JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. (2024)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12787

ORIGINAL ARTICLE

MIXED ORTHOGONALITY GRAPHS FOR CONTINUOUS-TIME STATE SPACE MODELS AND ORTHOGONAL PROJECTIONS

VICKY FASEN-HARTMANN AND LEA SCHENK
Institute of Stochastics, Karlsruhe Institute of Technology, Karlsruhe, Germany

In this article, we derive (local) orthogonality graphs for the popular continuous-time state space models, including in particular multivariate continuous-time ARMA (MCARMA) processes. In these (local) orthogonality graphs, vertices represent the components of the process, directed edges between the vertices indicate causal influences and undirected edges indicate contemporaneous correlations between the component processes. We present sufficient criteria for state space models to satisfy the assumptions of Fasen-Hartmann and Schenk (2024a) so that the (local) orthogonality graphs are well-defined and various Markov properties hold. Both directed and undirected edges in these graphs are characterised by orthogonal projections on well-defined linear spaces. To compute these orthogonal projections, we use the unique controller canonical form of a state space model, which exists under mild assumptions, to recover the input process from the output process. We are then able to derive some alternative representations of the output process and its highest derivative. Finally, we apply these representations to calculate the necessary orthogonal projections, which culminate in the characterisations of the edges in the (local) orthogonality graph. These characterisations are given by the parameters of the controller canonical form and the covariance matrix of the driving Lévy process.

Received 7 November 2023; Accepted 4 October 2024

Keywords: Contemporaneous correlation; controller canonical form; Granger causality; Markov property; MCARMA process; graphs; orthogonal projections; state space models.

JEL. Primary 62H22; 62M20; Secondary 62M10; 60G25.

1. INTRODUCTION

State space models are important tools in many scientific and engineering disciplines, including control theory, statistics and computational neuroscience. In this article, we study the time-invariant $\mathbb{R}^k$-valued state space model $(\mathcal{A}, \mathcal{B}, \mathcal{C}, L)$ of dimension kp, that is characterised by a driving $\mathbb{R}^k$-valued Lévy process $L = (L(t))_{t\in\mathbb{R}}$, a state transition matrix $\mathcal{A} \in \mathbb{R}^{kp\times kp}$ with $p \in \mathbb{N}$, an input matrix $\mathcal{B} \in \mathbb{R}^{kp\times k}$, and an observation matrix $\mathcal{C} \in \mathbb{R}^{kp\times k}$. Note that an $\mathbb{R}^k$-valued Lévy process L is a stochastic process with stationary and independent increments, it is continuous in probability, and satisfies $L(0) = 0_k \in \mathbb{R}^k$ almost surely (Sato, 2007). A continuous-time state space model then consists of a state equation

$$dX(t) = \mathcal{A}X(t)dt + \mathcal{B}dL(t), \tag{1.1}$$

and an observation equation

$$Y(t) = \mathcal{C}X(t).$$

* Correspondence to: Vicky Fasen-Hartmann, Institute of Stochastics, Karlsruhe Institute of Technology, Englerstr. 2, D-76131 Karlsruhe, Germany.
E-mail: vicky.fasen@kit.edu

The $\mathbb{R}^{kp}$-valued process $X = (X(t))_{t\in\mathbb{R}}$ is the input process and the $\mathbb{R}^k$-valued process $Y = (Y(t))_{t\in\mathbb{R}}$ is the output process. It is well known that the solution of the state equation (1.1) is

$$X(t) = e^{\mathcal{A}(t-s)}X(s) + \int_s^t e^{\mathcal{A}(t-u)}\mathcal{B}dL(u), \quad s < t. \tag{1.2}$$

A special subclass of such state space models is the popular multivariate continuous-time ARMA (MCARMA) models (Marquardt and Stelzer, 2007; Schlemm and Stelzer, 2012a, 2012b). In contrast to Schlemm and Stelzer (2012a), we speak here of a subclass instead of the equivalence of these classes because in our opinion there is an argument in the proof that is not clearly verifiable; the details are given in Section 3.2.

This article aims to construct a graphical model for such state space models. The interest in graphical models for stochastic processes has increased significantly in recent years, see, for example, Mogensen and Hansen (2020, 2022); Basu *et al.* (2015); Eichler (2007, 2012); Didelez (2007, 2008); Fasen-Hartmann and Schenk (2024a, 2024b), although the use of graphical models to visualise and analyse dependence structures in stochastic models is quite old (Wright, 1921, 1934). A major reason for this surge in interest is the simplicity and clarity of graphical models in representing the dependence structure in stochastic models such that examples of practical applications are ubiquitous. Another big advantage is their ease of implementation on computers, making them a powerful tool for the analysis of high-dimensional time series, as demonstrated, for example, in Eichler (2007). The state of the art of graphical models is presented in Maathuis et al. (2019).

In this article, we use the approach of Fasen-Hartmann and Schenk (2024a) to construct orthogonality graphs and local orthogonality graphs for state space models, and to derive analytic representations of the edges in these graphs by the model parameters. (Local) orthogonality graphs are mixed graphs where the vertices $V = \{1, \dots, k\}$ represent the different component series $Y_v = (Y_v(t))_{t\in\mathbb{R}}$, $v \in V$, of an $\mathbb{R}^k$-valued process $Y_V = Y$. Directed edges reflect (local) Granger causality and undirected edges reflect (local) contemporaneous correlation between the component series of the stochastic process. An attractive property of the (local) orthogonality graph is that it satisfies several types of Markov properties under fairly general assumptions.

The mathematical notion of causality can be traced back to Granger (1969) and Sims (1972) and has since then been extended and applied in various fields, see Shojaie and Fox (2022) for an excellent survey. In our context of continuous-time stochastic processes, the notion of Granger causality and contemporaneous correlation, as defined in Fasen-Hartmann and Schenk (2024a), are based on conditional orthogonality relations of linear subspaces generated by subprocesses, similarly to Eichler (2007) in discrete time. This setup is perfectly suitable for stationary stochastic processes. At the same time, the approaches of Eichler (2012) using conditional independence relations for stochastic processes in discrete time and that of Mogensen and Hansen (2020, 2022); Didelez (2007, 2008); Eichler *et al.* (2017) using conditional local independence are suitable for semimartingales and point processes. We refer to Fasen-Hartmann and Schenk (2024a) for a detailed overview of graphical models for stochastic processes and the advantages of the different approaches.

The conditional orthogonality relations in the definition of (local) Granger causality and (local) contemporaneous correlation can be expressed equivalently by orthogonal projections of the component processes $Y_v(t + h)$ ($t \in \mathbb{R}$, $h \geq 0$, $v \in V$) and their highest derivative respectively, on well-defined linear subspaces. To the best of our knowledge, the orthogonal projections for multivariate state space models and their derivatives required in this article have not yet been addressed in the existing literature. Although Rozanov (1967), III, 5, is devoted to the topic of predictions for general stationary processes, the representations in that book are based on a specific maximal decomposition of the spectral density matrix of that process. This decomposition is generally not expressible as a simple formula, and so he only considers univariate examples. The orthogonal projections of univariate CARMA processes were discussed in the previous paper by Brockwell and Lindner (2015). They provide representations for the linear projection of a CARMA process $Y(h)$ onto the entire linear space generated by the CARMA process up to time $t = 0$, and for the conditional expectation of $Y(h)$ on the σ-algebra generated by the CARMA process up to time $t = 0$. A multivariate generalisation of the conditional expectation result using the σ-algebra generated by the whole multivariate CARMA (MCARMA) process up to time $t = 0$ can be found in Basse-O'Connor

DOI: 10.1111/jtsa.12787

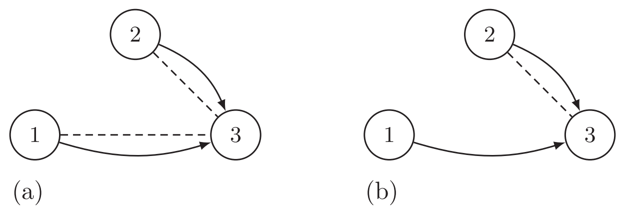

Figure 1. In the left figure is the orthogonality graph and in the right figure the local orthogonality graph of the three-dimensional ICCSS(2, 1) process defined in Example 3.8, (a) Orthogonality graph, (b) Local orthogonality graph

et al. (2019), Corollary 4.11, but the statement is not consistent with the comprehensible result of Brockwell and Lindner (2015). In any case, in this article, we require not only projections of $Y_v(t+h)$ on the linear space of the past of the process Y up to time t, but also on linear subspaces generated by subprocesses, and additionally the projections of the highest derivatives of $Y_v(t+h)$.

In the context of multivariate continuous-time AR (MCAR) processes, which are a subclass of MCARMA processes, the topic of orthogonal projections and the corresponding (local) orthogonality graphs are discussed in Fasen-Hartmann and Schenk (2024a). Although MCAR processes are state space models, the techniques used there are not applicable to MCARMA(p, q) processes with $q > 0$, because MCAR models have a much simpler structure. This structure allows, e.g., the direct recovery of the input process X from the output process Y. In particular, Fasen-Hartmann and Schenk (2024a) use the orthogonal projections to develop (local) orthogonality graphs for MCAR processes and give interpretations for the edges which correspond to other results in the literature; see, e.g., Comte and Renault (1996) for causality relations of MCAR processes and Eichler (2007) for discrete-time AR processes. This article can therefore be seen as an extension of Fasen-Hartmann and Schenk (2024a) to a broader class of models and of course, we compare our results with those of that article. We would like to point out here that, according to our knowledge, even for stochastic processes in discrete time, the literature on mixed graphical models is restricted to AR processes (Eichler, 2007, 2012), there is not much known on mixed graphical models for the more complex ARMA processes satisfying some types of Markov properties.

In this article, the controller canonical form of a state space model plays an important role in calculating the orthogonal projections of $Y_v(t+h)$ and its highest derivative, as highlighted in Basse-O'Connor *et al.* (2019). The controller canonical form of an MCARMA process has been studied in Brockwell and Schlemm (2013) and is the multivariate generalisation of the definition of a univariate CARMA process (Brockwell, 2014). In the case of the existence of a controller canonical form, we show that this representation is unique, which is essential for the unique characterisations of the edges in the (local) orthogonality graph later in this article; we will prove that the edges depend only on the model parameters of the controller canonical form and the covariance matrix of the driving Lévy process. A special feature of the controller canonical form is that under very general assumptions it can be used to recover the input process X from the output process Y, i.e., $X(t)$ lies in the closed linear space generated by $(Y(s))_{s\leq t}$ and which has only been known for univariate CARMA processes (Brockwell and Lindner, 2015, Theorem 2.2). In this case, we call Y an invertible controller canonical state space (ICCSS) process and we are able to calculate the necessary orthogonal projections to describe the (local) Granger causality and (local) contemporaneous correlation relations of the underlying process, resulting in the characterisations of the edges in the (local) orthogonality graph. An example of both graphs for a three-dimensional ICCSS(2, 1) process is given in Figure 1. This example is used for illustration throughout the article.

In conclusion, we show in this article that not only for ICCSS processes but also for most state space models, the (local) orthogonality graph exists and satisfies several of the preferred Markov properties. In our opinion, this is the first (mixed) graphical model for this popular and broad class of stochastic processes. However, for the explicit representation of the edges via (local) Granger causality and (local) contemporaneous correlation, we need the invertibility of the state space model and hence, the restriction to ICCSS processes. In addition, we derive new and important results for state space models, such as sufficient criteria for being an ICCSS model, alternative representations for $Y_v(t+h)$ and its highest derivative depending on the linear past of Y up to time t

and an independent noise term, and, in particular, their orthogonal projections onto linear subspaces, which are the basis for linear predictions.

The article is structured as follows. In Section 3, we introduce (local) orthogonality graphs, the controller canonical form of a state space model, and ICCSS processes. We also present their basic properties that are important for this article. In Section 4, we consider orthogonal projections of ICCSS processes and their highest derivatives onto linear subspaces generated by subprocesses, since the characterisations of the edges of the (local) orthogonality graph are based on these orthogonal projections. These results then lead to the existence of the (local) orthogonality graph and the characterisation of the directed and the undirected edges of an ICCSS process by its model parameters in Section 5. The proofs of the article are given in their own Section 6.

2. NOTATION

From now on we call the space of all real and 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. For $\boldsymbol{A} \in M_{k,d}(\mathbb{C})$ we write $\boldsymbol{A}^\top$ for the transpose of $\boldsymbol{A}$ and for $\boldsymbol{A} \in M_k(\mathbb{C})$ we write $\boldsymbol{A} \geq 0$ if $\boldsymbol{A}$ is positive semi-definite, and $\boldsymbol{A} > 0$ if $\boldsymbol{A}$ is positive definite. Furthermore, $\sigma(\boldsymbol{A})$ are the eigenvalues of $\boldsymbol{A}$. I_k is the $(k \times k)$-dimensional identity matrix, $0_{k\times d}$ is the $(k \times d)$-dimensional zero matrix and 0_k is either the k-dimensional zero vector or the $(k \times k)$-dimensional zero matrix which should be clear from the context. The vector $e_a \in \mathbb{R}^k$ is the a-th unit vector, as well as

$$\mathbf{E}_j = \begin{pmatrix} 0_{k(j-1)\times k} \\ I_k \\ 0_{k(p-j)\times k} \end{pmatrix} \in M_{kp\times k}(\mathbb{R}), \quad j = 1, \dots, p. \tag{2.1}$$

Furthermore, we write for $p > q$,

$$\boldsymbol{E}^\top = (I_{kq}, 0_{k(p-q)\times kq}) \in M_{kq\times kp}(\mathbb{R}) \quad \text{and} \quad \mathsf{E}^\top = (0_{k\times k(q-1)}, I_k) \in M_{k\times kq}(\mathbb{R}). \tag{2.2}$$

For a matrix polynomial $P(z)$ we define the set of zeros of the polynomial $\det(P(z))$ as $\mathcal{N}(P) = \{z \in \mathbb{C} : \det(P(z)) = 0\}$ and $\deg(\det(P(z)))$ denotes the degree of the polynomial $\det(P(z))$. $Y = (Y(t))_{t\in\mathbb{R}}$ is a k-dimensional stationary and mean-square continuous stochastic process with index set $V = \{1, \dots, k\}$ and expectation zero. The corresponding components are denoted by $Y_v = (Y_v(t))_{t\in\mathbb{R}}$ for $v \in V$ and multivariate subprocesses are denoted by $Y_S = (Y_s)_{s\in S} = (Y_S(t))_{t\in\mathbb{R}}$ for $S \subseteq V$. A special case is $S = V$ where $Y_V = Y$. Finally, l.i.m. is the mean square limit.

3. PRELIMINARIES

The topic of this article is mixed (local) orthogonality graphs for state space models. Therefore, in Section 3.1, we introduce (local) orthogonality graphs and present the main results of Fasen-Hartmann and Schenk (2024a) that are relevant to this article. Then, in Sections 3.2 and 3.3, the controller canonical form of a state space model is introduced, and the invertibility of these processes is discussed.

3.1. Mixed Orthogonality Graphs

Orthogonality graphs are graphical models $G_{\mathrm{OG}} = (V, E_{\mathrm{OG}})$ where the vertices $V = \{1, \dots, k\}$ represent the different component series $Y_v = (Y_v(t))_{t\in\mathbb{R}}$, $v \in V$, of a k-dimensional wide-sense stationary and mean-square continuous stochastic process $Y = (Y(t))_{t\in\mathbb{R}}$ with expectation zero. The vertices are connected with both directed and undirected edges E_{OG}. A directed edge represents a Granger causal relationship while an undirected edge represents a contemporaneous correlation between the components. To make this clear, we first present some notations that we require for the definitions of Granger causality and contemporaneous correlation. For any set $S \subseteq V$, $t, \tilde{t} \in \mathbb{R}$ with $t < \tilde{t}$, we define the closed linear spaces generated by the subprocess $Y_S = (Y_s)_{s\in S} = (Y_S(t))_{t\in\mathbb{R}}$ as

$$\mathcal{L}_{Y_S}(t,\tilde{t}) = \overline{\left\{ \sum_{i=1}^{n} \sum_{s\in S} \gamma_{s,i} Y_s(t_i) \,:\, \gamma_{s,i} \in \mathbb{C}, t \le t_1 \le \cdots \le t_n \le \tilde{t}, n \in \mathbb{N} \right\}},$$
$$\mathcal{L}_{Y_S}(t) = \overline{\left\{ \sum_{i=1}^{n} \sum_{s\in S} \gamma_{s,i} Y_s(t_i) \,:\, \gamma_{s,i} \in \mathbb{C}, -\infty < t_1 \le \cdots \le t_n \le t, n \in \mathbb{N} \right\}}.$$

We further denote the orthogonal projection of $Z \in L^2$ on the closed linear subspace $\mathcal{L} \subseteq L^2$ by $P_{\mathcal{L}}(Z) = P_{\mathcal{L}}Z$. Now, we establish definitions of Granger causality, which characterises the directed edges in the (local) orthogonality graph.

Definition 3.1. Let $a, b \in S \subseteq V$ and $a \neq b$.

(a) Y_a is Granger non-causal for Y_b with respect to Y_S, if and only if, for all $t \in \mathbb{R}$ and $0 \le h \le 1$,

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

We shortly write $Y_a \nrightarrow Y_b \mid Y_S$.

(b) Suppose Y_v 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 $D^{(0)}Y_v = Y_v$. Then Y_a is locally Granger non-causal for Y_b with respect to Y_S, if and only if, for all $t \in \mathbb{R}$,

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

We shortly write $Y_a \nrightarrow_0 Y_b \mid Y_S$.

In words, Y_a is Granger non-causal for Y_b with respect to Y_S if the prediction of $Y_b(t+h)$ based on the linear information available at time t provided by the past and present of Y_S is not diminished by removing the linear information provided by the past and present values of Y_a. Local Granger non-causality considers the limiting case $h \to 0$, where the highest existing derivative of the process must be examined to obtain a non-trivial criterion; see as well the discussion in Fasen-Hartmann and Schenk (2024a) for the motivation of these definitions.

Remark 3.2. Fasen-Hartmann and Schenk (2024a) originally defined Granger causality by conditional orthogonality of linear spaces generated by subprocess, and then showed that these definitions are equivalent to the definitions based on the orthogonal projections given above (see Theorem 3.5 and Remark 3.12 therein).

Next, the undirected edges are characterised by (linear) contemporaneous correlation. The idea is simple: there is no undirected influence between Y_a and Y_b with respect to Y_S if and only if, given the amount of information provided by the past of Y_S up to time t, Y_a and Y_b are uncorrelated in the future. Fasen-Hartmann and Schenk (2024a) introduce the following definitions.

Definition 3.3. Let $a, b \in S \subseteq V$ and $a \neq b$.

(a) Y_a and Y_b are contemporaneously uncorrelated with respect to Y_S, if and only if, for all $t \in \mathbb{R}$ and $0 \le h, \tilde{h} \le 1$,

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

We shortly write $Y_a \nsim Y_b \mid Y_S$.

(b) Suppose Y_v is j_v-times mean-square differentiable but the $(j_v + 1)$-derivative does not exist for $v \in V$. Then Y_a and Y_b are locally contemporaneously uncorrelated with respect to Y_S, if and only if, for all $t \in \mathbb{R}$,

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

We shortly write $Y_a \nsim_0 Y_b \,|\, Y_S$.

Remark 3.4. The original definition of contemporaneous uncorrelatedness in Fasen-Hartmann and Schenk (2024a) is again by conditional orthogonality of linear spaces generated by subprocesses, and equivalent characterisations using orthogonal projections are given there (see Lemma 4.3, Theorem 4.5, and Remark 4.9 therein).

Before defining the orthogonality graphs with these terms and definitions, we introduce assumptions on the stochastic process Y, which are fulfilled, in particular, by most state space models (see Section 5).

Assumption A.

(A.1) Y is a k-dimensional wide-sense stationary and mean-square continuous stochastic process with expectation zero and index set $V = \{1, \dots, k\}$.
(A.2) Y has a spectral density matrix $f_{YY}(\lambda) > 0$ and there exists an $0 < \varepsilon < 1$, such that

$$(1-\varepsilon)I_\alpha - f_{Y_AY_A}(\lambda)^{-1/2}f_{Y_AY_B}(\lambda)f_{Y_BY_B}(\lambda)^{-1}f_{Y_BY_A}(\lambda)f_{Y_AY_A}(\lambda)^{-1/2} \geq 0,$$

for almost all $\lambda \in \mathbb{R}$ and for all disjoint subsets $A, B \subseteq V$, where α is the cardinality of A.
(A.3) Y is purely non-deterministic, i.e., for all $a \in V$ and $t \in \mathbb{R}$,

$$\underset{h\to\infty}{\text{l.i.m.}} P_{\mathcal{L}_{Y}(t)}Y_a(t+h) = 0 \quad \mathbb{P}\text{-a.s.}$$

Remark 3.5.

(a) Assumption (A.1) is a basic requirement, otherwise, e.g., the spectral density in Assumption (A.2) is not well defined.
(b) Assumption (A.2) ensures the linear independence and closedness of sums of linear spaces generated by subprocesses, i.e., for $t \in \mathbb{R}$ and disjoint subsets $A, B \subseteq V$,

$$\mathcal{L}_{Y_A}(t) \cap \mathcal{L}_{Y_B}(t) = \{0\} \quad \text{and} \quad \mathcal{L}_{Y_A}(t) + \mathcal{L}_{Y_B}(t) = \overline{\mathcal{L}_{Y_A}(t) + \mathcal{L}_{Y_B}(t)} \quad \mathbb{P}\text{-a.s.} \tag{3.1}$$

(c) Any process that is wide-sense stationary can be uniquely decomposed into a deterministic and a purely non-deterministic process, which are orthogonal to each other (Gladyshev, 1958, Theorem 1). From the point of view of applications, deterministic processes are not important, so (A.3) is a natural assumption.

With this assumption, we are able to define orthogonality graphs.

Definition 3.6. Suppose Ysatisfies Assumption A.

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

(i) $a \longrightarrow b \notin E_{\mathrm{OG}} \quad \Leftrightarrow \quad Y_a \nrightarrow Y_b \,|\, Y_V$,
(ii) $a \text{ --- } b \notin E_{\mathrm{OG}} \quad \Leftrightarrow \quad Y_a \nsim Y_b \,|\, Y_V$,

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

DOI: 10.1111/jtsa.12787

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

(i) $a \longrightarrow b \notin E^0_{\mathrm{OG}} \quad \Leftrightarrow \quad Y_a \nrightarrow_0 Y_b \,|\, Y_V$,
(ii) $a \,\text{---}\, b \notin E^0_{\mathrm{OG}} \quad \Leftrightarrow \quad Y_a \nsim_0 Y_b \,|\, Y_V$,

for $a, b \in V$ with $a \neq b$, then $G^0_{\mathrm{OG}} = (V, E^0_{\mathrm{OG}})$ is called local orthogonality graph for $Y = Y_V$.

Remark 3.7.

(a) As discussed in Fasen-Hartmann and Schenk (2024a), the assumptions are not necessary for the definition of the graphs, but they ensure that the usual Markov properties for mixed graphs are satisfied. Specifically, the orthogonality graph satisfies the pairwise, local, block-recursive, global AMP, and global Granger-causal Markov properties. In particular, Assumption (A.3) ensures that the global AMP Markov property holds. The local orthogonality graph satisfies the pairwise, local, and block-recursive Markov property; for global Markov properties of the local orthogonality graph additional assumptions are required; see Fasen-Hartmann and Schenk (2024a), Propositions 5.20 and 5.21 respectively.

(b) The local orthogonality graph has fewer edges than the orthogonality graph and, in general, the graphs are not equal. An explicit example of a (local) orthogonality graph, illustrating this property, is given in Figure 1. The advantage of the local orthogonality graph over the orthogonality graph is that it allows for modelling more general graphs, whereas the orthogonality graph satisfies the global AMP and the global causal Markov property, the local orthogonality graph does not satisfy them in general, additional assumptions are necessary. For more details see Fasen-Hartmann and Schenk (2024a) again.

3.2. Controller Canonical State Space Models

The article aims to derive (local) orthogonality graphs for state space models. Therefore, we use the controller canonical form of a state space model and its uniqueness, which results in the unique characterisation of the edges in the (local) orthogonality graph in Section 5. To define the controller canonical form of a state space model, we need some definitions and terminology. Therefore, note that each state space model $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \boldsymbol{C}, L)$ is associated with a rational matrix function

$$H(z) = \boldsymbol{C}\left(zI_{kp} - \boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{B}}, \quad z \in \mathbb{C}\backslash\sigma(\boldsymbol{\mathcal{A}}), \tag{3.2}$$

called the transfer function of the state space model, and the triple $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \boldsymbol{C})$ is an algebraic realisation of the transfer function of dimension kp (characterising the dimension $(kp \times kp)$ of $\boldsymbol{\mathcal{A}}$). The triple $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \boldsymbol{C})$ is said to be minimal if there is no other algebraic realisation of $H(z)$ with dimension less than kp. The transfer function is of importance because, due to the spectral representation theorem (Lax, 2002, Theorem 17.5), we are able to recover the kernel function $\boldsymbol{C}\mathrm{e}^{\boldsymbol{\mathcal{A}}t}\boldsymbol{\mathcal{B}}\mathbf{1}_{\{t\geq 0\}}$ of the output process Y of the state space model via

$$\boldsymbol{C}\mathrm{e}^{\boldsymbol{\mathcal{A}}t}\boldsymbol{\mathcal{B}} = \frac{1}{2\pi i}\int_\Gamma \mathrm{e}^{zt}H(z)\,\mathrm{d}z, \quad t \geq 0, \tag{3.3}$$

where Γ is a closed contour in the complex numbers that winds around each eigenvalue of $\boldsymbol{\mathcal{A}}$ exactly once. The transfer function even uniquely determines the function $\boldsymbol{C}\mathrm{e}^{\boldsymbol{\mathcal{A}}t}\boldsymbol{\mathcal{B}}$, $t \in \mathbb{R}$ (Schlemm and Stelzer, 2012b, Lemma 3.2). Kailath (1980) provides in Lemma 6.3-8 that there exist $(k \times k)$-dimensional matrix polynomials $P(z)$ and $Q(z)$ such that

$$H(z) = Q(z)P(z)^{-1}, \quad z \in \mathbb{C}\backslash\sigma(\boldsymbol{\mathcal{A}}), \tag{3.4}$$

is a coprime right polynomial fraction description of the transfer function, which in turn means that the matrix $[P(z)\;\; Q(z)]$ has full rank for all $z \in \mathbb{C}$. In Lemma 6.3-3 Kailath (1980) even gives a construction for such a decomposition. However, without any additional assumption, the coprime polynomials $P(z)$ and $Q(z)$ that satisfy (3.4) are not unique. Even if we assume additionally that $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ is minimal, resulting in $\deg(\det(P(z)) = kp$ (Rugh, 1996, Theorem 17.5), then we can take any invertible matrix $S \in M_k(\mathbb{R})$ such that the matrix polynomials $P(z)S$ and $Q(z)S$ also satisfy $H(z) = Q(z)SS^{-1}P(z)^{-1}$.

Despite the many different coprime polynomials $P(z)$ and $Q(z)$ that satisfy (3.4), to the best of our knowledge, it remains unclear whether there is a coprime right polynomial fraction description with

$$P(z) = I_k z^p + A_1 z^{p-1} + \cdots + A_p \quad \text{and} \quad Q(z) = C_0 + C_1 z + \cdots + C_q z^q, \tag{3.5}$$

$A_1, A_2, \ldots, A_p, C_0, C_1, \ldots, C_q \in M_k(\mathbb{R})$, and $p, q \in \mathbb{N}_0$, $p > q$, i.e., z^p is the highest power with prefactor I_k. In representation (3.5), the assumption $\deg(\det(P(z)) = kp$ obviously holds. Note, the construction method of Kailath (1980) often gives a polynomial $P(z)$ with higher order than p, but the prefactor of the highest power has a zero determinant. Brockwell and Schlemm (2013), Theorem 3.2, and Schlemm and Stelzer (2012a), Corollary 3.4, implicitly assume such a right polynomial fraction description (3.5) without discussing its existence. Since the existence of such a coprime right polynomial fraction description is essential for the forthcoming results, we always assume it additionally. In Examples 3.8 and 3.14, we present likewise examples where this assumption is fulfilled.

Example 3.8. Consider a state space model $(\mathcal{A}, \mathcal{B}, \mathcal{C}, L)$ with $k = 3$, $p = 2$, $q = 1$, and $\Sigma_L = I_3$, where we set

$$\mathcal{A} = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & -1 & 1 \\ 0 & -1 & 1 \end{pmatrix} \quad \text{and} \quad \mathcal{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{3}{2} & 0 & 1 \end{pmatrix}.$$

Note that this is the state space model that generates the (local) orthogonality graph in Figure 1 in the introduction to this article, and we will look at this example in more detail in the course of this article.

In this example, a straightforward calculation shows that there exists the right polynomial fraction description of the transfer function

$$H(z) = \mathcal{C}\left(zI_6 - \mathcal{A}\right)^{-1}\mathcal{B} = \begin{pmatrix} z+1 & -z-1 & 0 \\ z+1 & z+1 & 0 \\ z+1 & z+1 & z+1 \end{pmatrix} \begin{pmatrix} z^2+z+1 & 0 & 0 \\ 0 & z^2+z+1 & 0 \\ 0 & 1 & z^2+z+1 \end{pmatrix}^{-1} =: Q(z)P(z)^{-1},$$

for $z \in \mathbb{C} \setminus \sigma(\mathcal{A})$, where $\sigma(\mathcal{A}) = \{-\frac{1}{2} + \frac{\sqrt{3}}{2}i, -\frac{1}{2} - \frac{\sqrt{3}}{2}i\}$. Furthermore, this decomposition is coprime, since $[P(z) \;\; Q(z)]$ has full rank 3 for all $z \in \mathbb{C}$. Thus, in this example, there exists a coprime right polynomial fraction description (3.4) with polynomials as in (3.5).

For the purpose of this article, not only the existence of a coprime right polynomial fraction description of the transfer function with polynomials $P(z)$ and $Q(z)$ as in (3.5) is important but also its uniqueness. In the next proposition, we derive that this requirement is immediately satisfied under the existence assumption of $P(z)$ and $Q(z)$.

DOI: 10.1111/jtsa.12787

Proposition 3.9. Let $(\mathcal{A}, \mathcal{B}, \boldsymbol{C}, L)$ be a state space model with transfer function $H(z)$. Suppose there exists a coprime right polynomial fraction description of $H(z)$ with polynomials $P(z)$ and $Q(z)$ as in (3.5) such that

$$H(z) = \boldsymbol{C}\left(zI_{kp} - \mathcal{A}\right)^{-1}\mathcal{B} = Q(z)P(z)^{-1}, \quad z \in \mathbb{C} \setminus \sigma(\mathcal{A}).$$

Then $P(z)$ and $Q(z)$ are unique. Moreover, defining

$$\boldsymbol{A} = \begin{pmatrix} 0_k & I_k & 0_k & \cdots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \cdots & \cdots & 0_k & I_k \\ -A_p & -A_{p-1} & \cdots & \cdots & -A_1 \end{pmatrix} \in M_{kp}(\mathbb{R}), \quad \boldsymbol{B} = \begin{pmatrix} 0_k \\ \vdots \\ 0_k \\ I_k \end{pmatrix} \in M_{kp\times k}(\mathbb{R}),$$
$$\boldsymbol{C} = \left(C_0,\ C_1,\ \cdots,\ C_q,\ 0_k,\ \cdots,\ 0_k\right) \in M_{k\times kp}(\mathbb{R}), \tag{3.6}$$

then $\sigma(\mathcal{A}) = \sigma(\boldsymbol{A})$ and

$$H(z) = \boldsymbol{C}\left(zI_{kp} - \boldsymbol{A}\right)^{-1}\boldsymbol{B}, \quad z \in \mathbb{C} \setminus \sigma(\boldsymbol{A}).$$

Finally, Y is a solution of the state space model $(\mathcal{A}, \mathcal{B}, \boldsymbol{C}, L)$, if and only if, it is a solution of the state space model $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$. The state space model $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$ is called controller canonical form.

In particular, of course, this implies that there exists no other minimal state space representation with matrices of the structure as in (3.6); this representation is unique. Since the solution of these two state space models is equal, we will henceforth assume, without loss of generality, that the state space model is given in the unique controller canonical form $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$ as in (3.6).

Example 3.10. In Example 3.8, the state space model $(\mathcal{A}, \mathcal{B}, \boldsymbol{C}, L)$ has the unique controller canonical form $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$, where

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 0 & 0 & -1 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

A particular example of a state space model is the MCARMA model whose output process Y is an MCARMA process. The definition of an MCARMA process Y is motivated by the idea that Y solves the stochastic differential equation

$$\mathcal{P}(D)Y(t) = \mathcal{Q}(D)DL(t),$$

where D is the differential operator with respect to t and

$$\mathcal{P}(z) = I_k z^{\mathfrak{p}} + \mathcal{P}_1 z^{\mathfrak{p}-1} + \cdots + \mathcal{P}_{\mathfrak{p}} \quad \text{and} \quad \mathcal{Q}(z) = \mathcal{Q}_0 z^{\mathfrak{q}} + \mathcal{Q}_1 z^{\mathfrak{q}-1} + \cdots + \mathcal{Q}_{\mathfrak{q}}, \tag{3.7}$$

are the AR (autoregressive) and the MA (moving average) polynomial with $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_{\mathfrak{p}}, \mathcal{Q}_0, \mathcal{Q}_1, \dots, \mathcal{Q}_{\mathfrak{q}} \in M_k(\mathbb{R})$. However, a Lévy process is not differentiable, so this is not a formal definition. The formal definition is given by a state space model (Marquardt and Stelzer, 2007) as follows.

Definition 3.11. Define

$$\mathcal{A} = \begin{pmatrix} 0_k & I_k & 0_k & \cdots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \cdots & \cdots & 0_k & I_k \\ -\mathcal{P}_{\mathfrak{p}} & -\mathcal{P}_{\mathfrak{p}-1} & \cdots & \cdots & -\mathcal{P}_1 \end{pmatrix} \in M_{k\mathfrak{p}}(\mathbb{R}), \quad \mathcal{B} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{\mathfrak{p}} \end{pmatrix} \in M_{k\mathfrak{p}\times k}(\mathbb{R}),$$
$$\mathcal{C} = (I_k, 0_k, \cdots, 0_k) \in M_{k\times k\mathfrak{p}}(\mathbb{R}),$$

where $\beta_1 = \cdots = \beta_{\mathfrak{p}-\mathfrak{q}-1} = 0_k$ and $\beta_{\mathfrak{p}-j} = -\sum_{i=1}^{\mathfrak{p}-j-1} \mathcal{P}_i \beta_{\mathfrak{p}-j-i} + \mathcal{Q}_{\mathfrak{q}-j}$ for $j = \mathfrak{q}, \ldots, 0$. Then $(\mathcal{A}, \mathcal{B}, \mathcal{C}, L)$ is called a multivariate continuous-time moving average model of order $(\mathfrak{p}, \mathfrak{q})$, shortly an MCARMA$(\mathfrak{p}, \mathfrak{q})$ model.

Remark 3.12. A comparison of the triplets $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ and $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ shows that the MCAR(p) = MCARMA$(p, 0)$ model is already in controller canonical form. For MCARMA(p, q) models, Schlemm and Stelzer (2012a) show the equivalence between state space models and MCARMA models in Corollary 3.4, and Brockwell and Schlemm (2013), Theorem 3.2, show the equivalence between MCARMA models and controller canonical state space models. However, as mentioned above, both implicitly assume the existence of a coprime left (right) polynomial fractional description (3.4) with polynomials $P(z)$ and $Q(z)$ as in (3.5), which is in our opinion not obvious. However, for univariate state space processes with $k = 1$, the existence of a coprime right polynomial fractional description is apparent (see proof of Proposition 4.1), so that any univariate state space model is a CARMA model and vice versa; additionally, any univariate state space model has a representation in controller canonical form.

A peculiarity of MCARMA models is that the AR polynomial $\mathcal{P}(z)$ and the MA polynomial $\mathcal{Q}(z)$ provide a left polynomial fraction description of the transfer function, i.e., $H(z) = \mathcal{P}(z)^{-1}\mathcal{Q}(z)$ (Marquardt and Stelzer, 2007 or Brockwell and Schlemm, 2013, Lemma 3.1). If the MCARMA model is minimal, this left polynomial fractional description is even coprime (Kailath, 1980, Theorem 6.5-1). The connection to the coprime right polynomial fraction description (3.4) with $P(z)$ and $Q(z)$ as in (3.5) is given in the next lemma.

Lemma 3.13. Let an MCARMA$(\mathfrak{p}, \mathfrak{q})$ model be given with state space representation $(\mathcal{A}, \mathcal{B}, \mathcal{C}, L)$ as in Definition 3.11 and polynomials $\mathcal{P}(z)$ and $\mathcal{Q}(z)$ as in (3.7). Suppose $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ is minimal and there exists a coprime right polynomial fraction description (3.4) of the transfer function with polynomials $P(z)$ and $Q(z)$ as in (3.5). Then $P(z)$ and $Q(z)$ are unique, $\mathfrak{p} = p$, $\mathfrak{q} = q$, $\mathcal{Q}_0 = C_q$, $\mathcal{N}(\mathcal{P}) = \mathcal{N}(P)$ and $\mathcal{N}(\mathcal{Q}) = \mathcal{N}(Q)$.

Example 3.14.

(a) For MCAR(p) models, $\mathcal{Q}(z) = I_k$ holds. Thus, $P(z) = \mathcal{P}(z)$ and $Q(z) = I_k$ always provide a coprime right polynomial fractional description of $H(z)$.

(b) Consider an MCARMA(2, 1) model with coprime AR polynomial and MA polynomial given by

$$\mathcal{P}(z) = \begin{pmatrix} (z+2)^2 & 0 \\ 0 & (z+2)^2 \end{pmatrix} \quad \text{and} \quad \mathcal{Q}(z) = \begin{pmatrix} z+1 & 0 \\ 0 & z+1 \end{pmatrix}.$$

Since $\mathcal{P}(z)$ and $\mathcal{Q}(z)$ are diagonal matrix polynomials and are right coprime, the unique coprime right polynomial fraction description of the transfer function $H(z)$ is given through $P(z) = \mathcal{P}(z)$ and $Q(z) = \mathcal{Q}(z)$.

(c) Consider an MCARMA(3, 1) model with coprime AR polynomial and MA polynomial given by

$$\mathcal{P}(z) = \begin{pmatrix} \frac{1}{4}(2z+3)(2z^2+7z+7) & -\frac{1}{4}(z+2)(3z+5) \\ -(z+1)^2 & (z+1)^2(z+2) \end{pmatrix} \quad \text{and} \quad \mathcal{Q}(z) = -\begin{pmatrix} z+1 & \frac{1}{4} \\ 0 & z+3 \end{pmatrix}.$$

DOI: 10.1111/jtsa.12787

Then the coprime right polynomial fractional description of $H(z)$ is given by

$$P(z) = \begin{pmatrix} (z+2)^3 & 0 \\ 0 & (z+1)^3 \end{pmatrix} \quad \text{and} \quad Q(z) = -\begin{pmatrix} z+2 & 1 \\ 1 & z+2 \end{pmatrix}.$$

(d) The controller canonical state space model $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$ in Example 3.8 with coprime right polynomial fraction description $P(z)$ and $Q(z)$ is as well an MCARMA(2, 1) model with coprime AR and MA polynomial given by

$$\mathcal{P}(z) = \begin{pmatrix} z^2+z+1 & 0 & 0 \\ 0 & z^2+z+1 & 0 \\ -\frac{1}{2} & \frac{1}{2} & z^2+z+1 \end{pmatrix} \quad \text{and} \quad \mathcal{Q}(z) = \begin{pmatrix} z+1 & -z-1 & 0 \\ z+1 & z+1 & 0 \\ z+1 & z+1 & z+1 \end{pmatrix}.$$

In particular, these examples further emphasise the existence of coprime right polynomial fraction descriptions (3.4) with polynomials as in (3.5). Since the aim of this article is not the investigation of polynomial fraction descriptions, but the application of (local) orthogonality graphs to state space models, we do not investigate this further and move on to the topic of the invertibility of a state model.

3.3. Invertible Controller Canonical State Space Models

Suppose Y is a solution to a state space model that has a controller canonical representation $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$, and we assume that the driving Lévy process satisfies the following common assumption.

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

Then the second moments of $X(t)$ and thus of $Y(t)$ also exist (Brockwell and Schlemm, 2013, Lemma A.4), which is a basic requirement for the forthcoming considerations on the existence of (local) orthogonality graphs.

Due to the state equation $Y(t) = \boldsymbol{C}X(t)$, we obtain the output process $Y(t)$ directly from the input process $X(t)$. However, the recovery of $X(t)$ from the output process $(Y(s))_{s\leq t}$, is not as obvious, because $\boldsymbol{C}$ is not invertible (cf. Example 3.10). Only for $q = 0$, corresponding to an MCAR(p) model, the simple structure of $\boldsymbol{C}$ allows to use the relation

$$\begin{aligned} D^{(j-1)}Y(t) &= X^{(j)}(t), \qquad j = 1, \dots, p, \qquad \text{where} \\ X^{(j)}(t) &= \left(X_{(j-1)k+1}(t), \ \cdots \ , \ X_{jk}(t)\right)^\top, \end{aligned} \tag{3.8}$$

is the jth k-block of $X(t)$ and $D^{(1)}Y(t), \dots, D^{(p-1)}Y(t)$ denote the mean-square derivatives of $Y(t)$. Therefore, in this case it is possible to recover $X(t)$ from $Y(t)$ and its derivatives $D^{(1)}Y(t), \dots, D^{(p-1)}Y(t)$ via (3.8). However, for controller canonical state space models with $q > 0$, we cannot apply this approach. Indeed, the structure of $\boldsymbol{A}$ still yields

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

and together with $Y(t) = \boldsymbol{C}X(t) = C_0X^{(1)}(t) + \cdots + C_qX^{(q+1)}(t)$, we obtain that

$$D^{(j-1)}Y(t) = \sum_{i=0}^{q} C_i X^{(j+i)}(t), \qquad j = 1, \dots, p-q. \tag{3.9}$$

Consequently, the p k-blocks of X cannot generally be recovered from these $(p-q)$ equations.

Remark 3.15.

(a) For the reader's convenience, we define

$$\mathsf{C} := (0_k, \dots, 0_k, C_0, \dots, C_q) \in M_{k\times kp}(\mathbb{R}). \tag{3.10}$$

From (3.9) we then receive the shorthands

$$Y(t) = \boldsymbol{C}X(t) \quad \text{and} \quad D^{(p-q-1)}Y(t) = \mathsf{C}X(t). \tag{3.11}$$

In particular, this implies that Y and its components Y_v, $v \in V$, are $(p-q-1)$ times mean-square differentiable.

(b) A conclusion from (a) and Fasen-Hartmann and Schenk (2024a), Remark 2.6, is then that for $v \in V$ and $t \geq 0$,

$$D^{(1)}Y_v(t), \dots, D^{(p-q-1)}Y_v(t) \in \mathcal{L}_{Y_v}(t).$$

For controller canonical state space models as in (3.6) with $q > 0$, we overcome the challenge of recovering the state process from the output process under some mild assumptions. Of course, due to $q > 0$, the class of MCAR(p) models are excluded in the following considerations. However, this is not an essential limitation, because the (local) orthogonality graphs and the orthogonal projections respectively for this case are already known (Fasen-Hartmann and Schenk, 2024a). We first define causal invertible controller canonical state space models, which are a special subclass of controller canonical state space models.

Definition 3.16. Let $(\mathcal{A}, \mathcal{B}, \boldsymbol{C}, L)$ be a state space model with controller canonical form $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$ as in (3.6) and right coprime polynomials $P(z)$ and $Q(z)$ as in (3.5) with $p > q > 0$. Suppose that

$$\operatorname{rank}(C_q) = k, \mathcal{N}(Q) \subseteq (-\infty, 0) + i\mathbb{R} \quad \text{and} \quad \mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}. \tag{3.12}$$

Then $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$ is called a causal invertible controller canonical state space (ICCSS) model of order (p, q) and the stationary solution $Y = (Y(t))_{t\in\mathbb{R}}$ of the ICCSS(p, q) model is called ICCSS(p, q) process.

Remark 3.17.

(a) Since $\mathcal{N}(P) = \sigma(\boldsymbol{A}) \subseteq (-\infty, 0) + i\mathbb{R}$ (Marquardt and Stelzer, 2007, Corollary 3.8), there exists a unique stationary solution $X(t)$ of the observation equation (Sato and Yamazato, 1984, Theorem 4.1) which has the representation

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

Hence, there exists as well a stationary version of the output process Y, which has the moving average representation

$$Y(t) = \int_{-\infty}^{\infty} g(t-u)\mathrm{d}L(u) \quad \text{with} \quad g(t) = \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{B}\mathbf{1}_{\{t\geq 0\}}, \quad t \in \mathbb{R}.$$

Throughout this article, we are working with these stationary versions of X and Y.

(b) The assumptions on $Q(z)$ are necessary to recover X from Y and to motivate the name ICCSS model, as we see in the remainder of this section.

(c) In the running Example 3.8, we have $p = 2 > q = 1 > 0$, $\operatorname{rank}(C_1) = 3$, as well as $\mathcal{N}(Q) = \{-1\} \subseteq (-\infty, 0) + i\mathbb{R}$, and $\mathcal{N}(P) = \sigma(\boldsymbol{A}) = \{-\frac{1}{2} + \frac{\sqrt{3}}{2}i, -\frac{1}{2} - \frac{\sqrt{3}}{2}i\} \subseteq (-\infty, 0) + i\mathbb{R}$. Thus, all of the assumptions in (3.12) are satisfied. $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$ is a causal invertible controller canonical state space model of order (2, 1), and the stationary solution Y is an ICCSS(2, 1) process. Furthermore, the assumptions in (3.12) are also satisfied in Example 3.14(b,c).

DOI: 10.1111/jtsa.12787

Under Assumption (3.12), Brockwell and Schlemm (2013), Lemma 4.1, derive a stochastic differential equation for the first (kq) components of X. This follows simply from combining the first q k-blocks of the state transition equation $\mathrm{d}X(t) = \boldsymbol{A}X(t)\mathrm{d}t + \boldsymbol{B}\mathrm{d}L(t)$ with the observation equation $Y(t) = \boldsymbol{C}X(t)$ having the special structure of $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ in mind.

Lemma 3.18. Let Y be an ICCSS(p,q) process with $p > q > 0$. Denote the (kq)-dimensional upper truncated state vector $X^q = (X^q(t))_{t\in\mathbb{R}}$ of X by

$$X^q(t) = \begin{pmatrix} X^{(1)}(t) \\ \vdots \\ X^{(q)}(t) \end{pmatrix}, \quad t \in \mathbb{R},$$

where $X^{(1)}(t), \dots, X^{(q)}(t)$ are the k-dimensional random vectors as defined in (3.8). Then X^q satisfies

$$\mathrm{d}X^q(t) = \boldsymbol{\Lambda} X^q(t)\mathrm{d}t + \boldsymbol{\Theta} Y(t)\mathrm{d}t, \tag{3.13}$$

where $\sigma(\boldsymbol{\Lambda}) \subseteq (-\infty, 0) + i\mathbb{R}$,

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0_k & I_k & 0_k & \cdots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \cdots & \cdots & 0_k & I_k \\ -C_q^{-1}C_0 & -C_q^{-1}C_1 & \cdots & \cdots & -C_q^{-1}C_{q-1} \end{pmatrix} \in M_{kq}(\mathbb{R}) \quad \text{and} \quad \boldsymbol{\Theta} = \begin{pmatrix} 0_k \\ \vdots \\ 0_k \\ C_q^{-1} \end{pmatrix} \in M_{kq\times k}(\mathbb{R}).$$

Remark 3.19.

(a) Assumption (3.12) corresponds to the minimum-phase assumption in classical time series analysis (Hannan and Deistler, 2012) and implies Assumption A2 in Brockwell and Schlemm (2013), who even allow for rectangular matrices $C_0, \dots, C_q$. To see this, note that Assumption (3.12) yields

$$\mathcal{N}(C_q^{-1}Q) = \{z \in \mathbb{C} : \det(C_q^{-1}Q(z)) = 0\} = \{z \in \mathbb{C} : \det(Q(z)) = 0\} = \mathcal{N}(Q) \subseteq (-\infty, 0) + i\mathbb{R},$$

which is one of their assumptions. Furthermore, $\sigma(\boldsymbol{\Lambda}) = \mathcal{N}(C_q^{-1}Q)$ (Marquardt and Stelzer, 2007, Lemma 3.8). Thus, $\boldsymbol{\Lambda}$ has full rank and, due to the structure of $\boldsymbol{\Lambda}$, we obtain that $C_q^{-1}C_0$ has full rank. It follows that C_0 and $(C_q)^\top C_0$ have full rank as well, which is the second assumption in Brockwell and Schlemm (2013).

(b) If the AR polynomial $\mathcal{P}(z)$ and the MA polynomial $\mathcal{Q}(z)$of an MCARMA model are left coprime, Assumption (3.12) can equally be made for $\mathcal{P}(z)$ and $\mathcal{Q}(z)$, respectively. Indeed, $\mathcal{N}(\mathcal{Q}) = \mathcal{N}(Q)$ and $\mathcal{N}(\mathcal{P}) = \mathcal{N}(P)$ by Lemma 3.13. Further, straightforward calculations of $\mathcal{Q}(z)P(z) = \mathcal{P}(z)Q(z)$ give $\mathcal{Q}_q A_p = \mathcal{P}_p C_q$. For $\mathcal{N}(\mathcal{P}) = \mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}$ we have $0 \notin \mathcal{N}(\mathcal{P})$ and thus, $\det(\mathcal{P}_p) = \det(\mathcal{P}(0)) \neq 0$. Similarly $\det(A_p) \neq 0$ follows, so $\mathcal{P}_p$ and A_p are invertible. Hence, if $\mathcal{N}(\mathcal{P}) = \mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}$, then $\mathcal{Q}_q$ has full rank, if and only if, C_q has full rank.

The differential equation (3.13) has the solution

$$X^q(t) = \mathrm{e}^{\boldsymbol{\Lambda}(t-s)}X^q(s) + \int_s^t \mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} Y(u)\mathrm{d}u, \quad s < t, \tag{3.14}$$

(Brockwell and Schlemm, 2013, (4.3)). Therefore, we can compute $X^q(t)$ based on the knowledge of the initial value $X^q(s)$ and $(Y(u))_{s\le u\le t}$. In Propositions 4.1 and 4.5 we even show the integral representation

$$X^q(t) = \int_{-\infty}^{t} \mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} Y(u)\mathrm{d}u.$$

$\mathbb{P}$-a.s. and in the mean square, respectively. Hence, $X^q(t)$ is even uniquely determined by the entire past $(Y(s))_{s\le t}$, implying that the truncated state vector X^q can be recovered from Y. The remaining k-blocks $X^{(q+j)}, j = 1, \dots, p-q$, are obtained from X^q and Y by differentiation as in Brockwell and Schlemm (2013), Lemma 4.2:

Lemma 3.20. Let Y be an ICCSS(p, q) process with $p > q > 0$. Then

$$X^{(q+j)}(t) = \mathbf{E}^\top \left[\boldsymbol{\Lambda}^j X^q(t) + \sum_{m=0}^{j-1} \boldsymbol{\Lambda}^{j-1-m}\boldsymbol{\Theta} D^{(m)} Y(t)\right], \quad j = 1, \dots, p-q, t \in \mathbb{R}.$$

Note that there is a duplication of notation in Brockwell and Schlemm (2013), which can be seen by recalculating the induction start. We therefore give the corrected result in Lemma 3.20.

Example 3.21. Coming back to Example 3.8, the output process Y has the representation as the linear combinations of X via

$$Y(t) = \boldsymbol{C}X(t) = \begin{pmatrix} X_1(t) - X_2(t) + X_4(t) - X_5(t) \\ X_1(t) + X_2(t) + X_4(t) + X_5(t) \\ X_1(t) + X_2(t) + X_3(t) + X_4(t) + X_5(t) + X_6(t) \end{pmatrix}.$$

From this representation, it is not immediately obvious how X can be recovered from Y. However, we can define

$$\boldsymbol{\Lambda} = -C_1^{-1}C_0 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Theta} = C_1^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -2 & 2 \end{pmatrix},$$

with $\sigma(\boldsymbol{\Lambda}) = \mathcal{N}(Q) = \{-1\} \subseteq (-\infty, 0) + i\mathbb{R}$. Due to $\mathbf{E}^\top = I_3$ and the simple form of $\boldsymbol{\Lambda}$ and its matrix exponential respectively, we can recover the input process X from the output process Y by

$$X^{(1)}(t) = X^q(t) = \int_{-\infty}^{t} \mathrm{e}^{-(t-u)}\frac{1}{2}\begin{pmatrix} Y_1(u) + Y_2(u) \\ -Y_1(u) + Y_2(u) \\ -2Y_2(u) + 2Y_3(u) \end{pmatrix}\mathrm{d}u,$$

$$X^{(2)}(t) = -X^{(1)}(t) + \frac{1}{2}\begin{pmatrix} Y_1(t) + Y_2(t) \\ -Y_1(t) + Y_2(t) \\ -2Y_2(t) + 2Y_3(t) \end{pmatrix}.$$

In summary, in the example as well as in the general setting, we are able to compute not only the truncated state vector $X^q(t)$ but also the full state vector $X(t)$ based on the knowledge of $(Y(s))_{s\le t}$. This justifies calling the ICCSS process Y invertible if Assumption (3.12) holds.

4. ORTHOGONAL PROJECTIONS OF ICCSS PROCESSES

Here, we derive the orthogonal projections of ICCSS processes and their derivatives which we require to characterise (local) Granger causality and (local) contemporaneous correlation for ICCSS processes. First, we give alternative representations of $Y_a(t+h)$ as well as $D^{(p-q-1)}Y_a(t+h)$, $a \in V = \{1, \dots, k\}$, suitable for the calculation of orthogonal projections in Section 4.1. Note that we consider the process $D^{(p-q-1)}Y_a(t+h)$ since, by Remark 4.7 below, it is the highest existing derivative of the ICCSS process which we require for the definition of local Granger causality and local contemporaneous correlation, respectively. In Section 4.2, we then present the corresponding orthogonal projections of both random variables on $\mathcal{L}_{Y_S}(t)$ for $S \subseteq V$. Furthermore, we discuss the limit of the projections of difference quotients.

4.1. Representations of ICCSS Processes and Their Derivatives

The aim of this subsection is to develop a $\mathbb{P}$-a.s. representation of $Y_a(t+h)$ and $D^{(p-q-1)}Y_a(t+h)$, $a \in V$. Therefore, we first introduce the $\mathbb{P}$-a.s. integral representation of the upper q-block truncation X^q, which is a multivariate generalisation of Brockwell and Lindner (2015), Theorem 2.2.

Proposition 4.1. Let Y be an ICCSS(p,q) process with $p > q > 0$. Then, for all $t \in \mathbb{R}$, we have

$$X^q(t) = \int_{-\infty}^{t} \mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} Y(u)\mathrm{d}u \quad \mathbb{P}\text{-a.s.}$$

Due to the well-definedness of this integral, it is obvious that the following representations of Y and its derivatives are well-defined as well.

Theorem 4.2. Let Y be an ICCSS(p,q) process with $p > q > 0$. Then, for $h \geq 0$, $t \in \mathbb{R}$, and $a \in V$, it holds that

$$Y_a(t+h) = \int_{-\infty}^{t} \mathrm{e}_a^\top \boldsymbol{M}(h) e^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} Y(u)\mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta} D^{(m)}Y(t) + \mathrm{e}_a^\top \varepsilon(t,h)\mathbb{P}\text{-a.s.and}$$

$$D^{(p-q-1)}Y_a(t+h) = \int_{-\infty}^{t} \mathrm{e}_a^\top \mathsf{M}(h) e^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} Y(u)\mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \mathsf{M}_m(h)\boldsymbol{\Theta} D^{(m)}Y(t) + \mathrm{e}_a^\top \epsilon(t,h)\mathbb{P}\text{-a.s.}$$

Here, we abbreviate

$$\boldsymbol{M}(h) = \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}h}\left(\boldsymbol{E} + \sum_{j=1}^{p-q}\mathbf{E}_{q+j}\mathsf{E}^\top\boldsymbol{\Lambda}^j\right), \qquad \mathsf{M}(h) = \mathsf{C}\mathrm{e}^{\boldsymbol{A}h}\left(\boldsymbol{E} + \sum_{j=1}^{p-q}\mathbf{E}_{q+j}\mathsf{E}^\top\boldsymbol{\Lambda}^j\right),$$

$$\boldsymbol{M}_m(h) = \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}h}\sum_{j=m+1}^{p-q}\mathbf{E}_{q+j}\mathsf{E}^\top\boldsymbol{\Lambda}^{j-1-m}, \qquad \mathsf{M}_m(h) = \mathsf{C}e^{\boldsymbol{A}h}\sum_{j=m+1}^{p-q}\mathbf{E}_{q+j}\mathsf{E}^\top\boldsymbol{\Lambda}^{j-1-m},$$

$$\varepsilon(t,h) = \boldsymbol{C}\int_{t}^{t+h} e^{\boldsymbol{A}(t+h-u)}\boldsymbol{B}dL(u), \qquad \epsilon(t,h) = \mathsf{C}\int_{t}^{t+h} e^{\boldsymbol{A}(t+h-u)}\boldsymbol{B}dL(u).$$

where C is defined in (3.10), $\boldsymbol{E}$ and E are defined in (2.2), and $\mathbf{E}_j$ is defined in (2.1), $j = 1, \dots, p$. Finally, $\varepsilon(t,0) = \epsilon(t,0) = 0_k \in \mathbb{R}^k$.

Remark 4.3. For an MCAR(p) process, Fasen-Hartmann and Schenk (2024a) state in Lemma 6.8 that

$$Y_a(t+h) = \mathrm{e}_a^\top \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}h}\sum_{m=0}^{p-1}\mathbf{E}_{m+1}D^{(m)}Y(t) + \mathrm{e}_a^\top \boldsymbol{C}\int_{t}^{t+h}\mathrm{e}^{\boldsymbol{A}(t+h-u)}\boldsymbol{B}\mathrm{d}L(u) \quad \mathbb{P}\text{-a.s.} \tag{4.1}$$

Thus, if we want to compare our Theorem 4.2 to the results for MCAR(p) processes, we have to interpret

$$\boldsymbol{M}_m(h)\boldsymbol{\Theta} \mathrel{\hat=} \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}h}\mathbf{E}_{m+1}, \quad m = 0, \dots, p-1, \quad \text{and} \quad \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} \mathrel{\hat=} 0_k \text{ for } u < t.$$

Then (4.1) can be seen as a special case of Theorem 4.2. Let us briefly heuristically justify that this interpretation is reasonable. First of all, in $\boldsymbol{M}_m(h)\boldsymbol{\Theta}$ the summand $j = m+1$ is mainly relevant. For this summand we have with $\boldsymbol{\Lambda}^0 = I_{kq}$ that

$$\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}h}\mathbf{E}_{q+m+1}\mathbf{E}^\top\boldsymbol{\Theta} = \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}h}\mathbf{E}_{q+m+1}C_q^{-1}. \tag{4.2}$$

If $q = 0$ is inserted into $\boldsymbol{M}_m(h)\boldsymbol{\Theta}$, all summands disappear due to the zero dimensionality of $\boldsymbol{\Lambda}^{j-1-m}$, $j = m+2, \dots, p-q$, except for (4.2). With $C_q = I_k$ it remains as claimed $\boldsymbol{M}_m(h)\boldsymbol{\Theta} \mathrel{\hat=} \boldsymbol{C}e^{\boldsymbol{A}h}\mathbf{E}_{m+1}$ for $m = 0, \dots, p-1$. For the second matrix function $\boldsymbol{M}(h)e^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta}$, $u < t$, we use similar arguments to show that it can be interpreted as a zero matrix. Although we get a non-zero matrix for $t = u$, this event is a Lebesgue null-set.

Example 4.4. In Example 3.8, we have $m = p - q - 1 = 0$, so

$$Y(t+h) = D^{(0)}Y(t+h) = \int_{-\infty}^{t} \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta}Y(u)\mathrm{d}u + \boldsymbol{M}_0(h)\boldsymbol{\Theta}Y(t) + \varepsilon(t,h) \quad \mathbb{P}\text{-a.s.}$$

If we abbreviate $c(h) := 3\cos(\sqrt{3}h/2)$ and $s(h) := \sqrt{3}\sin(\sqrt{3}h/2)$, the three addends can be specified as follows.

$$\begin{aligned}
&\int_{-\infty}^{t} \boldsymbol{M}(h)e^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta}Y(u)\mathrm{d}u \\
&= \frac{\mathrm{e}^{-\frac{h}{2}}}{3}\int_{-\infty}^{t}\mathrm{e}^{-(t-u)}\begin{pmatrix} -2s(h)Y_1(u) \\ -2s(h)Y_2(u) \\ -2s(h)Y_3(u) + \frac{1}{3}[hc(h)+s(h)]\big(Y_1(u)-Y_2(u)\big)\end{pmatrix}\mathrm{d}u, \\
&\boldsymbol{M}_0(h)\boldsymbol{\Theta}Y(t) \\
&= \frac{\mathrm{e}^{-\frac{h}{2}}}{3}\begin{pmatrix} [c(h)+s(h)]Y_1(t) \\ [c(h)+s(h)]Y_2(t) \\ [c(h)+s(h)]Y_3(t) + \frac{1}{6}[-hc(h)+(3h+2)s(h)]\big(Y_1(t)-Y_2(t)\big)\end{pmatrix}, \\
&\varepsilon(t,h) \\
&= \frac{\mathrm{e}^{-\frac{h}{2}}}{3}\int_{t}^{t+h}\mathrm{e}^{-\frac{(t-u)}{2}}\begin{pmatrix} [c(t+h-u)+s(t+h-u)]\big(dL_1(u)-dL_2(u)\big) \\ [c(t+h-u)+s(t+h-u)]\big(dL_1(u)+dL_2(u)\big) \\ [c(t+h-u)+s(t+h-u)]\big(dL_1(u)+dL_3(u)\big)\end{pmatrix} \\
&\quad + \begin{pmatrix} 0 \\ 0 \\ \frac{\mathrm{e}^{-\frac{h}{2}}}{3}\int_{t}^{t+h}\mathrm{e}^{-\frac{(t-u)}{2}}\left[\left(1+\frac{t+h-u}{3}\right)c(t+h-u)+\left(\frac{1}{3}-(t+h-u)\right)s(t+h-u)\right]\mathrm{d}L_2(u)\end{pmatrix}.
\end{aligned}$$

4.2. Orthogonal Projections of ICCSS Processes and Their Derivatives

The representations of $Y_a(t+h)$ and $D^{(p-q-1)}Y_a(t+h)$ in Theorem 4.2 suggest that for the orthogonal projection of the ath component one time step into the future, on the one hand, the past $(Y_V(s))_{s\le t}$ of all components and on the other hand, the future of the Lévy process $(L(s) - L(t))_{t \le s \le t+h}$ is relevant. However, for a formal proof, we require that all integrals are defined in L^2. Therefore, we show that the integral representation of X^q in Proposition 4.1 holds in L^2. The proof is based on the ideas of the proof of Theorem 2.8 in Brockwell and Lindner (2015) in the univariate setting.

Proposition 4.5. Let Y be an ICCSS(p, q) process with $p > q > 0$. Then, for $a, v \in V$and $t \in \mathbb{R}$, the integral

$$\int_{-\infty}^{t} \mathrm{e}_a^\top \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u \in \mathcal{L}_{Y_v}(t),$$

exists as L^2-limit. In particular, $X^q(t) = \int_{-\infty}^{t} \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) \mathrm{d}u$exists as L^2-limit.

Before finally moving on to the orthogonal projections, we introduce one last alternative representation, this time for the difference quotient $(D^{(p-q-1)} Y_a(t+h) - D^{(p-q-1)} Y_a(t))/h$. With this representation we can argue that $D^{(p-q-1)} Y_a(t)$ is indeed the maximum derivative of $Y_a(t)$ which we need for local Granger causality and local contemporaneous correlation.

Lemma 4.6. Let Y be an ICCSS(p, q) process with $p > q > 0$. Then for $h \ge 0$, $t \in \mathbb{R}$, and $a \in V$ the representation

$$\begin{aligned}\frac{D^{(p-q-1)} Y_a(t+h) - D^{(p-q-1)} Y_a(t)}{h} =& \int_{-\infty}^{t} \mathrm{e}_a^\top \mathrm{M}'(0) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) \mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \mathrm{M}_m'(0) \mathbf{\Theta} D^{(m)} Y(t) \\ &+ \mathrm{e}_a^\top O(h) R_1 + \mathrm{e}_a^\top O(h) R_2 + \mathrm{e}_a^\top \frac{\epsilon(t,h)}{h},\end{aligned}$$

holds, where R_1, R_2 are random vectors in $\mathcal{L}_Y(t) \subseteq L^2$. $\mathbf{M}'(0)$ and $\mathbf{M}_m'(0)$ denote the first derivatives of $\mathbf{M}(\cdot)$ and $\mathbf{M}_m(\cdot)$ in zero. The random variable $\mathrm{e}_a^\top \epsilon(t,h)/h$is independent of the former summands and

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}\left[(\mathrm{e}_a^\top \epsilon(t,h))^2\right] = \mathrm{e}_a^\top \mathrm{C} \boldsymbol{B} \Sigma_L \boldsymbol{B}^\top \mathrm{C}^\top e_a \neq 0 \text{ but } \lim_{h \downarrow 0} \frac{1}{h^2} \mathbb{E}\left[(\mathrm{e}_a^\top \epsilon(t,h))^2\right] = \infty.$$

Remark 4.7. An important consequence of Lemma 4.6 is that the mean-square limit of the difference quotient does not exist, and hence, for all components of the ICCSS process no mean-square derivatives higher than $(p - q - 1)$ exist. Thus, for local Granger causality and local contemporaneous correlation, we must always analyse the $(p - q - 1)$th derivative. It also becomes clear that in the definition of local contemporaneous correlation, one must divide by h and not by h^2.

Now, we specify the orthogonal projections.

Theorem 4.8. Let Y be an ICCSS(p, q) process with $p > q > 0$. Suppose $S \subseteq V$ and $a \in V$. Then, for $h \ge 0$ and $t \in \mathbb{R}$, we have

$$\begin{aligned}P_{\mathcal{L}_{Y_S}(t)} Y_a(t+h) =& \sum_{v \in S} \int_{-\infty}^{t} \mathrm{e}_a^\top \boldsymbol{M}(h) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u \\ &+ \sum_{v \in S} \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \boldsymbol{M}_m(h) \mathbf{\Theta} e_v D^{(m)} Y_v(t) \\ &+ P_{\mathcal{L}_{Y_S}(t)} \left(\sum_{v \in V \setminus S} \int_{-\infty}^{t} \mathrm{e}_a^\top \boldsymbol{M}(h) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u \right) \\ &+ P_{\mathcal{L}_{Y_S}(t)} \left(\sum_{v \in V \setminus S} \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \boldsymbol{M}_m(h) \mathbf{\Theta} e_v D^{(m)} Y_v(t) \right) \quad \mathbb{P}\text{-a.s.}\end{aligned}$$

and

$$\begin{aligned} P_{\mathcal{L}_{Y_S}(t)} D^{(p-q-1)} Y_a(t+h) &= \sum_{v \in S} \int_{-\infty}^{t} \mathrm{e}_a^{\top} \mathbf{M}(h) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u \\ &+ \sum_{v \in S} \sum_{m=0}^{p-q-1} \mathrm{e}_a^{\top} \mathbf{M}_m(h) \mathbf{\Theta} e_v D^{(m)} Y_v(t) \\ &+ P_{\mathcal{L}_{Y_S}(t)} \left(\sum_{v \in V \setminus S} \int_{-\infty}^{t} \mathrm{e}_a^{\top} \mathbf{M}(h) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u \right) \\ &+ P_{\mathcal{L}_{Y_S}(t)} \left(\sum_{v \in V \setminus S} \sum_{m=0}^{p-q-1} \mathrm{e}_a^{\top} \mathbf{M}_m(h) \mathbf{\Theta} e_v D^{(m)} Y_v(t) \right) \quad \mathbb{P}\text{-a.s.,} \end{aligned}$$

where $\boldsymbol{M}(\cdot)$, $\mathbf{M}(\cdot)$, $\boldsymbol{M}_m(\cdot)$, and $\mathbf{M}_m(\cdot)$ are defined in Theorem 4.2.

The basic idea of the proof is simple: In the representation in Theorem 4.2, the terms $Y_a(t)$, its derivatives and integrals over the past are already in the linear space $\mathcal{L}_{Y_S}(t)$ if $a \in S$ (Remark 3.15 and Proposition 4.5) and are therefore projected onto themselves. Furthermore, $(Y_S(s))_{s \leq t}$ and $(L(s) - L(t))_{t \leq s \leq t+h}$ are independent such that $e_a^{\top} \varepsilon(t, h)$ and $e_a^{\top} \epsilon(t, h)$ respectively, are independent of $\mathcal{L}_{Y_S}(t)$ and are projected onto zero. Of course, this argument can also be used in Example 3.8 (Example 4.4) to display the desired projections directly and we refrain from specifying them.

Remark 4.9. When calculating the orthogonal projections, it becomes clear why we require Assumption (3.12), a sufficient assumption to recover $X(t)$ from $(Y(s))_{s \leq t}$. Only then are we able to project the input process $X(t)$ onto the linear space of the output process $\mathcal{L}_{Y_S}(t)$.

To apply local Granger causality and local contemporaneous correlation to ICCSS processes, we also need the following orthogonal projections.

Theorem 4.10. Let Y be an ICCSS(p, q) process with $p > q > 0$. Suppose $S \subseteq V$ and $a \in V$. Then, for $t \in \mathbb{R}$, we have

$$\begin{aligned} &\underset{h \to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_S}(t)} \left(\frac{D^{(p-q-1)} Y_a(t+h) - D^{(p-q-1)} Y_a(t)}{h} \right) \\ &= \sum_{v \in S} \int_{-\infty}^{t} \mathrm{e}_a^{\top} \mathbf{M}'(0) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u + \sum_{v \in S} \sum_{m=0}^{p-q-1} \mathrm{e}_a^{\top} \mathbf{M}'_m(0) \mathbf{\Theta} e_v D^{(m)} Y_v(t) \\ &+ P_{\mathcal{L}_{Y_S}(t)} \left(\sum_{v \in V \setminus S} \int_{-\infty}^{t} \mathrm{e}_a^{\top} \mathbf{M}'(0) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u \right) \\ &+ P_{\mathcal{L}_{Y_S}(t)} \left(\sum_{v \in V \setminus S} \sum_{m=0}^{p-q-1} \mathrm{e}_a^{\top} \mathbf{M}'_m(0) \mathbf{\Theta} e_v D^{(m)} Y_v(t) \right) \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

and for $h \geq 0$,

$$D^{(p-q-1)} Y_a(t+h) - P_{\mathcal{L}_{Y}(t)} D^{(p-q-1)} Y_a(t+h) = \mathrm{e}_a^{\top} \epsilon(t, h) \quad \mathbb{P}\text{-a.s.,}$$

where $\mathbf{M}(\cdot)$, $\mathbf{M}_m(\cdot)$, and $\epsilon(\cdot, \cdot)$ are defined in Theorem 4.2.

In this article, for the derivation of the (local) orthogonality graph for ICCSS processes, the special case $S = V$ is most relevant, where a few terms are simplified.

DOI: 10.1111/jtsa.12787

Corollary 4.11. Let Y be an ICCSS(p, q) process with $p > q > 0$. Then, for $t \in \mathbb{R}$, $h \geq 0$, and $a \in V$ the following projections hold.

(a) $P_{\mathcal{L}_Y(t)} Y_a(t+h) = \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} X(t) \quad \mathbb{P}\text{-a.s.}$,
(b) $P_{\mathcal{L}_Y(t)} D^{(p-q-1)} Y_a(t+h) = \mathrm{e}_a^\top \mathbf{C} \mathrm{e}^{\boldsymbol{A}h} X(t) \quad \mathbb{P}\text{-a.s.}$,
(c) $\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_Y(t)} \left(\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h} \right) = \mathrm{e}_a^\top \mathbf{C} \boldsymbol{A} X(t) \quad \mathbb{P}\text{-a.s.}$

From Corollary 4.11(c), not only the existence of the limit becomes clear, but also that of the limit

$$\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_S}(t)} \left(\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h} \right) = P_{\mathcal{L}_{Y_S}(t)} \left(\mathrm{e}_a^\top \mathbf{C} \boldsymbol{A} X(t) \right).$$

The existence of these limits is essential for the well-definedness of local Granger causality and local contemporaneous correlation for ICCSS processes.

Remark 4.12.

(a) Although the derivation of the orthogonal projections for MCAR(p) processes differs from that for ICCSS(p, q) processes with $q > 0$, the results are consistent with Fasen-Hartmann and Schenk (2024a), Proposition 6.9 and Lemma 6.11 for MCAR processes, if we interpret $\boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} \mathrel{\hat{=}} 0_k$ for $u < t$ and $\boldsymbol{M}_m(h)\boldsymbol{\Theta} \mathrel{\hat{=}} \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}h}\mathbf{E}_m$ as in Remark 4.3.
(b) The linear projections in Corollary 4.11(a) match the linear projections for univariate CARMA processes in Brockwell and Lindner (2015), Theorem 2.8. Basse-O'Connor *et al.* (2019) derives as well linear projections for MCARMA processes, but the results there differ from Brockwell and Lindner (2015).

5. ORTHOGONALITY GRAPHS FOR ICCSS PROCESSES

Here, we derive (local) orthogonality graphs for state space models and obtain as the main result the characterisation of the directed and the undirected edges of the (local) orthogonality graph by the model parameters of the unique controller canonical form if this is an ICCSS(p, q) model with $p > q > 0$. To define the (local) orthogonality graph for ICCSS processes according to Definition 3.6, certain requirements for the well-definedness must be met. We have already assumed that we use the stationary version of the ICCSS process throughout the article and it has expectation zero. Furthermore, the continuity in the mean square of an ICCSS process is well known, it follows directly from Cramér (1940), Lemma 1, since the covariance function is continuous in 0. Therefore, we only need to make sure that the Assumptions (A.2) and (A.3) are satisfied.

Theorem 5.1. Let Y be an k-dimensional ICCSS(p, q) process with $\Sigma_L > 0$. Then Y satisfies Assumptions (A.2) and (A.3) and thus, the orthogonality graph and the local orthogonality graph are well defined and the Markov properties in Remark 3.7 hold.

Remark 5.2.

(a) In principle, more general state space models $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}}, L)$ also satisfy Assumption A. The proof of Theorem 5.1 shows that sufficient assumptions for the stationary state space process are that the driving Lévy process satisfies Assumption B, $\sigma(\boldsymbol{\mathcal{A}}) \subseteq (-\infty, 0) + i\mathbb{R}$, $f_{YY}(\lambda) > 0\ \forall \lambda \in \mathbb{R}$, and $\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{B}}\Sigma_L \boldsymbol{\mathcal{B}}^\top \boldsymbol{\mathcal{C}}^\top > 0$. Then the (local) orthogonality graphs are well defined as well, and the Markov properties in Remark 3.7 hold. However, in this general context, we are not able to calculate the orthogonal projections needed to characterise the edges, which is our main goal.
(b) In our running Example 3.8, we already know that Y is an ICCSS(2,1) process and, since $\Sigma_L = I_3 > 0$, the orthogonality graph and the local orthogonality graph are well defined and the Markov properties hold.

DOI: 10.1111/jtsa.12787

Let us now focus on the main results, i.e., the characterisations for the directed and the undirected edges in the (local) orthogonality graph for ICCSS processes. First, we present the characterisations of the (local) Granger non-causality.

Theorem 5.3. Let $Y = Y_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. Let $a, b \in V$ and $a \neq b$. Then the following characterisations hold:

(a) $Y_a \nrightarrow Y_b \mid Y_V \quad \Leftrightarrow \quad \mathrm{e}_b^\top \boldsymbol{C}\boldsymbol{A}^\alpha \left(\boldsymbol{E} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \mathsf{E}^\top \boldsymbol{\Lambda}^j\right) \boldsymbol{\Lambda}^\beta \boldsymbol{\Theta} e_a = 0$ and
$\mathrm{e}_b^\top \boldsymbol{C}\boldsymbol{A}^\alpha \left(\sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \mathsf{E}^\top \boldsymbol{\Lambda}^{j-1-m}\right) \boldsymbol{\Theta} e_a = 0,$
for $\alpha = 0, \ldots, kp-1$, $\beta = 0, \ldots, kq-1$, $m = 0, \ldots, p-q-1$.

(b) $Y_a \nrightarrow_0 Y_b \mid Y_V \quad \Leftrightarrow \quad \mathrm{e}_b^\top \mathbf{C}\boldsymbol{A} \left(\boldsymbol{E} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \mathsf{E}^\top \boldsymbol{\Lambda}^j\right) \boldsymbol{\Lambda}^\beta \boldsymbol{\Theta} e_a = 0$ and
$\mathrm{e}_b^\top \mathbf{C}\boldsymbol{A} \left(\sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \mathsf{E}^\top \boldsymbol{\Lambda}^{j-1-m}\right) \boldsymbol{\Theta} e_a = 0,$
for $\beta = 0, \ldots, kq-1$, $m = 0, \ldots, p-q-1$.

The basis for the proof of Theorem 5.3 are the following characterisations of the directed edges in Proposition 5.4. The characterisations in Proposition 5.4 are in turn developed from the definition of the directed edges in Section 3.1 and the orthogonal projections of the ICCSS process and its derivatives in Section 4.

Proposition 5.4. Let $Y = Y_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. Let $a, b \in V$ and $a \neq b$. Then the following characterisations hold:

(a) $Y_a \nrightarrow Y_b \mid Y_V \quad \Leftrightarrow \quad \mathrm{e}_b^\top \boldsymbol{M}(h) e^{\boldsymbol{\Lambda} t} \boldsymbol{\Theta} e_a = 0 \quad$ and $\quad \mathrm{e}_b^\top \boldsymbol{M}_m(h) \boldsymbol{\Theta} e_a = 0,$
for $m = 0, \ldots, p-q-1$, $0 \leq h \leq 1$, $t \geq 0$.

(b) $Y_a \nrightarrow_0 Y_b \mid Y_V \quad \Leftrightarrow \quad \mathrm{e}_b^\top \mathbf{M}'(0) \mathrm{e}^{\boldsymbol{\Lambda} t} \boldsymbol{\Theta} e_a = 0 \quad$ and $\quad \mathrm{e}_b^\top \mathbf{M}'_m(0) \boldsymbol{\Theta} e_a = 0,$
for $m = 0, \ldots, p-q-1$, $t \geq 0$.

Remark 5.5.

(a) Except for the assumption $\Sigma_L > 0$, the characterisations of the directed edges in the (local) orthogonality graph are independent of the chosen Lévy process, which is quite surprising. Thus, for example, the characterisations of (local) Granger causality are the same for a Brownian motion driven ICCSS process and a Poisson driven ICCSS process with the same controller canonical triple $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$, even if they have different covariance matrices and even though the path properties of these processes are significantly different. Note that in Example 3.8 we did not specify which Lévy process we are using.

(b) The characterisation in Proposition 5.4(a) seems to depend on h. However, this is not the case as can be seen from Theorem 5.3(a). So it does not matter whether we define directed edges by looking at the period $0 \leq h \leq 1$ or by looking at the entire future $h \geq 0$. In terms of Fasen-Hartmann and Schenk (2024a), there is no difference between Granger causality and global Granger causality for ICCSS processes.

Next, we present the characterisations of the undirected edges, i.e., contemporaneous uncorrelatedness.

Proposition 5.6. Let $Y = Y_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. Let $a, b \in V$ and $a \neq b$. Then the following characterisations hold:

(a) $Y_a \nsim Y_b \mid Y_V \;\Leftrightarrow\; \mathrm{e}_a^\top \int_0^{\min(h,\tilde{h})} \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}(h-s)} \boldsymbol{B} \Sigma_L \boldsymbol{B}^\top \mathrm{e}^{\boldsymbol{A}^\top (\tilde{h}-s)} \boldsymbol{C}^\top ds \; e_b = 0$, for $0 \leq h, \tilde{h} \leq 1$.
$\Leftrightarrow\; \mathrm{e}_a^\top \boldsymbol{C}\boldsymbol{A}^\alpha \boldsymbol{B} \Sigma_L \boldsymbol{B}^\top \left(\boldsymbol{A}^\top\right)^\beta \boldsymbol{C}^\top e_b = 0$, for $\alpha, \beta = 0, \ldots, kp-1$.

(b) $Y_a \nsim_0 Y_b \mid Y_V \;\Leftrightarrow\; \mathrm{e}_a^\top \mathbf{C} \boldsymbol{B} \Sigma_L \boldsymbol{B}^\top \mathbf{C}^\top e_b = \mathrm{e}_a^\top C_q \Sigma_L C_q^\top e_b = 0.$

The proof of this result again uses the orthogonal projections of ICCSS processes and its derivatives of Section 4 and the definition of undirected edges of Section 3.1. The assumption $\Sigma_L > 0$ is only used for the second characterisation in Proposition 5.6(a). However, it was also important for the proof of Assumption A.

DOI: 10.1111/jtsa.12787

Remark 5.7.

(a) The characterisations and thus the undirected edges in the (local) orthogonality graph depend on the chosen Lévy process only by Σ_L. For example, the characterisations of the (local) contemporaneous correlation and thus the (local) orthogonality graph are the same for a Brownian motion driven ICCSS process and a Poisson driven ICCSS process with the same controller canonical triple $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ if both Lévy processes have the same covariance matrix Σ_L. However, in contrast to (local) Granger causality, it is necessary that the Brownian motion and the Poisson process have the same covariance matrix. Again, in our running Example 3.8, we did not specify which Lévy process we were using, but we did specify that $\Sigma_L = I_3$.

(b) The second characterisation in Proposition 5.6(a) shows that there is indeed no dependence on the lag h again. As for the directed edges, it does not matter whether we define undirected edges by looking at the period $0 \leq h, \tilde{h} \leq 1$ or by looking at the entire future $h, \tilde{h} \geq 0$. In terms of Fasen-Hartmann and Schenk (2024a), there is no difference between contemporaneous correlation and global contemporaneous correlation for ICCSS processes.

We make some further comments on the characterisations in Propositions 5.4 and 5.6. In particular, we compare the characterisations with each other and with the results in the literature, additionally we give some interpretations.

Remark 5.8.

(a) The uniqueness of the polynomials $P(z)$ and $Q(z)$ in (3.4) (see Proposition 3.9) leads to the uniqueness of the controller canonical state space representation, which in turn leads to the uniqueness of the edges in the (local) orthogonality graph.

(b) It can be shown by a simple calculation that $\boldsymbol{C}\boldsymbol{A}^{p-q} = \mathbf{C}\boldsymbol{A}$. Comparing Theorem 5.3(a) and (b), we receive that Granger non-causality implies local Granger non-causality, which we know as well from the theory in Fasen-Hartmann and Schenk (2024a). Similarly, $\boldsymbol{C}\boldsymbol{A}^{p-q-1} = \mathbf{C}$, so comparing Proposition 5.6(a) and (b), we get that contemporaneous uncorrelatedness implies local contemporaneous uncorrelatedness, which is again in agreement with the theory. The relationships between Granger non-causality and local Granger non-causality, as well as contemporaneous uncorrelatedness and local contemporaneous uncorrelatedness in a general setting, are discussed in Fasen-Hartmann and Schenk (2024a), Lemma 3.13 and Lemma 4.10.

(c) In Example 3.8, straightforward computations yield the (local) Granger causality relations and (local) contemporaneous correlations, visualised in the corresponding (local) causality graphs in Figure 1. Again, the relationships between Granger non-causality and local Granger non-causality, and between contemporaneous uncorrelation and local contemporaneous uncorrelation are evident.

Interpretation 5.9. (Orthogonality graph)**.** To interpret the directed and the undirected edges in the orthogonality graph G_{OG}, we recall the representation of the bth component

$$Y_b(t+h) = \int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h) \mathrm{e}^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} Y_V(u) \mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h) \boldsymbol{\Theta} D^{(m)} Y_V(t) + \mathrm{e}_b^\top \varepsilon(t,h)$$

from Theorem 4.2.

(a) Directed edges: A direct application of Proposition 5.4 gives that $a \to b \notin E_{\mathrm{OG}}$, if and only if neither $Y_a(t)$, $D^{(1)} Y_a(t), \ \dots, D^{(p-q-1)} Y_a(t)$nor the integral over the past have any influence on $Y_b(t+h)$. In the representation of the bth component $Y_b(t+h)$, the ath component always vanishes because its coefficient functions are zero. This observation is also evident in Example 3.8 as derived in Example 4.4.

(b) Undirected edges: Proposition 5.6 yields

$$a \dashrightarrow b \notin E_{\mathrm{OG}} \quad \Leftrightarrow \quad \mathbb{E}[\mathrm{e}_a^\top \varepsilon(t,h) \mathrm{e}_b^\top \varepsilon(t,\tilde{h})] = \mathbb{E}[\mathrm{e}_a^\top \varepsilon(0,h) \mathrm{e}_b^\top \varepsilon(0,\tilde{h})] = 0, \quad 0 \leq h, \tilde{h} \leq 1.$$

This means that the noise terms $e_a^\top \varepsilon(t,h)$ and $e_b^\top \varepsilon(t,\tilde{h})$ of $Y_a(t+h)$ and $Y_b(t+\tilde{h})$ respectively, are uncorrelated for any $t \geq 0$ and $0 \leq h, \tilde{h} \leq 1$. Again, this observation is also evident in Example 3.8 (Example 4.4). However, due to the complexity of the expression $\mathbb{E}[\varepsilon(t,h)\varepsilon(t,\tilde{h})^\top] = \int_0^{\min(h,\tilde{h})} \boldsymbol{C} e^{\boldsymbol{A}(h-s)} \boldsymbol{B} \Sigma_L \boldsymbol{B}^\top e^{\boldsymbol{A}^\top(\tilde{h}-s)} \boldsymbol{C}^\top ds$, we do not specify the latter.

Interpretation 5.10. (Local orthogonality graph)**.** The interpretation of the directed and the undirected edges in the local orthogonality graph G_{OG}^0 is a lot more intricate since the mean square limit of the difference quotient does not exist by definition and Remark 4.7 respectively, but the limit of the projections does. Therefore we again use the representation for $b \in V$ of Lemma 4.6,

$$
\begin{aligned}
\mathrm{D}_h^{(p-q-1)} Y_b(t,h) &:= \frac{D^{(p-q-1)} Y_b(t+h) - D^{(p-q-1)} Y_b(t)}{h} \\
&= \int_{-\infty}^t \mathrm{e}_b^\top \mathbf{M}'(0) \mathrm{e}^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} Y_V(u) \mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \mathbf{M}'_m(0) \boldsymbol{\Theta} D^{(m)} Y_V(t) \\
&\quad + \mathrm{e}_b^\top O(h) R_1 + \mathrm{e}_b^\top O(h) R_2 + \frac{\mathrm{e}_b^\top \epsilon(t,h)}{h},
\end{aligned}
$$

and hence,

$$
\begin{aligned}
P_{\mathcal{L}_{Y_V}(t)} \mathrm{D}_h^{(p-q-1)} Y_b(t,h) &= \int_{-\infty}^t \mathrm{e}_b^\top \mathbf{M}'(0) e^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} Y_V(u) \mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \mathbf{M}'_m(0) \boldsymbol{\Theta} D^{(m)} Y_V(t) \\
&\quad + \mathrm{e}_b^\top O(h) R_1 + \mathrm{e}_b^\top O(h) R_2.
\end{aligned}
$$

Despite the fact that the L^2-limit of $\mathrm{D}_h^{(p-q-1)} Y_b(t,h)$ does not exist, the L^2-limits of $\sqrt{h}\mathrm{D}_h^{(p-q-1)} Y_b(t,h)$ and $P_{\mathcal{L}_{Y_V}(t)} \mathrm{D}_h^{(p-q-1)} Y_b(t,h)$ exist.

(a) Directed edges: By Proposition 5.4 we receive that $a \longrightarrow b \notin E_{\mathrm{OG}}^0$, if and only if, neither $Y_a(t), D^{(1)} Y_a(t), \dots, D^{(p-q-1)} Y_a(t)$ nor the integral over the past have any influence on $\mathrm{D}_h^{(p-q-1)} Y_b(t,h)$ if h is small. The same holds for $P_{\mathcal{L}_{Y_V}(t)} \mathrm{D}_h^{(p-q-1)} Y_b(t,h)$. Given $\mathcal{L}_{Y_V}(t)$, the ath component does not influence the bth component in the limit, because the corresponding coefficients are zero. Note that in Example 3.8, we have

$$
\int_{-\infty}^t \mathrm{M}'(0) \mathrm{e}^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} Y_V(u) \mathrm{d}u + \mathrm{M}'_0(0) \boldsymbol{\Theta} Y_V(t) = \int_{-\infty}^t -\mathrm{e}^{-(t-u)} \begin{pmatrix} Y_1(u) \\ Y_2(u) \\ Y_3(u) - \frac{1}{2}(Y_1(u) - Y_2(u)) \end{pmatrix} \mathrm{d}u + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},
$$

which explains the directed edges in the local orthogonality graph in Figure 1.

(b) Undirected edges: By Proposition 5.6 we receive that $a \,\text{---}\, b \notin E_{\mathrm{OG}}^0$, if and only if,

$$
\begin{aligned}
& h\mathbb{E}\Big[\Big(\mathrm{D}_h^{(p-q-1)} Y_a(t,h) - P_{\mathcal{L}_V(t)} \mathrm{D}_h^{(p-q-1)} Y_a(t,h)\Big)\Big(\mathrm{D}_h^{(p-q-1)} Y_b(t,h) - P_{\mathcal{L}_V(t)} \mathrm{D}_h^{(p-q-1)} Y_b(t,h)\Big)\Big] \\
&= \frac{1}{h} \mathbb{E}\big[\mathrm{e}_a^\top \epsilon(t,h) e_b^\top \epsilon(t,h)\big] \overset{h\downarrow 0}{\longrightarrow} \mathrm{e}_a^\top \mathrm{C} \boldsymbol{B} \Sigma_L \boldsymbol{B}^\top \mathrm{C}^\top e_b,
\end{aligned}
$$

DOI: 10.1111/jtsa.12787

is zero. Hence, given $\mathcal{L}_Y(t)$, $\sqrt{h}\mathbb{D}_h^{(p-q-1)}Y_a(t,h)$ and $\sqrt{h}\mathbb{D}_h^{(p-q-1)}Y_b(t,h)$ are uncorrelated in the limit. Equivalently, the noise terms $e_a^\top \epsilon(t,h)/\sqrt{h}$ and $e_b^\top \epsilon(t,h)/\sqrt{h}$ are uncorrelated in the limit. Note that in Example 3.8, we have

$$\mathbf{C}\boldsymbol{B}\Sigma_L\boldsymbol{B}^\top\mathbf{C}^\top = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 3 \end{pmatrix},$$

which explains the undirected edges in the local orthogonality graph in Figure 1.

Remark 5.11. We establish the relationship between our results for ICCSS processes and the results for MCAR processes in Fasen-Hartmann and Schenk (2024a).

(a) Since the undirected edges are characterised only by the noise terms $\epsilon(t,h)$ and $\epsilon(t,h)$ and thus, have nothing to do with the inversion of the process, it is not surprising that the characterisations for the undirected edges of the ICCSS(p,q) processes and for the undirected edges of the MCAR(p) processes coincide.
(b) In the characterisations of the directed edges of the ICCSS(p,q) process, the case $q = 0$ cannot simply be inserted because several matrices become zero-dimensional. However, if we interpret $\boldsymbol{M}(h)\mathrm{e}^{\mathbf{A}(t-u)}\boldsymbol{\Theta} \mathrel{\widehat{=}} 0_{k\times k}$ if $u < t$, and $\boldsymbol{M}_m(h)\boldsymbol{\Theta} \mathrel{\widehat{=}} \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}h}\mathbf{E}_{m+1}$ for $m = 0, \dots, p-1$, as in Remark 4.3, the characterisations of the directed edges for MCAR(p) processes can be seen as special case of Theorem 5.3 and Proposition 5.4.

6. PROOFS

6.1. Proofs of Section 3.2

Proof of Proposition 3.9. Assume that there exist two coprime right polynomial fraction descriptions of $H(z)$ as in (3.5), so that

$$Q_1(z)P_1(z)^{-1} = H(z) = Q_2(z)P_2(z)^{-1}.$$

Then, due to the coprimeness, there exists a matrix polynomial $U(z)$, where $\det(U(z))$ is a non-zero real number (Rugh, 1996, Theorem 16.10), such that

$$P_1(z) = P_2(z)U(z). \tag{6.1}$$

Both $P_1(z)$ and $P_2(z)$ have the highest power $I_k z^p$, so $U(z) = I_k$. Hence $P_1(z) = P_2(z)$ and finally, $Q_1(z) = Q_2(z)$, which results in the uniqueness of the decomposition.

The fact that $H(z)$ is equal to $\boldsymbol{C}\left(zI_{kp} - \boldsymbol{A}\right)^{-1}\boldsymbol{B}$ follows from the proof of Theorem 3.2 in Brockwell and Schlemm (2013).

Furthermore, the realisations $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C})$ and $(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}},\boldsymbol{C})$ are minimal because $P(z)$ and $Q(z)$ are right coprime and $\deg(\det(P(z))) = kp$, see Theorem 6.5-1 of Kailath (1980). Then a consequence of Theorem 2.3.4 in Hannan and Deistler (2012) is that there exists a non-singular matrix T such that

$$\boldsymbol{A} = T\boldsymbol{\mathcal{A}}T^{-1}, \quad \boldsymbol{B} = T\boldsymbol{\mathcal{B}}, \quad \text{and} \quad \boldsymbol{C} = \boldsymbol{C}T^{-1},$$

and for $s < t$,

$$Y(t) = \boldsymbol{C}\mathrm{e}^{\boldsymbol{\mathcal{A}}(t-s)}X(s) + \int_s^t \boldsymbol{C}\mathrm{e}^{\boldsymbol{\mathcal{A}}(t-u)}\boldsymbol{\mathcal{B}}\mathrm{d}L(u)$$

$$= \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}(t-s)}(TX(s)) + \int_s^t \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}(t-u)}\boldsymbol{B}\mathrm{d}L(u).$$

Thus, Y is a solution of the state space model $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}}, L)$, if and only if it is a solution of $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, L)$. Finally, $\sigma(\boldsymbol{A}) = \sigma(T\boldsymbol{\mathcal{A}}T^{-1}) = \sigma(\boldsymbol{\mathcal{A}})$. ■

Proof of Lemma 3.13. The uniqueness follows directly from Proposition 3.9. Furthermore, $\mathfrak{p} = p$ holds by Lemma 6.5-6 of Kailath (1980). Since $\mathcal{P}(z)^{-1}\mathcal{Q}(z) = Q(z)P(z)^{-1}$ we have $\mathfrak{p} - \mathfrak{q} = p - q$ and therefore $\mathfrak{q} = q$. Comparing the highest-order coefficient in $\mathcal{Q}(z)P(z) = \mathcal{P}(z)Q(z)$ gives $\mathcal{Q}_0 = C_q$. Finally, Lemma 6.3-8 in Kailath (1980) states that $\mathcal{N}(\mathcal{P}) = \mathcal{N}(P)$ and $\mathcal{N}(\mathcal{Q}) = \mathcal{N}(Q)$. ■

6.2. Proofs of Section 4

Proof of Proposition 4.1. The proof is divided into four steps. In the first three steps, we derive some auxiliary results which lead in Step 4 to the proof of the statement.

Step 1: First, we prove for all $\varepsilon > 0$ and $v \in V$ the asymptotic behaviour

$$\lim_{|u|\to\infty} e^{-\varepsilon|u|}|Y_v(u)| = 0 \quad \mathbb{P}\text{-a.s.} \tag{6.2}$$

Thus, we relate (6.2) back to Brockwell and Lindner (2015), Proposition 2.6, who prove this convergence for stationary univariate CARMA processes that are driven by univariate Lévy processes and whose AR polynomial has no zeros on the imaginary axis. Therefore, let $\varepsilon > 0$ and $v \in V$. Note that for $t \in \mathbb{R}$,

$$Y_v(t) = \int_{-\infty}^t \mathrm{e}_v^\top \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}(t-u)}\boldsymbol{B}\mathrm{d}L(u) = \sum_{\ell=1}^k \int_{-\infty}^t \mathrm{e}_v^\top \boldsymbol{C}\mathrm{e}^{\boldsymbol{A}(t-u)}\boldsymbol{B}e_\ell \mathrm{d}L_\ell(u) = \sum_{\ell=1}^k Y_v^\ell(t).$$

The process $Y_v^\ell = (Y_v^\ell(t))_{t\in\mathbb{R}}$ is the stationary solution of the state space model

$$\mathrm{d}X(t) = \boldsymbol{A}X(t)\mathrm{d}t + \boldsymbol{B}e_\ell \mathrm{d}L_\ell(t), \qquad Y_v^\ell(t) = \mathrm{e}_v^\top \boldsymbol{C}X(t),$$

and has the transfer function

$$H_v^\ell(z) = \mathrm{e}_v^\top \boldsymbol{C}(zI_{kp} - \boldsymbol{A})^{-1}\boldsymbol{B}e_\ell.$$

Then Kailath (1980) provides in Lemma 6.3-8 the existence of (right) coprime polynomials $P_v^\ell(z)$ and $Q_v^\ell(z)$ (polynomials with no common zeros) as in (3.5) so that $H_v^\ell(z) = Q_v^\ell(z)/P_v^\ell(z)$. Note that in the univariate setting the problem of the existence of a coprime right polynomial fraction description of the form (3.5) does not arise. Indeed, here $1 \cdot p = \deg(\det(P_v^\ell(z))) = \deg(P_v^\ell(z))$ follows immediately, and the constant before the p-th power can be included in $Q_v^\ell(z)$ without loss of generality so that $P_v^\ell(z)$ is a polynomial of degree p that has a 1 as the leading coefficient. Thus, the classes of univariate CARMA processes and univariate causal continuous-time state space models are equivalent (Schlemm and Stelzer, 2012a, Corollary 3.4) implying that Y_v^ℓ is a univariate CARMA process driven by a univariate Lévy process. Now, Bernstein (2009), Definition 4.7.1, provides that the poles of $H_v^\ell(z)$ are the roots of $P_v^\ell(z)$ including multiplicity. In addition, Bernstein (2009), Theorem 12.9.16, yields that the poles of $H_v^\ell(z)$ are a subset of $\sigma(\boldsymbol{A})$ resulting in

$$\mathcal{N}(P_v^\ell) = \{z \in \mathbb{C} : P_v^\ell(z) = 0\} \subseteq \sigma(\boldsymbol{A}) \subseteq (-\infty, 0) + i\mathbb{R},$$

DOI: 10.1111/jtsa.12787

which means that the AR polynomial $P_v^\ell(z)$ has no zeros on the imaginary axis. Thus, Y_v^ℓ satisfies the assumptions in Brockwell and Lindner (2015), Proposition 2.6, and we obtain for $\ell = 1, \dots, k$ that

$$\lim_{|u|\to\infty} \mathrm{e}^{-\varepsilon|u|} Y_v^\ell(u) = 0 \quad \mathbb{P}\text{-a.s.}$$

Therefore,

$$\lim_{|u|\to\infty} e^{-\varepsilon|u|} Y_v(u) = \sum_{\ell=1}^{k} \lim_{|u|\to\infty} e^{-\varepsilon|u|} Y_v^\ell(u) = 0 \quad \mathbb{P}\text{-a.s.},$$

and finally, the claim (6.2) follows.

Step 2: Next, we show that

$$\lim_{s\to-\infty} \int_s^t \mathrm{e}^{-\lambda(t-u)} |Y_v(u)| \mathrm{d}u, \tag{6.3}$$

exists $\mathbb{P}$-a.s. for $t \in \mathbb{R}$ and $\lambda > 0$.

From (6.2) we obtain that there exists some set $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$ and $\gamma > 0$ there exists a $u_0(\omega) < 0$ with

$$\mathrm{e}^{\frac{\lambda}{2}u} |Y_v(\omega, u)| = \mathrm{e}^{-\frac{\lambda}{2}|u|} |Y_v(\omega, u)| \leq \gamma \quad \forall u \leq u_0(\omega).$$

Then we obtain for $s < u_0(\omega)$ that

$$\begin{aligned}\int_s^t \mathrm{e}^{-\lambda(t-u)} |Y_v(\omega, u)| \mathrm{d}u &= \int_{u_0(\omega)}^t \mathrm{e}^{-\lambda(t-u)} |Y_v(\omega, u)| \mathrm{d}u + \int_s^{u_0(\omega)} \mathrm{e}^{-\lambda(t-u)} |Y_v(\omega, u)| \mathrm{d}u \\ &\leq \int_{u_0(\omega)}^t \mathrm{e}^{-\lambda(t-u)} |Y_v(\omega, u)| \mathrm{d}u + \gamma \mathrm{e}^{-\lambda t} \frac{2}{\lambda}.\end{aligned}$$

Thus, by dominated convergence the limit in (6.3) exists $\mathbb{P}$-a.s. for $t \in \mathbb{R}$ and $\lambda > 0$.

Step 3: Eventually, we derive that not only the univariate integral (6.3) exists, but also

$$\lim_{s\to-\infty} \int_s^t \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) \mathrm{d}u, \tag{6.4}$$

exists $\mathbb{P}$-a.s. for $t \in \mathbb{R}$. First, Assumption (3.12) provides that $\sigma(\mathbf{\Lambda}) \subseteq (-\infty, 0) + i\mathbb{R}$ and thus, $\operatorname{spabs}(\mathbf{\Lambda}) := \max\{\Re(\lambda) : \lambda \in \sigma(\mathbf{\Lambda})\} < 0$, where $\Re(\lambda)$ denotes the real part of λ. Therefore, there exists a $-\lambda \in (\operatorname{spabs}(\mathbf{\Lambda}), 0)$. Then Bernstein (2009), Proposition 11.18.8, provides a constant $c_1 > 0$ such that

$$\|\mathrm{e}^{\mathbf{\Lambda}t}\| \leq c_1 \mathrm{e}^{-\lambda t} \quad \forall t \geq 0. \tag{6.5}$$

Now, we obtain

$$\left\| \int_s^t \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) \mathrm{d}u \right\| \leq c_1 \|\mathbf{\Theta}\| \sum_{v\in V} \int_s^t \mathrm{e}^{-\lambda(t-u)} |Y_v(u)| \mathrm{d}u.$$

Due to (6.3) the limit of each of those addends exists, so (6.4) exists $\mathbb{P}$-a.s. for $t \in \mathbb{R}$.

Step 4: Finally, we are able to prove the statement of the proposition. Recall that due to (3.14) for $s, t \in \mathbb{R}$, $s < t$,

$$X^q(t) = \mathrm{e}^{\mathbf{\Lambda}(t-s)} X^q(s) + \int_s^t \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) \mathrm{d}u.$$

Since we assume that X is the unique stationary solution of the stochastic differential equation (1.1), X^q is also strictly stationary and $X^q(s)$ and $X^q(0)$ have the same distribution for all $s \in \mathbb{R}$. Moreover, it follows from Assumption (3.12) that $\sigma(\mathbf{\Lambda}) \subseteq (-\infty, 0) + i\mathbb{R}$. These properties lead to

$$e^{\mathbf{\Lambda}(t-s)} X^q(s) \to 0_{kq} \quad \text{as } s \to -\infty,$$

in distribution and in probability by Slutsky's lemma, since the limit is a degenerate random vector. In combination with (6.4) we receive for $t \in \mathbb{R}$ the statement

$$\lim_{s \to -\infty} \left(\mathrm{e}^{\mathbf{\Lambda}(t-s)} X^q(s) + \int_s^t \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) \mathrm{d}u \right) = \int_{-\infty}^t \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) \mathrm{d}u \quad \mathbb{P}\text{-a.s.}$$

■

Proof of Theorem 4.2. Let $t \in \mathbb{R}$, $h \geq 0$, and $a \in V$. First, due to (3.11), we receive

$$Y_a(t+h) = \mathrm{e}_a^\top \boldsymbol{C} X(t+h) \quad \text{and} \quad D^{(p-q-1)} Y_a(t+h) = \mathrm{e}_a^\top \mathbf{C} X(t+h).$$

From now on, the proofs for the two representations differ only in the choice of $\boldsymbol{C}$ and $\mathbf{C}$, respectively. Therefore, we will only continue with the representation of $Y_a(t+h)$. Due to (1.2) we have

$$Y_a(t+h) = \mathrm{e}_a^\top \boldsymbol{C} \left(\mathrm{e}^{\boldsymbol{A}h} X(t) + \int_t^{t+h} \mathrm{e}^{\boldsymbol{A}(t+h-u)} \boldsymbol{B} \mathrm{d}L(u) \right) = \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} X(t) + \mathrm{e}_a^\top \varepsilon(t,h).$$

Here,

$$\begin{aligned} \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} X(t) &= \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} \left(X^q(t),\ X^{(q+1)}(t),\ \cdots,\ X^{(p)}(t) \right)^\top \\ &= \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} \boldsymbol{E} X^q(t) + \sum_{j=1}^{p-q} \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} \mathbf{E}_{q+j} X^{(q+j)}(t). \end{aligned}$$

Lemma 3.20 and interchanging the summation order imply

$$\begin{aligned} \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} X(t) &= \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} \boldsymbol{E} X^q(t) + \sum_{j=1}^{p-q} \mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} \mathbf{E}_{q+j} \mathrm{E}^\top \left(\mathbf{\Lambda}^j X^q(t) + \sum_{m=0}^{j-1} \mathbf{\Lambda}^{j-1-m} \mathbf{\Theta} D^{(m)} Y(t) \right) \\ &= \mathrm{e}_a^\top \boldsymbol{M}(h) X^q(t) + \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \boldsymbol{M}_m(h) \mathbf{\Theta} D^{(m)} Y(t). \end{aligned} \tag{6.6}$$

Finally, we obtain due to Proposition 4.1,

$$\mathrm{e}_a^\top \boldsymbol{C} \mathrm{e}^{\boldsymbol{A}h} X(t) = \int_{-\infty}^t \mathrm{e}_a^\top \boldsymbol{M}(h) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) du + \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \boldsymbol{M}_m(h) \mathbf{\Theta} D^{(m)} Y(t) \quad \mathbb{P}\text{-a.s.}$$

■

DOI: 10.1111/jtsa.12787

Proof of Proposition 4.5. Let $a, v \in V$ and define $F(t) = \mathrm{e}_a^\top \mathrm{e}^{\boldsymbol{\Lambda} t} \boldsymbol{\Theta} \mathrm{e}_v$ for $t \geq 0$. First, for $s, t \in \mathbb{R}$, $s < t$,

$$\lim_{n\to\infty} \frac{t-s}{n} \sum_{\ell=1}^{n} F\Big(t - s - \ell \frac{t-s}{n}\Big) Y_v\Big(s + \ell \frac{t-s}{n}\Big) = \int_s^t F(t-u) Y_v(u) \mathrm{d}u \quad \mathbb{P}\text{-a.s.}$$

due to the definition of the integral. Using the theorem of dominated convergence, we show that this convergence also holds in the L^2 sense. Indeed, from the triangle inequality

$$\begin{aligned} &\left| \int_s^t F(t-u) Y_v(u) \mathrm{d}u - \frac{t-s}{n} \sum_{\ell=1}^{n} F\Big(t - s - \ell \frac{t-s}{n}\Big) Y_v\Big(s + \ell \frac{t-s}{n}\Big) \right| \\ &\leq \int_s^t |F(t-u)| |Y_v(u)| \mathrm{d}u + \frac{t-s}{n} \sum_{\ell=1}^{n} \left| F\Big(t - s - \ell \frac{t-s}{n}\Big) \right| \left| Y_v\Big(s + \ell \frac{t-s}{n}\Big) \right| \\ &\leq 2(t-s)(\sup_{u\in[0,t-s]} |F(u)|) \Big(\sup_{u\in[s,t]} |Y_v(u)| \Big) \end{aligned}$$

follows. This majorant is integrable, because

$$\sup_{u\in[0,t-s]} |Y_v(u)| = \sup_{u\in[0,t-s]} |\mathrm{e}_v^\top \boldsymbol{C} X(u)| \leq \sup_{u\in[0,t-s]} \|\mathrm{e}_v^\top \boldsymbol{C}\| \|X(u)\| \leq c \sup_{u\in[0,t-s]} \|X(u)\|,$$

for some constant $c \geq 0$ and thus,

$$\mathbb{E}\left[\Big(\sup_{u\in[0,t-s]} |Y_v(u)| \Big)^2 \right] \leq c^2 \mathbb{E}\left[(\sup_{u\in[0,t-s]} \|X(u)\|)^2 \right] < \infty, \tag{6.7}$$

due to Assumption B and Brockwell and Schlemm (2013), Lemma A.4. Furthermore, $\sup_{u\in[0,t-s]} |F(u)| < \infty$ since F is a continuous function. In summary,

$$\int_s^t F(t-u) Y_v(u) \mathrm{d}u = \underset{n\to\infty}{\mathrm{l.i.m.}} \frac{t-s}{n} \sum_{\ell=1}^{n} F\Big(t - s - \ell \frac{t-s}{n}\Big) Y_v\Big(s + \ell \frac{t-s}{n}\Big).$$

For the second step of this proof, we recall that for $t \in \mathbb{R}$,

$$\int_{-\infty}^t F(t-u) Y_v(u) \mathrm{d}u = \lim_{s\to-\infty} \int_s^t F(t-u) Y_v(u) \mathrm{d}u \quad \mathbb{P}\text{-a.s.},$$

due to Proposition 4.1. Again, using the theorem of dominated convergence, we show that this convergence holds in the L^2-sense. For $s < t$ it follows that

$$\begin{aligned} \left| \int_{-\infty}^t F(t-u) Y_v(u) \mathrm{d}u - \int_s^t F(t-u) Y_v(u) \mathrm{d}u \right| &\leq \int_{t-s}^{\infty} |F(u)| |Y_v(t-u)| \mathrm{d}u \\ &\leq \sum_{n=0}^{\infty} \sup_{u\in[n,n+1]} |F(u)| \sup_{u\in[n,n+1]} |Y_v(t-u)|. \end{aligned}$$

J. Time Ser. Anal. (2024) DOI: 10.1111/jtsa.12787

Journal of Time Series Analysis published by John Wiley & Sons Ltd. wileyonlinelibrary.com/journal/jtsa

To see that this majorant is in L^2, we use Fubini, Cauchy–Schwarz inequality and the stationarity of Y. This yields to

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{n=0}^{\infty}\sup_{u\in[n,n+1]}|F(u)|\sup_{u\in[n,n+1]}|Y_v(t-u)|\right)^2\right]\\
&\le \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sup_{u\in[n,n+1]}|F(u)|\sup_{u\in[m,m+1]}|F(u)|\\
&\qquad\times\left(\mathbb{E}\left[\left(\sup_{u\in[n,n+1]}|Y_v(t-u)|\right)^2\right]\mathbb{E}\left[\left(\sup_{u\in[m,m+1]}|Y_v(t-u)|\right)^2\right]\right)^{1/2}\\
&= \left(\sum_{n=0}^{\infty}\sup_{u\in[n,n+1]}|F(u)|\right)^2\mathbb{E}\left[\left(\sup_{u\in[0,1]}|Y_v(u)|\right)^2\right] < \infty,
\end{aligned}$$

where we used (6.7) and $\sum_{n=0}^{\infty}\sup_{u\in[n,n+1]}|F(u)| < \infty$ by the definition of F and (6.5). In summary, we obtain

$$\mathrm{e}_a^\top\int_{-\infty}^{t}\mathrm{e}^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta}\mathrm{e}_vY_v(u)\mathrm{d}u = \int_{-\infty}^{t}F(t-u)Y_v(u)\mathrm{d}u = \underset{s\to-\infty}{\mathrm{l.i.m.}}\int_{s}^{t}F(t-u)Y_v(u)\mathrm{d}u,$$

and the integral is in $\mathcal{L}_{Y_v}(t)$. The existence of $X^q(t)$ as an L^2-limit follows immediately from this. ■

Proof of Lemma 4.6. Recall that due to Theorem 4.2 and $\epsilon(t,0)=0_k\in\mathbb{R}^k$

$$\begin{aligned}
&\frac{D^{(p-q-1)}Y_a(t+h)-D^{(p-q-1)}Y_a(t)}{h}\\
&=\int_{-\infty}^{t}\mathrm{e}_a^\top\frac{\mathbf{M}(h)-\mathbf{M}(0)}{h}\mathrm{e}^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta}Y(u)\mathrm{d}u\\
&\quad+\sum_{m=0}^{p-q-1}\mathrm{e}_a^\top\frac{\mathbf{M}_m(h)-\mathbf{M}_m(0)}{h}\mathbf{\Theta}D^{(m)}Y(t)+\mathrm{e}_a^\top\frac{\epsilon(t,h)}{h}\quad\mathbb{P}\text{-a.s.}
\end{aligned}\tag{6.8}$$

Replacing the matrix exponential with its power series, it holds that

$$\begin{aligned}
\frac{\mathrm{M}(h)-\mathrm{M}(0)}{h}&=\mathrm{C}\frac{\mathrm{e}^{\mathbf{A}h}-I_{kp}}{h}\left(\boldsymbol{E}+\sum_{j=1}^{p-q}\mathbf{E}_{q+j}\mathrm{E}^\top\mathbf{\Lambda}^j\right)=\mathrm{M}'(0)+O(h)\left(\boldsymbol{E}+\sum_{j=1}^{p-q}\mathbf{E}_{q+j}\mathrm{E}^\top\mathbf{\Lambda}^j\right),\\
\frac{\mathrm{M}_m(h)-\mathrm{M}_m(0)}{h}&=\mathrm{C}\frac{\mathrm{e}^{\mathbf{A}h}-I_{kp}}{h}\sum_{j=m+1}^{p-q}\mathbf{E}_{q+j}\mathrm{E}^\top\mathbf{\Lambda}^{j-1-m}=\mathrm{M}_m'(0)+O(h)\sum_{j=m+1}^{p-q}\mathbf{E}_{q+j}\mathrm{E}^\top\mathbf{\Lambda}^{j-1-m}.
\end{aligned}\tag{6.9}$$

Furthermore, we define

$$\begin{aligned}
R_1&=\int_{-\infty}^{t}\left(\boldsymbol{E}+\sum_{j=1}^{p-q}\mathbf{E}_{q+j}\mathbf{E}^\top\mathbf{\Lambda}^j\right)\mathrm{e}^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta}Y(u)\mathrm{d}u,\\
R_2&=\sum_{m=0}^{p-q-1}\sum_{j=m+1}^{p-q}\mathbf{E}_{q+j}\mathbf{E}^\top\mathbf{\Lambda}^{j-1-m}\mathbf{\Theta}D^{(m)}Y(t).
\end{aligned}\tag{6.10}$$

If we plug (6.9) and (6.10) in (6.8) we obtain the stated representation. Moreover, from Proposition 4.5 we know that R_1 is in $\mathcal{L}_Y(t)$ and from Remark 3.15 we receive that R_2 is in $\mathcal{L}_Y(t)$. Since $\mathcal{L}_{Y(t)}$ and $(L(s)-L(t))_{t\le s\le t+h}$ are independent we receive that $R_1, R_2 \in \mathcal{L}_Y(t)$ are independent of $\epsilon(t,h)$. Finally,

$$\frac{1}{h}\mathbb{E}\left[(\mathrm{e}_a^\top \epsilon(t,h))^2\right] = \frac{1}{h}\mathrm{e}_a^\top \mathrm{C}\int_0^h \mathrm{e}^{\boldsymbol{A}u}\boldsymbol{B}\Sigma_L\boldsymbol{B}^\top \mathrm{e}^{\boldsymbol{A}^\top u}\mathrm{d}u\mathrm{C}^\top \mathrm{e}_a \overset{h\downarrow 0}{\to} \mathrm{e}_a^\top \mathrm{C}\boldsymbol{B}\Sigma_L\boldsymbol{B}^\top \mathrm{C}^\top \mathrm{e}_a.$$

$\mathrm{C}\boldsymbol{B}\Sigma_L\boldsymbol{B}^\top\mathrm{C}^\top$ is positive definite due to $\Sigma_L > 0$ and $\mathbf{C}$, $\boldsymbol{B}$ being of full rank by Assumption (3.12). Therefore, the limit $\mathrm{e}_a^\top \mathrm{C}\boldsymbol{B}\Sigma_L\boldsymbol{B}^\top \mathrm{C}^\top \mathrm{e}_a > 0$ and, of course, $\mathbb{E}[(\mathrm{e}_a^\top\epsilon(t,h))^2]/h^2$ converges then to infinity. ∎

Proof of Theorem 4.8. Based on Theorem 4.2, the proofs of the two orthogonal projections differ only in the choice of $\boldsymbol{M}(\cdot)$ or $\mathbf{M}(\cdot)$, $\boldsymbol{M}_m(\cdot)$ or $\mathbf{M}_m(\cdot)$, and $\varepsilon(\cdot,\cdot)$ or $\epsilon(\cdot,\cdot)$. Thus, we only prove the representation of $P_{\mathcal{L}_{Y_S}(t)}Y_a(t+h)$. Let $h \ge 0$, $t\in\mathbb{R}$, $S\subseteq V$, and $a\in V$. From Theorem 4.2recall that $\mathbb{P}$-a.s.

$$Y_a(t+h) = \int_{-\infty}^t \mathrm{e}_a^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta}Y(u)\mathrm{d}u + \sum_{m=0}^{p-q-1}\mathrm{e}_a^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta}D^{(m)}Y(t) + \mathrm{e}_a^\top \epsilon(t,h).$$

We calculate the projections of the summands separately. For the first summand we get

$$\begin{aligned} &P_{\mathcal{L}_{Y_S}(t)}\left(\int_{-\infty}^t \mathrm{e}_a^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta}Y(u)\mathrm{d}u\right)\\ &= \sum_{v\in S}\int_{-\infty}^t \mathrm{e}_a^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta}e_vY_v(u)\mathrm{d}u + P_{\mathcal{L}_{Y_S}(t)}\left(\sum_{v\in V\setminus S}\int_{-\infty}^t \mathrm{e}_a^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta}e_vY_v(u)\mathrm{d}u\right), \end{aligned}$$

since, because of Proposition 4.5, the integrals are in $\mathcal{L}_{Y_S}(t)$ for $v\in S$. For the second summand, we obtain

$$\begin{aligned} &P_{\mathcal{L}_{Y_S}(t)}\left(\sum_{m=0}^{p-q-1}\mathrm{e}_a^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta}D^{(m)}Y(t)\right)\\ &= \sum_{v\in S}\sum_{m=0}^{p-q-1}\mathrm{e}_a^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta}e_vD^{(m)}Y_v(t) + P_{\mathcal{L}_{Y_S}(t)}\left(\sum_{v\in V\setminus S}\sum_{m=0}^{p-q-1}\mathrm{e}_a^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta}\mathrm{e}_vD^{(m)}Y_v(t)\right), \end{aligned}$$

since, due to Remark 3.15, the derivatives of $Y_v(t)$ for $v\in S$ are in $\mathcal{L}_{Y_S}(t)$.

For the third summand $\mathrm{e}_a^\top\epsilon(t,h)$ we note that $(Y_S(s))_{s\le t}$ and $(L(s)-L(t))_{t\le s\le t+h}$ are independent. We obtain immediately that $P_{\mathcal{L}_{Y_S}(t)}\mathrm{e}_a^\top\epsilon(t,h) = 0$. If we put all three summands together, we get the assertion. ∎

Proof of Theorem 4.10. Let $S\subseteq V$, $a\in V$, $h\ge 0$, and $t\in\mathbb{R}$. First of all, due to Theorem 4.8 and similar ideas as in (6.6),

$$P_{\mathcal{L}_Y(t)}D^{(p-q-1)}Y_a(t+h) = \int_{-\infty}^t \mathrm{e}_a^\top \mathrm{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta}Y(u)\mathrm{d}u + \sum_{m=0}^{p-q-1}\mathrm{e}_a^\top \mathrm{M}_m(h)\boldsymbol{\Theta}D^{(m)}Y(t) = \mathrm{e}_a^\top \mathrm{C}\mathrm{e}^{\boldsymbol{A}h}X(t).$$

Then, due to

$$\lim_{h\to 0}\mathbb{E}\left[\left(P_{\mathcal{L}_Y(t)}\left(\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h}\right) - \mathrm{e}_a^\top \mathbf{C}\boldsymbol{A}X(t)\right)^2\right]$$

DOI: 10.1111/jtsa.12787 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

$$= \lim_{h\to 0} \mathbb{E}\left[\left(\mathrm{e}_a^\top \mathbf{C} \frac{\mathrm{e}^{Ah} - I_{kp}}{h} X(t) - \mathrm{e}_a^\top \mathbf{C} A X(t)\right)^2\right]$$

$$= \lim_{h\to 0} \mathrm{e}_a^\top \mathbf{C}\left(\frac{\mathrm{e}^{Ah} - I_{kp}}{h} - A\right) c_{XX}(0) \left(\frac{\mathrm{e}^{Ah} - I_{kp}}{h} - A\right)^\top \mathbf{C}^\top \mathrm{e}_a = 0,$$

we obtain

$$\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_Y(t)}\left(\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h}\right) = \mathrm{e}_a^\top \mathbf{C} A X(t) \quad \mathbb{P}\text{-a.s.}$$

Together with Brockwell and Davis (1991), Proposition 2.3.2.(iv, vii), it follows that

$$\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_S}(t)}\left(\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h}\right)$$

$$= \underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_S}(t)} P_{\mathcal{L}_Y(t)}\left(\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h}\right)$$

$$= P_{\mathcal{L}_{Y_S}(t)}\left(\mathrm{e}_a^\top \mathbf{C} A X(t)\right) \quad \mathbb{P}\text{-a.s.}$$

Again, similar to the proof of (6.6),

$$\mathrm{e}_a^\top \mathbf{C} A X(t) = \int_{-\infty}^t \mathrm{e}_a^\top \mathbf{M}'(0) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y(u) \mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \mathbf{M}_m'(0) \mathbf{\Theta} D^{(m)} Y(t) \quad \mathbb{P}\text{-a.s.}$$

We obtain replacing $\mathbf{M}(h)$ by $\mathbf{M}'(0)$ and $\mathbf{M}_m(h)$ by $\mathbf{M}_m'(0)$ in the proof of Theorem 4.8,

$$\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_S}(t)}\left(\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h}\right)$$

$$= \sum_{v\in S} \int_{-\infty}^t \mathrm{e}_a^\top \mathbf{M}'(0) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u + \sum_{v\in S} \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \mathbf{M}_m'(0) \mathbf{\Theta} e_v D^{(m)} Y_v(t)$$

$$+ P_{\mathcal{L}_{Y_S}(t)}\left(\sum_{v\in V\setminus S} \int_{-\infty}^t \mathrm{e}_a^\top \mathbf{M}'(0) \mathrm{e}^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) \mathrm{d}u\right)$$

$$+ P_{\mathcal{L}_{Y_S}(t)}\left(\sum_{v\in V\setminus S} \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \mathbf{M}_m'(0) \mathbf{\Theta} e_v D^{(m)} Y_v(t)\right) \quad \mathbb{P}\text{-a.s.}$$

as claimed.

The second assertion follows directly from Theorem 4.2 and Theorem 4.8, which give

$$D^{(p-q-1)}Y_a(t+h) - P_{\mathcal{L}_Y(t)} D^{(p-q-1)}Y_a(t+h)$$

$$
\begin{aligned}
&= \int_{-\infty}^{t} \mathrm{e}_a^\top \mathbf{M}(h) \mathrm{e}^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} Y(u)\mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \mathbf{M}_m(h)^\top \boldsymbol{\Theta} D^{(m)} Y(t) + \mathrm{e}_a^\top \epsilon(t,h) \\
&\quad - \int_{-\infty}^{t} \mathrm{e}_a^\top \mathbf{M}(h) \mathrm{e}^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} Y(u)\mathrm{d}u - \sum_{m=0}^{p-q-1} \mathrm{e}_a^\top \mathbf{M}_m(h) \boldsymbol{\Theta} D^{(m)} Y(t) \\
&= \mathrm{e}_a^\top \epsilon(t,h).
\end{aligned}
$$

■

6.3. Proofs of Section 5

Proof of Theorem 5.1.

(A.2) The proof of Assumption (A.2) is elaborate and has already been given in Fasen-Hartmann and Schenk (2024a), Proposition 6.5, for MCAR(p) processes. It can be directly generalised to ICCSS(p, q) processes, so we do not give the full proof. We simply note that we only require that $Q(i\lambda)P(i\lambda)^{-1}$ has full rank and $\Sigma_L > 0$ to obtain that $f_{YY}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Indeed, Assumption (3.12) provides that $Q(i\lambda)$ is of full rank and $\mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}$, so we directly receive that $Q(i\lambda)P(i\lambda)^{-1}$ has full rank as well. Furthermore, we require that $\sigma(\boldsymbol{A}) \subseteq (-\infty, 0) + i\mathbb{R}$, but this is also true due to Assumption (3.12). Finally, it is a necessity that $\boldsymbol{C}\boldsymbol{B}\Sigma_L\boldsymbol{B}^\top\boldsymbol{C}^\top > 0$. Again, $\Sigma_L > 0$, $\boldsymbol{C}$ is of full rank by Assumption (3.12), and $\boldsymbol{B}$ is of full rank by definition, so $\boldsymbol{C}\boldsymbol{B}\Sigma_L\boldsymbol{B}^\top\boldsymbol{C}^\top > 0$.

(A.3) For Assumption (A.3) we apply that $\sigma(\boldsymbol{A}) \subseteq (-\infty, 0) + i\mathbb{R}$ and hence

$$
\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_X(t)} X(t+h) = \underset{h\to 0}{\text{l.i.m.}}\, \mathrm{e}^{\boldsymbol{A}h} X(t) = 0,
$$

resulting in X being purely non-deterministic. By Rozanov (1967), III, (2.1) and Theorem 2.1 this is equivalent to $\bigcap_{t\in\mathbb{R}} \mathcal{L}_X(t) = \{0\}$. Since $\bigcap_{t\in\mathbb{R}} \mathcal{L}_Y(t) \subseteq \bigcap_{t\in\mathbb{R}} \mathcal{L}_X(t)$ the process Y is purely non-deterministic as well.

Finally, the Markov properties follow from Fasen-Hartmann and Schenk (2024a), Section 5; see also Fasen-Hartmann and Schenk (2024a), Propositions 6.6 and 6.7 for MCAR(p) processes. ■

Next, we prove Proposition 5.4, since the proof of Theorem 5.3 is based on Proposition 5.4.

Proof of Proposition 5.4. (a) Recall that due to Definition 3.1 we have no directed edge $a \to b \notin E_{\mathrm{OG}}$, if and only if, for $0 \le h \le 1$ and $t \in \mathbb{R}$,

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

From Theorem 4.8 we obtain for $0 \le h \le 1$ and $t \in \mathbb{R}$,

$$
P_{\mathcal{L}_Y(t)} Y_b(t+h) = \sum_{v\in V} \int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h) \mathrm{e}^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} e_v Y_v(u)\mathrm{d}u + \sum_{v\in V} \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h) \boldsymbol{\Theta} e_v D^{(m)} Y_v(t),
$$

$$
\begin{aligned}
P_{\mathcal{L}_{Y_{V\setminus\{a\}}}(t)} Y_b(t+h) &= \sum_{v\in V\setminus\{a\}} \int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h) \mathrm{e}^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} e_v Y_v(u)\mathrm{d}u + \sum_{v\in V\setminus\{a\}} \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h) \boldsymbol{\Theta} e_v D^{(m)} Y_v(t) \\
&\quad + P_{\mathcal{L}_{Y_{V\setminus\{a\}}}(t)} \left(\int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h) \mathrm{e}^{\boldsymbol{\Lambda}(t-u)} \boldsymbol{\Theta} e_a Y_a(u)\mathrm{d}u \right) + P_{\mathcal{L}_{Y_{V\setminus\{a\}}}(t)} \left(\sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h) \boldsymbol{\Theta} e_a D^{(m)} Y_a(t) \right) \mathbb{P}\text{-a.s.}
\end{aligned}
$$

We equate the two orthogonal projections and remove the coinciding terms. Then we receive that $a \to b \notin E_{\mathrm{OG}}$, if and only if, for $0 \le h \le 1$ and $t \in \mathbb{R}$,

$$\int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} e_a Y_a(u)\mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta} e_a D^{(m)} Y_a(t)$$
$$= P_{\mathcal{L}_{Y_{V\setminus\{a\}}}(t)}\left(\int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} e_a Y_a(u)\mathrm{d}u\right) + P_{\mathcal{L}_{Y_{V\setminus\{a\}}}(t)}\left(\sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta} e_a D^{(m)} Y_a(t)\right)\mathbb{P}\text{-a.s.}$$

The expression on the left-hand side is in $\mathcal{L}_{Y_a}(t)$ and the expression on the right side is in $\mathcal{L}_{Y_{V\setminus\{a\}}}(t)$. Since $\mathcal{L}_{Y_{V\setminus\{a\}}}(t) \cap \mathcal{L}_{Y_a}(t) = \{0\}$ due to (3.1), $a \to b \notin E_{\mathrm{OG}}$, if and only if, for $0 \le h \le 1$ and $t \in \mathbb{R}$,

$$\int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} e_a Y_a(u)\mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta} e_a D^{(m)} Y_a(t) = 0 \quad \mathbb{P}\text{-a.s.} \tag{6.11}$$

In the following, we show that this characterisation is in turn equivalent to

$$\mathrm{e}_b^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}t}\boldsymbol{\Theta} e_a = 0 \quad \text{and} \quad \mathrm{e}_b^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta} e_a = 0, \tag{6.12}$$

for $m = 0, \ldots, p-q-1$, $0 \le h \le 1$, and $t \ge 0$.

If (6.12) holds, we immediately obtain that (6.11) is valid. Now, suppose (6.11) holds. We convert the two summands in (6.11) into their spectral representation. Hence, note that due Bernstein (2009), Proposition 11.2.2, and $\sigma(\boldsymbol{\Lambda}) \subseteq (-\infty, 0) + i\mathbb{R}$ the equality

$$\int_{-\infty}^{\infty} \mathrm{e}^{-i\lambda s}\mathbf{1}_{\{s\ge 0\}}\mathrm{e}_b^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}s}\boldsymbol{\Theta} e_a ds = \mathrm{e}_b^\top \boldsymbol{M}(h)(i\lambda I_{kq} - \boldsymbol{\Lambda})^{-1}\boldsymbol{\Theta} e_a, \quad \lambda \in \mathbb{R},$$

holds. Now Rozanov (1967) I, Example 8.3, provides the spectral representation of the first summand

$$\int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} e_a Y_a(u)\mathrm{d}u = \int_{-\infty}^{\infty} \mathrm{e}^{i\lambda t}\mathrm{e}_b^\top \boldsymbol{M}(h)(i\lambda I_{kq} - \boldsymbol{\Lambda})^{-1}\boldsymbol{\Theta} e_a \Phi_a(\mathrm{d}\lambda),$$

where $\Phi_a(\cdot)$ is the random spectral measure from the spectral representation of Y_a. For the second summand, we substitute $Y_a(t)$ as well as its derivatives (cf. Fasen-Hartmann and Schenk, 2024a, Proposition 2.4) by their spectral representation. We obtain

$$0 = \int_{-\infty}^{t} \mathrm{e}_b^\top \boldsymbol{M}(h)\mathrm{e}^{\boldsymbol{\Lambda}(t-u)}\boldsymbol{\Theta} e_a Y_a(u)\mathrm{d}u + \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta} e_a D^{(m)} Y_a(t)$$
$$= \int_{-\infty}^{\infty} \mathrm{e}^{i\lambda t}\mathrm{e}_b^\top \boldsymbol{M}(h)(i\lambda I_{kq} - \boldsymbol{\Lambda})^{-1}\boldsymbol{\Theta} e_a \Phi_a(\mathrm{d}\lambda) + \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta} e_a \int_{-\infty}^{\infty} (i\lambda)^m \mathrm{e}^{i\lambda t}\Phi_a(\mathrm{d}\lambda).$$

Denoting $\psi(\lambda, h) = \mathrm{e}_b^\top \boldsymbol{M}(h)(i\lambda I_{kq} - \boldsymbol{\Lambda})^{-1}\boldsymbol{\Theta} e_a + \sum_{m=0}^{p-q-1} \mathrm{e}_b^\top \boldsymbol{M}_m(h)\boldsymbol{\Theta} e_a (i\lambda)^m$,for $\lambda \in \mathbb{R}$ and $0 \le h \le 1$, it follows that

$$0 = \mathbb{E}\left[\left|\int_{-\infty}^{\infty} \mathrm{e}^{i\lambda t}\psi(\lambda, h)\Phi_a(\mathrm{d}\lambda)\right|^2\right] = \int_{-\infty}^{\infty} |\psi(\lambda, h)|^2 f_{Y_a Y_a}(\lambda)\mathrm{d}\lambda,$$

and therefore $|\psi(\lambda,h)|^2 f_{Y_aY_a}(\lambda)=0$ for (almost) all $\lambda\in\mathbb{R}$. But by Theorem 5.1, $f_{Y_aY_a}(\lambda)>0$ for all $\lambda\in\mathbb{R}$, which yields $\psi(\lambda,h)=0$ for $0\le h\le 1$ and (almost) all $\lambda\in\mathbb{R}$. Bernstein (2009), (4.23), provides due to $i\lambda\in\mathbb{C}\setminus\sigma(\mathbf{A})$ that

$$(i\lambda I_{kq}-\mathbf{\Lambda})^{-1}=\frac{1}{\chi_{\mathbf{\Lambda}}(i\lambda)}\sum_{j=0}^{kq-1}(i\lambda)^j\Delta_j,$$

where $\Delta_j\in M_{kq}(\mathbb{R})$, $\Delta_{kq-1}=I_{kq}$, and $\chi_{\mathbf{\Lambda}}(z)=z^{kq}+\gamma_{kq-1}z^{kq-1}+\cdots+\gamma_1 z+\gamma_0, z\in\mathbb{C}$, is the characteristic polynomial of $\mathbf{\Lambda}$ with $\gamma_1,\ldots,\gamma_{kq-1}\in\mathbb{R}$, $\gamma_{kq}=1$, cf. Bernstein (2009), (4.4.3). Thus,

$$0=\psi(\lambda,h)=\frac{1}{\chi_{\mathbf{\Lambda}}(i\lambda)}\sum_{j=0}^{kq-1}(i\lambda)^j\mathrm{e}_b^\top\boldsymbol{M}(h)\Delta_j\mathbf{\Theta}e_a+\sum_{m=0}^{p-q-1}\mathrm{e}_b^\top\boldsymbol{M}_m(h)\mathbf{\Theta}e_a(i\lambda)^m,$$

and multiplication by the characteristic polynomial yields

$$0=\sum_{j=0}^{kq-1}(i\lambda)^j\mathrm{e}_b^\top\boldsymbol{M}(h)\Delta_j\mathbf{\Theta}e_a+\sum_{m=0}^{p-q-1}\sum_{\ell=0}^{kq}\mathrm{e}_b^\top\boldsymbol{M}_m(h)\mathbf{\Theta}e_a\gamma_\ell(i\lambda)^{\ell+m}.$$

In the first sum there are powers up to $kq-1$, while in the second sum there are powers up to $kq-1+p-q$. For $\ell=kq$ and $m=0,\ldots,p-q-1$ we receive in the second summand powers higher than kq and their prefactors have to be zero. Due to $\gamma_{kp}=1$we receive then for $m=0,\ldots,p-q-1$,

$$\mathrm{e}_b^\top\boldsymbol{M}_m(h)\mathbf{\Theta}e_a=0.$$

Inserting this result into $\psi(\lambda,h)=0$ yields

$$0=\mathrm{e}_b^\top\boldsymbol{M}(h)(i\lambda I_{kq}-\mathbf{\Lambda})^{-1}\mathbf{\Theta}e_a=\int_{-\infty}^{\infty}\mathrm{e}^{-i\lambda s}\mathbf{1}_{\{s\ge 0\}}\mathrm{e}_b^\top\boldsymbol{M}(h)\mathrm{e}^{\mathbf{\Lambda}s}\mathbf{\Theta}e_a\mathrm{d}s.$$

Together with the already known integrability, Pinsky (2009), Corollary 2.2.23, provides

$$\mathrm{e}_b^\top\boldsymbol{M}(h)\mathrm{e}^{\mathbf{\Lambda}t}\mathbf{\Theta}e_a=0,\quad t\ge 0,$$

which finally concludes the proof of (a).

(b) Due to the similarity of the results in Theorem 4.8 and Theorem 4.10, we just have to replace $\mathbf{M}(h)$ by $\mathbf{M}'(0)$ and $\mathbf{M}_m(h)$ by $\mathbf{M}'_m(0)$ in the proof of (a). ∎

Proof of Theorem 5.3. (a) Based on the characterisations in Proposition 5.4(a), the same ideas as in the proof of Theorem 6.19(a) in Fasen-Hartmann and Schenk (2024a) can be carried out and therefore, the proof is omitted. First, we replace the matrix exponential $\mathrm{e}^{\boldsymbol{A}h}$ in Proposition 5.4(a) by powers of the matrix $\boldsymbol{A}$ and second, we replace $\mathrm{e}^{\mathbf{\Lambda}h}$ by powers of $\mathbf{\Lambda}$.

(b) Follows in analogy to (a) using Proposition 5.4(b). ∎

Proof of Proposition 5.6. (a) Based on Corollary 4.11(a), the proof of the first characterisation in (a) can be done in the same way as the proof of Fasen-Hartmann and Schenk (2024a), proposition 6.13(a). The second characterisation in (a) follows along the lines of the proof of Fasen-Hartmann and Schenk (2024a), Theorem 6.19(b).

(b) Based on Theorem 4.2 and Corollary 4.11(b), statement (b) can be proven analogously to Fasen-Hartmann and Schenk (2024a), Proposition 6.13(b). ∎

7. CONCLUSION

In this article, we have applied the concept of (local) orthogonality graphs to state space models. For the state space models, we have assumed that they have a representation in controller canonical form satisfying the mild assumptions of (3.12) such that there exists a stationary invertible version of the state space model; the term invertible reflects that we are able to recover the state process from the observation process. These assumptions have been summarised under the acronym ICCSS(p, q) model with $p > q > 0$. The ICCSS processes satisfy the assumptions of the (local) orthogonality graphs defined in Fasen-Hartmann and Schenk (2024a) so that the graphical models are well-defined and several notions of causal Markov properties hold. However, the invertibility of the state process and the representation of the state space model in controller canonical form are not necessary for the existence of (local) orthogonality graphs. The orthogonality graphs exist for a much broader class of state space models, but for the analytic representations of the edges, these additional assumptions are useful. The characterisations of the edges of the ICCSS process require the knowledge of the orthogonal projections of the state process onto linear subspaces generated by subprocesses, and for the derivation of these orthogonal projections the invertibility of the state process is important. The orthogonal projections depend on the model parameters of the controller canonical form and therefore, the edges of the (local) orthogonality graph are also uniquely characterised by these model parameters.

ACKNOWLEDGEMENTS

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 programme is funded by the Federal Ministry of Education and Research. The authors are responsible for the content of this publication. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

REFERENCES

Basse-O’Connor A, Nielsen MS, Pedersen J, Rohde V. 2019. Multivariate stochastic delay differential equations and CAR representations of CARMA processes. *Stochastic Processes and their Applications* **129**:4119–4143. https://doi.org/10.1016/j.spa.2018.11.011.

Basu S, Shojaie A, Michailidis G. 2015. Network Granger causality with inherent grouping structure. *Journal of Machine Learning Research* **16**:417–453. https://doi.org/10.5555/2789272.2789285.

Bernstein DS. 2009. *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed. Princeton University Press, Princeton, NJ. https://doi.org/10.1515/9781400833344.

Brockwell P. 2014. Recent results in the theory and applications of CARMA processes. *Annals of the Institute of Statistical Mathematics* **66**:647–685. https://doi.org/10.1007/s10463-014-0468-7.

Brockwell P, Davis R. 1991. *Time Series: Theory and Methods*, 2nd ed. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-0320-4.

Brockwell P, Lindner A. 2015. Prediction of Lévy-driven CARMA processes. *Journal of Econometrics* **189**:263–271. https://doi.org/10.1016/j.jeconom.2015.03.021.

Brockwell P, Schlemm E. 2013. Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. *Journal of Multivariate Analysis* **115**:217–251. https://doi.org/10.1016/j.jmva.2012.09.004.

Comte F, Renault E. 1996. Noncausality in continuous time models. *Economic Theory* **12**:215–256. https://doi.org/10.1017/S0266466600006575.

Cramér H. 1940. On the theory of stationary random processes. *Annals of Mathematics* **41**:215–230. https://doi.org/10.2307/1968827.

Didelez V. 2007. Graphical models for composable finite Markov processes. *Scandinavian Journal of Statistics* **34**:169–185. https://doi.org/10.1111/j.1467-9469.2006.00528.x.

Didelez V. 2008. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society, Series B* **70**:245–264. https://doi.org/10.1111/j.1467-9868.2007.00634.x.

Eichler M. 2007. Granger causality and path diagrams for multivariate time series. *Journal of Econometrics* **137**:334–353. https://doi.org/10.1016/j.jeconom.2005.06.032.

Eichler M. 2012. Graphical modelling of multivariate time series. *Probability Theory and Related Fields* **153**:233–268. https://doi.org/10.1007/s00440-011-0345-8.

Eichler M, Dahlhaus R, Dueck J. 2017. Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *Journal of Time Series Analysis* **38**:225–242. https://doi.org/10.1111/jtsa.12213.

Fasen-Hartmann V, Schenk L. 2024a. Mixed orthogonality graphs for continuous-time stationary processes. *Stochastic Processes and their Applications*, to appear. https://doi.org/10.1016/j.spa.2024.10.4501X.

Fasen-Hartmann V, Schenk L. 2024b. Partial correlation graphs for continuous-parameter time series. https://doi.org/10.48550/arXiv.2401.16970.

Gladyshev EG. 1958. On multi-dimensional stationary random processes. *Theory of Probability and its Applications* **3**:425–428. https://doi.org/10.1137/1103035.

Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**:424–438. https://doi.org/10.2307/1912791.

Hannan EJ, Deistler M. 2012. *The Statistical Theory of Linear Systems* Society for Industrial and Applied Mathematics, Philadelphia, PA. https://doi.org/10.1137/1.9781611972191.

Kailath T. 1980. *Linear Systems* Prentice-Hall, Englewood Cliffs, NJ.

Lax PD. 2002. *Functional Analysis* Wiley, New York, NY.

Maathuis M, Drton M, Lauritzen SL, Wainwright M. 2019. *Handbook of Graphical Models* CRC Press, Boca Raton, FL. https://doi.org/10.1201/9780429463976.

Marquardt T, Stelzer R. 2007. Multivariate CARMA processes. *Stochastic Processes and their Applications* **117**:96–120. https://doi.org/10.1016/j.spa.2006.05.014.

Mogensen SW, Hansen NR. 2020. Markov equivalence of marginalized local independence graphs. *Annals of Mathematical Statistics* **48**:539–559. https://doi.org/10.1214/19-AOS1821.

Mogensen SW, Hansen NR. 2022. Graphical modeling of stochastic processes driven by correlated noise. *Bernoulli* **28**:3023–3050. https://doi.org/10.3150/21-bej1446.

Pinsky MA. 2009. *Introduction to Fourier Analysis and Wavelets* American Mathematical Society, Providence, RI.

Rozanov JA. 1967. *Stationary Random Processes* Holden-Day, San Francisco, CA.

Rugh WJ. 1996. *Linear System Theory*, 2nd ed. Prentice Hall, Upper Saddle River, NJ.

Sato K. 2007. *Lévy Processes and Infinitely Divisible Distributions*, 2nd ed. Cambridge University Press, Cambridge.

Sato K, Yamazato M. 1984. Operator-self-decomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stochastic Processes and their Applications* **17**:73–100. https://doi.org/10.1016/0304-4149(84)90312-0.

Schlemm E, Stelzer R. 2012a. Multivariate CARMA processes, continuous-time state space models and complete regularity of the innovations of the sampled processes. *Bernoulli* **18**:46–63. https://doi.org/10.3150/10-BEJ329.

Schlemm E, Stelzer R. 2012b. Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electronic Journal of Statistics* **6**:2185–2234. https://doi.org/10.1214/12-EJS743.

Shojaie A, Fox EB. 2022. Granger causality: a review and recent advances. *Annual Review of Statistics and Its Application* **9**:289–319. https://doi.org/10.1146/annurev-statistics-040120-010930.

Sims CA. 1972. Money, income, and causality. *The American Economic Review* **62**:540–552.

Wright S. 1921. Correlation and causation. *Journal of Agricultural Research* **20**:557–585.

Wright S. 1934. The method of path coefficients. *Annals of Mathematical Statistics* **5**:161–215. https://doi.org/10.1214/aoms/1177732676.