 3rd ed. Boston: Thomson Course Technology.
[55] Sokol, A. and Hansen, N.R. (2014). Causal interpretation of stochastic differential equations. *Electron. J. Probab.* **19** 1–24. MR3275852 https://doi.org/10.1214/ejp.v19-2891
[56] Sonntag, D. and Peña, J.M. (2015). Chain graph interpretations and their relations revisited. *Internat. J. Approx. Reason.* **58** 39–56. MR3311407 https://doi.org/10.1016/j.ijar.2014.12.001
[57] Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI)*.
[58] Stein, E.M. and Stein, J.C. (1991). Stock price distributions with stochastic volatility: An analytic approach. *Rev. Financ. Stud.* **4** 727–752.

[59] Varando, G. and Hansen, N.R. (2020). Graphical continuous Lyapunov models. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[60] Verma, T. and Pearl, J. (1991). Equivalence and synthesis of causal models Technical Report R-150, Univ. California, Los Angeles.
[61] Wright, S. (1921). Correlation and causation. *J. Agric. Res.* **20** 557–585.
[62] Wright, S. (1934). The method of path coefficients. *Ann. Math. Stat.* **5** 161–215.

Received September 2020 and revised November 2021

Supplement to "Graphical modeling of stochastic processes driven by correlated noise"

Søren Wengel Mogensen[1,2,3,⋆] and Niels Richard Hansen[1,†]

[1]Department of Mathematical Sciences, University of Copenhagen
[2]Section for Cognitive Systems, Technical University of Denmark
[3]Current affiliation: Department of Automatic Control, Lund University

[⋆]soren.wengel_mogensen@control.lth.se, [†]niels.r.hansen@math.ku.dk

November 13, 2021

This document contains proofs of results in the paper "Graphical modeling of stochastic processes driven by correlated noise" as well as an augmentation criterion for deciding μ-separation in cDGs. The concept of m-separation is found in Definition S2.

All cross-references to objects in this document are prepended with an 'S' and all other cross-references are to objects in the main text.

S1 Augmentation criterion for μ-separation in cDGs

In this section, we argue that we can decide μ-separation in a cDG by considering an *augmented graph*, a generalization of a so-called *moral graph* [S4]. Richardson and Spirtes [S18], Richardson [S17] use a similar approach to decide m-separation in *ancestral graphs* and *acyclic directed mixed graphs* (m-separation is defined in the supplementary material [S13]). Didelez [S5] uses moral graphs to decide δ-separation in DGs. The augmentation criterion has also been established in DMGs equipped with μ-separation [S14] and the analogous result in cDGs follows from almost identical arguments.

An *undirected graph* is a graph such that every edge is *undirected*, $\alpha - \beta$. The augmented graph of a cDG on nodes V is the undirected graph on nodes V, $\mathcal{U} = (V, E_{\mathcal{U}})$, such that α and β are adjacent in $\mathcal{U}$ if and only if α and β are collider connected in the cDG (omitting loops). Given an undirected graph and three disjoint subsets of nodes A, B, and C, we say that A and B are *separated* by C if every path between $\alpha \in A$ and $\beta \in B$ intersects C.

Proposition S1 (Augmentation criterion for μ-separation)**.** *Let $\mathcal{D} = (V, E)$ be a cDG. Let $A, B, C \subseteq V$, and assume that $B = \{\beta_1, \dots, \beta_j\}$. Let $B^p = \{\beta_1^p, \dots, \beta_j^p\}$ and define the graph $\mathcal{D}(B)$ with node set $V \dot\cup B^p$ such that $\mathcal{D}(B)_V = \mathcal{D}$ and*

$$\alpha \to_{\mathcal{D}(B)} \beta_i^p \text{ if } \alpha \to_{\mathcal{D}} \beta_i \text{ and } \alpha \in V, \beta_i \in B.$$

We have that $A \perp_\mu B \mid C\ [\mathcal{D}]$ *if and only if* $A \setminus C$ *and* B^p *are separated by* C *in the augmented graph of* $\mathcal{D}(B)_{an(A \cup B^p \cup C)}$.

Proof. The proofs of Propositions D.2 and D.4 by Mogensen and Hansen [S14] give the result. One can show that for $A, B, C \subseteq V$, $A \perp_\mu B \mid C\ [\mathcal{D}]$ holds if and only if $A \setminus C \perp_m B^p \mid C\ [\mathcal{D}(B)]$ holds (m-separation is found in Definition S2 and makes no distinction between blunt and bidirected edges). The second statement is then shown to be equivalent to separation in the relevant augmented graph using Theorem 1 in [S17]. Richardson [S17] studies acyclic graphs, however, the proof also applies to cyclic graphs as noted in the paper. □

S2 Proof of Theorem 10

We assume X is a regular Ornstein-Uhlenbeck process (see Example 9) with drift

$$\lambda(x) = M(x - \mu)$$

and diffusion matrix σ and let $\Sigma = \sigma\sigma^T$. We let $a = -M\mu$ and $V = U \dot\cup W$. We will use the following notation which is similar to that of Liptser and Shiryayev [S12],

$$s \circ s = \sigma_{UU}\sigma_{UU}^T + \sigma_{UW}\sigma_{UW}^T \tag{S1}$$

$$s \circ S = \sigma_{UU}\sigma_{WU}^T + \sigma_{UW}\sigma_{WW}^T \tag{S2}$$

$$S \circ S = \sigma_{WU}\sigma_{WU}^T + \sigma_{WW}\sigma_{WW}^T, \tag{S3}$$

where A_{UW}^T denotes $(A_{UW})^T$ for a matrix A and row and column indices U and W. Note that the above matrices are simply the block components of $\Sigma = \sigma\sigma^T$,

$$\Sigma = \begin{bmatrix} \sigma_{UU} & \sigma_{UW} \\ \sigma_{WU} & \sigma_{WW} \end{bmatrix} \begin{bmatrix} \sigma_{UU}^T & \sigma_{WU}^T \\ \sigma_{UW}^T & \sigma_{WW}^T \end{bmatrix} = \begin{bmatrix} s \circ s & s \circ S \\ (s \circ S)^T & S \circ S \end{bmatrix}. \tag{S4}$$

Let

$$m_t = E\left(X_t^U \mid \mathcal{F}_t^W\right).$$

The following integral equation holds [S12, Theorem 10.3],

$$m_t = m_0 + \int_0^t a_U + M_{UU} m_r + M_{UW} X_r^W \, \mathrm{d}r \tag{S5}$$

$$+ \int_0^t (s \circ S + \gamma_r M_{WU}^T)(S \circ S)^{-1}(\, \mathrm{d}X_r^W - (a_W + M_{WU} m_r + M_{WW} X_r^W)\, \mathrm{d}r) \tag{S6}$$

where $m_0 = \mathrm{E}\big[X_0^U \mid \mathcal{F}_0^W\big]$ and γ_t is the solution of a differential equation given below. We can write this as

$$m_t = m_0 + \int_0^t a_U + (M_{UU} - (s \circ S + \gamma_r M_{WU}^T)(S \circ S)^{-1} M_{WU}) m_r + M_{UW} X_r^W \, \mathrm{d}r + \int_0^t (s \circ S + \gamma_r M_{WU}^T)(S \circ S)^{-1} (\, \mathrm{d}X_r^W - (a_W + M_{WW} X_r^W) \, \mathrm{d}r).$$

The process γ_t is given by the following equation [S12, Theorem 10.3],

$$\dot{\gamma}(t) = M_{UU}\gamma_t + \gamma_t M_{UU}^T + s \circ s \tag{S7}$$

$$- \left(s \circ S + \gamma_t M_{WU}^T\right)[S \circ S]^{-1} \left(s \circ S + \gamma_t M_{WU}^T\right)^T \tag{S8}$$

$$= (M_{UU} - (s \circ S)[S \circ S]^{-1} M_{WU})\gamma_t + \gamma_t (M_{UU}^T - M_{WU}^T [S \circ S]^{-1} (s \circ S)^T) \tag{S9}$$

$$+ s \circ s - (s \circ S)[S \circ S]^{-1} (s \circ S)^T - \gamma_t M_{WU}^T [S \circ S]^{-1} M_{WU} \gamma_t, \tag{S10}$$

with initial condition $\gamma_0 = \mathrm{E}[(X_0^U - m_0)(X_0^U - m_0)^T]$. This is known as a *differential Riccati equation*. The solution of these equations is unique when we restrict our attention to solutions such that γ_t is symmetric and positive semidefinite [S12, Theorem 10.3]. Essentially, we will show the global Markov property by arguing about the measurability of m_t, using the sparsity of the matrices that go into the integral equation. We will achieve this by first describing the sparsity in the solution of an associated *algebraic* Riccati equation and this will allow us to describe the sparsity in the solution of the differential Riccati equation.

For ease of notation, we define matrices

$$D = M_{UU}^T - M_{WU}^T [S \circ S]^{-1} (s \circ S)^T, \tag{S11}$$

$$E = M_{WU}^T [S \circ S]^{-1} M_{WU}, \tag{S12}$$

$$F = s \circ s - (s \circ S)[S \circ S]^{-1} (s \circ S)^T, \tag{S13}$$

and this allows us to write the equation as

$$\dot{\gamma}(t) = \gamma_t D + D^T \gamma_t - \gamma_t E \gamma_t + F.$$

Note that F is the Schur complement of $S \circ S$ in Σ. The matrix Σ is positive definite by assumption, and therefore so are F [S7, p. 472] and $S \circ S$.

Proof of Theorem 10. Let $\beta \in B$ and let $t \in I$. We need to show that

$$E\left(\lambda_t^\beta \mid \mathcal{F}_t^{A \cup C}\right)$$

is almost surely equal to an $\mathcal{F}_t^C$-measurable random variable. We can without loss of generality assume that A and C are disjoint. The fact that B is μ-separated from A given C implies that $M_{\beta A} = 0$,

$$E\left(\lambda_t^\beta \mid \mathcal{F}_t^{A\cup C}\right) = -M_{\beta V}\mu + \sum_{\gamma\in A\cup C} M_{\beta\gamma}X_t^\gamma + \sum_{\delta\notin A\cup C} M_{\beta\delta}E\left(X_t^\delta \mid \mathcal{F}_t^{A\cup C}\right)$$
$$= -M_{\beta V}\mu + \sum_{\gamma\in C} M_{\beta\gamma}X_t^\gamma + \sum_{\delta\in \mathrm{pa}_{\mathcal{D}}(\beta)\setminus(A\cup C)} M_{\beta\delta}E\left(X_t^\delta \mid \mathcal{F}_t^{A\cup C}\right)$$

where $\mathcal{D}$ is the canonical local independence graph. Let $U = V \setminus A\cup C$. Consider now the partition of V given in Proposition S3. We see that $\mathrm{pa}_{\mathcal{D}}(\beta)\setminus(A\cup C) \subseteq V_1$. The matrix $M_{UU} - (s\circ S + \gamma_t M_{WU}^T)(S\circ S)^{-1}M_{WU}$ in the integral equation for the conditional expectation process has the sparsity of D^T (i.e., if $(D^T)_{ij} = 0$ then the ij-entry of that matrix is also zero, see Proposition S3, Corollary S4, and Lemma S15) and it follows that one can solve for $m_t^{V_1}$ independently of $m_t^{U\setminus V_1}$ as the solution of the smaller system is unique and continuous [S12, S1]. We see that processes X_t^A do not enter into these equations. This follows from the sparsity of $s\circ S$, $S\circ S$, and of $\gamma_t M_{WU}^T$, and the fact that $M_{V_4A} = 0$ and $M_{V_1A} = 0$, noting that $A\cap V_4 = \varnothing$. □

S2.1 Sparsity of the solution of the algebraic Riccati equation

In order to solve the differential Riccati equation, we will first solve an algebraic Riccati equation (Equation (S14)) — or rather argue that its solution has a certain sparsity structure.

$$0 = \Gamma D + D^T\Gamma - \Gamma E\Gamma + F \tag{S14}$$

The concept of μ-separation is similar to that of m-separation [S20, S9, S17] which has been used in acyclic graphs.

Definition S2 (m-separation)**.** *In a graph and for disjoint node sets A, B, and C, we say that A and B are* m-separated *given C (and write $A \perp_m B \mid C$) if there is no path between any $\alpha\in A$ and any $\beta\in B$ such that every collider is in $an(C)$ and no noncollider is in C.*

A path is said to be *m-connecting* given C if every collider is in $\mathrm{an}(C)$ and no noncollider is in C. m-connecting walks are defined analogously. m-separation is, in contrast to μ-separation, a symmetric notion of separation in the sense that if B is m-separated from A given C, then A is also m-separated from B given C. We will use m-separation as a technical tool in our study of cDGs as some statements are more easily expressed using this symmetric notion.

In the following proposition and its proof, we write $A \nrightarrow B \mid C$ if there exists $\alpha\in A$ and $\beta\in B$ such that there is walk between α and β with every collider in $\mathrm{an}(C)$ and no noncollider in C and furthermore there is a neck on the final edge at β.

Proposition S3. *Consider a regular Ornstein-Uhlenbeck process. Assume* $V = U \dot\cup W$ *and* $W = A \dot\cup C$ *and define*

$$
\begin{aligned}
V_1 &= \{u \in U : u \perp_m A \mid C\},\\
V_2 &= \{u \in U : u \perp_m V_1 \mid A \cup C,\ u \not\perp_m A \mid C\},\\
V_3 &= \{u \in U : u \not\perp_m V_1 \mid A \cup C,\ u \not\perp_m A \mid C\},\\
V_4 &= \{w \in W : V_1 \rightharpoonup w \mid W\},\\
V_5 &= \{w \in W : V_2 \rightharpoonup w \mid W\},\\
V_6 &= W \setminus (V_4 \cup V_5).
\end{aligned}
$$

If B *is* μ*-separated from* A *given* C *in the canonical local independence graph,* $\mathcal{D}$*, then* $U = V_1 \dot\cup V_2 \dot\cup V_3, W = V_4 \dot\cup V_5 \dot\cup V_6$*,* $pa_{\mathcal{D}}(B) \setminus (A \cup C) \subseteq V_1$*, and furthermore after a reordering of the rows and columns such that the order is consistent with* $V_1, \dots, V_6$*, we have the following sparsity of the matrices* M *and* Σ*,*

$$
M = \begin{bmatrix}
M_{11} & 0 & 0 & M_{14} & M_{15} & M_{16}\\
0 & M_{22} & 0 & M_{24} & M_{25} & M_{26}\\
M_{31} & M_{32} & M_{33} & M_{34} & M_{35} & M_{36}\\
M_{41} & 0 & 0 & M_{44} & M_{45} & M_{46}\\
0 & M_{52} & 0 & M_{54} & M_{55} & M_{56}\\
0 & 0 & 0 & M_{64} & M_{65} & M_{66}
\end{bmatrix},
$$

$$
\Sigma = \sigma\sigma^T = \begin{bmatrix}
\Sigma_{11} & 0 & \Sigma_{13} & \Sigma_{14} & 0 & 0\\
0 & \Sigma_{22} & \Sigma_{23} & 0 & \Sigma_{25} & 0\\
\Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} & \Sigma_{35} & \Sigma_{36}\\
\Sigma_{41} & 0 & \Sigma_{43} & \Sigma_{44} & 0 & 0\\
0 & \Sigma_{52} & \Sigma_{53} & 0 & \Sigma_{55} & 0\\
0 & 0 & \Sigma_{63} & 0 & 0 & \Sigma_{66}
\end{bmatrix}.
$$

For both matrices, the subscript ij corresponds to rows V_i and columns V_j.

Proof. We have that $U = V_1 \dot\cup V_2 \dot\cup V_3$. If $w \in V_4 \cap V_5 \neq \varnothing$, then there is an m-connecting walk between V_1 and V_2 given $A \cup C$ which would be a contradiction, and thus, $W = V_4 \dot\cup V_5 \dot\cup V_6$. Note that Σ is symmetric so we only need to argue that the lower triangular part has the postulated sparsity pattern. Whenever we mention an m-connecting walk in this proof without specifying a conditioning set we tacitly mean 'given W'.

Any edge $V_1 \sim V_2$ would create an m-connecting walk and therefore $M_{21} = 0, M_{12} = 0, \Sigma_{21} = 0$. An edge $V_1 \to w \in V_5$ would also create an m-connecting walk between V_1 and V_2 as $V_5 \subseteq W$, and therefore $M_{51} = 0$. Similarly, we see that $M_{42} = 0$, $\Sigma_{51} = 0$, and $\Sigma_{42} = 0$. If $V_1 \to w \in V_6$, then w would have to be in V_4, and thus, $M_{61} = 0$. Similarly, $M_{62} = 0$, $\Sigma_{61} = 0$, $\Sigma_{62} = 0$. Let $u \in V_3$. Then there exists an m-connecting walk between u and A given C, and composing this walk with an edge $u \to V_1$ would give an m-connecting walk between A and V_1

given C as $u \notin C$. This is a contradiction and $M_{13} = 0$. Similarly, $M_{23} = 0$, using the m-connecting walk between u and V_1. Consider again $u \in V_3$. There exists m-connecting walks between u and V_1 (given $A \cup C$) and u and A (given C). None of them can have a tail at u as otherwise we could find an m-connecting walk between A and V_1 given C. Therefore, u is a collider on their composition, and from this it follows that $M_{43} = 0, M_{53} = 0, M_{63} = 0$. If $V_4 \mapsto V_5$, it would follow that there is an m-connecting walk between V_1 and V_2, a contradiction. It follows that $\Sigma_{54} = 0$. If $V_4 \mapsto w \in W$, then $w \in V_4$, and it follows that $\Sigma_{64} = 0$. Similarly, $\Sigma_{65} = 0$. □

The matrices D, E, and F all have their rows and columns indexed by $U = V_1 \dot\cup V_2 \dot\cup V_3$. The above proposition and the definitions of the matrices D, E, and F give the following.

Corollary S4. *Under the conditions of Theorem 10, the matrix D has the sparsity structure*

$$\begin{bmatrix} * & 0 & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix},$$

i.e., $D_{V_2V_1} = 0, D_{V_3V_1} = 0, D_{V_1V_2} = 0$, and $D_{V_3V_2} = 0$. The matrix F is such that $F_{V_1V_2} = 0$ and $F_{V_2V_1} = 0$. The matrix E is block diagonal and $E_{V_3V_3} = 0$.

Lemma S5. *If N is an invertible matrix with the sparsity of D, then so is N^{-1}.*

Proof. The matrices on the (block) diagonal of N must also be invertible, and the result follows from the Schur complement representation of N^{-1}, using the first two blocks as one component, and the third as the second component. □

Lemma S6. *Consider the Lyapunov equation for square matrices L, Z, and Q such that Q is symmetric,*

$$LZ + ZL^T + Q = 0,$$

and let Z_0 denote its solution. If L is stable and has the sparsity pattern of D^T and Q is such that $Q_{V_1V_2} = 0$, $Q_{V_2V_1} = 0$, then $(Z_0)_{V_1V_2} = 0$ and $(Z_0)_{V_2V_1} = 0$.

Proof. The result follows from the explicit solution of a Lyapunov equation when L is stable [S11],

$$Z_0 = \int_0^\infty e^{Ls} Q e^{L^T s} \, \mathrm{d}s.$$

□

Definition S7 (Stabilizable pair of matrices)**.** *Let G and H be matrices, $n \times n$ and $n \times m$, respectively. We say that the pair (G, H) is* stabilizable *if there exists an $m \times n$ matrix, K, such that $G + HK$ is stable.*

In the literature, stabilizability is used in the context of both continuous-time and discrete-time systems. The above definition is that of a continuous-time system [S11, p. 90]. The following is proven in [S8].

Lemma S8. *The pair (A, B) is stabilizable if and only if for every eigenvector, v, of the matrix A^T with eigenvalue λ such that $Re(\lambda) \geq 0$ it holds that $v^T B \neq 0$.*

We let k and l denote the cardinalities of U and W, respectively.

Lemma S9. *The pair (D, E) is stabilizable.*

Proof. We will prove this using Lemma S8. To obtain a contradiction, assume that there exists an eigenvector v of D^T with corresponding eigenvalue λ such that $Re(\lambda) \geq 0$, and assume furthermore that $v^T E = 0$. The matrix $(S \circ S)^{-1}$ is positive definite (since Σ is positive definite), and $v^T M_{WU}^T (S \circ S)^{-1} M_{WU} v = 0$. It follows that $M_{WU} v = 0$. Let o be the column vector of zeros of length l. Note that $\lambda v = D^T v = M_{UU} v$. Then,

$$M \begin{pmatrix} v \\ o \end{pmatrix} = \begin{pmatrix} M_{UU} & M_{UW} \\ M_{WU} & M_{WW} \end{pmatrix} \begin{pmatrix} v \\ o \end{pmatrix} = \lambda \begin{pmatrix} v \\ o \end{pmatrix}.$$

It follows that λ is an eigenvalue of M which is a contradiction as M is stable by assumption. □

Corollary S10. *There exists a symmetric $k \times k$ matrix Z_0 such that $(Z_0)_{V_1 V_2} = 0$, $(Z_0)_{V_2 V_1} = 0$ and such that $D - EZ_0$ is stable.*

Proof. From the above lemma it follows that there exists a $k \times k$ matrix $\bar{Z}$ such that $D + E\bar{Z}$ is stable. From the sparsity of D and E it follows that for any $k \times k$ matrix, Z, $D + EZ$ is stable if and only if $D_{\{V1,V2\}\{V1,V2\}} + E_{\{V1,V2\}\{V1,V2\}} Z_{\{V1,V2\}\{V1,V2\}}$ and $D_{V_3 V_3}$ are stable. The matrices $D_{\{V1,V2\}\{V1,V2\}}$ and $E_{\{V1,V2\}\{V1,V2\}}$ are both block diagonal and thus both pairs of blocks are stabilizable using the existence of $\bar{Z}$ and Lemma S8. It follows that Z_0 can be chosen as block diagonal. We need to argue that Z_0 can be chosen to be symmetric. The blocks in the diagonal of E are positive semidefinite and stabilizable (when paired with their corresponding D blocks). Therefore Z_0 can be chosen to also be positive semidefinite (and therefore symmetric) and such that $D + EZ_0$ is stable [S11, Lemma 4.5.4], see also [S6]. □

Matrices E and F are both positive semidefinite and there exist unique positive semidefinite matrices $\bar{E}$ and $\bar{F}$ such that $E = \bar{E}\bar{E}$ and $F = \bar{F}\bar{F}$ [S7, Theorem 7.2.6].

Corollary S11. *The pair $(D, \bar{E})$ is stabilizable.*

Proof. This follows from the fact that (D, E) is stabilizable (Lemma S9). □

Definition S12 (Detectable pair of matrices)**.** *Let G and H be matrices, $m \times n$ and $n \times n$ respectively. We say that the pair (G, H) is* detectable *if there exists an $n \times m$ matrix, Z, such that $ZG + H$ is stable.*

Proposition S13. *The pair $(\bar{F}, D)$ is detectable. The pair (F, D) is also detectable.*

Proof. Observe that $\bar{F}$ is invertible. This means that we can choose $Z = (-I - D)\bar{F}^{-1}$. With this choice of Z, the matrix $Z\bar{F} + D$ is stable. The same argument works for the pair (F, D). □

We argue now that there is a unique positive semidefinite solution of the algebraic Riccati equation (S14) by showing that the conditions of Theorem 2 in [S10] are fulfilled. The pair $(D, \bar{E})$ is stabilizable (Corollary S11) and the pair $(\bar{F}, D)$ is detectable (Proposition S13) and we just need to show that

$$\tilde{M} = \begin{pmatrix} D & -E \\ -F & -D^T \end{pmatrix}$$

is such that $Re(\lambda) \neq 0$ for all eigenvalues, λ, of $\tilde{M}$. Assume to obtain a contradiction that λ is a eigenvalue of $\tilde{M}$ such that $Re(\lambda) = 0$,

$$\lambda v = \tilde{M} v, \qquad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Similarly to what is done in [S15], we left-multiply by $(v_2^* \;\; v_1^*)$ where $*$ denotes conjugate transpose to obtain

$$\begin{pmatrix} v_2^* & v_1^* \end{pmatrix} \begin{pmatrix} D & -E \\ -F & -D^T \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda (v_2^* v_1 + v_1^* v_2).$$

By taking real parts on both sides of the above equation, we obtain $Re(-v_2^* E v_2 - v_1^* F v_1) = 0$. Matrices E and F are both positive semidefinite so $v_2^* E v_2 = 0$ and $v_1^* F v_1 = 0$. The matrix F is positive definite so $v_1 = 0$. Lemma S8 gives a contradiction to the fact that (D, E) is stabilizable. In conclusion, it follows from Theorem 2 in [S10] that there exists a unique positive semidefinite solution of the algebraic Riccati equation.

Lemma S14 (Sparsity in solution of algebraic Riccati equation)**.** *Under the conditions of Theorem 10, it holds that $\bar{\Gamma}_{V_1 V_2} = 0$ when $\bar{\Gamma}$ is the unique, positive semidefinite solution of Equation (S14).*

Proof. Theorem 1.1 of Guo and Lancaster [S6] applies as E is positive semidefinite. We know from above that there is a unique positive semidefinite solution and this must necessarily be the same as the maximal symmetric solution of Theorem 1.1 in [S6].

Using Corollary S10, there exists a symmetric $k \times k$ matrix, Z_0, such that $(Z_0)_{V_1 V_2} = 0$, $(Z_0)_{V_2 V_1} = 0$, and such that $D - EZ_0$ is stable. From this matrix, we will define a sequence of matrices that converge to the maximal symmetric solution, Z_+. With this purpose in mind, we define a Newton step as the operation that takes a matrix Z_i to the solution of (this is an equation in Z)

$$(D - EZ_i)^T Z + Z(D - EZ_i) + Z_i E Z_i + F = 0.$$

Assume now that Z_i is such that $(Z_i)_{V_1V_2} = 0$ and $(Z_i)_{V_2V_1} = 0$. Note first that by Corollary S4, $\bar{Q} = Z_iEZ_i + F$ is also such that $\bar{Q}_{V_1V_2} = 0$ and $\bar{Q}_{V_2V_1} = 0$. The matrix EZ_i has the sparsity pattern of D, and the matrix D does too. By induction and using Lemma S6, it follows that Z_i is such that $(Z_i)_{V_1V_2} = 0$ and $(Z_i)_{V_2V_1} = 0$ for all $i \geq 0$. Note that for all i it holds that $D - EZ_i$ is stable and that Z_i is symmetric [S6]. Theorem 1.2 of Guo and Lancaster [S6] now gives that $Z_+ = \lim Z_i$ is the solution of the algebraic Riccati equation, and it follows from the above that $(Z_+)_{V_1V_2} = 0$ and $(Z_+)_{V_2V_1} = 0$. □

S2.2 Sparsity in the solution of the differential Riccati equation

We will use the above results on the algebraic Riccati equation to describe zero entries in the solution to the differential Riccati equation. We have that $(D, \bar{E})$ is stabilizable (Corollary S11) and $(\bar{F}, D)$ is detectable (Proposition S13). We note that γ_0 is the covariance in the conditional distribution of X_0^U given X_0^W and therefore positive definite. From [S19, S3], it follows that

$$\gamma_t = \bar{\Gamma} + \mathrm{e}^{tK^T}(\gamma_0 - \bar{\Gamma})\left(I + \int_0^t \mathrm{e}^{sK}E\mathrm{e}^{sK^T}\,\mathrm{d}s(\gamma_0 - \bar{\Gamma})\right)^{-1}\mathrm{e}^{tK} \tag{S15}$$

where $K = D - E\bar{\Gamma}$ and $\bar{\Gamma}$ is the unique positive semidefinite solution of the algebraic Riccati equation (Equation (S14)). Existence of the inverse matrix in Equation (S15) follows from Lemma 2.1 in Chapter 2 of Reid [S16].

Lemma S15. *Let γ_t denote the solution of the differential Riccati equation (Equation (S15)) with initial condition γ_0. Under the conditions of Theorem 10, it holds that $(\gamma_t)_{V_1V_2} = 0$ for all $t \geq 0$.*

Proof. This follows directly from the expression in Equation (S15) and the sparsity of the matrices that go into that expression: e^{tK} has the sparsity of D and e^{tK^T} has that of D^T. From Lemma S14 we know that $\bar{\Gamma}_{V_1V_2} = 0$. We see from the definition of sets V_1 and V_2 that they are m-separated by $W = A \cup C$. The sets V_1, V_2, and W are disjoint and it follows that $X_0^{V_1}$ and $X_0^{V_2}$ are conditionally independent given X_0^W using the compatibility of the distribution of X_0 with $\mathcal{D}$. Therefore, $(\gamma_0)_{V_1V_2} = 0$. The matrix

$$I + \int_0^t \mathrm{e}^{sK}E\mathrm{e}^{sK^T}\,\mathrm{d}s(\gamma_0 - \bar{\Gamma})$$

has the sparsity of D and so does its inverse (Lemma S5). This result follows immediately by matrix multiplication. □

S3 Other proofs

Proof of Proposition 20. Assume first that there is no weak inducing path between α and β in $\mathcal{D}$, and define

$$D(\alpha, \beta) = \{\gamma \in \text{an}(\alpha, \beta) \mid \gamma \text{ and } \beta \text{ are collider connected } \} \setminus \{\alpha, \beta\}.$$

We will show that β is μ-separated from α by $D(\alpha, \beta)$. We can assume that $\alpha \neq \beta$ as we have assumed that all nodes have directed loops. If there is a μ-connecting walk from α to β given $C \subseteq V \setminus \{\alpha, \beta\}$, then there is also a μ-connecting walk which is a path composed with a directed edge, $\gamma \to \beta$. We must have that $\gamma \neq \alpha$, and if $\gamma \neq \beta$ then the walk is closed by $D(\alpha, \beta)$. Assume instead that $\gamma = \beta$. Let π denote some path between α and β. Blunt and directed edges are weak inducing paths (in either direction) so π must be of length 2 or more,

$$\alpha = \gamma_0 \overset{e_0}{\sim} \gamma_1 \overset{e_1}{\sim} \dots \overset{e_{j-1}}{\sim} \gamma_j \overset{e_j}{\sim} \beta.$$

There must exist $i \in \{0, 1, \dots, j\}$, $j \geq 1$, such that either γ_i is not collider connected to β along π or $\gamma_i \notin \text{an}(\alpha, \beta)$. Let i_+ denote the largest such number in $\{0, 1, \dots, j\}$. Assume first that γ_{i_+} is not collider connected to β along π. In this case, $i_+ \neq j$. Then $\gamma_{i_+ + 1}$ is a noncollider on π and it is in $D(\alpha, \beta)$, and it follows that π is not μ-connecting. Note that necessarily $\gamma_{i_+ + 1} \neq \alpha, \beta$. On the other hand, assume $\gamma_{i_+} \notin \text{an}(\alpha, \beta)$. Then $i_+ \neq 0$, and there is some collider, γ_k, on π, $k \in \{1, ..., i_+\}$. We have that $\gamma_k \notin \text{an}(\alpha, \beta)$ and π is closed in this collider.

On the other hand, assume that there is a weak inducing path between α and β and let $C \subseteq V \setminus \{\alpha, \beta\}$. Note that $\alpha \neq \beta$. If α and β are adjacent, then $\alpha \sim \beta \to \beta$ is μ-connecting given $C \subseteq V \setminus \{\alpha, \beta\}$. Consider the weak inducing path,

$$\alpha \sim \gamma_1 \sim \dots \gamma_j \sim \beta = \gamma_{j+1}.$$

Let k be the maximal number in the set $\{1, \dots, j\}$ such that there is a walk between α and γ_k with all colliders in $\text{an}(C)$, no noncolliders in C, and which has a neck at γ_k. We see that $\gamma_1 \neq \beta$ fits this description, i.e., k is well-defined. Let ω be the walk from α to γ_k. If $\gamma_k \in \text{an}(C)$, then the composition of ω with $\gamma_k \sim \gamma_{k+1}$ gives either a new such walk (if the edge is blunt) and by maximality of k we have that $\gamma_{k+1} = \beta$, or if the edge is directed then also $\gamma_{k+1} = \beta$ (the weak inducing path is a collider path), and composing either walk with $\beta \to \beta$ gives a connecting walk given C. Assume instead that $\gamma_k \notin \text{an}(C)$, and consider again ω. There is a directed path from γ_k to α or to β. Let $\bar{\pi}$ denote the subpath from γ_k to the first instance of either α or β. If α occurs first, we compose $\bar{\pi}^{-1}$ with $\gamma_k \sim \gamma_{k+1}$ and argue as in the case of $\gamma_j \in \text{an}(C)$ above. In β occurs first, ω composed with $\bar{\pi}$ is connecting. □

Proof of Lemma 29. Assume that $\mathcal{D}_1$ and $\mathcal{D}_2$ are not collider equivalent. If $\mathcal{D}_1$ and $\mathcal{D}_2$ do not have the same directed edges, then they are not Markov equivalent (Corollary 23), and we can therefore assume that the directed edges are the same. Assume that there exist $\alpha, \beta \in V$ such that there is a collider path between α and β in $\mathcal{D}_2$,

$$\alpha \sim \bar{\gamma}_1 \sim \ldots \sim \bar{\gamma}_k \sim \beta,$$

which is not covered in $\mathcal{D}_1$ (both graphs contain every loop, so $\alpha \neq \beta$). This means that on every collider path between α and β in $\mathcal{D}_1$, there exists a collider γ such that $\gamma \notin \text{an}(\alpha, \beta)$ and $\gamma \notin \cup_j \text{an}(\bar{\gamma}_j)$. Now consider the set $D = \text{an}(\alpha, \beta) \cup [\cup_j \text{an}(\bar{\gamma}_j)] \setminus \{\alpha, \beta\}$. Note that β is not μ-separated from α given D in $\mathcal{D}_2$ as $\beta \to_{\mathcal{D}_2} \beta$, and we will argue that β is μ-separated from α given D in $\mathcal{D}_1$ showing that these graphs are not Markov equivalent. Consider a walk between α and β in $\mathcal{D}_1$. It suffices to consider a path, π, between α and β composed with the edge $\beta \to \beta$ (as $\beta \notin D$). Assume first that π is a collider path. If it is open, then every non-endpoint node is an ancestor of α, β, or $\bar{\gamma}_j$ for some j, which is a contradiction. Assume instead that there exists a noncollider (different from α and β) on the path. There must also exist a collider (otherwise π is closed), and the collider is a descendant of the noncollider. The collider is either closed, or it is an ancestor of either $\{\alpha, \beta\}$ or of $\cup_i \bar{\gamma}_i$. In the latter case, the path is closed in the noncollider. □

Proof of Proposition 30. In the original graph, $\mathcal{D}$, we add directed edges such that every node in C is a parent of α. Now the path is a weak inducing path, in this larger graph, $\mathcal{D}^+$. Using Proposition 20, we can find a μ-connecting walk from α to β given C in $\mathcal{D}^+$, and therefore a walk between α and β such that every noncollider is not in C and every collider is in $\text{an}(C)$. This walk is also in $\mathcal{D}$ as it cannot contain an edge with a tail at $\gamma \in C$. In $\mathcal{D}$, we see that every collider is still in $\text{an}(C)$ and the result follows. □

Proof of Proposition 32. Let ω be a μ-connecting walk from δ to ε given C in $\mathcal{D} + e$. If e is not on ω, then ω is also in $\mathcal{D}$ and connecting as the ancestral relations are the same in $\mathcal{D}$ and $\mathcal{D} + e$. If e is on ω, then consider the weak inducing path between α and β in $\mathcal{D}$ that consists of blunt edges only. Using a proof similar to the second part of the proof of Proposition 20 (let k in the proof of that proposition fulfil the additional assumptions that the corresponding walk in that proof has necks at both endpoints, only contains one instance of α, and does not contain any instances of β), one can show that there exists an open walk between α and β given $C \setminus \{\alpha, \beta\}$ in $\mathcal{D}$ which has necks at both ends and which only contains one instance of both α and β. This means that replacing $\alpha \mapsto \beta$ with this walk gives a μ-connecting walk given C in $\mathcal{D}$. □

Proof of Proposition 35. The graphs $\mathcal{D}$ and $\mathcal{P}_\rho(\mathcal{D})$ have the same directed edges so it suffices to show that they are collider equivalent (Theorem 31). Any permutation can be written as a composition of transpositions (such that $\alpha \notin S$ is a fixed point) so it suffices to prove the result for a permutation, ρ, such that $\rho(\alpha) = \beta$, $\rho(\beta) = \alpha$, and $\rho(\gamma) = \gamma$ for all $\gamma \neq \alpha, \beta$. Let π be a collider path in $\mathcal{D}$,

$$\gamma \sim \delta_1 \sim \ldots \sim \delta_k \sim \varepsilon.$$

If $\gamma, \varepsilon \notin \{\alpha, \beta\}$, then the path

$$\gamma \sim \rho(\delta_1) \sim \ldots \sim \rho(\delta_k) \sim \varepsilon$$

is in the permutation graph and is covering, using that α and β have the same parent set. If, e.g., $\gamma = \alpha \rightarrowtail \delta_1$ on the original path, then we can substitute this for $\alpha \to \beta \rightarrowtail \delta_1$ to obtain a covering walk in the permutation graph. Similar arguments in each case show that any collider path in $\mathcal{D}$ is covered in the permutation graph. Repeating the above argument starting from the permutation graph and using the transposition $\rho = \rho^{-1}$ shows that the two graphs are Markov equivalent. □

Proof of Theorem 36. Assume that there exists an ancestral set $A \subseteq V$ such that α and β are collider connected in $(\mathcal{D}_1)_A$, but not in $(\mathcal{D}_2)_A$. There exists a collider path in $\mathcal{D}_1$ between α and β. Any covering path in $\mathcal{D}_2$ must by definition consist of nodes in $\mathrm{an}(A) = A$ and it follows that no such path can exists. By Lemma 29, it follows that $\mathcal{D}_1$ and $\mathcal{D}_2$ are not Markov equivalent.

On the other hand, assume that for every ancestral set $A \subseteq V$ and every $\alpha, \beta \in A$, it holds that α and β are collider connected in $(\mathcal{D}_1)_A$ if and only if α and β are collider connected in $(\mathcal{D}_2)_A$. Using Theorem 31, it suffices to show that $\mathcal{D}_1$ and $\mathcal{D}_2$ are collider equivalent. Consider a collider path between α and β in $\mathcal{D}_1$, and let C denote the set of nodes on this path. This path is also a collider path in $(\mathcal{D}_1)_{\mathrm{an}(C)}$ and by assumption we can find a collider path between α and β in $(\mathcal{D}_2)_{\mathrm{an}(C)}$ as well. This collider path is in $\mathcal{D}_2$ as well and is covering the path in $\mathcal{D}_1$. □

Proof of Proposition 37. Consider an ancestral set $A \subseteq V$. We can write this as a union of strongly connected components, $A = \bigcup C_i$. These strongly connected components must necessarily constitute an ancestral set in $\mathcal{C}(\mathcal{D})$.

On the other hand, consider an ancestral set in $\mathcal{C}(\mathcal{D})$, $\mathbf{A}$, and consider $\alpha \in A = \bigcup_{C \in \mathbf{A}} C$. Assume that $\alpha \in C \in \mathbf{A}$. If β is an ancestor of α in $\mathcal{D}$, then $\beta \in \tilde{C}$ such that $\tilde{C}$ is an ancestor of C in $\mathcal{C}(\mathcal{D})$. By assumption, $\mathbf{A}$ is ancestral, so $\tilde{C} \in \mathbf{A}$ and we see that A is ancestral. □

Proof of Proposition 40. The contraposition follows from the definition of a virtual collider tripath. Assume that ω is an m-connecting path between α and β given $\mathrm{an}(\{\alpha, \beta\} \cup C) \setminus \{\alpha, \beta\}$. If it is a single edge, then (α, β, C) is a virtual collider tripath for any C. Assume that it has length at least two. If there is a noncollider, δ, on ω, then δ must be an ancestor of $\{\alpha, \beta\}$ or of a collider. In the former case, ω is closed as δ is in the conditioning set. In the latter case, either ω is closed in the collider or in δ. Assume therefore that ω is a collider path. We see from the definition that (α, β, C) is a virtual collider tripath. □

Proof of Theorem 41. We show this by contraposition. If α is a parent of β in $\mathcal{D}_1$, but not in $\mathcal{D}_2$, then it follows from Corollary 23 that they are not Markov equivalent. Assume instead that $\mathcal{D}_1$ and $\mathcal{D}_2$ have the same directed edges, and that (α, β, C) is a maximal virtual collider tripath in $\mathcal{D}_1$, but not in $\mathcal{D}_2$. Then it follows from the definition of virtual collider tripaths that $\alpha \neq \beta$. There are

two cases; either (α, β, C) is not a virtual collider tripath in $\mathcal{D}_2$, or it is not maximal. In the first case, β is μ-separated from α by $\mathrm{an}(\{\alpha, \beta\} \cup C) \setminus \{\alpha, \beta\}$ (Proposition 40) which is seen to not be the case in $\mathcal{D}_1$. In the second case, in $\mathcal{D}_2$ there is a virtual collider tripath $(\alpha, \beta, \tilde{C})$ such that $\tilde{C} \to C$ in $\bar{\mathcal{C}}(\mathcal{D}_1)$ (note that $\bar{\mathcal{C}}(\mathcal{D}_1) = \bar{\mathcal{C}}(\mathcal{D}_2)$) and $(\alpha, \beta, \tilde{C})$ is not a virtual collider tripath in $\mathcal{D}_1$. Repeating the above argument, we see that $\mathcal{D}_1$ and $\mathcal{D}_2$ are not Markov equivalent in this case either. □

Proof of Theorem 42. We first argue that deciding Markov equivalence is in coNP. This is clear as given two graphs that are not Markov equivalent and a certificate indicating sets A,B, and C such that we have separation in one but not in the other, we can use the augmentation criterion for μ-separation in the supplementary material [S13] to verify this no-instance in polynomial time.

In order to show that deciding Markov equivalence is coNP-hard, we use a reduction similar to one by Böhler et al. [S2] who study complexity of deciding equivalence of Boolean circuits, see in particular the proof of their Lemma 4.3. Consider Boolean variables $x_1, \ldots, x_n$. We say that x_l and $\neg x_l$ are *literals*. A Boolean formula is in *disjunctive normal form* (DNF) if it is a disjunction of conjuctions of literals. It is a 3DNF, if each conjunction has at most three literals. 3DNF tautology is the problem of deciding if a 3DNF is satisfied for all inputs and this problem is known to be coNP-hard [S2]. We reduce 3DNF tautology to the problem of deciding Markov equivalence. Let H be a 3DNF formula on variables $x_1, \ldots, x_n$ consisting of literals

$$H = (z_1^1 \wedge z_2^1 \wedge z_3^1) \vee \ldots \vee (z_1^N \wedge z_2^N \wedge z_3^N)$$

such that z_i^j equals x_l or $\neg x_l$ for some $l = 1, \ldots, n$. In the former case, we say that z_i^j is a *positive* literal, and in the latter that z_i^j is a *negative* literal. We say that a conjunction, e.g., $z_1^j \wedge z_2^j \wedge z_3^j$, is a *term*. In the following, we will define graphs in which the nodes correspond to literals, variables, and negated variables in this problem. We will use Greek alphabet letters for the nodes. Now define

$$V^- = \{\zeta_i^j\}_{j=1,\ldots,N, i=1,2,3} \cup \{\chi_l, \upsilon_l\}_{l=1,\ldots,n},$$

such that ζ_i^j corresponds to z_i^j, χ_l to x_l, and υ_l to the negation of x_l. We also define

$$V = \{\alpha, \beta\} \cup V^- \cup \{\gamma_\delta : \delta \in V^-\}.$$

We use $\rho_1 \leftrightarrows \rho_2$ to denote that $\rho_1 \to \rho_2$ and $\rho_1 \leftarrow \rho_2$. We use $\rho_1, \ldots, \rho_k \mapsto \rho_{k+1}, \ldots, \rho_{k+m}$ to denote that there is a blunt edge between any pair (δ_1, δ_2) such that $\delta_1 \in \{\rho_1, \ldots, \rho_k\}$ and $\delta_2 \in \{\rho_{k+1}, \ldots, \rho_{k+m}\}$. We construct a cDG on nodes V with the following edge set. Every node has a directed loop. Furthermore, for $\delta \in V^-$,

$$\alpha \to \gamma_\delta \leftrightarrows \delta.$$

For every term (analogously if the term has fewer than three literals),

$$\alpha \to \zeta_1^j \rightarrowtail \zeta_2^j \rightarrowtail \zeta_3^j \rightarrowtail \chi_1$$

and also $\zeta_3^j \rightarrowtail \upsilon_1$. Furthermore, $\chi_l, \upsilon_l \rightarrowtail \chi_{l+1}, \upsilon_{l+1}$, $l = 1, \ldots, n-1$, and $\chi_n, \upsilon_n \rightarrowtail \beta$. We also include $\chi_1 \rightarrowtail \upsilon_1$. Finally, $\chi_l \leftrightarrows \zeta_i^j$ if and only if z_i^j is a positive literal of the variable x_l and $\upsilon_l \leftrightarrows \zeta_i^j$ if and only if z_i^j is a negative literal of the variable x_l. We let $\mathcal{D}$ denote the cDG on nodes V and with edges as described above. We also define $\mathcal{D}^+$ by adding edges $\alpha \rightarrowtail \chi_1, \upsilon_1$ to $\mathcal{D}$.

We now argue that H is a tautology (that is, true for all inputs) if and only if $\mathcal{D}$ and $\mathcal{D}^+$ are Markov equivalent. Assume that H is a tautology. To argue that $\mathcal{D}$ and $\mathcal{D}^+$ are Markov equivalent it suffices to show that every collider path of $\mathcal{D}^+$ is covered in $\mathcal{D}$ (Theorem 31). Every collider path in $\mathcal{D}^+$ which is not in $\mathcal{D}$ either contains the subpath $\chi_1 \rightarrowtail \alpha \rightarrowtail \upsilon_1$ or is of the below form. If it contains $\chi_1 \rightarrowtail \alpha \rightarrowtail \upsilon_1$, then we can substitute this for $\chi_1 \rightarrowtail \upsilon_1$ and obtain a covering path in $\mathcal{D}$. Assume instead a collider path in $\mathcal{D}^+$ of the following form,

$$\alpha \rightarrowtail \varepsilon_1 \rightarrowtail \ldots \sim \varepsilon_{k+1}.$$

If $\varepsilon_{k+1} \neq \beta$, then this is covered in $\mathcal{D}$ by $\alpha \to \gamma_{\varepsilon_{k+1}} \leftrightarrows \varepsilon_{k+1}$, or by $\alpha \to \varepsilon_{k+1}$. Assume instead that $\varepsilon_{k+1} = \beta$. In this case, for all $i = 1 \ldots, n$ either $\chi_i \in \{\varepsilon_1 \ldots, \varepsilon_k\}$ or $\upsilon_i \in \{\varepsilon_1 \ldots, \varepsilon_k\}$. Consider now the following assignment of truth values to the variables: $x_l = 1$ if and only if $\chi_l \in \{\varepsilon_1 \ldots, \varepsilon_k\}$. By assumption, H is a tautology, so there is a term which equals 1 for this assignment, say the j'th (we assume without loss of generality that the j'th term contains three literals),

$$z_1^j \wedge z_2^j \wedge z_3^j.$$

If z_i^j is a positive literal, then it must correspond to a x_l such that $\chi_l \in \{\varepsilon_1 \ldots, \varepsilon_k\}$, and then in $\mathcal{D}$, ζ_i^j is a parent of $\chi_l \in \{\varepsilon_1 \ldots, \varepsilon_k\}$. If it is a negative literal, then it must correspond to x_l such that $\chi_l \notin \{\varepsilon_1 \ldots, \varepsilon_k\}$. Then $\upsilon_l \in \{\varepsilon_1 \ldots, \varepsilon_k\}$, and therefore ζ_i^j is a parent of $\{\varepsilon_1 \ldots, \varepsilon_k\}$. This means that the walk

$$\alpha \to \zeta_1^j \rightarrowtail \zeta_2^j \rightarrowtail \zeta_3^j \rightarrowtail \phi_1 \rightarrowtail \ldots \rightarrowtail \phi_n \rightarrowtail \beta,$$

where $\phi_l = \chi_l$ if $\chi_l \in \{\varepsilon_1 \ldots, \varepsilon_k\}$ and $\phi_l = \upsilon_l \in \{\varepsilon_1 \ldots, \varepsilon_k\}$ else, is a covering path in $\mathcal{D}$. This implies that $\mathcal{D}$ and $\mathcal{D}^+$ are Markov equivalent.

On the other hand, assume that H is not a tautology. In this case, there exists some assignment of truth values such that every term of H is 0, and let I denote this assignment. We now define the following subset of nodes,

$$C = \{\chi_l : x_l = 1 \text{ in } I\} \cup \{\upsilon_l : x_l = 0 \text{ in } I\}.$$

We see that for all $l = 1, \ldots, n$, either $\chi_l \in C$ or $\upsilon_l \in C$, and this means that β is not μ-separated from α by C in $\mathcal{D}^+$. If we consider a term (again, without loss of generality assuming that the term has three literals),

$$z_1^j \wedge z_2^j \wedge z_3^j.$$

We know that (under assignment I) one of them must equal 0, say z_i^j. If it is a positive literal, then the corresponding variable equals 0 in the assignment and ζ_i^j is not an ancestor of C. If it is a negative literal, then the corresponding variable x_l equals 1 in the assignment, and therefore υ_l is not in C, and ζ_i^j is not an ancestor of C. In either case, we see that every path

$$\alpha \to \zeta_1^j \mapsto \zeta_2^j \mapsto \zeta_3^j \mapsto \phi_1$$

such that $\phi_1 \in \{\chi_1, \upsilon_1\}$ contains a non-endpoint node which is not an ancestor of C. This implies that the collider path in $\mathcal{D}^+$ between α and β which traverses exactly the nodes in C is not covered in $\mathcal{D}$ and therefore $\mathcal{D}$ and $\mathcal{D}^+$ are not Markov equivalent (Theorem 31).

The reduction from 3DNF tautology to the Markov equivalence problem is clearly done in polynomial time and is a many-one reduction. □

References

[S1] Paul Beesack. Systems of multidimensional Volterra integral equations and inequalities. *Nonlinear Analysis*, 9(12):1451–1486, 1985.

[S2] Elmar Böhler, Nadia Creignou, Matthias Galota, Steffen Reith, Henning Schnoor, and Heribert Vollmer. Complexity classifications for different equivalence and audit problems for Boolean circuits. *Logical Methods in Computer Science*, 8(3:27):1–25, 2012.

[S3] Chiu H. Choi. A survey of numerical methods for solving matrix Riccati differential equations. In *IEEE Proceedings on Southeastcon*, 1990.

[S4] Robert G. Cowell, Philip Dawid, Steffen L. Lauritzen, and David J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. New York: Springer, 1999.

[S5] Vanessa Didelez. *Graphical Models for Event History Analysis based on Local Independence*. PhD thesis, Universität Dortmund, 2000.

[S6] Chun-Hua Guo and Peter Lancaster. Analysis and modification of Newton's method for algebraic Riccati equations. *Mathematics of Computation*, 67(223):1089–1105, 1998.

[S7] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[S8] Birgit Jacob and Hans J. Zwart. *Linear Port-Hamiltonian Systems on Infinite-dimensional Spaces*. Basel: Springer, 2012.

[S9] Jan T.A. Koster. On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*, 26:413–431, 1999.

[S10] Vladimír Kučera. On nonnegative definite solutions to matrix quadratic equations. *Automatica*, 8:413–423, 1972.

[S11] Peter Lancaster and Leiba Rodman. *Algebraic Riccati Equations*. Oxford: Clarendon Press, 1995.

[S12] R.S. Liptser and A.N. Shiryayev. *Statistics of Random Processes I: General Theory*. New York: Springer-Verlag, 1977.

[S13] Søren Wengel Mogensen and Niels Richard Hansen. Supplement to "Graphical modeling of stochastic processes driven by correlated noise".

[S14] Søren Wengel Mogensen and Niels Richard Hansen. Markov equivalence of marginalized local independence graphs. *The Annals of Statistics*, 48 (1):539–559, 2020.

[S15] B. P. Molinari. The stabilizing solution of the algebraic Riccati equation. *SIAM Journal on Control*, 11(2):262–271, 1973.

[S16] William T. Reid. *Riccati differential equations*. New York: Academic Press, 1972.

[S17] Thomas S. Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1):145–157, 2003.

[S18] Thomas S. Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

[S19] Ilan Rusnak. Almost analytic representation for the solution of the differential matrix Riccati equation. *IEEE Transactions on Automatic Control*, 33(2):191–193, 1988.

[S20] Peter Spirtes, Thomas Richardson, Christopher Meek, Richard Scheines, and Clark Glymour. Using path diagrams as a structural equation modeling tool. *Sociological Methods and Research*, 27(2):182–225, 1998.