



# Mixed orthogonality graphs for continuous-time stationary processes<sup>☆</sup>

Vicky Fasen-Hartmann<sup>\*</sup>, Lea Schenk

*Institute of Stochastics, Karlsruhe Institute of Technology, Germany*

## ARTICLE INFO

MSC:  
primary 62H22  
62M10  
secondary 62M20

Keywords:  
Granger causality  
Contemporaneous correlation  
Graphs  
Markov property  
MCAR processes  
Linear prediction

## ABSTRACT

In this paper, we introduce different concepts of Granger causality and contemporaneous correlation for multivariate stationary continuous-time processes to model different dependencies between the component processes. Several equivalent characterisations are given for the different definitions, in particular by orthogonal projections. We then define two mixed graphs based on different definitions of Granger causality and contemporaneous correlation, the (mixed) orthogonality graph and the local (mixed) orthogonality graph. In these graphs, the components of the process are represented by vertices, directed edges between the vertices visualise Granger causal influences and undirected edges visualise contemporaneous correlation between the component processes. Further, we introduce various notions of Markov properties in analogy to Eichler (2012), which relate paths in the graphs to different dependence structures of subprocesses, and we derive sufficient criteria for the (local) orthogonality graph to satisfy them. Finally, as an example, for the popular multivariate continuous-time AR (MCAR) processes, we explicitly characterise the edges in the (local) orthogonality graph by the model parameters.

## 1. Introduction

In this paper, we define new notions of Granger causality and contemporaneous correlation specifically for multivariate stochastic processes in continuous time and visualise them in mixed graphs. With the increasing interest in complex multivariate data sets and networks in diversified fields, the interest in graphical models develops rapidly, although the attempt to use graphical models for the visualisation and analysis of causal structures in stochastic models is quite old Wright [75,76]. The key advantage of graphical models is the simple and clear way to display the dependencies of stochastic processes. We refer to the nice overview in Maathuis et al. [54] for the state of the art on the mathematical and statistical aspects of graphical models. In our graphical models, vertices represent the different component series  $Y_v = (Y_v(t))_{t \in \mathbb{R}}$ ,  $v \in V := \{1, \dots, k\}$ , of an underlying continuous-time stochastic process  $Y_V = (Y_V(t))_{t \in \mathbb{R}}$ . The vertices are connected with directed and undirected edges, which represent Granger causalities and contemporaneous correlations, respectively.

The mathematical notion of causality was popularised by Clive W. J. Granger and Christopher A. Sims. In his original work, Granger [40] used a linear vector autoregressive (VAR) model, whereas Sims [71] used a moving average (MA) model to understand the causal effects in a bivariate model; a detailed discussion of the relationships between Granger and Sims causality

<sup>☆</sup> *Funding:* This work is supported by the project “digiMINT”, which is a part of the “Qualitätsoffensive Lehrerbildung”, a joint initiative of the Federal Government and the Länder which aims to improve the quality of teacher training. The program is funded by the Federal Ministry of Education and Research, Germany. The authors are responsible for the content of this publication.

<sup>\*</sup> Corresponding author.

*E-mail addresses:* [vicky.fasen@kit.edu](mailto:vicky.fasen@kit.edu) (V. Fasen-Hartmann), [lea.schenk@kit.edu](mailto:lea.schenk@kit.edu) (L. Schenk).

<https://doi.org/10.1016/j.spa.2024.104501>

Received 13 September 2023; Received in revised form 23 September 2024; Accepted 1 October 2024

Available online 9 October 2024

0304-4149/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

is given in Kuersteiner [48], see also Dufour and Renault [26], Eichler [30]. Since then, their ideas have been extended in various ways and have been applied in diversified fields, such as neuroscience (Bergmann and Hartwigsen [3]), econometrics (Imbens [46]), environmental science (Cox and Popken [16]), genomics (Heerah et al. [45]) and social systems (Kuzma et al. [49]). The recent publication of Shojaie and Fox [70] is an excellent review of Granger causality with its advances.

However, not every interesting relationship between two component series  $Y_a$  and  $Y_b$  is necessarily a causal relation and directed. But this does not diminish the importance of modelling such relationships. Some well-known examples are the correlation between the aggressive behaviour and the amount of time spent playing computer games each day (Lemmens et al. [51]) and the correlation between the number of infants who sleep with the light on and the number of people who develop myopia in later life (Zadnik et al. [77]). To model such undirected relationships, we use contemporaneous correlation, a symmetric relation between  $Y_a$  and  $Y_b$ .

Our novel approach is to define concepts of Granger causality and contemporaneous correlation for continuous-time multivariate processes by orthogonal projections onto linear spaces generated by subprocesses, resulting in *conditional orthogonality* relations. For processes in discrete time, this attempt was already studied in Florens and Mouchart [37], Dufour and Renault [26], Eichler [28]. In contrast to the other papers, Eichler [28] even represents the conditional orthogonality relations of a discrete-time VAR process in a graph, where Granger causality models the directed influences and contemporaneous correlation the undirected influences.

An alternative approach is to use *conditional independence* relations using conditional expectations given  $\sigma$ -fields generated by subprocesses, see Chamberlain [13], Florens and Mouchart [36], Eichler [29] for discrete-time processes and Comte and Renault [15], Florens and Fougère [35], Petrovic and Dimitrijevic [62] for continuous-time processes and especially for semimartingales. Comte and Renault [15] propose to model undirected influences by global instantaneous causality and local instantaneous causality in continuous time, however, the results are not related to graphical models. Again, Eichler [29] defines a graphical model for time series in discrete time representing the conditional independence relations using Granger causality for directed influences and contemporaneous conditional dependence for the undirected influences.

In principle, the two approaches of conditional orthogonality and of conditional independence to define causalities in continuous-time processes are both conceivable and of interest. For Gaussian random vectors, conditional independence and conditional orthogonality are even equivalent. However, in non-Gaussian time series models, conditional expectations are much more difficult to compute than linear predictions, which is the first reason for this paper to explore Granger causality concepts based on conditional orthogonality instead of conditional independence. The second reason is motivated from the background on discrete-time processes in Eichler [28,29], where the first paper derives a graphical model based on conditional orthogonality and the second paper on conditional independence relations. The proof of the Markov properties for discrete-time processes using conditional orthogonality of linear spaces of Eichler [28] requires fewer assumptions than for conditional independence in Eichler [29], which is based on non-linear spaces. There, some technical and difficult-to-verify assumptions are required in order to achieve the Markov properties. Since the aim of this paper is to propose an easy-to-use graphical model, these observations motivate our approach for using conditional orthogonality of linear spaces, in particular since linear causality concepts in continuous time have not yet been investigated. Comte and Renault [15], Florens and Fougère [35] both consider non-linear causality concepts, among others in the context of semimartingales.

A noteworthy extension of conditional independence is the concept of *local independence* for composable finite Markov processes of Schweder [69] which was generalised to semimartingales by Aalen [1]. This concept has been applied to define and analyse the *local independence graph*, e.g., in the context of composable finite Markov processes, point processes and physical systems in Didelez [21,22,23], Eichler et al. [31], Commenges and Gégout-Petit [14], Røysland et al. [66]. These definitions were recently taken up by Mogensen and Hansen [59,60] who study (canonical) local independence graphs for Itô processes. However, the results rely on the semimartingale property of such processes, but semimartingales do not seem to be the right tool for stationary time series models, especially for non-Gaussian models. Additionally, Mogensen and Hansen [60] assume continuous sample paths, which excludes Lévy-driven stochastic processes with jumps.

This paper is the first paper developing graphical models for conditional orthogonality relations of general stationary stochastic process in continuous-time. We also present several equivalent characterisations of our concepts of Granger causality and contemporaneous correlation and relate them to other definitions in the literature. These definitions do not require the stationarity of  $Y_V$ . Importantly, we define local versions of Granger causality and contemporaneous correlation, which are less strong. Based on the different definitions of Granger causality and contemporaneous correlation, we then introduce two mixed graphs, the *(mixed) orthogonality graph* and the *local (mixed) orthogonality graph* for such multivariate stochastic processes in continuous time. For example, for an Ornstein–Uhlenbeck process, the two graphs may look like in Fig. 1. We can already see from this picture that the edges of the local orthogonality graph are also edges in the orthogonality graph.



Fig. 1. In the left figure is the orthogonality graph and in the right figure the local orthogonality graph of the Ornstein–Uhlenbeck process defined in Example 3.15.

The causality structure of a graph is usually described by Markov properties. Eichler [28,29] discusses Markov properties for mixed graphical models, namely the pairwise, local, block-recursive and two global Markov properties, using  $m$ -separation (Richardson [65]) and  $p$ -separation (Levitz et al. [52]), respectively, for the global ones. For an asymmetric graph, Didelez [23] develops and investigates an asymmetric notion of separation and discusses different levels of Markov properties. In addition, Mogensen and Hansen [60] show that the multivariate Ornstein–Uhlenbeck process driven by a Brownian motion is the only process that satisfies their global Markov property. As the above literature shows, the derivation of global Markov properties might be quite challenging and often it is only valid under additional or even restrictive assumptions.

In our (local) orthogonality graph, we show the pairwise, local and block-recursive Markov property and then discuss global Markov properties in both graphs. Importantly, the orthogonality graph satisfies the global Andersson, Madigan and Perlman (AMP) Markov property (Andersson et al. [2]), which is a sufficient criterion for conditional orthogonality. The assumptions on our orthogonality graph are quite general. We only require a stationary mean-square continuous stochastic process in continuous time with expectation zero, which is purely non-deterministic, with some restriction on the spectral density, which is, e.g., satisfied for Ornstein–Uhlenbeck and, more general, for continuous-time moving average (MCAR) processes. Since the notion of  $m$ -separation in the AMP Markov property is strong, we present less restrictive alternatives and discuss the global Markov property of the orthogonality graph. Although the local orthogonality graph also satisfies the pairwise, local and block-recursive Markov properties, not surprisingly stronger assumptions are required for global Markov properties.

Finally, we derive the graphical structure of the popular multivariate continuous-time autoregressive (MCAR) processes driven by a general centred Lévy process with finite second moments, which are important extensions of their discrete-time counterparts. Different choices of the driving Lévy process and the model parameters, i.e., the parameters of the autoregressive polynomial and the covariance matrix of the driving Lévy process, allow quite flexible modelling of the margins, so MCAR processes form a broad class of processes. Special cases are the Gaussian MCAR processes, where the Brownian motion is the driving Lévy process and Ornstein–Uhlenbeck processes, which are MCAR(1) processes. For general MCAR models, we derive that the (local) orthogonality graph is well defined and we explicitly characterise the different types of edges by the model parameters. These characterisations differ for the orthogonality and local orthogonality graph. Finally, we find analogues to the edge characterisations for vector autoregressive processes in Eichler [28].

Remarkably, in the case of Gaussian MCAR processes, our characterisations of local Granger causality and local contemporaneous correlation given by the model parameters, respectively, coincide with the characterisations of local Granger causality and local instantaneous causality in the non-linear setting in Comte and Renault [15]. However, our approach has several advantages. On the one hand, their theory is developed for semimartingales and several characterisations even assume continuous sample paths. But non-Gaussian Lévy-driven MCAR models have jumps and can therefore not be covered by their theory. On the other hand, modelling the dependencies of the MCAR process in the local orthogonality graph allows to encode local Granger causalities and local contemporaneous correlations between multivariate subprocesses through the derived Markov properties. This is not content of Comte and Renault [15]. Similarly, for Gaussian Ornstein–Uhlenbeck models, the local independence graph of Mogensen and Hansen [60] coincides with our local causality graph. But their approach is based on Brownian motion driven Itô processes, again excluding Lévy driven models or MCAR( $p$ ) processes with  $p \geq 2$ . To the best of our knowledge, our paper is the first on graphical properties of Lévy-driven MCAR models. It provides a generalisation of the results known from the literature to non-Gaussian processes. In Fasen-Hartmann and Schenk [32] we even develop extensions to the more general class of multivariate state space models based on the present paper, and in Fasen-Hartmann and Schenk [33] we present an undirected graphical model and relate it to the (local) orthogonality graph.

### Structure of the paper

The paper is structured as follows. In Section 2, we first lay the foundation by introducing the conditional orthogonality relation as well as appropriate linear spaces generated by multivariate stochastic processes in continuous time and their properties which are important for this paper. We conclude the preliminaries with properties on mean-square differentiable stationary processes with expectation zero. In Sections 3 and 4, we then define, discuss, and relate different directed and undirected interactions between the component series of continuous-time stationary processes, i.e., Granger causality and contemporaneous correlation. This groundwork culminates in the definition of the orthogonality graph and the local orthogonality graph in Section 5. For these orthogonality graphs, we prove several Markov properties. Finally, in Section 6, we characterise the different graphical models for MCAR processes. The proofs of the paper are moved to the appendix.

### Notation

Throughout the paper,  $V = \{1, \dots, k\}$  and  $Y_V = (Y_V(t))_{t \in \mathbb{R}}$  denotes a  $k$ -dimensional (weakly) stationary stochastic process with expectation zero that is continuous in mean square. From now on we call the space of all real or complex  $(k \times k)$ -dimensional matrices  $M_k(\mathbb{R})$  and  $M_k(\mathbb{C})$ , respectively. Similarly,  $M_{k,d}(\mathbb{R})$  and  $M_{k,d}(\mathbb{C})$  denote real and complex  $(k \times d)$ -dimensional matrices. We write  $I_k$  for the  $k$ -dimensional identity matrix and  $0_k$  for the  $k$ -dimensional zero matrix ( $k \in \mathbb{N}$ ). With  $\|\cdot\|$  we denote some matrix norm. The vector  $e_\nu \in \mathbb{R}^k$  is the  $\nu$ -th unit vector and  $E_j^T := (0_{k \times k(j-1)}, I_k, 0_{k \times k(p-j)}) \in M_{k \times kp}(\mathbb{R})$ ,  $j = 1, \dots, p$ . For hermitian matrices  $A, B \in M_k(\mathbb{C})$ , we write  $A \geq_L B$  if and only if  $B - A$  is positive semi-definite, i.e.,  $B - A \geq 0$ . Similarly, we write  $A > 0$  if  $A$  is positive definite. Furthermore,  $\sigma(A)$  are the eigenvalues of  $A$ . Finally, by l.i.m. we denote the mean square limit.

## 2. Preliminaries

In these preliminaries, we present some basics about the conditional orthogonality relation, such as the semi-graphoid property. Furthermore, we define the important linear spaces of this paper and give properties of mean-square differentiable stationary processes with expectation zero, which we use throughout the paper. We start with some fundamentals on linear spaces in  $L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$ , the Hilbert space of square-integrable complex-valued random variables on a common probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . As usual, the inner product is  $\langle X, Y \rangle_{L^2} = \mathbb{E}[X\bar{Y}]$  for  $X, Y \in L^2$  and orthogonality with respect to this inner product is denoted by  $X \perp Y$ . We set  $\|X\|_{L^2} := \sqrt{\langle X, X \rangle_{L^2}}$  for  $X \in L^2$  and identify random variables that are equal  $\mathbb{P}$ -a.s. Note that if  $X_n \rightarrow_{L^2} X$  and  $Y \in L^2$ , then

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n Y) = \mathbb{E}(XY), \tag{2.1}$$

which can be shown by Cauchy–Schwarz inequality. Further, suppose  $\mathcal{L}_1$  and  $\mathcal{L}_2$  are closed linear subspaces of  $L^2$ , where the closure is formed in the mean square. Then

$$\mathcal{L}_1^\perp = \{X \in L^2 : \langle X, Y \rangle_{L^2} = 0 \text{ for all } Y \in \mathcal{L}_1\}$$

is the orthogonal complement of  $\mathcal{L}_1$ . The sum of  $\mathcal{L}_1$  and  $\mathcal{L}_2$  is the linear vector space

$$\mathcal{L}_1 + \mathcal{L}_2 = \{X + Y : X \in \mathcal{L}_1, Y \in \mathcal{L}_2\}.$$

Even when  $\mathcal{L}_1$  and  $\mathcal{L}_2$  are closed subspaces, this sum may fail to be closed if both are infinite-dimensional. A classic example of this can be found in Halmos [41], p. 28. Hence, the closed direct sum is denoted by

$$\mathcal{L}_1 \vee \mathcal{L}_2 = \overline{\{X + Y : X \in \mathcal{L}_1, Y \in \mathcal{L}_2\}}.$$

We further denote the orthogonal projection of  $X \in L^2$  on  $\mathcal{L}_1$  by  $P_{\mathcal{L}_1}(X) = P_{\mathcal{L}_1} X$ . A review of the properties of orthogonal projections can be found, e.g., in Weidmann [73], Brockwell and Davis [9], Lindquist and Picci [53].

### 2.1. Conditional orthogonality

With those notations in mind, we define the conditional orthogonality relation as in Eichler [28], p. 347.

**Definition 2.1.** Let  $\mathcal{L}_i, i = 1, 2, 3$ , be closed linear subspaces of  $L^2$ . Then  $\mathcal{L}_1$  and  $\mathcal{L}_2$  are *conditionally orthogonal* given  $\mathcal{L}_3$  if

$$X - P_{\mathcal{L}_3} X \perp Y - P_{\mathcal{L}_3} Y \quad \forall X \in \mathcal{L}_1, Y \in \mathcal{L}_2.$$

The conditional orthogonality relation is denoted by  $\mathcal{L}_1 \perp \mathcal{L}_2 \mid \mathcal{L}_3$ .

Moreover, we summarise properties of the conditional orthogonality relation as given in Eichler [28], Proposition A.1.

**Lemma 2.2.** Let  $\mathcal{L}_i, i = 1, \dots, 4$ , be closed linear subspaces of  $L^2$ . Then the conditional orthogonality relation defines a semi-graphoid, i.e., it satisfies the following properties:

- (C1) *Symmetry:*  $\mathcal{L}_1 \perp \mathcal{L}_2 \mid \mathcal{L}_3 \Rightarrow \mathcal{L}_2 \perp \mathcal{L}_1 \mid \mathcal{L}_3$ .
- (C2) *(De-) Composition:*  $\mathcal{L}_1 \perp \mathcal{L}_2 \mid \mathcal{L}_4$  and  $\mathcal{L}_1 \perp \mathcal{L}_3 \mid \mathcal{L}_4 \Leftrightarrow \mathcal{L}_1 \perp \mathcal{L}_2 \vee \mathcal{L}_3 \mid \mathcal{L}_4$ .
- (C3) *Weak union:*  $\mathcal{L}_1 \perp \mathcal{L}_2 \vee \mathcal{L}_3 \mid \mathcal{L}_4 \Rightarrow \mathcal{L}_1 \perp \mathcal{L}_2 \mid \mathcal{L}_3 \vee \mathcal{L}_4$ .
- (C4) *Contraction:*  $\mathcal{L}_1 \perp \mathcal{L}_2 \mid \mathcal{L}_4$  and  $\mathcal{L}_1 \perp \mathcal{L}_3 \mid \mathcal{L}_2 \vee \mathcal{L}_4 \Rightarrow \mathcal{L}_1 \perp \mathcal{L}_2 \vee \mathcal{L}_3 \mid \mathcal{L}_4$ .

If  $(\mathcal{L}_2 \vee \mathcal{L}_4) \cap (\mathcal{L}_3 \vee \mathcal{L}_4) = \mathcal{L}_4$  holds and  $\mathcal{L}_2 \vee \mathcal{L}_3$  is separable, then the conditional orthogonality relation defines a graphoid, i.e., additionally we have:

- (C5) *Intersection:*  $\mathcal{L}_1 \perp \mathcal{L}_2 \mid \mathcal{L}_3 \vee \mathcal{L}_4$  and  $\mathcal{L}_1 \perp \mathcal{L}_3 \mid \mathcal{L}_2 \vee \mathcal{L}_4 \Rightarrow \mathcal{L}_1 \perp \mathcal{L}_2 \vee \mathcal{L}_3 \mid \mathcal{L}_4$ .

Note that the definition of conditional orthogonality reduces to the usual orthogonality when  $\mathcal{L}_3 = \{0\}$ . For a more detailed discussion of the conditional orthogonality relation, we refer to Florens and Mouchart [37], who give the above results in terms of a general Hilbert space.

**Remark 2.3.** If  $(\mathcal{L}_2 \vee \mathcal{L}_4) \cap (\mathcal{L}_3 \vee \mathcal{L}_4) = \mathcal{L}_4$  holds, we say that  $\mathcal{L}_2$  are  $\mathcal{L}_3$  *conditionally linearly separated* by  $\mathcal{L}_4$  (cf. Eichler [28], p. 348).

### 2.2. Linear subspaces

To apply the concept of conditional orthogonality to a multivariate stochastic process  $Y_V$ , where  $V = \{1, \dots, k\}$ , we define suitable closed linear subspaces. Let  $A \subseteq V, s, t \in [-\infty, \infty]$  and  $s \leq t$ . Then we define the closed linear space

$$\mathcal{L}_{Y_A}(s, t) := \overline{\text{span}} \{Y_a(u) : a \in A, u \in [s, t] \cap \mathbb{R}\}$$

with  $\mathcal{L}_{Y_A}(-\infty, -\infty) := \mathcal{L}_{Y_A}(\infty, \infty) := \{0\}$  and use the shorthands

$$\mathcal{L}_{Y_A}(t) := \mathcal{L}_{Y_A}(-\infty, t), \quad \mathcal{L}_{Y_A}(-\infty) := \bigcap_{t \in \mathbb{R}} \mathcal{L}_{Y_A}(t), \quad \mathcal{L}_{Y_A} := \mathcal{L}_{Y_A}(-\infty, \infty).$$

Sometimes we use as well the linear space

$$\mathcal{L}_{Y_A}(s, t) := \text{span} \{Y_a(u) : a \in A, u \in [s, t] \cap \mathbb{R}\},$$

whose closure is  $\mathcal{L}_{Y_A}(s, t)$ . For further discussion and properties of such linear spaces, we refer to the early works of Cramér [18,19,20], but also to Rozanov [67], Lindquist and Picci [53], Brockwell and Lindner [11]. Furthermore, in Section 5.1 we derive sufficient criteria for conditional linear separation and separability of these linear spaces. The next lemma provides the basic properties of these linear spaces, which we use throughout the paper. The proof is given in the Supplementary Material D.

**Lemma 2.4.** *Let  $A, B \subseteq V$ ,  $s, t \in \mathbb{R}$ ,  $s \leq t$ . Then the following statements hold:*

- (a)  $\mathcal{L}_{Y_A}(s) \vee \mathcal{L}_{Y_A}(s, t) = \mathcal{L}_{Y_A}(t)$   $\mathbb{P}$ -a.s.
- (b)  $\mathcal{L}_{Y_A}(s, t) \vee \mathcal{L}_{Y_B}(s, t) = \mathcal{L}_{Y_{A \cup B}}(s, t)$   $\mathbb{P}$ -a.s.
- (c)  $\mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_B}(t) = \mathcal{L}_{Y_{A \cup B}}(t)$   $\mathbb{P}$ -a.s.
- (d)  $\bigcup_{n \in \mathbb{N}} \mathcal{L}_{Y_A}(n) = \mathcal{L}_{Y_A}$   $\mathbb{P}$ -a.s.

### 2.3. Mean-square differentiable stationary processes

To compute the mean-square derivative of a stationary continuous-time process  $Y_V$  with expectation zero, the following result of Gihman and Skorokhod [38], IV. §3, Corollary 2 is useful; see as well Brockwell and Lindner [11], Example 5.17 and Doob [25], XI. §9, Example 1.

**Proposition 2.5.** *Let  $Y_V$  be a stationary process with expectation zero, spectral density  $f_{Y_V Y_V}(\lambda)$ ,  $\lambda \in \mathbb{R}$ , and spectral representation*

$$Y_V(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \Phi_V(d\lambda), \quad t \in \mathbb{R}, \tag{2.2}$$

where  $\Phi_V(\lambda) = (\Phi_1(\lambda), \dots, \Phi_k(\lambda))^T$  is a random measure with

$$\mathbb{E}[\Phi_V(d\lambda)] = 0_k \in \mathbb{R}^k \quad \text{and} \quad \mathbb{E}[\Phi_V(d\lambda) \overline{\Phi_V(d\mu)}^T] = \delta_{\lambda=\mu} f_{Y_V Y_V}(\lambda) d\lambda.$$

Then

$$\text{L.i.m.}_{h \rightarrow 0} \frac{Y_V(t) - Y_V(t-h)}{h}$$

exists if and only if  $\int_{-\infty}^{\infty} \lambda^2 \|f_{Y_V Y_V}(\lambda)\| d\lambda < \infty$ . In this case,

$$D^{(1)}Y_V(t) := \text{L.i.m.}_{h \rightarrow 0} \frac{Y_V(t) - Y_V(t-h)}{h} = \int_{-\infty}^{\infty} i\lambda e^{i\lambda t} \Phi_V(d\lambda), \quad t \in \mathbb{R}.$$

Obviously, by recursion, we receive as well higher derivatives. Note that for a one-dimensional process  $Y = (Y(t))_{t \in \mathbb{R}}$ , the condition  $\int_{-\infty}^{\infty} \lambda^2 |f_{YY}(\lambda)| d\lambda < \infty$  is equivalent to the existence of  $c''_{YY}(0)$ , where  $c_{YY}(t)$ ,  $t \in \mathbb{R}$ , is the autocovariance function of  $Y$ .

**Remark 2.6.** Suppose  $Y_v$  is mean-square differentiable for some  $v \in V$ . Then

$$D^{(1)}Y_v(t) = \text{l.i.m.}_{h \searrow 0} \frac{Y_v(t) - Y_v(t-h)}{h} \in \mathcal{L}_{Y_v}(t).$$

Similarly, we are able to show by induction that if  $Y_v$  is  $j_v$ -times mean-square differentiable, then  $D^{(j_v)}Y_v(t) \in \mathcal{L}_{Y_v}(t)$ .

For more details on stationary processes, we refer to the comprehensive works of Doob [25], Rozanov [67], Lindquist and Picci [53], Brockwell and Lindner [11].

### 3. Directed influences: Granger causality for stationary continuous-time processes

In this section, we introduce and characterise directed influences between the component series of  $Y_V$  using different concepts of causality: *local Granger causality*, *Granger causality* and *global Granger causality*, where global Granger non-causality implies Granger non-causality which in turn implies local Granger non-causality. In Appendix A, we present the proofs of the present section.

The idea of a Granger causal influence of one component series  $Y_a$  on another component series  $Y_b$  goes back to Granger [40]. In discrete time, the general idea that one process  $Y_a$  is Granger non-causal for another process  $Y_b$  is based on the question of whether the prediction of  $Y_b(t+1)$  based on the information available at time  $t$  provided by the past and present values of  $Y_V$  is diminished by removing the information provided by the past and present values of  $Y_a$ . To transfer this approach to the continuous-time setting, we need to ask what it means to predict a time step into the future. As there is no obvious approach, we present the aforementioned three different concepts, motivated by other definitions of Granger causality in the literature. The first approach is the direct generalisation of Eichler [28], Definition 2.2, to continuous-time processes, considering one time step in the future.

**Definition 3.1.** Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A$  is *Granger non-causal* for  $Y_B$  with respect to  $Y_S$  if, for all  $t \in \mathbb{R}$ ,

$$\mathcal{L}_{Y_B}(t, t + 1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t).$$

We write  $Y_A \not\rightarrow Y_B \mid Y_S$ .

**Remark 3.2.** In the definition of Granger causality, we use the time step  $h = 1$  because this is also done for discrete-time processes in Eichler [28] and it is the natural choice. Of course, it is also plausible to take some step size  $h > 0$  and define that  $Y_A$  is Granger non-causal for  $Y_B$  with respect to  $Y_S$  by

$$\mathcal{L}_{Y_B}(t, t + h) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall t \in \mathbb{R}. \tag{3.1}$$

The results of this paper are straightforwardly transferable to this definition, but for ease of notation we stick to  $h = 1$ . For popular examples such as the MCAR processes, see Remark 6.20, and state space models (Fasen-Hartmann and Schenk [32]), we recognise that for different  $h$  these definitions are even equivalent. However, we believe that for non-linear processes the equivalence is in general not true anymore, see also Dufour and Renault [26], Kuersteiner [48], Eichler [30].

In the next lemma, we present some equivalent characterisations of Granger causality, for completeness the proof is given in the Supplementary Material D.

**Lemma 3.3.** Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then the following statements are equivalent:

- (a)  $Y_A \not\rightarrow Y_B \mid Y_S$
- (b)  $\mathcal{L}_{Y_B}(t + 1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall t \in \mathbb{R}$ .
- (c)  $\ell_{Y_B}(t, t + 1) \perp \ell_{Y_A}(-\infty, t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall t \in \mathbb{R}$ .
- (d)  $\ell_{Y_b}(s, s) \perp \ell_{Y_a}(s', s') \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall a \in A, b \in B, s \in [t, t + 1], s' \leq t, t \in \mathbb{R}$ .

The stationarity assumption is not necessary for the definition of Granger causality and its characterisations and can be neglected here. We first need it in Section 5, e.g., for the intersection property (C5).

**Remark 3.4.** The characterisation in Lemma 3.3(b) is analogous to Eichler [28], Definition 2.2. The other characterisations are useful for checking Granger non-causality. In particular, we have shown implicitly in Lemma 3.3(d) that

$$Y_A \not\rightarrow Y_B \mid Y_S \Leftrightarrow Y_A \not\rightarrow Y_b \mid Y_S \quad \forall b \in B. \tag{3.2}$$

From the characterisations in Lemma 3.3, the idea of Granger non-causality as equality of two predictions, as given, e.g., in Dufour and Renault [26] for discrete-time processes, is not yet clear. Therefore, we provide another characterisation of Granger non-causality using orthogonal projections.

**Theorem 3.5.** Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A$  is Granger non-causal for  $Y_B$  with respect to  $Y_S$  if for all  $h \in [0, 1], t \in \mathbb{R}$ , and  $b \in B$ ,

$$P_{\mathcal{L}_{Y_S}(t)} Y_b(t + h) = P_{\mathcal{L}_{Y_{S \setminus A}}(t)} Y_b(t + h) \quad \mathbb{P}\text{-a.s.}$$

In other words, the information given by the past process  $(Y_A(s), s \leq t)$  can be forgotten without any consequences for the optimal linear prediction of  $Y_B(t + h)$  for  $h \in [0, 1]$ . In particular, since  $\mathcal{L}_{Y_{S \setminus A}}(t) \subseteq \mathcal{L}_{Y_{S \setminus \{a\}}}(t) \subseteq \mathcal{L}_{Y_S}(t)$  for any  $a \in A$ , we receive

$$Y_A \not\rightarrow Y_B \mid Y_S \Rightarrow Y_a \not\rightarrow Y_b \mid Y_S \quad \forall a \in A, b \in B. \tag{3.3}$$

Under some additional model assumptions the opposite direction is also true. However, this is the topic of Section 5.

**Remark 3.6.** Florens and Fougère [35], Definition 2.1, and Comte and Renault [15], Definition 1, take a different approach to define Granger non-causality in continuous-time, using the equality of conditional expectations instead of orthogonal projections, and generated  $\sigma$ -fields instead of generated linear spaces. Comte and Renault [15], Definition 2, also define a local version of Granger causality, called local instantaneous causality, in the context of semimartingales. In Proposition 1, they further relate it to the definition of Renault and Szafarz [64], who study first-order stochastic differential equations and define local Granger causality as the equality of limits of predictions. That is, instead of looking at the entire prediction time interval  $[t, t + 1]$ , they examine  $[t, t + h]$  as  $h \rightarrow 0$ . To obtain non-trivial limits, they consider limits of difference quotients of the underlying process. Finally, Comte and Renault [15] discuss in detail in Section 2.5 that the highest existing derivative of the process must always be examined to obtain a non-trivial criterion which is the motivation for our approach. Therefore, in the style of their characterisation of local Granger causality and our Theorem 3.5, we define the following version of local Granger causality which is, as we derive in Lemma 3.13, weaker as Granger causality.

**Definition 3.7.** Suppose  $Y_v = (Y_v(t))_{t \in \mathbb{R}}$  is  $j_v$ -times mean-square differentiable but the  $(j_v + 1)$ -derivative does not exist for  $v \in V$ . The  $j_v$ -derivative is denoted by  $D^{(j_v)} Y_v$ , where for  $j_v = 0$  we define  $D^{(0)} Y_v = Y_v$ . Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A$  is *local*

Granger non-causal for  $Y_B$  with respect to  $Y_S$  if, for all  $t \in \mathbb{R}$  and  $b \in B$ ,

$$\lim_{h \rightarrow 0} \text{i.m. } P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)}Y_b(t+h) - D^{(j_b)}Y_b(t)}{h} \right) = \lim_{h \rightarrow 0} \text{i.m. } P_{\mathcal{L}_{Y_S \setminus A}(t)} \left( \frac{D^{(j_b)}Y_b(t+h) - D^{(j_b)}Y_b(t)}{h} \right) \quad \mathbb{P}\text{-a.s.}$$

We write  $Y_A \not\rightarrow_0 Y_B | Y_S$ .

**Remark 3.8.**

(a) Since  $Y_b$  is by assumption not  $(j_b + 1)$ -times mean-square differentiable, the  $L^2$ -limit of  $(D^{(j_b)}Y_b(t+h) - D^{(j_b)}Y_b(t))/h$  does not exist. However, it is still possible that the  $L^2$ -limit of

$$P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)}Y_b(t+h) - D^{(j_b)}Y_b(t)}{h} \right) \quad \text{and} \quad P_{\mathcal{L}_{Y_S \setminus A}(t)} \left( \frac{D^{(j_b)}Y_b(t+h) - D^{(j_b)}Y_b(t)}{h} \right) \tag{3.4}$$

exist and only then local Granger non-causality is possible.

(b) Typical examples of stochastic processes that are  $p$ -times but not  $(p + 1)$ -times mean-square differentiable and the  $L^2$ -limit in (3.4) for  $j_a = j_b = p$  exist are MCAR( $p$ ) processes (Section 6) and the more general class of state space models (Fasen-Hartmann and Schenk [32]). In addition, fractional MCAR processes also satisfy these assumptions (Comte and Renault [15], Marquardt [55]).

**Remark 3.9.** By definition we receive

$$Y_A \not\rightarrow_0 Y_B | Y_S \Leftrightarrow Y_A \not\rightarrow_0 Y_b | Y_S \quad \forall b \in B. \tag{3.5}$$

Moreover, for  $a \in A$ , the subset relation  $\mathcal{L}_{Y_S \setminus A}(t) \subseteq \mathcal{L}_{Y_S \setminus \{a\}}(t) \subseteq \mathcal{L}_{Y_S}(t)$  implies

$$Y_A \not\rightarrow_0 Y_B | Y_S \Rightarrow Y_a \not\rightarrow_0 Y_b | Y_S \quad \forall a \in A, b \in B. \tag{3.6}$$

Again, the opposite direction is valid under some additional assumption, see Section 5.

Local Granger causality implies a kind of local version of conditional orthogonality.

**Theorem 3.10.** Suppose  $Y_v = (Y_v(t))_{t \in \mathbb{R}}$  is  $j_v$ -times mean-square differentiable but the  $(j_v + 1)$ -derivative does not exist for  $v \in V$ . Further, let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A \not\rightarrow_0 Y_B | Y_S$  implies that, for all  $Y^A \in \mathcal{L}_{Y^A}(t)$  and  $t \in \mathbb{R}$ ,

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( D^{(j_b)}Y_b(t+h) - P_{\mathcal{L}_{Y_S \setminus A}(t)} D^{(j_b)}Y_b(t+h) \right) \overline{\left( Y^A - P_{\mathcal{L}_{Y_S \setminus A}(t)} Y^A \right)} \right] = 0.$$

A third concept of directed influence is to consider causality up to an arbitrary horizon. In discrete time, the concept of causality at any horizon goes back to the seminal work of Sims [71] and is also called Sims causality. We introduce the following definition as a generalisation of Eichler [28], Definition 4.4, to continuous-time processes.

**Definition 3.11.** Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A$  is global Granger non-causal for  $Y_B$  with respect to  $Y_S$  if, for all  $h \geq 0$  and  $t \in \mathbb{R}$ ,

$$\mathcal{L}_{Y_B}(t, t+h) \perp \mathcal{L}_{Y_A}(t) | \mathcal{L}_{Y_S \setminus A}(t).$$

We write  $Y_A \not\rightarrow_\infty Y_B | Y_S$ .

The study of such long-run effects is a useful complement to understanding the relationship between the component series and allows us to distinguish between short-run and long-run causality.

**Remark 3.12.** The characterisations are similar to those for Granger causality. In particular,  $Y_A$  is global Granger non-causal for  $Y_B$  with respect to  $Y_S$ , if and only if, for all  $h \geq 0, t \in \mathbb{R}$  and  $b \in B$ ,

$$P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h) = P_{\mathcal{L}_{Y_S \setminus A}(t)} Y_b(t+h) \quad \mathbb{P}\text{-a.s.} \tag{3.7}$$

On the one hand, note that the proof is similar to the proof of Theorem 3.5 and on the other hand, that analogue relationships as in (3.2) and (3.3) hold. The characterisation (3.7) is again consistent with the characterisation in Dufour and Renault [26] for discrete-time processes and with the definition of global Granger causality in Comte and Renault [15], who use generated  $\sigma$ -fields instead of linear spaces and conditional expectations instead of orthogonal projections. Of course, for Gaussian processes, the two definitions coincide.

In the following lemma, we state relations between Granger non-causality, local Granger non-causality and global Granger non-causality. See again Dufour and Renault [26], Eichler [30], Kuersteiner [48] for the relations between the different definitions for discrete-time processes.

**Lemma 3.13.** Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then the following implications hold:

- (a)  $Y_A \not\rightarrow_{\infty} Y_B | Y_S \Rightarrow Y_A \not\rightarrow Y_B | Y_S$ .
- (b)  $Y_A \not\rightarrow_{\infty} Y_{S \setminus A} | Y_S \Leftrightarrow Y_A \not\rightarrow Y_{S \setminus A} | Y_S$ .
- (c)  $Y_A \not\rightarrow Y_{S \setminus A} | Y_S \Rightarrow Y_A \not\rightarrow_{\infty} Y_B | Y_S$ .
- (d)  $Y_A \not\rightarrow Y_B | Y_S \Rightarrow Y_A \not\rightarrow_0 Y_B | Y_S$ .

**Remark 3.14.** The opposite direction in Lemma 3.13(a) does not hold in general. Dufour and Renault [26], p. 1106, present a counterexample in discrete time and explain the lack of equivalence between Granger non-causality and global Granger non-causality as follows. If there are auxiliary components,  $Y_A$  might not help to predict  $Y_B$  given  $Y_S$  one step ahead but  $Y_A$  might help to predict  $Y_B$  given  $Y_S$  several periods ahead. For example, the values of  $Y_A$  up to time  $t$  may help to predict  $\mathcal{L}_{Y_B}(t + 1, t + 2)$ , even though they are useless to predict  $\mathcal{L}_{Y_B}(t, t + 1)$ , because  $Y_A$  may help to predict the environment one period ahead, which in turn influences  $Y_A$  at a subsequent period. Therefore, it is also not surprising that we have equivalence in the case without environment in Lemma 3.13(b). This holds in particular for every bivariate process, i.e.,

$$Y_a \not\rightarrow Y_b | Y_{\{a,b\}} \Leftrightarrow Y_a \not\rightarrow_{\infty} Y_b | Y_{\{a,b\}}.$$

The similarities and differences between the various definitions of Granger causality can also be seen in examples, so we examine Ornstein–Uhlenbeck processes. In particular, we see that the opposite direction of Lemma 3.13(d) does not generally hold.

**Example 3.15.** Suppose  $Y_V = (Y_V(t))_{t \in \mathbb{R}}$  is an Ornstein–Uhlenbeck process driven by a two-sided  $k$ -dimensional Lévy process  $(L(t))_{t \in \mathbb{R}}$ . An one-sided Lévy process  $(L(t))_{t \geq 0}$  is an  $\mathbb{R}^k$ -valued stochastic process with  $L(0) = 0$ ,  $\mathbb{P}$ -a.s., stationary and independent increments and càdlàg sample paths. Now,  $L = (L(t))_{t \in \mathbb{R}}$  is obtained from two independent copies  $(L_1(t))_{t \geq 0}$  and  $(L_2(t))_{t \geq 0}$  of a one-sided Lévy process via  $L(t) = L_1(t)$  if  $t \geq 0$  and  $L(t) = -\lim_{s \nearrow -t} L_2(s)$  if  $t < 0$ . We assume that the Lévy process has a finite second moment with  $\Sigma_L := \mathbb{E}[L(1)L(1)^T]$  and expectation zero. Suppose further that  $A \in M_k(\mathbb{R})$  with  $\sigma(A) \subseteq (-\infty, 0) + i\mathbb{R}$ . Then the stochastic differential equation

$$dY_V(t) = AY_V(t)dt + dL(t)$$

has the unique stationary solution  $Y_V$  given by

$$Y_V(t) = \int_{-\infty}^t e^{A(t-u)} dL(u), \quad t \in \mathbb{R}.$$

The process  $Y_V$  is called (causal) *Ornstein–Uhlenbeck process* (cf. Masuda [57]). For the Ornstein–Uhlenbeck process, we derive in Section 6, in the more general context of (causal) MCAR processes, that

$$Y_a \not\rightarrow_{\infty} Y_b | Y_V \Leftrightarrow Y_a \not\rightarrow Y_b | Y_V \Leftrightarrow [A^\alpha]_{ab} = 0, \quad \alpha = 1, \dots, k - 1, \\ Y_a \not\rightarrow_0 Y_b | Y_V \Leftrightarrow [A]_{ab} = 0.$$

Of course,

$$Y_a \not\rightarrow Y_b | Y_V \Rightarrow [A^\alpha]_{ab} = 0, \alpha = 1, \dots, k - 1 \Rightarrow [A]_{ab} = 0 \Rightarrow Y_a \not\rightarrow_0 Y_b | Y_V,$$

but the opposite direction does not generally hold, an exception is the case where  $A$  is a diagonal matrix. A specific counterexample is the Ornstein–Uhlenbeck process with

$$A = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \quad \text{and} \quad \Sigma_L = \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1 \end{pmatrix}, \tag{3.8}$$

which is the underlying stochastic process of Fig. 1. Here,  $Y_1 \not\rightarrow_0 Y_2 | Y_{\{1,2,3\}}$  but  $Y_1 \rightarrow Y_2 | Y_{\{1,2,3\}}$ . It is clear from the example that Granger non-causality is much stronger than local Granger non-causality, and that in general there is no equivalence. Note that the special structure of  $\Sigma_L$  does not play a role in these directed influences, but the covariance structure has an impact on the undirected influences which we will define in the next section.

#### 4. Undirected influences: Contemporaneous correlation for stationary continuous-time processes

In this section, we introduce and characterise undirected influences between the component series of  $Y_V$  using different concepts of contemporaneous correlation. The idea is simple: There is no undirected influence between  $Y_a$  and  $Y_b$ , if and only if, given the amount of information provided by the past of  $Y_V$  up to time  $t$ ,  $Y_a$  and  $Y_b$  are uncorrelated in the future. Again, we need to specify what we mean by the future in continuous time. The first definition is a generalisation of Eichler [28], Definition 2.2, in discrete time, to continuous time, looking at the entire time interval  $[t, t + 1]$ .

**Definition 4.1.** Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A$  and  $Y_B$  are *contemporaneously uncorrelated* with respect to  $Y_S$  if, for all  $t \in \mathbb{R}$ ,

$$\mathcal{L}_{Y_A}(t, t + 1) \perp \mathcal{L}_{Y_B}(t, t + 1) | \mathcal{L}_{Y_S}(t).$$

We write  $Y_A \sim Y_B | Y_S$ .



**Remark 4.2.** Similarly, as for the definition of Granger causality, we defined contemporaneous uncorrelation by using the step size  $h = 1$ . However, it is also possible to use some arbitrary but fixed step size  $h > 0$  and define it via

$$\mathcal{L}_{Y_A}(t, t+h) \perp \mathcal{L}_{Y_B}(t, t+h) \mid \mathcal{L}_{Y_S}(t) \quad \forall t \in \mathbb{R}. \tag{4.1}$$

The choice of  $h$  has no effect on the characterisation of the undirected influences in certain models; see [Remark 6.20](#) for MCAR processes and Fasen-Hartmann and Schenk [\[32\]](#) for state space models. Again, we believe that this is generally not true for non-linear processes.

Unlike Granger causality, contemporaneous correlation is symmetric, reflecting an undirected influence. By analogy with [Lemma 3.3](#), we obtain the following equivalent characterisations of contemporaneous uncorrelation. Since the proof is very similar, it is not given here. Again, the stationarity assumption is not necessary for the definition of contemporaneous uncorrelation and its characterisations, it can be neglected.

**Lemma 4.3.** *Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then the following characterisations are equivalent:*

- (a)  $Y_A \approx Y_B \mid Y_S$ .
- (b)  $\mathcal{L}_{Y_A}(t+1) \perp \mathcal{L}_{Y_B}(t+1) \mid \mathcal{L}_{Y_S}(t) \quad \forall t \in \mathbb{R}$ .
- (c)  $\ell_{Y_A}(t, t+1) \perp \ell_{Y_B}(t, t+1) \mid \mathcal{L}_{Y_S}(t) \quad \forall t \in \mathbb{R}$ .
- (d)  $\ell_{Y_a}(s, s) \perp \ell_{Y_b}(s', s') \mid \mathcal{L}_{Y_S}(t) \quad \forall a \in A, b \in B, s, s' \in [t, t+1], t \in \mathbb{R}$ .

**Remark 4.4.** In the following, we make some remarks about [Lemma 4.3\(d\)](#).

- (a) In [Lemma 4.3\(d\)](#), we have implicitly shown that

$$Y_A \approx Y_B \mid Y_S \iff Y_a \approx Y_b \mid Y_S \quad \forall a \in A, b \in B,$$

which is useful for the verification of contemporaneous uncorrelation.

- (b) Given our [Lemma 4.3\(d\)](#) and Eichler [\[28\]](#), Definition 2.2, it would also be plausible to define contemporaneous uncorrelation by  $\ell_{Y_a}(s, s) \perp \ell_{Y_b}(s, s) \mid \mathcal{L}_{Y_S}(t) \quad \forall a \in A, b \in B, s \in [t, t+1], t \in \mathbb{R}$ . In this case, however, no global Markov property can be shown in the associated orthogonality graph (cf. Section 5), since the evidences rely heavily on [Definition 4.1](#) and [Lemma 2.2](#).

Similar to Granger non-causality, a characterisation of contemporaneous uncorrelation can be given, which allows for an interpretation as the correspondence of two linear predictions.

**Theorem 4.5.** *Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A$  and  $Y_B$  are contemporaneously uncorrelated with respect to  $Y_S$ , if and only if, for all  $b \in B, h \in [0, 1]$ , and  $t \in \mathbb{R}$ ,*

$$P_{\mathcal{L}_{Y_S}(t) \vee \mathcal{L}_{Y_A}(t, t+1)} Y_b(t+h) = P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h) \quad \mathbb{P}\text{-a.s.}$$

In words, the linear prediction of the information about  $Y_B$  in the near future based on  $\mathcal{L}_{Y_S}(t)$  cannot be improved by adding further information about  $Y_A$  in the near future (and vice versa). The proof is again similar to the proof of [Theorem 3.5](#) and we therefore skip the details.

To define a local version of contemporaneous uncorrelation, note that the characterisation in [Lemma 4.3\(b\)](#) means that for any  $Y^A \in \mathcal{L}_{Y_A}(t+1)$  and  $Y^B \in \mathcal{L}_{Y_B}(t+1)$

$$\mathbb{E} \left[ \left( Y^A - P_{\mathcal{L}_{Y_S}(t)} Y^A \right) \overline{\left( Y^B - P_{\mathcal{L}_{Y_S}(t)} Y^B \right)} \right] = 0. \tag{4.2}$$

So the motivation for the local version is that instead of taking all  $Y^A \in \mathcal{L}_{Y_A}(t+1)$ , we use only the highest derivative  $D^{(j_a)} Y_a(t+h)$  for each  $a \in A$  and consider  $h \rightarrow 0$ , similarly for  $\mathcal{L}_{Y_B}(t+1)$ . To get non-trivial limits we also have to divide by  $h$ .

**Definition 4.6.** Suppose  $Y_v = (Y_v(t))_{t \in \mathbb{R}}$  is  $j_v$ -times mean-square differentiable but the  $(j_v + 1)$ -derivative does not exist for  $v \in V$ . Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A$  and  $Y_B$  are *locally contemporaneously uncorrelated* with respect to  $Y_S$  if, for all  $t \in \mathbb{R}, a \in A, b \in B$ ,

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( D^{(j_a)} Y_a(t+h) - P_{\mathcal{L}_{Y_S}(t)} D^{(j_a)} Y_a(t+h) \right) \overline{\left( D^{(j_b)} Y_b(t+h) - P_{\mathcal{L}_{Y_S}(t)} D^{(j_b)} Y_b(t+h) \right)} \right] = 0.$$

We write  $Y_A \approx_0 Y_B \mid Y_S$ .

**Remark 4.7.**

- (a) Due to the definition, we receive directly

$$Y_A \approx_0 Y_B \mid Y_S \iff Y_a \approx_0 Y_b \mid Y_S \quad \forall a \in A, b \in B,$$

which is useful for verifying local contemporaneous uncorrelation.

- (b) Definition 4.6 is similar to the characterisation of local contemporaneous uncorrelation for semimartingales in Comte and Renault [15], Proposition 3, using linear predictions instead of conditional expectations and  $\sigma$ -fields instead of linear spaces. But Comte and Renault [15] assume additionally that the martingale part of the semimartingale is continuous, excluding Lévy-Itô processes that are not Brownian motion driven, such as Lévy-driven Ornstein-Uhlenbeck processes.
- (c) To give an equivalent characterisation as an equality of projections, restrictions on the linear derivative spaces are necessary. Thus, we do not include these characterisations here. Sufficient, however, is in any case that for all  $t \in \mathbb{R}$ ,  $a \in A$ ,  $b \in B$ ,

$$\text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{\sqrt{h}} \right) = \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t) \vee \mathcal{L}_a(t,t+h)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{\sqrt{h}} \right) \quad \mathbb{P}\text{-a.s.}$$

Finally, we introduce a global concept of contemporaneous correlation, in analogy to global Granger causality, to discuss short-run vs. long-run effects.

**Definition 4.8.** Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then  $Y_A$  and  $Y_B$  are globally contemporaneously uncorrelated with respect to  $Y_S$  if, for  $h \geq 0$  and  $t \in \mathbb{R}$ ,

$$\mathcal{L}_{Y_A}(t, t+h) \perp \mathcal{L}_{Y_B}(t, t+h) \mid \mathcal{L}_{Y_S}(t).$$

We write  $Y_A \approx_\infty Y_B \mid Y_S$ .

**Remark 4.9.** Again, projections can be used to characterise the global contemporaneous uncorrelation. Precisely,  $Y_A$  and  $Y_B$  are globally contemporaneously uncorrelated with respect to  $Y_S$ , if and only if, for all  $b \in B$ ,  $0 \leq h' \leq h$ ,  $h \geq 0$ , and  $t \in \mathbb{R}$

$$P_{\mathcal{L}_{Y_S}(t) \vee \mathcal{L}_{Y_A}(t,t+h)} Y_b(t+h') = P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h') \quad \mathbb{P}\text{-a.s.}$$

The proof is similar to the proof of Theorem 4.5 and is therefore not included in the paper. Also, the analogue statements to Remark 4.4 hold.

It is obvious that, by definition and due to Remark 2.6 and (4.2), the following relations between the three definitions of contemporaneous uncorrelation are valid.

**Lemma 4.10.** Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then the following implications hold:

- (a)  $Y_A \approx_\infty Y_B \mid Y_S \Rightarrow Y_A \approx Y_B \mid Y_S$ .
- (b)  $Y_A \approx Y_B \mid Y_S \Rightarrow Y_A \approx_0 Y_B \mid Y_S$ .

The similarities and differences between the various definitions again become apparent when looking at examples. In particular, we derive that the opposite direction in Lemma 4.10(b) does not hold in general.

**Example 4.11.** Suppose  $Y_V$  is the Ornstein-Uhlenbeck process as defined in Example 3.15 with  $A$  and  $\Sigma_L$  as in (3.8). Then we derive in Section 6 that

$$\begin{aligned} Y_a \approx_\infty Y_b \mid Y_V &\Leftrightarrow Y_a \approx Y_b \mid Y_V \Leftrightarrow [A^\alpha \Sigma_L (A^\top)^\beta]_{ab} = 0, \quad \alpha, \beta = 0, \dots, k-1, \\ Y_a \approx_0 Y_b \mid Y_V &\Leftrightarrow [\Sigma_L]_{ab} = 0. \end{aligned}$$

Of course, we obtain

$$Y_a \approx Y_b \mid Y_V \Rightarrow [A^\alpha \Sigma_L (A^\top)^\beta]_{ab} = 0, \quad \alpha, \beta = 0, \dots, k-1 \Rightarrow [\Sigma_L]_{ab} = 0 \Rightarrow Y_a \approx_0 Y_b \mid Y_V,$$

but the opposite direction does not generally hold, in turn, an exception is the case where  $A$  is a diagonal matrix. A specific counterexample is again the Ornstein-Uhlenbeck process from Example 3.15, which we see in Fig. 1. Here,  $Y_1 \approx_0 Y_2 \mid Y_{\{1,2,3\}}$  but  $Y_1 \not\approx Y_2 \mid Y_{\{1,2,3\}}$ .

### 5. Orthogonality graphs for stationary continuous-time processes

In this section, we introduce graphical models for stationary, mean-square continuous processes  $Y_V = (Y_V(t))_{t \in \mathbb{R}}$ . These graphical models visualise directed as well as undirected relations between the different component series  $Y_v = (Y_v(t))_{t \in \mathbb{R}}$ ,  $v = 1, \dots, k$ . The vertices represent the different component series  $Y_v$ ,  $v = 1, \dots, k$ , of the process. Furthermore, they are connected by directed and undirected edges, which represent certain directional and non-directional influences between them. The arising graphical models are then called (mixed) orthogonality graphs.

#### 5.1. Separability and conditional linear separation

For the definition of the graphical models, we first ensure that the conditional orthogonality relation satisfies the property of intersection (C5) in Lemma 2.2 for suitable linear subspaces and second, we show that the missing relations in (3.3) and (3.6) hold. Therefore, we investigate separability and conditional linear separation of linear spaces. The proofs of the lemmata of this subsection are the subject of the Supplementary Material E, and the proofs of the propositions and theorems are content of Appendix B.1.

**Lemma 5.1.** *Let  $A \subseteq V$  and  $s, t \in \mathbb{R}$  with  $s < t$ . Then  $\mathcal{L}_{Y_A}, \mathcal{L}_{Y_A}(t)$  and  $\mathcal{L}_{Y_A}(s, t)$  are separable.*

Furthermore, we require that  $\mathcal{L}_{Y_A}(t)$  and  $\mathcal{L}_{Y_B}(t)$  are conditionally linearly separated by  $\mathcal{L}_{Y_C}(t)$  if  $t \in \mathbb{R}$  and  $A, B, C \subseteq V$  are disjoint. This assumption is a lot more intricate because it is a very abstract definition and difficult to verify.

**Remark 5.2.** Unlike us, Eichler [29] uses conditional independence instead of conditional orthogonality. For the associated intersection property (C5) *measurable conditional separation* is required, corresponding to our *conditional linear separation* assumption. There, measurable conditional separation is also generally not valid, and sufficient assumptions are given.

To better understand conditional linear separation, we introduce a sufficient criterion.

**Lemma 5.3.** *Let  $t \in \mathbb{R}$ . Suppose that for all  $A, B \subseteq V$  with  $A \cap B = \emptyset$  we have*

$$\mathcal{L}_{Y_A}(t) \cap \mathcal{L}_{Y_B}(t) = \{0\} \quad \text{and} \quad \mathcal{L}_{Y_A}(t) + \mathcal{L}_{Y_B}(t) = \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_B}(t) \quad \mathbb{P}\text{-a.s.}$$

Then, for all disjoint subsets  $A, B, C \subseteq V$ , we get

$$\mathcal{L}_{Y_{A \cup C}}(t) \cap \mathcal{L}_{Y_{B \cup C}}(t) = \mathcal{L}_{Y_C}(t) \quad \mathbb{P}\text{-a.s.}$$

The first assumption is the linear independence of the two linear spaces, the second assumption is the closedness of the sum. It makes little sense to formulate these two properties as assumptions on  $Y_V$ , as they are still too abstract and difficult to verify. Therefore, we provide an easy-to-use criterion.

**Assumption 1.** Suppose  $Y_V$  has a spectral density matrix  $f_{Y_V Y_V}(\cdot) > 0$  and that there exists an  $0 < \varepsilon < 1$ , such that

$$d_{AB}(\lambda) := f_{Y_A Y_A}(\lambda)^{-1/2} f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda) f_{Y_A Y_A}(\lambda)^{-1/2} \leq_L (1 - \varepsilon) I_\alpha,$$

for almost all  $\lambda \in \mathbb{R}$  and for all disjoint subsets  $A, B \subseteq V$  with  $\#A = \alpha$ .

For  $A = \{a\}$  the function  $d_{AB}(\lambda)$ ,  $\lambda \in \mathbb{R}$ , is called multiple coherence; we refer to Priestley [63], Brillinger [7] for further reading. Assumption 1 is satisfied, e.g., for stationary causal MCAR processes and in particular Ornstein–Uhlenbeck processes, for details see Section 6, and for the more general family of state space models see Fasen-Hartmann and Schenk [32]. In our opinion, even fractional MCAR processes satisfy this assumption. Furthermore, the assumption is indeed sufficient for conditional linear separability.

**Proposition 5.4.** *Let  $Y_V$  satisfy Assumption 1. Then for all  $t \in \mathbb{R}$  and disjoint subsets  $A, B, C \subseteq V$ , we have*

$$\mathcal{L}_{Y_A}(t) \cap \mathcal{L}_{Y_B}(t) = \{0\}, \quad \mathcal{L}_{Y_A}(t) + \mathcal{L}_{Y_B}(t) = \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_B}(t), \quad \text{and} \quad \mathcal{L}_{Y_{A \cup C}}(t) \cap \mathcal{L}_{Y_{B \cup C}}(t) = \mathcal{L}_{Y_C}(t) \quad \mathbb{P}\text{-a.s.}$$

Recall that in Theorem 3.10 we already assume the closedness of the sum, and now Proposition 5.4 gives a sufficient criterion for this property.

**Remark 5.5.** First of all,  $d_{AB}(\lambda) \leq_L I_{\alpha \times \alpha}$  holds even without Assumption 1. Indeed, suppose  $\Phi_B(\cdot)$  is the random spectral measure from the spectral representation of  $Y_B$  in (2.2), then the spectral density matrix of

$$\varepsilon_{A|B}(t) = Y_A(t) - \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} \Phi_B(d\lambda)$$

is

$$f_{\varepsilon_{A|B} \varepsilon_{A|B}}(\lambda) = f_{Y_A Y_A}(\lambda) - f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda),$$

and it is non-negative definite according to Brockwell and Davis [9], p. 436. Furthermore, Assumption 1 especially forbids some purely linear relationships between the components, which can be seen as follows. Assume that  $d_{AB}(\lambda) = I_\alpha$  for almost all  $\lambda \in \mathbb{R}$ . Then  $f_{\varepsilon_{A|B} \varepsilon_{A|B}}(\lambda) = 0_\alpha$  for almost all  $t \in \mathbb{R}$  and thus,  $c_{\varepsilon_{A|B} \varepsilon_{A|B}}(t) = 0_\alpha$  for all  $t \in \mathbb{R}$ . Therefore,  $\varepsilon_{A|B}(t) = 0_\alpha$   $\mathbb{P}$ -a.s. and  $Y_A(t)$  is already a linear transformation of  $Y_B(t)$ . Somewhat loosely, one could say that Assumption 1 not only forbids a purely linear relationship between  $Y_A$  and  $Y_B$  but already requires some kind of distance between the subprocesses due to the uniform boundedness. This also fits with Brillinger [7], eq. (8.3.10), who calls the matrix function  $d_{AB}(\lambda)$  in discrete-time a measure of the linear association of  $Y_A$  and  $Y_B$  at frequency  $\lambda$ .

**Remark 5.6.** Let us compare Assumption 1 with Eichler [28], Eq. (2.1), who proposes a comparable assumption on the spectral density matrix in discrete time, also with the aim that the property of intersection (C5) is valid. Eichler [28] demands the existence of a constant  $c > 1$ , such that the spectral density matrix satisfies

$$\frac{1}{c} I_k \leq_L f_{Y_V Y_V}(\lambda) \leq_L c I_k, \tag{5.1}$$

for all  $\lambda \in [-\pi, \pi]$ . If this assumption is fulfilled, some matrix algebra calculations as in the proof of Lemma F.1 give that for any disjoint subsets  $A, B \subseteq V$ ,

$$f_{Y_A Y_A}(\lambda) - f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda) \geq_L \frac{1}{c} I_\alpha \geq_L \frac{1}{c^2} f_{Y_A Y_A}(\lambda).$$

Thus, on the interval  $[-\pi, \pi]$  Assumption 1 is satisfied with  $\varepsilon = 1/c^2$ . However, Eichler [28]’s assumption is stricter than ours since one must be able to place a diagonal matrix between  $1/c^2 f_{Y_A Y_A}(\lambda)$  and  $f_{Y_A Y_A}(\lambda) - f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda)$ . We further point out that we cannot generalise Eichler [28]’s assumption directly to continuous-time processes by assuming (5.1) for almost all  $\lambda \in \mathbb{R}$ . This requirement is too strict and, e.g., not satisfied for Ornstein–Uhlenbeck processes.

Assumption 1 now ensures, as desired, that the conditional orthogonality relation satisfies the property of intersection (C5) in Lemma 2.2 for suitable linear subspaces. Assumption 1 further provides us with the missing relations of the causality concepts in (3.3) and (3.6).

**Proposition 5.7.** *Let  $Y_V$  satisfy Assumption 1. Let  $A, B \subseteq S \subseteq V$  and  $A \cap B = \emptyset$ . Then*

- (a)  $Y_A \not\rightarrow Y_B \mid Y_S \Leftrightarrow Y_a \not\rightarrow Y_b \mid Y_S \quad \forall a \in A, b \in B.$
- (b)  $Y_A \not\rightarrow_0 Y_B \mid Y_S \Leftrightarrow Y_a \not\rightarrow_0 Y_b \mid Y_S \quad \forall a \in A, b \in B.$
- (c)  $Y_A \not\rightarrow_\infty Y_B \mid Y_S \Leftrightarrow Y_a \not\rightarrow_\infty Y_b \mid Y_S \quad \forall a \in A, b \in B.$

However, for the proof of the global Andersson, Madigan and Perlman (AMP) Markov property in our orthogonality graph, we require further assumptions. Any process that is wide sense stationary can be uniquely decomposed in a deterministic and a purely non-deterministic process that are mutually orthogonal (Gladyshev [39], Theorem 1). From the point of view of applications, deterministic processes are not important. Therefore, we assume that the given process is purely non-deterministic.

**Assumption 2.** Let  $Y_V$  be purely non-deterministic, that is  $\mathcal{L}_{Y_V}(-\infty) = \{0\}$   $\mathbb{P}$ -a.s.

Necessary and sufficient conditions for processes being purely non-deterministic can be found, e.g., in Gladyshev [39], Theorem 3, Rozanov [67], III, Theorem 2.4, Matveev [58], Theorem 1. Typical examples are MCAR processes and the more general class of state space models whose driving Lévy process has expectation zero.

Finally, we can deduce the following property from Assumptions 1 and 2, which we require for the proof of the global AMP Markov property. The property further stands in analogy to assumption (M) on  $\sigma$ -fields in Eichler [29] and equation (2.4) in Eichler [27]. Note that these assumptions are stronger than our Assumptions 1 and 2 and quite difficult to verify.

**Lemma 5.8.** *Let  $Y_V$  satisfy Assumptions 1 and 2. Let  $A \subseteq V$  and  $t \in \mathbb{R}$ . Then*

$$\bigcap_{k \in \mathbb{N}} \left( \mathcal{L}_{Y_A}(t-k) \vee \mathcal{L}_{Y_{V \setminus A}}(t) \right) = \mathcal{L}_{Y_{V \setminus A}}(t) \quad \mathbb{P}\text{-a.s.} \tag{5.2}$$

Note that Assumptions 1 and 2 are not necessary assumptions for the following Markov properties to hold. Sufficient and weaker assumptions are the conditional linear separation and (5.2), both are satisfied under Assumptions 1 and 2.

### 5.2. Introduction to (local) orthogonality graphs

Let us now visualise suitable concepts of directed and undirected influences in graphical models. In principle, it is possible to define a graph with any of the three definitions of Granger causality and contemporaneous correlation. However, our goal is to define a graph with concepts that are as strong as necessary, but as weak as possible, so that the usual Markov properties for mixed graphs hold. For MCAR processes, global Granger causality and Granger causality as well as global contemporaneous uncorrelation and contemporaneous uncorrelation coincide (see Section 6) and therefore we do not discuss a global graph.

**Definition 5.9.** Let  $Y_V$  satisfy Assumptions 1 and 2.

- (a) If we define  $V = \{1, \dots, k\}$  as the vertices and the edges  $E_{OG}$  via

- (i)  $a \longrightarrow b \notin E_{OG} \Leftrightarrow Y_a \not\rightarrow Y_b \mid Y_V,$
- (ii)  $a \text{ --- } b \notin E_{OG} \Leftrightarrow Y_a \not\sim Y_b \mid Y_V,$

for  $a, b \in V, a \neq b$ , then  $G_{OG} = (V, E_{OG})$  is called (mixed) orthogonality graph for  $Y_V$ .

- (b) If we define  $V = \{1, \dots, k\}$  as the vertices and the edges  $E_{OG}^0$  via

- (i)  $a \longrightarrow b \notin E_{OG}^0 \Leftrightarrow Y_a \not\rightarrow_0 Y_b \mid Y_V,$
- (ii)  $a \text{ --- } b \notin E_{OG}^0 \Leftrightarrow Y_a \not\sim_0 Y_b \mid Y_V,$

for  $a, b \in V, a \neq b$ , then  $G_{OG}^0 = (V, E_{OG}^0)$  is called local (mixed) orthogonality graph for  $Y_V$ .

In words, in both graphs each vertex  $v \in V$  represents one component series  $Y_v$ . Two vertices  $a$  and  $b$  are joined by a directed edge  $a \longrightarrow b$  whenever  $Y_a$  is (local) Granger causal for  $Y_b$  and by an undirected edge  $a \text{ --- } b$  whenever  $Y_a$  and  $Y_b$  are (locally) contemporaneously correlated given  $Y_V$ . We make some remarks on those graphical models.

**Remark 5.10.**

- (a) The motivation for the name (local) orthogonality graph arises from the fact that both the directed and undirected edges are defined by specific (local) conditional orthogonality relations. For a concise notation, we omit the word conditional. Furthermore, the name (local) orthogonality graph is also analogous to the local independence graph (Didelez [21,22,23], Mogensen and Hansen [59,60]). The graphical models are further named *mixed* orthogonality graphs because they contain two types of edges. Since we do not usually consider purely directed or undirected graphs, we omit the prefix mixed for ease of notation. Note that the orientation of the directed edge makes a difference and multiple edges of the same type and orientation are not allowed. Thus, two vertices  $a$  and  $b$  can be connected by up to three edges, namely  $a \longrightarrow b$ ,  $a \longleftarrow b$  and  $a --- b$ , as can also be seen in Fig. 1.
- (b) The Assumptions 1 and 2 as well as the stationarity and the mean square continuity are not necessary for the definition of the graphs, but they are essential for the usual Markov properties to hold. Wide sense stationarity is a basic requirement, otherwise, e.g., Assumption 1 is not well-defined, which is a sufficient assumption for conditional linear separation. The mean square continuity and Assumption 1 will already be used for the first time in the proof of the local Markov property. Assumption 2 is only required in the proof of the global AMP Markov property. Since we show global Markov properties for the local orthogonality graph only in special cases, Assumption 2 is not necessary there.
- (c) We already know that  $a \longrightarrow b \notin E_{OG}$  directly implies  $a \longrightarrow b \notin E_{OG}^0$  and similarly  $a --- b \notin E_{OG}$  also gives  $a --- b \notin E_{OG}^0$ . In summary,  $E_{OG}^0 \subseteq E_{OG}$ , the graph defined by the local versions of Granger causality and contemporaneous correlation has fewer edges than the graph  $G_{OG}$  based on the classical Granger causality and contemporaneous correlation, and in general the graphs are not equal. Again, this can be seen in Fig. 1. The advantage of the graph  $G_{OG}^0$  based on the local version is that it allows to model more general graphs than  $G_{OG}$ .
- (d) In Definition 5.9, we have defined the orthogonality graph and the local orthogonality graph. Of course, it is also possible to define the *global orthogonality graph* based on global Granger causality and global contemporaneous correlation, but this is not part of this work. There are various reasons for this. On the one hand, the sparsity structure of the global orthogonality graph is very weak. The global orthogonality graph has even more edges than the orthogonality graph and the local orthogonality graph. Moreover, the orthogonality graph already satisfies the global AMP Markov and the global Markov property, as we are going to derive later in Section 5.3.2. These Markov properties can easily be transferred to the global orthogonality graph, the proofs are even easier. On the other hand, in specific models such as MCAR processes and state space models, Granger causality corresponds to global Granger causality, and contemporaneous correlation corresponds to contemporaneous correlation (cf. Remark 6.20), so that the global orthogonality graph is equal to the orthogonality graph and does not give any additional information.

5.3. Markov properties of (local) orthogonality graphs

The (local) orthogonality graph decodes directed and undirected relations between component series of the process  $Y_V$ . Conversely, a mixed graph can be associated with a set of constraints imposed on the stochastic process  $Y_V$ . Such a set of causal relations encoded by a graph is commonly known as a Markov property of the graph (cf. Lauritzen [50], Whittaker [74]). In this section, we introduce various levels of Markov properties. We start with the pairwise, local and block-recursive Markov properties. We then move on to two global Markov properties, namely the global AMP Markov property and the global Markov property.

5.3.1. Pairwise, local and block-recursive Markov property

Let us start with a few simple Markov properties that we expect from a graph. First of all, the (local) orthogonality graph visualises pairwise relationships between the components of a process  $Y_V$  by definition, that is the pairwise Markov property.

**Proposition 5.11.**

- (a) Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$ . Then  $Y_V$  satisfies the pairwise Markov property with respect to  $G_{OG}$ , i.e., for all  $a, b \in V$ ,  $a \neq b$ :

- (i)  $a \longrightarrow b \notin E_{OG} \Rightarrow Y_a \not\rightarrow Y_b \mid Y_V$ ,
- (ii)  $a --- b \notin E_{OG} \Rightarrow Y_a \sim Y_b \mid Y_V$ .

- (b) Let  $G_{OG}^0 = (V, E_{OG}^0)$  be the local orthogonality graph for  $Y_V$ . Then  $Y_V$  satisfies the pairwise Markov property with respect to  $G_{OG}^0$ .

Further, define  $pa(a) = \{v \in V \mid v \longrightarrow a \in E\}$  and  $ne(a) = \{v \in V \mid v --- a \in E\}$  as the set of parents and neighbours of  $a \in V$ , respectively. If we consider a vertex  $a \in V$ , then all vertices  $b \in V \setminus (pa(a) \cup \{a\})$  are Granger non-causal for  $a$ , i.e.,  $Y_b \not\rightarrow Y_a \mid Y_V$ . A direct consequence of Proposition 5.7(a) is then that  $Y_{V \setminus (pa(a) \cup \{a\})} \not\rightarrow Y_a \mid Y_V$  holds. The same applies to neighbours of  $a$  and the components being contemporaneously uncorrelated. Let  $a \in V$  and  $b \in V \setminus (ne(a) \cup \{a\})$ , then  $a --- b \notin E_{OG}$  and  $Y_b \sim Y_a \mid Y_V$ . Remark 4.4 yields  $Y_{V \setminus (ne(a) \cup \{a\})} \sim Y_a \mid Y_V$ . This is the local Markov property. The same arguments work for the local orthogonality graph using Proposition 5.7(b) and Remark 4.7, respectively.

**Proposition 5.12.**

- (a) Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$ . Then  $Y_V$  satisfies the local Markov property with respect to  $G_{OG}$ , i.e., for all  $a \in V$ :
  - (i)  $Y_{V \setminus (\text{pa}(a) \cup \{a\})} \not\rightarrow Y_a \mid Y_V$ ,
  - (ii)  $Y_{V \setminus (\text{ne}(a) \cup \{a\})} \approx Y_a \mid Y_V$ .
- (b) Let  $G_{OG}^0 = (V, E_{OG}^0)$  be the local orthogonality graph for  $Y_V$ . Then  $Y_V$  satisfies the local Markov property with respect to  $G_{OG}^0$ .

Furthermore, let  $\text{pa}(A) = \bigcup_{a \in A} \text{pa}(a)$  and  $\text{ne}(A) = \bigcup_{a \in A} \text{ne}(a)$  denote the set of all parents and neighbours of vertices in  $A \subseteq V$ . Again, we expect components that are not parents of  $A$  to be Granger non-causal for  $A$  and components that are not neighbours of  $A$  to be contemporaneously uncorrelated to  $A$ . This is the block-recursive Markov property and it also follows directly from Proposition 5.7, Remark 4.4 and Remark 4.7.

**Proposition 5.13.**

- (a) Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$ . Then  $Y_V$  satisfies the block-recursive Markov property with respect to  $G_{OG}$ , i.e., for all  $A \subseteq V$ :
  - (i)  $Y_{V \setminus (\text{pa}(A) \cup A)} \not\rightarrow Y_A \mid Y_V$ ,
  - (ii)  $Y_{V \setminus (\text{ne}(A) \cup A)} \approx Y_A \mid Y_V$ .
- (b) Let  $G_{OG}^0 = (V, E_{OG}^0)$  be the local orthogonality graph for  $Y_V$ . Then  $Y_V$  satisfies the block-recursive Markov property with respect to  $G_{OG}^0$ .

In our (local) orthogonality graph all three Markov properties are fulfilled. Thus, for example, using the local Markov property, we can infer from Fig. 1 that  $Y_{\{2,3\}} \not\rightarrow Y_1 \mid Y_{\{1,2,3\}}$  and  $Y_{\{2,3\}} \not\rightarrow_0 Y_1 \mid Y_{\{1,2,3\}}$ . However, the validity of Markov properties is not self-evident. For more information, see Eichler [29], Theorem 2.1 and Definition 2.3, who proposes to specify graphical time series models that satisfy the block-recursive Markov property as graphical time series models. For the visualisation of the various Markov properties at more complex examples than the one in Fig. 1, we also refer to Eichler [29], Example 2.1.

5.3.2. Global Markov properties for the orthogonality graph  $G_{OG} = (V, E_{OG})$

The three Markov properties we have discussed so far only encode relations with respect to  $Y_V$ . However, for a better understanding of the causal structure, we are interested in relations with respect to partial information. An intuitive analysis of orthogonality graphs suggests that paths between vertices may be associated with relations between corresponding components given only the information provided by a subprocess. To this end, we first introduce the global AMP Markov property of Andersson et al. [2], Definition 6, which relates paths in a graph to conditional orthogonality relations between variables. We then introduce the global Markov property, which provides sufficient criteria for Granger non-causality and contemporaneous uncorrelation. As we have to make additional assumptions for the local orthogonality graph, the results for the local model are presented in the next subsection, and here we only consider the orthogonality graph.

Let us start with the global AMP Markov property, where for  $A, B, C \subseteq V$  disjoint, the fact that  $A$  and  $B$  are separated given  $S$  implies that  $\mathcal{L}_A$  and  $\mathcal{L}_B$  are conditionally orthogonal given  $\mathcal{L}_S$ . But there are two main approaches to defining separation. The first approach is based on the path-oriented criterion “ $m$ -separation”. The second approach uses separation in undirected graphs by applying the operation of augmentation or moralisation to appropriate subgraphs (Eichler [28], Section 3). Since the second approach to defining a global Markov property is not straightforward in the sense that the graph is modified during the test, we just discuss the concept of  $m$ -separation and refer to Fassen-Hartmann and Schenk [33], who compare the augmented causality graph, the augmentation of the causality graph, with the path diagram, an undirected graphical model for continuous-time stationary processes. To define the latter, we start with some definitions from graph theory, which can be found in Eichler [28,29].

**Definition 5.14.** Let  $G = (V, E)$  be a mixed graph. A path  $\pi$  between two vertices  $a$  and  $b$  is a sequence  $\pi = \langle e_1, \dots, e_n \rangle$  of edges  $e_i \in E$ , such that  $e_i$  is an edge between  $v_{i-1}$  and  $v_i$  for some sequence of vertices  $a = v_0, v_1, \dots, v_n = b$ . We say that  $a$  and  $b$  are the endpoints of the path, while  $v_1, \dots, v_{n-1}$  are intermediate vertices.  $n$  is called length of the path. An intermediate vertex  $c$  on a path  $\pi$  is said to be a collider on the path, if the edges preceding and succeeding  $c$  on the path both have an arrowhead or a dashed tail at  $c$ , i.e.,  $\rightarrow c \leftarrow$ ,  $\rightarrow c \text{ ---}$ ,  $\text{---} c \leftarrow$ ,  $\text{---} c \text{ ---}$ . Otherwise the vertex  $c$  is said to be a non-collider on the path. A path  $\pi$  between vertices  $a$  and  $b$  is said to be  $m$ -connecting given a set  $S$  if

- (a) every non-collider on the path is not in  $S$ , and
- (b) every collider on the path is in  $S$ ,

otherwise we say the path is  $m$ -blocked given  $S$ . If all paths between  $a$  and  $b$  are  $m$ -blocked given  $S$ , then  $a$  and  $b$  are said to be  $m$ -separated given  $S$ . Similarly, sets  $A$  and  $B$  are said to be  $m$ -separated in  $G$  given  $S$ , denoted by  $A \bowtie_m B \mid S [G]$ , if for every pair  $a \in A$  and  $b \in B$ ,  $a$  and  $b$  are  $m$ -separated given  $S$ .

The  $m$ -separation is the natural extension of the  $d$ -separation for directed graphs (cf. Pearl [61]) to mixed graphs (cf. Richardson [65]), and was earlier also called  $d$ -separation by Spirtes et al. [72], Koster [47]. Since we consider mixed graphs, which are generally not directed, we prefer the notion of  $m$ -separation. For a motivation and visualisation of the respective definitions, we also refer to these papers. Note that condition (a) differs from the original definition of  $m$ -connecting paths in Richardson [65] and takes into account that we consider paths that can intersect themselves, as in Eichler [28]. Nevertheless, the concepts of  $m$ -separation here and in Richardson [65] are equivalent. In contrast, Eichler [29] uses another natural extension of  $d$ -separation, called  $p$ -separation and introduced by Levitz et al. [52] for chain graphs, where  $---c---$  is considered a non-collider. Let us present the main result, the global AMP Markov property.

**Theorem 5.15.** *Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$ . Then  $Y_V$  satisfies the global AMP Markov property with respect to  $G_{OG}$ , i.e., for all disjoint subsets  $A, B, C \subseteq V$ ,*

$$A \bowtie_m B \mid C [G_{OG}] \Rightarrow \mathcal{L}_{Y_A} \perp \mathcal{L}_{Y_B} \mid \mathcal{L}_{Y_C}.$$

In words, if the sets  $A$  and  $B$  are  $m$ -separated given  $C$ , then  $Y^A \in \mathcal{L}_{Y_A}$  and  $Y^B \in \mathcal{L}_{Y_B}$  are uncorrelated after removing all of the (linear) information provided by  $\mathcal{L}_{Y_C}$ . A visualisation of the global AMP Markov property at a typical mixed graph is illustrated in Eichler [29], Example 2.1, which can also be found in several of his articles. The proof of Theorem 5.15 is structured into three auxiliary statements that culminate in the actual proof, see Appendix B.2. Note that in the latter we need Assumption 2 for the first time.

**Remark 5.16.** Similar statements can be found, e.g., in Eichler [27], Theorem 4.8, Eichler [28], Theorem 3.1 or Eichler [29], Theorem 4.1. However, the graphs defined there are based on different definitions of the edges and on processes in discrete time. The definition of the undirected edges in Eichler [29] further differs from our definition. The linear continuous-time analogue of his definition is that  $\mathcal{L}_{Y_A}(t, t+1) \perp \mathcal{L}_{Y_B}(t, t+1) \mid \mathcal{L}_{Y_S}(t) \vee \mathcal{L}_{S \setminus (A \cup B)}(t, t+1)$ . Still most of the proofs can be carried over because it makes no difference whether one adds  $\mathcal{L}_{S \setminus (A \cup B)}(t, t+1)$  or not.

The concept of  $m$ -separation provides a sufficient criterion for conditional orthogonality. However, we would also like to derive sufficient graphical conditions for Granger non-causality and processes being contemporaneously uncorrelated. An obvious first idea would be to start again with  $m$ -separation. However, this condition is stronger than necessary. A motivating example to only consider paths that point in the “right” direction is provided by Eichler [28], p. 341. We introduce further graph-theoretic notions and then provide the main result.

**Definition 5.17.** Let  $G = (V, E)$  be a mixed graph. A path  $\pi$  between vertices  $a$  and  $b$  is called  $b$ -pointing if it has an arrowhead at the endpoint  $b$ . More generally, a path  $\pi$  between  $A$  and  $B$  is said to be  $B$ -pointing if it is  $b$ -pointing for some  $b \in B$ . Furthermore, a path  $\pi$  between vertices  $a$  and  $b$  is said to be  $bi$ -pointing if it has an arrowhead at both endpoints  $a$  and  $b$ .

**Theorem 5.18.** *Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$ . Then  $Y_V$  satisfies the global Markov property with respect to  $G_{OG}$ , i.e., for all disjoint subsets  $A, B, C \subseteq V$  the following conditions hold:*

- (a) *If every  $B$ -pointing path in  $G_{OG}$  between  $A$  and  $B$  is  $m$ -blocked given  $B \cup C$  then  $Y_A \not\rightarrow Y_B \mid Y_{A \cup B \cup C}$ .*
- (b) *If  $a --- b \notin E_{OG}$  for all  $a \in A$  and  $b \in B$ , and if every  $bi$ -pointing path in  $G_{OG}$  between  $A$  and  $B$  is  $m$ -blocked given  $A \cup B \cup C$ , then  $Y_A \approx Y_B \mid Y_{A \cup B \cup C}$ .*

A similar result in discrete time can be found in Eichler [28], Theorems 4.1 and 4.2, and Eichler [29], Theorem 4.2. For the visualisation of the global AMP Markov property at some mixed graph, we also refer to Eichler [29], Example 2.1. Because of the properties of a graphoid in Lemma 2.2, the block-recursive Markov property in Proposition 5.13 and Lemma B.2, the proof can be carried out similarly as in Eichler [28,29], respectively, and is therefore skipped.

As a consequence of the global Markov property, we find that the  $m$ -separation  $A \bowtie_m B \mid C [G_{OG}]$  is indeed too strong implying causality in both directions between  $Y_A$  and  $Y_B$  as well as their contemporaneous uncorrelation. We refer to Eichler [29], Corollary 4.1, and Eichler [28], Corollary 4.3 for the proof.

**Corollary 5.19.** *Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$  and let  $A, B, C \subseteq V$  be disjoint subsets. Then  $A \bowtie_m B \mid C [G_{OG}]$  implies*

$$Y_A \not\rightarrow Y_B \mid Y_{A \cup B \cup C}, \quad Y_B \not\rightarrow Y_A \mid Y_{A \cup B \cup C}, \quad \text{and} \quad Y_A \approx Y_B \mid Y_{A \cup B \cup C}.$$

### 5.3.3. Global Markov properties for the local orthogonality graph $G_{OG}^0 = (V, E_{OG}^0)$

For the local orthogonality graph, the global Markov properties are, as expected, much more difficult due to the weaker definition of the edges. However, we still derive sufficient graphical conditions for local Granger non-causality and local contemporaneous uncorrelation. At least under additional assumptions, the property of  $m$ -separation implies local Granger non-causality in both directions between  $Y_A$  and  $Y_B$ , and that they are locally contemporaneously uncorrelated. We start with a special case where  $C = V \setminus (A \cup B)$ . The proofs of this subsection are given in Appendix B.3.

**Proposition 5.20.** Let  $G_{OG}^0 = (V, E_{OG}^0)$  be the local orthogonality graph for  $Y_V$  and  $A, B \subseteq V$  with  $A \cap B = \emptyset$ . Then  $A \bowtie_m B \mid V \setminus (A \cup B) [G_{OG}^0]$  implies

$$Y_A \not\rightarrow_0 Y_B \mid Y_V, \quad Y_B \not\rightarrow_0 Y_A \mid Y_V, \quad \text{and} \quad Y_A \approx_0 Y_B \mid Y_V.$$

We consider a second special case where the block-recursive Markov property already leads to local Granger non-causality and local contemporaneous uncorrelation.

**Proposition 5.21.** Let  $G_{OG}^0 = (V, E_{OG}^0)$  be the local orthogonality graph for  $Y_V$  and let  $A, B, C \subseteq V$  be disjoint subsets. Suppose  $pa(A) \cup pa(B) \subseteq A \cup B \cup C$ . Then  $A \bowtie_m B \mid C [G_{OG}^0]$  implies

$$Y_A \not\rightarrow_0 Y_B \mid Y_{A \cup B \cup C}, \quad Y_B \not\rightarrow_0 Y_A \mid Y_{A \cup B \cup C}, \quad \text{and} \quad Y_A \approx_0 Y_B \mid Y_{A \cup B \cup C}.$$

**Remark 5.22.**

- (a)  $an(A \cup B \cup C) = A \cup B \cup C$  implies  $pa(A) \cup pa(B) \subseteq A \cup B \cup C$ . Therefore, we also have a graphical condition for non-causality and contemporaneous uncorrelation for ancestral subsets.
- (b)  $pa(B) \subseteq A \cup B \cup C$  is sufficient for  $Y_A \not\rightarrow_0 Y_B \mid Y_{A \cup B \cup C}$ .  
For the proof of [Proposition 5.21](#), we need the left decomposition property of local Granger non-causality.

**Lemma 5.23.** Let  $A, B, C, D \subseteq V$  be disjoint subsets. Then

$$Y_{A \cup B} \not\rightarrow_0 Y_C \mid Y_{A \cup B \cup C \cup D} \Rightarrow Y_A \not\rightarrow_0 Y_C \mid Y_{A \cup C \cup D}.$$

**Remark 5.24.**

- (a) The right decomposition property, which is that

$$Y_A \not\rightarrow_0 Y_{B \cup C} \mid Y_{A \cup B \cup C \cup D} \Rightarrow Y_A \not\rightarrow_0 Y_B \mid Y_{A \cup B \cup D}$$

cannot be expected. This can be explained as follows: It is possible that  $Y_A$  is non-causal for  $Y_{B \cup C}$  given  $Y_{A \cup B \cup C \cup D}$ , since the corresponding information of  $Y_A$  is already present in  $Y_C$ . However, if  $Y_C$  is omitted, there may be causal influence of  $Y_A$  on  $Y_B$ . This topic has been addressed, e.g., by Didelez [\[21\]](#) in the context of directed graphs.

- (b) The lack of right decomposability is the key problem when trying to derive the global Markov property from the block-recursive Markov property. In the case that  $A \cup B \cup C \subset V$ , Corollary 1 and Proposition 2 of Koster [\[47\]](#) yield

$$A \bowtie_m B \mid C [G_{OG}^0] \Leftrightarrow A' \bowtie_m B' \mid C [G_{OG, an(A \cup B \cup C)}^0],$$

for disjoint subsets  $A'$  and  $B'$  with  $A \subseteq A'$ ,  $B \subseteq B'$  and  $A' \cup B' \cup C = an(A \cup B \cup C)$  as in the proof of [Theorem 5.15](#). According to [Proposition 5.20](#), we can conclude

$$Y_{A'} \not\rightarrow_0 Y_{B'} \mid Y_{A' \cup B' \cup C}, \quad Y_{B'} \not\rightarrow_0 Y_{A'} \mid Y_{A' \cup B' \cup C} \quad \text{and} \quad Y_{A'} \approx_0 Y_{B'} \mid Y_{A' \cup B' \cup C},$$

in  $[G_{OG, an(A \cup B \cup C)}^0]$ . Since the definition of local Granger non-causality and local contemporaneous uncorrelation does not depend on whether we choose the subgraph with vertices in  $A' \cup B' \cup C$  or the whole graph with vertices in  $V$ , the statements also hold for  $[G_{OG}^0]$ . But to obtain from this, e.g.,  $Y_A \not\rightarrow_0 Y_B \mid Y_{A \cup B \cup C}$ , we not only need the left decomposability but also the right decomposability.

## 6. Orthogonality graphs for MCAR processes

To gain a deeper understanding of the theoretical concept of a (local) orthogonality graph, we apply the graphical models to the class of causal MCAR processes. We not only give theoretical results but also interpret them and relate them to the results of Eichler [\[28\]](#) in discrete time. First, we give a brief introduction to MCAR processes and show that they satisfy the assumptions of the (local) orthogonality graph. We then derive linear predictors of MCAR processes, which we require to characterise the edges; which is the ultimate goal of this section. The details of the proofs of this section are moved to [Appendix C](#).

### 6.1. MCAR processes

A multivariate  $k$ -dimensional continuous-time AR (MCAR) process is a continuous-time version of the well-known vector AR (VAR) process in discrete time. The driving process is a  $k$ -dimensional Lévy process  $(L(t))_{t \in \mathbb{R}}$  as defined in [Example 3.15](#) and satisfies the following assumption throughout the paper.

**Assumption 3.** The two-sided Lévy process  $L = (L(t))_{t \in \mathbb{R}}$  satisfies  $\mathbb{E}L(1) = 0_k$  and  $\mathbb{E}\|L(1)\|^2 < \infty$  with  $\Sigma_L = \mathbb{E}[L(1)L(1)^\top]$ .



The idea is then that a  $k$ -dimensional MCAR( $p$ ) process is the solution of the stochastic differential equation

$$P(D)Y(t) = DL(t) \quad \text{for } t \in \mathbb{R}, \tag{6.1}$$

where  $D$  is the differential operator with respect to  $t$  and

$$P(\lambda) := I_k \lambda^p + A_1 \lambda^{p-1} + \dots + A_p, \quad \lambda \in \mathbb{C}, \tag{6.2}$$

is the autoregressive polynomial, respectively with  $A_1, \dots, A_p \in M_k(\mathbb{R})$ . However, this is not the formal definition of an MCAR process, since a Lévy process is not differentiable. The formal definition of a Lévy-driven causal MCAR process used here goes back to Marquardt and Stelzer [56], Definition 3.20. However, one-dimensional Gaussian CARMA processes were already investigated by Doob [24] (cf. Doob [25]) and Lévy-driven CARMA processes were propagated by Peter Brockwell at the beginning of this century, see Brockwell [8], Brockwell and Lindner [11] for an overview. Very early Gaussian MCAR processes were already studied in the economics literature, e.g., in Harvey and Stock [42,43,44] and were further explored in the well-known paper of Bergstrom [4].

**Definition 6.1.** Let  $(L(t))_{t \in \mathbb{R}}$  be a two sided  $k$ -dimensional Lévy process. Further, let  $\mathbf{A} \in M_{kp}(\mathbb{R})$ ,  $p \geq 1$  with  $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$ , such that

$$\mathbf{A} = \begin{pmatrix} 0_k & I_k & 0_k & \dots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \dots & \dots & 0_k & I_k \\ -A_p & -A_{p-1} & \dots & \dots & -A_1 \end{pmatrix},$$

$\mathbf{B}^\top = (0_k, \dots, 0_k, I_k) \in M_{k \times kp}(\mathbb{R})$  and  $\mathbf{C} = (I_k, 0_k, \dots, 0_k) \in M_{kp \times k}(\mathbb{R})$ . Then the process  $Y_V = (Y_V(t))_{t \in \mathbb{R}}$  given by

$$Y_V(t) = \mathbf{C}X(t),$$

where  $X = (X(t))_{t \in \mathbb{R}}$  is the unique  $kp$ -dimensional stationary solution of the state equation

$$dX(t) = \mathbf{A}X(t)dt + \mathbf{B}dL(t), \tag{6.3}$$

is called (causal) MCAR( $p$ ) process.

Indeed, if  $p = 1$ , the MCAR(1) process corresponds to the Ornstein–Uhlenbeck process of Example 3.15. We summarise important properties of causal MCAR processes used in this paper. Details are given in Marquardt and Stelzer [56], Schlemm and Stelzer [68].

**Lemma 6.2.** Let  $Y_V$  be a causal MCAR( $p$ ) process. Then the following results hold:

(a) The unique stationary solution  $X$  of the state Eq. (6.3) has the representation

$$X(t) = \int_{-\infty}^t e^{\mathbf{A}(t-u)} \mathbf{B}dL(u), \quad t \in \mathbb{R},$$

and

$$X(t) = e^{\mathbf{A}(t-s)} X(s) + \int_s^t e^{\mathbf{A}(t-u)} \mathbf{B}dL(u), \quad s, t \in \mathbb{R}, s < t.$$

(b) We denote the  $j$ -th  $k$ -block of  $X$  by

$$X^{(j)}(t) = \begin{pmatrix} X_{(j-1)k+1}(t) \\ \vdots \\ X_{jk}(t) \end{pmatrix}, \quad t \in \mathbb{R}, j = 1, \dots, p, \tag{6.4}$$

such that  $X(t) = (X^{(1)}(t)^\top, \dots, X^{(p)}(t)^\top)^\top$ ,  $t \in \mathbb{R}$ . Suppose  $\Phi_L(\cdot)$  is the  $k$ -dimensional random orthogonal measure of the Lévy process  $L$ , i.e.,

$$\Phi_L([a, b]) = \int_{-\infty}^{\infty} \frac{e^{-i\lambda a} - e^{-i\lambda b}}{2\pi i \lambda} dL(\lambda), \quad -\infty < a < b < \infty,$$

with spectral measure  $F_L(d\lambda) = \Sigma_L/2\pi d\lambda$  and  $\mathbb{E}(\Phi_L([a, b])) = 0_k$ . Then

$$X^{(j)}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} (i\lambda)^{j-1} P(i\lambda)^{-1} \Phi_L(d\lambda), \quad t \in \mathbb{R},$$

and in particular,  $Y_V(t) = X^{(1)}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} P(i\lambda)^{-1} \Phi_L(d\lambda)$ ,  $t \in \mathbb{R}$ .

(c) The covariance function  $(c_{XX}(t))_{t \in \mathbb{R}}$  of  $X$  is

$$c_{XX}(t) = c_{XX}(-t)^\top = \mathbb{E}[X(t+h)\overline{X(h)}^\top] = e^{\mathbf{A}t} \Gamma(0), \quad t \geq 0, \tag{6.5}$$

where  $\Gamma(0) = \int_0^\infty e^{\mathbf{A}u} \mathbf{B} \Sigma_L \mathbf{B}^\top e^{\mathbf{A}^\top u} du$  satisfies

$$\mathbf{A} \Gamma(0) + \Gamma(0) \mathbf{A}^\top = -\mathbf{B} \Sigma_L \mathbf{B}^\top. \tag{6.6}$$

(d) The spectral density of the causal MCAR process  $Y_V$  is

$$f_{Y_V Y_V}(\lambda) = \frac{1}{2\pi} P(i\lambda)^{-1} \Sigma_L (P(-i\lambda)^{-1})^\top = \frac{1}{2\pi} \mathbf{C} (i\lambda I_{kp} - \mathbf{A})^{-1} \mathbf{B} \Sigma_L \mathbf{B}^\top (-i\lambda I_{kp} - \mathbf{A}^\top)^{-1} \mathbf{C}^\top, \quad \lambda \in \mathbb{R}.$$

We point out some more properties that we use later in the paper.

**Remark 6.3.**

- (a) If  $\Sigma_L > 0$ , then  $c_{XX}(0) > 0$ . Indeed,  $\mathbf{B}$  is of full rank and thus the assumptions of Schlemm and Stelzer [68], Corollary 3.9, are satisfied.
- (b) Since the matrix exponential is continuous, we have  $c_{XX}(t) \rightarrow c_{XX}(0)$  for  $t \rightarrow 0$ . Now,  $c_{Y_V Y_V}(\cdot)$  corresponds to the upper left  $k \times k$  block of  $c_{XX}(\cdot)$ . Thus,  $c_{Y_V Y_V}(t) \rightarrow c_{Y_V Y_V}(0)$  for  $t \rightarrow 0$ . Cramér [17], Lemma 1, then gives that the causal MCAR process  $Y_V$  is mean-square continuous.

For the definition of the local orthogonality graph and, in particular, the local Granger non-causality and the local contemporaneous uncorrelation, respectively, we need some knowledge about the existence and the description of the mean-square derivatives of the MCAR process. Therefore, we note the following.

**Remark 6.4.** Due to the spectral representation of  $X^{(j)}$  given in (6.4), we directly obtain the spectral density

$$f_{X^{(j)} X^{(j)}}(\lambda) = \frac{1}{2\pi} (i\lambda)^{j-1} P(i\lambda)^{-1} \Sigma_L (P(-i\lambda)^{-1})^\top (-i\lambda)^{j-1}, \quad \lambda \in \mathbb{R}.$$

Therefore, it holds that  $\int_{-\infty}^{\infty} \lambda^2 \|f_{X^{(j)} X^{(j)}}(\lambda)\| d\lambda < \infty$  for  $j = 1, \dots, p-1$ , but  $\int_{-\infty}^{\infty} \lambda^2 \|f_{X^{(p)} X^{(p)}}(\lambda)\| d\lambda = \infty$ . Thus, a conclusion of Proposition 2.5 is that the process  $X^{(j)}$  is mean-square differentiable with derivative

$$D^{(1)} X^{(j)}(t) = X^{(j+1)}(t), \quad j = 1, \dots, p-1, \tag{6.7}$$

while for  $X^{(p)}$  the mean-square derivative does not exist. With  $Y_V(t) = X^{(1)}(t)$  in mind, we receive iteratively from (6.7) that  $Y_V$  is  $(p-1)$ -times mean-square differentiable with

$$D^{(j)} Y_V(t) = X^{(j+1)}(t), \quad j = 1, \dots, p-1, \tag{6.8}$$

but the  $p$ -th derivative does not exist. By the same arguments, we receive that for any component  $Y_v$ ,  $v \in V$ , of  $Y_V$  there is no derivative higher than  $(p-1)$ .

6.2. Orthogonality graph for MCAR processes

In the following, we verify that the (local) orthogonality graph for the MCAR process is well-defined. Therefore, we have to check that the Assumptions 1 and 2 are satisfied.

**Proposition 6.5.** Let  $Y_V$  be a causal MCAR( $p$ ) process with  $\Sigma_L > 0$ . Then  $Y_V$  satisfies Assumptions 1 and 2.

The proof of Assumption 1 is elaborate, using results from matrix analysis and functional analysis (Bernstein [5], Bhatia [6], Bühler and Salamon [12]), and is therefore presented in the Supplementary Material F. However, the basic idea is simple. Note,  $\Sigma_L > 0$  results in  $f_{Y_V Y_V}(\cdot) > 0$ . On the one hand, we prove that an epsilon bound can always be found on compact intervals. On the other hand, the matrix function converges to a boundary matrix which can also be bounded. Together this then gives Assumption 1. The proof of Assumption 2 is also given in the Supplementary Material F and is based on a characterisation of purely non-deterministic processes by limits of orthogonal projections. It was expected that the MCAR( $p$ ) process would satisfy this assumption since in our case the driving Lévy process has no drift term. Since Assumptions 1 and 2 hold, a direct consequence of Section 5 is then the following.

**Proposition 6.6.** Let  $Y_V$  be a causal MCAR( $p$ ) process with  $\Sigma_L > 0$ . If we define  $V = \{1, \dots, k\}$  as the vertices and the edges  $E_{OG}$  via

- (i)  $a \longrightarrow b \notin E_{OG} \iff Y_a \not\rightarrow Y_b \mid Y_V$ ,
- (ii)  $a \dashrightarrow b \notin E_{OG} \iff Y_a \approx Y_b \mid Y_V$ ,

for  $a, b \in V$ ,  $a \neq b$ , then the orthogonality graph  $G_{OG} = (V, E_{OG})$  for the MCAR process  $Y_V$  is well-defined and satisfies the pairwise, local, block-recursive, global AMP and global Markov property.

If we look at the local orthogonality graph, we also get the following from Section 5.

**Proposition 6.7.** Let  $Y_V$  be a causal MCAR( $p$ ) process with  $\Sigma_L > 0$ . If we define  $V = \{1, \dots, k\}$  as the vertices and the edges  $E_{OG}^0$  via

- (i)  $a \longrightarrow b \notin E_{OG}^0 \Leftrightarrow Y_a \not\rightarrow_0 Y_b \mid Y_V$ ,
- (ii)  $a \dashrightarrow b \notin E_{OG}^0 \Leftrightarrow Y_a \approx_0 Y_b \mid Y_V$ ,

for  $a, b \in V$ ,  $a \neq b$ , then the local orthogonality graph  $G_{OG}^0 = (V, E_{OG}^0)$  for the MCAR process  $Y_V$  is well-defined and satisfies the pairwise, local and block-recursive Markov property. Furthermore, the statements of [Propositions 5.20](#) and [5.21](#) hold.

### 6.3. Prediction of MCAR processes

To characterise the different Granger causalities and contemporaneous correlations as is done, e.g., in [Theorems 3.5](#) and [4.5](#), respectively, we need to compute the linear predictions of the MCAR process and its derivatives on the different subspaces. To do this, we first give a suitable representation for  $Y_v(t+h)$ . [Appendix C.1](#) contains all proofs of this subsection.

**Lemma 6.8.** Let  $Y_V$  be a causal MCAR( $p$ ) process. Further, let  $t \in \mathbb{R}$ ,  $h \geq 0$ , and  $v \in V$ . Then

$$Y_v(t+h) = e_v^\top C e^{Ah} \sum_{j=1}^p E_j D^{(j-1)} Y_V(t) + e_v^\top C \int_t^{t+h} e^{A(t+h-u)} \mathbf{B} dL(u) \quad \mathbb{P}\text{-a.s.}$$

From this representation of  $Y_v(t+h)$  we conclude that on the one hand, the past ( $Y_V(s)$ ,  $s \leq t$ ) of all components and on the other hand, the future of the Lévy process ( $L(t+h) - L(s)$ ,  $t \leq s \leq t+h$ ) are relevant for  $Y_v(t+h)$ . Based on this knowledge, we specify the orthogonal projections.

**Proposition 6.9.** Let  $Y_V$  be a causal MCAR( $p$ ) process. Further, let  $t \in \mathbb{R}$ ,  $h \geq 0$ ,  $S \subseteq V$ , and  $v \in V$ . Then

$$P_{\mathcal{L}_{Y_S}(t)} Y_v(t+h) = e_v^\top C e^{Ah} \sum_{s \in S} \sum_{j=1}^p E_j e_s D^{(j-1)} Y_s(t) + e_v^\top C e^{Ah} \sum_{s \in V \setminus S} \sum_{j=1}^p E_j e_s P_{\mathcal{L}_{Y_S}(t)} D^{(j-1)} Y_s(t) \quad \mathbb{P}\text{-a.s.}$$

According to [Lemma 6.8](#), the basic idea of the proof is simple:  $Y_s(t)$  and its derivatives are already in  $\mathcal{L}_{Y_S}(t)$  (see [Remark 2.6](#)) and are therefore projected onto themselves. Additionally,  $\sigma(Y_S(s) : s \leq t)$  and  $\sigma(L(t+h) - L(s) : t \leq s \leq t+h)$  are independent and thus,  $e_v^\top C \int_t^{t+h} e^{A(t+h-u)} \mathbf{B} dL(u)$  is projected on zero.

**Remark 6.10.** For  $S = V$  we get the explicit representation

$$P_{\mathcal{L}_{Y_V}(t)} Y_v(t+h) = e_v^\top C e^{Ah} \sum_{s \in V} \sum_{j=1}^p E_j e_s D^{(j-1)} Y_s(t) = e_v^\top C e^{Ah} X(t),$$

as in Brockwell and Lindner [[10](#)] for univariate CARMA processes. For an explicit representation in the case  $S \subset V$  the methods in Rozanov [[67](#)], III, 5, can be applied but this is quite elaborate.

Next, we calculate the projections of  $D^{(p-1)} Y_V$ , which we require for the characterisation of local Granger causality and local contemporaneous correlation.

**Lemma 6.11.** Let  $Y_V$  be a causal MCAR( $p$ ) process. Further, let  $t \in \mathbb{R}$ ,  $h \geq 0$ ,  $S \subseteq V$ , and  $v \in V$ . Then

$$\begin{aligned} & P_{\mathcal{L}_{Y_S}(t)} (D^{(p-1)} Y_v(t+h) - D^{(p-1)} Y_v(t)) \\ &= e_v^\top \mathbf{E}_p^\top (e^{Ah} - I_{kp}) \sum_{s \in S} \sum_{j=1}^p E_j e_s D^{(j-1)} Y_s(t) + e_v^\top \mathbf{E}_p^\top (e^{Ah} - I_{kp}) \sum_{s \in V \setminus S} \sum_{j=1}^p E_j e_s P_{\mathcal{L}_{Y_S}(t)} D^{(j-1)} Y_s(t) \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

and

$$D^{(p-1)} Y_v(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(p-1)} Y_v(t+h) = e_v^\top \mathbf{E}_p^\top \int_t^{t+h} e^{A(t+h-u)} \mathbf{B} dL(u) \quad \mathbb{P}\text{-a.s.}$$

### 6.4. Characterisation of the directed and undirected influences for the MCAR process

In this subsection, we focus on criteria for the directed and undirected influences for causal MCAR( $p$ ) processes. All proofs of this subsection are carried out in [Appendix C.2](#). We start with a characterisation of (local) Granger causality for an MCAR process, which is well suited for interpretation and for comparison with Eichler [[28](#)] in discrete time. The proofs are based on the characterisation of (local) Granger causality in [Theorem 3.5](#) using the orthogonal projections from Section 6.3. Note that for the definition of local Granger causality and local contemporaneous correlation, we use that all components of  $Y_V$  are  $(p-1)$ -times mean square differentiable, but the  $p$ -th derivative does not exist (cf. [Remark 6.4](#)), so that  $j_v = p-1$  for any  $v \in V$ .

**Proposition 6.12.** Let  $Y_V$  be a causal MCAR( $p$ ) process with  $\Sigma_L > 0$ . Further, let  $a, b \in V$  and  $a \neq b$ . Then the following holds.

- (a)  $Y_a \not\rightarrow Y_b \mid Y_V \Leftrightarrow \left[ \mathbf{C}e^{A\bar{h}}\mathbf{E}_j \right]_{ba} = \left[ e^{A\bar{h}} \right]_{b, k(j-1)+a} = 0 \quad \forall h \in [0, 1], j = 1, \dots, p.$
- (b)  $Y_a \not\rightarrow_0 Y_b \mid Y_V \Leftrightarrow \left[ \mathbf{E}_p^T \mathbf{A} \mathbf{E}_j \right]_{ba} = \left[ A_j \right]_{ba} = 0 \quad \forall j=1, \dots, p.$

These characterisations of (local) Granger causality are convenient since we no longer need to compute and compare orthogonal projections. Moreover, the deterministic criteria depend only on the state transition matrix  $\mathbf{A}$  and not on the driving Lévy process.

Let us now move on to contemporaneous uncorrelation and also give a first characterisation specifically related to the structure of an MCAR( $p$ ) process. Similar to Proposition 6.12, the proof is based on the characterisation of contemporaneous uncorrelation by orthogonal projections from Section 6.3 and (4.2).

**Proposition 6.13.** Let  $Y_V$  be a causal MCAR( $p$ ) process. Further, let  $a, b \in V$  and  $a \neq b$ . Then the following holds.

- (a)  $Y_a \approx Y_b \mid Y_V \Leftrightarrow \left[ \int_0^{\min(h, \bar{h})} \mathbf{C}e^{A(h-u)} \mathbf{B} \Sigma_L \mathbf{B}^T e^{A^T(\bar{h}-u)} \mathbf{C}^T du \right]_{ab} = 0 \quad \forall h, \bar{h} \in [0, 1].$
- (b)  $Y_a \approx_0 Y_b \mid Y_V \Leftrightarrow \left[ \Sigma_L \right]_{ab} = 0.$

**Remark 6.14.**

- (a) Comte and Renault [15] investigate non-stationary Brownian motion driven MCAR processes on local Granger causality and local instantaneous causality, which are similar to our concepts of local Granger causality and local contemporaneous correlation. In their Proposition 20, Comte and Renault [15] obtain that  $Y_a$  does not locally Granger cause  $Y_b$  if and only if  $[A_j]_{ba} = 0$ , for  $j = 1, \dots, p$ , as in our Proposition 6.12. Furthermore, there is no local instantaneous causality between  $Y_a$  and  $Y_b$  if and only if  $[\Sigma_L]_{ab} = 0$ , as in Proposition 6.13 for the local orthogonality graph. Statements about local Granger causality and local instantaneous causality for subprocesses under possible partial information, as we present with the Markov properties in Section 5.3, are not available there.
- (b) Furthermore, as a generalisation of Didelez [21], Mogensen and Hansen [60] study the local independence graph for Itô processes where the graph models the local independence structure of the underlying stochastic process; in contrast, we model local orthogonality. A special case is the Brownian motion driven Ornstein–Uhlenbeck process. The edges of the local independence graph of a Brownian motion driven Ornstein–Uhlenbeck process (cf. Proposition 7 in Mogensen and Hansen [60]) are the same as given here in Propositions 6.12 and 6.13, i.e., there is no directed edge from  $a$  to  $b$  if and only if  $[A]_{ba} = 0$ , and there is no undirected edge between  $a$  and  $b$  if and only if  $[\Sigma_L]_{ab} = 0$ . Thus, in the case of a Brownian motion driven Ornstein–Uhlenbeck process, the local independence graph and our conditional orthogonality graph coincide.
- (c) In both papers Comte and Renault [15] and Mogensen and Hansen [60], it is important to have Brownian motion driven Itô processes to receive the dependence structure of the underlying processes. Since for Gaussian models conditional orthogonality and conditional independence are equivalent, it is not surprising that we obtain the same edge characterisations as there for Gaussian driven Ornstein–Uhlenbeck processes. However, it will be a challenging task to extend the results in Comte and Renault [15] and Mogensen and Hansen [60] to Lévy-driven Itô processes. Our approach is able to fill this gap by presenting a graphical model for Lévy-driven MCAR( $p$ ) processes that moves away from the Gaussian assumption and  $p \geq 2$  but is still consistent with the existing literature and satisfies some Markov properties.

Let us compare our results for the continuous-time multivariate AR process with the results for discrete-time vector AR (VAR) processes of Eichler [28], whose article provided the basis for our considerations. We start with the local orthogonality graph because the comparison is obvious there.

**Remark 6.15.** The  $k$ -dimensional VAR( $p$ ) process  $Z_V = (Z_V(t))_{t \in \mathbb{Z}}$  is defined as

$$Z_V(t+1) = \sum_{n=1}^p \Phi_n Z_V(t+1-n) + \varepsilon(t+1), \quad t \in \mathbb{Z}, \tag{6.9}$$

where  $\varepsilon = (\varepsilon(t))_{t \in \mathbb{Z}}$  is a  $k$ -dimensional white noise process with non-singular covariance matrix  $\Sigma_\varepsilon \in M_k(\mathbb{R})$  and autoregressive coefficients  $\Phi_n \in M_k(\mathbb{R})$ ,  $n = 1, \dots, p$ . Further, define the AR-polynomial  $\Phi(\lambda) = I_k + \Phi_1 \lambda + \dots + \Phi_p \lambda^p$ ,  $\lambda \in \mathbb{C}$ , and denote by  $B$  the backshift operator. Then

$$\Phi(B)Z_V(t) = \varepsilon(t),$$

which corresponds to the idea for an MCAR( $p$ ) process to be the solution of the stochastic differential equation

$$P(D)Y_V(t) = DL(t),$$

where  $P(\lambda) = I_k \lambda^p + A_1 \lambda^{p-1} + \dots + A_p$ ,  $\lambda \in \mathbb{C}$ . Let  $G = (V, E)$  be the path diagram of  $Z_V$  as defined in Eichler [28].

- (a) *Directed edges:* Lemma 2.3 and Definition 2.1 in Eichler [28] state that the directed edges in the path diagram  $G$  of the discrete-time VAR( $p$ ) process  $Z_V$  satisfy

$$Z_a \not\rightarrow Z_b \mid Z_V \Leftrightarrow a \longrightarrow b \notin E \Leftrightarrow [\Phi_j]_{ba} = 0, \quad j = 1, \dots, p.$$

However, this is again in analogy to the characterisation of directed edges in the local orthogonality graph  $G_{OG}^0$  of an MCAR( $p$ ) processes where

$$Y_a \not\rightarrow_0 Y_b \mid Y_V \Leftrightarrow a \longrightarrow b \notin E_{OG}^0 \Leftrightarrow [A_j]_{ba} = 0, \quad j = 1, \dots, p.$$

In summary, both continuous and discrete-time models have in common that there is no directed edge between components  $a$  and  $b$  if and only if the  $ba$ -th components of the autoregressive coefficients are zero.

(b) *Undirected edges:* On the other hand, for the undirected edges in the path diagram  $G$  of the VAR( $p$ ) process  $Z_V$ , Lemma 2.3 and Definition 2.1 in Eichler [28] give the equivalence

$$Z_a \sim Z_b \mid Z_V \Leftrightarrow a \dashrightarrow b \notin E \Leftrightarrow [\Sigma_\varepsilon]_{ab} = 0.$$

However, this is again in analogy to the condition for the undirected edges in the local orthogonality graph  $G_{OG}^0$  where

$$Y_a \sim_0 Y_b \mid Y_V \Leftrightarrow a \dashrightarrow b \notin E_{OG}^0 \Leftrightarrow [\Sigma_L]_{ab} = 0.$$

Thus, a common feature of the continuous-time and discrete-time model is that there is no undirected edge between components  $a$  and  $b$  if and only if the  $a$ -th and  $b$ -th components of the driving process are uncorrelated.

Next, we compare the path diagram of the VAR model with the orthogonality graph of the MCAR model. Before doing so, we need to give some interpretations for the orthogonality graph.

**Remark 6.16.** For the purpose of interpretation of the directed and undirected edges in the orthogonality graph  $G_{OG}$ , recall from Lemma 6.8 the representation of the component  $Y_v$  of the MCAR process  $Y_V$  as

$$Y_v(t+h) = \sum_{j=1}^p e_v^\top \Theta_j^{(h)} D^{(j-1)} Y_V(t) + e_v^\top \varepsilon^{(h)}(t), \quad v \in V, \tag{6.10}$$

with

$$\Theta_j^{(h)} := \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j \in M_k(\mathbb{R}) \quad \text{and} \quad \varepsilon^{(h)}(t) := \int_t^{t+h} \mathbf{C} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \in \mathbb{R}^k.$$

(a) *Directed edges:* A direct application of Proposition 6.12 gives the condition for the directed edges in the orthogonality graph  $G_{OG}$  as

$$Y_a \not\rightarrow Y_b \mid Y_V \Leftrightarrow [\Theta_j^{(h)}]_{ba} = 0 \quad \forall h \in [0, 1], j = 1, \dots, p. \tag{6.11}$$

This means that the components  $Y_a(t)$ ,  $D^{(1)}Y_a(t), \dots, D^{(p-1)}Y_a(t)$  in the representation of the  $b$ th component  $Y_b(t+h)$  vanish due to the corresponding prefactors being zero.  $Y_a(t)$  and its derivatives do not matter to predict  $Y_b(t+h)$ .

(b) *Undirected edges:* A consequence of Proposition 6.13 is the condition for the undirected edges in the orthogonality graph  $G_{OG}$  as

$$Y_a \sim Y_b \mid Y_V \Leftrightarrow [\mathbb{E}[\varepsilon^{(h)}(t)\varepsilon^{(\tilde{h})}(t)^\top]]_{ab} = [\mathbb{E}[\varepsilon^{(h)}(0)\varepsilon^{(\tilde{h})}(0)^\top]]_{ab} = 0 \quad \forall h, \tilde{h} \in [0, 1], \tag{6.12}$$

i.e., the noise terms  $e_a^\top \varepsilon^{(h)}(t)$  and  $e_b^\top \varepsilon^{(\tilde{h})}(t)$  of  $Y_a(t+h)$  and  $Y_b(t+\tilde{h})$  are uncorrelated for any  $t \geq 0$ .

**Remark 6.17.** The characterisations of the directed and undirected edges of the orthogonality graph in Remark 6.16 are well suited for comparison with VAR( $p$ ) processes in Eichler [28]. The challenge here is that in representation (6.10) of  $Y_v(t+h)$  appear derivatives which have to be related to appropriate differences in the discrete-time process (6.9). Thus, our goal is to replace the backshifts  $Z_V(t+1-n)$ ,  $n = 1, \dots, p$ , by appropriate differences. To do this, we define a discrete-time difference operator iteratively by

$$D^{(1)}Z_V(t) = Z_V(t) - Z_V(t-1), \quad D^{(j)}Z_V(t) = D^{(j-1)}(Z_V(t) - Z_V(t-1)),$$

$j = 1, \dots, p-1$ , where we set  $D^{(0)}Z_V(t) = Z_V(t)$ . Furthermore, define

$$\Theta_j := \sum_{n=j}^p \binom{n-1}{j-1} (-1)^{j-1} \Phi_n, \quad j = 1, \dots, p.$$

Then some direct calculations show (see the Supplementary Material F) that

$$Z_b(t+1) = \sum_{j=1}^p e_b^\top \Theta_j D^{(j-1)}Z_V(t) + e_b^\top \varepsilon(t+1). \tag{6.13}$$

This representation is now in analogy to (6.10) for MCAR( $p$ ) processes.

(a) *Directed edges:* In the former Remark 6.15 we just saw that for the discrete-time VAR( $p$ ) process  $Z_V$  the directed edges in the path diagram  $G$  satisfy

$$Z_a \not\rightarrow Z_b \mid Z_V \Leftrightarrow a \longrightarrow b \notin E \Leftrightarrow [\Phi_j]_{ba} = 0, \quad j = 1, \dots, p.$$

But

$$[\Phi_j]_{ba} = 0, \quad j = 1, \dots, p \iff [\Theta_j]_{ba} = \sum_{n=j}^p \binom{n-1}{j-1} (-1)^{j-1} [\Phi_n]_{ba} = 0, \quad j = 1, \dots, p.$$

However, this is again analogous to the characterisation of directed edges in the orthogonality graph  $G_{OG}$  for the MCAR( $p$ ) process in (6.11) where

$$Y_a \not\rightarrow Y_b \mid Y_V \iff a \longrightarrow b \notin E \iff [\Theta_j^{(h)}]_{ba} = 0 \quad \forall h \in [0, 1], j = 1, \dots, p.$$

(b) *Undirected edges:* For the path diagram  $G$  for the VAR( $p$ ) process  $Z_V$  we have

$$Z_a \sim Z_b \mid Z_V \iff a \text{ --- } b \notin E \iff [\mathbb{E}[\varepsilon(0)\varepsilon(0)^\top]]_{ab} = 0.$$

Here we have the similarity to the condition (6.12) for the undirected edges of the MCAR( $p$ ) in the orthogonality graph  $G_{OG}$

$$Y_a \sim Y_b \mid Y_V \iff a \text{ --- } b \notin E_{OG} \iff [\mathbb{E}[\varepsilon^{(h)}(0)\varepsilon^{(\tilde{h})}(0)^\top]]_{ab} = 0 \quad \forall h, \tilde{h} \in [0, 1].$$

Since a continuous-time Ornstein–Uhlenbeck process sampled at discrete equidistant time points is a discrete-time VAR(1) process, we study the results for an Ornstein–Uhlenbeck process in more detail and, in particular, relate them to the results for VAR models in Eichler [28].

**Remark 6.18.** Let  $Y_V$  be a causal Ornstein–Uhlenbeck process as given in Example 3.15 with  $\Sigma_L > 0$ . Then the continuous-time process  $Y_V$  sampled at discrete-time points of distance  $h$  is a discrete-time VAR(1) process with representation

$$Y_V((k+1)h) = e^{A^h} Y_V(kh) + \int_{kh}^{(k+1)h} e^{A((k+1)h-u)} dL(u) = e^{A^h} Y_V(kh) + \varepsilon^{(h)}(kh), \quad k \in \mathbb{Z},$$

which we denote by  $Y_V^{(h)} = (Y_V((k+1)h))_{k \in \mathbb{Z}}$  and the corresponding discrete-time path diagram by  $G^{(h)} = (V, E^{(h)})$ . Then a direct conclusion of Remark 6.15 is that for  $a, b \in V$  and  $a \neq b$ :

$$\begin{aligned} \text{(a)} \quad Y_a \not\rightarrow Y_b \mid Y_V &\iff [e^{A^h}]_{ba} = 0 &\iff Y_a^{(h)} \not\rightarrow Y_b^{(h)} \mid Y_V^{(h)}. \\ \text{(b)} \quad Y_a \sim Y_b \mid Y_V &\iff [\mathbb{E}[\varepsilon^{(h)}(0)\varepsilon^{(h)}(0)^\top]]_{ab} = 0 &\iff Y_a^{(h)} \sim Y_b^{(h)} \mid Y_V^{(h)}. \end{aligned}$$

This means that a directed (undirected) edge  $a \longrightarrow b \in E^{(h)}$  ( $a \text{ --- } b \in E^{(h)}$ ) in the discrete-time model  $Y_V^{(h)}$  implies a (undirected) directed edge  $a \longrightarrow b \in E_{OG}$  ( $a \text{ --- } b \in E_{OG}$ ) in the continuous-time model  $Y_V$ . In summary,  $E^{(h)} \subseteq E_{OG}$  for every  $h \in [0, 1]$ . We believe that this result may hold for general MCAR( $p$ ) processes. This phenomenon is an advantage of the orthogonality graph over the local orthogonality graph, where there is generally no relationship between the edges  $E_{OG}^{(0)}$  and  $E^{(h)}$ .

The characterisation of the directed edges in Proposition 6.12 and the characterisation of the undirected edges in Proposition 6.13 are nice for interpretation, but depend on the lags  $h, \tilde{h}$ . We provide simpler necessary and sufficient criteria for the directed and undirected edges, respectively, where the lags  $h, \tilde{h}$  no longer play a role.

**Theorem 6.19.** Let  $Y_V$  be a causal MCAR( $p$ ) process with  $\Sigma_L > 0$ . Further, let  $a, b \in V$ ,  $a \neq b$ . Then the following holds.

$$\begin{aligned} \text{(a)} \quad Y_a \not\rightarrow Y_b \mid Y_V &\iff [\mathbf{CA}^\alpha \mathbf{E}_j]_{ba} = [\mathbf{A}^\alpha]_{b, k(j-1)+a} = 0, \quad \alpha = 1, \dots, kp-1, j = 1, \dots, p. \\ \text{(b)} \quad Y_a \sim Y_b \mid Y_V &\iff [\mathbf{CA}^\alpha \mathbf{B} \Sigma_L \mathbf{B}^\top (\mathbf{A}^\top)^\beta \mathbf{C}^\top]_{ab} = 0, \quad \alpha, \beta = 0, \dots, kp-1. \end{aligned}$$

**Remark 6.20.** The proof of Theorem 6.19 shows that in the definition of Granger causality and contemporaneous correlation the choice of the step size  $h$  as defined in Remark 3.2 (cf. (3.1)) and Remark 4.2 (cf. (4.1)), respectively, has no influence on the final characterisations of the edges in the MCAR model. For any choice  $h > 0$  we obtain the characterisations as in Theorem 6.19. In particular, it follows that Granger causality and global Granger causality as well as contemporaneous correlation and global contemporaneous correlation are equivalent for MCAR( $p$ ) processes, and hence the global orthogonality graph also satisfies the different Markov properties.

We obtain the following direct conclusion from Propositions 6.12, 6.13 and Theorem 6.19, setting  $\alpha = p$  in Theorem 6.19(a) and  $\alpha = \beta = p-1$  in Theorem 6.19(b).

**Corollary 6.21.** Let  $Y_V$  be a causal MCAR( $p$ ) process with  $\Sigma_L > 0$ , orthogonality graph  $G_{OG} = (V, E_{OG})$ , and local orthogonality graph  $G_{OG} = (V, E_{OG})$ . Then  $E_{OG}^{(0)} \subseteq E_{OG}$ , and in general the sets are not equal.

In particular, in the case of an Ornstein–Uhlenbeck process, the characterisation of the edges in an orthogonality graph can be reduced to the following.

**Corollary 6.22.** Let  $Y_V$  be a causal Ornstein–Uhlenbeck process with  $\Sigma_L > 0$ . Further, let  $a, b \in V$ ,  $a \neq b$ . Then the following holds.

$$\begin{aligned} \text{(a)} \quad Y_a \not\rightarrow Y_b \mid Y_V &\iff [\mathbf{A}^\alpha]_{ba} = 0, \quad \alpha = 1, \dots, k-1. \\ \text{(b)} \quad Y_a \sim Y_b \mid Y_V &\iff [\mathbf{A}^\alpha \Sigma_L (\mathbf{A}^\top)^\beta]_{ab} = 0, \quad \alpha, \beta = 0, \dots, k-1. \end{aligned}$$

**Remark 6.23.** Suppose  $\Sigma_L$  is a diagonal matrix and  $Y_V$  is a causal Ornstein–Uhlenbeck process. Then [Corollary 6.22](#) implies that from  $Y_a \sim Y_b \mid Y_V$  directly follows  $Y_a \not\rightarrow Y_b \mid Y_V$ . Thus, a directed edge in such an orthogonality graph of an Ornstein–Uhlenbeck process induces an undirected edge.

### 7. Conclusion

In this paper, we have introduced concepts of directed and undirected influences for stochastic processes in continuous time, defined (local) orthogonality graphs, discussed their properties, and applied them to MCAR processes. The main results are as follows:

- (a) (Local) orthogonality graphs provide a simple visualisation and a concise way to communicate directed and undirected (local) conditional orthogonality structures of the process.
- (b) (Local) orthogonality graphs are defined using the pairwise Markov property to represent the pairwise relationships between variables. However, the associated orthogonality graph can be interpreted using the global AMP Markov and the global Markov property. In this way, new Granger non-causality relations and contemporaneous uncorrelations between subprocesses can be obtained.
- (c) For MCAR models the (local) orthogonality graphs are closely related to the moving average parameters and the covariance matrix of the driving Lévy process. Any local orthogonality graph can be constructed by an MCAR model, but this is generally not true for an orthogonality graph. However, if there is no edge in the orthogonality graph, then there is no edge in the discrete-time sampled model.

### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lea Schenk reports financial support was provided by Federal Ministry of Education and Research. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Appendix A. Proofs of Section 3

**Proof of Theorem 3.5.** Due to Lindquist and Picci [53], Proposition 2.4.2,  $\mathcal{L}_{Y_B}(t, t + 1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t)$  is equivalent to  $P_{\mathcal{L}_{Y_S}(t)} Y^B = P_{\mathcal{L}_{Y_{S \setminus A}}(t)} Y^B$   $\mathbb{P}$ -a.s. for all  $Y^B \in \mathcal{L}_{Y_B}(t, t + 1)$ . Due to the linearity and continuity of orthogonal projections, this is in turn equivalent to  $P_{\mathcal{L}_{Y_S}(t)} Y_b(t + h) = P_{\mathcal{L}_{Y_{S \setminus A}}(t)} Y_b(t + h)$   $\mathbb{P}$ -a.s. for all  $h \in [0, 1]$ ,  $t \in \mathbb{R}$  and  $b \in B$ .  $\square$

**Proof of Theorem 3.10.** First assume that  $Y_A \not\rightarrow_0 Y_B \mid Y_S$ , i.e.,  $\mathbb{P}$ -a.s.

$$\liminf_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t + h) - D^{(j_b)} Y_b(t)}{h} \right) = \liminf_{h \rightarrow 0} P_{\mathcal{L}_{Y_{S \setminus A}}(t)} \left( \frac{D^{(j_b)} Y_b(t + h) - D^{(j_b)} Y_b(t)}{h} \right), \tag{A.1}$$

for all  $t \in \mathbb{R}$  and  $b \in B$ . Now let  $Y^A \in \mathcal{L}_{Y_A}(t)$ ,  $b \in B$ , and  $t \in \mathbb{R}$ . Then as well  $Y^A \in \mathcal{L}_{Y_S}(t)$  and  $D^{(j_b)} Y_b(t + h) - P_{\mathcal{L}_{Y_S}(t)} D^{(j_b)} Y_b(t + h) \in \mathcal{L}_{Y_S}(t)^\perp$ , so

$$\frac{1}{h} \mathbb{E} \left[ \left( D^{(j_b)} Y_b(t + h) - P_{\mathcal{L}_{Y_S}(t)} D^{(j_b)} Y_b(t + h) \right) \overline{Y^A} \right] = 0.$$

Adding and subtracting  $P_{\mathcal{L}_{Y_{S \setminus A}}(t)} D^{(j_b)} Y_b(t + h)$  in the first factor and then forming the limit gives

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( D^{(j_b)} Y_b(t + h) - P_{\mathcal{L}_{Y_{S \setminus A}}(t)} D^{(j_b)} Y_b(t + h) \right) \overline{Y^A} \right] + \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( P_{\mathcal{L}_{Y_{S \setminus A}}(t)} D^{(j_b)} Y_b(t + h) - P_{\mathcal{L}_{Y_S}(t)} D^{(j_b)} Y_b(t + h) \right) \overline{Y^A} \right] = 0. \tag{A.2}$$

Due to [Remark 2.6](#) and  $A \cap B = \emptyset$ , we already know that  $D^{(j_b)} Y_b(t) \in \mathcal{L}_{Y_{S \setminus A}}(t) \subseteq \mathcal{L}_{Y_S}(t)$ . Then it follows together with (A.1) and (2.1) that the second summand in (A.2) is zero and thus, the first summand is zero as well, i.e.,

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( D^{(j_b)} Y_b(t + h) - P_{\mathcal{L}_{Y_{S \setminus A}}(t)} D^{(j_b)} Y_b(t + h) \right) \overline{Y^A} \right] = 0.$$

Further,  $D^{(j_b)} Y_b(t + h) - P_{\mathcal{L}_{Y_{S \setminus A}}(t)} D^{(j_b)} Y_b(t + h) \in \mathcal{L}_{Y_{S \setminus A}}(t)^\perp$  and  $P_{\mathcal{L}_{Y_{S \setminus A}}(t)} Y^A \in \mathcal{L}_{Y_{S \setminus A}}(t)$  give

$$\frac{1}{h} \mathbb{E} \left[ \left( D^{(j_b)} Y_b(t + h) - P_{\mathcal{L}_{Y_{S \setminus A}}(t)} D^{(j_b)} Y_b(t + h) \right) \overline{P_{\mathcal{L}_{Y_{S \setminus A}}(t)} Y^A} \right] = 0.$$

Adding the limit, the last two equations yield as claimed

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( D^{(j_b)} Y_b(t + h) - P_{\mathcal{L}_{Y_{S \setminus A}}(t)} D^{(j_b)} Y_b(t + h) \right) \overline{\left( Y^A - P_{\mathcal{L}_{Y_{S \setminus A}}(t)} Y^A \right)} \right] = 0. \quad \square$$

**Proof of Lemma 3.13.** (a) This is obvious by definitions.

(b) The implication  $\Rightarrow$  follows instantly. For the proof of  $\Leftarrow$  we use mathematical induction and show that

$$\mathcal{L}_{Y_{S \setminus A}}(t+k) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall t \in \mathbb{R}, k \in \mathbb{N}. \tag{A.3}$$

First, we note that  $Y_A \dashrightarrow Y_{S \setminus A} \mid Y_S$  and Lemma 3.3(b) yield the initial case

$$\mathcal{L}_{Y_{S \setminus A}}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall t \in \mathbb{R}. \tag{A.4}$$

Now, replacing  $t$  by  $t+1$  in the induction hypothesis gives

$$\mathcal{L}_{Y_{S \setminus A}}(t+k+1) \perp \mathcal{L}_{Y_A}(t+1) \mid \mathcal{L}_{Y_{S \setminus A}}(t+1) \quad \forall t \in \mathbb{R}.$$

Since by Lemma 2.4 we have  $\mathcal{L}_{Y_A}(t+1) = \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_A}(t, t+1)$ , the property of decomposition (C2) from Lemma 2.2 implies

$$\mathcal{L}_{Y_{S \setminus A}}(t+k+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t+1) \quad \forall t \in \mathbb{R},$$

which is by Lemma 2.4 again

$$\mathcal{L}_{Y_{S \setminus A}}(t+k+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \vee \mathcal{L}_{Y_{S \setminus A}}(t, t+1) \quad \forall t \in \mathbb{R}.$$

This result together with the initial case (A.4) and the properties of decomposition (C2) and contraction (C4) from Lemma 2.2 yield

$$\mathcal{L}_{Y_{S \setminus A}}(t+k+1) \vee \mathcal{L}_{Y_{S \setminus A}}(t, t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall t \in \mathbb{R}.$$

Finally, the property of decomposition (C2) gives the induction step

$$\mathcal{L}_{Y_{S \setminus A}}(t+k+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall t \in \mathbb{R}.$$

To bring the proof to an end, let  $\lceil \cdot \rceil$  be the ceiling function. Then  $\mathcal{L}_{Y_{S \setminus A}}(t+h) \subseteq \mathcal{L}_{Y_{S \setminus A}}(t + \lceil h \rceil)$ . Now it follows from (A.3) and the decomposition property (C2) that

$$\mathcal{L}_{Y_{S \setminus A}}(t+h) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t) \quad \forall t \in \mathbb{R}, h \geq 0.$$

(c) This follows directly due to (b), the decomposition property (C2), and  $B \subseteq S \setminus A$ .

(d) Let  $Y_A \dashrightarrow Y_B \mid Y_S$ , i.e.,  $\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S \setminus A}}(t)$  for all  $t \in \mathbb{R}$  due to Lemma 3.3(b). Then, as in the proof of Theorem 3.5 (cf. Proposition 2.4.2 in Lindquist and Picci [53]), we have

$$P_{\mathcal{L}_{Y_S}(t)} Y^B = P_{\mathcal{L}_{Y_{S \setminus A}}(t)} Y^B \quad \mathbb{P}\text{-a.s.}$$

for all  $Y^B \in \mathcal{L}_{Y_B}(t+1)$  and  $t \in \mathbb{R}$ . Furthermore, Remark 2.6 provides that, for  $b \in B$  and  $h \in [0, 1]$ , we have  $D^{(j_b)} Y_b(t+h) \in \mathcal{L}_{Y_B}(t+h) \subseteq \mathcal{L}_{Y_B}(t+1)$ . All together result in

$$P_{\mathcal{L}_{Y_S}(t)} D^{(j_b)} Y_b(t+h) = P_{\mathcal{L}_{Y_{S \setminus A}}(t)} D^{(j_b)} Y_b(t+h) \quad \mathbb{P}\text{-a.s.}$$

Since, in addition,  $D^{(j_b)} Y_b(t) \in \mathcal{L}_{Y_{S \setminus A}}(t) \subseteq \mathcal{L}_{Y_S}(t)$  by Remark 2.6 again, we have

$$P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) = P_{\mathcal{L}_{Y_{S \setminus A}}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right).$$

Letting  $h \rightarrow 0$ , we receive the statement.  $\square$

## Appendix B. Proofs of Section 5

### B.1. Proofs of Section 5.1

**Proof of Proposition 5.4.** Let  $A, B \subseteq V$  be disjoint with  $\#A = \alpha$ ,  $\#B = \beta$ . First, according to Assumption 1, there exists an  $0 < \varepsilon < 1$  such that

$$f_{Y_A Y_A}(\lambda)^{-1/2} f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda) f_{Y_A Y_A}(\lambda)^{-1/2} \leq_L (1 - \varepsilon) I_\alpha,$$

for almost all  $\lambda \in \mathbb{R}$  and hence,

$$(1 - \varepsilon) f_{Y_A Y_A}(\lambda) - f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda) \geq 0,$$

for almost all  $\lambda \in \mathbb{R}$ . If we choose  $0 < \tilde{\varepsilon} < 1$ , such that  $(1 - \tilde{\varepsilon})^2 = (1 - \varepsilon)$ , we obtain

$$(1 - \tilde{\varepsilon}) f_{Y_A Y_A}(\lambda) - f_{Y_A Y_B}(\lambda) \left( (1 - \tilde{\varepsilon}) f_{Y_B Y_B}(\lambda) \right)^{-1} f_{Y_B Y_A}(\lambda) \geq 0,$$

for almost all  $\lambda \in \mathbb{R}$ . Since  $(1 - \tilde{\varepsilon}) f_{Y_B Y_B}(\lambda) \geq 0$ , Bernstein [5], Proposition 8.2.4., provides

$$\begin{pmatrix} (1 - \tilde{\varepsilon}) f_{Y_A Y_A}(\lambda) & f_{Y_A Y_B}(\lambda) \\ f_{Y_B Y_A}(\lambda) & (1 - \tilde{\varepsilon}) f_{Y_B Y_B}(\lambda) \end{pmatrix} \geq 0,$$



respectively

$$\begin{pmatrix} f_{Y_A Y_A}(\lambda) & f_{Y_A Y_B}(\lambda) \\ f_{Y_B Y_A}(\lambda) & f_{Y_B Y_B}(\lambda) \end{pmatrix} \geq_L \tilde{\varepsilon} \begin{pmatrix} f_{Y_A Y_A}(\lambda) & 0_{\alpha \times \beta} \\ 0_{\beta \times \alpha} & f_{Y_B Y_B}(\lambda) \end{pmatrix}, \tag{B.1}$$

for almost all  $\lambda \in \mathbb{R}$ . With this preliminary work in mind, we can now provide the actual proof of the assertion. Let  $Y^A \in \mathcal{L}_{Y_A}(t)$  and  $Y^B \in \mathcal{L}_{Y_B}(t)$ ,  $t \in \mathbb{R}$ . Then  $Y^A \in \mathcal{L}_{Y_A}$  and  $Y^B \in \mathcal{L}_{Y_B}$ . Due to Rozanov [67], I, (7.2), the spectral representation

$$Y^A = \int_{-\infty}^{\infty} \varphi(\lambda) \Phi_A(d\lambda) \quad \text{and} \quad Y^B = \int_{-\infty}^{\infty} \psi(\lambda) \Phi_B(d\lambda) \quad \mathbb{P}\text{-a.s.}$$

holds, where  $\Phi_A(\cdot)$  and  $\Phi_B(\cdot)$  are the random spectral measures from the subprocesses  $Y_A$  and  $Y_B$  from (2.2). Furthermore,  $\varphi(\cdot) \in C^{1 \times \alpha}$  and  $\psi(\cdot) \in C^{1 \times \beta}$  are measurable vector functions that satisfy

$$\int_{-\infty}^{\infty} \varphi(\lambda) f_{Y_A Y_A}(\lambda) \overline{\varphi(\lambda)}^\top d\lambda < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} \psi(\lambda) f_{Y_B Y_B}(\lambda) \overline{\psi(\lambda)}^\top d\lambda < \infty.$$

Using (B.1) and the monotonicity of the integral in the inequality, we obtain

$$\begin{aligned} \|Y^A + Y^B\|_{L^2}^2 &= \int_{-\infty}^{\infty} (\varphi(\lambda) \ \psi(\lambda)) \begin{pmatrix} f_{Y_A Y_A}(\lambda) & f_{Y_A Y_B}(\lambda) \\ f_{Y_B Y_A}(\lambda) & f_{Y_B Y_B}(\lambda) \end{pmatrix} \overline{(\varphi(\lambda) \ \psi(\lambda))}^\top d\lambda \\ &\geq \tilde{\varepsilon} \int_{-\infty}^{\infty} (\varphi(\lambda) \ \psi(\lambda)) \begin{pmatrix} f_{Y_A Y_A}(\lambda) & 0_{\alpha \times \beta} \\ 0_{\beta \times \alpha} & f_{Y_B Y_B}(\lambda) \end{pmatrix} \overline{(\varphi(\lambda) \ \psi(\lambda))}^\top d\lambda = \tilde{\varepsilon} (\|Y^A\|_{L^2}^2 + \|Y^B\|_{L^2}^2). \end{aligned}$$

Then Feshchenko [34], Proposition 2.3, provides that for  $t \in \mathbb{R}$ ,

$$\mathcal{L}_{Y_A}(t) \cap \mathcal{L}_{Y_B}(t) = \{0\} \quad \text{and} \quad \mathcal{L}_{Y_A}(t) + \mathcal{L}_{Y_B}(t) = \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_B}(t) \quad \mathbb{P}\text{-a.s.}$$

Thus, Lemma 5.3 yields the final statement  $\mathcal{L}_{Y_{A \cup C}}(t) \cap \mathcal{L}_{Y_{B \cup C}}(t) = \mathcal{L}_{Y_C}(t)$   $\mathbb{P}$ -a.s.  $\square$

**Proof of Proposition 5.7.** (a) The direction  $\Rightarrow$  is already given in (3.3). Thus, let us prove  $\Leftarrow$  and assume that  $Y_a \dashrightarrow Y_b \mid Y_S$  for all  $a \in A$ ,  $b \in B$ . Then we receive due to Theorem 3.5 that

$$P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h) = P_{\mathcal{L}_{Y_S \setminus \{a\}}(t)} Y_b(t+h) \quad \mathbb{P}\text{-a.s.}$$

for all  $h \in [0, 1]$ ,  $t \in \mathbb{R}$ ,  $a \in A$ ,  $b \in B$ . This implies that

$$P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h) \in \mathcal{L}_{Y_S \setminus \{a\}}(t) \quad \forall a \in A.$$

Now, from Proposition 5.4, which requires Assumption 1, we conclude that

$$P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h) \in \bigcap_{a \in A} \mathcal{L}_{Y_S \setminus \{a\}}(t) = \mathcal{L}_{Y_S \setminus A}(t),$$

implying due to Brockwell and Davis [9], Proposition 2.3.2. (vii) that

$$P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h) = P_{\mathcal{L}_{Y_S \setminus A}(t)} P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h) = P_{\mathcal{L}_{Y_S \setminus A}(t)} Y_b(t+h) \quad \mathbb{P}\text{-a.s.}$$

for all  $b \in B$ ,  $t \in \mathbb{R}$ , and  $h \in [0, 1]$ . We apply Theorem 3.5 again and obtain  $Y_A \dashrightarrow Y_B \mid Y_S$ .

(b) The direction  $\Rightarrow$  is already given in (3.6) and we just prove  $\Leftarrow$ . Thus assume that  $Y_a \dashrightarrow_0 Y_b \mid Y_S$  for all  $a \in A$ ,  $b \in B$ . By Definition 3.7 that is

$$\text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) = \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S \setminus \{a\}}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) \quad \mathbb{P}\text{-a.s.}$$

for all  $t \in \mathbb{R}$ ,  $a \in A$ ,  $b \in B$ . Since  $\mathcal{L}_{Y_S \setminus \{a\}}(t)$  is closed in the mean-square sense, we obtain

$$\text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) \in \mathcal{L}_{Y_S \setminus \{a\}}(t) \quad \forall a \in A.$$

As in (a), Proposition 5.4, which requires Assumption 1, yields

$$\text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) \in \mathcal{L}_{Y_S \setminus A}(t).$$

Due to Brockwell and Davis [9], Proposition 2.3.2. (iv) and (vii), it follows

$$\begin{aligned} &\text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) \\ &= P_{\mathcal{L}_{Y_S \setminus A}(t)} \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) \\ &= \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S \setminus A}(t)} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) \end{aligned}$$

$$= \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S \setminus A}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) \quad \mathbb{P}\text{-a.s.}$$

for all  $b \in B$ ,  $t \in \mathbb{R}$ . By Definition 3.7 that is  $Y_A \dashrightarrow_0 Y_B \mid Y_S$ .

(c) The proof is the same as in (a).  $\square$

**B.2. Proof of Theorem 5.15**

The proof of the global AMP Markov property is structured in three auxiliary lemmata and is based on the ideas of Eichler [28,29]. At the end, we present the proof of Theorem 5.15.

**Lemma B.1.** Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$ . Suppose  $A, B \subseteq V$  are disjoint subsets,  $t \in \mathbb{R}$ , and  $k \in \mathbb{N}$ . Then

$$A \bowtie_m B \mid V \setminus (A \cup B) [G_{OG}] \Rightarrow \mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t) \vee \mathcal{L}_{Y_{A \cup B}}(t-k).$$

**Proof.** The proof can be done step by step as in Eichler [29], proof of Lemma 4.1, by induction over  $k$ , using the properties of a semi-graphoid given in our Lemma 2.2.  $\square$

**Lemma B.2.** Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$ . Suppose  $A, B \subseteq V$  are disjoint subsets and  $t \in \mathbb{R}$ . Then

$$A \bowtie_m B \mid V \setminus (A \cup B) [G_{OG}] \Rightarrow \mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t).$$

**Proof.** First,  $\mathcal{L}_{Y_{A \cup B}}(t-k) \vee \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t) \supseteq \mathcal{L}_{Y_{A \cup B}}(t-k-1) \vee \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)$  for  $k \in \mathbb{N}$  and

$$\bigcap_{k \in \mathbb{N}} \left( \mathcal{L}_{Y_{A \cup B}}(t-k) \vee \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t) \right) = \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t),$$

due to Lemma 5.8. Theorems 4.31(b) and 4.32 in Weidmann [73] provide

$$\text{l.i.m.}_{k \rightarrow \infty} P_{\mathcal{L}_{Y_{A \cup B}}(t-k) \vee \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)} Y = P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)} Y, \quad Y \in L^2.$$

Let  $Y^A \in \mathcal{L}_{Y_A}(t)$  and  $Y^B \in \mathcal{L}_{Y_B}(t)$ . Then, using (2.1),

$$\begin{aligned} & \mathbb{E} \left[ \left( Y^A - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)} Y^A \right) \overline{\left( Y^B - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)} Y^B \right)} \right] \\ &= \lim_{k \rightarrow \infty} \mathbb{E} \left[ \left( Y^A - P_{\mathcal{L}_{Y_{A \cup B}}(t-k) \vee \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)} Y^A \right) \overline{\left( Y^B - P_{\mathcal{L}_{Y_{A \cup B}}(t-k) \vee \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)} Y^B \right)} \right]. \end{aligned}$$

The expression on the right-hand side is zero since, due to Lemma B.1,  $\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t) \vee \mathcal{L}_{Y_{A \cup B}}(t-k)$  for  $t \in \mathbb{R}$ ,  $k \in \mathbb{N}$ . Thus, the expression on the left-hand side is also zero and  $\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)$ .  $\square$

**Lemma B.3.** Let  $G_{OG} = (V, E_{OG})$  be the orthogonality graph for  $Y_V$  and suppose  $A, B \subseteq V$  are disjoint subsets. Then

$$A \bowtie_m B \mid V \setminus (A \cup B) [G_{OG}] \Rightarrow \mathcal{L}_{Y_A} \perp \mathcal{L}_{Y_B} \mid \mathcal{L}_{Y_{V \setminus (A \cup B)}}.$$

**Proof.** First, note from Lemma 2.4 that  $\overline{\bigcup_{n \in \mathbb{N}} \mathcal{L}_{Y_S}(n)} = \mathcal{L}_{Y_S}$   $\mathbb{P}$ -a.s. for any  $S \subseteq V$ . Let  $Y^A \in \mathcal{L}_{Y_A}$  and  $Y^B \in \mathcal{L}_{Y_B}$ . Then analogue arguments as in the proof of Lemma B.2 give

$$\begin{aligned} Y^A - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}} Y^A &= \text{l.i.m.}_{n \rightarrow \infty} P_{\mathcal{L}_{Y_A}(n)} Y^A - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}(n)} P_{\mathcal{L}_{Y_A}(n)} Y^A, \\ Y^B - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}} Y^B &= \text{l.i.m.}_{n \rightarrow \infty} P_{\mathcal{L}_{Y_B}(n)} Y^B - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}(n)} P_{\mathcal{L}_{Y_B}(n)} Y^B. \end{aligned}$$

Further, (2.1) yields

$$\begin{aligned} & \mathbb{E} \left[ \left( Y^A - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}} Y^A \right) \overline{\left( Y^B - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}} Y^B \right)} \right] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left[ \left( P_{\mathcal{L}_{Y_A}(n)} Y^A - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}(n)} P_{\mathcal{L}_{Y_A}(n)} Y^A \right) \overline{\left( P_{\mathcal{L}_{Y_B}(n)} Y^B - P_{\mathcal{L}_{Y_{V \setminus (A \cup B)}}(n)} P_{\mathcal{L}_{Y_B}(n)} Y^B \right)} \right]. \end{aligned}$$

The expression on the right-hand side is zero, since  $\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_{V \setminus (A \cup B)}}(t)$  for  $t \in \mathbb{R}$  due to Lemma B.2. Thus, the left-hand side is zero and  $\mathcal{L}_{Y_A} \perp \mathcal{L}_{Y_B} \mid \mathcal{L}_{Y_{V \setminus (A \cup B)}}$ .  $\square$

**Proof of Theorem 5.15.** For the proof of Theorem 5.15, we refer to the proof of Theorem 3.1 in Eichler [28], since it is based only on Lemma B.3, properties of mixed graphs, and Lemma 2.2.  $\square$

B.3. Proofs of Section 5.3.3

**Proof of Proposition 5.20.** For a graph  $G = (V, E)$  let

$$\text{ch}(a) = \{v \in V \mid a \longrightarrow v \in E\} \quad \text{and} \quad \text{dis}(a) = \{v \in V \mid v \dashrightarrow \dots \dashrightarrow a \text{ or } v = a\},$$

denote the set of children and the district of  $a \in V$ , respectively. For  $A \subseteq V$  let  $\text{ch}(A) = \bigcup_{a \in A} \text{ch}(a)$  and  $\text{dis}(A) = \bigcup_{a \in A} \text{dis}(a)$ . Due to Eichler [28], Lemma B.1,  $A \bowtie_m B \mid V \setminus (A \cup B) [G_{OG}^0]$  yields  $\text{dis}(A \cup \text{ch}(A)) \cap \text{dis}(B \cup \text{ch}(B)) = \emptyset$ . In particular,  $\text{ch}(A) \cap B = \emptyset$ ,  $A \cap \text{ch}(B) = \emptyset$ , and  $\text{ne}(A) \cap B = \emptyset$ . Thus, as claimed,  $Y_A \dashrightarrow_0 Y_B \mid Y_V$ ,  $Y_B \dashrightarrow_0 Y_A \mid Y_V$ , and  $Y_A \approx_0 Y_B \mid Y_V$ .  $\square$

**Proof of Lemma 5.23.** The assumption  $Y_{A \cup B} \dashrightarrow_0 Y_C \mid Y_{A \cup B \cup C \cup D}$  states that for all  $t \in \mathbb{R}$  and  $c \in C$ ,

$$\text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_{A \cup B \cup C \cup D}}(t)} \left( \frac{D^{(j_c)} Y_c(t+h) - D^{(j_c)} Y_c(t)}{h} \right) = \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_{C \cup D}}(t)} \left( \frac{D^{(j_c)} Y_c(t+h) - D^{(j_c)} Y_c(t)}{h} \right) \quad \mathbb{P}\text{-a.s.}$$

An application of  $P_{\mathcal{L}_{Y_{A \cup C \cup D}}(t)}$  on the left and the right hand side, Brockwell and Davis [9], Proposition 2.3.2.(iv,vii), and

$$P_{\mathcal{L}_{Y_{A \cup C \cup D}}(t)} P_{\mathcal{L}_{Y_{A \cup B \cup C \cup D}}(t)} = P_{\mathcal{L}_{Y_{A \cup C \cup D}}(t)} \quad \text{and} \quad P_{\mathcal{L}_{Y_{A \cup C \cup D}}(t)} P_{\mathcal{L}_{Y_{C \cup D}}(t)} = P_{\mathcal{L}_{Y_{C \cup D}}(t)},$$

respectively, give for  $t \in \mathbb{R}$  and  $c \in C$ ,

$$\text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_{A \cup C \cup D}}(t)} \left( \frac{D^{(j_c)} Y_c(t+h) - D^{(j_c)} Y_c(t)}{h} \right) = \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_{C \cup D}}(t)} \left( \frac{D^{(j_c)} Y_c(t+h) - D^{(j_c)} Y_c(t)}{h} \right) \quad \mathbb{P}\text{-a.s.}$$

By definition that is  $Y_A \dashrightarrow_0 Y_C \mid Y_{A \cup C \cup D}$ .  $\square$

**Proof of Proposition 5.21.** The block-recursive Markov property Proposition 5.13 says that  $Y_{V \setminus (B \cup \text{pa}(B))} \dashrightarrow_0 Y_B \mid Y_V$ . By assumption,  $B \cup \text{pa}(B) \subseteq A \cup B \cup C$ . However,  $A \cap \text{pa}(B) = \emptyset$ . Otherwise, there are vertices  $a \in A$  and  $b \in B$  such that  $a \longrightarrow b \in E_{OG}^0$  is a  $m$ -connecting path between  $A$  and  $B$  given  $C$  which is a contradiction to  $A \bowtie_m B \mid C [G_{OG}^0]$ . Thus,  $B \cup \text{pa}(B) \subseteq B \cup C$  and Proposition 5.7 yields  $Y_{V \setminus (B \cup C)} \dashrightarrow_0 Y_B \mid Y_V$ . The property of left decomposition (Lemma 5.23) gives  $Y_A \dashrightarrow_0 Y_B \mid Y_{A \cup B \cup C}$ . By symmetry of  $m$ -separation  $Y_B \dashrightarrow_0 Y_A \mid Y_{A \cup B \cup C}$  follows.

It remains to show that  $Y_A \approx_0 Y_B \mid Y_{A \cup B \cup C}$ . Proposition 5.13 provides  $Y_{V \setminus (B \cup \text{ne}(B))} \approx_0 Y_B \mid Y_V$ . Here,  $A \cap \text{ne}(B) = \emptyset$ . Else there are vertices  $a \in A$  and  $b \in B$  such that  $a \dashrightarrow \dots \dashrightarrow b \in E_{OG}^0$  is a  $m$ -connecting path between  $A$  and  $B$  given  $C$  which is again a contradiction to  $A \bowtie_m B \mid C [G_{OG}^0]$ . So Remark 4.7 yields  $Y_A \approx_0 Y_B \mid Y_V$ . By definition and  $D^{(j_a)} Y_a(t), D^{(j_b)} Y_b(t) \in \mathcal{L}_{Y_{A \cup B \cup C}}(t) \subseteq \mathcal{L}_{Y_V}(t)$  we get

$$\begin{aligned} 0 &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( D^{(j_a)} Y_a(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(j_a)} Y_a(t+h) \right) \overline{\left( D^{(j_b)} Y_b(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(j_b)} Y_b(t+h) \right)} \right] \\ &= \lim_{h \rightarrow 0} h \mathbb{E} \left[ \left( \frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} - P_{\mathcal{L}_{Y_V}(t)} \frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} \right) \right. \\ &\quad \left. \times \overline{\left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} - P_{\mathcal{L}_{Y_V}(t)} \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right)} \right], \end{aligned} \tag{B.2}$$

for  $t \in \mathbb{R}$ ,  $a \in A$ ,  $b \in B$ . Due to Proposition 5.13 and  $\text{pa}(A) \cup \text{pa}(B) \subseteq A \cup B \cup C$  we receive, as in the first part of this proof,

$$Y_{V \setminus (A \cup B \cup C)} \dashrightarrow_0 Y_B \mid Y_V \quad \text{and} \quad Y_{V \setminus (A \cup B \cup C)} \dashrightarrow_0 Y_A \mid Y_V,$$

which means that  $\mathbb{P}$ -a.s.

$$\begin{aligned} \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_V}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) &= \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_{A \cup B \cup C}}(t)} \left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right) \quad \text{and} \\ \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_V}(t)} \left( \frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} \right) &= \text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_{A \cup B \cup C}}(t)} \left( \frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} \right), \end{aligned}$$

for  $t \in \mathbb{R}$ ,  $a \in A$ ,  $b \in B$ . Similar arguments as in the proof of Theorem 3.10 and (B.2) yield

$$\begin{aligned} 0 &= \lim_{h \rightarrow 0} h \mathbb{E} \left[ \left( \frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} - P_{\mathcal{L}_{Y_{A \cup B \cup C}}(t)} \frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} \right) \right. \\ &\quad \left. \times \overline{\left( \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} - P_{\mathcal{L}_{Y_{A \cup B \cup C}}(t)} \frac{D^{(j_b)} Y_b(t+h) - D^{(j_b)} Y_b(t)}{h} \right)} \right] \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( D^{(j_a)} Y_a(t+h) - P_{\mathcal{L}_{Y_{A \cup B \cup C}}(t)} D^{(j_a)} Y_a(t+h) \right) \overline{\left( D^{(j_b)} Y_b(t+h) - P_{\mathcal{L}_{Y_{A \cup B \cup C}}(t)} D^{(j_b)} Y_b(t+h) \right)} \right] \end{aligned}$$

for  $t \in \mathbb{R}$ ,  $a \in A$ ,  $b \in B$ , which says that  $Y_A \approx_0 Y_B \mid Y_{A \cup B \cup C}$ .  $\square$

**Appendix C. Proofs of Section 6**

**C.1. Proofs of Section 6.3**

**Proof of Lemma 6.8.** Let  $t \in \mathbb{R}$ ,  $h \geq 0$ , and  $v \in V$ . First of all, due to Lemma 6.2,

$$Y_v(t+h) = e_v^\top \mathbf{C} X(t+h) = e_v^\top \mathbf{C} \left( e^{\mathbf{A}h} X(t) + \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \right).$$

With the definition of the  $j$ -th  $k$ -block  $X^{(j)}$  of  $X$  as in (6.4) and with (6.8) it follows

$$Y_v(t+h) = e_v^\top \mathbf{C} e^{\mathbf{A}h} \sum_{j=1}^p \mathbf{E}_j X^{(j)}(t) + e_v^\top \mathbf{C} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) = e_v^\top \mathbf{C} e^{\mathbf{A}h} \sum_{j=1}^p \mathbf{E}_j D^{(j-1)} Y_V(t) + e_v^\top \mathbf{C} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u). \quad \square$$

**Proof of Lemma 6.11.** For the proof of the first equation note that the MCAR( $p$ ) process  $Y_V$  is  $(p-1)$ -times differentiable with  $D^{(p-1)} Y_V(t) = X^{(p)}(t) = \mathbf{E}_p^\top X(t)$ , see Remark 6.4. Then, as in the proof of Lemma 6.8,

$$\begin{aligned} & D^{(p-1)} Y_v(t+h) - D^{(p-1)} Y_v(t) \\ &= e_v^\top \mathbf{E}_p^\top \left( (e^{\mathbf{A}h} - I_{kp}) X(t) + \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \right) \\ &= e_v^\top \mathbf{E}_p^\top (e^{\mathbf{A}h} - I_{kp}) \sum_{j=1}^p \mathbf{E}_j D^{(j-1)} Y_V(t) + e_v^\top \mathbf{E}_p^\top \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u). \end{aligned}$$

Remark 2.6 states that  $Y_s(t)$  and its derivatives are already in  $\mathcal{L}_{Y_S}(t)$  and are therefore projected onto themselves. Additionally,  $\sigma(Y_S(t'), t' \leq t)$  and  $\sigma(L(t+h) - L(t'), t \leq t' \leq t+h)$  are independent and thus,  $e_v^\top \mathbf{E}_p^\top \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u)$  is projected on zero. It follows

$$\begin{aligned} & P_{\mathcal{L}_{Y_S}(t)} (D^{(p-1)} Y_v(t+h) - D^{(p-1)} Y_v(t)) \\ &= e_v^\top \mathbf{E}_p^\top (e^{\mathbf{A}h} - I_{kp}) \sum_{s \in S} \sum_{j=1}^p \mathbf{E}_j e_s D^{(j-1)} Y_s(t) + e_v^\top \mathbf{E}_p^\top (e^{\mathbf{A}h} - I_{kp}) \sum_{s \in V \setminus S} \sum_{j=1}^p \mathbf{E}_j e_s P_{\mathcal{L}_{Y_S}(t)} (D^{(j-1)} Y_s(t)) \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

For the proof of the second equation, we apply the same arguments to receive

$$\begin{aligned} & D^{(p-1)} Y_v(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(p-1)} Y_v(t+h) \\ &= e_v^\top \mathbf{E}_p^\top \left( e^{\mathbf{A}h} X(t) + \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \right) - P_{\mathcal{L}_{Y_V}(t)} \left( e_v^\top \mathbf{E}_p^\top \left( e^{\mathbf{A}h} X(t) + \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \right) \right) \\ &= e_v^\top \mathbf{E}_p^\top \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \quad \mathbb{P}\text{-a.s.} \quad \square \end{aligned}$$

**C.2. Proofs of Section 6.4**

**Proof of Proposition 6.12.** (a) Recall that, due to Theorem 3.5,  $Y_a \not\rightarrow Y_b | Y_V$  if and only if,

$$P_{\mathcal{L}_{Y_V}(t)} Y_b(t+h) = P_{\mathcal{L}_{Y_V \setminus \{a\}}(t)} Y_b(t+h) \quad \mathbb{P}\text{-a.s.} \quad \forall h \in [0, 1], t \in \mathbb{R}.$$

From Proposition 6.9 we know that

$$\begin{aligned} & P_{\mathcal{L}_{Y_V}(t)} Y_b(t+h) = \sum_{j=1}^p \sum_{s \in V} e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_s D^{(j-1)} Y_s(t), \\ & P_{\mathcal{L}_{Y_V \setminus \{a\}}(t)} Y_b(t+h) = \sum_{j=1}^p \sum_{s \in V \setminus \{a\}} e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_s D^{(j-1)} Y_s(t) + \sum_{j=1}^p e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_a P_{\mathcal{L}_{Y_V \setminus \{a\}}(t)} D^{(j-1)} Y_a(t) \quad \forall h \in [0, 1], t \in \mathbb{R}. \end{aligned}$$

We equate the two orthogonal projections and remove the coinciding terms. Then we receive  $Y_a \not\rightarrow Y_b | Y_V$  if and only if

$$\sum_{j=1}^p e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_a D^{(j-1)} Y_a(t) = \sum_{j=1}^p e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_a P_{\mathcal{L}_{Y_V \setminus \{a\}}(t)} D^{(j-1)} Y_a(t) \quad \mathbb{P}\text{-a.s.}$$

for  $h \in [0, 1]$ ,  $t \in \mathbb{R}$ . The expression on the right side is in  $\mathcal{L}_{Y_V \setminus \{a\}}(t)$  and the expression on the left side is in  $\mathcal{L}_{Y_a}(t)$ . Due to their equality, they are in  $\mathcal{L}_{Y_V \setminus \{a\}}(t) \cap \mathcal{L}_{Y_a}(t) = \{0\}$ , making use of Proposition 5.4. Thus,  $Y_a \not\rightarrow Y_b | Y_V$  if and only if

$$\sum_{j=1}^p e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_a D^{(j-1)} Y_a(t) = 0 \quad \mathbb{P}\text{-a.s.} \quad \forall h \in [0, 1], t \in \mathbb{R}. \tag{C.1}$$

In the following, we show that (C.1) is equivalent to

$$e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_a = 0 \quad \forall h \in [0, 1], j = 1, \dots, p. \tag{C.2}$$

Clearly, (C.2) implies (C.1). For the opposite direction, suppose (C.1) holds. Define the  $kp$ -dimensional vector  $\mathbf{y} = (y_1, \dots, y_{kp})$  with entries

$$y_i = \begin{cases} e_b^\top \mathbf{C} e^{A h} \mathbf{E}_j e_a & \text{if } i = (j-1)k + a, j = 1, \dots, p, \\ 0 & \text{else.} \end{cases}$$

Then (C.1) implies  $\mathbb{P}$ -a.s.

$$0 = \sum_{j=1}^p e_b^\top \mathbf{C} e^{A h} \mathbf{E}_j e_a D^{(j-1)} Y_a(t) = \sum_{j=1}^p e_b^\top \mathbf{C} e^{A h} \mathbf{E}_j e_a X_{(j-1)k+a}(t) = \mathbf{y}^\top X(t)$$

and, in particular,

$$0 = \mathbb{E} \left[ (\mathbf{y}^\top X(t))^2 \right] = \mathbf{y}^\top c_{XX}(0) \mathbf{y}.$$

But  $c_{XX}(0) > 0$  (cf. Remark 6.3(a)) such that  $\mathbf{y}$  is the zero vector and (C.2) is valid.

(b) Let  $S \subseteq V, v \in V, t \in \mathbb{R}$ , and  $h \geq 0$ . From Lemma 6.11 we already know that

$$\begin{aligned} & \frac{1}{h} P_{\mathcal{L}_{Y_S}(t)}(D^{(p-1)} Y_v(t+h) - D^{(p-1)} Y_v(t)) \\ &= \sum_{j=1}^p \sum_{s \in S} e_v^\top \mathbf{E}_p^\top \frac{(e^{A h} - I_{kp})}{h} \mathbf{E}_j e_s D^{(j-1)} Y_s(t) + \sum_{j=1}^p \sum_{s \in V \setminus S} e_v^\top \mathbf{E}_p^\top \frac{(e^{A h} - I_{kp})}{h} \mathbf{E}_j e_s P_{\mathcal{L}_{Y_S}(t)}(D^{(j-1)} Y_s(t)) \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

But  $\lim_{h \rightarrow 0} (e^{A h} - I_{kp}) / h = \mathbf{A}$  implies that

$$\text{l.i.m.}_{h \rightarrow 0} P_{\mathcal{L}_{Y_S}(t)} \left( \frac{D^{(p-1)} Y_v(t+h) - D^{(p-1)} Y_v(t)}{h} \right) = \sum_{j=1}^p \sum_{s \in S} e_v^\top \mathbf{E}_p^\top \mathbf{A} \mathbf{E}_j e_s D^{(j-1)} Y_s(t) + \sum_{j=1}^p \sum_{s \in V \setminus S} e_v^\top \mathbf{E}_p^\top \mathbf{A} \mathbf{E}_j e_s P_{\mathcal{L}_{Y_S}(t)} D^{(j-1)} Y_s(t).$$

Then the remaining proof is similar to the proof of (a).  $\square$

**Proof of Proposition 6.13.** (a) A combination of Remark 6.10 and Lemma 6.2(a) results in

$$Y_v(t+h) - P_{\mathcal{L}_{Y_V}(t)} Y_v(t+h) = e_v^\top \mathbf{C} \int_t^{t+h} e^{A(t+h-u)} \mathbf{B} dL(u).$$

Thus,  $Y_a \approx Y_b \mid Y_V$  if and only if

$$\begin{aligned} 0 &= \mathbb{E} \left[ \left( Y_a(t+h) - P_{\mathcal{L}_{Y_V}(t)} Y_a(t+h) \right) \left( Y_b(t+\tilde{h}) - P_{\mathcal{L}_{Y_V}(t)} Y_b(t+\tilde{h}) \right) \right] \\ &= \mathbb{E} \left[ \left( e_a^\top \mathbf{C} \int_t^{t+h} e^{A(t+h-u)} \mathbf{B} dL(u) \right) \left( e_b^\top \mathbf{C} \int_t^{t+\tilde{h}} e^{A(t+\tilde{h}-u)} \mathbf{B} dL(u) \right) \right] \\ &= e_a^\top \mathbf{C} \int_0^{\min(h, \tilde{h})} e^{A(h-u)} \mathbf{B} \Sigma_L \mathbf{B}^\top e^{A^\top(\tilde{h}-u)} du \mathbf{C}^\top e_b \end{aligned}$$

for  $h, \tilde{h} \in [0, 1], t \in \mathbb{R}$ .

(b) Let  $a, b, v \in V, t \in \mathbb{R}$ , and  $h \geq 0$ . An application of Lemma 6.11 gives that

$$D^{(p-1)} Y_v(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(p-1)} Y_v(t+h) = e_v^\top \mathbf{E}_p^\top \int_t^{t+h} e^{A(t+h-u)} \mathbf{B} dL(u) \quad \mathbb{P}\text{-a.s.}$$

Thus,

$$\mathbb{E} \left[ \left( D^{(p-1)} Y_a(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(p-1)} Y_a(t+h) \right) \overline{\left( D^{(p-1)} Y_b(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(p-1)} Y_b(t+h) \right)} \right] = e_a^\top \mathbf{E}_p^\top \int_0^h e^{A u} \mathbf{B} \Sigma_L \mathbf{B}^\top e^{A^\top u} du \mathbf{E}_p e_b.$$

Setting  $f(u) = e^{A u} \mathbf{B} \Sigma_L \mathbf{B}^\top e^{-A^\top u}$  and  $F(\cdot)$  as its primitive function, we obtain

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[ \left( D^{(p-1)} Y_a(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(p-1)} Y_a(t+h) \right) \overline{\left( D^{(p-1)} Y_b(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(p-1)} Y_b(t+h) \right)} \right] \\ &= e_a^\top \mathbf{E}_p^\top \left[ \lim_{h \rightarrow 0} \frac{F(h) - F(0)}{h} \right] \mathbf{E}_p e_b = e_a^\top \Sigma_L e_b. \quad \square \end{aligned}$$

**Proof of Theorem 6.19.** (a)  $\Leftarrow$ : Suppose  $e_b^\top \mathbf{C} A^\alpha \mathbf{E}_j e_a = 0$  for  $\alpha = 1, \dots, kp-1$  and  $j = 1, \dots, p$ . Bernstein [5], (11.2.1) provides

$$e^{A h} = \sum_{\alpha=0}^{kp-1} \psi_\alpha(h) \mathbf{A}^\alpha, \quad h \in \mathbb{R}, \quad \text{where } \psi_\alpha(h) = \frac{1}{2\pi i} \oint_C \frac{\chi_A^{[\alpha+1]}(z)}{\chi_A(z)} e^{t z} dz, \tag{C.3}$$

$\chi_A^{[1]}(\cdot), \dots, \chi_A^{[kp]}(\cdot)$  are polynomials defined by recursion and  $C$  is a simple, closed contour in the complex plane enclosing  $\sigma(\mathbf{A})$ . With  $e_b^\top \mathbf{C} A^\alpha \mathbf{E}_j e_a = 0$  we can conclude then that

$$e_b^\top \mathbf{C} e^{A h} \mathbf{E}_j e_a = \sum_{\alpha=0}^{kp-1} \psi_\alpha(h) e_b^\top \mathbf{C} A^\alpha \mathbf{E}_j e_a = 0 \quad \forall h \in [0, 1],$$

such that Proposition 6.12 results in  $Y_a \not\rightarrow Y_b | Y_V$ .

$\Rightarrow$ : Assume  $Y_a \not\rightarrow Y_b | Y_V$ . Thus,  $e_b^\top C e^{A_h} E_j e_a = 0$  for  $h \in [0, 1]$  and  $j = 1, \dots, p$  by Proposition 6.12. Define

$$f(h) = e_b^\top C e^{A_h} E_j e_a, \quad h \in \mathbb{R},$$

and differentiate this function using Bernstein [5], Proposition 11.1.4. Then

$$f^{(\alpha)}(h) = e_b^\top C A^\alpha e^{A_h} E_j e_a, \quad h \in \mathbb{R}, \alpha = 1, \dots, kp - 1.$$

Since  $f(h) = 0$  for  $h \in [0, 1]$  and  $f^{(\alpha)}(\cdot)$  is continuous, we obtain  $f^{(\alpha)}(h) = 0$  for  $h \in [0, 1]$ . Putting  $h = 0$ , we get as claimed

$$0 = e_b^\top C A^\alpha E_j e_a, \quad \alpha = 1, \dots, kp - 1, j = 1, \dots, p.$$

(b)  $\Leftarrow$ : Let  $e_a^\top C A^\alpha B \Sigma_L B^\top (A^\top)^\beta C^\top e_b = 0$  for  $\alpha, \beta = 0, \dots, kp - 1$ . We apply the representation (C.3) and obtain

$$e_a^\top C \int_0^{\min(h, \tilde{h})} e^{A(h-s)} B \Sigma_L B^\top e^{A^\top(\tilde{h}-s)} ds C^\top e_b = \sum_{\alpha=0}^{kp-1} \sum_{\beta=0}^{kp-1} \int_0^{\min(h, \tilde{h})} \psi_\alpha(h-s) \varphi_\beta(\tilde{h}-s) e_a^\top C A^\alpha B \Sigma_L B^\top (A^\top)^\beta C^\top e_b ds = 0,$$

for  $h, \tilde{h} \in [0, 1], t \in \mathbb{R}$ , by assumption. Proposition 6.13 yields then  $Y_a \approx Y_b | Y_V$ .

$\Rightarrow$ : Assume  $Y_a \approx Y_b | Y_V$ . Due to Theorem 4.5 we have for  $h \in [0, 1]$  and  $t \in \mathbb{R}$ ,

$$P_{\mathcal{L}_{Y_V(t)} \vee \mathcal{L}_{Y_b(t, t+1)}} Y_a(t+h) = P_{\mathcal{L}_{Y_V(t)}} Y_a(t+h) \quad \mathbb{P}\text{-a.s.}$$

In addition, we know from Proposition 6.9 that  $P_{\mathcal{L}_{Y_V(t)}} Y_a(t+h) = e_a^\top C e^{A_h} X(t)$ . Both together provide

$$P_{\mathcal{L}_{Y_V(t)} \vee \mathcal{L}_{Y_b(t, t+1)}} Y_a(t+h) = e_a^\top C e^{A_h} X(t) \quad \mathbb{P}\text{-a.s.} \tag{C.4}$$

for  $h \in [0, 1]$  and  $t \in \mathbb{R}$ . Since  $Y_b(t+\tilde{h}) \in \mathcal{L}_{Y_V(t)} \vee \mathcal{L}_{Y_b(t, t+1)}$  for  $\tilde{h} \in [0, 1]$  as well as  $Y_a(t+h) - P_{\mathcal{L}_{Y_V(t)} \vee \mathcal{L}_{Y_b(t, t+1)}} Y_a(t+h) \in (\mathcal{L}_{Y_V(t)} \vee \mathcal{L}_{Y_b(t, t+1)})^\perp$ , we obtain

$$0 = \mathbb{E} \left[ \left( Y_a(t+h) - P_{\mathcal{L}_{Y_V(t)} \vee \mathcal{L}_{Y_b(t, t+1)}} Y_a(t+h) \right) Y_b(t+\tilde{h}) \right].$$

Plugging in (C.4) gives

$$0 = \mathbb{E} \left[ \left( Y_a(t+h) - e_a^\top C e^{A_h} X(t) \right) Y_b(t+\tilde{h}) \right] = e_a^\top C \mathbb{E} \left[ \left( X(t+h) - e^{A_h} X(t) \right) X(t+\tilde{h}) \right] C^\top e_b = e_a^\top C \left( c_{XX}(h-\tilde{h}) - e^{A_h} c_{XX}(-\tilde{h}) \right) C^\top e_b,$$

for  $h, \tilde{h} \in [0, 1]$ . If we only consider the case  $0 \leq \tilde{h} \leq h \leq 1$  then (6.5) provides

$$0 = e_a^\top C \left( e^{A(h-\tilde{h})} c_{XX}(0) - e^{A_h} c_{XX}(0) e^{A^\top \tilde{h}} \right) C^\top e_b = e_a^\top C e^{A_h} \left( e^{-A \tilde{h}} c_{XX}(0) - c_{XX}(0) e^{A^\top \tilde{h}} \right) C^\top e_b,$$

using Bernstein [5], Corollary 11.1.6. Now, we define

$$\gamma(h, \tilde{h}) = e_a^\top C e^{A_h} \left( e^{-A \tilde{h}} c_{XX}(0) - c_{XX}(0) e^{A^\top \tilde{h}} \right) C^\top e_b, \quad 0 \leq \tilde{h} \leq h \leq 1.$$

Differentiating this function several times (cf. Bernstein [5], Proposition 11.1.4) provides

$$\frac{\partial^m}{\partial h^m} \frac{\partial^n}{\partial \tilde{h}^n} \gamma(h, \tilde{h}) = e_a^\top C A^m e^{A_h} \left( (-A)^n e^{-A \tilde{h}} c_{XX}(0) - c_{XX}(0) (A^\top)^n e^{A^\top \tilde{h}} \right) C^\top e_b.$$

Furthermore, since  $\gamma(h, \tilde{h}) = 0$  for  $0 \leq \tilde{h} \leq h \leq 1$  and due to the continuity of the function under consideration, we obtain that the derivatives are zero for  $0 \leq \tilde{h} \leq h \leq 1$ . Now, plugging in  $h = \tilde{h} = 0$  yields

$$e_a^\top C A^m c_{XX}(0) (A^\top)^n C^\top e_b = e_a^\top C A^m (-A)^n c_{XX}(0) C^\top e_b, \quad m, n \in \mathbb{N}_0. \tag{C.5}$$

Finally, (6.6) leads to

$$\begin{aligned} & e_a^\top C A^\alpha B \Sigma_L B^\top (A^\top)^\beta C^\top e_b \\ &= e_a^\top C A^\alpha \left( -A c_{XX}(0) - c_{XX}(0) A^\top \right) (A^\top)^\beta C^\top e_b \\ &= -e_a^\top C A^{\alpha+1} c_{XX}(0) (A^\top)^\beta C^\top e_b - e_a^\top C A^\alpha c_{XX}(0) (A^\top)^{\beta+1} C^\top e_b. \end{aligned}$$

Applying (C.5) gives then

$$e_a^\top C A^\alpha B \Sigma_L B^\top (A^\top)^\beta C^\top e_b = -e_a^\top C (-1)^\beta A^{\alpha+\beta+1} c_{XX}(0) C^\top e_b - e_a^\top C (-1)^{\beta+1} A^{\alpha+\beta+1} c_{XX}(0) C^\top e_b = 0,$$

for  $\alpha, \beta = 0, \dots, kp - 1$ , the desired statement.  $\square$

#### Appendix D. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.spa.2024.104501>.

## References

- [1] O.O. Aalen, Dynamic modelling and causality, *Scand. Actuar. J.* 1987 (3–4) (1987) 177–190.
- [2] S.A. Andersson, D. Madigan, M.D. Perlman, Alternative Markov properties for chain graphs, *Scand. Stat. Theory Appl.* 28 (1) (2001) 33–85.
- [3] T. Bergmann, G. Hartwigsen, Inferring causality from noninvasive brain stimulation in cognitive neuroscience, *J. Cogn. Neurosci.* 33 (2) (2021) 195–224.
- [4] A.R. Bergstrom, Gaussian estimation of mixed-order continuous-time dynamic models with unobservable stochastic trends from mixed stock and flow data, *Econom. Theory* 13 (1997) 467–505.
- [5] D.S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*, second ed., Princeton University Press, Princeton, 2009.
- [6] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.
- [7] D.R. Brillinger, *Time Series: Data Analysis and Theory*, Society for Industrial and Applied Mathematics, Philadelphia, 2001.
- [8] P. Brockwell, Recent results in the theory and applications of CARMA processes, *Ann. Inst. Stat. Math.* 66 (2014) 647–685.
- [9] P. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, second ed., Springer, New York, 1991.
- [10] P. Brockwell, A. Lindner, Prediction of Lévy-driven CARMA processes, *J. Econ.* 189 (2) (2015) 263–271.
- [11] P. Brockwell, A. Lindner, *Continuous-Parameter Time Series*, De Gruyter, 2024.
- [12] T. Bühler, D.A. Salamon, *Functional Analysis*, American Mathematical Society, Providence, 2018.
- [13] G. Chamberlain, The general equivalence of Granger and Sims causality, *Econometrica* 50 (3) (1982) 569–581.
- [14] D. Commenges, A. Gégout-Petit, A general dynamical statistical model with causal interpretation, *J. R. Stat. Soc. Ser. B.* 71 (3) (2009) 719–736.
- [15] F. Comte, E. Renault, Noncausality in continuous time models, *Econom. Theory* 12 (2) (1996) 215–256.
- [16] L. Cox, D. Popken, Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? *Ann. Epidemiol.* 25 (3) (2015) 162–173.
- [17] H. Cramér, On the theory of stationary random processes, *Ann. of Math.* 41 (1) (1940) 215–230.
- [18] H. Cramér, On the structure of purely non-deterministic stochastic processes, *Ark. Mat.* 4 (1961) 249–266.
- [19] H. Cramér, Stochastic processes as curves in Hilbert space, *Theory Probab. Appl.* 9 (2) (1964) 169–179.
- [20] H. Cramér, Structural and statistical problems for a class of stochastic processes: The first Samuel Stanley Wilks lecture at Princeton university, March 7, 1970, Princeton University Press, 1971.
- [21] V. Didelez, Asymmetric separation for local independence graphs, in: R. Dechter, T. Richardson (Eds.), *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, AUAI Press, Arlington, 2006, pp. 130–137.
- [22] V. Didelez, Graphical models for composable finite Markov processes, *Scand. J. Stat.* 34 (1) (2007) 169–185.
- [23] V. Didelez, Graphical models for marked point processes based on local independence, *J. R. Stat. Soc. Ser. B.* 70 (1) (2008) 245–264.
- [24] J.L. Doob, The elementary Gaussian processes, *Ann. Math. Stat.* 15 (3) (1944) 229–282.
- [25] J.L. Doob, *Stochastic Processes*, third ed., Wiley, New York, 1953.
- [26] J.-M. Dufour, E. Renault, Short run and long run causality in time series: Theory, *Econometrica* 66 (5) (1998) 1099–1126.
- [27] M. Eichler, *Markov Properties for Graphical Time Series Models*, Technical Report, University of Heidelberg, Heidelberg, 2001.
- [28] M. Eichler, Granger causality and path diagrams for multivariate time series, *J. Econ.* 137 (2) (2007) 334–353.
- [29] M. Eichler, Graphical modelling of multivariate time series, *Probab. Theory Related Fields* 153 (2012) 233–268.
- [30] M. Eichler, Causal inference with multiple time series: Principles and problems, *Phil. Trans. R. Soc. A* 371 (1997) (2013).
- [31] M. Eichler, R. Dahlhaus, J. Dueck, Graphical modeling for multivariate Hawkes processes with nonparametric link functions, *J. Time Series Anal.* 38 (2) (2017) 225–242.
- [32] V. Fasen-Hartmann, L. Schenk, Mixed orthogonality graphs for continuous-time state space models and orthogonal projections, *J. Time Series Anal.* (2024+) to appear.
- [33] V. Fasen-Hartmann, L. Schenk, *Partial correlation graphs for continuous-parameter time series*, 2024, arXiv:2401.16970.
- [34] I. Feshchenko, On closeness of the sum of  $n$  subspaces of a Hilbert space, *Ukrainian Math. J.* 63 (2012) 1566–1622.
- [35] J.-P. Florens, D. Fougère, Noncausality in continuous time, *Econometrica* 64 (5) (1996) 1195–1212.
- [36] J.-P. Florens, M. Mouchart, A note on noncausality, *Econometrica* 50 (3) (1982) 583–591.
- [37] J.-P. Florens, M. Mouchart, A linear theory for noncausality, *Econometrica* 53 (1) (1985) 157–175.
- [38] I.I. Gihman, A.V. Skorokhod, *The Theory of Stochastic Processes I*, Springer, Berlin, 2004.
- [39] E.G. Gladyshev, On multi-dimensional stationary random processes, *Theory Probab. Appl.* 3 (4) (1958) 425–428.
- [40] C. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* 37 (3) (1969) 424–438.
- [41] P.R. Halmos, *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, second ed., Chelsea Publ., New York, 1957.
- [42] A.C. Harvey, J.H. Stock, Continuous time autoregressive models with common stochastic trends, *J. Econom. Dynam. Control* 12 (2–3) (1985) 365–384.
- [43] A.C. Harvey, J.H. Stock, The estimation of higher-order continuous time autoregressive models, *Econom. Theory* 1 (1) (1985) 97–112.
- [44] A.C. Harvey, J.H. Stock, Estimating integrated higher-order continuous time autoregressions with an application to money-income causality, *J. Econ.* 42 (3) (1989) 319–336.
- [45] S. Heerah, R. Molinari, S. Guerrier, A. Marshall-Colon, Granger-causal testing for irregularly sampled time series with application to nitrogen signalling in Arabidopsis, *Bioinformatics* 37 (16) (2021) 2450–2460.
- [46] G.W. Imbens, Causality in econometrics: Choice vs. chance, *Econometrica* 90 (6) (2022) 2541–2566.
- [47] J.T.A. Koster, On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors, *Scand. J. Stat.* 26 (3) (1999) 413–431.
- [48] G.M. Kuersteiner, Granger-sims causality, in: S.N. Durlauf, L.E. Blume (Eds.), *Macroeconometrics and Time Series Analysis*, Palgrave Macmillan, 2010, pp. 5413–5425.
- [49] R. Kuzma, I.J. Cruickshank, K.M. Carley, Influencing the influencers: Evaluating person-to-person influence on social networks using granger causality, in: R.M. Benito, C. Cherifi, H. Cherifi, E. Moro, L.M. Rocha, M. Sales-Pardo (Eds.), *Complex Networks & their Applications X*, Springer, Cham, 2022, pp. 89–99.
- [50] S.L. Lauritzen, *Graphical Models*, second ed., Clarendon Press, Oxford, 2004.
- [51] J. Lemmens, P. Valkenburg, J. Peter, The effects of pathological gaming on aggressive behavior, *J. Youth Adolesc.* 40 (2011) 38–47.
- [52] M. Levitz, M.D. Perlman, D. Madigan, Separation and completeness properties for AMP chain graph Markov models, *Ann. Statist.* 29 (6) (2001) 1751–1784.
- [53] A. Lindquist, G. Picci, *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*, Springer, Berlin, Heidelberg, 2015.
- [54] M. Maathuis, M. Drton, S. Lauritzen, M. Wainwright, *Handbook of Graphical Models*, CRC Press, Boca Raton, 2019.
- [55] T. Marquardt, Multivariate fractionally integrated CARMA processes, *J. Multivariate Anal.* 98 (9) (2007) 1705–1725.
- [56] T. Marquardt, R. Stelzer, Multivariate CARMA processes, *Stoch. Process. Appl.* 117 (1) (2007) 96–120.
- [57] H. Masuda, On multidimensional ornstein-uhlenbeck processes driven by a general Lévy process, *Bernoulli* 10 (1) (2004) 97–120.
- [58] R. Matveev, On multi-dimensional regular stationary processes, *Theory Probab. Appl.* 6 (2) (1961) 149–165.
- [59] S.W. Mogensen, N.R. Hansen, Markov equivalence of marginalized local independence graphs, *Ann. Math. Stat.* 48 (1) (2020) 539–559.
- [60] S.W. Mogensen, N.R. Hansen, Graphical modeling of stochastic processes driven by correlated noise, *Bernoulli* 28 (4) (2022) 3023–3050.
- [61] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, third ed., Morgan Kaufmann, San Francisco, 1994.

- [62] L. Petrovic, S. Dimitrijevic, Causality with finite horizon of the past in continuous time, *Statist. Probab. Lett.* 82 (7) (2012) 1219–1223.
- [63] M.B. Priestley, *Spectral Analysis and Time Series*, Vol. II, Academic Press, London, 1981.
- [64] E. Renault, A. Szafarz, *True Versus Spurious Instantaneous Causality*, Universite Libre de Bruxelles, Bruxelles, 1991.
- [65] T. Richardson, Markov properties for acyclic directed mixed graphs, *Scand. J. Stat.* 30 (1) (2003) 145–157.
- [66] K. Røysland, P. Ryalen, M. Nygård, V. Didelez, Graphical criteria for the identification of marginal causal effects in continuous-time survival and event-history analyses, *J. R. Stat. Soc. Ser. B.* (2024) 1–24.
- [67] J.A. Rozanov, *Stationary Random Processes*, Holden-Day, San Francisco, 1967.
- [68] E. Schlemm, R. Stelzer, Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes, *Electron. J. Stat.* 6 (2012) 2185–2234.
- [69] T. Schweder, Composable Markov processes, *J. Appl. Probab.* 7 (2) (1970) 400–410.
- [70] A. Shojaie, E.B. Fox, Granger causality: A review and recent advances, *Annu. Rev. Stat. Appl.* 9 (1) (2022) 289–319.
- [71] C.A. Sims, Money, income, and causality, *Amer. Econ. Rev.* 62 (4) (1972) 540–552.
- [72] P. Spirtes, T. Richardson, C. Meek, R. Scheines, C. Glymour, Using path diagrams as a structural equation modeling tool, *Sociol. Methods Res.* 27 (2) (1998) 182–225.
- [73] J. Weidmann, *Linear Operators in Hilbert Spaces*, Springer, New York, 1980.
- [74] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, second ed., Wiley, Chichester, 2008.
- [75] S. Wright, Correlation and causation, *J. Agric. Res.* 20 (3) (1921) 557–585.
- [76] S. Wright, The method of path coefficients, *Ann. Math. Stat.* 5 (3) (1934) 161–215.
- [77] K. Zadnik, L. Jones, B. Irvin, R.N. Kleinstein, R.E. Manny, J.A. Shin, D.O. Mutti, Myopia and ambient night-time lighting, *Nature* 404 (2000) 143–144.